



*agronomy*

# Molecular Marker Technology for Crop Improvement

---

Edited by

José Miguel Soriano

Printed Edition of the Special Issue Published in *Agronomy*

# **Molecular Marker Technology for Crop Improvement**



# Molecular Marker Technology for Crop Improvement

Editor

**José Miguel Soriano**

MDPI • Basel • Beijing • Wuhan • Barcelona • Belgrade • Manchester • Tokyo • Cluj • Tianjin



*Editor*

José Miguel Soriano

Institute for Food and Agricultural Research and Technology (IRTA)

Spain

*Editorial Office*

MDPI

St. Alban-Anlage 66

4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *Agronomy* (ISSN 2073-4395) (available at: <https://www.mdpi.com/journal/agronomy/special-issues/molecular-marker>).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

LastName, A.A.; LastName, B.B.; LastName, C.C. Article Title. <i>Journal Name</i> <b>Year</b> , Volume Number, Page Range.
--

**ISBN 978-3-03943-863-1 (Hbk)**

**ISBN 978-3-03943-864-8 (PDF)**

Cover image courtesy of Sustainable Field Crops Programme (IRTA).

© 2020 by the authors. Articles in this book are Open Access and distributed under the Creative Commons Attribution (CC BY) license, which allows users to download, copy and build upon published articles, as long as the author and publisher are properly credited, which ensures maximum dissemination and a wider impact of our publications.

The book as a whole is distributed by MDPI under the terms and conditions of the Creative Commons license CC BY-NC-ND.

# Contents

About the Editor . . . . .	vii
<b>Jose Miguel Soriano</b> Molecular Marker Technology for Crop Improvement Reprinted from: <i>Agronomy</i> 2020, 10, 1462, doi:10.3390/agronomy10101462 . . . . .	1
<b>Martina Roselló, Conxita Royo, Miguel Sanchez-Garcia and Jose Miguel Soriano</b> Genetic Dissection of the Seminal Root System Architecture in Mediterranean Durum Wheat Landraces by Genome-Wide Association Study Reprinted from: <i>Agronomy</i> 2019, 9, 364, doi:10.3390/agronomy9070364 . . . . .	7
<b>Jiantao Wu, Qinnan Wang, Jing Xie, Yong-Bao Pan, Feng Zhou, Yuqiang Guo, Hailong Chang, Huanying Xu, Wei Zhang, Chuiming Zhang and Yongsheng Qiu</b> SSR Marker-Assisted Management of Parental Germplasm in Sugarcane ( <i>Saccharum</i> spp. hybrids) Breeding Programs Reprinted from: <i>Agronomy</i> 2019, 9, 449, doi:10.3390/agronomy9080449 . . . . .	25
<b>Yu Ge, Xiaoping Zang, Lin Tan, Jiashui Wang, Yuanzheng Liu, Yanxia Li, Nan Wang, Di Chen, Rulin Zhan and Weihong Ma</b> Single-Molecule Long-Read Sequencing of Avocado Generates Microsatellite Markers for Analyzing the Genetic Diversity in Avocado Germplasm Reprinted from: <i>Agronomy</i> 2019, 9, 512, doi:10.3390/agronomy9090512 . . . . .	41
<b>Vadim G. Lebedev, Natalya M. Subbotina, Oleg P. Maluchenko, Konstantin V. Krutovsky and Konstantin A. Shestibratov</b> Assessment of Genetic Diversity in Differently Colored Raspberry Cultivars Using SSR Markers Located in Flavonoid Biosynthesis Genes Reprinted from: <i>Agronomy</i> 2019, 9, 518, doi:10.3390/agronomy9090518 . . . . .	59
<b>Carlos Maldonado, Freddy Mora, Filipe Augusto Bengosi Bertagna, Maurício Carlos Kuki and Carlos Alberto Scapim</b> SNP- and Haplotype-Based GWAS of Flowering-Related Traits in Maize with Network-Assisted Gene Prioritization Reprinted from: <i>Agronomy</i> 2019, 9, 725, doi:10.3390/agronomy9110725 . . . . .	73
<b>Hui Zhang, Yu-Jun Zhu, An-Dong Zhu, Ye-Yang Fan, Ting-Xu Huang, Jian-Fu Zhang, Hua-An Xie and Jie-Yun Zhuang</b> Identification and Verification of Quantitative Trait Loci Affecting Milling Yield of Rice Reprinted from: <i>Agronomy</i> 2020, 10, 75, doi:10.3390/agronomy10010075 . . . . .	83
<b>Xiaoxia Yu, Mingfei Zhang, Zhuo Yu, Dongsheng Yang, Jingwei Li, Guofang Wu and Jiaqi Li</b> An SNP-Based High-Density Genetic Linkage Map for Tetraploid Potato Using Specific Length Amplified Fragment Sequencing (SLAF-Seq) Technology Reprinted from: <i>Agronomy</i> 2020, 10, 114, doi:10.3390/agronomy10010114 . . . . .	97
<b>Rosa Mérida-García, Alison R. Bentley, Sergio Gálvez, Gabriel Dorado, Ignacio Solís, Karim Ammar and Pilar Hernandez</b> Mapping Agronomic and Quality Traits in Elite Durum Wheat Lines under Differing Water Regimes Reprinted from: <i>Agronomy</i> 2020, 10, 144, doi:10.3390/agronomy10010144 . . . . .	111

<b>Rubén Rufo, Silvio Salvi, Conxita Royo and Jose Miguel Soriano</b> Exploring the Genetic Architecture of Root-Related Traits in Mediterranean Bread Wheat Landraces by Genome-Wide Association Analysis Reprinted from: <i>Agronomy</i> <b>2020</b> , <i>10</i> , 613, doi:10.3390/agronomy10050613 . . . . .	135
<b>Korachan Thanasilungura, Sukanya Kranto, Tidarat Monkham, Sompong Chankaew and Jirawat Sanitchon</b> Improvement of a RD6 Rice Variety for Blast Resistance and Salt Tolerance through Marker-Assisted Backcrossing Reprinted from: <i>Agronomy</i> <b>2020</b> , <i>10</i> , 1118, doi:10.3390/agronomy10081118 . . . . .	155
<b>Baffour Badu-Apraku, Samuel Adewale, Agre Paterne, Melaku Gedil and Robert Asiedu</b> Identification of QTLs Controlling Resistance/Tolerance to <i>Striga hermonthica</i> in an Extra-Early Maturing Yellow Maize Population Reprinted from: <i>Agronomy</i> <b>2020</b> , <i>10</i> , 1168, doi:10.3390/agronomy10081168 . . . . .	173
<b>Manuel Blasco, Francisco Gil-Muñoz, María del Mar Naval and María Luisa Badenes</b> Molecular Assisted Selection for Pollination-Constant and Non-Astringent Type without Male Flowers in Spanish Germplasm for Persimmon Breeding Reprinted from: <i>Agronomy</i> <b>2020</b> , <i>10</i> , 1172, doi:10.3390/agronomy10081172 . . . . .	191
<b>Sirjan Sapkota, J. Lucas Boatwright, Kathleen Jordan, Richard Boyles and Stephen Kresovich</b> Multi-Trait Regressor Stacking Increased Genomic Prediction Accuracy of Sorghum Grain Composition Reprinted from: <i>Agronomy</i> <b>2020</b> , <i>10</i> , 1221, doi:10.3390/agronomy10091221 . . . . .	203
<b>Yuliya Genievskaya, Saule Abugalieva, Aralbek Rsaliyev, Gulbahar Yskakova and Yerlan Turuspekov</b> QTL Mapping for Seedling and Adult Plant Resistance to Leaf and Stem Rusts in Pamyati Azieva × Paragon Mapping Population of Bread Wheat Reprinted from: <i>Agronomy</i> <b>2020</b> , <i>10</i> , 1285, doi:10.3390/agronomy10091285 . . . . .	219
<b>Ángela Polo-Oltra, Carlos Romero, Inmaculada López, María Luisa Badenes and Elena Zuriaga</b> Cost-Effective and Time-Efficient Molecular Assisted Selection for Ppv Resistance in Apricot Based on <i>ParPMC2</i> Allele-Specific PCR Reprinted from: <i>Agronomy</i> <b>2020</b> , <i>10</i> , 1292, doi:10.3390/agronomy10091292 . . . . .	237
<b>Rosa Mérida-García, Sergio Gálvez, Etienne Paux, Gabriel Dorado, Laura Pascual, Patricia Giraldo and Pilar Hernandez</b> High Resolution Melting and Insertion Site-Based Polymorphism Markers for Wheat Variability Analysis and Candidate Genes Selection at Drought and Heat MQTL Loci Reprinted from: <i>Agronomy</i> <b>2020</b> , <i>10</i> , 1294, doi:10.3390/agronomy10091294 . . . . .	263

## About the Editor

**José Miguel Soriano** is a researcher in the Sustainable Field Crops Programme at IRTA. He has a PhD in Molecular and Evolutionary Genetics. His research has been focused on the development and identification of molecular markers linked to traits of interest in crop plants, the construction of linkage maps, and the use of genomic technologies in plant breeding. His early research focused on fruit breeding to incorporate pathogen resistance assisted by molecular techniques, and in recent years his focus has shifted to cereal research for drought adaptation in Mediterranean environments.





# Molecular Marker Technology for Crop Improvement

Jose Miguel Soriano

Sustainable Field Crops Programme, IRTA (Institute for Food and Agricultural Research and Technology),  
25198 Lleida, Spain; josemiguel.soriano@irta.cat

Received: 3 September 2020; Accepted: 23 September 2020; Published: 24 September 2020

**Abstract:** Since the 1980s, agriculture and plant breeding have changed with the development of molecular marker technology. In recent decades, different types of molecular markers have been used for different purposes: mapping, marker-assisted selection, characterization of genetic resources, etc. These have produced effective genotyping, but the results have been costly and time-consuming, due to the small number of markers that could be tested simultaneously. Recent advances in molecular marker technologies such as the development of high-throughput genotyping platforms, genotyping by sequencing, and the release of the genome sequences of major crop plants open new possibilities for advancing crop improvement. This Special Issue collects sixteen research studies, including the application of molecular markers in eleven crop species, from the generation of linkage maps and diversity studies to the application of marker-assisted selection and genomic prediction.

**Keywords:** crop breeding; genetic maps; QTL mapping; GWAS; marker assisted selection; genomic selection; DNA sequencing

---

## 1. Introduction

Classical breeding was the main approach used by breeders to increase crop productivity during the 20th century. It implies the selection of cultivars with the desired characteristics for the target trait, usually morphological or visual characteristics. The best genotypes were selected and used as parents in a backcross scheme with a recurrent parent to dilute the irrelevant or undesired traits [1]. However, the long time to get a commercial cultivar and limitations related to traits highly dependent on the environment or with low heritability make necessary the use of complementary approaches to assist the breeding process. The development of molecular biology made possible the appearance of a new type of marker based on polymorphisms in the DNA sequence, the molecular markers, which broaden the possibilities for new challenges in plant breeding. Molecular markers are widely distributed in the genome, they are not affected by the environment, and they can be identified in any tissue and developmental stage. From their development, their use in agriculture increased through the construction of genetic maps in crop species, the association between molecular markers and important agronomic traits, the dissection of quantitative traits, and the positional cloning of genes of interest. Besides the estimation of genetic distances and molecular cloning, molecular markers provide the most suitable tool for the evaluation of genetic diversity, allowing for the selection of the most suitable parental lines in breeding programs, the management of germplasm collections, and varietal identification [2].

Once the association between a marker and a trait is detected, it can be deployed into a breeding program through marker-assisted selection (MAS). The success of this technique relies on the identification of markers tightly linked with the genome region of a target trait. MAS improves the efficiency of the selection of Mendelian traits, facilitating the introgression of single genes with the desired alleles into elite cultivars and removing the undesirable genome of the donor parent in a backcrossing program, and allows for the identification and protection of commercial cultivars through fingerprinting [3]. Although MAS has been effectively used before for Mendelian traits or

those regulated by a low number of genes, many of the agronomic traits show a quantitative nature and are influenced by the environment [4]. In the last decade, the development of high-throughput genotyping platforms has allowed for the screening of whole genomes for the selection of desired traits. In addition, novel statistical approaches for the use of large amounts of genetic and phenotypic data have been developed. Altogether, this has allowed for the development of new selection strategies, such as the genomic prediction or selection (GS), which attempts to skip the limitations of MAS [5].

Coupled with the rapid development of high-throughput genotypic technology, the next DNA sequencing technologies and their applications in genetic and physical mapping have made significant progress in accelerating plant breeding with a lower cost.

## 2. Overview of the Special Issue

This special issue of *Agronomy* with the title “Molecular marker technology for crop improvement” publishes 16 articles providing insights into the different applications of molecular markers in plant breeding. Eleven crop species were analyzed with six different approaches.

Although the development of linkage maps is the first step for gene identification and many maps have been developed in several crop species, novel marker approaches involving high-throughput genotyping are evolving for the construction of highly saturated maps. A high-density single nucleotide polymorphism (SNP) linkage map in potato was developed using a recently developed strategy for the discovery of SNPs, the specific length amplified fragment sequencing (SLAF-seq) approach [6].

Biparental quantitative trait loci (QTL) mapping is a classical approach to identify multi-genic traits. The success of detecting a QTL depends on several factors: (1) marker density, (2) population size, and (3) trait heritability. As a classical approach in crop breeding, within the Special Issue it was applied for the identification of the loci controlling milling yield in rice [7], resistance to *Striga hermonthica* in maize [8], and leaf rust and stem rust resistance in wheat [9]. These studies identified new loci for important traits in breeding and will be the starting point for a deeper analysis of candidate gene identification.

Genome-wide association studies (GWASs) have become a valuable tool in recent years as a complementary approach to biparental mapping, providing broader allelic coverage and higher mapping resolution. In this issue, GWASs were performed for the analysis of seminal roots in landraces of wheat and durum wheat from the Mediterranean basin [10,11], agronomic and quality traits in elite durum wheat [12], and for flowering time in maize inbred lines [13]. The studies of Roselló et al. [10] and Rufo et al. [11] pointed out the usefulness of the old germplasm to be used as genetic resources for improving drought-related traits in the breeding programs to broaden the genetic variability. Merida-García et al. [12] combined a GWAS with a candidate gene approach to successfully identify gene clusters involved in important traits for wheat breeding. The study of Maldonado et al. [13] revealed that the use of a GWAS based on haplotype blocks was more efficient than the standard approach to identify major effect loci, and the network-assisted gene prioritization used identified four genes influencing flowering time in tropical maize.

Genetic diversity is crucial for crop improvement, as it allow breeders to identify appropriate parents to be included in breeding programs for broadening genetic variability. Within this Special Issue, the genetic diversity of wheat, avocado, and raspberry was assessed by high-resolution melting (HRM), and insertion site-based polymorphism (ISBP) markers, simple sequence repeats (SSR) developed from single-molecule long-read sequences, and SSRs from flavonoid biosynthesis genes, respectively [14–16]. Merida-García et al. [14] developed ISBP markers for the wheat genome as an alternative to SSRs and SNPs. The authors concluded that these HRM-ISBPs are a cost-effective and efficient marker approach for wheat breeding programs, being also useful for gene tagging. The studies of Ge et al. [15] and Lebedev et al. [16] demonstrated the power of SSR markers. Although they were developed three decades ago [17], they are still commonly used because of their codominant and multi-allelic nature and high reproducibility.

The abovementioned studies in this Special Issue have taken into account the use of molecular markers for the development of linkage maps, the analysis of genetic diversity, and the mapping of quantitative traits, by means of the classic biparental QTL mapping or GWASs. The rest of the studies represent direct applications of molecular markers after these previous steps of development, location in a linkage map, and genotype–phenotype association: MAS and GS.

Marker-assisted selection was applied in four studies for (1) the selection of parental germplasm in sugarcane breeding programs [18], (2) improving blast resistance and salt tolerance in rice [19], (3) the selection of the pollination constant non-astringent (PCNA) type in Spanish germplasms of persimmon [20], and (4) the selection of resistance to plum pox virus (PPV) in apricot by allele-specific PCR [21]. Wu et al. [18] identified two groups among the 150 most widely used sugarcane parental clones. Based on these results, the authors could identify the most appropriate cultivars to broaden the genetic base of breeding germplasm. Thanasilungura et al. [19] improved the rice cultivar RD6 for salt tolerance with the QTL “Saltol” and blast resistance with four different QTLs by marker-assisted backcrossing and phenotypic selection. The authors found that one of the introgression lines showed superior salt tolerance and blast resistance, maintaining higher quality and agronomic performance than RD6. In the study of the selection of the PCNA type of persimmon, Blasco et al. [20] identified in the Spanish germplasm of persimmon the previously developed markers DISx-AF4, linked to the production of male flowers, and AST, linked to fruit astringency. The screening of these markers in different progenies of backcrosses demonstrated a very low rate of selection of both traits together and is thus a valuable tool in a breeding program. The last example of MAS in this Special Issue corresponds to the selection of PPV resistance in the apricot breeding program at IVIA (Valencia, Spain) [21]. In this study, the authors present a high-throughput method for a rapid test of PPV resistance, thus improving the efficiency of apricot breeding programs at a low cost. MAS is of special interest in fruit trees due to the long time needed to obtain a new generation.

Finally, GS was used to estimate the breeding values for grain composition in sorghum [22]. Although GS was developed initially for animal breeding, its use in plant breeding has been extended in recent years. GS emerged as a valuable tool for improving complex traits controlled by QTLs with small effects. Together with high-throughput phenotyping techniques, it has brought a revolution in breeding by enhancing the accuracy level of selection. Sapkota et al. [22] report the use of GS for grain compositional traits. The authors found that the prediction accuracy for single trait prediction was moderate to high in respect to the phenotypic measurements obtained from near infra-red spectroscopy (NIRS) prediction.

### **3. Concluding Remarks**

The Special Issue covers the use of different types of molecular markers, from SSR markers developed in 1989 to the newest high-throughput marker technology, as well as different approaches for genetic mapping and the use of molecular markers to assist crop breeding as a single marker with MAS or at genome level with GS.

To meet the needs of a growing world population, crop yields must be increased under the climate change scenario predicted for the coming decades and the threat of the emergence of new pathogens. However, the information obtained by genome sequencing and its availability at low cost, the continuous development of new molecular markers, the implementation of high-throughput phenotyping tools, and speed breeding techniques will make it possible to face these new challenges. Although not directly related with the molecular marker technology of this Special Issue, the new systems based on gene editing will play an important role in the future of agriculture.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

1. Hou, H.; Atlihan, N.; Lu, Z.X. New biotechnology enhances the application of cisgenesis in plant breeding. *Front. Plant Sci.* **2014**, *5*, 389. [[CrossRef](#)]
2. Phougat, D.; Panwar, I.; Puina, M.; Sethi, S. Microsatellite markers based characterization in advance breeding lines and cultivars of bread wheat. *J. Environ. Biol.* **2018**, *39*, 339–346. [[CrossRef](#)]
3. Shimelis, H.; Laing, M. Timelines in conventional crop improvement: Pre-breeding and breeding procedures. *Aust. J. Crop Sci.* **2012**, *6*, 1542–1549.
4. Goutam, U.; Kukreja, S.; Tiwari, R.; Chaudhury, A.; Gupta, R.K.; Dholakia, B.B.; Yadav, R. Biotechnological approaches for grain quality improvement in wheat: Present status and future possibilities. *Aust. J. Crop Sci.* **2013**, *7*, 469–483.
5. Talukder, S.K.; Saha, M.C. Toward Genomics-Based Breeding in C3 Cool-Season Perennial Grasses. *Front. Plant Sci.* **2017**, *8*, 1317. [[CrossRef](#)]
6. Yu, X.; Zhang, M.; Yu, Z.; Yang, D.; Li, J.; Wu, G.; Li, J. An SNP-Based High-Density Genetic Linkage Map for Tetraploid Potato Using Specific Length Amplified Fragment Sequencing (SLAF-Seq) Technology. *Agronomy* **2020**, *10*, 114. [[CrossRef](#)]
7. Zhang, H.; Zhu, Y.J.; Zhu, A.D.; Fan, Y.Y.; Huang, T.X.; Zhang, J.F.; Xie, H.A.; Zhuang, J.Y. Identification and Verification of Quantitative Trait Loci Affecting Milling Yield of Rice. *Agronomy* **2020**, *10*, 75. [[CrossRef](#)]
8. Badu-Apraku, B.; Adewale, S.; Paterne, A.; Gedil, M.; Asiedu, R. Identification of QTLs Controlling Resistance/Tolerance to *Striga hermonthica* in an Extra-Early Maturing Yellow Maize Population. *Agronomy* **2020**, *10*, 1168. [[CrossRef](#)]
9. Genievskaia, Y.; Abugalieva, S.; Rsaliyev, A.; Yskakova, G.; Turuspekov, Y. QTL Mapping for Seedling and Adult Plant Resistance to Leaf and Stem Rusts in Pamyati Azieva × Paragon Mapping Population of Bread Wheat. *Agronomy* **2020**, *10*, 1285. [[CrossRef](#)]
10. Roselló, M.; Royo, C.; Sanchez-Garcia, M.; Soriano, J.M. Genetic Dissection of the Seminal Root System Architecture in Mediterranean Durum Wheat Landraces by Genome-Wide Association Study. *Agronomy* **2019**, *9*, 364. [[CrossRef](#)]
11. Rufo, R.; Salvi, S.; Royo, C.; Soriano, J.M. Exploring the Genetic Architecture of Root-Related Traits in Mediterranean Bread Wheat Landraces by Genome-Wide Association Analysis. *Agronomy* **2020**, *10*, 613.
12. Mérida-García, R.; Bentley, A.R.; Gálvez, S.; Dorado, G.; Solís, I.; Ammar, K.; Hernandez, P. Mapping Agronomic and Quality Traits in Elite Durum Wheat Lines under Differing Water Regimes. *Agronomy* **2020**, *10*, 144.
13. Maldonado, C.; Mora, F.; Bertagna, F.A.B.; Kuki, M.C.; Scapim, C.A. SNP- and Haplotype-Based GWAS of Flowering-Related Traits in Maize with Network-Assisted Gene Prioritization. *Agronomy* **2019**, *9*, 725.
14. Mérida-García, R.; Gálvez, S.; Paux, E.; Dorado, G.; Pascual, L.; Giraldo, P.; Hernandez, P. High Resolution Melting and Insertion Site-Based Polymorphism Markers for Wheat Variability Analysis and Candidate Genes Selection at Drought and Heat MQTL Loci. *Agronomy* **2020**, *10*, 1294.
15. Ge, Y.; Zang, X.; Tan, L.; Wang, J.; Liu, Y.; Li, Y.; Wang, N.; Chen, D.; Zhan, R.; Ma, W. Single-Molecule Long-Read Sequencing of Avocado Generates Microsatellite Markers for Analyzing the Genetic Diversity in Avocado Germplasm. *Agronomy* **2019**, *9*, 512.
16. Lebedev, V.G.; Subbotina, N.M.; Maluchenko, O.P.; Krutovsky, K.V.; Shestibratov, K.A. Assessment of Genetic Diversity in Differently Colored Raspberry Cultivars Using SSR Markers Located in Flavonoid Biosynthesis Genes. *Agronomy* **2019**, *9*, 518.
17. Tautz, D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res.* **1989**, *17*, 6463–6471.
18. Wu, J.; Wang, Q.; Xie, J.; Pan, Y.-B.; Zhou, F.; Guo, Y.; Chang, H.; Xu, H.; Zhang, W.; Zhang, C.; et al. SSR Marker-Assisted Management of Parental Germplasm in Sugarcane (*Saccharum* spp. hybrids) Breeding Programs. *Agronomy* **2019**, *9*, 449.
19. Thanasilungura, K.; Kranto, S.; Monkham, T.; Chankaew, S.; Sanitchon, J. Improvement of a RD6 Rice Variety for Blast Resistance and Salt Tolerance through Marker-Assisted Backcrossing. *Agronomy* **2020**, *10*, 1118.
20. Blasco, M.; Gil-Muñoz, F.; Naval, M.M.; Badenes, M.L. Molecular Assisted Selection for Pollination-Constant and Non-Astringent Type without Male Flowers in Spanish Germplasm for Persimmon Breeding. *Agronomy* **2020**, *10*, 1172.

21. Polo-Oltra, Á.; Romero, C.; López, I.; Badenes, M.L.; Zuriaga, E. Cost-Effective and Time-Efficient Molecular Assisted Selection for Ppv Resistance in Apricot Based on *ParPMC2* Allele-Specific PCR. *Agronomy* **2020**, *10*, 1292. [[CrossRef](#)]
22. Sapkota, S.; Boatwright, J.L.; Jordan, K.; Boyles, R.; Kresovich, S. Multi-Trait Regressor Stacking Increased Genomic Prediction Accuracy of Sorghum Grain Composition. *Agronomy* **2020**, *10*, 1221. [[CrossRef](#)]



© 2020 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).





Article

# Genetic Dissection of the Seminal Root System Architecture in Mediterranean Durum Wheat Landraces by Genome-Wide Association Study

Martina Roselló <sup>1</sup>, Conxita Royo <sup>1</sup>, Miguel Sanchez-Garcia <sup>2</sup> and Jose Miguel Soriano <sup>1,\*</sup>

<sup>1</sup> Sustainable Field Crops Programme, Institute for Food and Agricultural Research and Technology (IRTA), 25198 Lleida, Spain

<sup>2</sup> International Centre for Agricultural Research in Dry Areas (ICARDA), Rabat 10112, Morocco

\* Correspondence: josesmiguel.soriano@irta.cat; Tel.: +34-973-032850

Received: 19 June 2019; Accepted: 8 July 2019; Published: 9 July 2019

**Abstract:** Roots are crucial for adaptation to drought stress. However, phenotyping root systems is a difficult and time-consuming task due to the special feature of the traits in the process of being analyzed. Correlations between root system architecture (RSA) at the early stages of development and in adult plants have been reported. In this study, the seminal RSA was analysed on a collection of 160 durum wheat landraces from 21 Mediterranean countries and 18 modern cultivars. The landraces showed large variability in RSA, and differences in root traits were found between previously identified genetic subpopulations. Landraces from the eastern Mediterranean region, which is the driest and warmest within the Mediterranean Basin, showed the largest seminal root size in terms of root length, surface, and volume and the widest root angle, whereas landraces from eastern Balkan countries showed the lowest values. Correlations were found between RSA and yield-related traits in a very dry environment. The identification of molecular markers linked to the traits of interest detected 233 marker-trait associations for 10 RSA traits and grouped them in 82 genome regions named marker-trait association quantitative trait loci (MTA-QTLs). Our results support the use of ancient local germplasm to widen the genetic background for root traits in breeding programs.

**Keywords:** durum wheat; landraces; marker-trait association; root system architecture

## 1. Introduction

Wheat is estimated to have been first cultivated around 10,000 years before present (BP) in the Fertile Crescent region. It spread to the west of the Mediterranean Basin and reached the Iberian Peninsula around 7000 years BP [1]. During this migration, both natural and human selection resulted in the development of local landraces considered to be very well adapted to the regions where they were grown and containing the largest genetic diversity within the species [2]. From the middle of the 20th century, as a consequence of the Green Revolution, the cultivation of local landraces was progressively abandoned and replaced by the improved, more productive, and genetically uniform semi-dwarf cultivars. However, scientists are convinced that local landraces may provide new alleles to improve commercially valuable traits [3]. Introgression of these alleles into modern cultivars can be very useful, especially in breeding for suboptimal environments.

Drought is the most important environmental factor limiting wheat productivity in many parts of the world. Therefore, improving yield under water-limited conditions is one of the major challenges for wheat production worldwide. Breeding for adaptation to drought is extremely challenging due to the complexity of the target environments and the stress-adaptive mechanisms adopted by plants to withstand and mitigate the negative effects of a water deficit [4]. These mechanisms allow the plant to escape (e.g., early flowering date), avoid (e.g., root system), and/or tolerate (e.g., osmolyte accumulation)

the negative effects of drought, which plays a role in determining final crop performance [5]. The crop traits to be considered as selection targets under drought conditions must be genetically correlated with yield and should have a greater heritability than yield itself [6,7]. Among these traits, early vigour, leaf area duration, crop water status, radiation use efficiency, and root architecture have been identified to be associated with yield under rainfed conditions (reviewed by Reference [8]).

Root system architecture (RSA) is crucial for wheat adaptation to drought stress. Roots exhibit a high level of morphological plasticity in response to soil conditions, which allows plants to adapt better, which is particularly under drought conditions. However, evaluating root architecture in the field is very difficult, expensive, and time-consuming, especially when a large number of plants need to be phenotyped. Several studies have reported a correlation of RSA in the early stages of development with RSA in adult plants [9], Manschadi et al. [10] reported that adult root geometry is strongly related to seminal root angle (SRA). Wasson et al. [11] described a relationship of root vigor between plants grown in the field and controlled conditions. Several systems have been adopted to enable early screening of the RSA in wheat [12].

Identifying quantitative trait loci (QTLs) and using marker-assisted selection is an efficient way to increase selection efficiency and boost genetic gains in breeding programs. However, while numerous studies have reported QTLs for RSA in bi-parental crosses [13], very few of them were based on association mapping [12,14–18]. Association mapping is a complementary approach to bi-parental linkage analysis and provides broader allelic coverage with higher mapping resolution. Association mapping is based on linkage disequilibrium, defined as the non-random association of alleles at different loci, and is used to detect the relationship between phenotypic variation and genetic polymorphism.

The main objectives of the present study were a) to identify differences in RSA among genetic subpopulations of durum wheat Mediterranean landraces, b) to find correlations of RSA with yield-related traits in different rainfed Mediterranean environments, and c) to identify molecular markers linked to RSA in the old Mediterranean germplasm through a genome-wide association study.

## **2. Materials and Methods**

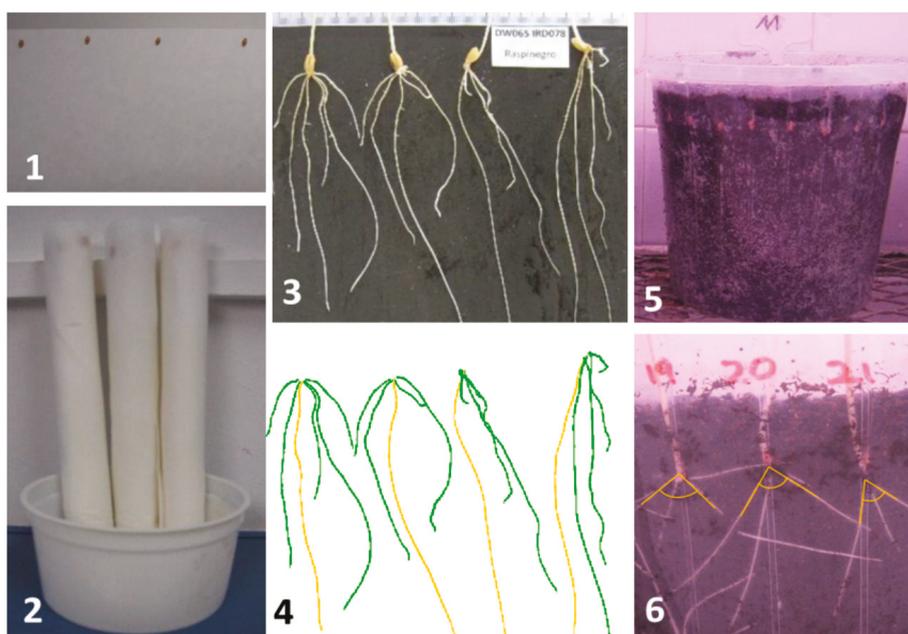
### *2.1. Plant Material*

The germplasm used in the current study consisted of a set of 160 durum wheat landraces from 21 Mediterranean countries and 18 modern cultivars from a previously structured collection [2,19]. The landraces were classified into four genetic subpopulations (SPs) that matched their geographical origin as follows: the eastern Mediterranean (19 genotypes), the eastern Balkans and Turkey (20 genotypes), the western Balkans and Egypt (31 genotypes), the western Mediterranean (71 genotypes), and 19 genotypes that remained as admixed (Supplementary Materials Table S1).

### *2.2. Phenotyping*

Eight uniform seeds per genotype were cultured following the paper roll method [20,21] in two replicates of four seeds. The seeds were placed at the top of a filter paper (420 × 520 mm) with the embryo facing down and sprayed with a 0.4% sodium hypochlorite solution. Subsequently, the papers were folded in half to obtain a 210 × 520 mm rectangle with the seeds fixed at the top. The papers were misted with deionized water and rolled by hand. The rolls were placed in plastic pots with deionized water at the bottom that was regularly checked to ensure it did not evaporate. The experiment was conducted in a growth chamber at 25 °C and darkness conditions. One week after sowing, the seeds were transferred to a black surface to take digital images that were processed by SmartRoot software [22] (Figure 1). Nine traits for the seminal root system architecture (RSA) were measured: total root number (TRN), primary root length (PRL, cm), total lateral root length (LRL, cm), primary root surface (PRS, cm<sup>2</sup>), total lateral root surface (LRS, cm<sup>2</sup>), primary root volume (PRV, cm<sup>3</sup>), total

lateral root volume (LRV, cm<sup>3</sup>), primary root diameter (PRD, cm), and mean lateral root diameter (LRD, cm).



**Figure 1.** Experimental setup for root system architecture analysis. First, seeds were placed on humid filter paper (1) and rolled. Paper rolls were placed in plastic pots with deionized water at the bottom for root growth (2). One week after sowing, the seeds were transferred to a black surface for digital imaging (3) that were processed by SmartRoot software [22] (4). The seminal root angle was measured using the clear pots (5,6).

Additionally, the SRA (°) was measured at the facilities of the International Center for Agricultural Research in the Dry Areas (ICARDA) in Rabat (Morocco) using the clear pot method described by Richard et al. [23] (Figure 1). Using a randomized complete block design, eight seeds per genotype were grown in 4 L clear pots filled with peat. The seeds were placed with the embryo facing down and close to the pot wall to facilitate root growth along the transparent wall. The pots were then watered, placed inside 4 L black pots, and kept at 20 °C and darkness conditions in a growth chamber. Five days after sowing, digital images were taken and processed with ImageJ software [24].

Data from field experiments conducted under rainfed conditions during two years of contrasting water input from sowing to physiological maturity (285 mm in 2008 and 104 mm in 2014) in Lleida, North-eastern Spain [25] were used to assess the relationships between RSA traits and yield-related traits.

The experiments were carried out in a non-replicated modified augmented design with three replicated checks (the cultivars ‘Claudio,’ ‘Simeto,’ and ‘Vitron’) and plots of 6 m<sup>2</sup> (8 rows, 5 m long with a 0.15 m spacing). Sowing density was adjusted to 250 viable seeds m<sup>-2</sup> and the plots were maintained free of weeds and diseases.

### 2.3. Statistical Analysis

Combined analyses of variance (ANOVA) were performed for the RSA traits of the structured accessions (141 landraces and 18 modern cultivars), considering the accessions and the replicate as random effects. The sum of squares of the cultivar effect was partitioned into differences between SPs and differences within them. The Kenward-Roger correction was used due to the unbalanced

number of genotypes within the SPs. Since the experiment was divided into six sets with one check, least squared means were calculated using Simeto as a check and compared using the Tukey test [26] at  $p < 0.01$ .

Raw field data were fitted to a linear mixed model with the check cultivars as fixed effects and the row number, column number, and genotype as random effects [27]. Restricted maximum likelihood was used to estimate the variance components and to produce the best linear unbiased predictors (BLUPs) for yield and yield components. The relationships between RSA traits and yield-related traits were assessed through correlation analyses. All calculations were carried out using the SAS statistical package [28].

#### 2.4. Genotyping

DNA isolation was performed from leaf samples following the method reported by Doyle and Doyle [29]. High throughput genotyping was performed at Diversity Arrays Technology Pty Ltd. (Canberra, Australia) (<http://www.diversityarrays.com>) with the genotyping by sequencing (GBS) DARtSeq platform [30]. A total of 46,161 markers were used to genotype the association mapping panel, including 35,837 presence/absence variants (PAVs) and 10,324 single nucleotide polymorphisms (SNPs). Markers were ordered according to the consensus map of wheat v4 available at <https://www.diversityarrays.com/>.

#### 2.5. Linkage Disequilibrium

Linkage disequilibrium (LD) among markers was calculated for the A and B genomes using markers with a map position on the wheat v4 consensus map, and a minor allele frequency greater than 5%, using TASSEL 5.0 [31]. Pair-wise LD was measured using the squared allele frequency correlations  $r^2$  and the values for genomes A and B were plotted against the genetic distance to determine how fast the LD decays. A LOESS curve was fitted to the plot using the JMP v12Pro statistical package (SAS Institute Inc, Cary, NC, USA).

#### 2.6. Genome-Wide Association Study

A genome-wide association study (GWAS) was performed with 160 landraces for the mean of measured traits with TASSEL 5.0 software [31]. A mixed linear model was conducted using the population structure determined by Soriano et al. [19] as the fixed effect and a kinship (K) matrix as the random effect (Q + K) at the optimum compression level. A false discovery rate threshold [32] was established at  $-\log_{10}p > 4.6$  ( $p < 0.05$ ), using 2135 markers according to the results of the LD decay, to consider a marker-trait association (MTA) significant. Moreover, a second, less restrictive threshold was established at  $-\log_{10}p > 3$ . To simplify the MTA information, those associations located within LD blocks were considered to belong to the same QTL and were named marker-trait association quantitative trait loci (MTA-QTLs). Graphical representation of the genetic position of MTA-QTLs was carried out using MapChart 2.3 [33].

#### 2.7. Gene Annotation

Gene annotation for the target region of significant MTAs was performed using the gene models for high-confidence genes reported for the wheat genome sequence [34] available at <https://wheat-urgi.versailles.inra.fr/Seq-Repository/>.

### 3. Results

#### 3.1. Phenotypic Analyses

The ANOVA showed that, for all traits, the phenotypic variability was mainly explained by the cultivar effect, since it accounted for 63.41% (PRD) to 90.57% (LRD) of the total sum of squares (Table 1). A summary of the genetic variation of the RSA traits is shown in Supplementary Materials

Table S2. The partitioning of the sum of squares of the cultivar effect into differences between and within SPs revealed that the variability induced by the genotype was mainly explained by differences within SPs on a range from 70.1% for TRN to 91.5 for PRV (Table 1). Differences between SPs were statistically significant for all traits, accounting for 8.5% (PRV) to 30.5% (TRN) of the sum of squares of the genotype effect (Table 1). Western Mediterranean landraces showed the highest number of seminal roots and the narrowest root angle, whereas the eastern Balkans and Turkey SP showed the widest angle (Table 2). The highest values for root size-related traits (length, surface and volume) in both primary and lateral roots were recorded in the eastern Mediterranean landraces. The western Balkans and Egypt subpopulation showed the largest root diameter (Table 2). The comparison of mean values of eastern Balkans and Turkish landraces revealed that the Turkish ones had high values for all traits except TRN, LRL, and root diameter (Supplementary Materials Table S3). The modern cultivars showed intermediate values for all RSA traits (Table 2).

**Table 1.** Percentage of the sum of squares of the ANOVA model for the seminal root system architecture traits in a set of 159 Mediterranean durum wheat genotypes structured into five genetic subpopulations by Soriano et al. [19].

Source of Variation	df	TRN	SRA	PRL	LRL	PRS	LRS	PRV	LRV	PRD	LRD
Genotype	158	84.1 ***	69.4 ***	87.0 ***	86.8 ***	85.3 ***	86.1 ***	82.9 ***	86.8 ***	63.41 ***	90.57 ***
Between subpopulations	4	30.5 ***	16.6 ***	15.4 ***	22.7 ***	11.6 ***	18.5 ***	8.5 **	12.7 ***	10.44 **	17.96 ***
Within subpopulations	154	70.1 ***	83.4 ***	85.0 ***	77.8 ***	88.6 ***	82.0 ***	91.5 ***	87.6 ***	89.33 **	81.87 ***
Replicate	1	0.001	1.83 **	0.00	0.43 *	0.36 *	0.36 *	0.86 **	0.33 **	1.35 *	0.12
Error	157	15.9	28.8	13.0	12.9	14.3	13.6	16.1	12.9	35.25	9.33
Total	316										

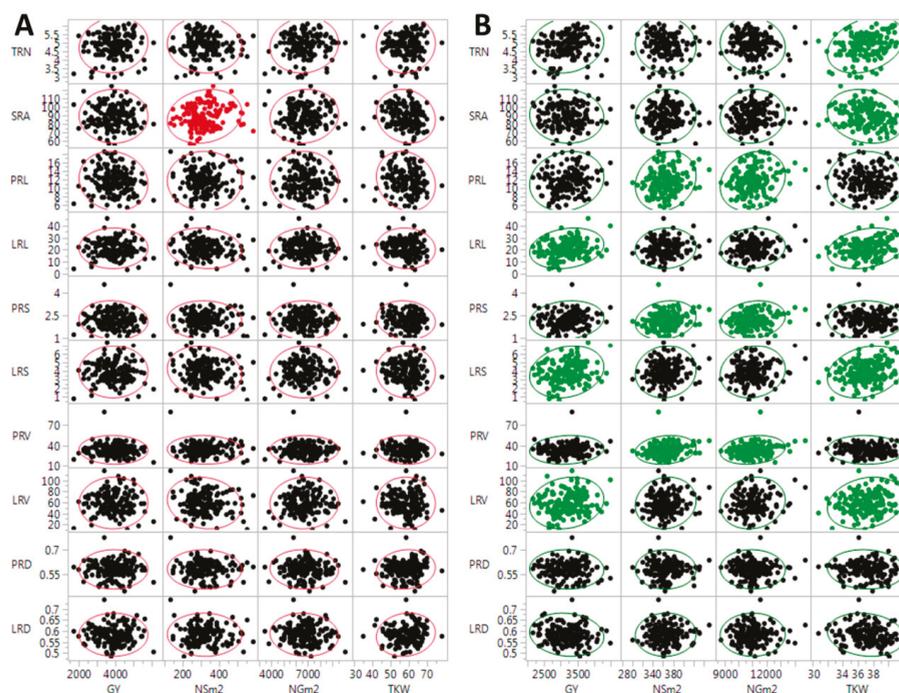
TRN, total root number. SRA, seminal root angle. PRL, primary root length. LRL, total lateral root length. PRS, primary root surface. LRS, total lateral root surface. PRV, primary root volume. LRV, total lateral root volume. PRD, primary root diameter. LRD, mean lateral root diameter. \*  $p < 0.05$ . \*\*  $p < 0.01$ . \*\*\*  $p < 0.001$ .

**Table 2.** Means comparison of seminal root system architecture traits measured in a set of 159 Mediterranean durum wheat genotypes structured into five genetic subpopulations [19]. Means within columns with different letters are significantly different at  $p < 0.01$  following a Tukey test.

	TRN	SRA	PRL	LRL	PRS	LRS	PRV	LRV	PRD	LRD
EM	4.8 <sup>b</sup>	94.7 <sup>ab</sup>	13.8 <sup>a</sup>	25.1 <sup>a</sup>	2.5 <sup>a</sup>	4.6 <sup>a</sup>	38.3 <sup>a</sup>	67.2 <sup>a</sup>	0.57 <sup>b</sup>	0.57 <sup>b</sup>
EB + T	4.8 <sup>b</sup>	98.2 <sup>a</sup>	10.3 <sup>c</sup>	17.2 <sup>b</sup>	1.9 <sup>c</sup>	3.2 <sup>b</sup>	28.5 <sup>b</sup>	47.6 <sup>b</sup>	0.57 <sup>b</sup>	0.58 <sup>b</sup>
WB + E	4.3 <sup>c</sup>	87.6 <sup>bc</sup>	10.4 <sup>c</sup>	16.5 <sup>b</sup>	2.1 <sup>bc</sup>	3.2 <sup>b</sup>	33.6 <sup>ab</sup>	52.3 <sup>b</sup>	0.61 <sup>a</sup>	0.62 <sup>a</sup>
WM	5.2 <sup>a</sup>	84.5 <sup>c</sup>	11.8 <sup>ab</sup>	23.5 <sup>a</sup>	2.2 <sup>bc</sup>	4.3 <sup>a</sup>	33.0 <sup>ab</sup>	63.9 <sup>a</sup>	0.58 <sup>b</sup>	0.58 <sup>b</sup>
Modern	4.5 <sup>bc</sup>	93.9 <sup>ab</sup>	12.8 <sup>ab</sup>	20.8 <sup>ab</sup>	2.4 <sup>ab</sup>	3.7 <sup>ab</sup>	35.2 <sup>ab</sup>	54.5 <sup>ab</sup>	0.56 <sup>b</sup>	0.57 <sup>b</sup>

TRN, total root number. SRA, seminal root angle. PRL, primary root length. LRL, total lateral root length. PRS, primary root surface. LRS, total lateral root surface. PRV, primary root volume. LRV, total lateral root volume. PRD, primary root diameter. LRD, mean lateral root diameter. EM, Eastern Mediterranean. EB + T, Eastern Balkans and Turkey. WB + E, Western Balkans and Egypt. WM, Western Mediterranean.

Correlation coefficients between RSA traits and yield-related traits were calculated for two field experiments with contrasting water input (285 and 104 mm of rainfall from sowing to physiological maturity). Whereas, for the rainiest environment, only the relationship between SRA and number of spikes per square meter (NSm<sup>2</sup>) was statistically significant ( $p = 0.043$ ,  $r^2 = 0.16$ ). For the driest environment, 14 correlations involving all the yield-related traits and RSA traits except root diameter were statistically significant (Figure 2) ( $r^2$  between 0.17 for NSm<sup>2</sup> and PRL and PRS to 0.30 for TKW and TRN). Most of the significant correlations were positive. Only the relationship between SRA and thousand kernel weight (TKW) was negative.



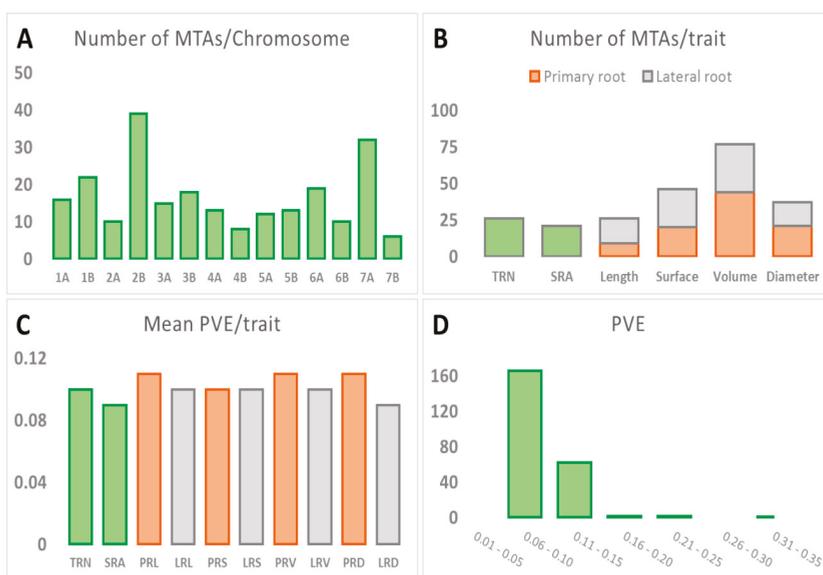
**Figure 2.** Correlations between seminal root system architecture traits and yield-related traits determined in field experiments receiving high (density ellipse in red, **A**) and low (density ellipse in green, **B**) water input from sowing to physiological maturity. Significant correlation coefficients ( $p < 0.05$ ) are indicated with red and green points. TRN, total root number. SRA, seminal root angle. PRL, primary root length. LRL, total lateral root length. PRS, primary root surface. LRS, total lateral root surface. PRV, primary root volume. LRV, total lateral root volume. PRD, primary root diameter. LRD, mean lateral root diameter. GY, grain yield. NSm<sup>2</sup>, number of spikes per square meter. NGm<sup>2</sup>, number of grains per square meter. TKW, thousand kernel weight.

### 3.2. Marker-Trait Associations

A total of 46,161 DArTseq markers, including PAVs and SNPs, were used to genotype the set of 160 durum wheat landraces. To reduce the risk of false positives, markers and accessions were analyzed for the presence of duplicated patterns and missing values. Of 35,837 PAVs, 24,188 were placed on the wheat v4 consensus map. Of these, those with more than 30% of missing data and those with a minor allele frequency lower than 5% were removed from the analysis, leaving 19,443 PAVs. A total of 6957 SNPs were mapped, leaving a total of 4686 SNPs after marker filtering as before. Additionally, 413 markers were duplicated between PAVs and SNPs, so the corresponding PAVs were eliminated. A total of 23,716 markers remained for the subsequent analysis.

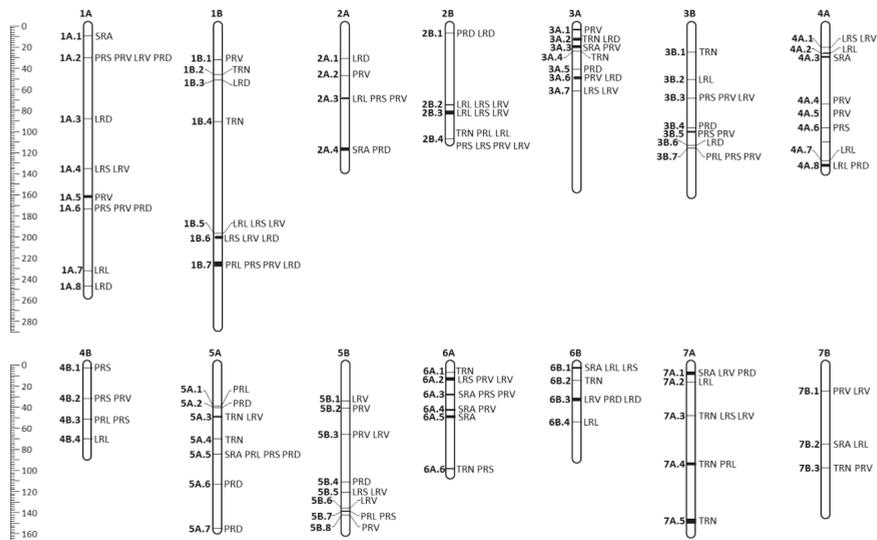
Linkage disequilibrium was estimated for locus pairs in genomes A and B using a sliding window of 50 cM. A total of 471,319 and 681,389 possible pair-wise loci were observed for genomes A and B, respectively. Of these locus pairs, 52% and 43% showed significant linkage disequilibrium at  $p < 0.01$  and  $p < 0.001$ , respectively. Mean  $r^2$  was 0.12 for genome A and 0.11 for genome B. These means were used as a threshold for estimating the intercept of the LOESS curve to determine the distance at which LD decays in each genome. Markers were in LD in a range from less than 1 cM in genome B to 1 cM in genome A (Supplementary Materials Table S4).

Results of the GWAS are reported in Figure 3 and in Supplementary Materials Table S5. Using a restrictive threshold based on a false discovery rate at  $p < 0.05$  ( $-\log_{10}p > 4.6$ ) and the LD decay, only 12 MTAs corresponding to seven markers were significant. Using a common threshold of  $-\log_{10}p > 3$ , as previously reported by other authors [35–38], a total of 233 MTAs involving 176 markers were identified. MTAs were equally distributed in both genomes (50.2% in the A genome and 49.8% in the B genome). Chromosomes 2B and 7A harbored the highest number of MTAs (39 and 32 respectively), carrying 30% of the total number of MTAs, whereas chromosomes 4B and 7B harbored the lowest number of MTAs, 8 and 6 MTAs, respectively (Figure 3A). Root volume was the trait showing the highest number of MTAs (77), followed by root surface (46), root diameter (37), root length and number (26), and SRA (21) (Figure 3B). The mean percentage of phenotypic variance explained (PVE) per MTA was similar for all traits, ranging from 0.09 to 0.11 (Figure 3C). Most of the MTAs showed low PVE, in agreement with the quantitative nature of the analyzed traits. The percentage of MTAs with a PVE lower than 0.1 was 71%, whereas that of MTAs with a PVE lower than 0.15 was 98% (Figure 3D).



**Figure 3.** Summary of marker trait associations (MTA). (A) Number of MTAs per chromosome. (B) Number of MTAs per trait. (C) Mean PVE per trait. (D) PVE. TRN, total root number. SRA, seminal root angle. PRL, primary root length. LRL, total lateral root length. PRS, primary root surface. LRS, total lateral root surface. PRV, primary root volume. LRV, total lateral root volume. PRD, primary root diameter. LRD, mean lateral root diameter.

To simplify the MTA information, those MTAs located within a region of 1 cM, as reported by the LD decay, were considered part of the same QTL. Thus, the 233 associations were restricted to 81 MTA-QTLs (Figure 4 and Table 3). Of the 82 MTA-QTLs, 33 had only one MTA, whereas, for the remaining 49, the number of MTAs per MTA-QTL ranged from 2 in 19 MTA-QTLs to 15 in mtaq-7A.1. When several consecutive pairs of MTAs were separated for a distance of 1 cm, the whole block was considered as the same MTA-QTL. The genomic distribution of MTA-QTLs showed that chromosome 1A, 4A, and 5B harbored 8 MTA-QTLs, chromosomes 1B, 3A, 3B, and 5A 7 MTA-QTLs, 6A 6 MTA-QTLs, 7A 5 MTA-QTLs, 2A, 2B, 4B, and 6B 4 MTA-QTLs and chromosome 7B harbored 3 MTA-QTLs. For the 48 MTA-QTLs with more than one MTA, 10 were related to one trait. Of these, mtaq-1A.5, mtaq-3A.1, mtaq-4A.4, and mtaq-4A.5 carried associations related to root volume, mtaq-2B.1, and mtaq-3B.7 to root diameter, mtaq-3B.1, and mtaq-7A.5 to root number, and mtaq-4A.3 and mtaq-6A.5 to the root angle.



**Figure 4.** MTA-QTL map. MTA-QTLs are indicated in bold on the left side of the chromosome and traits involved in each MTA-QTL are on the right side. The rule on the left indicates genetic distance in cM. TRN, total root number. SRA, seminal root angle. PRL, primary root length. LRL, total lateral root length. PRS, primary root surface. LRS, total lateral root surface. PRV, primary root volume. LRV, total lateral root volume. PRD, primary root diameter. LRD, mean lateral root diameter.

**Table 3.** MTA-QTLs.

MTA-QTLs	Chromosome	Position (cM)	MTAs	Trait
mtaq-1A.1	1A	9.24	1	SRA
mtaq-1A.2	1A	29.71	4	PRS PRV LRV PRD
mtaq-1A.3	1A	88.15	1	LRD
mtaq-1A.4	1A	135.37	2	LRS LRV
mtaq-1A.5	1A	160.75–163.11	3	PRV
mtaq-1A.6	1A	173.41	3	PRS PRV PRD
mtaq-1A.7	1A	231.76	1	LRL
mtaq-1A.8	1A	246.3	1	LRD
mtaq-1B.1	1B	31.69	1	PRV
mtaq-1B.2	1B	45.68	1	TRN
mtaq-1B.3	1B	51.29	1	LRD
mtaq-1B.4	1B	90.37	1	TRN
mtaq-1B.5	1B	196.56	3	LRL LRS LRV
mtaq-1B.6	1B	199.9–201.49	3	LRS LRV LRD
mtaq-1B.7	1B	223.51–227.36	12	PRL PRS PRV LRD
mtaq-2A.1	2A	31.13	1	LRD
mtaq-2A.2	2A	46.78	1	PRV
mtaq-2A.3	2A	68.39–68.96	4	LRL PRS PRV
mtaq-2A.4	2A	115.8–118.32	4	SRA PRD
mtaq-2B.1	2B	6.7	2	PRD LRD
mtaq-2B.2	2B	75.09–75.13	13	LRL LRS LRV
mtaq-2B.3	2B	80.79–83.84	16	LRL LRS LRV
mtaq-2B.4	2B	106.98–107.03	8	TRN PRL LRL PRS LRS PRV LRV
mtaq-3A.1	3A	3.32–3.58	3	PRV
mtaq-3A.2	3A	11.88–12.93	2	TRN LRD
mtaq-3A.3	3A	18.37–20.39	3	SRA PRV
mtaq-3A.4	3A	23.99	1	TRN

Table 3. Cont.

MTA-QTLs	Chromosome	Position (cM)	MTAs	Trait
mtaq-3A.5	3A	40.97	1	PRD
mtaq-3A.6	3A	48.06–49.67	3	PRV LRD
mtaq-3A.7	3A	61.57	2	LRS LRV
mtaq-3B.1	3B	24.98–25	2	TRN
mtaq-3B.2	3B	50.7	1	LRL
mtaq-3B.3	3B	68.36	4	PRS PRV LRV
mtaq-3B.4	3B	96.48	1	PRD
mtaq-3B.5	3B	100.07–101.44	3	PRS PRV
mtaq-3B.6	3B	112.86	4	LRD
mtaq-3B.7	3B	115.61	3	PRL PRS PRV
mtaq-4A.1	4A	20.42–26.03	2	LRS LRV
mtaq-4A.2	4A	26.03	1	LRL
mtaq-4A.3	4A	28.85–28.87	2	SRA
mtaq-4A.4	4A	74.09	2	PRV
mtaq-4A.5	4A	96.08	2	PRV
mtaq-4A.6	4A	109.72	1	PRS
mtaq-4A.7	4A	127.56	1	LRL
mtaq-4A.8	4A	131.42–132.72	2	LRL PRD
mtaq-4B.1	4B	2.79	1	PRS
mtaq-4B.2	4B	31.93	4	PRS PRV
mtaq-4B.3	4B	51.22	3	PRL PRS
mtaq-4B.4	4B	70.04	1	LRL
mtaq-5A.1	5A	38.83	1	PRL
mtaq-5A.2	5A	40.51	1	PRD
mtaq-5A.3	5A	48.57–48.65	2	TRN LRV
mtaq-5A.4	5A	69.82	1	TRN
mtaq-5A.5	5A	84.51	5	SRA PRL PRS PRD
mtaq-5A.6	5A	112.96	1	PRD
mtaq-5A.7	5A	155.41	1	PRD
mtaq-5B.1	5B	33.99	1	LRV
mtaq-5B.2	5B	40.83	1	PRV
mtaq-5B.3	5B	65.51	2	PRV LRV
mtaq-5B.4	5B	111.15	1	PRD
mtaq-5B.5	5B	120.34	2	LRS LRV
mtaq-5B.6	5B	135.45	1	LRV
mtaq-5B.7	5B	138.69	4	PRL PRS
mtaq-5B.8	5B	142.12	1	PRV
mtaq-6A.1	6A	7.11	1	TRN
mtaq-6A.2	6A	11.95–14.24	8	LRS PRV LRV
mtaq-6A.3	6A	27.82–28.69	3	SRA PRS PRV
mtaq-6A.4	6A	42.36	3	SRA PRV
mtaq-6A.5	6A	48.39–50.08	2	SRA
mtaq-6A.6	6A	98.51–98.82	2	TRN PRS
mtaq-6B.1	6B	2.41–3.31	5	SRA LRL LRS
mtaq-6B.2	6B	14.26	1	TRN
mtaq-6B.3	6B	31.49–33.46	3	LRV PRD LRD
mtaq-6B.4	6B	53.66	1	LRL
mtaq-7A.1	7A	5.7–9.43	15	SRA LRV PRD
mtaq-7A.2	7A	16.28	1	LRL
mtaq-7A.3	7A	47.85	3	TRN LRS LRV
mtaq-7A.4	7A	92.69–94.34	2	TRN PRL
mtaq-7A.5	7A	145.94–150.31	11	TRN
mtaq-7B.1	7B	24.48	2	PRV LRV
mtaq-7B.2	7B	74.86–75.24	2	SRA LRL
mtaq-7B.3	7B	97.45	2	TRN PRV

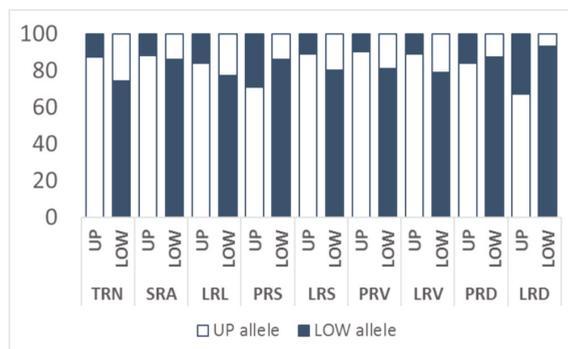
TRN, total root number. SRA, seminal root angle. PRL, primary root length. LRL, total lateral root length. PRS, primary root surface. LRS, total lateral root surface. PRV, primary root volume. LRV, total lateral root volume. PRD, primary root diameter. LRD, mean lateral root diameter.

Among all significant MTAs, markers with different alleles between extreme genotypes for each trait (i.e., the upper and lower 10th percentile) were identified except for PRL (Table 4, Figure 5). Frequency of the most common allele among genotypes from the upper 10th percentile ranged from 67% for LRD to 90% for PRV, whereas, for the lower 10th percentile, they ranged from 74% for TRN to 93% for LRD (Figure 5).

**Table 4.** Selected significant markers from the GWAS with different allele composition for the upper (UP) and lower (LOW) 10th percentile of genotypes. Different letters on the UP and LOW 10th phenotype indicate that means are significantly different at  $p < 0.01$  following a Tukey test.

Trait	Phenotype			Marker	Chromosome	Position	R <sup>3</sup>	Most Frequent Allele			
	Mean	UP 10th	LOW 10th					UP	Frequency	LOW	Frequency
TRN (N)	4.9	5.8 <sup>a</sup>	3.7 <sup>b</sup>	2260740_SNP	7A	148.38	0.09	T	0.80	C	0.81
				1252655_PAV	7B	97.45	0.11	1	0.94	0	0.67
SRA (°)	88.5	111.0 <sup>a</sup>	67.1 <sup>b</sup>	1125557_PAV	2A	115.80	0.09	0	1.00	1	1.00
				1117775_PAV	2A	118.32	0.10	1	0.75	0	0.71
LRL (cm)	21.8	36.5 <sup>a</sup>	9.3 <sup>b</sup>	4408432_PAV	6B	3.31	0.09	1	0.88	0	0.73
				4408958_PAV	6B	3.31	0.09	1	0.88	0	0.73
				1098568_PAV	6B	53.66	0.08	1	0.77	0	0.86
PRS (cm <sup>2</sup> )	2.2	3.2 <sup>a</sup>	1.3 <sup>b</sup>	4406631_PAV	4B	31.93	0.09	0	0.71	0	0.86
				4406980_PAV	4B	31.93	0.09	1	0.71	1	0.86
LRS (cm <sup>2</sup> )	4.0	6.5 <sup>a</sup>	1.8 <sup>b</sup>	1201756_PAV	2B	107.03	0.15	1	1.00	0	0.73
				987263_PAV	3A	61.57	0.10	0	0.92	1	0.88
				4408432_PAV	6B	3.31	0.09	1	0.81	0	0.79
				4408958_PAV	6B	3.31	0.09	1	0.81	0	0.79
PRV (mm <sup>3</sup> )	33.7	49.8 <sup>a</sup>	20.2 <sup>b</sup>	997799_SNP	1B	31.69	0.12	A	0.86	G	0.77
				1201756_PAV	2B	107.03	0.11	1	0.87	0	0.71
				4406631_PAV	4B	31.93	0.09	0	0.93	1	0.87
				4406980_PAV	4B	31.93	0.09	0	0.93	1	0.87
LRV (mm <sup>3</sup> )	60.5	99 <sup>a</sup>	27.7 <sup>b</sup>	1201756_PAV	2B	107.03	0.15	1	0.94	0	0.73
				987263_PAV	3A	61.57	0.10	0	0.93	1	0.81
				1126050_SNP	5B	33.99	0.07	A	0.81	M	0.81
				1149356_PAV	7B	24.48	0.08	0	0.87	1	0.81
PRD (mm)	0.58	0.66 <sup>a</sup>	0.48 <sup>b</sup>	1113225_SNP	5A	84.51	0.09	G	0.87	C	0.92
				1864057_SNP	6B	33.46	0.07	C	0.81	M	0.81
LRD (mm)	0.58	0.66 <sup>a</sup>	0.51 <sup>b</sup>	4005012_PAV	1B	51.29	0.10	0	0.67	1	0.93

TRN, total root number. SRA, seminal root angle. PRL, primary root length. LRL, total lateral root length. PRS, primary root surface. LRS, total lateral root surface. PRV, primary root volume. LRV, total lateral root volume. PRD, primary root diameter. LRD, mean lateral root diameter.



**Figure 5.** Marker allele frequency means from landraces within the upper and lower 10th percentile for the analyzed traits. All significant markers shown in Table 4 are included. TRN, total root number. SRA, seminal root angle. PRL, primary root length. LRL, total lateral root length. PRS, primary root surface. LRS, total lateral root surface. PRV, primary root volume. LRV, total lateral root volume. PRD, primary root diameter. LRD, mean lateral root diameter.

### 3.3. Gene Annotation

Of the 176 markers showing significant associations, 31 were identified in the reference sequence of the wheat genome [34] (Table 5). Eight of them were positioned within gene models, whereas, for the rest, the closest gene model to the corresponding marker was taken into consideration. The gene models described in Table 5 included molecules related to abiotic stress resistance, seed formation, carbohydrate remobilization, disease resistance proteins, and other genes involved in different cellular metabolic pathways.

**Table 5.** Gene models within MTA-QTL positions. Only MTAs with markers mapped against the genome sequence are included. Genome position of the gene model is indicated in Mb.

DARtseq Marker	MTA-QTL	Gene Model	Position	Description
1109244_SNP	mtaq-1A.5	TraesCS1A01G363600	540.1	Jacalin lectin family protein
1210090_SNP	mtaq-1A.7	TraesCS1A01G424800	579.8	Cellulose synthase
997799_SNP	mtaq-1B.1	TraesCS1B01G022500	10.1	Protein trichome birefringence
1003552_SNP	mtaq-1B.7	TraesCS1B01G430400	654.8	F-box domain protein
1085277_SNP *	mtaq-2A.3	TraesCS2A01G250600	378.4	9-cis-epoxycarotenoid dioxygenase
1083104_SNP	mtaq-2A.3	TraesCS2A01G281000	469.4	Dynammin-like family protein
1117775_PAV	mtaq-2A.4	TraesCS2A01G541700	752.9	LEA hydroxyproline-rich glycoprotein family
1075469_SNP	mtaq-2B.1	TraesCS2B01G004500	2.4	Cytochrome P450 family protein
1256467_PAV	mtaq-3A.1	TraesCS3A01G018600	11.5	F-box domain protein
1082068_PAV	mtaq-3A.2	TraesCS3A01G034100	19.3	Receptor-like kinase
1130621_PAV	mtaq-3A.5	TraesCS3A01G132300	108.9	Blue copper protein
987263_PAV *	mtaq-3A.7	TraesCS3A01G393600	641.6	Pectin lyase-like superfamily protein
1101009_SNP	mtaq-3B.4	TraesCS3B01G516800	759.9	Ribosomal protein S4
3034109_PAV	mtaq-4A.6	TraesCS4A01G419000	688.9	Histone acetyltransferase of the CBP family 5
1250077_PAV *	mtaq-4B.3	TraesCS4B01G345800	639.4	Basic helix-loop-helix DNA-binding protein
1240561_PAV	mtaq-6A.3	TraesCS6A01G041500	21.7	Transmembrane protein 97
1047867_PAV	mtaq-6A.3	TraesCS6A01G415600	615.3	Cobryric acid synthase
1105573_PAV	mtaq-6A.5	TraesCS6A01G242300	453.9	50S ribosomal protein L19
989287_PAV *	mtaq-6A.6	TraesCS6A01G417400	615.8	F-box domain protein
1129380_PAV *	mtaq-6B.1	TraesCS6B01G000200	0.1	NBS-LRR resistance-like protein
1864057_SNP *	mtaq-6B.3	TraesCS6B01G335600	590.9	Hexosyltransferase
1098568_PAV *	mtaq-6B.4	TraesCS6B01G399700	675.2	bZIP transcription factor family protein
1130796_PAV	mtaq-7A.1	TraesCS7A01G015100	0.0	Mitochondrial pyruvate carrier
2253648_PAV	mtaq-7A.1	TraesCS7A01G016700	7.3	Transmembrane protein DUF594
1139027_PAV	mtaq-7A.1	TraesCS7A01G015400	6.7	Signal peptidase complex catalytic subunit SEC11
1076865_PAV	mtaq-7A.1	TraesCS7A01G024800	9.7	WAT1-related protein
1059554_SNP *	mtaq-7A.3	TraesCS7A01G100600	61.8	GD5L esterase/lipase
1665955_PAV	mtaq-7A.4	TraesCS7A01G442400	636.7	BTB/POZ domain
1149356_PAV	mtaq-7B.1	TraesCS7B01G058300	60.6	Glutamate receptor
1075278_SNP	mtaq-7B.2	TraesCS7B01G378200	642.6	Receptor-like kinase
1252655_PAV	mtaq-7B.3	TraesCS7B01G421300	690.2	NBS-LRR resistance-like protein

\* Markers located within gene models.

## 4. Discussion

Roots exhibit a high level of morphological plasticity in response to soil conditions, which allows plants to better adapt, particularly under drought conditions. Several authors have reported the role of RSA traits in response to drought stress [39,40]. Wasson et al. [11] suggested that a deep root system with the appropriate density along the soil profile would confer an advantage on wheat grown in rainfed agricultural systems. Therefore, identifying new alleles for improving root architecture under drought conditions and introgressing them into adapted phenotypes is a desirable approach for breeding purposes. The current study analyzed a collection of durum wheat landraces representative of the variability existing within the Mediterranean Basin in an attempt to broaden the genetic background present in commercial cultivars.

Evaluating root architecture in the field is a difficult, expensive, and time-consuming assignment, especially when a large number of plants need to be phenotyped. It has been reported that the root geometry of adult plants is strongly related to the seminal root angle (SRA), with deeply rooted wheat genotypes showing a narrower SRA [10]. Different systems have been adopted to enable early screening

of the root system architecture in wheat, assuming that genotypes that differ in root architecture at an early developmental stage would also differ in the field at stages when nutrient and/or water capture become critical for grain yield [12].

#### *4.1. Phenotypic Variation*

The germplasm analyzed in the present study, including mostly durum wheat landraces from the Mediterranean Basin, showed wide variability in RSA traits. The variability found was higher than that observed in other studies using elite accessions [12,14] or even landraces, as reported by Ruiz et al. [41] analyzing a collection of Spanish durum wheat landraces. These results, and the intermediate values obtained for all traits in modern cultivars, support the use of ancient local germplasm for widening the genetic background in breeding programs.

Means comparison of phenotypic traits revealed large differences among SPs associated with their geographical origin. Eastern Mediterranean landraces, collected in the area closest to the origin of tetraploid wheat, showed the largest root size in terms of length, surface, and volume, and the widest root angle. The wheat-growing areas of this region, which comprises Syria, Jordan, Israel, and Egypt, are the warmest and driest within the Mediterranean Basin [42]. In addition, when SRA traits were analyzed separately for the two components of the eastern Balkans and Turkey subpopulation, large differences appeared between them, with Turkish landraces being much more similar to the eastern Mediterranean ones than to the eastern Balkan ones, since the latter showed the lowest values for root length, surface, and volume. Turkish landraces also showed a wide root angle, as did the eastern Mediterranean ones. The differences found in SRA between the eastern Balkans and Turkish landraces are sustained by two lines of evidence. One is the contrasting environmental conditions of the wheat-growing areas of northern Balkan countries and Turkey, since the analysis of long-term climate data demonstrated less rainfall and higher temperatures and solar radiation in the latter [42]. The other is that the northern Balkan landraces likely originated in the steps of southern Russia and the Volga region [2,43], which also suggests contrasting environmental conditions in the zones of origin of the eastern Balkan and Turkish landraces. The phenotypic analysis carried out in the current study revealed that landraces from regions where drought stress is prevalent have a larger root size and a wider root angle. This architecture should allow a larger proportion of the soil to be covered for more efficient water capture, and this hypothesis is supported by correlations between RSA and yield traits. Although low, likely due to the very early stage when the root traits were measured, differences in the number of significant correlations were observed between the two environments with the highest and lowest water input reported by Roselló et al. [25]. Root size-related traits were positively correlated with the number of grains and spikes per unit area (primary roots) and with grain yield and grain weight (lateral roots) in the driest environment. SRA was negatively correlated with TKW, as reported previously by Canè et al. [12], who concluded that it was due to the influence of the root angle on the distribution of roots in the soil layers, which affects the water uptake from deeper layers. In our study, the genotypes with the narrowest angle corresponded to those from the western Mediterranean countries, which Royo et al. [42] and Soriano et al. [19] reported to have heavier grains.

#### *4.2. Marker-Trait Associations*

The current study attempts to dissect the genetic architecture controlling the seminal root system in a collection of landraces from the Mediterranean Basin by association analysis. A mixed linear model accounting for the genetic relatedness between cultivars (random effect) and their population structure (fixed effect) (K + Q model) was used in order to reduce the number of spurious associations.

A total of 233 significant associations were identified for the 10 RSA traits underlying the complex genetic control of RSA. However, in order to simplify this information and to integrate closely linked MTAs in the same QTL, those MTAs located within LD blocks were considered as belonging to the same MTA-QTL. As a result, the number of genome regions involved in RSA was reduced to 82. The relationships between RSA and yield-related traits was also suggested by the presence of pleiotropic

MTA-QTLs. The comparison of the genome regions identified in the current study with those related to yield and yield components by Roselló et al. [44] showed that 45% of the RSA MTA-QTLs were located with yield-related trait MTA-QTLs. These results are in agreement with the findings of Canè et al. [12], who found that 30% of the RSA-QTLs affected agronomic traits, which provided evidence of the implications of RSA in field performance of durum wheat at early growth stages.

In the last few years, GWAS for RSA have been limited in comparison with QTL mapping for root traits based on bi-parental populations (see Soriano and Álvaro [13] for a review). A comparison with previous studies reporting MTAs for RSA resulted in several common regions with the current study. Three common regions were found with the study of Canè et al. [12], but different traits were included for MTAs in those QTLs (mtaq-3A.3, mtaq-3A.5, mtaq-3A.6, and mtaq-6B.2). Two MTAs were in common with those reported by Ayalew et al. [15], who identified five significant associations with root length under stress (2) and non-stress (3) conditions. The MTA reported under stress conditions in chromosome 2B may correspond with mtaq-2B.2, which also shows an association with LRL. However, the association on chromosome 3B, although in a common region with mtaq-3B.4, differed in RSA. When MTA-QTLs were compared with QTLs from bi-parental populations, twelve genomic regions were located within the meta-QTL positions defined by Soriano and Álvaro [13] after the compilation of 754 QTLs from 30 studies.

Candidate genes at the MTA peak were sought using the high-confidence gene annotation from the wheat genome sequence [34]. Among these genes, those involved in plant growth and development as well as tolerance to abiotic stresses may be of special interest. On chromosome 1A, the marker 1210090\_SNP in mtaq-1A.7 is located close to a cellulose synthase gene. This type of gene is involved in plant cell growth and structure [45]. A trichome birefringence (TB) protein was identified in mtaq-1B.1. According to Zhu et al. [46], the TB-like27 protein mutants in *Arabidopsis* increased aluminium accumulation in cell walls, which inhibited root elongation through structural and functional damage. Three peaks corresponded with F-box domains located in mtaq-1B.7, mtaq-3A.1, and mtaq-6A.6. According to Hua et al. [47], this is the protein subunit of E3 ubiquitin ligases involved in the response to abiotic stresses. Li et al. [48] overexpressed the F-box *TaFBA1* in transgenic tobacco to improve heat tolerance, and one of the results was increased root length in the transgenic plants. 9-cis-epoxycarotenoid dioxygenase (NCED) is a key enzyme in the biosynthesis of ABA in higher plants, which regulates the response to various environmental stresses [49]. This enzyme is located within mtaq-2A.3. In mtaq-2A.4, the marker 1117775\_PAV corresponded with a late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein. These proteins play a role in the response to abiotic stresses. They are mainly accumulated in seeds, but have been found in roots during the whole developmental cycle [50]. The marker 1098568\_PAV, in mtaq-6B.4, is located within a gene coding a bZIP transcription factor family protein. This type of transcription factor is involved in abiotic stresses [51]. Zhang et al. [52] observed that the root growth of transgenic plants overexpressing the gene *TabZIP14-B* was hindered more severely than that of the control plants. Another gene involved in abiotic stress tolerance is the mitochondrial pyruvate carrier (MPC) located in mtaq-7A.1 [53]. This gene is involved in cadmium tolerance in *Arabidopsis*, which prevents its accumulation. Roots are the predominant plant tissue for cadmium absorption or exclusion. He et al. [53] found that the root length of mutant plants of *Arabidopsis* for MPC genes was substantially shorter than the wild-type plants. A protein related to WAT1 (WALLS ARE THIN1) involved in secondary cell wall thickness [54] is located in the peak of mtaq-7A.1.

## 5. Conclusions

Including local landraces in breeding programs is a useful approach to broadening the genetic variability of crops [3]. The variability for root system architecture traits found in Mediterranean landraces and the high number of genome regions controlling them—most of them not reported previously—makes this germplasm a valuable source for root architecture improvement. The identification of extreme genotypes for root architecture traits can help identify parents for the

development of new mapping populations to tackle a map-based cloning approach to the genes of interest. In the present study, we identified the molecular markers linked to these genotypes with different allele composition that facilitate the introgression of the corresponding traits through marker-assisted breeding.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4395/9/7/364/s1>, Table S1: Accessions included in the study, Table S2: Statistics of the seminal RSA traits, Table S3: Means comparison of seminal root system architecture traits for the eastern Balkans (EB) and Turkish durum wheat landraces. Means within columns with different letters are significantly different at  $p < 0.05$  following a Tukey test, Table S4: Linkage disequilibrium decay plots. (A) Genome A. (B) Genome B. The LOESS curve is represented in blue. The horizontal red line corresponds to the  $r^2$  mean for each genome, Table S5: Significant markers associated with seminal root system architecture traits obtained in 160 durum wheat Mediterranean landraces.

**Author Contributions:** Conceptualization, M.R. and J.M.S.; Methodology, M.R., M.S.-G., and J.M.S.; Formal Analysis, M.R.; Investigation, M.R., M.S.-G., J.M.S.; Resources, C.R.; Data Curation, M.R., C.R.; Writing—Original Draft Preparation, M.R., J.M.S.; Writing—Review & Editing, C.R., M.S.-G., and J.M.S.; Visualization, J.M.S.; Supervision, C.R., M.S.-G., and J.M.S.; Project Administration, C.R. and J.M.S.; Funding Acquisition, C.R. and J.M.S.

**Funding:** This research was funded by Spanish Ministry of Science, Innovation and Universities (<http://www.ciencia.gob.es/>), grant numbers AGL-2012-37217 (C.R.) and AGL2015-65351-R (J.M.S. and C.R.).

**Acknowledgments:** Projects AGL-2012-37217 (C.R.) and AGL2015-65351-R (J.M.S. and C.R.) of the Spanish Ministry of Science, Innovation and Universities (<http://www.ciencia.gob.es/>) funded this study. The authors acknowledge the contribution of the CERCA Program (Generalitat de Catalunya). M.R. is a recipient of a PhD grant from the Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA). J.M.S. was hired by the INIA-CCAA program funded by INIA and the Generalitat de Catalunya.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

BP	Before Present
DARtseq	Diversity Arrays Technology sequencing
EB + T	Eastern Balkans and Turkey
EM	Eastern Mediterranean
FDR	False Discovery Rate
GWAS	Genome Wide Association Study
GY	Grain Yield
LRD	Lateral Roots Diameter
LRL	Lateral Roots Length
LRS	Lateral Roots Surface
LRV	Lateral Roots Volume
MTA	Marker-Trait Association
NGm <sup>2</sup>	Number of Grains per square meter
NSm <sup>2</sup>	Number of Spikes per square meter
PAV	Presence/Absence Variants
PRD	Primary Root Diameter
PRL	Primary Root Length
PRS	Primary Root Surface
PRV	Primary Root Volume
PVE	Phenotypic Variance Explained
QTL	Quantitative Trait Loci
RSA	Root System Architecture
SNP	Single Nucleotide Polymorphism
SP	Subpopulation
SRA	Seminal Root Angle
TKW	Thousand Kernel Weight
TRN	Total Root Number
WB + E	Western Balkans and Egypt
WM	Western Mediterranean

## References

1. Feldman, M. Origin of cultivated wheat. In *The World Wheat Book: A History of Wheat Breeding*; Bonjean, A.P., Angus, W.J., Eds.; Lavoisier Publishing: Paris, France, 2001; pp. 3–56. ISBN 1898298726.
2. Nazco, R.; Villegas, D.; Ammar, K.; Peña, R.J.; Moragues, M.; Royo, C. Can Mediterranean durum wheat landraces contribute to improved grain quality attributes in modern cultivars? *Euphytica* **2012**, *185*, 1–17. [[CrossRef](#)]
3. Lopes, M.S.; El-Basyoni, I.; Baenziger, P.S.; Singh, S.; Royo, C.; Ozbek, K.; Aktas, H.; Ozer, E.; Ozdemir, F.; Manickavelu, A.; et al. Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change. *J. Exp. Bot.* **2015**, *66*, 3477–3486. [[CrossRef](#)] [[PubMed](#)]
4. Reynolds, M.P.; Mujeeb-Kazi, A.; Sawkins, M. Prospects for utilising plant-adaptive mechanisms to improve wheat and other crops in drought- and salinity-prone environments. *Ann. Appl. Biol.* **2005**, *146*, 239–259. [[CrossRef](#)]
5. Maccaferri, M.; Sanguineti, M.C.; Demontis, A.; El-Ahmed, A.; García del Moral, L.; Maalouf, F.; Nachit, M.; Nserallah, N.; Ouabbou, H.; Rhouma, S.; et al. Association mapping in durum wheat grown across a broad range of water regimes. *J. Exp. Bot.* **2011**, *62*, 409–438. [[CrossRef](#)] [[PubMed](#)]
6. Royo, C.; García del Moral, L.; Slafer, G.A.; Nachit, M.; Araus, J.L. Selection tools for improving yield-associated physiological traits. In *Durum Wheat Breeding: Current Approaches and Future Strategies*; Royo, C., Nachit, M.M., Di Fonzo, N., Araus, J., Pfeiffer, W., Slafer, G., Eds.; Food Products Press: New York, NY, USA, 2005; pp. 563–598.
7. Royo, C.; Villegas, D. Field measurements of canopy spectra for biomass assessment of small-grain cereals. In *Biomass—Detection, Production and Usage*; Darko, M., Ed.; InTech Open: London, UK, 2011; pp. 27–52. ISBN 978-953-307-492-4.
8. Tuberosa, R. Phenotyping for drought tolerance of crops in the genomics era. *Front. Physiol.* **2012**, *3*, 347. [[CrossRef](#)] [[PubMed](#)]
9. Løes, A.-K.; Gahoonia, T.S. Genetic variation in specific root length in Scandinavian wheat and barley accessions. *Euphytica* **2004**, *137*, 243–249. [[CrossRef](#)]
10. Manschadi, A.M.; Hammer, G.L.; Christopher, J.T.; Manschadi, A.M.; Hammer, G.L.; Christopher, J.T. Genotypic variation in seedling root architectural traits and implications for drought adaptation in wheat (*Triticum aestivum* L.). *Plant Soil* **2008**, *303*, 115–129. [[CrossRef](#)]
11. Wasson, A.P.; Richards, R.A.; Chatrath, R.; Misra, S.C.; Prasad, S.V.S.; Rebetzke, G.J.; Kirkegaard, J.A.; Christopher, J.; Watt, M. Traits and selection strategies to improve root systems and water uptake in water-limited wheat crops. *J. Exp. Bot.* **2012**, *63*, 3485–3498. [[CrossRef](#)]
12. Canè, M.A.; Maccaferri, M.; Nazemi, G.; Salvi, S.; Francia, R.; Colalongo, C.; Tuberosa, R. Association mapping for root architectural traits in durum wheat seedlings as related to agronomic performance. *Mol. Breed.* **2014**, *34*, 1629–1645. [[CrossRef](#)]
13. Soriano, J.M.; Álvaro, F. Discovering consensus genomic regions in wheat for root-related traits by QTL meta-analysis. *Sci. Rep.* **2019**, in press.
14. Sanguineti, M.C.; Li, S.; Maccaferri, M.; Corneti, S.; Rotondo, F.; Chiari, T.; Tuberosa, R. Genetic dissection of seminal root architecture in elite durum wheat germplasm. *Ann. Appl. Biol.* **2007**, *151*, 291–305. [[CrossRef](#)]
15. Ayalew, H.; Liu, H.; Börner, A.; Kobiljski, B.; Liu, C.; Yan, G. Genome-wide association mapping of major root length QTLs under PEG induced water stress in wheat. *Front. Plant Sci.* **2018**, *9*, 1759. [[CrossRef](#)] [[PubMed](#)]
16. Alahmad, S.; El Hassouni, K.; Bassi, F.M.; Dinglasan, E.; Youssef, C.; Quarry, G.; Aksoy, A.; Mazzucotelli, E.; Juhász, A.; Able, J.A.; et al. A major root architecture QTL responding to water limitation in durum wheat. *Front. Plant Sci.* **2019**, *10*, 436. [[CrossRef](#)] [[PubMed](#)]
17. Beyer, S.; Daba, S.; Tyagi, P.; Bockelman, H.; Brown-Guedira, G.; IWGSC; Mohammadi, M. Loci and candidate genes controlling root traits in wheat seedlings—A wheat root GWAS. *Funct. Integr. Genom.* **2019**, *19*, 91–107. [[CrossRef](#)] [[PubMed](#)]
18. Li, L.; Peng, Z.; Mao, X.; Wang, J.; Chang, X.; Reynolds, M.; Jing, R. Genome-wide association study reveals genomic regions controlling root and shoot traits at late growth stages in wheat. *Ann. Bot.* **2019**, 1–14. [[CrossRef](#)]

19. Soriano, J.; Villegas, D.; Aranzana, M.; García del Moral, L.; Royo, C. Genetic structure of modern durum wheat cultivars and mediterranean landraces matches with their agronomic performance. *PLoS ONE* **2016**, *11*, e0160983. [[CrossRef](#)] [[PubMed](#)]
20. Rahnama, A.; Munns, R.; Poustini, K.; Watt, M. A screening method to identify genetic variation in root growth response to a salinity gradient. *J. Exp. Bot.* **2011**, *62*, 69–77. [[CrossRef](#)] [[PubMed](#)]
21. Watt, M.; Moosavi, S.; Cunningham, S.C.; Kirkegaard, J.A.; Rebetzke, G.J.; Richards, R.A. A rapid, controlled-environment seedling root screen for wheat correlates well with rooting depths at vegetative, but not reproductive, stages at two field sites. *Ann. Bot.* **2013**, *112*, 447–455. [[CrossRef](#)] [[PubMed](#)]
22. Lobet, G.; Pagès, L.; Draye, X. A Novel image-analysis toolbox enabling quantitative analysis of root system architecture. *Breakthr. Technol.* **2011**, *157*, 29–39. [[CrossRef](#)] [[PubMed](#)]
23. Richard, C.A.I.; Hickey, L.T.; Fletcher, S.; Jennings, R.; Chenu, K.; Christopher, J.T. High-throughput phenotyping of seminal root traits in wheat. *Plant Methods* **2015**, *11*, 13. [[CrossRef](#)] [[PubMed](#)]
24. Abramoff, M.D.; Magalhaes, P.J.; Ram, S.J. Image Processing with ImageJ. *Biophotonics Int.* **2004**, *11*, 36–42.
25. Roselló, M.; Villegas, D.; Álvaro, F.; Soriano, J.M.; Lopes, M.S.; Nazco, R.; Royo, C. Unravelling the relationship between adaptation pattern and yield formation strategies in Mediterranean durum wheat landraces. *Eur. J. Agron.* **2019**, *107*, 43–52. [[CrossRef](#)]
26. Tukey, J.W. Comparing individual means in the analysis of variance. *Biometrics* **1949**, *5*, 99–114. [[CrossRef](#)] [[PubMed](#)]
27. Little, R.; Milliken, G.; Stroup, W.; Wolfinger, R. SAS system for mixed models. *Technometrics* **1997**, *39*, 344.
28. SAS Institute. *SAS Enterprise Guide*; SAS Institute: Cary, NC, USA, 2014.
29. Doyle, J.; Doyle, J. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **1987**, *19*, 11–15.
30. Sansaloni, C.; Petrolini, C.; Jaccoud, D.; Carling, J.; Detering, F.; Grattapaglia, D.; Kilian, A. Diversity arrays technology (DArT) and next-generation sequencing combined: Genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. *BMC Proc.* **2011**, *5*, 54. [[CrossRef](#)]
31. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [[CrossRef](#)] [[PubMed](#)]
32. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **1995**, *57*, 289–300. [[CrossRef](#)]
33. Voorrips, R.E. MapChart: Software for the graphical presentation of linkage maps and QTLs. *J. Hered.* **2002**, *93*, 77–78. [[CrossRef](#)]
34. IWGSC. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **2018**, *361*, eaar7191. [[CrossRef](#)]
35. Mwadzingeni, L.; Shimelis, H.; Rees, D.J.G.; Tsilo, T.J. Genome-wide association analysis of agronomic traits in wheat under drought-stressed and non-stressed conditions. *PLoS ONE* **2017**, *12*, e0171692. [[CrossRef](#)] [[PubMed](#)]
36. Wang, S.-X.; Zhu, Y.-L.; Zhang, D.-X.; Shao, H.; Liu, P.; Hu, J.-B.; Zhang, H.; Zhang, H.-P.; Chang, C.; Lu, J.; et al. Genome-wide association study for grain yield and related traits in elite wheat varieties and advanced lines using SNP markers. *PLoS ONE* **2017**, *12*, e0188662. [[CrossRef](#)] [[PubMed](#)]
37. Mangini, G.; Gadaleta, A.; Colasuonno, P.; Marcotuli, I.; Signorile, A.M.; Simeone, R.; De Vita, P.; Mastrangelo, A.M.; Laidò, G.; Pecchioni, N.; et al. Genetic dissection of the relationships between grain yield components by genome-wide association mapping in a collection of tetraploid wheats. *PLoS ONE* **2018**, *13*, e0190162. [[CrossRef](#)] [[PubMed](#)]
38. Sukumaran, S.; Reynolds, M.P.; Sansaloni, C. Genome-wide association analyses identify QTL hotspots for yield and component traits in durum wheat grown under yield potential, drought, and heat stress environments. *Front. Plant Sci.* **2018**, *9*, 81. [[CrossRef](#)] [[PubMed](#)]
39. Christopher, J.; Christopher, M.; Jennings, R.; Jones, S.; Fletcher, S.; Borrell, A.; Manschadi, A.M.; Jordan, D.; Mace, E.; Hammer, G. QTL for root angle and number in a population developed from bread wheats (*Triticum aestivum*) with contrasting adaptation to water-limited environments. *Theor. Appl. Genet.* **2013**, *126*, 1563–1574. [[CrossRef](#)] [[PubMed](#)]
40. Paez-García, A.; Motes, C.; Scheible, W.-R.; Chen, R.; Blancaflor, E.; Monteros, M. Root traits and phenotyping strategies for plant improvement. *Plants* **2015**, *4*, 334–355. [[CrossRef](#)] [[PubMed](#)]

41. Ruiz, M.; Giraldo, P.; González, J.M. Phenotypic variation in root architecture traits and their relationship with eco-geographical and agronomic features in a core collection of tetraploid wheat landraces (*Triticum turgidum* L.). *Euphytica* **2018**, *214*, 54. [[CrossRef](#)]
42. Royo, C.; Nazco, R.; Villegas, D. The climate of the zone of origin of Mediterranean durum wheat (*Triticum durum* Desf.) landraces affects their agronomic performance. *Genet. Resour. Crop Evol.* **2014**, *61*, 1345–1358. [[CrossRef](#)]
43. Dedkova, O.S.; Badaeva, E.D.; Amosova, A.V.; Martynov, S.P.; Ruanet, V.V.; Mitrofanova, O.P.; Pukhal'skiy, V.A. Diversity and the origin of the European population of *Triticum dicoccum* (Schrank) Schuebl. As revealed by chromosome analysis. *Russ. J. Genet.* **2009**, *45*, 1082–1091. [[CrossRef](#)]
44. Roselló, M.; Royo, C.; Sansaloni, C.; Soriano, J.M. GWAS for yield and related traits under rainfed mediterranean environments revealed different genetic architecture in pre- and post-green revolution durum wheat collections. *Front. Plant Sci.* **2019**, in press.
45. Lei, L.; Li, S.; Gu, Y. Cellulose synthase complexes: Composition and regulation. *Front. Plant Sci.* **2012**, *3*, 75. [[CrossRef](#)] [[PubMed](#)]
46. Zhu, X.F.; Sun, Y.; Zhang, B.C.; Mansoori, N.; Wan, J.X.; Liu, Y.; Wang, Z.W.; Shi, Y.Z.; Zhou, Y.H.; Zheng, S.J. *Trichome Birefringence-Like27* affects aluminum sensitivity by modulating the O-acetylation of xyloglucan and aluminum-binding capacity in *Arabidopsis*. *Plant Physiol.* **2014**, *166*, 181–189. [[CrossRef](#)] [[PubMed](#)]
47. Hua, Z.; Vierstra, R.D. The cullin-RING ubiquitin-protein ligases. *Annu. Rev. Plant Biol.* **2011**, *62*, 299–334. [[CrossRef](#)] [[PubMed](#)]
48. Li, Q.; Wang, W.; Wang, W.; Zhang, G.; Liu, Y.; Wang, Y.; Wang, W. Wheat f-box protein gene *TaFBA1* is involved in plant tolerance to heat stress. *Front. Plant Sci.* **2018**, *9*, 521. [[CrossRef](#)] [[PubMed](#)]
49. Zhang, S.J.; Song, G.Q.; Li, Y.L.; Gao, J.; Liu, J.J.; Fan, Q.Q.; Huang, C.Y.; Sui, X.X.; Chu, X.S.; Guo, D.; et al. Cloning of 9-*cis*-epoxycarotenoid dioxygenase gene (*TaNCED1*) from wheat and its heterologous expression in tobacco. *Biol. Plant.* **2014**, *58*, 89–98. [[CrossRef](#)]
50. Gao, J.; Lan, T. Functional characterization of the late embryogenesis abundant (LEA) protein gene family from *Pinus tabulaeformis* (Pinaceae) in *Escherichia coli*. *Sci. Rep.* **2016**, *6*, 19467. [[CrossRef](#)]
51. Somaraj, P.; Luang, S.; Lopato, S.; Hrmova, M. Basic leucine zipper (bZIP) transcription factors involved in abiotic stresses: A molecular model of a wheat bZIP factor and implications of its structure in function. *Biochim. Biophys. Acta* **2016**, *1860*, 46–56. [[CrossRef](#)]
52. Zhang, W.; Chen, S.; Abate, Z.; Nirmala, J.; Rouse, M.N.; Dubcovsky, J. Identification and characterization of *Sr13*, a tetraploid wheat gene that confers resistance to the Ug99 stem rust race group. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E9483–E9492. [[CrossRef](#)]
53. He, L.; Jing, Y.; Shen, J.; Li, X.; Liu, H.; Geng, Z.; Wang, M.; Li, Y.; Chen, D.; Gao, J.; et al. Mitochondrial pyruvate carriers prevent cadmium toxicity by sustaining the TCA cycle and glutathione synthesis. *Plant Physiol.* **2019**, *180*, 198–211. [[CrossRef](#)]
54. Ranocha, P.; Dima, O.; Nagy, R.; Felten, J.; Corratgé-Faillie, C.; Novák, O.; Morreel, K.; Lacombe, B.; Martinez, Y.; Pfrunder, S.; et al. *Arabidopsis* WAT1 is a vacuolar auxin transport facilitator required for auxin homeostasis. *Nat. Commun.* **2013**, *4*, 2625. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



Article

# SSR Marker-Assisted Management of Parental Germplasm in Sugarcane (*Saccharum* spp. hybrids) Breeding Programs

Jiantao Wu <sup>1,\*†</sup>, Qinnan Wang <sup>1,†</sup>, Jing Xie <sup>1</sup>, Yong-Bao Pan <sup>2</sup>, Feng Zhou <sup>1</sup>, Yuqiang Guo <sup>1</sup>, Hailong Chang <sup>1</sup>, Huanying Xu <sup>1</sup>, Wei Zhang <sup>1</sup>, Chuiming Zhang <sup>1</sup> and Yongsheng Qiu <sup>1,\*</sup>

<sup>1</sup> Guangdong Provincial Bioengineering Institute (Guangzhou Sugarcane Industry Research Institute), Guangzhou 510316, China

<sup>2</sup> Sugarcane Research Unit, USDA-ARS, Houma, LA 70360, USA

\* Correspondence: wujiantao2010@163.com (J.W.); siriwu@126.com (Y.Q.); Tel.: +86-18789656100 (J.W.); +86-089888704031 (Y.Q.)

† These authors contributed equally to this work.

Received: 11 July 2019; Accepted: 10 August 2019; Published: 14 August 2019

**Abstract:** Sugarcane (*Saccharum* spp. hybrids) is an important sugar and bioenergy crop with a high aneuploidy, complex genomes and extreme heterozygosity. A good understanding of genetic diversity and population structure among sugarcane parental lines is a prerequisite for sugarcane improvement through breeding. In order to understand genetic characteristics of parental lines used in sugarcane breeding programs in China, 150 of the most popular accessions were analyzed with 21 fluorescence-labeled simple sequence repeats (SSR) markers and high-performance capillary electrophoresis (HPCE). A total of 226 SSR alleles of high-resolution capacity were identified. Among the series obtained from different origins, the YC-series, which contained eight unique alleles, had the highest genetic diversity. Based on the population structure analysis, the principal coordinate analysis (PCoA) and phylogenetic analysis, the 150 accessions were clustered into two distinct sub-populations (Pop1 and Pop2). Pop1 contained the majority of clones introduced to China (including 28/29 CP-series accessions) while accessions native to China clustered in Pop2. The analysis of molecular variance (AMOVA), fixation index (*F<sub>st</sub>*) value and gene flow (*N<sub>m</sub>*) value all indicated the very low genetic differentiation between the two groups. This study illustrated that fluorescence-labeled SSR markers combined with high-performance capillary electrophoresis (HPCE) could be a very useful tool for genotyping of the polyploidy sugarcane. The results provided valuable information for sugarcane breeders to better manage the parental germplasm, choose the best parents to cross, and produce the best progeny to evaluate and select for new cultivar(s).

**Keywords:** sugarcane; parental line; population structure; plant breeding; genetic diversity; simple sequence repeats (SSR)

## 1. Introduction

Sugarcane cultivars are allopolyploids with highly heterozygous and complex genomes, which render a slow progress in breeding. To date, most commercial sugarcane varieties can be traced back to a limited number of popular cultivars belonging to either the POJ- or Co-series, which represent a very narrow genetic base [1]. Therefore, it is important for sugarcane breeders to fully understand the genetic relationship among parental lines and to choose elite parents of different genetic background for crossing in order to broaden the genetic diversity of sugarcane population [2].

Hainan sugarcane breeding station (HSBS) is the primary sugarcane crossing facility in Mainland China. It produces nearly all the seeds for sugarcane breeders in China every year [3]. HSBS has

more than 2000 germplasm materials. Currently, thousands of new elite sugarcane genotypes are created by breeders each year. The utilization of these ever-increasing germplasm materials is a daunting challenge. Parental selection is a crucial step for good quality cross-breeding. Therefore, breeding materials should be adequately evaluated by different analytical methods to ensure their genetic suitability.

In the past, sugarcane breeders studied the genetic differences of parents mainly from the aspects of the genetic relationship, geographical origin and morphology. The genetic differences of sugarcane parents cannot really be reflected by pedigree because of mixed pollen, selfing and seed admixture [4]. Although morphological traits can be evaluated, these traits are easily influenced by the environment and may not reflect the real genetic diversity of sugarcane germplasm resources [5]. DNA molecular markers with high stability, multiple quantity and high polymorphism are more suitable for evaluating sugarcane germplasm collection [1]. With the rapid development of biotechnology, sugarcane researchers have utilized different types of DNA molecular markers, including amplified fragment length polymorphisms (AFLP) [1,5], restriction fragment length polymorphisms (RFLP) [6,7], random amplification of polymorphic DNAs (RAPD) [8,9], single nucleotide polymorphism (SNP) [10], simple sequence repeats (SSRs) [11], inter simple sequence repeat (ISSRs) [12,13], expressed sequence tag-simple sequence repeat (EST-SSRs) [14–16], 5S rRNA intergenic spacers [17], start codon targeted (SCoT) [18], target region amplification polymorphism (TRAP) [5,19,20], and cleaved amplified polymorphism sequences (CAPS) [21] for evaluating sugarcane germplasm.

Among PCR-based markers, SSR (microsatellite) markers are considered one of the most efficient markers for plant breeding due to large quantity, low dosage, co-dominant, reliability and multi-allele detecting [22]. SSR markers have been used widely to study sugarcane genetic diversity and population structure [22–24], variety identity [25], genetic map [26,27], and genetic association [28–30]. Furthermore, fluorescence-labeled SSR markers combined with high-performance capillary electrophoresis (HPCE) have manifested better performance in genotyping of polyploid sugarcane, due to higher accuracy and better detection power [22–24,31–37].

Now, this paper reports a study that was designed to manage the parental germplasm of the sugarcane breeding programs in China through the microsatellite (SSR) DNA fingerprinting using fluorescence-labeled SSR primers and the high-performance capillary electrophoresis (HPCE) system. The results will help sugarcane breeders better manage the parental germplasm, choose cross parents, design cross combinations, and produce high quality seedlings for the selection and development of elite varieties.

## **2. Materials and Methods**

### *2.1. Plant Materials*

One hundred and fifty parental clones were chosen for this study, based on the number of lines used most often in crossing from 2014 to 2018 in all Chinese sugarcane breeding programs (Table 1 and S1). These included 32 of clones from foreign origin, 109 clones from the China Mainland, and nine ROC-series clones from China Taiwan. Among the 32 foreign clones, one was from India (Co-series), 29 were from the U.S. (CP-series) and two were from Thailand (K-series). Among the 109 clones from China Mainland, four were from the Dehong Sugarcane Research Institute, Yunnan Province (DZ-series); 11 were from the Fujian Agriculture and Forestry University, Fujian Province (FN-series); two were from the Jiangxi Sugarcane Research Institute, Jiangxi Province (GN-series); 21 were from the Guangxi Academy of Agricultural Sciences, Guangxi Province (GT-series); six were from the Liucheng Academy of Agricultural Sciences, Guangxi Province (LC-series); six were from the Neijiang Academy of Agricultural Sciences, Sichuan Province (NJ-series); 18 were from the Hainan Sugarcane Breeding Station of Guangzhou Sugarcane Industry Research Institute, Hainan Province (YC-series); 29 were from the Guangzhou Sugarcane Industry Research Institute, Guangdong Province (YT-series); 10 were from the Yunnan Academy of Agricultural Sciences, Yunnan Province (YZ-series) and two were from

other breeding units in China Mainland (one from Sichuan Research Institute of Sugar Crops, Sichuan Province and one from the Guangdong Academy of Agricultural Sciences, Guangdong Province).

**Table 1.** The 150 sugarcane accessions used in the experiment.

No.	Accession	Series	No.	Accession	Series	No.	Accession	Series
1	Co1001	Co	51	GZ75-65	GN	101	YC06-92	YC
2	CP57-614	CP	52	HoCP00-1142	CP	102	YC07-65	YC
3	CP67-412	CP	53	HoCP00-2218	CP	103	YC07-71	YC
4	CP72-1210	CP	54	HoCP01-517	CP	104	YC09-13	YC
5	CP72-2086	CP	55	HoCP01-564	CP	105	YC71-374	YC
6	CP80-1827	CP	56	HoCP02-610	CP	106	YC94-46	YC
7	CP81-1254	CP	57	HoCP02-623	CP	107	YC97-24	YC
8	CP84-1198	CP	58	HoCP03-704	CP	108	YC97-40	YC
9	CP89-2143	CP	59	HoCP03-708	CP	109	YC98-2	YC
10	CP93-1382	CP	60	HoCP03-716	CP	110	YC98-27	YC
11	CP93-1634	CP	61	HoCP05-902	CP	111	YN73-204	YN
12	CP94-1100	CP	62	HoCP07-612	CP	112	YT00-236	YT
13	CT89-103	CT	63	HoCP07-613	CP	113	YT00-318	YT
14	DZ03-83	DZ	64	HoCP07-617	CP	114	YT00-319	YT
15	DZ05-61	DZ	65	HoCP91-555	CP	115	YT01-120	YT
16	DZ06-51	DZ	66	HoCP92-648	CP	116	YT01-125	YT
17	DZ93-88	DZ	67	HoCP93-746	CP	117	YT01-71	YT
18	FN02-6404	FN	68	HoCP95-988	CP	118	YT03-373	YT
19	FN02-6427	FN	69	K5	K	119	YT03-393	YT
20	FN05-2848	FN	70	K86-110	K	120	YT85-177	YT
21	FN0711	FN	71	LC03-1137	LC	121	YT86-368	YT
22	FN0712	FN	72	LC03-182	LC	122	YT89-240	YT
23	FN0713	FN	73	LC04-256	LC	123	YT91-976	YT
24	FN0717	FN	74	LC05-128	LC	124	YT92-1287	YT
25	FN91-23	FN	75	LC05-136	LC	125	YT93-124	YT
26	FN92-4621	FN	76	LC05-291	LC	126	YT93-159	YT
27	FN95-1702	FN	77	LCP85-384	CP	127	YT94-128	YT
28	FN99-20169	FN	78	NJ00-118	NJ	128	YT96-86	YT
29	GN95-108	GN	79	NJ00-15	NJ	129	YT97-20	YT
30	GT00-122	GT	80	NJ03-218	NJ	130	YT97-76	YT
31	GT02-1156	GT	81	NJ07-13	NJ	131	YT99-66	YT
32	GT02-208	GT	82	NJ86-117	NJ	132	YZ02-2540	YZ
33	GT02-281	GT	83	NJ92-244	NJ	133	YZ02-588	YZ
34	GT02-467	GT	84	ROC1	ROC	134	YZ03-194	YZ
35	GT02-761	GT	85	ROC10	ROC	135	YZ07-100	YZ
36	GT02-901	GT	86	ROC16	ROC	136	YZ07-49	YZ
37	GT03-11	GT	87	ROC20	ROC	137	YZ89-7	YZ
38	GT03-1403	GT	88	ROC22	ROC	138	YZ94-343	YZ
39	GT03-2112	GT	89	ROC23	ROC	139	YZ94-375	YZ
40	GT03-3005	GT	90	ROC25	ROC	140	YZ99-601	YZ
41	GT03-3089	GT	91	ROC26	ROC	141	YZ99-91	YZ
42	GT03-8	GT	92	ROC28	ROC	142	ZZ33	YT
43	GT03-91	GT	93	YC04-55	YC	143	ZZ41	YT
44	GT05-3084	GT	94	YC05-64	YC	144	ZZ43	YT
45	GT05-3595	GT	95	YC06-111	YC	145	ZZ45	YT
46	GT73-167	GT	96	YC06-140	YC	146	ZZ49	YT
47	GT89-5	GT	97	YC06-166	YC	147	ZZ50	YT
48	GT92-66	GT	98	YC06-61	YC	148	ZZ80-101	YT
49	GT94-119	GT	99	YC06-63	YC	149	ZZ90-76	YT
50	GT96-154	GT	100	YC06-91	YC	150	ZZ92-126	YT

## 2.2. SSR Genotyping

Young leaf tissues were collected from three individual clones, rinsed with 75% ethanol, and kept at  $-80^{\circ}\text{C}$  prior to DNA extraction. The genomic DNA was extracted from leaf tissues using the cetyl trimethyl ammonium bromide (CTAB) method [38] with minor modifications. The quality and concentration of DNA were measured using the UV-Vis Spectrophotometer Q5000 of Quawell (Quawell Technology, Inc. San Jose, CA, USA) and diluted to  $20\text{ ng}/\mu\text{L}$ . A set of 21 SSR primer pairs (Table 1) with stable and clear amplification was selected from previous reports [3,11,33,39–42]. All forward primers were labeled with a fluorescence dye, 6-carboxy-fluorescein (FAM) or Hexachlorofluorescein (HEX). PCR reactions were performed with the following cycling condition:  $95^{\circ}\text{C}$  for 2 min, followed by 40 cycles of  $94^{\circ}\text{C}$  for 30 s, then primer-specific annealing temperature ( $T_m$ ) for 90 s,  $65^{\circ}\text{C}$  for 30 s, followed by one cycle at  $65^{\circ}\text{C}$  for 10 min. The annealing temperatures for the 21 primer pairs were optimized separately, ranging from  $49^{\circ}\text{C}$  to  $62^{\circ}\text{C}$  (Table 2). The amplified PCR products were checked by a 3% agarose gel electrophoresis. High-performance capillary electrophoreses (HPCE) was conducted on the ABI 3730XL DNA analyzer (Applied Biosystems, Inc. Foster City, CA, USA) to generate GeneScan files. The GeneScan files were analyzed using the GeneMarker V2.2 software (SoftGenetics, LLC. State College, PA, USA) to show SSR DNA fragments (alleles) and the sizes of these fragments were calibrated automatically against the GeneScan500 size standards. Due to the polyploidy nature of sugarcane, the SSR alleles had to be manually called first and the score sheet was manually rechecked according to Pan [43]. The presence of an allele was scored as “1” and its absence scored as “0”. SSR alleles were named using a combination of primer name and allele size.

**Table 2.** The 21 simple sequence repeat (SSR) markers used in this study.

Primer Name	Type <sup>a</sup>	Repeat Motif	Primer Sequence (5'-3')	Annealing Temperatures ( $^{\circ}\text{C}$ )
mSSCIR36	G-SSR	(GA) <sub>18</sub> GT (GA) <sub>4</sub>	CAACAATAACTTAACCTGGTA CTGTCCTTTTATCTCTTT	52
mSSCIR46	G-SSR	(GT) <sub>10</sub>	ATGCTCCGCTTCTCACTC AAGGGGAAAATGAAAACC	52
mSSCIR74	G-SSR	(CGC) <sub>9</sub>	GCGCAAGCCACACTGAGA ACGCAACGCAAAAACAACG	56
SCM4	E-SSR	(CGGAT) <sub>4</sub>	CATTGTCTGTGCTGCT CCGTTTCCCTTCTCTCC	52
SCM7	E-SSR	(GCAC) <sub>4</sub>	ACGGTGCTCTTCACTGCT GGGCATACTTCTCTCTAC	60
SCM18	E-SSR	(ATAC) <sub>3</sub>	CATCAGTATCATTTCATCTTGG CAGTCACAGTCGGGTAGA	60
SMC1825LA	G-SSR	(TG) <sub>11</sub>	CACGTCCTTCCGCCTTGA TCATCGTTCTGCTGCACTG	56
SMC286CS	G-SSR	(TG) <sub>43</sub>	TCAAATGGGACCTTATTGGAG TCCCTCGATCTCCGTTGTT	52
SMC477CG	G-SSR	(CA) <sub>31</sub>	CAAACAACGAATGTGCAATGT CCTGGTTGGCTACCTGTCTTCA	60
SMC486CG	G-SSR	(CA) <sub>14</sub>	GAAATTGCCTCCAGGATTA CCAACCTTGAGAATTGAGATTTCG	60
SMC569CS	G-SSR	(TG) <sub>37</sub>	GCGATGGTTCCTATGCAACTT TTCGTGGCTGAGATTCACTA	60
SMC597CS	G-SSR	(AG) <sub>31</sub>	GCACACCACTCGAATAACGGAT AGTATATCGTCCCTGGCATCA	52
SMC334BS	G-SSR	(TG) <sub>36</sub>	CAATTCTGACCGTCAAAGAT CGATGAGCTTGATGCGAATG	60
SMC36BUQ	G-SSR	(TTG) <sub>7</sub>	GGGTTTCATCTTAGCTACC TCAGTAGCAGAGTCAGACGCTT	56

Table 2. Cont.

Primer Name	Type <sup>a</sup>	Repeat Motif	Primer Sequence (5'-3')	Annealing Temperatures (°C)
SMC7CUQ	G-SSR	(CA) <sub>10</sub> (C) <sub>4</sub>	GCCAAAGCAAGGGTCACTAGA AGTCTATCAGTTGAAACCGA	60
SEGM285	G-SSR	(GCAC) <sub>4</sub>	AAGAAGAAGACTGAGAAGAACACT TAGCAACAACCTAATTTAGCAATC	56
UGSM345	E-SSR	(TG) <sub>6</sub>	CTGTACTGGTATTACATGTGACCT TCTACTAATCACAAGAGAAGATGC	60
UGSM10	E-SSR	(GGC) <sub>11</sub>	GCTACTATGGACAACAGGG ATGAAGAGACGAGACGAAGA	56
UGSuM50	E-SSR	(TC) <sub>14</sub>	CTACTGCCGAGGAAAAGATCG GGAAAAGTTTGTGGCAAGGA	56
MCSA068G08	E-SSR	(CAG) <sub>6</sub>	CTAATGCCATGCCCCAGAGG GCTGGTGATGTCGCCCATCT	56
MCSA176C01	E-SSR	(GGT) <sub>5</sub>	GAGTCAGTTGGTGCCGAGATTG GAACAGGTTAAAGCCCATGTC	56

<sup>a</sup> G-SSR: SSR primer pair designed from genomic sequence; E-SSR: SSR primer pair designed from UniGene or cDNA sequences.

### 2.3. Genetic Diversity Analysis

Qualitative allelic data matrix was constructed and formatted using the DataFormatter software [44]. The PowerMarker v3.25 software [45] was used to calculate allele frequency, number of alleles per locus, polymorphism information content (PIC), the gene diversity index ( $h$ ), Shannon's information index ( $I$ ), and percentage of polymorphic loci ( $PPL$ ) of each marker. The resolving power of the primer ( $R_p$ ) [46] was calculated using allele frequencies. The probability of identity ( $PI$ ) [23] was computed using the CERVUS v3.0 software [47]. Unique (Series-specific) alleles were estimated using GeneALEx v6.502 [48,49].

### 2.4. Population Structure Analysis

The model-based program Structure v2.3.4 [50] was used to analyze the population structure involving the 226 alleles amplified by the 21 SSR primer pairs. The number of populations ( $K$ ) was set from one to 10, and at each  $K$  value, ten runs were conducted separately with 50,000 iterations of burn-in length and 50,000 Markov Chain Monte Carlo (MCMC). Then, the best  $K$  value was estimated using Evanno's  $\Delta K$  method [51] with an online tool, Structure Harvester [52]. An individual  $Q$  matrix was generated by CLUMPP v1.1.2 [53]. Parental clones with membership probabilities greater than 0.5 were identified as the same group [54]. A Principal Coordinate Analysis (PCoA) map was generated based on the genetic distances between pairs of clones by GeneALEx v6.502 [48,49]. An unrooted phylogenetic tree was constructed based on the neighbor-joining (NJ) method and the genetic distance matrix using PowerMarker v3.25 [45] and adjusted with MEGA v6.06 [55].

### 2.5. Differentiation Analysis and Genetic Diversity Indices

Analysis of Molecular Variance (AMOVA) was conducted to find the genetic differentiation within and among subpopulations using GeneALEx v6.502 [48,49]. From AMOVA, the fixation index ( $F_{st}$ ) and gene flow ( $Nm$ ) within the population was also acquired. In addition, genetic diversity indices, including number of different alleles ( $N_a$ ), number of effective alleles ( $N_e$ ), Shannon's information index ( $I$ ), observed heterozygosity ( $H_o$ ), expected heterozygosity ( $H_e$ ), unbiased expected heterozygosity ( $uHe$ ), and percentage of polymorphic loci ( $PPL$ ) of different sub-groups were also calculated using GeneALEx v6.502 [48,49].

### 3. Results

#### 3.1. Polymorphism Revealed by SSR Genotyping

The 21 SSR primer pairs amplified a total of 226 alleles with an average of 10.8 alleles per primer pair (Table 2). Of the 226 alleles, 220 alleles were polymorphic and the other six alleles could be amplified in each clone. The number of alleles amplified by one primer pair ranged from five by MCSA176C01 to 25 by SCM4. The mean PIC value of each SSR primer pair ranged from 0.15 to 0.29 with an average of 0.23. The probability of identity (*PI*) of the 21 markers was all very low, which ranged from 0.000001 (mSSCIR36) to 0.071332 (SMC569CS) with an average of 0.015532. For the 21 primers pairs, the resolving power of the primer (*RP*) was relatively high, ranging from 3.68 (SMC569CS) to 21.01 (mSSCIR36) with an average of 9.14. The mean number of alleles and the mean PIC value of genomic SSRs were 10.6 and 0.23, and were 9.8 and 0.23 for EST SSRs, respectively (Table 3).

**Table 3.** Genetic diversity parameters of 150 of the most popular parental clones from sugarcane hybrid breeding programs.

Primer Name	Allele (No.)	Product Size (bp)	Range of PIC <sup>a</sup> Values	Mean of PIC Values	<i>PI</i> <sup>b</sup>	<i>RP</i> <sup>c</sup>
mSSCIR36	21	127–168	0.01–0.38	0.15	0.000001	7.09
mSSCIR46	12	146–177	0.01–0.37	0.15	0.002858	13.04
mSSCIR74	6	215–228	0.00–0.37	0.17	0.042135	4.69
SCM4	25	92–209	0.01–0.37	0.17	0.000087	4.16
SCM7	7	155–188	0.03–0.37	0.18	0.048672	3.68
SCM18	9	226–251	0.00–0.38	0.19	0.010157	8.67
SMC1825LA	10	91–119	0.01–0.37	0.20	0.001240	6.53
SMC286CS	13	128–152	0.01–0.37	0.21	0.000411	7.31
SMC477CG	15	115–134	0.00–0.36	0.21	0.000125	4.11
SMC486CG	7	222–243	0.06–0.36	0.22	0.051066	4.88
SMC569CS	6	166–220	0.04–0.38	0.24	0.071332	14.05
SMC597CS	14	143–166	0.03–0.37	0.24	0.000034	10.99
SMC334BS	12	145–163	0.01–0.38	0.24	0.000140	6.27
SMC36BUQ	12	103–251	0.00–0.37	0.25	0.010448	7.49
SMC7CUQ	7	156–170	0.00–0.37	0.26	0.024118	9.76
SEGM285	13	306–389	0.03–0.38	0.26	0.000143	21.01
UGSM345	8	320–334	0.01–0.38	0.27	0.005772	13.68
UGSM10	10	97–125	0.00–0.38	0.28	0.005289	9.31
UGSuM50	6	123–139	0.05–0.38	0.28	0.023095	6.24
MCSA068G08	8	179–202	0.06–0.38	0.29	0.003035	15.57
MCSA176C01	5	427–440	0.11–0.38	0.29	0.026013	13.31

<sup>a</sup> PIC: Polymorphism information content; <sup>b</sup> *PI*: Probability of identity; <sup>c</sup> *RP*: Resolving power.

#### 3.2. Genetic Diversity

The gene diversity (*h*) of the polymorphic allele ranged from 0.013 to 0.500 with an average of 0.282. The Shannon's information index (*I*) of the polymorphic allele ranged from 0.010 to 0.534 with an average of 0.261. Among the different series of sugarcane parental lines, the highest values of both gene diversity (*h*) and Shannon's information index (*I*) were found in the YC-series (0.261, 0.397), followed by the YT-series (0.254, 0.386,) and the GT-series (0.251, 0.376) (Table 3), indicating that the YC-series is genetically more diverse than the other series. The average percentages of polymorphic allele for the YT-, YC-, and CP-series were 0.814, 0.805 and 0.743, respectively. Alleles were identified that were unique to the 12 distinct germplasm groups (Table 4).

**Table 4.** Gene diversity, Shannon's information index, percentage of polymorphic loci and series-specific alleles of different series.

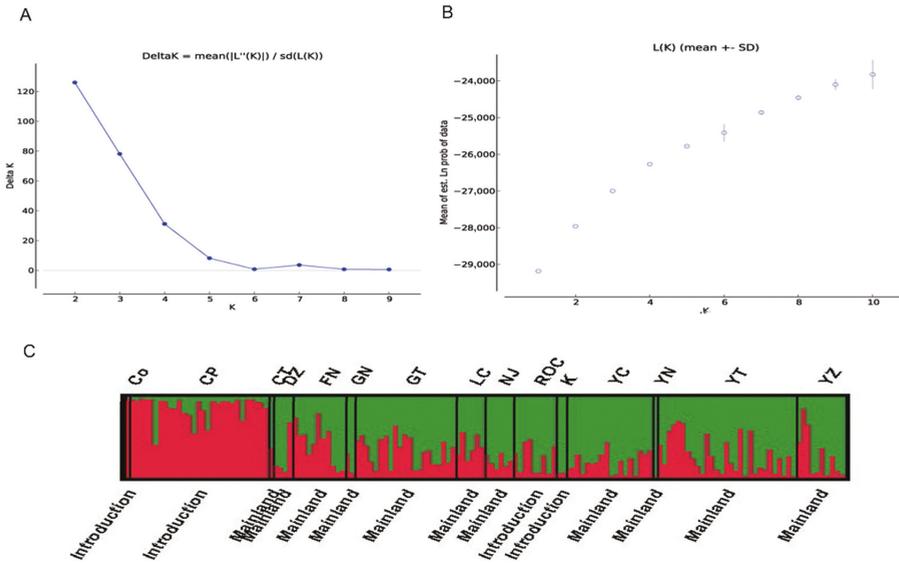
Series	Sample Size	$h^a$	$I^b$	PPL <sup>c</sup>	Series-Specific Alleles
CP	29	0.239	0.361	0.743	SCM7-188, SCM18-238, SMC486CG-225, SMC486CG-233
DZ	4	0.235	0.341	0.562	
FN	11	0.245	0.365	0.677	mSSCIR46-153
GN	2	0.148	0.205	0.296	
GT	21	0.251	0.376	0.721	
LC	6	0.197	0.293	0.522	
NJ	6	0.205	0.302	0.527	SMC36BUQ-125
ROC	9	0.201	0.301	0.558	SMC36BUQ-184, SEGM285-359
K	2	0.164	0.227	0.327	
YC	18	0.261	0.397	0.805	mSSCIR46-146, mSSCIR46-149, SCM7-175, SMC569CS-174, SMC569CS-202, SMC36BUQ-106, SMC36BUQ-132, UGSM10-113
YT	29	0.254	0.386	0.814	SMC36BUQ-105, SMC36BUQ-139
YZ	10	0.241	0.358	0.650	
Mean		0.176	0.261	0.480	

<sup>a</sup>  $h$ , Gene diversity; <sup>b</sup>  $I$ , Shannon's information index; <sup>c</sup> PPL, percentage of polymorphic loci.

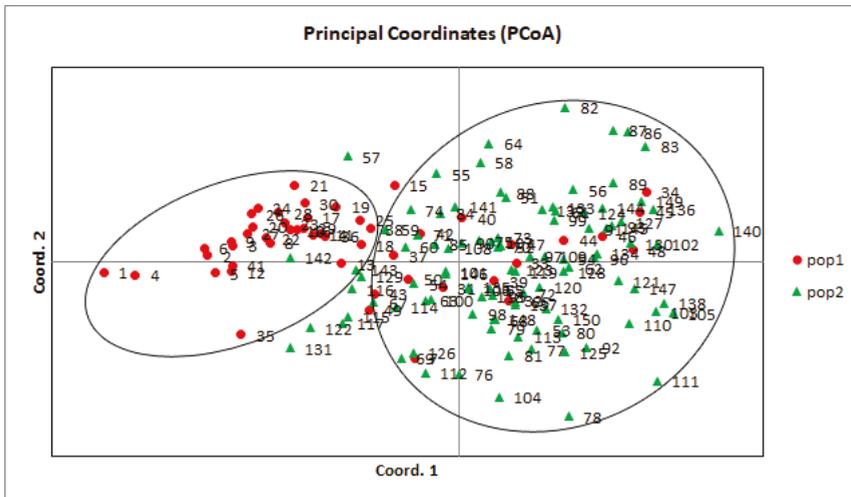
### 3.3. Population Structure and Phylogeny

The  $K$ -value was used to estimate the number of clusters of the clones based on the genotypic data. A continuous gradual increase was observed in the log-likelihood of  $K$ -value ( $\text{LnP}(K)$ ) with the increase of  $K$ -value (Figure 1B and Table S2). The number of clusters ( $K$ ) was plotted against Delta  $K$  ( $\Delta K$ ), which revealed a sharp peak at  $K = 2$  (Figure 1A and Table S2). The optimal  $K$ -value was  $K = 2$ , which revealed that the highest probability for the presence of two sub-populations (Pop1 and Pop2) among the 150 sugarcane clones (Figure 1C); Pop1 consisted of 50 clones and Pop2 contained 100 clones (Table S3). Pop1 clones were mainly introduction accessions and most of the Pop2 clones were from Mainland China.

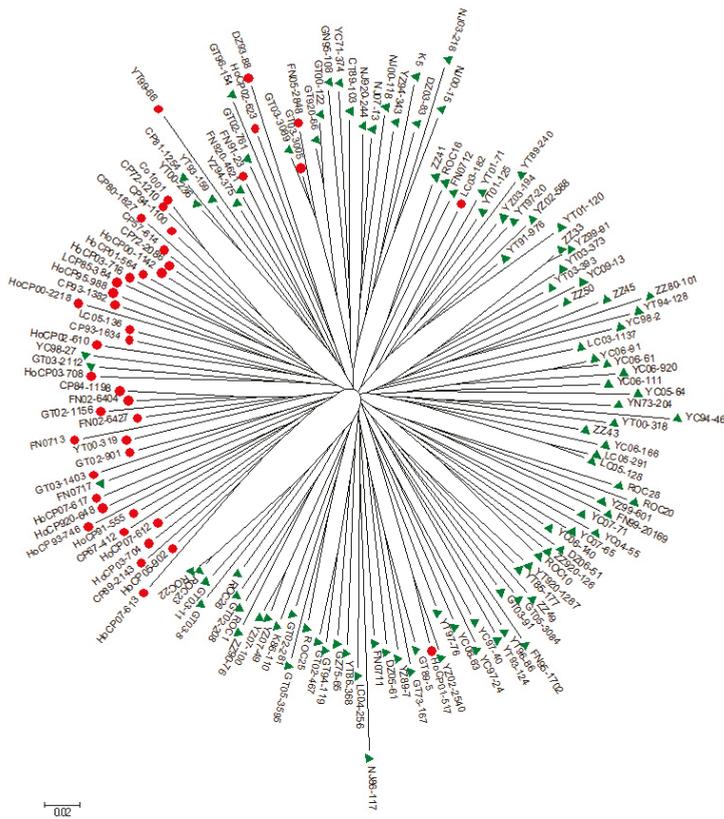
In accordance with the population structure results, PCoA also showed two clusters with the first three axes together explained 20.04% of cumulative variation. In the PCoA plot, the first and second principal coordinates accounted for 8.41% and 6.71% of the total variations, respectively (Figure 2). Furthermore, the unrooted neighbor-joining phylogenetic tree (Figure 3) also showed two clusters. One cluster contained most of the clones of Pop1; the other cluster contained most of the clones of Pop2. However, the admixture of clones between the two sub-populations does exist. Few accessions (YC98-27, GT03-2112 and FN0717) native to China were clustered into Pop1 while several others (HoCP01-517, ROC10, ROC16, K5, ROC25, ROC22, ROC1) introduced to China Mainland were grouped into Pop2.



**Figure 1.** (A) Delta K ( $\Delta K$ ) for different numbers of subpopulations (K); (B) average log-likelihood K-value ( $\ln P(K)$ ) against the number of K; (C) the population structure of 150 most popular parental clones in the hybrid breeding programs in China based on the distribution of 226 SSR alleles among these clones. Pop1 clones are coded in red and Pop2 clones in green.



**Figure 2.** Principal coordinates analysis (PCoA) scatter plots. Red circles represent the Pop1 clones and green triangles the Pop2 clones.



**Figure 3.** A neighbor-joining phylogenetic tree based on the pair-wise genetic distance between 150 most popular parental clones from hybrid breeding programs in China. Red circles represent the Pop1 clones and green triangles the Pop2 clones.

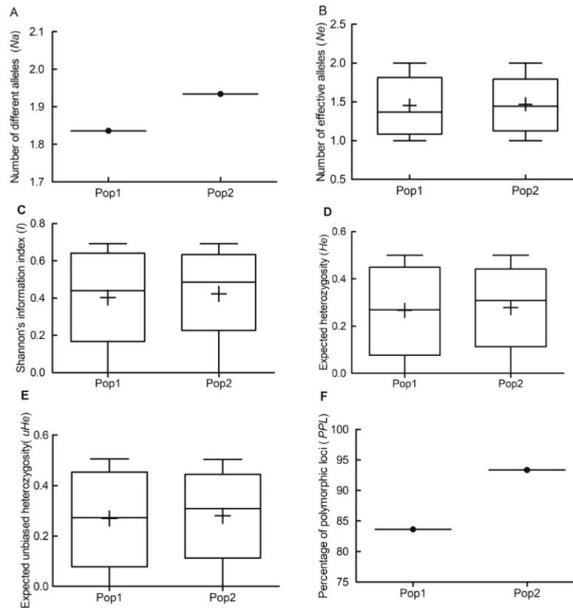
3.4. Genetic Differentiation and Allelic Pattern Across Populations

The two sub-populations Pop1 and Pop2 identified by the Structure analysis were subjected to the GeneALEX analysis to calculate the values of Analysis of Molecular Variance (AMOVA), *Nei's* genetic distance and genetic diversity indices (Table 5). The variation value within the sub-populations (95% of total variation) was significantly higher than that between the sub-populations (5% of total variation). In addition, a high gene flow ( $Nm = 4.981$ ) and a low fixation index value ( $Fst = 0.048$ ) were obtained on the basis of *Nei's* genetic distance analysis.

**Table 5.** Analysis of molecular variance (AMOVA) of SSR-based genetic variation between and within two sub-populations of Pop1 and Pop2.

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Sum of Squares	Estimated Variance	Percentage of Variation
Between sub-Pops	1	546.240	546.240	6.308	5%
Within sub-Pop	148	18,601.600	125.686	125.686	95%
Total	149	19,147.840		131.995	100%
Fixation Index	$Fst = 0.048$				
Gene Flow	$Nm = 4.981$				

The mean value of the number of different alleles ( $N_a$ ) and effective alleles ( $N_e$ ) of the two sub-populations were  $1.885 \pm 0.015$  and  $1.462 \pm 0.017$ , respectively. The mean values for  $I$ ,  $He$  and  $uHe$  among the 150 parental clones were  $0.413 \pm 0.011$ ,  $0.272 \pm 0.008$  and  $0.274 \pm 0.009$ , respectively. Pop2 ( $I = 0.423 \pm 0.016$ ,  $He = 0.278 \pm 0.012$ , and  $uHe = 0.278 \pm 0.012$ ) showed higher levels of genetic diversity than Pop1 ( $I = 0.403 \pm 0.017$ ,  $He = 0.267 \pm 0.012$ , and  $uHe = 0.269 \pm 0.012$ ). The percentage of polymorphic loci per population ( $PPL$ ) ranged from 83.63% (Pop1) to 93.36% (Pop2) with an average of 88.50% (Figure 4).



**Figure 4.** Allelic pattern of SSR across the two sub-populations Pop1 and Pop2. (A) Number of SSR alleles ( $N_a$ ); (B) number of effective SSR alleles ( $N_e$ ); (C) Shannon's information index ( $I$ ); (D) expected heterozygosity ( $He$ ); (E) expected unbiased heterozygosity ( $uHe$ ); and (F) percentage of polymorphic loci ( $PPL$ ).

#### 4. Discussion

Cross hybridization has become the main breeding method for the sugarcane variety improvement. In the traditional sugarcane cross-breeding process, selecting parental clones for crossing is the most important step. Only parental clones sharing a highly level of genetic diversity and complementarity can generate high quality seedling populations [56,57]. Since the 1950s, some sugarcane cultivars from America and China Taiwan have played a very important role in China's sugarcane cross-breeding programs [3]. Meanwhile, some new elite sugarcane parents are being created and utilized by the breeders every year. To make informed crossing choices, the genetic relationship among the parental clones involved in the latest sugarcane cross-breeding programs should be clarified.

In this study, we used 21 pairs of SSR primers to investigate the genetic diversity and population structure of 150 of the most commonly used parental clones. These primer pairs amplified 226 alleles, of which 97.3% were polymorphic. The mean PIC and the gene diversity ( $h$ ) of the polymorphic alleles were 0.23 and 0.28, respectively, which were lower than the values reported on the "World Collections of Sugarcane and Related Grasses" (WGSRG) ( $PIC = 0.2568$ ,  $h = 0.310$ ) [23]. This may be largely due to the number of accessions involved in the world collection study. The WCSRG study involved 1002 highly diverse accessions, belonging to nine species, whereas only 150 clones were used in this study.

Since 2000, a large number of genomic SSR and EST-SSR markers has been developed and applied effectively in estimating genetic diversity in the sugarcane [16,35,39,41,58]. After a lot of screening and identification (unpublished), we selected the best 21 primer pairs from these reports, including eight EST-SSR and 13 genomic SSR. We found that the number and mean PIC value of the EST-SSR alleles were lower than those of the genomic SSR alleles (Table 2). This can be due to the fact that the EST-SSR alleles are located in more conserved regions of the genome [16].

The probability of identity (*PI*) is an individual identification estimator that shows the probability of two different accessions sharing the same genotypes at one specific locus in a population [23]. In this study, the *PI* values of all SSR primer pairs were very low, ranging from 0.000001 (mSSCIR36) to 0.071332 (SMC569CS) (Table 2). The combined *PI* value for all markers was  $9.04 \times 10^{-57}$ , indicating that these 21 SSR primer pairs are able to distinguish the 150 parental clones. The resolving power of the primer pair (*Rp*) is an index, which explains the primer pair's ability to identify different genotypes. *Rp* is related to the distribution of alleles within the sampled genotypes [46] and has been found to correlate strongly with the genotype in evaluating 34 potato cultivars using four primers [46]. The mean *Rp* value (9.135) of the 21 SSR primer pairs is much higher than other studies, such as 2.37 by [59] and 2.2 by [12], indicating these primer pairs are more informative and could identify more cultivars.

Based on geographic origin, the 150 clones were sorted into 15 series. Among these series, the genetic diversity (*h*) indices ranged from 0 to 0.261 and the Shannon's information index (*I*) ranged from 0 to 0.397. At the series level, the YC-series had the highest genetic diversity ( $h = 0.261, I = 0.397$ ), which was similar to the previous results reported by You et al. [35,60]. The YC-series clones are from the Hainan Sugarcane Breeding Station of Guangzhou Sugarcane Industry Research Institute in Sanya city, Hainan province, where the primary sugarcane crossing facility of China is located. The YC-series clones were selected from crosses involving indigenous clones, foreign clones, and clones of closely related *Saccharum* species and genera [35]. Furthermore, the YC-series also had the greatest number of eight series-specific alleles. Only four, two, one, and one unique alleles were found in the CP-series, YT-series, ROC-series, FN-series and NJ-series clones, respectively. Series-specific alleles are the alleles found only in a single population among a broader collection of populations [61,62]. These alleles have been proven to be informative for population genetic studies [63,64] and we may use these alleles for variety identification and marker assisted selection.

The 150 parental clones were classified into two groups (Pop1 and Pop2) based on the PCoA, phylogenetic analysis and population structure analysis. Pop1 contained the majority of foreign accessions with the membership probabilities of  $>0.5$ , while most accessions from Mainland China were assigned to Pop2. Certain specific target traits intentionally selected by different germplasm collectors or breeders might also contribute to the population structure [54]. However, admixture of clones between the two sub-populations do exist (Figures 1–3). For example, one out of the 29 CP-series clones, nine ROC-series clones and two K-series clones clustered into Pop2, but the majority of introduction clones clustered into Pop1. Likewise, one out of four DZ-series, five out of 11 FN-series, four out of 21 GT-series, two out of six LC-series, seven out of 29 YT-series, and two out of 10 YZ-series clones clustered into Pop1, while the majority of the clones from Mainland China clustered into Pop2. This might be due to genetic exchange among different series, or the similar threshold (Pop1: 0.5098, Pop2: 0.4902) (Table S3) resulting in several clones to be clustered completely into a certain group (Pop1 or Pop2), while others being clustered into both groups.

The utilization data was based the most widely used 150 parental clones of sugarcane breeding programs in China during the recent five years. These included 32 of clones from foreign origin, 109 clones from the China Mainland, and nine ROC-series clones from China Taiwan. Among the 32 foreign clones, only one was from India (Co1001), two were from Thailand (K5 and K86-110) while the majority of them (29/32) were from the US (CP-series). Co1001 has been used as parental line extensively in the sugarcane breeding programs in the world. Some sugarcane cultivars, including the CP-series and China Mainland clones, were the progenies of Co-series varieties. Compared to clones from China Mainland, the CP-series clones may have closer genetic distance with the Co-series. So CP-series clones

and Co-series clone can be clustered into Pop1. K5 and K86-110, which were from Thailand, were two of the most widely used parental clones in China. Some clones from China Mainland were the progenies of K5 and K86-110. Clones from China mainland may have the closer genetic distance with the two clones to be clustered into Pop2. The ROC-series varieties have been used as major cultivars in China Mainland accounting for greater than 80% of sugarcane planting areas [24]. In addition, the ROC-series accessions were also the most widely used parents in China Mainland during the recent five years (Table S1). In our study, the ROC-series accessions were clustered into Pop2 because of their closer genetic distance with China Mainland's clones. It is suggested that less attention be continually paid on the utilization of ROC-series accessions in China Mainland's sugarcane breeding programs.

Fixation index ( $F_{st}$ ) measures the genetic distance between populations. An  $F_{st}$  value of zero indicates no differentiation between the sub-populations, while one indicates complete differentiation [65]. An  $F_{st}$  value less than 0.05 is considered no differentiation, while an  $F_{st}$  value greater than 0.15 is considered significant in differentiating populations [66]. In this study, the  $F_{st}$  value between the two sub-populations was 0.048 (Table 5), which was low and would indicate a very low genetic differentiation. This is consistent with the results obtained from the AMOVA, where the genetic variation within sub-populations (95%) was significantly higher than between sub-populations (5%). Gene flow ( $Nm$ ) is the transfer of genetic variation from one population to another. If the value is less than one, then the gene exchange would be limited between sub-populations [67]. In this study, the  $Nm$  value was high, 4.981 suggesting that a high level of genetic exchange may have occurred and this can result in a low genetic differentiation between the two sub-populations. Since the genetic diversity indices of Pop2, such as the number of different alleles ( $N_a$ ), effective alleles ( $N_e$ ),  $I$ ,  $H_e$  and  $uH_e$ , were all higher than those of Pop1, Pop2 is more diverse than Pop1.

Selecting genetically distant accessions from Pop1 and Pop2 for crossing parents in sugarcane breeding programs will potentially lead to elite varieties with broadened genetic bases. Almost all the CP-series clones from the US were clustered into Pop1. These clones have been used extensively as parental lines in the sugarcane breeding programs in China; some have become or are elite progenitors of Chinese cultivars [67]. In addition, this study shows that several YC-series clones are also good crossing parents with a high level of genetic diversity.

## 5. Conclusions

Using a high-performance capillary electrophoresis (HPCE) detection system, the most widely used 150 sugarcane parental clones from 15 different series were fingerprinted with 21 SSR primer pairs. A total of 226 SSR alleles were identified and the distribution of these SSR alleles were subjected to genetic variation, phylogeny, population structure, and principal coordinate analyses. The results showed that the parental lines were clustered into two distinct groups, Pop1 and Pop2. Pop1 contained the majority of foreign clones, while Pop2 consisted of the majority of accessions from Mainland China. Genetic differentiation between the two groups was low. The YC-series clones of Pop2 displayed a high level of genetic diversity and the CP-series clones were elite parents of several Chinese cultivars. The introduction and utilization of more clones of the YC- and CP-series into China's sugarcane breeding programs will broaden the genetic base of breeding germplasm and produce high quality seedlings for selection and development of elite varieties.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4395/9/8/449/s1>, Table S1: Utilization data of the most widely used 150 parental clones from sugarcane hybrid breeding programs in China during the recent five years. Table S2: Tabulated  $K$  values of 150 most popular parental clones from sugarcane hybrid breeding programs in China at  $K = 1$  to 10. Table S3: Sub-population assignment of the 150 most popular parental clones from the sugarcane breeding programs in China based on the  $Q$  values.

**Author Contributions:** Methodology, J.W.; Validation, J.W. and Q.W.; Formal Analysis, J.W. and Y.-B.P.; Investigation, J.X., H.X. and J.W.; Resources, F.Z., C.Z., and W.Z.; Data Curation, J.X., Y.G. and H.C.; Writing—Original Draft Preparation, J.W., Y.-B.P. and J.X.; Writing—Review and Editing, Q.W. and Y.-B.P. Funding Acquisition, Q.W. and Y.Q.

**Funding:** This research was funded by the National Natural Science Foundation of China (31701488), the Earmarked Fund for China Agriculture Research System (CARS-170107), the Science and Technology Project of Guangdong Province (2017A030303049) and the Guangdong Provincial Team of Technical System Innovation for Sugarcane Sisal Industry (2019KJ104-02).

**Acknowledgments:** We thank Perng-Kuang Chang, James Todd and Yunlin Jia for their review comments and language editing.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Lima, M.L.A.; Garcia, A.A.F.; Oliveira, K.M.; Matsuoka, S.; Arizono, H.; de Souza, C.L.; de Souza, A.P. Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugar cane (*Saccharum* spp.). *Theor. Appl. Genet.* **2002**, *104*, 30–38. [[CrossRef](#)] [[PubMed](#)]
2. Ming, R.; Moore, P.H.; Wu, K.; D’Hont, A.; Glaszmann, J.C.; Tew, T.L.; Mirkov, T.E.; Da Silva, J.; Jifon, J.; Rai, M.; et al. *Sugarcane Improvement through Breeding and Biotechnology*; John Wiley & Sons, Ltd.: New York, NY, USA, 2010; pp. 15–118.
3. Qi, Y.W.; Pan, Y.B.; Fang, Y.L.; Zhang, C.M.; Fan, L.N.; He, H.Y.; Liu, R.; Wang, Q.N.; Liu, S.M.; Liu, F.Y.; et al. Genetic structure and diversity of parental cultivars involved in China mainland sugarcane breeding programs as inferred from DNA microsatellites. *J. Integr. Agric.* **2012**, *11*, 1794–1803. [[CrossRef](#)]
4. Nair, N.V.; Selvi, A.; Sreenivasan, T.V.; Pushpalatha, K.N. Molecular diversity in Indian sugarcane cultivars as revealed by Randomly Amplified DNA polymorphisms. *Euphytica* **2002**, *127*, 219–225. [[CrossRef](#)]
5. Creste, S.; Sansoli, D.M.; Tardiani, A.C.S.; Silva, D.N.; Gonçalves, F.K.; Fávero, T.M.; Medeiros, C.N.F.; Festucci, C.S.; Carlini-Garcia, L.A.; Landell, M.G.A.; et al. Comparison of AFLP, TRAP and SSRs in the estimation of genetic relationships in sugarcane. *Sugar Tech.* **2010**, *12*, 150–154. [[CrossRef](#)]
6. Jannoo, N.; Grivet, L.; Seguin, M.; Paulet, F.; Domaingue, R.; Rao, P.S.; Dookun, A.; D’Hont, A.; Glaszmann, J.C. Molecular investigation of the genetic base of sugarcane cultivars. *Theor. Appl. Genet.* **1999**, *99*, 171–184. [[CrossRef](#)]
7. Silva, J.A.G.D.; Sorrells, M.E.; Burnquist, W.L.; Tanksley, S.D. RFLP linkage map and genome analysis of *Saccharum spontaneum*. *Genome* **1993**, *36*, 782–791. [[CrossRef](#)]
8. Pan, Y.B.; Burner, D.M.; Legendre, B.L.; Grisham, M.P.; White, W.H. An assessment of the genetic diversity within a collection of *Saccharum spontaneum* L. with RAPD-PCR. *Genet. Resour. Crop Evol.* **2004**, *51*, 895–903. [[CrossRef](#)]
9. Singh, P.; Singh, S.P.; Tiwari, A.K.; Sharma, B.L. Genetic diversity of sugarcane hybrid cultivars by RAPD markers. *3 Biotech* **2017**, *7*, 222. [[CrossRef](#)]
10. Garcia, A.A.F.; Mollinari, M.; Marconi, T.G.; Serang, O.R.; Silva, R.R.; Vieira, M.L.C.; Vicentini, R.; Costa, E.A.; Mancini, M.C.; Garcia, M.O.S.; et al. SNP genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autopolyploids. *Sci. Rep. UK* **2013**, *3*, 3399. [[CrossRef](#)]
11. Parida, S.K.; Kalia, S.K.; Kaul, S.; Dalal, V.; Hemaprabha, G.; Selvi, A.; Pandit, A.; Singh, A.; Gaikwad, K.; Sharma, T.R.; et al. Informative genomic microsatellite markers for efficient genotyping applications in sugarcane. *Theor. Appl. Genet.* **2009**, *118*, 327–338. [[CrossRef](#)]
12. Devarumath, R.M.; Kalwade, S.B.; Kawar, P.G.; Sushir, K.V. Assessment of Genetic Diversity in Sugarcane Germplasm Using ISSR and SSR Markers. *Sugar Tech.* **2012**, *14*, 334–344. [[CrossRef](#)]
13. Oliveira, L.A.R.; Machado, C.A.; Cardoso, M.N.; Oliveira, A.C.A.; Amaral, A.L.; Rabbani, A.R.C.; Silva, A.V.C.; Ledo, A.S. Genetic diversity of *Saccharum* complex using ISSR markers. *Genet. Mol. Res.* **2017**, *16*, 1–9. [[CrossRef](#)]
14. Pinto, L.R.; Oliveira, K.M.; Marconi, T.; Garcia, A.A.F.; Ulian, E.C.; de Souza, A.P. Characterization of novel sugarcane expressed sequence tag microsatellites and their comparison with genomic SSRs. *Plant Breed.* **2006**, *125*, 378–384. [[CrossRef](#)]
15. James, B.T.; Chen, C.; Rudolph, A.; Swaminathan, K.; Murray, J.E.; Na, J.; Spence, A.K.; Smith, B.; Hudson, M.E.; Moose, S.P.; et al. Development of microsatellite markers in autopolyploid sugarcane and comparative analysis of conserved microsatellites in sorghum and sugarcane. *Mol. Breed.* **2012**, *30*, 661–669. [[CrossRef](#)]

16. Parthiban, S.; Govindaraj, P.; Senthilkumar, S. Comparison of relative efficiency of genomic SSR and EST-SSR markers in estimating genetic diversity in sugarcane. *3 Biotech* **2018**, *8*, 144. [[CrossRef](#)]
17. Pan, Y.B.; Burner, D.M.; Legendre, B.L. An Assessment of the Phylogenetic Relationship Among Sugarcane and Related Taxa Based on the Nucleotide Sequence of 5S rRNA Intergenic Spacers. *Genetica* **2000**, *108*, 285–295. [[CrossRef](#)]
18. Que, Y.X.; Pan, Y.B.; Lu, Y.H.; Yang, C.; Yang, Y.T.; Huang, N.; Xu, L.P. Genetic analysis of diversity within a Chinese local sugarcane germplasm based on start codon targeted polymorphism. *Biomed Res. Int.* **2014**, *2014*, 468375. [[CrossRef](#)]
19. Que, Y.X.; Chen, T.; Xu, L.; Chen, R.K. Genetic diversity among key sugarcane clones revealed by TRAP markers. *J. Agric. Biotechnol.* **2009**, *17*, 496–503.
20. Alwala, S.; Suman, A.; Arro, J.A.; Veremis, J.C.; Kimbeng, C.A. Target region amplification polymorphism (TRAP) for assessing genetic diversity in sugarcane germplasm collections. *Crop Sci.* **2006**, *46*, 448. [[CrossRef](#)]
21. Khan, M.; Pan, Y.B.; Iqbal, J. Development of an RAPD-based SCAR marker for smut disease resistance in commercial sugarcane cultivars of Pakistan. *Crop Prot.* **2017**, *94*, 166–172. [[CrossRef](#)]
22. Ali, A.; Pan, Y.; Wang, Q.; Wang, J.; Chen, J.; Gao, S. Genetic diversity and population structure analysis of *Saccharum* and *Erianthus* genera using microsatellite (SSR) markers. *Sci. Rep. UK* **2019**, *9*, 395. [[CrossRef](#)]
23. Nayak, S.N.; Song, J.; Villa, A.; Pathak, B.; Ayala-Silva, T.; Yang, X.; Todd, J.; Glynn, N.C.; Kuhn, D.N.; Glaz, B.; et al. Promoting Utilization of *Saccharum* spp. Genetic Resources through Genetic Diversity Analysis and Core Collection Construction. *PLoS ONE* **2014**, *9*, e110856. [[CrossRef](#)]
24. Liu, H.L.; Yang, X.P.; You, Q.; Song, J.; Wang, L.P.; Zhang, J.S.; Deng, Z.H.; Ming, R.; Wang, J.P. Pedigree, marker recruitment, and genetic diversity of modern sugarcane cultivars in China and the United States. *Euphytica* **2018**, *214*, 48. [[CrossRef](#)]
25. Pan, Y. Development and Integration of an SSR-Based Molecular Identity Database into Sugarcane Breeding Program. *Agronomy* **2016**, *6*, 28. [[CrossRef](#)]
26. Marconi, T.G.; Costa, E.A.; Miranda, H.R.; Mancini, M.C.; Cardososilva, C.B. Functional markers for gene mapping and genetic diversity studies in sugarcane. *BMC Res. Notes* **2011**, *4*, 264. [[CrossRef](#)]
27. Andru, S.; Pan, Y.; Thongthawee, S.; Burner, D.M.; Kimbeng, C.A. Genetic analysis of the sugarcane (*Saccharum* spp.) cultivar ‘LCP 85-384’. I. Linkage mapping using AFLP, SSR, and TRAP markers. *Theor. Appl. Genet.* **2011**, *123*, 77–93. [[CrossRef](#)]
28. Banerjee, N.; Siraree, A.; Yadav, S.; Kumar, S.; Singh, J.; Kumar, S.; Pandey, D.K.; Singh, R.K. Marker-trait association study for sucrose and yield contributing traits in sugarcane (*Saccharum* spp. hybrid). *Euphytica* **2015**, *205*, 185–201. [[CrossRef](#)]
29. Racedo, J.; Gutiérrez, L.; Perera, M.F.; Ostengo, S.; Pardo, E.M.; Cuenya, M.I.; Welin, B.; Castagnaro, A.P. Genome-wide association mapping of quantitative traits in a breeding population of sugarcane. *BMC Plant Biol.* **2016**, *16*, 142. [[CrossRef](#)]
30. Ukoskit, K.; Posudsavang, G.; Pongsiripat, N.; Chatwachirawong, P.; Klomsa-ard, P.; Poomipant, P.; Tragoonrun, S. Detection and validation of EST-SSR markers associated with sugar-related traits in sugarcane using linkage and association mapping. *Genomics* **2019**, *111*, 1–9. [[CrossRef](#)]
31. Pan, Y.B. Highly polymorphic microsatellite DNA markers for sugarcane germplasm evaluation and variety identity testing. *Sugar Tech.* **2006**, *8*, 246–256. [[CrossRef](#)]
32. Chen, P.H.; Pan, Y.B.; Chen, R.K.; Xu, L.P.; Chen, Y.Q. SSR marker-based analysis of genetic relatedness among sugarcane cultivars (*Saccharum* spp. hybrids) from breeding programs in China and other countries. *Sugar Tech.* **2009**, *11*, 347–354. [[CrossRef](#)]
33. Liu, P.W.; Que, Y.X.; Pan, Y.B. Highly Polymorphic Microsatellite DNA Markers for Sugarcane Germplasm Evaluation and Variety Identity Testing. *Sugar Tech.* **2011**, *13*, 129–136. [[CrossRef](#)]
34. Pan, Y.B.; Liu, P.W.; Que, Y.X. Independently Segregating Simple Sequence Repeats (SSR) Alleles in Polyploid Sugarcane. *Sugar Tech.* **2015**, *17*, 235–242. [[CrossRef](#)]
35. You, Q.; Pan, Y.; Xu, L.; Gao, S.; Wang, Q.; Su, Y.; Yang, Y.; Wu, Q.; Zhou, D.; Que, Y. Genetic Diversity Analysis of Sugarcane Germplasm Based on Fluorescence-Labeled Simple Sequence Repeat Markers and a Capillary Electrophoresis-based Genotyping Platform. *Sugar Tech.* **2016**, *18*, 380–390. [[CrossRef](#)]
36. Ali, A.; Wang, J.; Pan, Y.; Deng, Z.; Chen, Z.; Chen, R.; Gao, S. Molecular identification and genetic diversity analysis of Chinese sugarcane (*Saccharum* spp. Hybrids) varieties using SSR markers. *Trop. Plant Biol.* **2017**, *10*, 194–203. [[CrossRef](#)]

37. Fu, Y.; Pan, Y.; Lei, C.; Grisham, M.P.; Yang, C.; Meng, Q. Genotype-Specific Microsatellite (SSR) Markers for the Sugarcane Germplasm from the Karst Region of Guizhou, China. *Am. J. Plant Sci.* **2016**, *7*, 2209–2220. [[CrossRef](#)]
38. Rogers, S.O.; Bendich, A.J. *Extraction of DNA from plant tissues. Plant Molecular Biology Manual*; Gelvin, S.B., Schilperoort, R.A., Verma, D.P.S., Eds.; Springer: Dordrecht, The Netherlands, 1989.
39. Pan, Y.B.; Scheffler, B.E.; Richard, E., Jr. High throughput genotyping of commercial sugarcane clones with microsatellite (SSR) DNA markers. *Sugar Tech.* **2007**, *9*, 176–181.
40. Singh, R.K.; Mishra, S.K.; Singh, S.P.; Mishra, N.; Sharma, M.L. Evaluation of microsatellite markers for genetic diversity analysis among sugarcane species and commercial hybrids. *Aust. J. Crop Sci.* **2010**, *4*, 116–125.
41. Parida, S.K.; Pandit, A.; Gaikwad, K.; Sharma, T.R.; Srivastava, P.; Singh, N.K.; Mohapatra, T. Functionally relevant microsatellites in sugarcane unigenes. *BMC Plant Biol.* **2010**, *10*, 251. [[CrossRef](#)]
42. Singh, R.K.; Singh, R.B.; Singh, S.P.; Sharma, M.L. Identification of sugarcane microsatellites associated to sugar content in sugarcane and transferability to other cereal genomes. *Euphytica* **2011**, *182*, 335–354. [[CrossRef](#)]
43. Pan, Y.B.; Tew, T.L.; Schnell, R.J.; Viator, R.P.; Richard, E.P.; Grisham, M.P.; White, W.H. Microsatellite DNA marker-assisted selection of *Saccharum spontaneum* cytoplasm-derived germplasm. *Sugar Tech.* **2006**, *8*, 23–29. [[CrossRef](#)]
44. Fan, W.Q.; Ge, H.M.; Sun, X.S.; Yang, A.G.; Zhang, Z.F.; Ren, M. DataFormater, A software for SSR data formatting to develop population genetics analysis. *Mol. Plant Breed.* **2016**, *14*, 1029–1034.
45. Liu, K.; Muse, S.V. PowerMaker: An integrated analysis environment for genetic maker analysis. *Bioinformatics* **2005**, *21*, 2128–2129. [[CrossRef](#)]
46. Prevost, A.; Wilkinson, M.J. A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. *Theor. Appl. Genet.* **1999**, *98*, 107–112. [[CrossRef](#)]
47. Kalinowski, S.T.; Taper, M.L.; Marshall, T.C. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* **2007**, *16*, 1099–1106. [[CrossRef](#)]
48. Peakall, R.; Smouse, P.E. GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* **2006**, *6*, 288–295. [[CrossRef](#)]
49. Peakall, R.; Smouse, P.E. GenALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* **2012**, *28*, 2537–2539. [[CrossRef](#)]
50. Falush, D.; Stephens, M.; Pritchard, J.K. Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. *Genetics* **2003**, *164*, 1567–1587.
51. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **2010**, *14*, 2611–2620. [[CrossRef](#)]
52. Earl, D.A.; VonHoldt, B.M. Structure harvester: A website and program for visualizing structure output and implementing the Evanno method. *Conserv. Genet. Resour.* **2012**, *4*, 359–361. [[CrossRef](#)]
53. Jakobsson, M.; Rosenberg, N.A. Clumpp: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **2007**, *23*, 1801–1806. [[CrossRef](#)]
54. Luo, Z.; Brock, J.; Dyer, J.M.; Kutchan, T.; Schachtman, D.; Augustin, M.; Ge, Y.; Fahlgren, N.; Abdel-Haleem, H. Genetic Diversity and Population Structure of a Camelina sativa Spring Panel. *Front. Plant Sci.* **2019**, *10*, 184. [[CrossRef](#)]
55. Tamura, K.; Stecher, G.; Peterson, D.; Filipiński, A.; Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* **2016**, *30*, 2725–2729. [[CrossRef](#)]
56. Jackson, P.A. Breeding for improved sugar content in sugarcane. *Field Crop. Res.* **2005**, *92*, 277–290. [[CrossRef](#)]
57. Stevenson, G.C. *Genetic and Breeding of Sugarcane*; Longmans: London, UK, 1965.
58. Cordeiro, G.M.; Casu, R.; McIntyre, C.L.; Manners, J.M.; Henry, R.J. Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. *Plant Sci.* **2001**, *160*, 1115–1123. [[CrossRef](#)]
59. Hameed, U.; Pan, Y.; Muhammad, K.; Afghan, S.; Iqbal, J. Use of simple sequence repeat markers for DNA fingerprinting and diversity analysis of sugarcane (*Saccharum* spp.) cultivars resistant and susceptible to red rot. *Genet. Mol. Res.* **2012**, *11*, 1195. [[CrossRef](#)]
60. You, Q.; Xu, L.P.; Zheng, Y.F.; Que, Y.X. Genetic diversity analysis of sugarcane parents in Chinese breeding programmes using gSSR markers. *Sci. World J.* **2013**, *2013*, 613062. [[CrossRef](#)]

61. Szpiech, Z.A.; Rosenberg, N.A. On the size distribution of private microsatellite alleles. *Theor. Popul. Biol.* **2011**, *80*, 100–113. [[CrossRef](#)]
62. Slatkin, M. Rare alleles as indicators of gene flow. *Evolution* **1985**, *39*, 53–65. [[CrossRef](#)]
63. Kalinowski, S.T. Counting Alleles with Rarefaction: Private Alleles and Hierarchical Sampling Designs. *Conserv. Genet.* **2004**, *5*, 539–543. [[CrossRef](#)]
64. Schroeder, K.B.; Schurr, T.G.; Long, J.C.; Rosenberg, N.A.; Crawford, M.H.; Tarskaia, L.A.; Osipova, L.P.; Zhadanov, S.I.; Smith, D.G. A private allele ubiquitous in the Americas. *Biol. Lett.* **2007**, *3*, 218–223. [[CrossRef](#)]
65. Bird, K.A.; An, H.; Gazave, E.; Gore, M.A.; Pires, J.C.; Robertson, L.D.; Labate, J.A. Population structure and phylogenetic relationships in a diverse panel of *Brassica rapa* L. *Front. Plant Sci.* **2017**, *8*, 321. [[CrossRef](#)]
66. Wright, S. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* **1965**, *19*, 395–420. [[CrossRef](#)]
67. Deng, H.H.; Li, Q.W. Utilization of CP72-1210 in sugarcane breeding program in mainland China. *Guangdong Agric. Sci.* **2007**, *11*, 18–21.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



Article

# Single-Molecule Long-Read Sequencing of Avocado Generates Microsatellite Markers for Analyzing the Genetic Diversity in Avocado Germplasm

Yu Ge, Xiaoping Zang, Lin Tan, Jiashui Wang, Yuanzheng Liu, Yanxia Li, Nan Wang, Di Chen, Rulin Zhan \* and Weihong Ma \*

Haikou Experimental Station, Chinese Academy of Tropical Agricultural Sciences, Haikou 570102, China

\* Correspondence: zhanrulin@catas.cn (R.Z.); zjwhma@catas.cn (W.M.); Tel.: +86-898-6679-4563 (R.Z.); +86-898-6677-3067 (W.M.)

Received: 9 July 2019; Accepted: 1 September 2019; Published: 5 September 2019

**Abstract:** Avocado (*Persea americana* Mill.) is an important fruit crop commercially grown in tropical and subtropical regions. Despite the importance of avocado, there is relatively little available genomic information regarding this fruit species. In this study, we functionally annotated the full-length avocado transcriptome sequence based on single-molecule real-time sequencing technology, and predicted the coding sequences (CDSs), transcription factors (TFs), and long non-coding RNA (lncRNA) sequences. Moreover, 76,777 simple sequence repeat (SSR) loci detected among the 42,096 SSR-containing transcript sequences were used to develop 149,733 expressed sequence tag (EST)-SSR markers. A subset of 100 EST-SSR markers was randomly chosen for an analysis that detected 15 polymorphic EST-SSR markers, with an average polymorphism information content of 0.45. These 15 markers were able to clearly and effectively characterize 46 avocado accessions based on geographical origin. In summary, our study is the first to generate a full-length transcriptome sequence and develop and analyze a set of EST-SSR markers in avocado. The application of third-generation sequencing techniques for developing SSR markers is a potentially powerful tool for genetic studies.

**Keywords:** *Persea americana*; SMRT sequencing; simple sequence repeat; genetic relationship

## 1. Introduction

Avocado (*Persea americana* Mill.) belonging to the family Lauraceae of the order Laurales is native to Mexico and Central and South America, and is one of the most economically important subtropical/tropical fruit crops worldwide [1]. Taxonomic treatments differ considerably in terms of the circumscription and defining of infraspecific avocado entities [2–5]. Additionally, researchers have long considered that geographical isolation has likely resulted in the following three ecological races of avocado: Mexican (*P. americana* var. *drymifolia*), Guatemalan (*P. americana* var. *guatemalensis*), and West Indian (*P. americana* var. *americana*) [1]. The Mexican race adapted to a Mediterranean climate, whereas the Guatemalan race originated in a tropical highland climate, and the West Indian race adapted to humid tropical lowland conditions [1].

Avocado is rich in lipids, sugars, proteins, minerals, vitamins, and other active ingredients [6–8]. Moreover, avocado production has increased worldwide [1]. One factor contributing to the increases in production and consumption is the expansion of avocado products into new global markets where avocado was previously unknown or scarce, including China, which is an emerging market for the production and consumption of avocado [1,9]. After avocado was first introduced and cultivated in China in the late 1950s, selective breeding by some national scientific research bodies and other state farms have resulted in the development of more than 10 superior avocado accessions [9,10]. Additionally, natural crosses among avocado accessions have generated new hybrids on state and

private farms, and some native accessions are increasingly produced in somewhat remote areas with distinct local environmental conditions [9,10]. Avocado is broadly grown and exploited in some provinces in southern China, including Hainan, Guangxi, Yunnan, and Taiwan [9,10]. The climatic conditions in these provinces are subtropical to tropical, which are ideal conditions for the cultivation of avocado [9,10].

The avocado germplasm should be precisely characterized to maximize its utility to breeders worldwide [1]. Specifically, a molecular characterization is required for analyses of the genetic relationships among avocado germplasm. Over the past two decades, studies involving various types of molecular markers have examined the genetic relationships among avocado germplasm [11–20]. Of the many available DNA markers, simple sequence repeats (SSRs) are commonly used for investigating plant genetics and breeding because they are widely distributed and abundant in plant genomes. They are also genetically codominant, highly reproducible, multi-allelic, and perfectly suitable for high-throughput genotyping [21–25]. Expressed sequence tag (EST)-derived markers in the genomic coding regions have an advantage over genomic DNA-derived markers, and can be efficiently amplified to reveal conserved sequences among related species [26]. There has recently been increasing interest in developing EST-SSR markers via high-throughput transcriptome sequencing. Thus, there has been rapid progress in the development of EST-SSR markers based on transcriptome data produced with second-generation sequencing technology for *Lilium brownii* var. *viridulum* Baker [27], *Crataegus pinnatifida* Bunge [28], *Acer miaotaiense* P. C. Tsong [29], and *Rosa hybrida* hort. ex Lavalle [30]. Among the third-generation sequencing platforms, PacBio RS II, which is regarded as the first commercialized third-generation sequencer, is based on single-molecule real-time (SMRT) technology [31]. The PacBio RS II system can produce much longer reads than second-generation sequencing platforms, and has been applied to effectively capture full-length transcript sequences for EST-derived marker development [32]. However, there are few reports regarding the application of EST-SSR markers developed with SMRT technology for crop breeding.

Single-molecule real-time technology has the following three main advantages over second-generation sequencing options: it generates longer reads, it has higher consensus accuracy, and it is less biased [33]. A previous study revealed that SMRT technology can precisely ascertain alternative polyadenylation sites and full-length splice isoforms, and also detect a higher isoform density than that for the reference genome [34]. The application of SMRT technology for nearly 3 years has helped to elucidate the complexity of the transcriptome and molecular mechanism underlying the metabolite synthesis in safflower [31], *Zanthoxylum bungeanum* Maxim. [32], *Trifolium pratense* L. [34], *Saccharum officinarum* L. [35], *Panicum virgatum* L. [36], *Medicago sativa* L. [37], *Zanthoxylum planispinum* Sieb. [38], *Cynodon dactylon* L. Pers. [39], *Camellia sinensis* L. O. Ktze. [40], and *Cassia obtusifolia* L. [41].

In the previous study, we had generated the first full-length transcriptome sequence of avocado based on SMRT technology and the short-reads obtained in this previous study involving second-generation transcriptome sequencing were used to correct the transcripts that were obtained with SMRT technology [42]. In this study, we functionally annotated sequences and completed SSR mining experiments from SMRT technology in avocado mesocarp. We also predicted the coding sequences (CDSs), transcription factors (TFs), and long non-coding RNA (lncRNA) sequences. Furthermore, we identified a set of EST-SSR markers, and assessed their utility for determining the genetic diversity among 46 selected avocado accessions from various locations in southern China. The generated data enabled the broad and distinct visualization of the genetic diversity in the analyzed avocado germplasm. The results of this study represent useful genetic and transcriptome information to support future research on avocado.

## **2. Materials and Methods**

### *2.1. Sample Collection, DNA Extraction, and RNA Extraction*

For transcriptome analyses, avocado fruits (cultivar ‘Hass’) were harvested from April to September 2018 from six 10-year-old trees (grafted onto Zutano clonal rootstock) growing at the Chinese Academy of Tropical Agricultural Sciences (CATAS; Danzhou, Hainan, China; latitude 19°31′ N, longitude 109°34′ E, and 20 m above sea level). Each biological replicate comprised samples from two trees. Specifically, fruits that developed during the main flowering season (i.e., February 2018) were marked, after which samples were collected at five time-points (75, 110, 145, 180, and 215 days after full bloom) until the fruits reached physiological maturity (i.e., able to ripen after harvest). The fruits were randomly collected for each biological replicate during each developmental stage. Fruits were quickly brought to the laboratory, after which the mesocarp (pulp) was separated from the seed and then immediately frozen at  $-80^{\circ}\text{C}$  for subsequent transcriptome analyses. Total RNA was extracted with a Plant RNA Kit (OMEGA Bio-Tek, Norcross, GA, USA).

For competitive allele-specific PCR (KASP) genotyping and EST-SSR detection, seven commercial cultivars and 39 native accessions were selected. These native accessions were obtained from the CATAS (Danzhou, Hainan, China; latitude 19°31′ N, longitude 109°34′ E, and 20 m above sea level), Daling State Farm (DLSE; Baisha, Hainan, China; latitude 19°14′ N, longitude 109°14′ E, and 60 m above sea level), Mengmao State Farm (MMSF; Ruili, Yunnan, China; latitude 24°00′ N, longitude 97°50′ E, and 240 m above sea level), and Guangxi Vocational and Technical College (GVTC; Nanning, Guangxi, China; latitude 22°29′ N, longitude 108°11′ E, and 79 m above sea level). Details regarding the avocado germplasm are provided in Table S1. Genomic DNA was extracted from fresh leaves as described by Ge [43].

### *2.2. PacBio cDNA Library Construction and Sequencing*

Poly-T oligo-attached magnetic beads were used to purify the mRNA from the total RNA extracted from 15 mesocarp (pulp) samples collected at each analyzed developmental stage. The mRNA from all five developmental stages was combined to serve as the template to synthesize cDNA with the SMARTer PCR cDNA Synthesis Kit (Clontech, Mountain View, CA, USA). After a PCR amplification, quality control check, and purification, full-length cDNA fragments were acquired according to the BluePippin Size Selection System protocol, ultimately resulting in the construction of a cDNA library (1–6 kb). Selected full-length cDNA sequences were ligated to the SMRT bell hairpin loop. The concentration of the cDNA library was then determined with the Qubit 2.0 fluorometer, whereas the quality of the cDNA library was assessed with the 2100 Bioanalyzer (Agilent). Finally, one SMRT cell was sequenced with the PacBio RSII system (Pacific Biosciences, Menlo Park, CA, USA).

### *2.3. Illumina cDNA Library Construction and Sequencing*

Oligo-(dT) magnetic beads were used to purify the mRNA from the total RNA extracted from 15 mesocarp (pulp) samples from five developmental stages. Three replicates were analyzed for each developmental stage. Samples from each developmental stage underwent an RNA-sequencing analysis, with three biological replicates per sample. The fragmentation step was completed with divalent cations in heated 5× NEBNext First Strand Synthesis Reaction Buffer. First-strand cDNA was synthesized with a series of random hexamer primers and reverse transcriptase, after which the second-strand cDNA was generated with DNA polymerase I and RNase H. The cDNA libraries were constructed by ligating the cDNA fragments to sequencing adapters and amplifying the fragments by PCR. The libraries were then sequenced with the Illumina HiSeq 2000 platform (Nanxin Bioinformatics Technology Co., Ltd., Guangzhou, China).

#### 2.4. Quality Filtering and Correction of PacBio Long-Reads

Raw reads were processed into error-corrected reads of insert (ROIs) using an isoform sequencing pipeline, with minimum full pass = 0.00 and minimum predicted accuracy = 0.80. Next, full-length, non-chimeric transcripts were detected by searching for the poly-A tail signal and the 5' and 3' cDNA primer sequences in the ROIs. Iterative clustering for error correction was used to obtain high-quality consensus isoforms, which were then polished with QuiverVersion 1.0. The low-quality full-length transcript isoforms were corrected based on Illumina short-reads with the default setting of the Proovread program. High-quality and corrected low-quality transcript isoforms were confirmed as nonredundant with the CD-HIT software.

#### 2.5. Functional Annotation

Genes were functionally annotated based on a BLASTX search (*E*-value threshold of  $10^{-5}$ ) of the following databases: Clusters of Orthologous Groups of proteins (KOG/COG) (available online: <http://www.ncbi.nlm.nih.gov/KOG/>; available online: <http://www.ncbi.nlm.nih.gov/COG/>), Non-supervised Orthologous Groups (eggNOG) (available online: <http://eggnogdb.embl.de/#/app/home>), Swiss-Prot (a manually annotated and reviewed protein sequence database, available online: <http://www.uniprot.org/>), Pfam (assigned with the HMMER3.0 package, available online: <https://pfam.xfam.org/>), and NCBI nonredundant protein sequence (Nr) (available online: <http://www.ncbi.nlm.nih.gov/>). Additionally, the KEGG Automatic Annotation Server [44] was used to assign these genes to Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic pathways (available online: <http://www.genome.jp/kegg/>). The unigenes were annotated with gene ontology (GO) terms (available online: <http://www.geneontology.org/>) with the Blast2GO (version 2.5) program [45] based on the BLASTX matches in the Pfam and Nr databases (*E*-value threshold of  $10^{-6}$ ).

#### 2.6. Mining of EST-SSR Markers

The MISA (version 1.0) program, with the following default settings, was used to locate SSRs: a minimum of five repeats; a minimum motif length of 5 for tri- and hexanucleotides, 6 for dinucleotides, and 10 for single nucleotides.

#### 2.7. Analyses of Detected Coding Sequences, Transcription Factors, and Long Non-Coding RNA Features

The open reading frames (ORFs) detected with the TransDecoder (version 3.0.0) program were designated as putative CDSs if they satisfied the following criteria: (1) An ORF was detected in a transcript sequence; (2) the log-likelihood score was  $>0$ , and was similar to what was calculated with the GeneID software; (3) the score was higher when the ORF was in the first reading frame than when the ORF was in the other five reading frames; (4) if a candidate ORF was within another candidate ORF, the longer one was reported. However, a single transcript could be associated with multiple ORFs (because of operons and chimeras); and (5) the putative encoded peptide matched a Pfam domain.

Transcription factor gene families were identified based on categorically defined TF families and criteria from the KO, KOG, GO, Swiss-Prot, Pfam, Nr, and Nt databases. Specifically, the default parameters of the iTAK (version 1.2) program were used. The methods used to identify and classify TFs were previously described by Perez-Rodriguez [46].

The following four computational tools were combined to sort non-protein-coding RNA candidates from putative protein-coding RNAs among the transcripts: the Coding Potential Calculator (CPC), Coding-Non-Coding Index (CNCI), Coding Potential Assessment Tool (CPAT), and Pfam database. Transcripts longer than 200 nt, with more than two exons, were selected as lncRNA candidates and were further screened with CPC/CNCI/CPAT/Pfam, which distinguished the protein-coding genes from the non-coding genes.

## 2.8. Assignment of the Native Avocado Accessions with an Unknown Race

To validate the origins of the 33 native accessions with an unknown race, six primers for race-specific single nucleotide polymorphism (SNP) loci were used for KASP genotyping listed in Table S2 [47]. The primer mix, which was prepared and used as described by KBioscience (<http://www.kbioscience.co.uk>), comprised 46  $\mu\text{L}$  dH<sub>2</sub>O, 30  $\mu\text{L}$  common primer (100  $\mu\text{M}$ ), and 12  $\mu\text{L}$  each tailed primer (100  $\mu\text{M}$ ). The SNPs were amplified by PCR in a thermal cycler with a 5- $\mu\text{L}$  solution consisting of 1 $\times$  KASP Master mix, 10 ng genomic DNA, and the SNP-specific KASP assay mix. The following PCR amplification conditions were the same as those used for each SNP assay: 94 °C for 15 min; 10 touchdown cycles of 94 °C for 20 s, and 58–61 °C for 60 s (decreasing by 0.8 °C per cycle); 35 cycles of 94 °C for 20 s and 57 °C for 60 s. The resulting data were analyzed with the Roche LightCycler 480 (version 1.50.39) program.

## 2.9. Identification of EST-SSR Markers

To screen the EST-SSR loci, primers based on the sequences flanking the selected microsatellite loci were designed with the Primer3 program; the PCR products ranged from 100 to 300 bp. All assigned marker names included Pa-eSSR to indicate their association with *P. Americana* and EST-SSRs. A subset of 100 EST-SSR primer pairs was randomly selected for validation by a PCR amplification with the same conditions as those described by Ge [43]. The PCR products were analyzed with the 96-capillary 3730xl DNA Analyzer (Applied Biosystems, Foster City, CA, USA). The detection system included 8.9  $\mu\text{L}$  HIDI (Applied Biosystems), 0.1  $\mu\text{L}$  LIZ (Applied Biosystems), and 1  $\mu\text{L}$  PCR products (1:10 dilution). A lack of amplification was considered indicative of a null allele.

## 2.10. Data Analysis

The number of observed alleles (Na), effective number of alleles (Ne), observed heterozygosity (Ho), expected heterozygosity (He), and polymorphism information content (PIC) of each EST-SSR was assessed with the POPGEN (version 1.32) program [48]. A cluster analysis was performed with PowerMarker (version 3.25) [49]. The cophenetic correlation coefficient was computed for the dendrogram after the construction of a cophenetic matrix to measure the goodness of fit between the original similarity matrix and the dendrogram. Bootstrap support values were obtained from 1000 replicates. A neighbor-joining tree was constructed based on shared alleles, and visualized with the MEGA6.0 software [50].

## 3. Results

### 3.1. General Properties and Functional Annotations Based on Public Databases of Single-Molecule Long-Reads

Figure 1 presents the length distribution of 651,260 reads of insert in avocado mesocarp, and the classification of the reads of insert in avocado mesocarps listed in Figure 2. The SMRT and Illumina HiSeq 2000 sequencing data were deposited in the GenBank database (accession numbers PRJNA551932 and PRJNA541745, respectively). Gene annotations according to a BLASTX algorithm indicated that the 71,627 avocado transcripts significantly matched sequences in the COG, GO, KEGG, KOG, Pfam, Swiss-Prot, eggNOG, and Nr databases, respectively (Table S3). The species with the most matches for the transcripts were *Nelumbo nucifera* Gaertn. (41.18% of transcripts), *Vitis vinifera* L. (10.76% of transcripts), *Elaeis guineensis* Jacq. (8.88% of transcripts), and *Phoenix dactylifera* L. (6.90% of transcripts). The homology with the other species was relatively low (1.14%–2.54% of transcripts; Figure 3). To further predict and classify the functions of the annotated transcripts, we analyzed their matching GO terms, eggNOG classifications, and KEGG pathway assignments. A total of 45,134 transcripts were assigned to 51 subcategories of the three main GO functional categories as follows: 106,390 transcripts for biological processes, 45,931 transcripts for cellular components, and 69,120 for molecular functions (Figure 4a, Table S4). Next, 70,205 transcripts were functionally classified into 25 eggNOG categories (Figure 4b, Table S5). Among the 26 categories, the most heavily represented group was posttranslational modification, protein turnover, chaperones (6410 transcripts, 8.94%), followed by

signal transduction mechanisms (4189 transcripts, 5.84%) and transcription (3868 transcripts, 5.39%). Only 20 and 6 transcripts belonged to the cell motility and nuclear structure categories, respectively. Finally, 33,310 transcripts were assigned to 129 KEGG pathways (Table S6). The most represented pathways were related to carbon metabolism (1678 transcripts), protein processing in endoplasmic reticulum (1649 transcripts), and biosynthesis of amino acids (1503 transcripts).

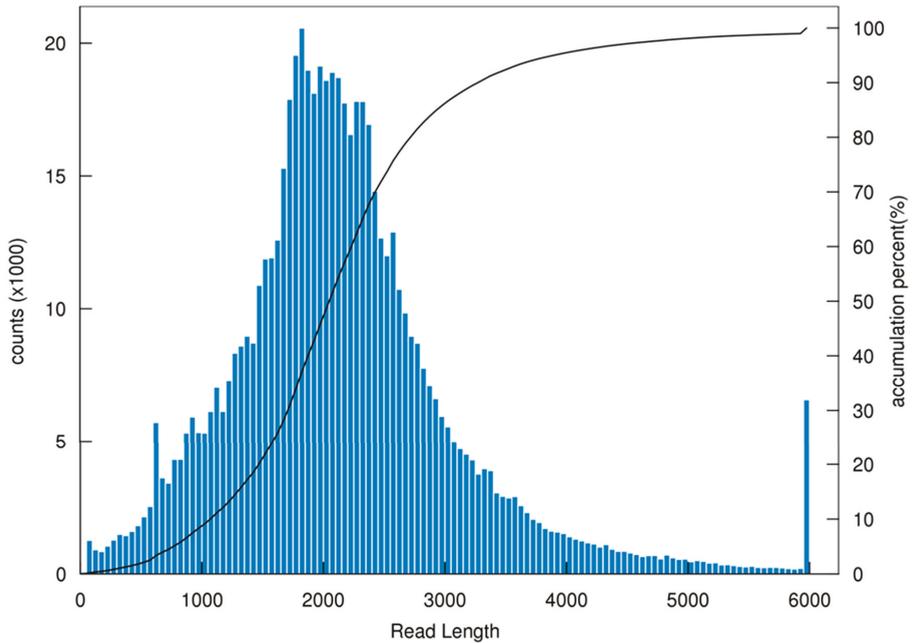


Figure 1. Length distribution of 651,260 reads of insert in avocadomesocarp.

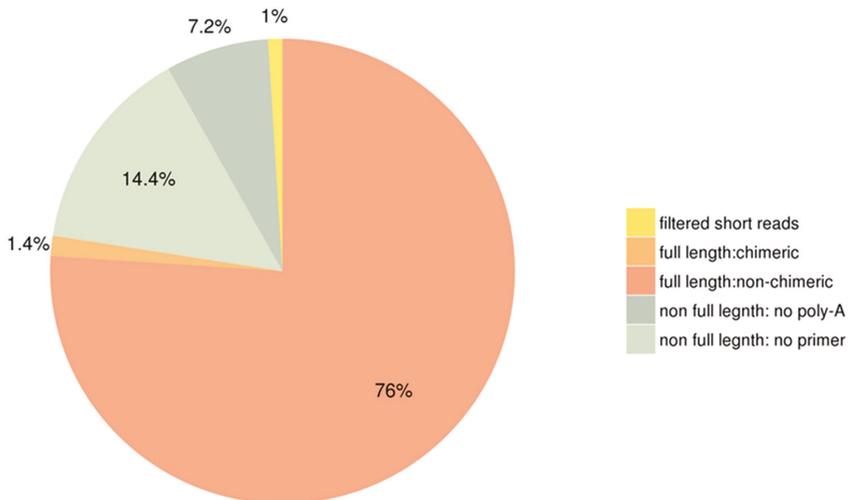


Figure 2. Classification of reads of insert in avocadomesocarp.

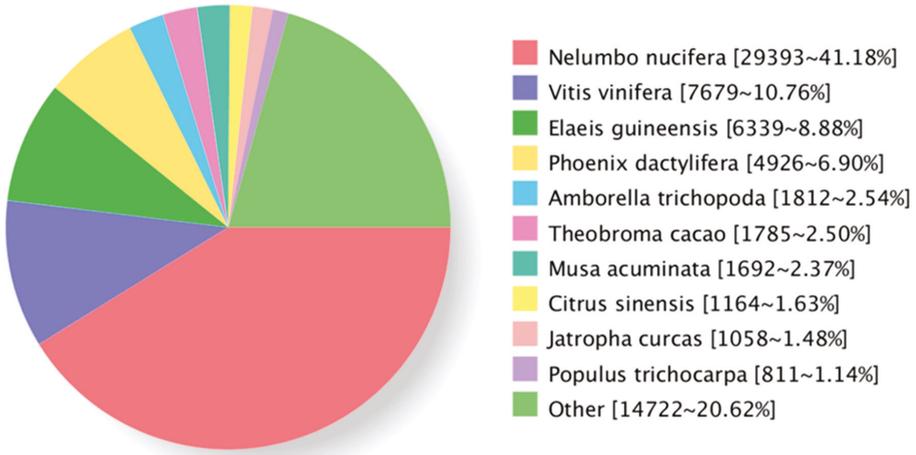
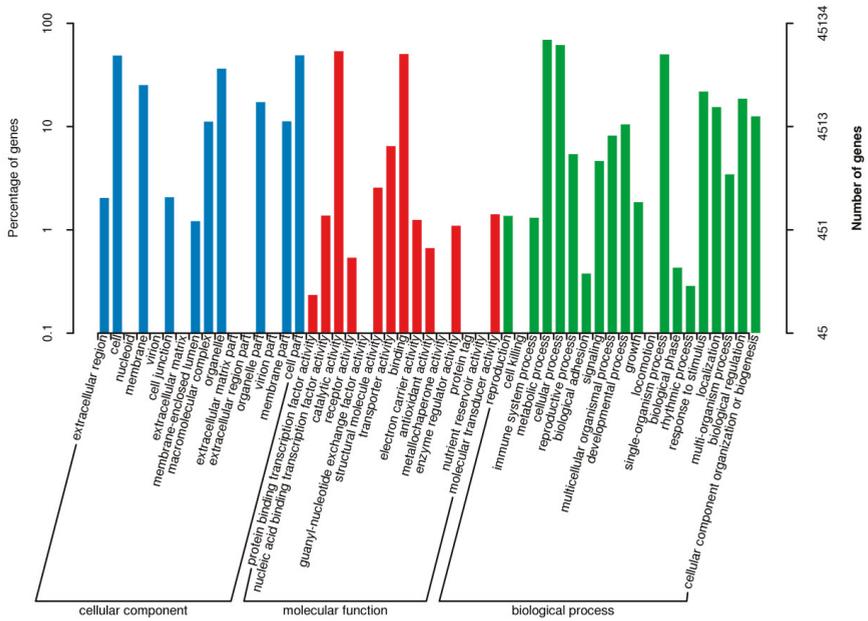
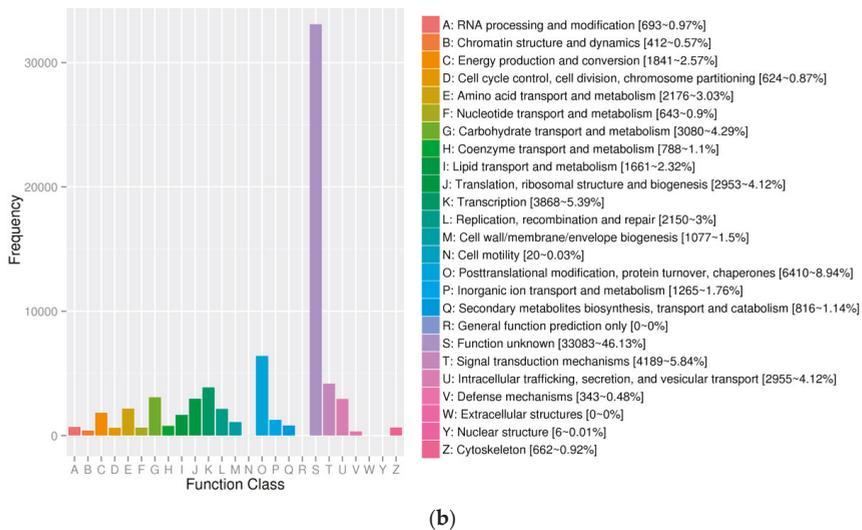


Figure 3. Species most closely related to avocado based on the NCBI nonredundant protein sequence database.



(a)

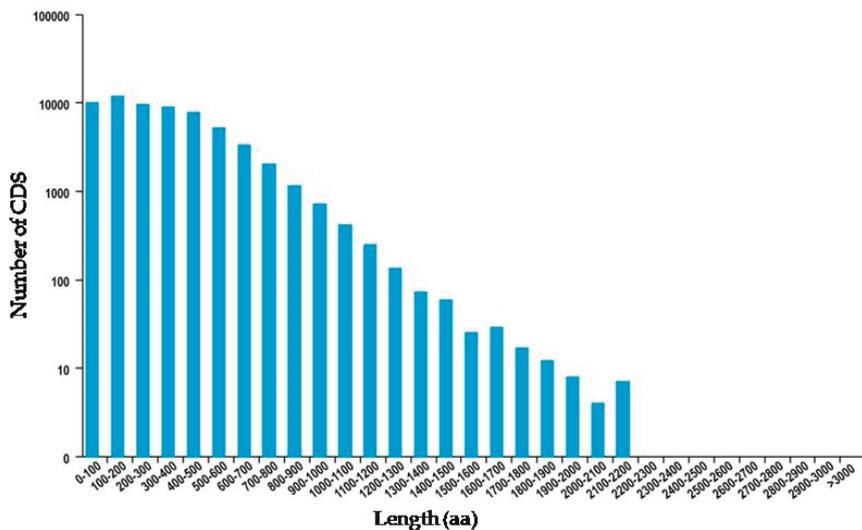
Figure 4. Cont.



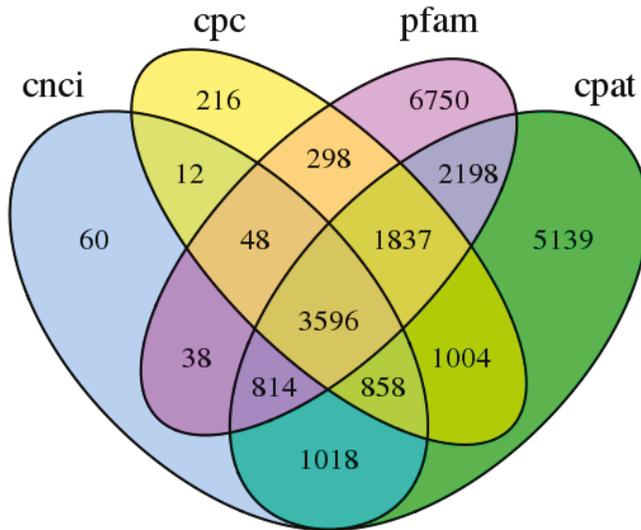
**Figure 4.** Functional classification of transcripts. The predicted functions were based on Gene Ontology (a) and Non-supervised Orthologous Groups (b) databases.

### 3.2. Predictions of ORFs, TFs, and lncRNAs

A total of 73,946 ORFs were predicted, 61,523 of which were complete CDSs. The number and length distribution of proteins encoded by the CDS regions are presented in Figure 5 and Additional file 1. A total of 7969 putative avocado TFs distributed in 203 families were identified (Table S7). The most abundant TF categories included RLK-Pelle\_DLSV (241) and C3H (240). Additionally, the CPC, CNCI, CPAT, and Pfam database were combined to distinguish lncRNA candidates from putative protein-coding RNAs among the unannotated transcripts. Analyses with the CPC, CNCI, CPAT, and Pfam database revealed 7869, 6444, 16,464, and 15,579 transcripts longer than 200 nt with more than two exons as lncRNA candidates. A total of 3596 lncRNA transcripts were predicted (Figure 6).



**Figure 5.** Distribution of 61,523 complete coding sequences for the avocado open reading frames.



**Figure 6.** The number of long non-coding RNA transcripts predicted in avocado based on the Coding Potential Calculator, Coding-Non-Coding Index, Coding Potential Assessment Tool, and Pfam database.

3.3. Frequency and Distribution of Various Types of EST-SSR Loci

The 75,946 transcript sequences comprising 170,959,769 bp detected in this study included 42,096 sequences containing 76,777 SSR loci (Table 1). Of these SSR-containing transcript sequences, 19,825 harbored more than one SSR locus. Mononucleotide motifs were the most abundant (44,800, 58.35%), followed by di- (18,903; 24.62%), tri- (11,724, 15.27%), tetra- (788, 0.01%), hexa- (321, 0.00%), and pentanucleotide (241, 0.00%) motif repeats (Table 2).

**Table 1.** Details regarding the simple sequence repeats (SSRs) identified from single-molecule real-time (SMRT) sequencing in avocado mesocarp.

Source	Number
Total number of sequences examined	75,956
Total size of examined sequences (bp)	170,959,769
Total number of identified SSRs	76,777
Number of SSR containing sequences	42,096
Number of sequences containing more than 1 SSR	19,825
Number of SSRs present in compound formation	12,675

**Table 2.** Details regarding the number of repeating units at avocado expressed sequence tag-simple sequence repeat (EST-SSR) loci.

SSR Motif Length	Repeat Unit Number							Total	%
	5	6	7	8	9	10	>10		
Mono-	-	-	-	-	-	8951	35,849	44,800	58.35
Di-	-	4017	2773	2489	2013	1547	6064	18,903	24.62
Tri-	6129	2838	1237	720	403	193	204	11,724	15.27
Tetra-	541	175	41	21	8	-	2	788	0.01
Penta-	172	67	1	1	-	-	-	241	0.00
Hexa-	228	72	15	3	2	-	1	321	0.00
Total	7070	7169	4067	3234	2426	10,691	42,120	76,777	
%	9.21	9.34	5.30	4.21	3.16	13.92	45.14		

There were 5–1343 SSRs per locus. Moreover, SSRs with more than 10 repeats were the most abundant, followed by those with 10, 6, and 5 random repeats. Among the 139 different repeat types, (A/T)<sub>n</sub> was the most common (56.63%). The six other main motif types were (AG/CT)<sub>n</sub> (19.14%), (AAG/CTT)<sub>n</sub> (5.97%), (AT/AT)<sub>n</sub> (3.35%), (AGC/CTG)<sub>n</sub> (2.18%), and (AC/GT)<sub>n</sub> (2.04%) (Table S8).

### 3.4. Development of Polymorphic EST-SSR Markers, Analysis of Genetic Diversity, and KASP genotyping

Using Primer3, we developed 149,733 EST-SSR markers from the 49,911 SSR loci (Table S9). To verify the amplification of the EST-SSR markers, a subset of 100 EST-SSR markers was randomly chosen and tested with seven accessions from various regions in southern China (Table S10). The primers for 30 of the tested markers generated amplification products, whereas 37 primer pairs amplified nonpolymorphic products and 33 did not produce clear amplicons. The 30 polymorphic EST-SSR markers, which included 15 di-, 5 tri-, 5 tetra-, 2 penta-, and 3 hexanucleotidemotif-based markers, were further verified with 46 avocado accessions. Finally, 15 polymorphic EST-SSR markers, with missing allele frequencies <10% for all 46 avocado accessions, were selected for subsequent analyses of genetic diversity (Table S11). A total of 71 alleles in the 46 avocado accessions carried the 15 polymorphic EST-SSR markers. Eight of these alleles were considered to be accession-specific and the other 63 alleles were generally found in multiple accessions (Table S11). The eight accession-specific alleles were from the following accessions: Renong No. 4, Renong, No. 5; Renong No. 6, Guiyan No. 8, Daling No. 5, Daling No. 6, RL chang, and RL yuan.

The 15 polymorphic EST-SSRs were applied to evaluate diversity parameters (Table 3). The Na amplified per SSR locus varied from 2 to 10, with a mean of 4.73. The Ne varied from 1.04 to 4.39, with an average of 2.31, and Ho ranged from 0.04 to 0.93, with an average of 0.49. The He ranged from 0.04 to 0.77, with an average of 0.50, and PIC values ranged from 0.04 to 0.74, with an average of 0.45.

**Table 3.** Diversity parameters associated with 15 polymorphic EST-SSRs analyzed in 46 avocado accessions.

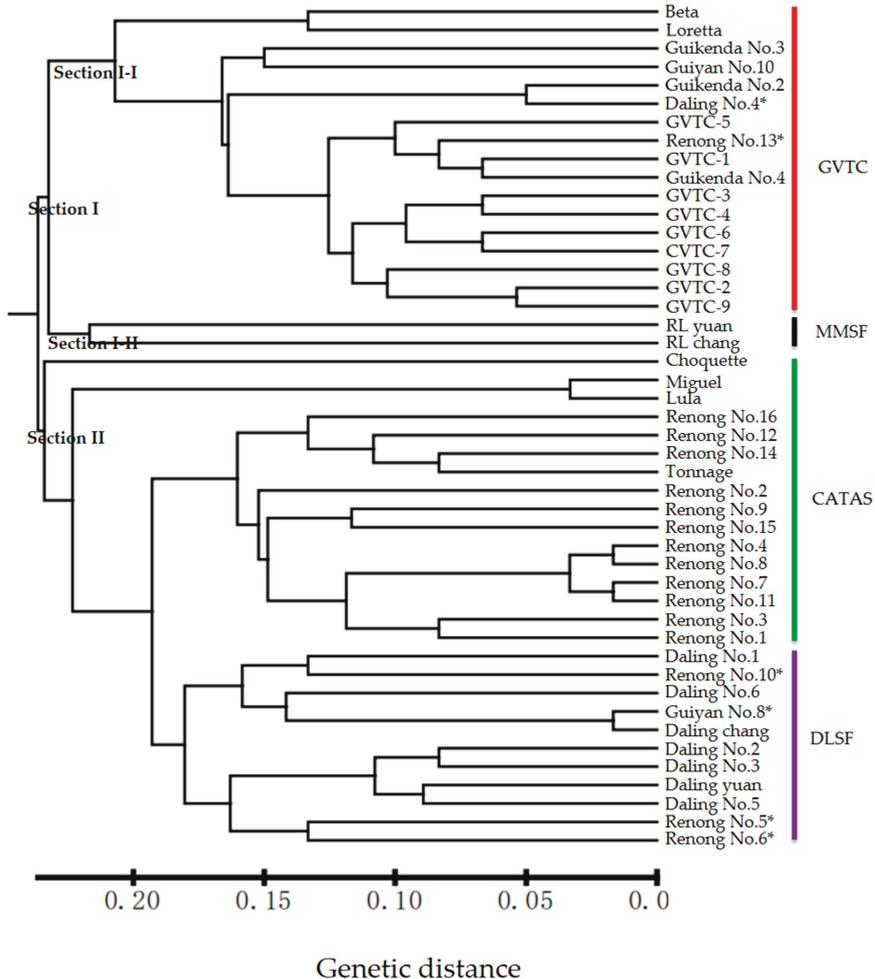
Marker Name	Transcript ID	Na <sup>1</sup>	Ne <sup>2</sup>	Ho <sup>3</sup>	He <sup>4</sup>	PIC <sup>5</sup>
Pa-eSSR-17	F01_cb7709_c10/f1p0/2063	8	3.02	0.61	0.67	0.62
Pa-eSSR-18	F01_cb7876_c2/f1p0/2226	10	3.09	0.61	0.68	0.65
Pa-eSSR-19	F01_cb1803_c26/f1p0/2838	6	2.04	0.63	0.51	0.46
Pa-eSSR-20	F01_cb10663_c1/f1p0/2458	3	1.87	0.50	0.46	0.40
Pa-eSSR-21	F01_cb15691_c2/f1p0/2049	3	2.41	0.50	0.58	0.50
Pa-eSSR-22	F01_cb3034_c12/f2p0/2705	5	2.85	0.67	0.65	0.60
Pa-eSSR-23	F01_cb12182_c0/f6p2/1774	3	1.47	0.28	0.32	0.29
Pa-eSSR-24	F01_cb13109_c0/f3p0/1635	5	2.80	0.48	0.64	0.58
Pa-eSSR-25	F01_cb1901_c3/f1p1/2722	2	1.04	0.04	0.04	0.04
Pa-eSSR-26	F01_cb7204_c7/f10p1/2700	3	2.65	0.93	0.62	0.55
Pa-eSSR-27	F01_cb10594_c1/f1p0/4058	3	1.40	0.33	0.29	0.27
Pa-eSSR-28	F01_cb9432_c36/f1p2/1811	5	1.56	0.43	0.36	0.33
Pa-eSSR-29	F01_cb15387_c0/f3p0/1548	8	4.39	0.49	0.77	0.74
Pa-eSSR-30	F01_cb12814_c24/f1p0/3423	4	2.67	0.53	0.62	0.55
Pa-eSSR-31	F01_cb10835_c0/f4p0/2019	3	1.33	0.28	0.25	0.22
	Total	71				
	Mean	4.73	2.31	0.49	0.50	0.45

<sup>1</sup> Number of observed alleles; <sup>2</sup> effective number of alleles; <sup>3</sup> observed heterozygosity; <sup>4</sup> expected heterozygosity; <sup>5</sup> polymorphism information content.

Six race-specific KASP markers were used to determine the race of 33 avocado accessions with an unknown race. The KASP genotyping results demonstrated that all 33 avocado accessions were Guatemalan × West Indian hybrids based on the corresponding genotype of each racial avocado (Table S2).

*3.5. Analyses of Genetic Relationships Based on Polymorphic EST-SSRs from SMRT Sequencing Data*

A cluster analysis grouped the 46 accessions into two major sections (Figure 7). The dendrogram revealed a clear separation between the native avocado accessions from Hainan province and those from Guangxi and Yunnan provinces. In cluster I, 19 Guatemalan × West Indian hybrids were clustered into two sub-sections. Sub-cluster I-I consisted of 13 native Guatemalan × West Indian hybrids from Guangxi province. Sub-cluster I-II contained two native Guatemalan × West Indian hybrids from Yunnan province. Cluster II comprised 27 Guatemalan × West Indian hybrids from Hainan province. Among these hybrids, 15 and 6 were obtained from the CATAS and DLSE, respectively.



**Figure 7.** Neighbor-joining consensus tree of 1000 bootstrap replicates revealing the phylogenetic relationships among the 46 analyzed avocado accessions based on the shared alleles for the 15 EST-SSR markers. GVTC, native avocado accessions from Guangxi Vocational and Technical College; MMSF, native avocado accessions from Mengmao State Farm; CATAS, native avocado accessions from the Chinese Academy of Tropical Agricultural Sciences; and DLSE, native avocado accessions from Daling State Farm. The native avocado accessions labeled with an asterisk originated from other regions.

Figure 8 presents the distribution of the 46 avocado accessions for the first two principal coordinates of a principal coordinate analysis (PCoA). On the basis of the first coordinate, which accounted for 21.71% of the total variation, the accessions were generally distributed in two groups. The native avocado accessions from Hainan and Yunnan provinces were basically grouped separately from the native avocado accessions from Guangxi province. The second coordinate accounted for 10.06% of the total variation. Finally, we observed that the native avocado accessions were generally grouped according to their geographical origins.



**Figure 8.** Principal coordinate analysis of 46 avocado accessions based on the 15 EST-SSR markers. POP1, avocado accessions from Florida, USA; POP2, native avocado accessions from the Chinese Academy of Tropical Agricultural Sciences; POP3, native avocado accessions from Mengmao State Farm; POP4, native avocado accessions from Daling State Farm; and POP5, native avocado accessions from Guangxi Vocational and Technical College.

**4. Discussion**

Transcriptome sequencing is a useful technique for obtaining a large number of transcripts for organisms lacking a reference sequence, at least partly because it is inexpensive and can be completed rapidly [51–53]. To date, several short-read next-generation sequencing (NGS) transcriptome databases have been developed for avocado mesocarp samples [54,55] and avocado mixed tissue samples [18,56]. However, both the number and length of the transcript sequences derived from these short-read NGS studies have hampered their application in genetics and molecular biology research [41]. One of the advances in sequencing technology has been the development of the long-read SMRT sequencing

technique, which enables researchers to obtain a substantial number of full-length sequences from a cDNA library [32]. In the current study, we applied the PacBio SMRT system to generate and analyze the full-length transcriptome of avocado mixed mesocarp samples collected at various developmental stages. The 25.79 Gb SMRT data produced in this study provide the first comprehensive insights into the avocado mesocarp, which is the most economically valuable organ of this fruit species, and might serve as the genetic basis for future research on avocado. Interestingly, the full-length transcriptome sequence described herein is also the first such sequence for a plant species from the family Lauraceae.

In this study, 93.82% (71,627 of 76,345) of the nonredundant transcripts were annotated based on similarities with sequences in public databases. Thus, a greater proportion of transcripts were annotated in this study than in previous investigations involving NGS data for various avocado races (49.00%) [18] and for avocado mesocarp samples (57.50%) [55]. We determined that the mean length of the avocado nonredundant transcripts was 2330 bp, implying that our sequences were long enough to represent full-length transcripts. Additionally, this mean length was in between the mean lengths obtained for other species, including *Z. bungeanum* (3414 bp) [32], *T. pretense* (2789 bp) [34], *M. sativa* (1706 bp) [37], *Z. planispinum* (1781 bp) [38], *C. sinensis* (1781 bp) [40], and *Arabidopsis pumila* (2194 bp) [57]. Moreover, the 76,345 nonredundant transcripts derived from the 25.79 Gb clean PacBio SMRT data produced in this study may facilitate future research on the physiology, biochemistry, and molecular genetics of avocado and related species.

A previous study indicated that lncRNAs may be important for the gene regulation in eukaryotic cells, especially during some key biological processes [58]. However, the number of lncRNAs encoded in genomes as well as their characteristics remains largely unknown [59]. Predicting and functionally annotating lncRNAs is challenging, but valuable because they are not orthologous and there is a lack of homologous sequences between closely related species [38]. Unfortunately, very few of the lncRNA functions have been elucidated [60,61]. Hence, the lncRNA information for one species is not suitable for predicting the lncRNAs in another species. In this study, 3596 avocado transcript sequences (accounting for 4.71% of the total number of nonredundant transcripts) were putatively predicted as lncRNAs. This almost completely uncharacterized gene pool may include genes associated with agronomically relevant traits related to the most economically valuable organ (mesocarp).

The accurate identification of avocado germplasm races is needed to ensure that germplasm collections are optimally used by plant breeders and farmers worldwide [1]. The traditional assignment of avocado races based on morphological traits is imprecise because of environmental effects and a limited number of applicable characteristics [17]. Molecular-based characterizations are more consistent and valid for assigning avocado genotypes. We previously confirmed the universality of six race-specific KASP markers [47]. These markers were used in the current study to identify avocado accessions with an unknown race, with implications for the application of available avocado germplasms for breeding and resource conservation. Interestingly, the KASP genotyping results revealed that all of the native avocado accessions included in this study are Guatemalan × West Indian hybrids. The reason for this observation might be related to the introduction of avocado cultivars and the climates of the sample collection regions. First, the major avocado cultivars grown commercially are typically hybrids of three races (i.e., mainly Guatemalan × West Indian and Guatemalan × Mexican hybrids) [1]. Since the late 1950s, Guatemalan × West Indian and Guatemalan × Mexican hybrids have been brought into China from other countries for cultivation in Southern China [9]. Second, the native avocado accessions included in the present study are mainly from three geographical regions, namely Nanning located in the central and southern region of Guangxi province, Danzhou and Baisha located in the central and western region of Hainan province, and Ruili located in the western region of Yunnan province. These locations are characterized by a warm and humid oceanic climate with a relatively low altitude in the central and southern region of Guangxi province and the central and western region of Hainan province. Although Ruili is located in the western region of Yunnan province and far from the ocean, it still has a subtropical monsoon climate. The climates of these three regions resemble that of the areas in which the West Indian races originated, and are favorable for the growth of Guatemalan ×

West Indian hybrids. Therefore, Guatemalan  $\times$  West Indian hybrids may have gradually become the dominant native avocado accessions because of artificial selection or via naturally occurring crosses.

The 100 EST-SSR markers randomly selected for validation in the present study had an amplification rate of 67%, and 30 were determined to be polymorphic. This polymorphism level is generally consistent with that of our previous study [18]. In subsequent analyses of the genetic diversity of these polymorphic EST-SSR markers among 46 avocado accessions, 15 markers produced 4.73 alleles per locus, which was fewer than the 6.13 alleles per locus of Ge [18], the 11.40 alleles per SSR locus of Gross-German and Viruel [17], the 18.8 alleles per SSR locus of Schnell [16], and the 9.75 alleles per SSR locus of Alcaraz and Hormaza [62]. Additionally, a PIC value  $> 0.5$  is generally considered to represent a high polymorphism rate [63]. In this study, 7 of 15 polymorphic EST-SSRs had a PIC value  $< 0.5$ . This result may have been because the 46 avocado accessions in this study are genotypically the same (Guatemalan  $\times$  West Indian hybrids), with relatively low genetic diversity.

In this study, a cluster analysis and a PCoA grouped the native avocado accessions according to where they originated. Additionally, some of the native avocado accessions derived from different regions was included in the same sub-cluster. For example, Renong No. 13 from Hainan province clustered with the native accessions from Guangxi province. One factor leading to this promiscuous clustering is the fact that avocado germplasm resources have been exchanged among researchers and breeders since the late 1980s. The CATAS, which is a national scientific research unit, was commissioned to popularize superior avocado accessions among breeders at adjacent state farms or at other national scientific research units. Some superior native accessions from the CATAS may be the male or female parent of other native accessions from various state farms or other national scientific research units, which is consistent with our study results. Furthermore, a cluster analysis grouped two native avocado accessions from Yunnan province with the native avocado accessions from Guangxi province. In contrast, our PCoA indicated that these two native avocado accessions from Yunnan province belong to the same group as the native avocado accessions from Hainan province. We speculate that the relatively few native avocado accessions from Yunnan province (i.e., two) may have led to these contradictory results based on two statistical analyses. At many avocado plantations in Yunnan province, the local avocado accessions have been replaced by “Hass,” which is the most economically valuable avocado cultivar, ultimately making it difficult to collect local avocado accessions. Thus, maximizing the economic benefits of cultivating specific avocado cultivars, while ensuring avocado genetic resources are conserved will need to be addressed.

## 5. Conclusions

We annotated SMRT sequencing data based on the COG, GO, KEGG, KOG, Pfam, Swiss-Prot, eggNOG, and Nr databases. Among 71,627 transcripts, 45,134, 52,125, and 33,310 were annotated according to GO, eggNOG, and KEGG classifications, respectively. We detected 76,777 SSR loci in 42,096 transcript sequences and used them to develop 149,733 EST-SSR markers. From a randomly selected subset comprising 100 EST-SSR markers, we finally identified 15 polymorphic EST-SSR markers on 71 alleles, which had 2–10 of these markers per locus. A cluster analysis and a PCoA separated the 46 avocado accessions according to their geographical origins. These 15 newly developed EST-SSR markers may be useful for future analyses of avocado accessions and may contribute to the improved management of avocado resources for germplasm conservation and breeding programs.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4395/9/9/512/s1>. Table S1. Sources of the 46 avocado accessions evaluated in this study. Table S2. KASP primer information and KASP genotyping results. Table S3. Gene annotations of the 71,627 avocado transcripts. Table S4. Characteristics of the GO annotation of avocado transcripts. Table S5. Characteristics of eggNOG classifications of avocado transcripts. Table S6. Characteristics of KEGG pathways of avocado transcripts. Table S7. Transcription factors identified in the avocado transcripts. Table S8. Frequencies of different repeat motifs in EST-SSRs from avocado. Table S9. Characteristics of avocado EST-SSR markers in this study. Table S10. Summary of 100 EST-SSR markers used for amplification. Table S11. Summary of 15 EST-SSRs in 46 avocado accessions. Additional file 1. Coding sequences predicted with TransDecoder.

**Author Contributions:** Y.G., R.Z., and W.M. conceived and designed the experiments; J.W., Y.L. (Yuanzheng Liu), and N.W. performed the experiments; L.T. and D.C. analyzed the data; Y.L. (Yanxia Li) helped complete the experiments; X.Z. contributed materials; and Y.G. wrote the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (grant number 31701883) and the Natural Science Foundation of Hainan Province of China (grant number 319QN266).

**Acknowledgments:** We gratefully acknowledge Pingzhen Lin from the Haikou Experimental Station of the Chinese Academy of Tropical Agricultural Sciences for supporting the collection of avocado resources. We thank Yajima for editing the English text of a draft of this manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Schaffer, B.; Wolstenholme, B.N.; Whiley, A.W. *The Avocado: Botany, Production and Uses*, 2nd ed.; CPI Group (UK) Ltd.: Croydon, UK, 2012.
2. Kopp, L.E. A taxonomic revision of the genus *Persea* in the western hemisphere (*Persea-Lauraceae*). *Mem. N. Y. Bot. Gard.* **1966**, *14*, 1–120.
3. Williams, L.O. The avocado, a synopsis of the genus *Persea*, subg. *Persea* *Econ. Bot.* **1977**, *31*, 315–320. [[CrossRef](#)]
4. Schaffer, B.; Wolstenholme, B.N. *The Avocado: Botany, Production and Uses*; CAB International: Wallingford, UK, 2002.
5. Van der Werff, H. A synopsis of *Persea* (Lauraceae) in Central America. *Novon* **2002**, *12*, 575–586. [[CrossRef](#)]
6. Dreher, M.L.; Davenport, A.J. Hass avocado composition and potential health effects. *Crit. Rev. Food Sci.* **2013**, *53*, 738–750. [[CrossRef](#)]
7. Galvão, M.D.S.; Narain, N.; Nigam, N. Influence of different cultivars on oil quality and chemical characteristics of avocado fruit. *Food Sci. Technol.* **2014**, *34*, 539–546. [[CrossRef](#)]
8. Ge, Y.; Si, X.Y.; Cao, J.Q.; Zhou, Z.X.; Wang, W.L.; Ma, W.H. Morphological characteristics, nutritional quality, and bioactive constituents in fruits of two avocado (*Perseaamericana*) varieties from hainan province, China. *J. Agric. Sci.* **2017**, *9*, 8–17. [[CrossRef](#)]
9. Ge, Y.; Si, X.Y.; Lin, X.E.; Wang, J.S.; Zang, X.P.; Ma, W.H. Advances in avocado (*Perseaamericana* Mill.). *South China Fruit* **2017**, *46*, 63–70. [[CrossRef](#)]
10. Zhang, L.; Zhang, D.S.; Liu, K.D. Environmental analysis and countermeasures for industrial development of Hainan avocado. *Chin. J. Agric. Resour. Reg. Plan.* **2015**, *36*, 78–84.
11. Fiedler, J.; Bufler, G.; Bangerth, F. Genetic relationships of avocado (*Perseaamericana* Mill.) using RAPD markers. *Euphytica* **1998**, *101*, 249–255. [[CrossRef](#)]
12. Mhameed, S.; Sharon, D.; Kaufman, D.; Lahav, E.; Hillel, J.; Degani, C.; Lavi, U. Genetic relationships within avocado (*Perseaamericana* Mill.) cultivars and between *Persea* species. *Theor. Appl. Genet.* **1997**, *94*, 279–286. [[CrossRef](#)]
13. Furnier, G.R.; Cummings, M.P.; Clegg, M.T. Evolution of the avocados as revealed by DNA restriction site variation. *J. Hered.* **1990**, *81*, 183–188. [[CrossRef](#)]
14. Davis, J.; Henderson, D.; Kobayashi, M. Genealogical relationships among cultivated avocado as revealed through RFLP analysis. *J. Hered.* **1998**, *89*, 319–323. [[CrossRef](#)]
15. Ashworth, V.E.T.M.; Clegg, M.T. Microsatellite markers in avocado (*Perseaamericana* Mill.). genealogical relationships among cultivated avocado genotypes. *J. Hered.* **2003**, *94*, 407–415. [[CrossRef](#)]
16. Schnell, R.J.; Brown, J.S.; Olano, C.T.; Power, E.J.; Krol, C.A.; Kuhn, D.N.; Motamayor, J.C. Evaluation of avocado germplasm using microsatellite markers. *J. Am. Soc. Hortic. Sci.* **2003**, *128*, 881–889. [[CrossRef](#)]
17. Gross-German, E.; Viruel, M.A. Molecular characterization of avocado germplasm with a new set of SSR and EST-SSR markers: Genetic diversity, population structure, and identification of race-specific markers in a group of cultivated genotypes. *Tree Genet. Genomes* **2013**, *9*, 539–555. [[CrossRef](#)]
18. Ge, Y.; Tan, L.; Wu, B.; Wang, T.; Zhang, T.; Chen, H.; Zou, M.; Ma, F.; Xu, Z.; Zhan, R. Transcriptome sequencing of different avocado ecotypes: De novo transcriptome assembly, annotation, identification and validation of EST-SSR markers. *Forests* **2019**, *10*, 411. [[CrossRef](#)]
19. Chen, H.; Morrel, P.L.; Ashworth, V.E.T.M.; De la Cruz, M.; Clegg, M.T. Nucleotide diversity and linkage disequilibrium in wild avocado (*Perseaamericana* Mill.). *J. Hered.* **2008**, *99*, 382–389. [[CrossRef](#)]

20. Chen, H.; Morrel, P.L.; Ashwoth, V.E.T.M.; De la Cruz, M.; Clegg, M.T. Tracing the geographic origins of mayor avocado cultivars. *J. Hered.* **2009**, *100*, 56–65. [[CrossRef](#)]
21. Hou, M.Y.; Mu, G.J.; Zhang, Y.J.; Cui, S.L.; Yang, X.L.; Liu, L.F. Evaluation of total flavonoid content and analysis of related EST-SSR in Chinese peanut germplasm. *Crop Breed. Appl. Biotechnol.* **2017**, *17*, 221–227. [[CrossRef](#)]
22. Azevedo, A.O.N.; Azevedo, C.D.O.; Santos, P.H.A.D.; Ramos, H.C.C.; Boechat, M.S.B.; Arêdes, F.A.S.; Ramos, S.R.R.; Mirizola, L.A.; Perera, L.; Aragão, W.M.; et al. Selection of legitimate dwarf coconut hybrid seedlings using DNA fingerprinting. *Crop Breed. Appl. Biotechnol.* **2018**, *18*, 409–416. [[CrossRef](#)]
23. Ahmad, A.; Wang, J.D.; Pan, Y.B.; Sharif, R.; Gao, S.J. Development and use of simple sequence repeats (SSRs) markers for sugarcane breeding and genetic studies. *Agronomy* **2018**, *8*, 260. [[CrossRef](#)]
24. Ferreira, F.; Scapim, C.A.; Maldonado, C.; Mora, F. SSR-based genetic analysis of sweet corn inbred lines using artificial neural networks. *Crop Breed. Appl. Biotechnol.* **2018**, *18*, 309–313. [[CrossRef](#)]
25. Ge, Y.; Hu, F.C.; Tan, L.; Wu, B.; Wang, T.; Zhang, T.; Ma, F.N.; Cao, J.Q.; Xu, Z.N.; Zhan, R.L. Molecular diversity in a germplasm collection of avocado accessions from the tropical and subtropical regions of China. *Crop Breed. Appl. Biotechnol.* **2019**, *19*, 153–160. [[CrossRef](#)]
26. Wu, J.; Cai, C.F.; Cheng, F.Y.; Cui, H.L.; Zhou, H. Characterization and development of EST-SSR markers in tree peony using transcriptome sequences. *Mol. Breed.* **2014**, *34*, 1853–1866. [[CrossRef](#)]
27. Biswas, M.K.; Nath, U.K.; Howlader, J.; Bagchi, M.; Natarajan, S.; Kayum, M.A.; Kim, H.T.; Park, J.I.; Kang, J.G.; Nou, I.S. Exploration and exploitation of novel SSR markers for candidate transcription factor genes in *Lilium* species. *Genes* **2018**, *9*, 97. [[CrossRef](#)]
28. Ma, S.L.Y.; Dong, W.S.; Lyu, T.; Lyu, Y.M. An RNA sequencing transcriptome analysis and development of EST-SSR markers in Chinese hawthorn through Illumina sequencing. *Forests* **2019**, *10*, 82. [[CrossRef](#)]
29. Li, X.; Li, M.; Hou, L.; Zhang, Z.Y.; Pang, X.M.; Li, Y.Y. De novo transcriptome assembly and population genetic analyses for an endangered Chinese endemic *Acer miaotaiense* (Aceraceae). *Genes* **2018**, *9*, 378. [[CrossRef](#)]
30. Qi, W.C.; Chen, X.; Fang, P.H.; Shi, S.C.; Li, J.J.; Liu, X.T.; Cao, X.Q.; Zhao, N.; Hao, H.Y.; Li, Y.J.; et al. Genomic and transcriptomic sequencing of *Rosa hybrida* provides microsatellite markers for breeding, flower trait improvement and taxonomy studies. *BMC Plant Biol.* **2018**, *18*, 119. [[CrossRef](#)]
31. Chen, J.; Tang, X.H.; Ren, C.X.; Wei, B.; Wu, Y.Y.; Wu, Q.H.; Pei, J. Full-length transcriptome sequences and the identification of putative genes for flavonoid biosynthesis in safflower. *BMC Genom.* **2018**, *19*, 548. [[CrossRef](#)]
32. Tian, J.Y.; Feng, S.J.; Liu, Y.L.; Zhao, L.L.; Tian, L.; Hu, Y.; Yang, T.X.; Wei, A.Z. Single-molecule long-read sequencing of *Zanthoxylum bungeanum* maxim. transcriptome: Identification of aroma-related genes. *Forests* **2018**, *9*, 765. [[CrossRef](#)]
33. Roberts, R.J.; Carneiro, M.O.; Schatz, M.C. The advantages of SMRT sequencing. *Genome Biol.* **2013**, *14*, 405–409. [[CrossRef](#)]
34. Chao, Y.H.; Yuan, J.B.; Li, S.F.; Jia, S.Q.; Han, L.B.; Xu, L.X. Analysis of transcripts and splice isoforms in red clover (*Trifolium pratense* L.) by single-molecule long-read sequencing. *BMC Plant Biol.* **2018**, *18*, 300. [[CrossRef](#)]
35. Hoang, N.V.; Furtado, A.; Mason, P.J.; Marquardt, A.; Kasirajan, L.; Thirugnanasambandam, P.P.; Botha, F.C.; Henry, R.J. A survey of the complex transcriptome from the highly polyploid sugarcane genome using full-length isoform sequencing and de novo assembly from short read sequencing. *BMC Genom.* **2017**, *18*, 395. [[CrossRef](#)]
36. Zuo, C.M.; Blow, M.; Sreedasyam, A.; Kuo, R.C.; Ramamoorthy, G.K.; Torres-Jerez, I.; Li, G.F.; Wang, M.; Dilworth, D.; Barry, K.; et al. Revealing the transcriptomic complexity of switchgrass by PacBio long-read sequencing. *Biotechnol. Biofuels* **2018**, *11*, 170. [[CrossRef](#)]
37. Chao, Y.H.; Yuan, J.B.; Guo, T.; Xu, L.X.; Mu, Z.Y.; Han, L.B. Analysis of transcripts and splice isoforms in *Medicago sativa* L. by single-molecule long-read sequencing. *Plant Mol. Biol.* **2019**, *99*, 219–235. [[CrossRef](#)]
38. Kim, J.A.; Roy, N.S.; Lee, I.H.; Choi, A.Y.; Choi, B.S.; Yu, Y.S.; Park, N.I.; Park, K.C.; Kim, S.; Yang, H.S.; et al. Genome-wide transcriptome profiling of the medicinal plant *Zanthoxylum planispinum* using a single-molecule direct RNA sequencing approach. *Genomics* **2019**, *111*, 973–979. [[CrossRef](#)]
39. Zhang, B.; Liu, J.X.; Wang, X.S.; Wei, Z.W. Full-length RNA sequencing reveals unique transcriptome composition in bermudagrass. *Plant Physiol. Biochem.* **2018**, *132*, 95–103. [[CrossRef](#)]

40. Xu, Q.S.; Zhu, J.Y.; Zhao, S.Q.; Hou, Y.; Li, F.D.; Tai, Y.L.; Wan, X.C.; Wei, C.L. Transcriptome profiling using single-molecule direct RNA sequencing approach for in-depth understanding of genes in secondary metabolism pathways of *Camellia sinensis*. *Front. Plant Sci.* **2017**, *8*, 1205. [[CrossRef](#)]
41. Deng, Y.; Zheng, H.; Yan, Z.C.; Liao, D.Y.; Li, C.L.; Zhou, J.Y.; Liao, H. Full-length transcriptome survey and expression analysis of *Cassia obtusifolia* to discover putative genes related to auranio-obtusin biosynthesis, seed formation and development, and stress response. *Int. J. Mol. Sci.* **2018**, *19*, 2476. [[CrossRef](#)]
42. Ge, Y.; Cheng, Z.H.; Si, X.Y.; Ma, W.H.; Tan, L.; Zang, X.P.; Wu, B.; Xu, Z.N.; Wang, N.; Zhou, Z.X.; et al. Transcriptome profiling provides insight into the genes in carotenoid biosynthesis during the mesocarp and seed developmental stages of avocado (*Persea Americana*). *Int. J. Mol. Sci.* **2019**, *20*, 4117. [[CrossRef](#)]
43. Ge, Y.; Ramchiary, N.; Wang, T.; Liang, C.; Wang, N.; Wang, Z.; Choi, S.R.; Lim, Y.P.; Piao, Z.Y. Development and linkage mapping of unigene-derived microsatellite markers in *Brassica rapa* L. *Breed. Sci.* **2011**, *61*, 160–167. [[CrossRef](#)]
44. Kanehisa, M.; Araki, M.; Goto, S.; Hattori, M.; Hirakawa, M.; Itoh, M.; Katayama, T.; Kawashima, S.; Okuda, S.; Tokimatsu, T.; et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **2008**, *36*, 480–484. [[CrossRef](#)]
45. Götz, S.; García-Gómez, J.M.; Terol, J.; Williams, T.D.; Nagaraj, S.H.; Nueda, M.J.; Robles, M.; Talon, M.; Dopazo, J.; Conesa, A. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* **2008**, *36*, 3420–3435. [[CrossRef](#)]
46. Perez-Rodriguez, P.; Riano-Pachon, D.M.; Correa, L.G.; Rensing, S.A.; Kersten, B.; Mueller-Roeber, B. PlnTFDB: Updated content and new features of the plant transcription factor database. *Nucleic Acids Res.* **2010**, *38*, 822–827. [[CrossRef](#)]
47. Ge, Y.; Zhang, T.; Wu, B.; Tan, L.; Ma, F.N.; Zou, M.H.; Chen, H.H.; Pei, J.L.; Liu, Y.Z.; Chen, Z.H.; et al. Genome-wide assessment of avocado germplasm determined from specific length amplified fragment sequencing and transcriptomes: Population structure, genetic diversity, identification, and application of race-specific markers. *Genes* **2019**, *10*, 215. [[CrossRef](#)]
48. Krawczak, M.; Nikolaus, S.; von Eberstein, H.; Croucher, P.J.; El Mokhtari, N.E.; Schreiber, S. PopGen: Population based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet.* **2006**, *9*, 55–61. [[CrossRef](#)]
49. Liu, K.; Muse, S.V. PowerMarker: An integrated analysis environment for genetic marker analysis. *Bioinformatics* **2005**, *21*, 2128–2129. [[CrossRef](#)]
50. Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **2013**, *30*, 2725–2729. [[CrossRef](#)]
51. Du, M.; Li, N.; Niu, B.; Liu, Y.; You, D.; Jiang, D.; Ruan, C.Q.; Qin, Z.Q.; Song, T.W.; Wang, W.T. De novo transcriptome analysis of *Bagarius yarrelli* (Siluriformes: Sisoridae) and the search for potential SSR markers using RNA-Seq. *PLoS ONE* **2018**, *13*, e0190343. [[CrossRef](#)]
52. Liu, F.M.; Hong, Z.; Yang, Z.J.; Zhang, N.N.; Liu, X.J.; Xu, D.P. De Novo transcriptome analysis of *Dalbergia odorifera* T. Chen (Fabaceae) and transferability of SSR markers developed from the transcriptome. *Forests* **2019**, *10*, 98. [[CrossRef](#)]
53. Li, W.; Zhang, C.P.; Jiang, X.Q.; Liu, Q.C.; Liu, Q.H.; Wang, K.L. De Novotranscriptomic analysis and development of EST-SSRs for *Styrax japonicas*. *Forests* **2018**, *9*, 748. [[CrossRef](#)]
54. Kilaru, A.; Cao, X.; Dabbs, P.B.; Sung, H.J.; Rahman, M.M.; Thrower, N.; Zynda, G.; Podicheti, R.; Ibarra-Laclette, E.; Herrera-Estrella, L.; et al. Oil biosynthesis in a basal angiosperm: Transcriptome analysis of *Persea Americana* mesocarp. *BMC Plant Biol.* **2015**, *15*, 203. [[CrossRef](#)]
55. Vergara-Pulgar, C.; Rothkegel, K.; González-Agüero, M.; Pedreschi, R.; Campos-Vargas, R.; Defilippi, B.G.; Meneses, C. De novo assembly of *Persea americana* cv. ‘Hass’ transcriptome during fruit development. *BMC Genom.* **2019**, *20*, 108. [[CrossRef](#)]
56. Ibarra-Laclette, E.; Méndez-Bravo, A.; Pérez-Torres, C.A.; Albert, V.A.; Mockaitis, K.; Kilaru, A.; López-Gómez, R.; Cervantes-Luevano, J.I.; Herrera-Estrella, L. Deep sequencing of the Mexican avocado transcriptome, an ancient angiosperm with a high content of fatty acids. *BMC Genom.* **2015**, *16*, 599. [[CrossRef](#)]
57. Yang, L.F.; Jin, Y.H.; Huang, W.; Sun, Q.; Liu, F.; Huang, X.Z. Full-length transcriptome sequences of ephemeral plant *Arabidopsis pumila* provides insight into gene expression dynamics during continuous salt stress. *BMC Genom.* **2018**, *19*, 717. [[CrossRef](#)]

58. Liu, J.; Wang, H.; Chua, N.H. Long noncoding RNA transcriptome of plants. *Plant Biotechnol. J.* **2015**, *13*, 319–328. [[CrossRef](#)]
59. Yandell, M.; Ence, D. A beginner's guide to eukaryotic genome annotation. *Nat. Rev. Genet.* **2012**, *13*, 329–342. [[CrossRef](#)]
60. Liu, J.; Jung, C.; Xu, J.; Wang, H.; Deng, S.; Bernad, L.; Arenas-Huertero, C.; Chua, N.H. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. *Plant Cell* **2012**, *24*, 4333–4345. [[CrossRef](#)]
61. Ochogavía, A.; Galla, G.; Seijo, J.G.; González, A.M.; Bellucci, M.; Pupilli, F.; Barcaccia, G.; Albertini, E.; Pessino, S. Structure, target-specificity and expression of *PN\_LNC\_N13*, a long non-coding RNA differentially expressed in apomictic and sexual *Paspalumnotatum*. *Plant Mol. Biol.* **2018**, *96*, 53–67. [[CrossRef](#)]
62. Alcaraz, M.L.; Hormaza, J.I. Molecular characterization and genetic diversity in an avocado collection of cultivars and local Spanish genotypes using SSRs. *Heredity* **2007**, *144*, 244–253. [[CrossRef](#)]
63. Botstein, D.; White, R.L.; Skolnick, M.; Davis, R.W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **1980**, *32*, 314–331.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



Article

# Assessment of Genetic Diversity in Differently Colored Raspberry Cultivars Using SSR Markers Located in Flavonoid Biosynthesis Genes

Vadim G. Lebedev <sup>1,2</sup>, Natalya M. Subbotina <sup>1,2</sup>, Oleg P. Maluchenko <sup>3</sup>,  
Konstantin V. Krutovsky <sup>4,5,6,7,8,\*</sup> and Konstantin A. Shestibratov <sup>2</sup>

<sup>1</sup> Pushchino State Institute of Natural Sciences, Prospekt Nauki 3, 142290 Pushchino, Moscow Region, Russia

<sup>2</sup> Branch of the Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry of the Russian Academy of Sciences, Prospekt Nauki 6, 142290 Pushchino, Moscow Region, Russia

<sup>3</sup> All-Russian Research Institute of Agricultural Biotechnology, Timiriazevskaya Str. 42, 127550 Moscow, Russia

<sup>4</sup> Department of Forest Genetics and Forest Tree Breeding, Faculty of Forest Sciences and Forest Ecology, Georg-August University of Göttingen, Büsgenweg 2, D-37077 Göttingen, Germany

<sup>5</sup> Center for Integrated Breeding Research (CiBreed), Georg-August University of Göttingen, Albrecht-Thaer-Weg 3, D-37075 Göttingen, Germany

<sup>6</sup> Laboratory of Population Genetics, N. I. Vavilov Institute of General Genetics, Russian Academy of Sciences, Gubkin Str. 3, 119333 Moscow, Russia

<sup>7</sup> Laboratory of Forest Genomics, Genome Research and Education Center, Siberian Federal University, 660036 Krasnoyarsk, Russia

<sup>8</sup> Department of Ecosystem Science and Management, Texas A&M University, 2138 TAMU, College Station, TX 77843-2138, USA

\* Correspondence: konstantin.krutovsky@forst.uni-goettingen.de; Tel.: +49-551-339-3537

Received: 4 August 2019; Accepted: 4 September 2019; Published: 6 September 2019

**Abstract:** Raspberry is a valuable berry crop containing a large amount of antioxidants that correlates with the color of the berries. We evaluated the genetic diversity of differently colored raspberry cultivars by the microsatellite markers developed using the flavonoid biosynthesis structural and regulatory genes. Among nine tested markers, seven were polymorphic. In total, 26 alleles were found at seven loci in 19 red (*Rubus idaeus* L.) and two black (*R. occidentalis* L.) raspberry cultivars. The most polymorphic marker was *RiMY01* located in the MYB10 transcription factor intron region. Its polymorphic information content (PIC) equalled 0.82. The *RiG001* marker that previously failed to amplify in blackberry also failed in black raspberry. The raspberry cultivar clustering in the UPGMA dendrogram was unrelated to geographical and genetic origin, but significantly correlated with the color of berries. The black raspberry cultivars had a higher homozygosity and clustered separately from other cultivars, while at the same time they differed from each other. In addition, some of the raspberry cultivars with a yellow-orange color of berries formed a separate cluster. This suggests that there may be not a single genetic mechanism for the formation of yellow-orange berries. The data obtained can be used prospectively in future breeding programs to improve the nutritional qualities of raspberry fruits.

**Keywords:** flavonoid biosynthesis; fruit coloration; marker-assisted selection; microsatellites; *Rubus*

## 1. Introduction

The genus *Rubus* L. (Rosaceae, Rosoideae) is one of the most diverse in the plant kingdom and contains between 600 and 800 species grouped in 12 subgenera, which are widely distributed throughout the world from the lowland tropics to subarctic regions [1]. Among these species, red

raspberry (*Rubus idaeus* L.) and blackberry (several species in the genus *Rubus*) grown world-wide, and black raspberry (*R. occidentalis* L.) grown mainly in the United States, are of the greatest economic importance. Their berries are in great demand due to their flavor, color, and taste. In addition, they are very healthy providing a good source of antioxidants, including phenolic acids, flavonoids, anthocyanins, and carotenoids [2]. Berries contain four times more antioxidants than non-berry fruits, 10 times more than vegetables, and 40 times than cereals [3]. For this reason, berries and their products (i.e., berry juice and jam) are very often recognized as “superfoods” [4]. The popularity of this crop can be indicated by the fact that their harvest increased 1.5 times from 2010 to 2017 worldwide and exceeded 800,000 tons [5]. Russia consistently ranks first in the world for the raspberry production. The growing interest in raspberry has led not only to an increase in its production, but also to the expansion of breeding programs for the development of new cultivars. However, classical selection takes a lot of time: in red raspberry, it can take up to 15 years for development and release of a new cultivar [6]. Moreover, a specific feature in the *Rubus* spp. breeding system is that multiple species are often utilized in breeding programs [7]. Scientific achievements in molecular biology, and use of molecular markers, in particular, can accelerate the selection process, as they will allow for the assessment of the seedlings with valuable traits at a much earlier stage. Molecular genetic markers provide more reliable cultivar identification of *Rubus* species than morphological markers [8].

In order to speed up the breeding process, it is useful to have genetic linkage maps containing information about the markers associated with the most important traits, including disease and pest resistance, plant habitus, nutritional and sensory fruit quality, and plant architecture. The first genetic linkage map of *Rubus* was constructed from a cross between two *Rubus* subspecies, *R. idaeus* (cv. Glen Moy) × *R. strigosus* (cv. Latham), in 2004 [9]. After that, other molecular maps for red raspberry [10–12], black raspberry [13] and tetraploid blackberry [14] appeared. Quantitative trait loci (QTL) have been identified for important traits including resistance to diseases [11,15] and pests [10], fruit anthocyanin content [16], growth characteristics [10,17], fruit color and quality traits [18]. Currently, molecular markers are routinely used in breeding raspberries for resistance to the *Phytophthora* root rot at the James Hutton Institute (UK), and two promising genotypes are under commercial trials, as well as markers for the quality of berries are in the process of validation [19]. If in the first reports a combination of various types of molecular markers such as AFLP and simple sequence repeat (SSR) [9,10], RAPD, and RGAP [11] were used, then the most recent molecular maps were produced using only molecular markers designed from sequenced DNA such as microsatellites or SSR markers [13,20]. SSRs are DNA tandem repeats of the 1–6 nucleotide long motifs that are very frequent in genomes. They are very polymorphic with high information content, co-dominant inheritance, locus specificity, extensive genome coverage and simple detection using labelled primers that flank the microsatellite [9,21], and their ability to distinguish even closely related individuals is particularly important for many crop species [21]. Raspberry researchers have noted the benefits of the SSR markers, but very few molecular markers still exist for *Rubus* [7,22]. It should be also acknowledged that the breeding process can be accelerated using genomic selection (e.g., [23]), an approach under rapid adoption in many species, which is based on multiple marker–trait associations and does not require linkage maps.

The color of the berries not only affects their attractiveness but also serves as an indicator of the content of biologically active compounds. For example, the content of anthocyanins in raspberry berries varies widely from 2 to 325 mg/100 g depending on the color of the berries [24]. Flavonols and anthocyanins are synthesized in the flavonoid pathway, and its enzymes are well characterized. Kassim et al. [16] mapped QTLs for individual anthocyanin pigments in raspberry. The genes of various enzymes of flavonoid biosynthesis were also identified in red [18] and black [25] raspberry and blackberry [26]. Besides the structural genes, regulatory genes are important in the biosynthesis of flavonoids. The late flavonoid biosynthetic genes are activated by the ternary transcriptional MYB-bHLH-WD40 (MBW) complex comprising three classes of regulatory proteins including R2R3-MYBs, bHLHs, and TTG1 (WD40) [27]. Transcription factor genes, such as *MYB10*, *bHLH* and *bZIP*, have also been identified in the *Rubus* species [18,26].

There are several studies that used random genomic SSR markers to assess genetic diversity in cultivars within [8,28] and between [29] different species. However, we are unaware of studies in which genetic diversity would be assessed using markers located in genes of any metabolic pathway and the biosynthesis of flavonoids, in particular. In this study, we developed SSR markers using nucleotide sequences of structural and regulatory genes of flavonoid biosynthesis in *Rubus* and *Fragaria* (strawberry) available at the National Center for Biotechnology Information (NCBI) GenBank database to test whether genetic variation associated with these genes correlate with a variation of berry colors. These markers were genotyped in 19 raspberry cultivars from different geographic regions (Russia, Poland, Italy, Switzerland, UK, and USA) and two cultivars of black raspberry. If alleles at these loci correlate with the content of biologically active substances, they could subsequently be used to optimize selection for valuable traits associated with color and, indirectly, with the content of flavonoids, by accelerating selection via screening genotypes at early stages.

## 2. Materials and Methods

### 2.1. Plant Materials

Nineteen cultivars of red raspberry (Amira, Anne, Babye Leto II, Beglyanka, Brilliantovaya, Bryanskoe Divo, Gerakl, Glen Ample, Marosejka, Meteor, Oranzhevoe Chudo, Pingvin, Polka, Poranna Rosa, Solnyshko, Sugana, Tarusa, Zheltyj Gigant, and Zolotaya Osen) and two cultivars of black raspberry (Cumberland and Jewel) were chosen to genotype SSR loci located in the flavonoid biosynthesis genes. These cultivars have a wide range of fruit color from yellow to black with various geographic and genetic origins, but cultivars of Russian origin from two raspberry breeding centers (Bryansk and Moscow) dominated in the list (Table 1). Raspberry plants used in this study were kindly provided by Dr. I. A. Pozdniakov (OOO Mikroklon, Pushchino, Russia). Each cultivar represented a microclonally vegetatively propagated line containing practically genetically identical plants. Therefore, a single specimen per culture was used for further DNA isolation and genotyping.

**Table 1.** Parentage and fruit color of the *Rubus* cultivars used in the study.

Cultivar	Abbr.	Genetic Origin and Background	Fruit Color	Origin
<i>R. idaeus</i> (red raspberry)				
Amira	Ami	Polka × Tulameen	red	Italy
Anne	Ann	Amity × Glenn Garry	yellow	USA
Babye Leto II	BL2	Autumn Bliss × Babye Leto	red	Russia (Bryansk)
Beglyanka	Beg	Kostinbrodskaya × Novost Kuzmina	orange	Russia (Bryansk)
Brilliantovaya	Bri	open pollination of interspecific hybrids	red	Russia (Bryansk)
Bryanskoe Divo	BrD	47-18-4 (open pollination)	light-red	Russia (Bryansk)
Gerakl	Ger	Autumn Bliss × 14-205-4	red	Russia (Bryansk)
Glen Ample	GAm	SCRI7326EI × SCRI7412H16	dark red	UK
Marosejka	Mar	7324/50 × 7331/3	light-red	Russia (Moscow)
Meteor	Met	Kostinbrodskaya × Novost Kuzmina	red	Russia (Bryansk)
Oranzhevoe Chudo	OrC	Shapka Monomaha (open pollination)	orange	Russia (Bryansk)
Pingvin	Pin	interspecific hybrid	dark red	Russia (Bryansk)
Polka	Pol	P89141 (open pollination)	red	Poland
Poranna Rosa	PoR	83291 × ORUS 1098-1	yellow	Poland
Solnyshko	Sol	Kostinbrodskaya × Novost Kuzmina	red	Russia (Bryansk)
Sugana	Sug	Autumn Bliss × Tulameen	light-red	Switzerland
Tarusa	Tar	Stolichnaya × Shtambovyj-1	red	Russia (Moscow)
Zheltyj Gigant	ZhG	Marosejka × Ivanovskaya	yellow	Russia (Moscow)
Zolotaya Osen	ZOs	13-39-11 (open pollination)	yellow	Russia (Bryansk)
<i>R. occidentalis</i> (black raspberry)				
Cumberland	Cum	Gregg selfed	blue-black	USA
Jewel	Jew	(Bristol × Dundee) × Dundee	black	USA

## 2.2. Simple Sequence Repeat (SSR) Marker and Polymerase Chain Reaction (PCR) Primer Development

The WebSat software [30] was used to detect SSR loci in the nucleotide sequences of *Rubus* and *Fragaria* × *ananassa* (the garden strawberry or simply strawberry, a widely grown hybrid species of the genus *Fragaria*) flavonoid biosynthesis genes available at the NCBI GenBank database (<http://www.ncbi.nlm.nih.gov>) (Table 2). The Primer 3 software (<http://primer3.org>) was used to design appropriate polymerase chain reaction (PCR) primers based on the sequences flanking the SSR loci. The minimum number of motifs used to select the SSR locus was nine for mono-nucleotide repeats, five for di-nucleotide motifs, three for tri-, and tetra-, and two for penta-, and hexa-nucleotide repeats. Primers were designed using the following criteria: primer length of 18–27 bp (optimally 22 bp), GC content of 40%–80%, annealing temperature of 57–68 °C (optimally 60 °C), and expected amplified product size of 100–400 bp. Primers for the *RiG001* locus were as in [8]. Primers were synthesized by Syntol Company (Moscow, Russia) and are summarized in Table 2.

## 2.3. DNA Isolation, PCR Amplification and Fragment Analysis

A single DNA sample per each cultivar was produced from young expanding leaves representing a single plant per each cultivar. Total genomic DNA was extracted using the STAB method [31]. The quality and quantity of extracted DNA were determined by the NanoDrop 2000 spectrophotometer (ThermoFisher). The final concentration of each DNA sample was adjusted to 50 ng/μL in TE buffer before the PCR amplification.

For genotyping, PCR was performed separately for each primer pair using a forward primer labeled with the fluorescent dye 6-FAM and an unlabeled reverse primer (Syntol, Russia). The PCR amplification was performed in a total volume of 20 μL consisted of 50 ng of genomic DNA, 10 pmol of the labeled forward primer, 10 pmol of an unlabeled reverse primer, and PCR Mixture Screenmix (Eurogen, Russia). After an initial denaturation at 95 °C for 3 min, DNA was amplified during 33 cycles in a gradient thermal cycler (Bio-Rad, Hercules, CA, USA) programmed for a 30 s denaturation step at 95 °C, a 20 s annealing step at the optimal annealing temperature of the primer pair and a 35 s extension step at 72 °C. A final extension step was done at 72 °C for 5 min.

The PCR generating clear, stable, and specific DNA fragments within an expected length (200–400 bp) were considered as successful PCR amplifications. If a primer pair failed three times to amplify template DNA that was amplified with other primers, then it was scored as a null genotype.

Separation of amplified DNA fragments was performed in an ABI 3130xl Genetic Analyzer using S450 LIZ size standard (Syntol Company, Moscow, Russia). Peak identification and fragment sizing were done using the Gene Mapper v4.0 software (Applied Biosystems, Foster, CA, USA).

## 2.4. Genetic Data Analysis

Genetic parameters were calculated for 21 raspberry cultivars based on seven SSR polymorphic loci. The allele frequencies, number of alleles, observed ( $H_o$ ) and expected ( $H_e$ ) heterozygosities, and polymorphic information content (PIC) were calculated using the PowerMarker v.3.25 software [32]. This software was also used to estimate pairwise Nei's standard genetic distances between each pair of cultivars and to generate a UPGMA dendrogram, which was visualized using the Statistica software (TIBCO Software Inc., Palo Alto, CA, USA).

**Table 2.** Data on nine simple sequence repeat (SSR) loci located in the flavonoid biosynthesis genes and their polymerase chain reaction (PCR) primer pairs used to study genetic diversity in *Rubus* cultivars.

Locus	Gene	Species	NCBI GenBank Accession Number	Motif and Number of Repeats	Location in the Gene	PCR Primer Nucleotide Sequence		T °, °C	Allele Size, bp		
						Forward	Reverse		Expected	Observed	
RIG001	aromatic polyketide synthase (PKS3)	<i>R. idaeus</i>	AF292369	(AT) <sub>6</sub>	intron	TGTCGATCCTTTTC TTTGG	CGCTTCTGATCCTT GACTTGT	55	345	349, 350, 351	0
RcFH01	flavanone-3-hydroxylase (F3H)	<i>R. coraninus</i>	EU255776	(TATG) <sub>3</sub>	intron	GGTCCAAGTGCATTC TATTAC	GTTCCTGAACTCCC GTTCCT	60	262	255, 265, 271	255, 271
FafS01	flavonol synthase (FLS)	<i>Fragaria × ananassa</i>	DQ834905	(CT) <sub>12</sub>	intron	CATCCCTAATGCCCTA GTCATC	TGTACTTCGGTGA TTCCTCTT	60	304	323, 328	323
FafS02	flavonol synthase (FLS)	<i>F. manassai</i>	DQ834905	(GGAAG) <sub>2</sub>	exon	AAAGCTCCTCAAACAA ATCTTCG	GTAGTTAATGGCAGAA GGTGGC	60	273	255, 271	271
RiA501	anthocyanidin synthase (ANS)	<i>R. idaeus</i>	KV950789	(ATCTC) <sub>2</sub>	exon	TCAAACAAGGAGAAAGGT GAGGAT	CCGTTACGAGAGATCAA ACCAAG	60	334	309, 333, 358	309
FaA R01 **	anthocyanidin reductase (ANR)	<i>F. manassai</i>	DQ664193	(TGCCTG) <sub>2</sub> (CATTT) <sub>2</sub>	exon intron	AATCTCTCTGTGGTCG GTACAT	AGAGAGATATGGTCTTC GCCTTG	60	244	250	250
RhLF01 **	UDP-glucose flavonoid 3-O-glycosyltransferase-like protein (UFGT)	<i>R. hybrid</i>	JF764808	(GAG) <sub>7</sub> (ACAAGC) <sub>2</sub>	exon	AGGAGCTGAAGAAAAG ACTCCA	AAAGTCTCTTAGGTT TCCCCTC	60	275	267, 270	269, 270
RmMY01 **	transcription factor MYB10	<i>R. idaeus</i>	EU155165	(TAATA) <sub>2</sub> (CT) <sub>7</sub> (AT) <sub>15</sub>	introns	GTTCCTCTCCAAGCA GGTTATT	TGCAAAAGTCTCTCTC TTGATG	59	330	323, 325, 327, 329, 331, 333, 341, 342	358
RiT01	transparent testa glabra 1 (TTG1) protein	<i>R. idaeus</i>	HM579852	(CAC) <sub>5</sub>	exon	ACTCCACACAAGAATC CCATCT	CTGTTTCAAGACCG AAATTC	60	379	379	379

\* Optimal annealing temperature. \*\* Two or three SSRs in these loci were amplified simultaneously by a single pair of primers.

### 3. Results

#### 3.1. Polymorphism and Genetic Diversity Analysis

Nine SSR markers (six based on *Rubus* and three on *Fragaria* nucleotide sequences of the flavonoid biosynthesis genes) were used to estimate genetic diversity in 19 raspberry (*R. idaeus*) and two black raspberry (*R. occidentalis*) cultivars. All PCR primer pairs amplified one or two alleles. In raspberries, two loci (*RiTT01* and *FaAR01*) were monomorphic, and other seven were polymorphic. In black raspberry cultivars, the *RiG001* was not amplified at all, six loci were monomorphic and only two polymorphic (Table 2). In total, 26 alleles were found in seven polymorphic microsatellite loci. The number of alleles per locus varied from two per locus (*FaFS02* and *FaFL01*) to nine per locus (*RiMY01*) with an average number of 3.7 alleles per locus (Table 3). The *RiMY01* locus was the most polymorphic. In general, the SSR loci located in introns were more polymorphic than loci in exons.

**Table 3.** Parameters of genetic variation for seven polymorphic SSR loci in 21 *Rubus* cultivars.

Locus	Location in the Gene	Major Allele Frequency	Number of Alleles	Heterozygosity		Polymorphism Information Content (PIC)
				Expected ( $H_e$ )	Observed ( $H_o$ )	
<i>RiG001</i>	intron	0.81	4	0.33	0.19	0.31
<i>RcFH01</i>	intron	0.74	3	0.41	0.52	0.35
<i>FaFS01</i>	intron	0.76	2	0.36	0.48	0.30
<i>FaFS02</i>	exon	0.98	2	0.05	0.05	0.05
<i>RiAS01</i>	exon	0.79	3	0.36	0.19	0.33
<i>RhUF01</i>	exon	0.90	3	0.18	0.00	0.17
<i>RiMY01</i>	introns	0.29	9	0.84	0.57	0.82
Mean		0.75	3.71	0.36	0.29	0.33

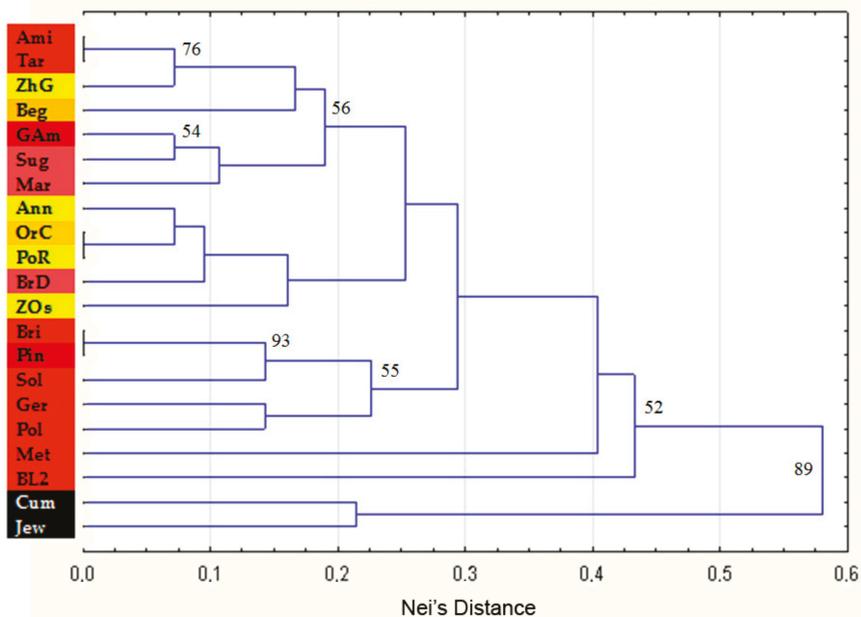
There were cultivar-specific alleles, such as a unique allele 358 at the *RiMY01* locus found only in black raspberry, and alleles 267 and 269 at the *RhUF01* locus found only in the red raspberry Meteor and Jewel cultivars, respectively. Meteor contained also a unique allele 333 at the *RiMY01* locus.

Parameters of genetic variation for seven polymorphic SSR loci in 21 *Rubus* cultivars are presented in Table 3. Expected heterozygosity ( $H_e$ ) ranged from 0.05 in the *RiMY01* locus up to 0.84 in the *RiMY01* locus with an average value of 0.36. Observed heterozygosity was zero in the *RhUF01* locus and ranged from 0.05 in the *FaFS02* locus to 0.57 in the *RiMY01* locus with an average value of 0.29. The observed heterozygosity was lower than expected in four microsatellite loci and on average (Table 3). On average, the expected and observed heterozygosities were higher for the SSRs in introns (0.49 and 0.44, respectively) compared to the SSRs in exons (0.20 and 0.08, respectively). The average PIC was 0.332 and varied from 0.05 in the *FaFS02* locus to 0.82 in the *RiMY01* locus (Table 3).

#### 3.2. Cluster Analysis

A UPGMA dendrogram was constructed for 21 raspberry cultivars based on seven SSR markers located in the genes of the flavonoid biosynthesis (Figure 1). The dendrogram clearly separates red and black raspberries. Among the red raspberry cultivars, there is a group of cultivars with yellow-orange colored berries (Anne, Poranna Rosa, Orangevoe Chudo, and Zolotaya Osen), which forms a separate cluster. The same group includes also the Bryanskoe Divo cultivar with light red berries. At the same time, the Zheltyj Gigant (yellow berries) and Beglyanka (orange berries) were not included in this group. Separation of cultivars did not follow their genetic origin. The cultivars Beglyanka, Solnyshko, and Meteor having the same genetic origin from the Kostinbrodskaya × Novost Kuzmina cross were completely separated from each other. In addition, the Babye Leto 2 also having an ancestral hybrid (Autumn Bliss × (September × (Kostinbrodskaya × Novost Kuzmina))) turned out to differ mostly from other raspberry cultivars. Gerakl and Sugana both also having Autumn Bliss as their parent species

were significantly separated. At the same time, close similarities have been observed for cultivars from different geographic regions. No genetic differences were found between the Orangevoe Chudo (Russia) and Poranna Rosa (Poland) cultivars, and between the Amira (Italy) and Tarusa (Russia) cultivars, although they have different genetic origins. The Brilliantovaya and Pingvin cultivars were also identical and were obtained with the use of interspecific hybrids.



**Figure 1.** The UPGMA dendrogram of the 21 *Rubus* cultivars based on pairwise Nei’s standard genetic distances calculated using seven SSR markers located in the flavonoid biosynthesis genes. Left column shows the colors of the cultivar berries. Only bootstrap values larger than 50% are presented. See Table 1 for the full cultivar names.

#### 4. Discussion

SSR markers (microsatellites) are widely used in genetic diversity studies, QTL and genetic mapping, molecular-assisted selection (MAS), and cultivar identification, because they are multi-allelic, co-dominant, highly informative, relatively accurate and easily detected [33]. SSR markers have been often used to map different types of *Rubus* [9,13], fingerprinting germplasm [34], and in studies of the genetic diversity and population structure within [28] and among [29] *Rubus* species. However, genetic diversity has not previously been studied in terms of any specific metabolic pathway genes that determine valuable breeding traits.

In this study, we report on the evaluation of a number of red and black raspberry cultivars using SSR loci representing known sequences of the flavonoid biosynthesis pathway genes, which synthesize biologically active substances with high antioxidant activity—flavonols and anthocyanins. Among these microsatellite loci, six (*RcFH01*, *FaFS01*, *FaFS02*, *RiAS01*, *FaAR01*, and *RhUF01*) were located in the structural genes of the flavonoid biosynthesis (*F3H*, *FLS*, *ANS*, *ANR*, and *UFGT*) and two (*RiMY01* and *RiTT01*) in the regulatory genes (*MYB10* and *TTG1*). Flavanone-3-hydroxylase (F3H) is a key enzyme in the flavonoid biosynthesis in plants, as it catalyzes formation of 3-hydroxy flavonol, a common precursor of anthocyanins, flavanols, and proanthocyanidins [35]. Particular attention was paid to the flavonol synthase gene, for which two loci were used. Flavonol synthase (FLS) is an important enzyme

of flavonoid pathway that catalyzes the formation of flavonols from dihydroflavonols, and thus may influence anthocyanin levels, as dihydroflavonols are intermediates in the production of both colored anthocyanins and colorless flavonols [36]. The anthocyanidin synthase (ANS) leads to the synthesis of the anthocyanidin, the first colored compound in the anthocyanin biosynthetic pathway, from which anthocyanidin reductase catalyzes the formation of proanthocyanidins (condensed tannins) [37]. The last common step for the production of stable anthocyanins is the glycosylation by the enzyme UDP-glucose/flavonoid 3-O-glucosyl transferase (UFGT) [38].

In addition, loci were used on the sequence of two transcription factors (MYB 10 and TTG1) that belong to the MBW complex, which regulates the production of the late biosynthetic genes [27]. For comparison, we also used a pair of primers designed for the *RiG001* locus using the sequence of the *R. idaeus* aromatic polyketide synthase (*PiPKS3*) gene, which was not amplified in blackberry cultivars [8]. The *RiPKS3* gene differed from the *RiPKS1* gene, encoding a typical chalcone synthase (CHS) catalyzing the first step of flavonoid biosynthesis, in four amino acid positions and produced in vitro predominantly p-coumaryltriacetic acid lactone and low levels of chalcone [39]. Within the PCR fragment amplified by the primers for the *RiG001* locus the sequence of the *RiPKS3* gene (NCBI GenBank AF292369) differed from the *RiPKS1* gene sequence (AF292367) by a two nucleotide long deletion (2 bp) and a single nucleotide insertion. Three alleles (349, 350, and 351 bp) were obtained for this locus (Table 2).

In addition to the sequences of the genes of the *Rubus* plants (*R. idaeus*, *R. coreanus*, and *R. hybrid*), we used the sequences of the genes from *Fragaria* × *ananassa*, which is a close relative of *Rubus* from the same sub-family, Rosoideae. The *Rubus* and *Fragaria* both have the same base chromosome number  $1n = 7$ , similar morphology and chloroplast and nuclear DNA phylogenies [13].

Among three most economically important types of raspberry, 19 cultivars of red raspberry with a wide range of berry color from various world breeding centers and two cultivars of black raspberry are mostly used. Both species, red (*R. idaeus*) and black (*R. occidentalis*) raspberry belong to the same subgenus *Idaobatus* (raspberries) and are diploids ( $2n = 2x = 14$ ), while blackberry species vary greatly in ploidy [34].

In our study, the average number of alleles for seven polymorphic SSR loci in the flavonoid biosynthesis genes was 3.71, the mean  $H_o$  and  $H_e$  were 0.286 and 0.360, respectively, and the mean PIC was 0.332. These values were generally lower than previously reported for *R. idaeus* [8] and *R. coreanus* [29], but quite comparable with the data for black raspberry cultivars [28]. Perhaps, this is due to the fact that red raspberry cultivars are, for the most part, complex hybrids with a limited genetic pool [34], and the selection for berries quality has further reduced their diversity. The level of expected heterozygosity ( $H_e$ ) was higher than observed ( $H_o$ ) both on average and in most individual loci. These data are different from other studies of the *Rubus* species, where these parameters were approximately equal [8,29], or even higher [28]. However, unlike those studies, where population samples were used, a collection of different cultures was used in this study, which is not a population sample, but a mixture of genotypes with different genetic background and origin. Therefore, it is expected to observe excess of expected heterozygosity in comparison to observed heterozygosity due to Wahlund effect.

Only the *RiMY01* locus was highly polymorphic (PIC = 0.82). This locus had three SSR regions, two of which represent dinucleotide repeats. These data coincide with the results of Castillo et al. [8], in which all three highly informative markers (PIC = 0.78–0.82) represented dinucleotide repeats. In *R. coreanus*, among five highly polymorphic markers (PIC > 0.7), four represented dinucleotide repeats, and one trinucleotide repeats [29]. The high variation of the *RiMY01* locus can be explained by its location in the first intron of the transcription factor MYB10. SSR markers located in introns were more variable in comparison to those located in exons (expected and observed heterozygosities averaged 0.49 and 0.44 vs. 0.20 and 0.08, respectively). Our results are in agreement with those of Garcia-Gomez et al. [40], which showed that SSRs in introns had a higher level of heterozygosity compared to SSRs in exons in *Prunus* species—0.65 vs. 0.17, respectively. Similar results were also

obtained in maize [41]. Significantly higher variation was observed also for SNPs in noncoding regions compared to coding ones [42]. In general, introns are more variable than exons, as they are under less selection pressure during the evolutionary process [43].

The length of most alleles at the *RiMY01* locus differ from each other by two nucleotide-long steps, which is consistent with dinucleotide repeats of the SSR motifs in this locus. However, imperfect repeats also often occur in the raspberry SSR loci. For instance, Fernandez et al. [34] has previously reported the alleles with length different by consecutive one nucleotide-long steps in the *Rubus57a* and *Rub5a* markers. This single nucleotide stepwise variation is expected for *Rub5a*, which is a SSR marker with a mononucleotide motif, but *Rubus57a* is a SSR marker with a dinucleotide motif. We also observed a few alleles with imperfect repeats, such as the unique allele 267 of the *RhUF01* locus in the Meteor cultivar, for which the perfect allele size is 270 following the trinucleotide motif GAG stepwise allelic variation.

The black raspberry cultivars were highly homozygous: six out of eight loci were monomorphic (Table 2). High homozygosity in black raspberry has been also found earlier by Lewers and Weber [44]. They noticed that the level of homozygosity for the black raspberry was 80%, but only 40% for the red raspberry. The 21 SSR loci were unable to distinguish between six of the black raspberry cultivars [28]. However, the black raspberry cultivars Cumberland and Jewel were well discriminated in this study. Despite the small number of loci used in our study, these two cultivars were also separated by two loci: *RcFH01* and *RhUF01*. In our study the red raspberry cultivars were easily discriminated from the black raspberry cultivars by a unique black raspberry specific allele 358 at the *RiMY01* locus and the allele 309 at the *RiAS01* locus, which occurred almost exclusively in the black raspberry cultivars, except the red raspberry cultivar Babye Leto 2. In addition, the *RiG001* locus was not amplified in black raspberry. The same was observed also in 48 earlier tested blackberry cultivars [8]. Thus, in respect to this locus, the black raspberry is closer to the wild blackberry than to the red raspberry, although it belongs to different subgenera. No amplification of *RiG001* and the unique allele 358 at the *RiMY01* locus can be used to separate the red raspberry cultivars from the black ones.

Cluster analysis of the SSR markers located in the genes of the biosynthesis of flavonoids showed a clear separation of the black raspberry (*R. occidentalis*) cultivars with black colored berries from the red raspberry (*R. idaeus*) cultivars with berries colored from yellow to dark red (Figure 1). It is important to note also that five cultivars with berries of similar shades of light red color (three with yellow berries, one with orange, and another with light red color) having completely different origin still clustered together into one sub-group. Perhaps, gene-targeted markers [45] such as SSR loci in the genes of the biosynthesis of flavonoids reflect better their genetic similarity for traits, such as color of their berries, likely controlled or affected by these genes, than random genomic SSR markers.

Castillo et al. [8] found that the primocane fruiting (fall fruiting) raspberry cultivars were grouped into a separate cluster. In Fernandez et al. [34] studies, it was shown that the majority of primocane-fruiting material from various breeding programs, as well as some very early ripening florican-fruiting genotypes are grouped into one cluster. This shows that cultivars can be grouped according to a particular trait regardless of their origin. At the same time, two cultivars with yellow and orange-colored fruits (Zhelytyj Gigant and Beglyanka) fell into another group of red-colored fruits. Perhaps, for a clearer separation, it is necessary to use additionally more polymorphic markers, including other genes of the biosynthesis of flavonoids not represented in this study.

Moreover, it is possible that the yellow color of the raspberry fruits can be obtained by two or more mechanisms. For example, primocane fruiting cultivars were also distributed in two different groups [34]. The genetic mechanisms for the formation of yellow color in raspberry fruit have not yet been fully studied. Although assumptions on this topic were made back in the 1930s, it was not until 2016 when an inactive anthocyanidin synthase (ANS) allele was identified in yellow raspberry [46]. A 5 bp insertion in the coding region of gene creates a premature stop codon resulting in a truncated amino acid sequence of the defective ANS protein. However, other mechanisms are also possible, such

as the combinations of recessive and dominant alleles, or the transcription factors that may lead to a huge variety of berry colors in raspberry.

The clustering along the flavonoid pathway also showed that there is a lack of connections between cultivars of the related origin. This is exactly the opposite data compared to the analyses carried out on randomly selected SSR markers evenly distributed across the genome. For example, Fernandez et al. [34] demonstrated that one cluster is almost entirely composed of cultivars from the Scottish raspberry breeding program or cultivars based on their germplasm. From the point of view of MAS the use of gene-targeted markers to assess genotypes for particular breeding traits is preferable to the use of random SSR markers. Graham et al. [9] suggested in 2004 that *Rubus idaeus* due to the diploid set of chromosomes ( $2n = 2x = 14$ ) and a very small genome (275 Mb) may be used as a model species for the Rosaceae. For many years, this was impeded by the lack of the full-genome *Rubus* sequence, although the genomes of other Rosaceae species have been already sequenced, such as apple in 2010, strawberry in 2011, pear and peach in 2013 [47]. However, the situation is changing with genomes of *R. occidentalis* [48] and *R. idaeus* [49] having been recently published. This will facilitate developing gene-targeted markers that can advance breeding *Rubus* for important traits including those related to the nutritional value of their berries.

## 5. Conclusions

In this study, we demonstrated that a set of gene-targeted SSR markers representing structural and regulatory genes of flavonoid biosynthesis could potentially allow more informative and meaningful evaluation of the genetic relationship between different cultivars of red and black raspberries that reflect the color of their berries and possibly also their nutritional value. However, the study did not compare this set of gene-targeted markers with an analysis of the same germplasm set using neutral markers. A comparative analysis using a set of neutral SSR markers would seem to be important to support this particular conclusion. The developed primer set can be potentially used for MAS in the *Rubus* breeding programs for improving the nutritional quality of fruits. This first requires confirmation that the SSR alleles identified correlate with differences in the content of flavonoids. Additional studies and further development of these gene-targeted markers are needed to validate this approach.

**Author Contributions:** Conceptualization, V.G.L. and K.A.S.; Data curation, V.G.L., K.V.K. and K.A.S.; Formal Analysis, V.G.L. and O.P.M.; Funding Acquisition, V.G.L., K.V.K. and K.A.S.; Investigation, V.G.L., N.M.S., O.P.M. and K.A.S.; Methodology, V.G.L. and K.A.S.; Project Administration, V.G.L. and K.A.S.; Resources, V.G.L. and K.A.S.; Supervision, V.G.L. and K.A.S.; Writing, V.G.L., K.V.K. and K.A.S.

**Funding:** The work was financially supported by the Ministry of Education and Science of the Russian Federation (grant No. 14.574.21.0149 from 26.09.2017, unique project identifier RFMEFI57417X0149).

**Acknowledgments:** We thank I. A. Pozdniakov (OOO Mikroklon, Pushchino, Russia) for providing us with raspberry plants used in this study.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Thompson, M.M. Chromosome numbers of *Rubus* species at the National Clonal Germplasm Repository. *HortScience* **1995**, *30*, 1447–1452. [CrossRef]
2. Skrovankova, S.; Sumczynski, D.; Mlcek, J.; Jurikova, T.; Sochor, J. Bioactive compounds and antioxidant activity in different types of berries. *Int. J. Mol. Sci.* **2015**, *16*, 24673–24706. [CrossRef]
3. Halvorsen, B.L.; Myhrstad, M.C.W.; Wold, A.B.; Jacobs, D.R.; Haffner, K.; Holte, K.; Andersen, L.F.; Baugerod, H.; Barikmo, I.; Hvattum, E.; et al. A systematic screening of total antioxidants in dietary plants. *J. Nutr.* **2002**, *132*, 461–471. [CrossRef]
4. Olas, B. Berry phenolic antioxidants—Implications for human health? *Front. Pharmacol.* **2018**, *9*, 78. [CrossRef]
5. FAOSTAT. 2019. Available online: <http://www.fao.org/faostat/> (accessed on 29 July 2019).

6. Graham, J.; Jennings, S.N. Raspberry breeding. In *Breeding Tree Crops*; Jain, S.M., Priyadarshan, M., Eds.; IBH & Science Publication: Oxford, UK, 2009; pp. 233–248.
7. Bushakra, J.M.; Lewers, K.S.; Staton, M.E.; Zhebentyayeva, T.; Sasaki, C.A. Developing expressed sequence tag libraries and the discovery of simple sequence repeat markers for two species of raspberry (*Rubus* L.). *BMC Plant Biol.* **2015**, *15*, 258. [[CrossRef](#)]
8. Castillo, N.R.F.; Reed, B.M.; Graham, J.; Fernandez-Fernandez, F.; Bassil, N.V. Microsatellite markers for raspberry and blackberry. *J. Am. Soc. Hortic. Sci.* **2010**, *135*, 271–278. [[CrossRef](#)]
9. Graham, J.; Smith, K.; MacKenzie, K.; Jorgenson, L.; Hackett, C.; Powell, W. The construction of a genetic linkage map of red raspberry (*Rubus idaeus* subsp. *idaeus*) based on AFLPs, genomic-SSR and EST-SSR markers. *Theor. Appl. Genet.* **2004**, *109*, 740–749. [[CrossRef](#)]
10. Sargent, D.J.; Fernández-Fernández, F.; Rys, A.; Knight, V.H.; Simpson, D.W.; Tobutt, K.R. Mapping of A1 conferring resistance to the aphid *Amphorophora idaei* and *dw* (dwarfing habit) in red raspberry (*Rubus idaeus* L.) using AFLP and microsatellite markers. *BMC Plant Biol.* **2007**, *7*, 15. [[CrossRef](#)]
11. Pattison, J.A.; Samuelian, S.K.; Weber, C.A. Inheritance of *Phytophthora* root rot resistance in red raspberry determined by generation means and molecular linkage analysis. *Theor. Appl. Genet.* **2007**, *115*, 225–236. [[CrossRef](#)]
12. Woodhead, M.; McCallum, S.; Smith, K.; Cardle, L.; Mazzitelli, L.; Graham, J. Identification, characterisation and mapping of simple sequence repeat (SSR) markers from raspberry root and bud ESTs. *Mol. Breed.* **2008**, *22*, 555–563. [[CrossRef](#)]
13. Bushakra, J.M.; Stephens, M.J.; Atmadjaja, A.N.; Lewers, K.S.; Symonds, V.V.; Udall, J.A.; Chagne, D.; Buck, E.J.; Gardiner, S.E. Construction of black (*Rubus occidentalis*) and red (*R. idaeus*) raspberry linkage maps and their comparison to the genomes of strawberry, apple, and peach. *Theor. Appl. Genet.* **2012**, *125*, 311–327. [[CrossRef](#)]
14. Castro, P.; Stafne, E.T.; Clark, J.R.; Lewers, K.S. Genetic map of the primocane-fruiting and thornless traits of tetraploid blackberry. *Theor. Appl. Genet.* **2013**, *126*, 2521–2532. [[CrossRef](#)]
15. Graham, J.; Smith, K.; Tierney, I.; MacKenzie, K.; Hackett, C.A. Mapping gene *H* controlling cane pubescence in raspberry and its association with resistance to cane botrytis and spur blight, rust and cane spot. *Theor. Appl. Genet.* **2006**, *112*, 818–831. [[CrossRef](#)]
16. Kassim, A.; Poette, J.; Paterson, A.; Zait, D.; McCallum, S.; Woodhead, M.; Smith, K.; Hackett, C.; Graham, J. Environmental and seasonal influences on red raspberry anthocyanin antioxidant contents and identification of quantitative traits loci (QTL). *Mol. Nutr. Food Res.* **2009**, *53*, 625–634. [[CrossRef](#)]
17. Graham, J.; Hackett, C.A.; Smith, K.; Woodhead, M.; Hein, I.; McCallum, S. Mapping QTL for developmental traits in raspberry from bud break to ripe fruit. *Theor. Appl. Genet.* **2009**, *118*, 1143–1155. [[CrossRef](#)]
18. McCallum, S.; Smith, K.; Woodhead, M.; Hackett, C.; Paterson, A.; Graham, J. Developing molecular markers for quality traits in red raspberry. *Theor. Appl. Genet.* **2010**, *121*, 611–627. [[CrossRef](#)]
19. Jennings, S.N.; Graham, J.; Ferguson, L.; Young, V. New developments in raspberry breeding in Scotland. *Acta Hortic.* **2016**, *1133*, 23–28. [[CrossRef](#)]
20. Bushakra, J.M.; Bryant, D.B.; Dossett, M.; Vining, K.J.; VanBuren, R.; Gilmore, B.S.; Lee, J.; Mockler, T.C.; Finn, C.E.; Bassil, N.V. A genetic linkage map of black raspberry (*Rubus occidentalis*) and the mapping of Ag<sub>4</sub> conferring resistance to the aphid *Amphorophora agathonica*. *Theor. Appl. Genet.* **2015**, *128*, 1631–1646. [[CrossRef](#)]
21. Powell, W.; Machray, G.C.; Provan, J. Polymorphism revealed by simple sequence repeats. *Trends Plant Sci.* **1996**, *1*, 215–222. [[CrossRef](#)]
22. Graham, J.; Smith, K.; Woodhead, M.; Russell, J. Development and use of simple sequence repeat SSR markers in *Rubus* species. *Mol. Ecol. Notes* **2002**, *2*, 250–252. [[CrossRef](#)]
23. Gezan, S.; Osorio, L.; Verma, S.; Whitaker, V. An experimental validation of genomic selection in octoploid strawberry. *Hortic. Res.* **2017**, *4*, 16070. [[CrossRef](#)]
24. Bobinaite, R.; Viskelis, P.; Venskutonis, P.R. Variation of total phenolics, anthocyanins, ellagic acid and radical scavenging capacity in various raspberry (*Rubus* spp.) cultivars. *Food Chem.* **2012**, *132*, 1495–1501. [[CrossRef](#)]
25. Lee, S.S.; Lee, E.M.; An, B.C.; Barampuzam, S.; Kim, J.-S.; Cho, J.Y.; Lee, I.-C.; Chung, B.Y. Molecular cloning and characterization of a flavanone-3-hydroxylase gene from *Rubus occidentalis* L. *J. Radiat. Ind.* **2008**, *2*, 121–128.

26. Chen, Q.; Yu, H.W.; Wang, X.R.; Xie, X.L.; Yue, X.Y.; Tang, H.R. An alternative cetyltrimethylammonium bromide-based protocol for RNA isolation from blackberry (*Rubus* L.). *Genet. Mol. Res.* **2012**, *11*, 1773–1782. [[CrossRef](#)]
27. Li, S. Transcriptional control of flavonoid biosynthesis: Fine-tuning of the MYB-bHLH-WD40 (MBW) complex. *Plant Signal. Behav.* **2014**, *9*, e27522. [[CrossRef](#)]
28. Dossett, M.; Bassil, N.V.; Lewers, K.S.; Finn, C.E. Genetic diversity in wild and cultivated black raspberry (*Rubus occidentalis* L.) evaluated by simple sequence repeat markers. *Genet. Resour. Crop Evol.* **2012**, *59*, 1849–1865. [[CrossRef](#)]
29. Lee, G.-A.; Song, J.Y.; Choi, H.-R.; Chung, J.-W.; Jeon, Y.-A.; Lee, J.-R.; Ma, K.-H.; Lee, M.-C. Novel microsatellite markers acquired from *Rubus coreanus* Miq. and cross-amplification in other *Rubus* species. *Molecules* **2015**, *20*, 6432–6442. [[CrossRef](#)]
30. Martins, W.S.; Lucas, D.C.S.; Neves, K.F.S.; Bertioli, D.J. WebSat—A web software for microsatellite marker development. *Bioinformatics* **2009**, *3*, 282–283. [[CrossRef](#)]
31. Nunes, C.F.; Ferreira, J.L.; Nunes-Fernandes, M.C.; de Souza Breves, S.; Generoso, A.L.; Fontes-Soares, B.D.; Carvalho-Dias, M.S.; Pasqual, M.; Borem, A.; de Almeida Cancado, G.M. An improved method for genomic DNA extraction from strawberry leaves. *Ciência Rural* **2011**, *41*, 1383–1389. [[CrossRef](#)]
32. Liu, K.; Muse, S.V. PowerMarker: An integrated analysis environment for genetic marker analysis. *Bioinformatics* **2005**, *21*, 2128–2129. [[CrossRef](#)]
33. Ahmad, A.; Wang, J.-D.; Pan, Y.-B.; Rahat Sharif, R.; Gao, S.-J. Development and use of simple sequence repeats (SSRs) markers for sugarcane breeding and genetic studies. *Agronomy* **2018**, *8*, 260. [[CrossRef](#)]
34. Fernandez-Fernandez, F.; Antanaviciute, L.; Govan, C.L.; Sargent, D.J. Development of a multiplexed microsatellite set for fingerprinting red raspberry (*Rubus idaeus*) germplasm and its transferability to other *Rubus* species. *J. Berry Res.* **2011**, *1*, 177–187. [[CrossRef](#)]
35. Han, Y.; Huang, K.; Liu, Y.; Jiao, T.; Ma, G.; Qian, Y.; Wang, P.; Dai, X.; Gao, L.; Xia, T. Functional analysis of two flavanone-3-hydroxylase genes from *Camellia sinensis*: A critical role in flavonoid accumulation. *Genes* **2017**, *8*, 300. [[CrossRef](#)]
36. Tian, J.; Han, Z.; Zhang, J.; Hu, Y.; Song, T.; Yao, Y. The balance of expression of dihydroflavonol 4-reductase and flavonol synthase regulates flavonoid biosynthesis and red foliage coloration in crabapples. *Sci. Rep.* **2015**, *5*, 12228. [[CrossRef](#)]
37. Saito, K.; Yonekura-Sakakibara, K.; Nakabayashi, R.; Higashi, Y.; Yamazaki, M.; Tohge, T.; Fernie, A.R. The flavonoid biosynthetic pathway in *Arabidopsis*: Structural and genetic diversity. *Plant Physiol. Biochem.* **2013**, *72*, 21–34. [[CrossRef](#)]
38. Petrusa, E.; Braidot, E.; Zancani, M.; Peresson, C.; Bertolini, A.; Patui, S.; Vianello, A. Plant Flavonoids—Biosynthesis, transport and involvement in stress responses. *Int. J. Mol. Sci.* **2013**, *14*, 14950–14973. [[CrossRef](#)]
39. Zheng, D.; Schröder, G.; Schröder, J.; Hrazdina, G. Molecular and biochemical characterization of three aromatic polyketide synthase genes from *Rubus idaeus*. *Plant Mol. Biol.* **2001**, *46*, 1–15. [[CrossRef](#)]
40. García-Gómez, B.; Razi, M.; Salazar, J.A.; Prudencio, A.S.; Ruiz, D.; Dondini, L.; Martínez-Gómez, P. Comparative analysis of SSR markers developed in exon, intron, and intergenic regions and distributed in regions controlling fruit quality traits in *Prunus* species: Genetic diversity and association studies. *Plant Mol. Biol. Rep.* **2018**, *36*, 23–35. [[CrossRef](#)]
41. Holland, J.B.; Helland, S.J.; Sharopova, N.; Rhyne, D.C. Polymorphism of PCR-based markers targeting exons, introns, promoter regions, and SSRs in maize and introns and repeat sequences in oat. *Genome* **2001**, *44*, 1065–1076. [[CrossRef](#)]
42. Krutovsky, K.V.; Neale, D.B. Nucleotide diversity and linkage disequilibrium in cold-hardiness and wood quality-related candidate genes in Douglas-fir. *Genetics* **2005**, *171*, 2029–2041. [[CrossRef](#)]
43. Cai, C.; Wu, S.; Niu, E.; Cheng, C.; Guo, W. Identification of genes related to salt stress tolerance using intron-length polymorphic markers, association mapping and virus-induced gene silencing in cotton. *Sci. Rep.* **2017**, *7*, 528. [[CrossRef](#)]
44. Lewers, K.S.; Weber, C.A. The trouble with genetic mapping of raspberry. *HortScience* **2005**, *40*, 1108. [[CrossRef](#)]
45. Anderson, J.R.; Lübberstedt, T. Functional markers in plants. *Trends Plant Sci.* **2003**, *8*, 554–560. [[CrossRef](#)]

46. Rafique, M.Z.; Carvalho, E.; Stracke, R.; Palmieri, L.; Herrera, L.; Feller, A.; Malnoy, M.; Martens, S. Nonsense mutation inside anthocyanidin synthase gene controls pigmentation in yellow raspberry (*Rubus idaeus* L.). *Front. Plant Sci.* **2016**, *7*, 1892. [[CrossRef](#)]
47. Michael, T.P.; VanBuren, R. Progress, challenges and the future of crop genomes. *Curr. Opin. Plant Biol.* **2015**, *24*, 71–81. [[CrossRef](#)]
48. VanBuren, R.; Bryant, D.; Bushakra, J.M.; Vining, K.J.; Edger, P.P.; Rowley, E.R.; Priest, H.D.; Michael, T.P.; Lyons, E.; Filichkin, S.A.; et al. The genome of black raspberry (*Rubus occidentalis*). *Plant J.* **2016**, *87*, 535–547. [[CrossRef](#)]
49. Wight, H.; Zhou, J.; Li, M.; Hannenhalli, S.; Mount, S.M.; Liu, Z. Draft genome assembly and annotation of red raspberry *Rubus idaeus*. *BioRxiv* **2019**, 546135, 1–22. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



Communication

# SNP- and Haplotype-Based GWAS of Flowering-Related Traits in Maize with Network-Assisted Gene Prioritization

Carlos Maldonado <sup>1</sup>, Freddy Mora <sup>1,\*</sup>, Filipe Augusto Bengosi Bertagna <sup>2</sup>, Maurício Carlos Kuki <sup>2</sup> and Carlos Alberto Scapim <sup>3</sup>

<sup>1</sup> Institute of Biological Sciences, University of Talca, 2 Norte 685, Talca 3460000, Chile; camaldonado@utalca.cl

<sup>2</sup> Genetic and Plant Breeding Post-Graduate Program, Universidade Estadual de Maringá, Maringá PR 87020-900, Brazil; filipeabbertagna@gmail.com (F.A.B.B.); mcarloskuki@gmail.com (M.C.K.)

<sup>3</sup> Departamento de Agronomia, Universidade Estadual de Maringá, Maringá PR 87020-900, Brazil; cascapim@uem.br

\* Correspondence: fmora@utalca.cl; Tel.: +56-71-2200280

Received: 18 September 2019; Accepted: 4 November 2019; Published: 7 November 2019

**Abstract:** Maize (*Zea mays* L.) is one of the most crucial crops for global food security worldwide. For this reason, many efforts have been undertaken to address the efficient utilization of germplasm collections. In this study, 322 inbred lines were used to link genotypic variations (53,403 haplotype blocks (HBs) and 290,973 single nucleotide polymorphisms (SNPs)) to corresponding differences in flowering-related traits in two locations in Southern Brazil. Additionally, network-assisted gene prioritization (NAGP) was applied in order to better understand the genetic basis of flowering-related traits in tropical maize. According to the results, the linkage disequilibrium (LD) decayed rapidly within 3 kb, with a cut-off value of  $r^2 = 0.11$ . Total values of 45 and 44 marker-trait associations (SNPs and HBs, respectively) were identified. Another important finding was the identification of HBs, explaining more than 10% of the total variation. NAGP identified 44, 22, and 34 genes that are related to female/male flowering time and anthesis-silking interval, respectively. The co-functional network approach identified four genes directly related to female flowering time ( $p < 0.0001$ ): *GRMZM2G013398*, *GRMZM2G021614*, *GRMZM2G152689*, and *GRMZM2G117057*. NAGP provided new insights into the genetic architecture and mechanisms underlying flowering-related traits in tropical maize.

**Keywords:** gene prioritization; linkage disequilibrium; marker-trait association; tropical maize

## 1. Introduction

Maize (*Zea mays* L.) plays an important role in the human diet and accounts for a large proportion of the global cereal demand. Together with rice and wheat, these three cereals account for more than 40% and 35% of the world's calorie and protein supply, respectively [1,2]. Maize is among the few crops grown on almost every continent and has diverse uses, including food, animal feed, and ethanol production [3]. The United States, China, and Brazil are the top three largest maize-producing countries in the world, representing more than 70% of total maize production [4].

Since maize is one of the most important crops for global food security, several efforts have been undertaken addressing the efficient utilization of germplasm materials. In fact, the development of maize germplasm collections has been beneficial to capture and maintain the high levels of genetic diversity that exist locally and globally [5–9]. These efforts have allowed the methodical exploration of the genetic architecture of complex traits in maize, which benefit from the high diversity [8]. Liu [10], for instance, performed a genome-wide association study (GWAS; a standard forward genetic

technique) using a population comprised of a global core collection of maize inbred lines, and found several candidate genes associated with starch synthesis, of which one gene (*Glucose-1-phosphate adenyltransferase*) is known as an important regulator of kernel starch content. Li et al. [11] identified several genetic variants associated with maize flowering time using an extremely large multigenetic background population (>8000 maize lines). The associated single nucleotide polymorphisms (SNPs) detected in this large panel exhibited high accuracy for predicting flowering time.

In an effort to overcome certain limitations present in forward and reverse genetic techniques, for example lacking in functional clues of trait-associated candidate genes derived from forward genetics studies and in silico strategies for candidate gene selection in targeted mutagenesis in reverse genetics approaches, Lee et al. [12] recently presented a network-assisted gene prioritization system (MaizeNet), which facilitates genetic analysis through supporting candidate genes based on network neighbors with known traits or functions, and aids in identifying potential candidate genes that are highly likely to be causal to the phenotype of interest. This network-based resource provides new insights into the genetic architecture and mechanisms underlying complex traits in maize and promises to accelerate the discovery of trait-associated genes for crop improvement. In this study, an integrated approach using GWAS (based on 53,403 haplotype blocks (HBs) and 290,973 SNPs) and network-assisted gene prioritization was applied in order to better understand the genetic basis of flowering-related traits in tropical maize. To this end, marker-trait association analyses were performed using a multigenetic background population comprising 322 inbred lines of field corn, popcorn, and sweet corn.

## 2. Materials and Methods

### 2.1. Trial Conditions and Phenotyping

A total of 322 inbred lines of tropical maize were used in this genome-wide association study, which were derived from three genetic backgrounds collected in Brazil: Field corn (178), popcorn (128), and sweet corn (16). This maize panel was evaluated during the growing season of 2017–2018 in two locations (Cambira and Sabaudia) situated in Southern Brazil, Parana State. The experimental design was an alpha-lattice with 24 incomplete blocks and 3 replications per line. Female and male flowering time (FF and MF, respectively) were measured in each line as the number of days from sowing to anther extrusion from the tassel glumes (MF) or to visible silks (FF). Additionally, the anthesis-silking interval (ASI) was calculated as the difference between MF and FF.

### 2.2. Population Structure, Linkage Disequilibrium (LD), and Haplotype Blocks

Genomic DNA was isolated from young leaves of five plants from each inbred line of tropical maize (319 in Cambira and 293 in Sabaudia), approximately 30 days after germination. The DNA extraction was carried out by Cetyl trimethyl ammonium bromide (CTAB) according to the protocol established by Chen and Ronald [13]. The quality of DNA was evaluated and quantified using 1% agarose gel and Nanodrop, respectively. The DNA samples were sent to the University of Wisconsin-Madison—Biotechnology Center for SNP discovery via genotyping by sequencing (GBS), which is described in Elshire et al. [14] and Glaubitz et al. [15]. The raw database was filtered considering a minor allele frequency (MAF) > 0.05, resulting in a genotype file of 291,633 high-quality SNPs. The LD kNNi imputation (linkage disequilibrium k-nearest neighbor imputation) was performed to impute missing data in the dataset [16]. Finally, SNPs with a MAF < 0.01 and a proportion of missing data per location >90% were eliminated from the imputed dataset [17]. Subsequently, 290,973 SNPs were retained after filtering for MAF and missing data.

The population structure was inferred using the model-based Bayesian clustering approach implemented in the program InStruct [18]. For each K value (where K is the number of genetically differentiated groups, K = 1–6), 10 runs were performed separately, each with 100,000 Monte Carlo Markov Chain replicates and a burn-in period of 10,000 iterations. The optimal K value was determined with the highest  $\Delta K$  method [19] and the lowest deviance information criterion (DIC).

The extent of LD was estimated using the correlation coefficients of the allelic frequencies ( $r^2$ ) considering all the possible combinations of the alleles. The critical  $r^2$  value was calculated according to the method used by Brescaglio and Sorells [20].

The HBs were constructed for each chromosome according to the confidence interval algorithm developed by Gabriel [21], implemented in the software Haploview v.4.2 [22]. This method considers the 95% confidence intervals of the disequilibrium coefficient ( $D'$ ) values and builds a haplotype block if the LD is classified as a “strong LD” type ( $D'$  higher than 0.98 and lower interval limit of  $>0.7$ ). Finally, HBs were later transformed into multiallelic markers, considering the allelic combinations within each block to be independent alleles [5,23].

### 2.3. SNP- and Haplotype-Based GWAS

The HB- and SNP-based association analyses were performed using a mixed linear model (MLM) in TASSEL 3.0 and TASSEL 5.2, respectively [24], which considers the effects of the population structure (Q) and genetic relationships or matrix kinship (K) among inbred lines. The kinship matrix was calculated based on identity by state (IBS) [25] in TASSEL. The Adjusted-Entry Means of the general linear model (experimental design) were used as the adjusted phenotypes according to Contreras-Soto et al. [26] and Arriagada et al. [27]. Correlations between each pair of traits were calculated using a Bayesian bi-trait model [28–30]. The statistical analysis was performed using the R package MCMCglmm (version 3.6.1; <https://www.r-project.org>) [31].

### 2.4. Prioritization of GWAS Candidate Genes and Inference of Co-Functional Networks for Flowering Traits in Maize

The candidate genes were chosen from the genes around the significant loci (SNP or haplotype blocks) identified by GWAS. To this end, a window (or threshold) of twice the distance indicated by the LD analysis was established, placing the marker in the center of the window. The gene prioritization was performed using MaizeNet [12] based on the connections of the candidate genes to the genes in one estimated network with previously associated genes with flowering-time in *Zea mays*. New candidate genes were then ranked by closeness to the “guide genes” (derived from estimated network in MaizeNet) measured for each candidate gene (derived from GWAS) as the sum of network edge scores from that gene to the guide genes [12]. The estimated co-functional network was carried out through the association of genes (candidate genes and genes identified in prioritization of MaizeNet) with subnetworks enriched by gene ontology annotations related to the biological processes (GOBP) of flowering in MaizeNet. Finally, the given genes are related to the flowering-time if the subnetworks of MaizeNet significantly associated with these genes, and if are also enriched for on the relevant GOBP term for flowering.

## 3. Results and Discussion

### 3.1. Genetic Structure

The Bayesian clustering analysis (InStruct) of the population structure indicated that the 322 inbred lines from the Brazilian germplasm represent two main genetic clusters ( $k = 2$ ; Figure S1A), inferred from both the lowest DIC value and the second-order change rate of the probability function with respect to Q ( $\Delta Q$ ) [19]. Cluster I contained 221 lines (68.6%), over 80% (177/221) of which were genotypes of field corn, while all sweet corn lines (16) were within this cluster. On the other hand, cluster II consisted of 101 lines, over 99% (100/101) of which were genotypes of popcorn (Figure S1). Similar results were obtained by the PCA method for this association mapping panel, as shown in Figure S1B. The first component explained 12.1% of the total variation and most of the inbred lines were separated in the same genetically differentiated groups (Figure S1B). These results are in accordance with the previous findings of Maldonado et al. [5] and Coan et al. [6], in which tropical maize inbred lines were grouped into two genetically differentiated clusters, separating field corn and popcorn lines.

### 3.2. Linkage Disequilibrium

The genome-wide LD decay pattern is shown in Table 1 and Figure S2. The LD statistic  $r^2$  showed a clear nonlinear trend with physical distance. According to these results, the LD decayed rapidly within 3 kb, with a cut-off value of  $r^2 = 0.11$ . The average LD on all chromosomes (Chr) was  $r^2 = 0.09$ . On the other hand, 0.63% of the total pairs of linked SNPs were in a complete LD ( $r^2 = 1$ ), and 4.4% had an  $r^2$  value  $>0.5$  (strong LD). The LD of Chr 3 and 7 decayed faster than the other chromosomes, with  $\sim 2.2$  kb for a cut-off value of  $r^2 = 0.11$ . Past studies have found that this LD pattern (i.e., rapid decay with increasing physical distance) is typical in tropical maize germplasms [6,9,32]. The LD decay pattern in this study was similar to the findings of Yan et al. [33] and Coan et al. [6], who reported that the LD pattern (in tropical maize germplasms) decreases rapidly in the range of 0.1–10 kb.

**Table 1.** Summary of information on linkage disequilibrium (LD) and haplotype blocks (HBs) determined in inbred lines of tropical maize. Chr corresponds to the chromosome number;  $N^{\circ}_{\text{SNP}}$  indicates the number of single nucleotide polymorphisms (SNPs) detected on each chromosome;  $N^{\circ}_{\text{HB}}$  is the number of haplotype blocks;  $\text{Size}_{\text{HB}}$  and  $\text{Max}(\text{kb})$  correspond to the maximum number of SNPs forming a haplotype block and the maximum size (in kb) for a haplotype block, respectively.

Chr	LD	$N^{\circ}_{\text{SNP}}$	Position (pb) First-Last SNP	Cambira			Sabaudia		
				$N^{\circ}_{\text{HB}}$	$\text{Size}_{\text{HB}}$	$\text{Max}(\text{kb})$	$N^{\circ}_{\text{HB}}$	$\text{Size}_{\text{HB}}$	$\text{Max}(\text{kb})$
1	2.87	39,148	38,222–275,861,066	7180	32	498	7087	32	497
2	2.68	37,341	40,724–244,417,305	6874	35	500	6730	36	500
3	2.21	34,889	191,169–235,520,333	6370	35	487	6256	33	491
4	6.55	26,908	217,040–246,840,261	4790	42	500	4693	42	499
5	2.61	35,691	12,711–223,658,670	6584	33	493	6472	32	500
6	3.75	23,441	169,964–173,881,702	4368	53	466	4317	53	466
7	2.23	24,958	180,204–182,128,999	4683	28	500	4546	28	500
8	4.1	25,537	204,228–181,043,617	4703	50	498	4577	50	498
9	2.94	22,404	61,292–159,668,042	4091	35	500	3991	36	500
10	2.85	20,656	128,669–150,847,940	3760	52	498	3698	51	498
Mean	2.94	29,097	-	5340	40	494	5238	39	495

### 3.3. Haplotype Blocks

Total values of 53,403 and 52,377 HBs were identified in all chromosomes for Cambira and Sabaudia, respectively (Table 1), over 47%, 20%, and 33% of which contained two, three, and four (or more) SNPs, respectively (Figure S3). These HBs were constructed considering 319 and 293 inbred lines in Cambira and Sabaudia (respectively), and 290,973 SNPs distributed on all chromosomes (Table 1). An average of  $\sim 20,000$  SNPs per chromosome satisfied the criteria of the 95% confidence interval proposed by Gabriel et al. [21]. Particularly, the largest number of HBs in both locations was determined by combinations of SNPs located on Chr 1, while the smaller amount was constructed by SNPs located on Chr 10 (Table 1). In this study, several genomic regions were detected in strong disequilibrium, up to  $\sim 0.1$  Mb. Therefore, as these regions have a strong LD, it is possible to suggest that they will be inherited together across generations. Moreover, about 2.3% of the HBs formed in both locations had an extension over 0.1 Mb, with a  $D'$  value between 0.7 and 0.98 [21]. Analysis of the LD pattern enabled the identification and characterization of several HBs (or strongly linked genomic regions), because there is a strong LD among the SNPs that compose it. This indicates that recombination events within these HBs are unlikely, thus, these HBs should inherit together across generations.

### 3.4. Genome-Wide Association Study and Network-Assisted Gene Prioritization

Total values of 45 and 44 associations (SNPs and HBs, respectively) were identified for the studied traits, which are distributed in all chromosomes of maize (Table 2 and Table S1). Four SNPs were jointly associated with the FF and MF traits. In Cambira, four haplotype blocks—two loci on Chr 8 (bin

8.03) and two on Chr 9 (bin 9.06)—were jointly associated with FF and MF. In turn, Chr 3 presented two genomic regions associated with FF in Cambira (one SNP and one HB) and three in Sabaudia (one SNP and two HBs), while various SNPs (five) and HBs (three) were associated with some of the three traits on Chr 9 (bin 9.06). Interestingly, all associations were environment-specific, confirming the existence of a significant and complex genetic-by-environment interaction. The results from Bayesian bi-trait analyses showed a high correlation between FF and MF, which was significantly different from zero in both locations ( $r = 0.94$  and  $0.92$ ), justifying the fact that FF and MF share significant loci. In accordance with our findings, Xu et al. [34] found a very high amount of quantitative trait loci (QTL) significant on bins 1.03, 8.05, and 9.06 for photoperiod sensitivity and flowering time (traits highly correlated in maize; [35]), while Chardon et al. [36], through a meta-analysis, detected hot-spot QTL regions for flowering time on bins 8.03 and 8.05. On the other hand, 64 QTLs related to maize flowering time were identified by Liu et al. [37], which were distributed on chromosomal bins 1.01, 1.03, 1.1, 2.02, 3.02, 3.04, 4.05, 6.06, 7.02, 7.03, 7.04, 8.03, 8.05, 9.01, and 9.07. Like these previous studies, this study also identified significant marker-trait associations on bins 1.01, 1.03, 1.1, 2.02, 3.02, 3.04, 4.05, 6.06, 7.02, 7.03, 7.04, 8.03, 8.05, 9.01, and 9.07. This result suggests that these regions should contain important genes controlling the flowering time in maize. In addition, chromosomes 8 and 9 had the main associations for all three traits, which is consistent with studies that considered other environments and genetic materials [34,36–38].

**Table 2.** Summary of the associations detected by a genome-wide association study (GWAS), based on haplotype blocks and SNP for the traits of female/male flowering time (FF and MF, respectively) and anthesis–silking interval (ASI) measured in inbred lines of tropical maize.

Marker	Trait	Cambira			Sabaudia		
		N <sub>M</sub>	Chr(N <sub>M</sub> )	PV%	N <sub>M</sub>	Chr(N <sub>M</sub> )	PV%
SNP	FF	10	2(2), 3(1), 6(1), 7(4), 8(1), and 9(1)	5.6–6.3	7	2(1), 3(1), 5(3), and 6(2)	6.5–10.1
	MF	5	2(2), 8(1), and 9(2)	5.7–6.4	6	3(3), 5(1), and 6(2)	6.5–8.5
	ASI	8	1(2), 3(1), 5(1), 6(1), 7(1), and 8(2)	5.6–6.0	9	1(2), 3(2), 8(3), 9(1), and 10(1)	6.5–9.9
Haplotype Blocks	FF	11	1(1), 3(1), 5(1), 6(2), 7(1), 8(3), and 9(2)	5.6–17.0	3	3(2) and 9(1)	6.3–8.6
	MF	7	2(1), 4(1), 7(1), 8(2), and 9(2)	4.6–13.0	12	1(2), 4(4), 5(1), 7(3), 8(1), and 9(1)	5.0–6.9
	ASI	4	1(1), 3(1), 8(1), and 10(1)	5.6–7.8	7	2(2), 5(1), 7(2), 8(1), and 9(1)	5.1–9.5

PV%: Percentage of the phenotype variation explained by the marker; N<sub>M</sub>: Number of significant marker-trait associations.

In Cambira, the proportion of the phenotypic variance (PV%) explained by SNP markers was ~6%, while haplotype blocks explained 6–17%, 5–13%, and 6–8% of the phenotypic variation of FF, MF, and ASI, respectively (Table 2 and Table S1). On the other hand, in Sabaudia, the PV% explained by SNPs was similar to that detected by HBs. In Sabaudia, the PV% values were moderate (either SNPs or HBs), which varied between 5 and 10%, while in Cambira, HBs showed higher PV% values (>10%) in comparison with SNPs. Moreover, the HB HChr9B2943 (in Cambira) was jointly associated with FF and MF, accounting for 17% and 13% of the total variation of FF and MF, respectively (Table 2 and Table S1). Several studies reported PV% values of flowering time smaller than 10% [34,37,39,40]. In fact, numerous QTLs with small effects would be contributing to genetic variation in flowering time across diverse maize germplasms [34,37,41]. In accordance with this, 93% (41/44) of the significant HB and all SNP associations did not explain more than 10% of the total variation. Importantly, three HBs had PV% values higher than 10%, indicating the potential effectiveness of haplotypes over individual

SNP analysis, an aspect emphasized by Maldonado et al. [5] and Contreras-Soto et al. [26]. Twenty-five of the 45 SNPs detected by GWAS (i.e., 56%) were found to be part of a haplotype block, which in turn were significantly associated with a given trait. Moreover, 14 HBs contained at least 1 significant SNP, and 9 HBs contained 2 or 3 SNPs significantly associated with some trait. On the other hand, 68% (30/44) of the HBs detected did not contain any associated SNPs, which suggests that haplotype blocks are useful for discovering genomic regions that are not detected by SNP markers. On the other hand, the use of haplotype blocks in GWAS reduces the number of multiple tests, compared with SNP-based association analysis [5]. Moreover, the use of haplotype blocks as multiallelic markers might improve marker-trait association analyses, compensating the biallelic limitation of SNP markers [5,26].

Based on the physical position of the maize reference genome (<http://www.maizegdb.org/>), 51 candidate genes were identified neighboring the significant SNPs and HBs (Table S1), of which 11 were present in more than one trait (FF and MF) (Table S1). The network-assisted gene prioritization performed by MaizeNet [12] identified 100 additional genes based on biological processes of flowering and reproduction. Forty-four, 22, and 34 genes that were identified by MaizeNet are related to FF, MF, and ASI, respectively (Table S1). Co-functional networks determined by MaizeNet [12] are shown in Figure 1 and Figure S4. The co-functional networks identified the following genes directly related to FF—*GRMZM2G013398*, *GRMZM2G021614*, *GRMZM2G152689*, and *GRMZM2G117057*—with statistical significance of  $p < 0.0001$  (Figure 1). On the other hand, the co-functional networks of MF presented significances of  $2.2 \times 10^{-11}$  and  $2.3 \times 10^{-5}$  using HBs and SNPs, respectively. The gene *GRMZM2G013398* has an ortholog in *Arabidopsis thaliana* that encodes CONSTANS-LIKE 9 (COL9), which has light-controlled functions and is crucial to inducing the day-length specific expression of the *FLOWERING LOCUS T* (*FT*) gene in leaves [42]. FT protein is the main component of florigen that strongly influences the timing of flowering [43]. Notably, the CONSTANS protein strongly influences the performance of maize flowering time in response to photoperiod, directly inducing the transcription of *FT* genes in *Arabidopsis* [42,43]. On the other hand, the genes *GRMZM2G021614*, *GRMZM2G152689*, and *GRMZM2G117057* encode phosphatidylethanolamine-binding proteins (pebp9, pebp10, and pebp11, respectively), which play important roles in floral transition in angiosperms [44]. Moreover, Kikuchi et al. [45] and Wickland and Hanzawa [46] showed that the presence and structure of these genes, together with their roles in the regulation of flowering, are well conserved among cereal plants.

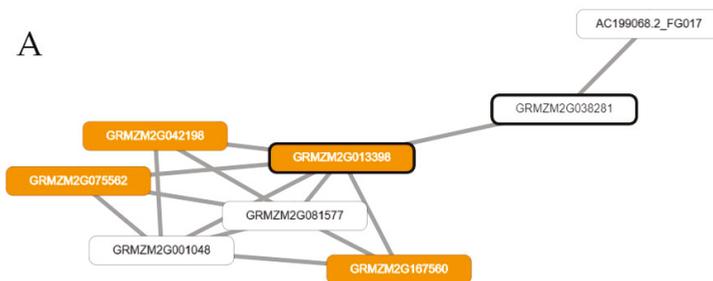
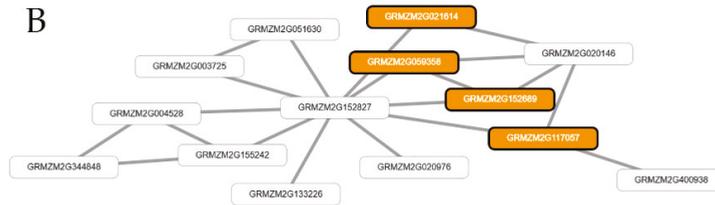


Figure 1. Cont.



**Figure 1.** Co-functional networks estimated using genes identified by SNP- and haplotype-based GWAS (A,B, respectively), genes identified by network-assisted gene prioritization (in MaizeNet) for flowering time and subnets enriched by gene ontology annotations related to the biological processes of female flowering (FF) in MaizeNet. (A) Gene *GRMZM2G013398* identified by the prioritization analysis (nodes orange highlighted with bold borderline) connected with all the subnetwork genes of MaizeNet (white nodes) and genes associated with the ontology annotations related to flowering time (orange nodes). (B) Genes *GRMZM2G117057*, *GRMZM2G021614*, *GRMZM2G059358*, and *GRMZM2G152689* identified by prioritization analysis (nodes orange highlighted with bold borderline) connected with all the subnetwork genes of MaizeNet (white nodes).

#### 4. Conclusions

In the present study, we identified several loci (SNPs and haplotype blocks) with variable contributions to phenotypic expression, which were located in regions that play important roles in the control of flowering time in maize. The GWAS based on haplotype blocks was beneficial to identify loci with major effects in comparison to SNP-based GWAS. The co-functional network approach identified four genes that strongly influence the timing of flowering in tropical maize. In general, network-assisted gene prioritization provides new insights into the genetic architecture and mechanisms underlying flowering-related traits in tropical maize.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4395/9/11/725/s1>: Figure S1. Inferred population structure in a collection of maize germplasm (322 inbred lines). (A) Genetic structure inferred by a Bayesian clustering model using InStruct and a dendrogram carried out using the neighbor-joining (Nei's genetic distances). The light gray and dark gray indicate the proportion of the genome extracted from the two main genetic clusters estimated by InStruct. (B) Principal components analysis (PCA) with two major groups identified, which correspond closely to InStruct results. Values in parentheses indicate the percentage of variation explained by each main component; Figure S2. Linkage disequilibrium (LD) decay pattern in all chromosomes of maize. Chromosomes 3 and 7 decayed faster than the other chromosomes, while chromosome 4 presented the slowest decay (lower and upper margins, respectively); Figure S3. Frequency distribution of the size of haplotype blocks consisting of two or more SNPs, in the locations Cambira and Sabaudia; Figure S4. Co-functional networks estimated using genes identified by SNP- and Haplotype-based GWAS (A and B, respectively), genes identified in prioritization of MaizeNet for flowering time and subnets enriched by gene ontology annotations related to the biological processes of male flowering (MF) in MaizeNet. White nodes represent all the subnetwork genes of MaizeNet, orange nodes are genes associated with the ontology annotations related to flowering time, and nodes highlighted with bold borderline correspond to genes identified by GWAS or the prioritization analysis; Table S1. Details of the associations and candidate genes detected in SNP- and Haplotype-based GWAS for the traits of Female/Male Flowering time (FF and MF, respectively) and Anthesis–Silking Interval (ASI) measured in inbred lines of tropical maize in two locations (Cambira and Sabaudia).

**Author Contributions:** Conceptualization, F.M., C.M., M.C.K., and F.A.B.B.; methodology, F.M., C.A.S., and C.M.; software, C.M., M.C.K., and F.A.B.B.; validation, F.M., C.M., M.C.K., F.A.B.B., and C.A.S.; formal analysis, F.M., C.A.S., and C.M.; investigation, M.C.K., F.A.B.B., and C.A.S.; resources, F.M. and C.M.; writing—original draft preparation, F.M. and C.M.; writing—review and editing, F.M. and C.M.; visualization, M.C.K., F.A.B.B., and C.A.S.; supervision, F.M.; project administration, C.A.S.; funding acquisition, C.A.S.

**Funding:** This research was funded by the National Council of Technological and Scientific Development (CNPq) and the Coordination for the Improvement of Higher Education Personnel (CAPES).

**Acknowledgments:** Freddy Mora thanks FONDECYT (grant number 1170695). Carlos Maldonado thanks CONICYT-PCHA/Doctorado Nacional/2017-21171466.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study, in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

## References

1. FAO. *Save and Grow in Practice: Maize, Rice, Wheat—A Guide to Sustainable Production*; Food and Agriculture Organization of the United Nations: Rome, Italy, 2016.
2. Begcy, K.; Sandhu, J.; Walia, H. Transient heat stress during early seed development primes germination and seedling establishment in rice. *Front. Plant. Sci.* **2019**, *9*, 1768. [[CrossRef](#)] [[PubMed](#)]
3. Rashid, S.; Rosentrater, K.A.; Bern, C. Effects of deterioration parameters on storage of maize: A review. *J. Nat. Sci. Res.* **2013**, *3*, 147.
4. Ranum, P.; Pena-Rosas, J.P.; Garcia-Casal, M.N. Global maize production, utilization, and consumption. *Ann. N. Y. Acad. Sci.* **2014**, *1312*, 105–112. [[CrossRef](#)] [[PubMed](#)]
5. Maldonado, C.; Mora, F.; Scapim, C.A.; Coan, M. Genome-wide haplotype-based association analysis of key traits of plant lodging and architecture of maize identifies major determinants for leaf angle: *hapLA4*. *PLoS ONE* **2019**, *14*, e0212925. [[CrossRef](#)]
6. Coan, M.; Senhorinho, H.J.; Pinto, R.J.; Scapim, C.A.; Tessmann, D.J.; Williams, W.P.; Warburton, M.L. Genome-Wide Association Study of Resistance to Ear Rot by *Fusarium verticillioides* in a Tropical Field Maize and Popcorn Core Collection. *Crop. Sci.* **2018**, *58*, 564–578. [[CrossRef](#)]
7. Xiao, Y.; Liu, H.; Wu, L.; Warburton, M.; Yan, J. Genome-Wide Association Studies in Maize: Praise and Stargaze. *Mol. Plant.* **2017**, *10*, 359–374. [[CrossRef](#)]
8. Zhu, X.M.; Shao, X.Y.; Pei, Y.H.; Guo, X.M.; Li, J.; Song, X.Y.; Zhao, M.A. Genetic Diversity and Genome-Wide Association Study of Major Ear Quantitative Traits Using High-Density SNPs in Maize. *Front. Plant. Sci.* **2018**, *9*, 966. [[CrossRef](#)]
9. Romay, M.C.; Millard, M.J.; Glaubitz, J.C.; Peiffer, J.A.; Swarts, K.L.; Casstevens, T.M.; Elshire, R.J.; Acharya, C.B.; Mitchell, S.E.; Flint-Garcia, S.A.; et al. Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol.* **2013**, *14*, R55. [[CrossRef](#)]
10. Liu, N.; Xue, Y.; Guo, Z.; Li, W.; Tang, J. Genome-Wide Association Study Identifies Candidate Genes for Starch Content Regulation in Maize Kernels. *Front. Plant. Sci.* **2016**, *7*, 1046. [[CrossRef](#)]
11. Li, Y.; Li, C.; Bradbury, P.; Liu, X.; Lu, F.; Romay, C.; Glaubitz, J.; Wu, X.; Peng, B.; Shi, Y.; et al. Identification of genetic variants associated with maize flowering time using an extremely large multi-genetic background population. *Plant. J.* **2016**, *86*, 391–402. [[CrossRef](#)]
12. Lee, T.; Lee, S.; Yang, S.; Lee, I. MaizeNet: A co-functional network for network-assisted systems genetics in *Zea mays*. *Plant. J.* **2019**, *99*, 571–582. [[CrossRef](#)] [[PubMed](#)]
13. Chen, D.H.; Ronald, P.C. A rapid DNA miniprep method suitable for AFLP and other PCR applications. *Plant. Mol. Biol. Rep.* **1999**, *17*, 53–57. [[CrossRef](#)]
14. Elshire, R.J.; Glaubitz, J.C.; Sun, Q.; Poland, J.A.; Kawamoto, K.; Buckler, E.S.; Mitchell, S.E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **2011**, *6*, e19379. [[CrossRef](#)] [[PubMed](#)]
15. Glaubitz, J.C.; Casstevens, T.M.; Lu, F.; Harriman, J.; Elshire, R.J.; Sun, Q.; Buckler, E.S. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS ONE* **2014**, *9*, e90346. [[CrossRef](#)]
16. Money, D.; Gardner, K.; Migicovsky, Z.; Schwaninger, H.; Zhong, G.Y.; Myles, S. LinkImpute: Fast and accurate genotype imputation for nonmodel organisms. *G3 Genes Genom. Genet.* **2015**, *5*, 2383–2390. [[CrossRef](#)]
17. Yu, L.X.; Zheng, P.; Bhamidimarri, S.; Liu, X.P.; Main, D. The Impact of Genotyping-by-Sequencing Pipelines on SNP Discovery and Identification of Markers Associated with Verticillium Wilt Resistance in Autotetraploid Alfalfa (*Medicago sativa* L.). *Front. Plant. Sci.* **2017**, *8*, 89. [[CrossRef](#)]
18. Gao, H.; Williamson, S.; Bustamante, C.D. A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. *Genetics* **2007**, *176*, 1635–1651. [[CrossRef](#)]
19. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **2005**, *14*, 2611–2620. [[CrossRef](#)]

20. Breseghello, F.; Sorrells, M.E. Association Mapping of Kernel Size and Milling Quality in Wheat (*Triticum aestivum* L.) Cultivars. *Genetics* **2006**, *172*, 1165–1177. [[CrossRef](#)]
21. Gabriel, S.B.; Schaffner, S.F.; Nguyen, H.; Moore, J.M.; Roy, J.; Blumenstiel, B.; Higgins, J.; DeFelice, M.; Lochner, A.; Faggart, M.; et al. The structure of haplotype blocks in the human genome. *Science* **2002**, *296*, 2225–2229. [[CrossRef](#)]
22. Barrett, J.C.; Fry, B.; Maller, J.D.M.J.; Daly, M.J. Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* **2004**, *21*, 263–265. [[CrossRef](#)] [[PubMed](#)]
23. Ballesta, P.; Maldonado, C.; Pérez-Rodríguez, P.; Mora, F. SNP and Haplotype-Based Genomic Selection of Quantitative Traits in *Eucalyptus globulus*. *Plants* **2019**, *8*, 331. [[CrossRef](#)] [[PubMed](#)]
24. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [[CrossRef](#)] [[PubMed](#)]
25. Endelman, J.B.; Jannink, J.L. Shrinkage estimation of the realized relationship matrix. *G3 Genes Genom. Genet.* **2012**, *2*, 1405–1413. [[CrossRef](#)]
26. Contreras-Soto, R.I.; Mora, F.; de Oliveira, M.A.R.; Higashi, W.; Scapim, C.A.; Schuster, I. A genome-wide association study for agronomic traits in soybean using SNP markers and SNP-based haplotype analysis. *PLoS ONE* **2017**, *12*, e0171105. [[CrossRef](#)]
27. Arriagada, O.; Amaral-Júnior, A.T.; Mora, F. Thirteen years under arid conditions: Exploring marker-trait associations in *Eucalyptus cladocalyx* for complex traits related to flowering, stem form and growth. *Breed. Sci.* **2018**, *68*, 367–374. [[CrossRef](#)]
28. Hadfield, J. MCMCglimm Course Notes. 2012. Available online: <http://cran.r-project.org/web/packages/MCMCglimm/vignettes/CourseNotes.pdf> (accessed on 7 July 2019).
29. Mora, F.; Zúñiga, P.E.; Figueroa, C.R. Genetic variation and trait correlations for fruit weight, firmness and color parameters in wild accessions of *Fragaria chiloensis*. *Agronomy* **2019**, *9*, 506. [[CrossRef](#)]
30. Mora, F.; Ballesta, P.; Serra, N. Bayesian analysis of growth, stem straightness and branching quality in full-sib families of *Eucalyptus globulus*. *Bragantia* **2019**, *78*. [[CrossRef](#)]
31. Hadfield, J.D. MCMC methods for multi-response generalized linear mixed models: The MCMCglimm R package. *J. Stat. Softw.* **2010**, *33*, 1–22. [[CrossRef](#)]
32. Paes, G.P.; Viana, J.M.S.; Mundim, G.B. Linkage disequilibrium, SNP frequency change due to selection, and association mapping in popcorn chromosome regions containing QTLs for quality traits. *Genet. Mol. Biol.* **2016**, *39*, 97–110. [[CrossRef](#)]
33. Yan, J.; Shah, T.; Warburton, M.L.; Buckler, E.S.; McMullen, M.D.; Crouch, J. Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PLoS ONE* **2009**, *4*, e8451. [[CrossRef](#)] [[PubMed](#)]
34. Xu, J.; Liu, Y.; Liu, J.; Cao, M.; Wang, J.; Lan, H.; Xu, Y.; Lu, Y.; Pan, G.; Rong, T. The genetic architecture of flowering time and photoperiod sensitivity in maize as revealed by QTL review and meta-analysis. *J. Integr. Plant. Biol.* **2012**, *54*, 358–373. [[CrossRef](#)] [[PubMed](#)]
35. Zhang, X.; Tang, B.; Liang, W.; Zheng, Y.; Qiu, F. Quantitative genetic analysis of flowering time, leaf number and photoperiod sensitivity in maize (*Zea mays* L.). *J. Plant. Breed. Crop. Sci.* **2011**, *3*, 168–184.
36. Chardon, F.; Virlon, B.; Moreau, L.; Falque, M.; Joets, J.; Decousset, L.; Murigneux, A.; Charcosset, A. Genetic architecture of flowering time in maize as inferred from quantitative trait loci meta-analysis and synteny conservation with the rice genome. *Genetics* **2004**, *168*, 2169–2185. [[CrossRef](#)] [[PubMed](#)]
37. Liu, S.; Zenda, T.; Wang, X.; Liu, G.; Jin, H.; Yang, Y.; Dong, A.; Duan, H. Comprehensive Meta-Analysis of Maize QTLs Associated with Grain Yield, Flowering Date and Plant Height Under Drought Conditions. *J. Agric. Sci.* **2019**, *11*, 1–19. [[CrossRef](#)]
38. Coles, N.D.; McMullen, M.D.; Balint-Kurti, P.J.; Pratt, R.C.; Holland, J.B. Genetic control of photoperiod sensitivity in maize revealed by joint multiple population analysis. *Genetics* **2010**, *184*, 799–812. [[CrossRef](#)]
39. Salvi, S.; Castelletti, S.; Tuberosa, R. An updated consensus map for flowering time QTLs in maize. *Maydica* **2009**, *54*, 501–512.
40. Frey, F.P.; Prestler, T.; Lecoq, P.; Orlik, A.; Stich, B. First steps to understand heat tolerance of temperate maize at adult stage: Identification of QTL across multiple environments with connected segregating populations. *Theor. Appl. Genet.* **2016**, *129*, 945–961. [[CrossRef](#)]

41. Buckler, E.S.; Holland, J.B.; Bradbury, P.J.; Acharya, C.B.; Brown, P.J.; Browne, C.; Ersoz, E.; Flint-Garcia, S.; Garcia, A.; Glaubitz, J.C.; et al. The genetic architecture of maize flowering time. *Science* **2009**, *325*, 714–718. [[CrossRef](#)]
42. Song, Y.H.; Shogo, I.; Takato, I. Flowering time regulation: Photoperiod-and temperature-sensing in leaves. *Trends Plant Sci.* **2013**, *18*, 575–583. [[CrossRef](#)]
43. Pin, P.A.; Nilsson, O. The multifaceted roles of FLOWERING LOCUS T in plant development. *Plant. Cell Environ.* **2012**, *35*, 1742–1755. [[CrossRef](#)] [[PubMed](#)]
44. Liu, Y.Y.; Yang, K.Z.; Wei, X.X.; Wang, X.Q. Revisiting the phosphatidylethanolamine-binding protein (PEBP) gene family reveals cryptic FLOWERING LOCUS T gene homologs in gymnosperms and sheds new light on functional evolution. *New Phytol.* **2016**, *212*, 730–744. [[CrossRef](#)] [[PubMed](#)]
45. Kikuchi, R.; Kawahigashi, H.; Ando, T.; Tonooka, T.; Handa, H. Molecular and functional characterization of PEBP genes in barley reveal the diversification of their roles in flowering. *Plant. Physiol.* **2009**, *149*, 1341–1353. [[CrossRef](#)]
46. Wickland, D.P.; Hanzawa, Y. The FLOWERING LOCUS T/TERMINAL FLOWER 1 gene family: Functional evolution and molecular mechanisms. *Mol. Plant* **2015**, *8*, 983–997. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



Article

# Identification and Verification of Quantitative Trait Loci Affecting Milling Yield of Rice

Hui Zhang <sup>1,2,3</sup>, Yu-Jun Zhu <sup>2</sup>, An-Dong Zhu <sup>2</sup>, Ye-Yang Fan <sup>2</sup>, Ting-Xu Huang <sup>3</sup>, Jian-Fu Zhang <sup>3,\*</sup>, Hua-An Xie <sup>1,3,\*</sup> and Jie-Yun Zhuang <sup>2,\*</sup>

<sup>1</sup> College of Crop Science, Fujian Agriculture and Forestry University, Fuzhou 350002, China; hzhangfafu@163.com

<sup>2</sup> State Key Laboratory of Rice Biology and Chinese National Center for Rice Improvement, China National Rice Research Institute, Hangzhou 310006, China; yjzhu2013@163.com (Y.-J.Z.); zhuandong1234@163.com (A.-D.Z.); fanyeyangcnrri@163.com (Y.-Y.F.)

<sup>3</sup> Rice Research Institute and Fuzhou Branch of the National Center for Rice Improvement, Fujian Academy of Agricultural Sciences, Fuzhou 350018, China; txhuang@sina.com

\* Correspondence: jianfzhang@163.com (J.-F.Z.); huaanxie@163.com (H.-A.X.); zhuangjieyun@caas.cn (J.-Y.Z.); Tel.: +86-591-83408726 (J.-F.Z. & H.-A.X.); +86-571-63370369 (J.-Y.Z.)

Received: 25 November 2019; Accepted: 2 January 2020; Published: 5 January 2020

**Abstract:** Rice is generally consumed in the form of milled rice. The yield of total milled rice and head mill rice is affected by both the paddy rice yield and milling efficiency. In this study, three recombinant inbred line (RIL) populations and one  $F_{4.5}$  population derived from a residual heterozygous (RH) plant were used to determine quantitative trait loci (QTLs) affecting milling yield of rice. Seven traits were analyzed, including recovery of brown rice (BR), milled rice (MR) and head rice (HR); grain yield (GY); and the yield of brown rice (BRY), milled rice (MRY) and head rice (HRY). A total of 77 QTLs distributed on 35 regions was detected in the three RIL populations. Four regions, where  $qBR5$ ,  $qBR7$ ,  $qBR10$ , and  $qBR12$  were located, were validated in the RH-derived  $F_{4.5}$  population. In the three RIL populations, all the 11 QTLs for GY detected were accompanied with QTLs for two or all the three milling yield traits. Not only the allele direction for milling yield traits was unchanged, but also the effects were consistent with GY. In the RH-derived  $F_{4.5}$  population, regions controlling GY also affected all three milling yield traits. Results indicated that variations of BRY and MR were mainly ascribed to GY, but HRY was determined by both GY and HR. Results also showed that the regions covering  $GW5-Chalk5$  and  $Wx$  loci had major effects on milling quality and milling yield of rice. These two regions, which have been known to affect multiple traits determining grain quality and yield of rice, provide good candidates for milled yield improvement.

**Keywords:** brown rice recovery; milled rice recovery; head rice recovery; milling yield traits; QTL mapping; rice (*Oryza sativa* L.)

## 1. Introduction

As a major cereal crop, rice (*Oryza sativa* L.) provides staple food for at least half of the global population. Rice food is mainly consumed in the form of cooked milled rice. Farmer's incomes are based on both the paddy rice yield and the milling efficiency. In the postharvest processing, paddy grains are firstly de-hulled into brown rice and then milled into milled rice. Milled rice is separated into head rice (also called whole rice) and broken rice. Milled rice whose length is longer than or equal to 3/4 of its unbroken length falls into the category of head rice, and the rest is called broken rice. Head rice has a higher price than broken rice [1]. Three parameters, recovery of brown rice (BR), milled rice (MR) and head rice (HR), are used to evaluate rice milling quality and efficiency of the milling processing [2–4]. Generally, BR is defined as the percentage of brown rice to grains, MR the percentage

of milled rice to grains, and HR the percentage of head rice to grains. Alternatively, MR is measured as the percentage of milled rice to brown rice, and HR the percentage of head rice to milled rice [5]. In some reports, these traits are called percent milling yield, brown rice yield, milled rice yield, total milled yield, or head rice yield [3,6–8].

In the past two decades, analysis of quantitative trait loci (QTLs) was employed to study the genetic basis of rice milling quality. A larger number of QTLs were identified using various populations developed from crosses of the same subspecies [3,9], between the *indica* and *japonica* subspecies [6,10–12], or between different species [2]. In these studies, clustering of QTLs for different milling traits was commonly observed. Some of them detected only one cluster. For example, Tan et al. [9] located QTLs for MR and HR in the C1087–RZ403 region on chromosome 3; Aluko et al. [2] detected *br8* and *hr8* in RM126–RM137 on chromosome 8; and Lou et al. [12] found *qMRR-3* and *qHRR-3* in RM3204–RM6283 on chromosome 3. Other studies detected more clusters. Li et al. [6] located *qBR-4* and *qMR-4* in the C975–C734 interval on chromosome 4, *qBR-9* and *qMR-9* in R1751–R2272 on chromosome 9, *qBR-10* and *qMR-10* in C488–R716 on chromosome 10, and *qBR11* and *qMR11* in R728–G202 on chromosome 11. Zheng et al. [11] detected *QBr6* and *QMr6* in the *Wx* region on chromosome 6, and *QBr7* and *QMr7* in RM505–RM118 on chromosome 7.

When QTL mapping for milling quality traits and components of grain yield was performed using the same population, a proportion of genomic regions were found to be associated with both types of traits. Using 240 backcross introgression lines derived from the Ce258/IR75862 cross, three QTL regions were found to simultaneously affect components of milling quality and grain yield [5]. In the RM71–RM300 interval harboring *qBR2*, the Ce258 allele increased BR and panicle weight. In RM348–RM349 harboring *qBR4*, the Ce258 allele increased BR and grain weight but decreased spikelet number. In RM250–RM482 harboring *qHR2*, the Ce258 allele increased HR, grain number and spikelet number but decreased grain weight. Using 205 recombinant inbred lines (RILs) developed from the L-204/01Y110 cross, two QTL regions were found to simultaneously affect milling traits and grain weight [8]. For QTLs located in RM5638–RM1361 on chromosome 1, the L-204 allele increased BR and HR but decreased grain weight. For QTLs linked to RM3283 on chromosome 10, the L-204 allele increased HR but decreased grain weight. These results provide evidences for genetic association between milling quality and grain yield in rice, but it remains unknown whether this association has a consequence on the yield of milled and head rice.

In the present study, QTL analysis was employed to determine the dependence of milled and head rice yield on milling quality and grain yield. Firstly, three RIL populations were used to detect QTLs for brown, milled and head rice recovery; grain yield; and brown, milled and head rice yield. Then, QTL validation was performed using one secondary population derived from a residual heterozygote (RH) identified from one of the RIL populations.

## 2. Materials and Methods

### 2.1. Plant Materials

Four populations of *indica* rice (*Oryza sativa* subsp. *indica*) were used, including three RIL populations and one RH-derived F<sub>4:5</sub> population.

The three RIL populations were previously used to detect QTLs for components of appearance quality and physiochemical traits for eating and cooking quality [13–15]. The TI population consisting of 204 lines was constructed from crosses between Teqing (TQ) and IRBB lines. The IRBB lines are near isogenic lines carrying different bacterial blight resistance genes [16] in the background of IR24, including IRBB50, IRBB51, IRBB52, IRBB54, IRBB55, and IRBB59. The ZM population consisting of 230 lines was constructed from a cross between Zhenshan 97 and Milyang 46 (MY46). The XM population consisting of 209 lines was constructed from a cross between Xieqingzao and MY46. All the parental lines of these populations have been widely used in the breeding and production of three-line hybrid rice in China.

The RH-derived F<sub>4.5</sub> population was previously used to validate minor QTL for gel consistency [15]. It was derived from one RH plant that was an F<sub>7</sub> progeny of the cross TQ/IRBB52. Of the 135 polymorphic markers included in the TI map, 33 were heterozygous and 102 were homozygous in the RH plant. This plant was selfed to produce an F<sub>2</sub>-type population consisting of 250 individuals. Single seed descent was applied to advance the population to F<sub>4</sub>. Seeds of the F<sub>4</sub> plants were harvested and a population consisting of 250 F<sub>4.5</sub> families was constructed.

## 2.2. Field Experiment and Trait Measurement

All the populations were planted in the middle rice growing season (from May to October) at the China National Rice Research Institute (CNIRRI), Hangzhou (30°04' N, 119°54' E), China. The three RIL populations were tested for 2 years, including 2008 and 2009 for TI, 2009 and 2010 for ZM, and 2003 and 2009 for XM. The RH-derived F<sub>4.5</sub> population was tested for 1 year in 2017. The experiments followed a randomized complete block design. Twelve plants per line were transplanted in a single row with 16.7 cm between plants and 26.7 cm between rows. Field management followed common practice in rice production.

At maturity, five plants from the 10 middle plants of each line were randomly sampled and harvested. The grains were dried and weighted to calculate grain yield per plant (GY, g). Dried grains were stored at room temperature for three months. Then, two replicates of filled grains were processed independently. Filled grain of 100 g was de-husked using a Satake Rice Machine (Suzhou, China). The brown rice was milled using a JNMJ3 rice miller (Taizhou, China). Head rice of which the length was longer than or equal to 3/4 of the full length was separated from broken rice. Three traits for milling quality, i.e., BR, MR and HR, were calculated as the percentage of the weight of brown, milled and head rice to grain weight, respectively. Three traits for milling yield, i.e., brown, milled and head rice yield per plant, were calculated as follows:

$$\text{Brown rice yield per plant (BRY, g)} = \text{BR} \times \text{GY} \quad (1)$$

$$\text{Milled rice yield per plant (MRY, g)} = \text{MR} \times \text{GY} \quad (2)$$

$$\text{Head rice yield per plant (HRY, g)} = \text{HR} \times \text{GY} \quad (3)$$

## 2.3. Marker Data and Genetic Maps

Marker data and genetic maps of the four populations have been available [15]. The TI, ZM and XM maps were 1345.3, 1814.7 and 2080.4 cM in length, consisting of 135, 256 and 240 DNA markers, respectively. Genomic coverage and distances between neighboring markers are satisfactory for primary QTL mapping in the ZM and XM populations. A number of large homozygous segments remain in the TI map due to low polymorphism between the female and male parents of the TI population. For the RH-derived F<sub>4.5</sub> population, the map consisted of 35 markers, including 28 simple sequence repeats, six InDels and one single nucleotide polymorphism.

## 2.4. Data Analysis

For the three RIL populations that were tested for 2 years, phenotypic data averaged over 2 years were used for computing the descriptive statistics, plotting the frequency distribution and calculating the Pearson correlation coefficient, and the data of each year were used for QTL mapping. QTL analysis was performed using the default setting of the MET (multi-environmental trials) approach in IciMapping V4.1 [17], taking the 2 years for each population as two environments. LOD thresholds for genome-wide type I error of  $p < 0.05$  were calculated with 1000 permutation test and used to claim a putative QTL. QTL effects and the proportions of phenotypic variance explained ( $R^2$ ) were estimated. When a QTL was shown to have a significant genotype-by-environment (GE) interaction, the effect and  $R^2$  due to GE interaction were also measured. QTLs were designated as proposed by McCouch and CGSNL [18].

For the RH-derived F<sub>4:5</sub> population that was tested for 1 year, QTLs were determined with the BIP (bi-parental populations) approach in IciMapping V4.1 [17].  $LOD > 2.0$  was used as the threshold to claim a putative QTL. QTLs were designated as proposed by McCouch and CGSNL [18].

### 3. Results

#### 3.1. Phenotypic Performance of the Three RIL Populations

Descriptive statistics of the seven traits in the three RIL populations are presented in Table S1. Two of the seven traits, BR and MR, showed similar coefficients of variation (CV) among the three populations, ranging from 0.0116 to 0.0119 and 0.0140 to 0.0180, respectively. The CV of the five other traits were higher in the ZM and XM populations than in the TI population, ranging from 0.1266 to 0.2642 in ZM, 0.1288 to 0.2942 in XM, and 0.0967 to 0.1871 in TI (Table S1). Continuous distributions were observed for all the traits in the three populations (Figure S1), suggesting polygenic inheritance of these traits.

Correlations between the seven traits were either non-significant or positive highly significant ( $p < 0.01$ ) (Table 1). Regarding the three traits for milling quality, the correlation was strong between BR and MR but weak between these two traits and HR. The correlation coefficients ( $r$ ) between BR and MR ranged from 0.764 to 0.815 in the three populations, but their correlations with HR were either non-significant or had low  $r$  values (0.230–0.363). These results suggest that the control of properties for maintaining whole milled rice may differ greatly from that for achieving high brown and milled rice recovery.

**Table 1.** Simple correlation coefficients between seven traits in three RIL populations of rice.

Population	Trait	MR	HR	GY	BRY	MRY	HRY
TI	BR	0.764 **	−0.157	0.229 **	0.292 **	0.287 **	0.107
	MR		0.160	0.205 **	0.253 **	0.287 **	0.275 **
	HR			−0.064	−0.074	−0.046	0.505 **
	GY				0.998 **	0.996 **	0.816 **
	BRY					0.998 **	0.809 **
	MRY						0.826 **
ZM	BR	0.815 **	0.148	−0.006	0.055	0.054	0.063
	MR		0.230 **	0.063	0.112	0.136	0.157
	HR			0.354 **	0.362 **	0.368 **	0.710 **
	GY				0.998 **	0.997 **	0.902 **
	BRY					0.999 **	0.905 **
	MRY						0.907 **
XM	BR	0.784 **	0.233 **	0.246 **	0.292 **	0.296 **	0.300 **
	MR		0.363 **	0.365 **	0.398 **	0.425 **	0.449 **
	HR			0.224 **	0.232 **	0.241 **	0.605 **
	GY				0.999 **	0.998 **	0.900 **
	BRY					0.999 **	0.903 **
	MRY						0.907 **

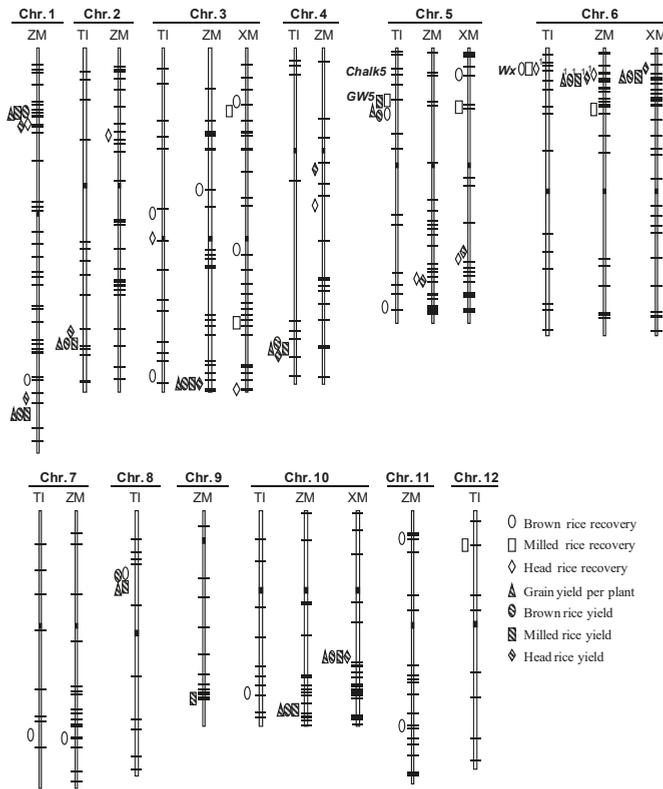
TI = Teqing/IRBB lines; ZM = Zhenshan 97/Milyang 46; XM = Xieqingzao/Milyang 46; BR = Brown rice recovery; MR = Milled rice recovery; HR = Head rice recovery; GY = Grain yield per plant; BRY = Brown rice yield per plant; MRY = Milled rice yield per plant; HRY = Head rice yield per plant; \*\*,  $p < 0.01$ .

Regarding the four yield traits, GY, BRY, MRY and HRY, not only the correlations were all significant but also the coefficients were all high. Near-perfect correlation was observed between GY, BRY and MRY, with the  $r$  values ranging from 0.996 to 0.999. Lower  $r$  values were found between these three traits and HRY, ranging from 0.816 to 0.826, 0.902 to 0.907 and 0.900 to 0.907 in the TI, ZM and XM populations, respectively. These results suggest that GY is the main source of variations for BRY, MRY and HRY, and the postharvest processing has a more significant influence on HRY than on BRY and MRY.

Regarding the three pairs of corresponding traits for milling quality and yield, the correlation was stronger between HR and HRY than between the other two pairs of traits. The  $r$  values ranged from 0.505 to 0.710 between HR and HRY in the three populations and decreased to 0.292–0.055 between BR and BRY, and 0.136–0.425 between MR and MRY. These results also suggest that the postharvest processing has a more significant influence on HRY than on BRY and MRY.

### 3.2. QTL Detected in the Three RIL Populations

In the TI, ZM and XM populations, a total of 27, 33 and 17 QTLs were detected for the seven traits analyzed, of which one, four and none showed significant GE interactions, respectively (Figure 1; Tables 2–4). Based on the physical position of DNA markers, it was found that all the five QTLs having significant GE effects were located in the region covering the *Wx* locus on the short arm of chromosome 6.



**Figure 1.** Genomic distribution of QTLs for seven traits detected in three RIL populations. TI = Teqing/IRBB lines; ZM = Zhenshan 97/Milyang 46; XM = Xieqingzao/Milyang 46. Marker positions in each chromosome are indicated by solid lines and the distances are in proportion to the physical length. Solid rectangles refer to the approximate positions of centromeres. QTLs are drawn on the left side of the corresponding interval. Significant genotype-by-environment interaction is indicated by the number “1”.

**Table 2.** QTLs for seven traits detected in the TI population.

Chr	Interval	QTL	LOD	LOD (A)	LOD (ge)	A	ge	R <sup>2</sup> (A)	R <sup>2</sup> (ge)
2	RM6–RM240	<i>qGY2</i>	8.84	8.25		−1.89		10.75	
		<i>qBRY2</i>	8.76	8.09		−1.54		10.37	
		<i>qMRY2</i>	8.15	7.61		−1.30		9.78	
		<i>qHRY2</i>	7.44	6.57		−1.21		8.38	
3	RM15139–RM15303	<i>qBR3.1</i>	26.64	25.34		0.41		14.16	
		<i>qHR3</i>	14.07	13.17		−2.51		17.81	
	RM16048–RM16184	<i>qBR3.2</i>	8.02	7.46		0.21		3.75	
4	RM6992–RM349	<i>qGY4</i>	4.90	4.88		1.44		6.23	
		<i>qBRY4</i>	4.59	4.58		1.14		5.64	
		<i>qMRY4</i>	4.53	4.52		0.99		5.69	
		<i>qHRY4</i>	3.90	3.85		0.89		4.55	
5	RM437–RM18189	<i>qBR5.1</i>	35.75	34.08		−0.51		20.20	
		<i>qMR5</i>	11.22	10.51		−0.39		11.99	
		<i>qGY5</i>	4.18	3.52		−1.30		4.52	
		<i>qBRY5</i>	5.66	4.81		−1.25		6.02	
	RM274–RM334	<i>qMRY5</i>	6.17	5.06		−1.13		6.48	
		<i>qBR5.2</i>	3.11	2.93		−0.13		1.38	
6	RM190–RM587	<i>qBR6</i>	15.18	14.32		−0.30		7.45	
		<i>qMR6</i>	6.30	5.96		−0.28		6.44	
		<i>qHR6</i>	10.08	4.35	5.73	−1.40	−1.20	5.59	4.28
7	RM70–RM18	<i>qBR7</i>	5.82	5.54		0.19		2.75	
8	RM547–RM22755	<i>qBR8</i>	8.00	6.82		−0.20		3.31	
		<i>qGY8</i>	2.85	2.82		−1.10		3.55	
		<i>qBRY8</i>	3.42	3.33		−0.98		4.09	
		<i>qMRY8</i>	2.85	2.82		−0.79		3.53	
10	RM6100–RM3773	<i>qBR10</i>	14.93	12.44		−0.28		6.60	
12	RM20–RM27610	<i>qMR12</i>	3.48	3.36		0.21		3.50	

QTLs are designated as proposed by McCouch and CGSNL [18]. A: additive effect of replacing a maternal with a paternal allele; ge: effect due to genotype-by-environment interaction; R<sup>2</sup>: percentage of phenotypic variance explained by the additive or GE effect.

**Table 3.** QTLs for seven traits detected in the ZM population.

Chr	Interval	QTL	LOD	LOD (A)	LOD (ge)	A	ge	R <sup>2</sup> (A)	R <sup>2</sup> (ge)
1	RG532–RM5359	<i>qHR1</i>	9.34	9.14		2.21		8.95	
		<i>qGY1.1</i>	5.25	4.35		0.89		4.79	
		<i>qBRY1.1</i>	5.53	4.56		0.74		4.93	
		<i>qMRY1.1</i>	5.68	4.73		0.69		5.01	
		<i>qHRY1.1</i>	6.16	4.60		0.70		4.59	
	RZ730–RG381	<i>qBR1</i>	7.34	6.67		−0.24		5.91	
		<i>qGY1.2</i>	6.63	6.61		1.12		7.51	
		<i>qBRY1.2</i>	5.96	5.93		0.86		6.59	
		<i>qMRY1.2</i>	5.58	5.58		0.76		6.02	
		<i>qHRY1.2</i>	4.69	4.30		0.67		4.34	
2	A5–RM71	<i>qHR2</i>	3.30	3.30		−1.29		3.00	
3	RM251–RG393 RZ613–RG418A	<i>qBR3</i>	6.00	5.77		−0.22		5.02	
		<i>qGY3</i>	4.17	3.14		−0.76		3.42	
		<i>qBRY3</i>	4.17	3.04		−0.61		3.24	
		<i>qMRY3</i>	4.00	2.98		−0.56		3.18	
		<i>qHRY3</i>	5.06	4.08		−0.65		4.03	
4	RZ69–RM3317	<i>qHRY4</i>	5.72	5.68		0.76		5.52	
	RM401–RM3643	<i>qHR4</i>	3.76	3.51		1.31		3.20	
5	CDO348–RG480	<i>qHR5</i>	4.95	4.94		−1.61		4.69	
		<i>qHRY5</i>	5.80	5.79		−0.79		5.81	

Table 3. Cont.

Chr	Interval	QTL	LOD	LOD (A)	LOD (ge)	A	ge	R <sup>2</sup> (A)	R <sup>2</sup> (ge)
6	RZ516–RM197	qHR6	5.72	4.78		−1.56		4.51	
		qGY6	7.97	3.30	4.67	−0.79	0.80	3.72	3.72
		qBRY6	7.74	3.22	4.52	−0.63	0.63	3.54	3.54
		qMRY6	7.38	3.13	4.24	−0.58	0.54	3.39	3.03
		qHRY6	7.41	4.15	3.27	−0.66	−0.64	4.11	3.89
	RM276–RZ667	qMR6	3.16	3.01		0.20		3.25	
7	RG650–RZ395	qBR7	4.08	3.50		0.18		3.25	
9	RG667–RM201	qMRY9	4.13	3.59		0.60		3.77	
10	RZ811–RZ583	qGY10	4.46	3.68		−0.83		4.13	
		qBRY10	5.21	4.34		−0.73		4.78	
		qMRY10	5.17	4.45		−0.68		4.82	
11	RZ816–RM332	qBR11.1	3.12	3.10		0.16		2.63	
	RM187–RM254	qBR11.2	3.20	3.18		−0.16		2.73	

QTLs are designated as proposed by McCouch and CGSNL [18]. A: additive effect of replacing a maternal with a paternal allele; ge: effect due to genotype-by-environment interaction; R<sup>2</sup>: percentage of phenotypic variance explained by the additive or GE effect.

Table 4. QTLs for seven traits detected in the XM population.

Chr	Interval	QTL	LOD	LOD (A)	A	R <sup>2</sup> (A)
3	RM6849–RM14629	qBR3.1	3.97	3.58	0.21	4.57
		qMR3.1	3.76	3.01	0.28	2.96
	RZ696–RG445A	qBR3.2	3.14	2.99	−0.19	3.58
	RZ519–RZ328	qMR3.2	3.35	3.26	−0.28	3.10
	RM85–RG418A	qHR3	3.39	3.12	−1.61	4.28
5	RM13–RM267	qBR5	3.94	3.02	0.19	3.53
	RG182–RG413	qMR5	6.86	6.86	0.43	7.05
	RM163–RG470	qHR5	3.78	3.28	−1.62	4.58
		qHRY5	3.02	3.00	−0.80	3.09
6	RM190–RM204	qGY6	7.14	6.64	−1.74	8.46
		qBRY6	7.14	6.65	−1.43	8.25
		qMRY6	7.31	6.73	−1.32	8.53
		qHRY6	4.97	4.30	−0.99	4.85
10	RM1859–RM184	qGY10	4.56	4.06	−1.36	5.15
		qBRY10	4.51	4.00	−1.11	4.96
		qMRY10	4.70	4.17	−1.04	5.28
		qHRY10	6.11	5.39	−1.11	6.05

QTLs are designated as proposed by McCouch and CGSNL [18]. A: additive effect of replacing a maternal with a paternal allele; R<sup>2</sup>: percentage of phenotypic variance explained by the additive effect.

### 3.3. QTLs Detected in the TI Population

The 27 QTLs identified in the TI population were distributed across nine of the 12 rice chromosomes (Figure 1, Table 2). Numbers of QTLs detected for BR, MR, HR, GY, BRY, MRY, and HRY were 8, 3, 2, 4, 4, 4, and 2, having overall R<sup>2</sup> of 59.55%, 21.93%, 23.40%, 25.05%, 26.13%, 25.49%, and 12.93%, respectively. Twenty-two of these QTLs formed six clusters distributed on chromosomes 2, 3, 4, 5, 6, and 8.

The largest cluster consisted of five QTLs, followed by three clusters of four QTLs. It was found that the 14 QTLs detected for grain yield and the three traits for milling yield were all included in these four clusters. In the RM437–RM18189 region on chromosome 5, the TQ allele increased BR, MR, GY, BRY, and MRY by 0.51%, 0.39%, 1.30 g, 1.25 g, and 1.13 g, respectively (Table 2). In the RM6–RM240 region on chromosome 2, the TQ allele increased GY, BRY, MRY, and HRY by 1.89 g, 1.54 g, 1.30 g, and 1.21 g, respectively. In the RM6992–RM349 region on chromosome 4, the TQ allele decreased

GY, BRY, MRY, and HRY by 1.44 g, 1.14 g, 0.99 g, and 0.89 g, respectively. In the RM547–RM22755 region on chromosome 8, the TQ allele increased BR, GY, BRY, and MRY by 0.20%, 1.10 g, 0.98 g, and 0.79 g, respectively.

The fifth cluster consisted of three QTLs, which were located in the RM190–RM587 region covering the *Wx* locus [19] on chromosome 6. The TQ allele increased BR, MR and HR by 0.30%, 0.28% and 1.40%, respectively. The sixth cluster consisted of two QTLs, which were located in the RM15139–RM15303 region covering the *GS3* locus [20] on chromosome 3. The TQ allele decreased BR by 0.41 g but increased HR by 2.51 g.

In the other five regions, one QTL was detected in each region. Included were *qBR3.2* located in the interval RM16048–RM16184 on chromosome 3, *qBR5.2* in RM274–RM334 on chromosome 5, *qBR7* in RM70–RM18 on chromosome 7, *qBR10* in RM6100–RM3773 on chromosome 10, and *qMR12* in RM20–RM27610 on chromosome 12.

### 3.4. QTLs Detected in the ZM Population

The 33 QTLs identified in the ZM population were distributed across 10 of the 12 rice chromosomes (Figure 1, Table 3). Numbers of QTLs detected for BR, MR, HR, GY, BRY, MRY, and HRY were 5, 1, 5, 5, 5, 6, and 6, having overall  $R^2$  of 19.55%, 3.25%, 24.34%, 23.57%, 23.08%, 26.20%, and 28.39%, respectively. Twenty-four of these QTLs formed six clusters distributed on chromosomes 1, 3, 5, 6, and 10.

Two clusters on chromosome 1 and one on chromosome 6 were the three largest clusters consisting of five QTLs. Each of them affected one milling quality trait and all the four yield traits. In the RG532–RM5359 region on the short-arm chromosome 1, the MY46 allele increased HR, GY, BRY, MRY, and HRY by 2.21%, 0.89 g, 0.74 g, 0.69 g, and 0.70 g, respectively (Table 3). In the RZ730–RG381 region on the long arm of chromosome 1, the MY46 allele decreased BR by 0.24% but increased GY, BRY, MRY, and HRY by 1.12 g, 0.86 g, 0.76 g, and 0.67 g, respectively. In the RZ516–RM197 region covering the *Wx* locus on chromosome 6, the MY46 allele decreased HR, GY, BRY, MRY, and HRY by 1.56%, 0.79 g, 0.63 g, 0.58 g, and 0.66 g, respectively.

The other three clusters consisted of four, three and two QTLs, respectively. The RZ613–RG418A region on chromosome 3 affected all four yield traits, with the MY46 allele decreasing GY, BRY, MRY, and HRY by 0.76 g, 0.61 g, 0.56 g, and 0.65 g, respectively. The RZ811–RZ583 region on chromosome 10 affected three yield traits, with the MY46 allele decreasing GY, BRY and MRY by 0.83 g, 0.73 g and 0.68 g, respectively. The CDO348–RG480 region on chromosome 5 affected the recovery and yield of head rice, with the MY46 allele decreasing HR and HRY by 1.61% and 0.79 g, respectively.

Two other QTLs, *qHRY4* and *qHR4*, were mapped in close positions on chromosome 4 (Figure 1). The MY46 allele increased HRY and HR by 0.76 g and 1.31%, respectively (Table 3). In the other seven regions, one QTL was detected in each region. Included were *qHR2* located in the interval A5–RM71 on chromosome 2, *qBR3* in RM251–RG393 on chromosome 3, *qMR6* in RM276–RZ667 on chromosome 6, *qBR7* in RG650–RZ395 on chromosome 7, *qMRY9* in RG667–RM201 on chromosome 9, *qBR11.1* in RZ816–RM332, and *qBR11.2* in RM187–RM254 on chromosome 11.

### 3.5. QTLs Detected in the XM Population

The 17 QTLs identified in the XM population were distributed on four of the 12 rice chromosomes (Figure 1, Table 4). Numbers of QTLs detected for BR, MR, HR, GY, BRY, MRY, and HRY were 3, 3, 2, 2, 2, 2, and 3, having overall  $R^2$  of 11.68%, 13.11%, 8.86%, 13.62%, 13.21%, 13.80%, and 13.98%, respectively. Twelve of these QTLs formed four clusters distributed on chromosomes 3, 5, 6, and 10.

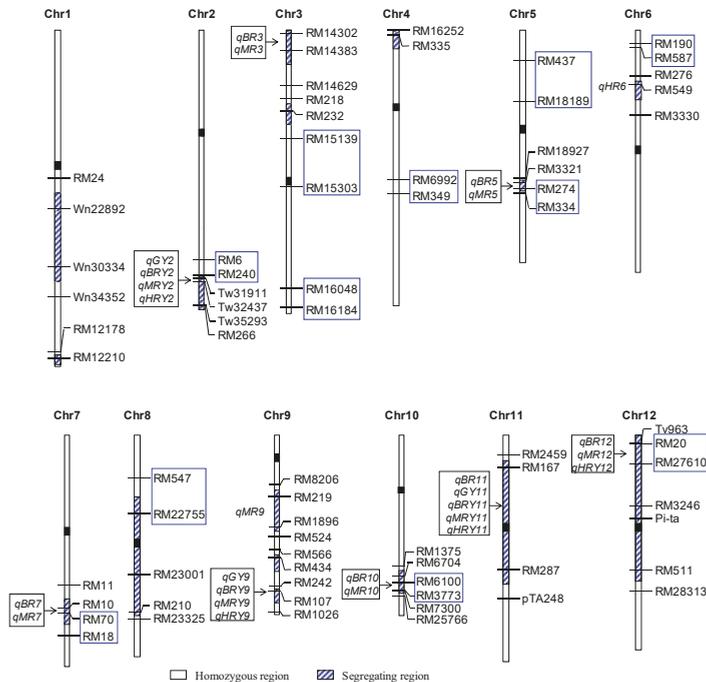
The clusters on chromosomes 6 and 10 were the two largest clusters consisting of four QTLs, both of which affected all four yield traits. In the RM190–RM204 region covering the *Wx* locus on chromosome 6, the MY46 allele decreased GY, BRY, MRY, and HRY by 1.74 g, 1.43 g, 1.32 g, and 0.99 g, respectively (Table 4). In the RM1859–RM184 region on chromosome 10, the MY46 allele decreased GY, BRY, MRY, and HRY by 1.36 g, 1.11 g, 1.04 g, and 1.11 g, respectively.

The other two clusters each consisted of two QTLs. In the RM6849–RM14629 region on the short arm of chromosome 3, the MY46 allele increased BR and MR by 0.21% and 0.28%, respectively. In the RM163–RG470 on the long arm of chromosome 5, the MY46 allele decreased HR and HRY by 1.62% and 0.80 g, respectively. Two other QTLs, *qBR5* and *qMR5*, were mapped in close positions on the short arm of chromosome 5 (Figure 1). The MY46 allele increased BR and MR by 0.19% and 0.43%, respectively (Table 4). The remaining three QTLs were loosely linked on the long arm of chromosome 3, including *qBR3.2* for brown rice recovery, *qMR3.2* for milled rice recovery, and *qHR3* for head rice recovery.

**3.6. Validation of Five QTL Regions in an RH-Derived F<sub>4:5</sub> Population**

In the Ti52-3 population that was derived from an RH-plant of TQ/IRBB52, correlations between the seven traits (Table S2) are much the same as in the three RIL populations. Regarding the three traits for milling quality, the correlation between BR and MR ( $r = 0.773$ ) was much stronger than between these two traits and HR ( $r = 0.243$  and  $0.266$ ). Regarding the four yield traits, near-perfect correlation was observed between GY, BRY and MRY ( $r$  values ranging as 0.993–0.998), and their correlations with HRY were slightly weaker ( $r$  values ranging as 0.923–0.929). Regarding the three pairs of traits for milling quality and yield, the correlation between HR and HRY ( $r = 0.440$ ) was much stronger than between the two others ( $r = 0.065$  and  $0.159$ ).

Among the 16 segregating regions distributed on 12 chromosomes in the Ti52-3 population, QTLs were detected in 10 regions across nine chromosomes (Figure 2; Table 5). A total of 26 QTLs were found, including 6, 6, 1, 3, 3, 3, and 4 for BR, MR, HR, GY, BRY, MRY, and HRY, which had overall  $R^2$  of 33.39%, 45.99%, 6.01%, 22.75%, 22.93%, 23.73%, and 23.27%, respectively.



**Figure 2.** Genomic distribution of QTLs for seven traits detected in the RH-F<sub>4:5</sub> population. BR = Brown rice recovery; MR = Milled rice recovery; HR = Head rice recovery; GY = Grain yield per plant; BRY = Brown rice yield per plant; MRY = Milled rice yield per plant; HRY = Head rice yield per plant. Markers within the blue rectangle are flanking markers of QTLs detected in the TI population.

**Table 5.** QTLs for seven traits detected in the RH-F<sub>4,5</sub> population.

Chr	Interval	QTL	LOD	A	D	R <sup>2</sup> (%)
2	Tw31911–Tw32437	<i>qGY2</i>	8.93	1.87	0.41	11.60
		<i>qBRY2</i>	8.88	1.52	0.29	11.59
		<i>qMRY2</i>	9.25	1.40	0.20	12.03
		<i>qHRY2</i>	7.35	1.29	0.16	10.67
3	RM14302–RM14383	<i>qBR3</i>	2.34	−0.13	0.05	2.82
		<i>qMR3</i>	6.41	−0.31	0.08	7.30
5	RM3321–RM274	<i>qBR5</i>	4.88	−0.17	−0.18	6.25
		<i>qMR5</i>	5.92	−0.26	−0.33	6.54
6	RM549	<i>qHR6</i>	8.05	1.34	0.31	6.01
7	RM10–RM70	<i>qBR7</i>	7.23	0.23	−0.09	9.28
		<i>qMR7</i>	12.89	0.44	−0.15	15.51
9	RM219–RM1896	<i>qMR9</i>	3.06	0.20	0.12	3.34
9	RM107	<i>qGY9</i>	2.83	−0.93	−0.84	3.35
		<i>qBRY9</i>	2.81	−0.76	−0.65	3.35
		<i>qMRY9</i>	3.09	−0.73	−0.52	3.68
		<i>qHRY9</i>	2.04	−0.62	−0.28	2.67
10	RM6704–RM7300	<i>qBR10</i>	2.20	−0.10	−0.21	2.52
		<i>qMR10</i>	2.92	−0.18	−0.24	3.29
11	RM167–RM287	<i>qBR11</i>	2.20	0.10	0.20	2.52
		<i>qGY11</i>	3.30	0.76	3.05	7.80
		<i>qBRY11</i>	3.48	0.63	2.52	8.00
		<i>qMRY11</i>	3.46	0.55	2.30	8.03
		<i>qHRY11</i>	2.43	0.36	2.34	7.19
12	Tv963–RM3246	<i>qBR12</i>	2.84	0.10	−0.57	9.99
		<i>qMR12</i>	3.74	0.12	−0.88	10.01
		<i>qHRY12</i>	2.04	0.62	−0.32	2.73

QTLs are designated as proposed by McCouch and CGSNL [18]. *A*: additive effect of replacing a maternal with a paternal allele; *D*: dominance effect; *R*<sup>2</sup>: proportion of the phenotypic variance explained by the QTL.

Of the QTLs detected in the TI population, four were covered by the segregating regions of the Ti52-3 population (Figure 2), including *qBR5.2* located in the interval RM274–RM334 on chromosome 5, *qBR7* in RM70–RM18 on chromosome 7, *qBR10* in RM6100–RM3773 on chromosome 10, and *qMR12* in RM20–RM27610 on chromosome 12. They were all well validated. The TQ alleles consistently increased BR in the *qBR5.2* and *qBR10* regions, decreased BR in the *qBR7* region, and decreased MR in the *qMR12* region (Tables 2 and 5). Additionally, significant effects were newly detected on MR in the *qBR5.2*, *qBR7* and *qBR10* regions, and on BR and HRY in the *qMR12* region. The QTL region *qBR8/qGY8/qBRY8/qMRY8* found in the TI population was overlapped with the segregating region RM22755–RM23001 in the Ti52-3 population (Figure 2). These QTLs were not detected in Ti52-3. Since one side of this putative QTL region was homozygous in the new population, it is possible that the QTLs were not segregated in Ti52-3.

The other six QTL regions found in the Ti52-3 population were not detected in the TI population. One of them, Tw31911–Tw32437 on chromosome 2, showed significant effects on four traits. In the neighboring region RM6–RM240, QTLs for the same four traits were detected in the TI population. However, the QTL directions were opposite between the two regions. It is noted that RM6–RM240 segregated in the TI population was homozygous in the Ti52-3 population (Figure 2). The gene underlying this QTL cluster may be located between RM6–RM240 and Tw31911–Tw32437, and crossover may have occurred between the gene and Tw31911.

Five other QTL regions detected in the Ti52-3 population included three QTL clusters and two regions affecting a single trait. The RM14302–RM14383 region on chromosome 3 affected two traits, in

which the TQ allele increased BR and MR by 0.13% and 0.31%, respectively. The RM107 region on chromosome 9 affected four traits, in which the TQ allele increased GY, BR, MRY, and HRY by 0.93 g, 0.76 g, 0.73 g, and 0.62 g, respectively. The RM167–RM287 region on chromosome 11 affected five traits, in which the TQ allele decreased BR, GY, BRY, MRY, and HRY by 0.10%, 0.76 g, 0.63 g, 0.55 g, and 0.36 g, respectively. The remaining two QTLs were *qHR6* and *qMR9*, of which the TQ allele decreased HR and MR by 1.34% and 0.20%, respectively.

#### 4. Discussion

Milled and head rice yield, two of the most important commercial traits in rice production, are determined by grain yield and milling quality. Understanding the genetic relationship among these traits is critical for the improvement of milled and head rice yield in breeding. In this study, QTL analysis for seven traits—brown, milled and head rice recovery, grain yield, and brown, milled and head rice yield—was performed using three RIL populations and one RH-derived F<sub>4:5</sub> population. New knowledge on the genetic basis underlying the control of brown, milled and head rice yield is provided.

In the four populations investigated in this study, correlations between the four yield traits were all highly significant. Near-perfect correlations were observed between GY, BRY and MRY, and their correlations with HRY were slightly weaker. These results were supported by QTLs detected for the four traits. Four, five, two, and three QTLs were detected for grain yield in the TI, ZM, XM, and Ti52-3 populations. It is worth noting that each of these QTL regions had significant effects on all or two of the three traits for milling yield. For multiple QTLs accompanied in the same region, not only the allelic direction remained unchanged, but also the effects were consistent. Of the four regions controlling GY in TI, the *qGY2* and *qGY4* regions controlled all four traits, but the *qGY5* and *qGY8* regions were non-significant for HRY (Table 2). No other QTLs for these traits were detected in TI. Of the five regions controlling GY in ZM, the *qGY1.1*, *qGY1.2*, *qGY3*, and *qGY6* regions controlled all four traits, but the *qGY10* region was non-significant for HRY (Table 3). One more QTL for MRY, *qMRY9*, was detected alone. One more QTL for HRY, *qHRY4*, was detected and accompanied with a QTL for HR, *qHR4*. In XM, the two regions controlling GY both affected all four traits (Table 4). One more QTL for HRY, *qHRY5*, was detected and accompanied with a QTL for HR, *qHR5*. Similarly, all three regions controlling GY in Ti52-3 affected all four traits (Table 5). These results have two implications. Firstly, variation on paddy grain yield might be the only main source of variation for brown and milled rice yield. Secondly, variations on the paddy grain yield and head rice recovery both make important contributions to the variation of head rice yield.

Significant correlations between different quality traits in rice have been commonly observed [7,8,12,21], which could be partly ascribed to the influence of a QTL region on multiple traits [7,12]. By comparing the locations of genes or QTLs reported for various grain quality traits in rice, it is found that some regions harboring QTLs for milling quality are associated with other traits that determine appearance quality or eating and cooking characteristics. Two typical examples are the *GW5-Chalk5* region on chromosome 5 [22,23] and the *Wx* region on chromosome 6 [19]. In the *GW5-Chalk5* region, QTLs having major effects for MR were detected in the TI and XM populations. In a previous study reported by Zheng et al. [11], one QTL for MR, *QMr5*, was also detected in this region, having an additive effect of 1.10% and *R*<sup>2</sup> of 11.5%. In addition, this region was reported to affect various traits for grain chalkiness, endosperm transparency and grain size in the TI and XM populations [13,14].

The *Wx* gene not only plays a key role in controlling eating and cooking quality of rice, but also influences other traits including protein content, head rice recovery, grain chalkiness, and grain weight [2,3,8,9,24]. The *Wx* locus was segregated in the TI, ZM and XM populations, having major effects on amylose content and gel consistency [15]. The *Wx* region also showed significant effects on grain chalkiness, grain width and endosperm transparency in TI; on grain chalkiness and grain length in ZM; and on grain chalkiness, grain length and endosperm transparency in XM [13,14]. In the

present study, this region was found to have significant effects on BR, MR and HR in TI; on HR, GY, BR<sub>Y</sub>, MR<sub>Y</sub>, and HR<sub>Y</sub> in ZM; and on GY, BR<sub>Y</sub>, MR<sub>Y</sub>, and HR<sub>Y</sub> in XM.

In conclusion, the *GW5-Chalk5* and *Wx* regions are good targets for studying the genetic control of multiple traits determining grain yield, appearance quality, eating and cooking quality, milling quality, and milling yield.

## 5. Conclusions

A total of 77 QTLs for seven traits affecting milling yield in rice were detected using three RIL populations. All the regions harboring QTLs for grain yield were found to affect two or all three milling yield traits. QTLs for head rice yield were usually accompanied with grain yield and head rice recovery. Variations of brown and milled rice yield were mainly ascribed to grain yield, but head rice yield was determined by both grain yield and head rice recovery. Two regions covering *GW5-Chalk5* and *Wx* loci, respectively, had a major contribution to milling quality and milling yield of rice.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4395/10/1/75/s1>, Figure S1: Frequency distribution of seven traits in three RIL populations. Table S1: Phenotypic performance of seven traits in the three RIL populations. Table S2: Simple correlation coefficients between seven traits in the RH-F<sub>4:5</sub> population.

**Author Contributions:** Conceptualization, J.-Y.Z., H.-A.X. and J.-F.Z.; investigation, H.Z., Y.-J.Z., A.-D.Z., Y.-Y.F. and T.-X.H.; writing—original draft preparation, H.Z.; writing—review and editing, J.-Y.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Science and Technology Innovation Program of the Fujian Academy of Agricultural Sciences (Grant No. STIT 2017-1-1), the National Research and Development Program (Grant No. 2016YFD0101801), the Special Foundation of Non-Profit Research Institutes of Fujian Province (Grant No. 2018R1101013-4) and the National Natural Science Foundation of China (Grant No. 31521064).

**Acknowledgments:** The authors would like to thank D.-P. Li for his assistance in field work. We acknowledge Y.-F. Sun and H.-Z. Lin for their technical assistance in laboratory works.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Nalley, L.; Tack, J.; Barkley, A.; Jagadish, K.; Brye, K. Quantifying the agronomic and economic performance of hybrid and conventional rice varieties. *Agron. J.* **2016**, *108*, 1514–1523. [\[CrossRef\]](#)
- Aluko, G.; Martinez, C.; Tohme, J.; Castano, C.; Bergman, C.; Oard, J.H. QTL mapping of grain quality traits from the interspecific cross *Oryza sativa* × *O. glaberrima*. *Theor. Appl. Genet.* **2004**, *109*, 630–639. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kepiro, J.L.; McClung, A.M.; Chen, M.H.; Yeater, K.M.; Fjellstrom, R.G. Mapping QTLs for milling yield and grain characteristics in a tropical *japonica* long grain cross. *J. Cereal Sci.* **2008**, *48*, 477–485. [\[CrossRef\]](#)
- Bao, J. Genes and QTLs for rice grain quality improvement. In *Rice-Germplasm, Genetics and Improvement*; Yan, W., Bao, J., Eds.; Intech: Rijeka, Croatia, 2014; pp. 239–278.
- Hu, X.; Shi, Y.-M.; Jia, Q.; Xu, Q.; Wang, Y.; Chen, K.; Sun, Y.; Zhu, L.-H.; Xu, J.-L.; Li, Z.-K. Analyses of QTLs for rice panicle and milling quality traits and their interaction with environment. *Acta Agron. Sin.* **2011**, *37*, 1175–1185.
- Li, Z.F.; Wan, J.M.; Xia, J.F.; Zhai, H.Q.; Ikehashi, H. Identification of quantitative trait loci underlying milling quality of rice (*Oryza sativa*) grains. *Plant Breed.* **2004**, *123*, 229–234. [\[CrossRef\]](#)
- Nelson, J.C.; McClung, A.M.; Fjellstrom, R.G.; Moldenhauer, K.A.K.; Boza, E.; Jodari, F.; Oard, J.H.; Linscombe, S.; Scheffler, B.E.; Yeater, K.M. Mapping QTL main and interaction influences on milling quality in elite US rice germplasm. *Theor. Appl. Genet.* **2011**, *122*, 291–309. [\[CrossRef\]](#)
- Nelson, J.C.; Jodari, F.; Roughton, A.I.; McKenzie, K.M.; McClung, A.M.; Fjellstrom, R.G.; Scheffler, B.E. QTL mapping for milling quality in elite western U.S. rice germplasm. *Crop Sci.* **2012**, *52*, 242–252. [\[CrossRef\]](#)
- Tan, Y.F.; Sun, M.; Xing, Y.Z.; Hua, J.P.; Sun, X.L.; Zhang, Q.F.; Corke, H. Mapping quantitative trait loci for milling quality, protein content and color characteristics of rice using a recombinant inbred line population derived from an elite rice hybrid. *Theor. Appl. Genet.* **2001**, *103*, 1037–1045. [\[CrossRef\]](#)

10. Dong, Y.; Tsuzuki, E.; Lin, D.; Kamiunten, H.; Terao, H.; Matsuo, M.; Cheng, S. Molecular genetic mapping of quantitative trait loci for milling quality in rice (*Oryza sativa* L.). *J. Cereal Sci.* **2004**, *40*, 109–114. [[CrossRef](#)]
11. Zheng, T.Q.; Xu, J.L.; Li, Z.K.; Zhai, H.Q.; Wan, J.M. Genomic regions associated with milling quality and grain shape identified in a set of random introgression lines of rice (*Oryza sativa* L.). *Plant Breed.* **2007**, *126*, 158–163. [[CrossRef](#)]
12. Lou, J.; Chen, L.; Yue, G.; Lou, Q.; Mei, H.; Xiong, L.; Luo, L. QTL mapping of grain quality traits in rice. *J. Cereal Sci.* **2009**, *50*, 145–151. [[CrossRef](#)]
13. Mei, D.-Y.; Zhu, Y.-J.; Yu, Y.-H.; Fan, Y.-Y.; Huang, D.-R.; Zhuang, J.-Y. Quantitative trait loci for grain chalkiness and endosperm transparency detected in three recombinant inbred line populations of *indica* rice. *J. Integr. Agric.* **2013**, *12*, 1–11. [[CrossRef](#)]
14. Wang, Z.; Chen, J.-Y.; Zhu, Y.-J.; Fan, Y.-Y.; Zhuang, J.-Y. Validation of *qGS10*, a quantitative trait locus for grain size on the long arm of chromosome 10 in rice (*Oryza sativa* L.). *J. Integr. Agric.* **2017**, *16*, 20–30. [[CrossRef](#)]
15. Zhang, H.; Zhu, Y.-J.; Fan, Y.-Y.; Huang, T.-X.; Zhang, J.-F.; Xie, H.-A.; Zhuang, J.-Y. Identification and verification of quantitative trait loci for eating and cooking quality of rice (*Oryza sativa*). *Plant Breed.* **2019**, *138*, 568–576. [[CrossRef](#)]
16. Huang, N.; Angeles, E.R.; Domingo, J.; Magpantay, G.; Singh, S.; Zhang, G.; Kumaravadivel, N.; Bennett, J.; Khush, G.S. Pyramiding of bacterial blight resistance genes in rice: Marker-assisted selection using RFLP and PCR. *Theor. Appl. Genet.* **1997**, *95*, 313–320. [[CrossRef](#)]
17. Meng, L.; Li, H.; Zhang, L.; Wang, J. QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J.* **2015**, *3*, 269–283. [[CrossRef](#)]
18. McCouch, S.R.; CGSNL (Committee on Gene Symbolization, Nomenclature and Linkage, Rice Genetic Cooperative). Gene nomenclature system for rice. *Rice* **2008**, *1*, 72–84. [[CrossRef](#)]
19. Wang, Z.Y.; Wu, Z.L.; Xing, Y.Y.; Zheng, F.G.; Guo, X.L.; Zhang, W.G.; Hong, M.M. Nucleotide sequence of rice *waxy* gene. *Nucleic Acids Res.* **1990**, *18*, 5898. [[CrossRef](#)]
20. Fan, C.; Xing, Y.; Mao, H.; Lu, T.; Han, B.; Xu, C.; Li, X.; Zhang, Q. *GS3*, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor. Appl. Genet.* **2006**, *112*, 1164–1171. [[CrossRef](#)]
21. Wang, D.-Y.; Zhang, X.-F.; Zhu, Z.-W.; Chen, N.; Min, J.; Yao, Q.; Yan, J.-L.; Liao, X.-Y. Correlation analysis of rice grain quality characteristics. *Acta Agron. Sin.* **2005**, *31*, 1086–1091. (In Chinese)
22. Weng, J.; Gu, S.; Wan, X.; Gao, H.; Guo, T.; Su, N.; Lei, C.; Zhang, X.; Cheng, Z.; Guo, X.; et al. Isolation and initial characterization of *GW5*, a major QTL associated with rice grain width and weight. *Cell Res.* **2008**, *18*, 1199–1209. [[CrossRef](#)] [[PubMed](#)]
23. Li, Y.; Fan, C.; Xing, Y.; Yun, P.; Luo, L.; Yan, B.; Peng, B.; Xie, W.; Wang, G.; Li, X.; et al. *Chalk5* encodes a vacuolar H<sup>+</sup>-translocating pyrophosphatase influencing grain chalkiness in rice. *Nat. Genet.* **2014**, *46*, 398–404. [[CrossRef](#)] [[PubMed](#)]
24. Zhou, P.H.; Tan, Y.F.; He, Y.Q.; Xu, C.G.; Zhang, Q. Simultaneous improvement for four quality traits of Zhenshan 97, an elite parent of hybrid rice, by molecular marker-assisted selection. *Theor. Appl. Genet.* **2003**, *106*, 326–331. [[CrossRef](#)] [[PubMed](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).





Article

# An SNP-Based High-Density Genetic Linkage Map for Tetraploid Potato Using Specific Length Amplified Fragment Sequencing (SLAF-Seq) Technology

Xiaoxia Yu <sup>†</sup>, Mingfei Zhang <sup>†</sup>, Zhuo Yu <sup>\*</sup>, Dongsheng Yang, Jingwei Li, Guofang Wu and Jiaqi Li

Agricultural College, Inner Mongolia Agricultural University, Hohhot 010000, China; yuxiaoxia1985@sina.com (X.Y.); zhangmingfei0207@163.com (M.Z.); yangdongsheng007@163.com (D.Y.); ljw2016409@163.com (J.L.); wuguofang25@163.com (G.W.); lijiaqi1127@sina.com (J.L.)

<sup>\*</sup> Correspondence: yuzhuo58@sina.com; Tel.: +86-155-9806-8434

<sup>†</sup> These authors contributed equally to this work.

Received: 30 December 2019; Accepted: 9 January 2020; Published: 13 January 2020

**Abstract:** Specific length amplified fragment sequencing (SLAF-seq) is a recently developed high-resolution strategy for the discovery of large-scale de novo genotyping of single nucleotide polymorphism (SNP) markers. In the present research, in order to facilitate genome-guided breeding in potato, this strategy was used to develop a large number of SNP markers and construct a high-density genetic linkage map for tetraploid potato. The genomic DNA extracted from 106 F<sub>1</sub> individuals derived from a cross between two tetraploid potato varieties YSP-4 × MIN-021 and their parents was used for high-throughput sequencing and SLAF library construction. A total of 556.71 Gb data, which contained 2269.98 million pair-end reads, were obtained after preprocessing. According to bioinformatics analysis, a total of 838,604 SLAF labels were developed, with an average sequencing depth of 26.14-fold for parents and 15.36-fold for offspring of each SLAF, respectively. In total, 113,473 polymorphic SLAFs were obtained, from which 7638 SLAFs were successfully classified into four segregation patterns. After filtering, a total of 7329 SNP markers were detected for genetic map construction. The final integrated linkage map of tetraploid potato included 3001 SNP markers on 12 linkage groups, and covered 1415.88 cM, with an average distance of 0.47 cM between adjacent markers. To our knowledge, the integrated map described herein has the best coverage of the potato genome and the highest marker density for tetraploid potato. This work provides a foundation for further quantitative trait loci (QTL) location, map-based gene cloning of important traits and marker-assisted selection (MAS) of potato.

**Keywords:** tetraploid potato; SNP markers; SLAF-seq technology; high-density genetic linkage map

## 1. Introduction

Potato, *Solanum tuberosum* L., is the fourth most important food crop in the world behind maize, wheat, and rice, with a total production of more than 388 million tons in 2017 [1]. Nevertheless, cultivated potato is a highly heterozygous outcrossing autotetraploid ( $2n = 4x = 48$ ), which causes complexities in genetic or genomic studies, and provides many challenges for breeding. Therefore, more breeding efforts have been focused on improving important traits, such as processing quality, nutritional value, as well as disease/pest resistance.

A high-density genetic linkage map can provide a large amount of information that facilitates map-based cloning, QTL identification, and comparative genomic researches, establishing a general tool for marker-assisted selection breeding (MAS). However, the construction of linkage maps in autopolyploids always has more difficulties than that in polyploids as well as allopolyploid species, due to their complicated segregation patterns and chromosomal pairing [2–5]. Over the past two

decades, multiple linkage maps have been constructed for potato (both diploid and autotetraploid potato) for the purpose of better understanding the potato genome, facilitating map-based cloning, and developing markers for MAS [6–11]. Gebhardt et al. (1991) [6] reported the first potato map in the world, including 135 restriction fragment length polymorphism (RFLP) molecular markers and defining 12 distinct linkage groups, which was drawn from segregation data derived from the interspecific cross of diploid potato ( $2n = 2x = 24$ ), *S. phureja* × (*S. tuberosum* × *S. chacoense*). Yamanaka et al. (2005) [10] constructed an integrated genetic linkage map of diploid potato, using 106 F<sub>1</sub> individuals from a cross of two wild and landrace germplasm 86.61.26 × 84.194.30 as the mapping population. This map included 13 newly developed P450-based analogue (PBA), 27 random amplified polymorphic DNA (RAPD), 4 inter-simple sequence repeat (ISSR), 22 simple sequence repeat (SSR), 9 restriction fragment length polymorphism-sequence-tagged sites (RFLP-STSSs), and 7 RFLP markers, with a coverage of 673 cM and an average marker distance of 8.2 cM. Van Os et al. (2006) [11] constructed an ultradense map of potato with more than 10,000 amplified fragment length polymorphism (AFLP) markers from a heterozygous diploid potato population. It is also the densest meiotic recombination map ever constructed.

With the rapid development of next-generation sequencing technologies, single nucleotide polymorphism (SNP) markers have been developed to construct high-density genetic linkage maps for many important crop species, such as maize [12,13], rice [14,15], and wheat [16,17]. For potato, Xun et al. (2011) [18] used a homozygous double-monoploid potato clone to sequence and assemble 86% of the 844-megabase genome, which bridged the gap between genomics and applied breeding with an in-depth understanding of the structure and function of the potato genome, and provided an effective tool and data to develop potato SNP markers. To date, several high-density genetic linkage maps based on SNP markers have been reported with the accomplishment and subsequent development of the potato's whole genome sequence. Felcher et al. (2012) [19] first used SNP markers and two diploid potato populations to create two linkage maps, where over 4400 markers were mapped, including 787 markers common to both populations, and the map sizes were 965 and 792 cM, respectively. Hackett et al. (2013) [20] constructed a high-density SNP map of tetraploid potato based on obtained Infinium 8300 Potato SNP Array data, which included 1130 markers with a coverage of 1087.5 cM, using a mapping population of 190 progenies from a cross between the breeding clone 12601ab1 and the cultivar *stirling*. Endelman et al. (2016) [21] first used a diploid inbred line-based F<sub>2</sub> population to construct a genetic linkage map of diploid potato with 2264 SNP markers. To sum up, most potato linkage maps are generated from diploid populations of wild species and primitive cultivars. Linkage mapping in tetraploid potato species is still a challenge despite the recent advances in mapping methodology, genotyping, and molecular marker technology.

Due to the advances in next generation sequencing (NGS) technologies, new high-throughput genotyping methods hold promise for the detection of a large number of SNPs in a short time, which include genotyping-by-sequencing (GBS) [22], complexity reduction of polymorphic sequences (CroPSs) [23], restriction site-associated DNA sequencing (RAD-seq) [24,25], and specific length amplified fragment sequencing (SLAF-seq) [26]. Specific-locus amplified fragment sequencing (SLAF-seq) technology, reported by Sun et al. (2013) [26], is an efficient strategy for the de novo SNP discovery and genotyping of large populations based on an enhanced reduced representation library (RRL) sequencing method. The advantages of SLAF-seq technology are: (i) Deep sequencing to ensure genotyping accuracy; (ii) a lower sequencing cost; (iii) pre-designed RRL scheme to optimize marker efficiency; (iv) and double barcode multiplexed sequencing system for large population and large numbers of loci. To date, this strategy has been applied to various species for SNP high-density genetic mapping, such as cucumber [27], *Agropyron gaertn* [28], and orchardgrass [29], due to its advantages of optimized marker efficiency, accurate genotyping, affordable price, and applicability for large populations. In the present research, an F<sub>1</sub> mapping population of 106 individuals was created from the cross between two tetraploid potato varieties, YSP-4 × MIN-021. We used the SLAF-seq approach to construct a high-density integrated SNP genetic linkage map of tetraploid potato, which

will expedite map-based cloning efforts, QTL location for important traits, as well as marker-assisted selection breeding for tetraploid potato.

## **2. Materials and Methods**

### *2.1. Plant Materials*

The F<sub>1</sub> mapping population consisted of 106 individuals from a cross between two tetraploid potato varieties, YSP-4 (female) and MIN-021 (male). YSP-4 is a wild tetraploid potato material, which has a short growth period, moderate tuber numbers per plant, high commodity potato rate, and high starch content (ca. 18%). This material is also highly resistant to early blight and virus disease. MIN-021 is a tetraploid potato material, which has a short growth period, high yield, and high starch content (ca. 19%). All the materials were planted in the potato breeding base of Inner Mongolia Agricultural University. The field trial was arranged in randomized complete block design (RCBD) with three replications per plot. Each plot contained 20 plants, which were grown in 2 rows with a spacing of 30 cm within rows and 90 cm between rows, and the planting depth was about 12 cm. The experiment field had sandy soil with pH 7.8 to 8.2, good irrigation conditions with annual precipitation from 300 to 400 mm, and the geographic position is 111°42' E, 45°57' N, with an altitude of 1063 m.

### *2.2. DNA Extraction*

At the potato squaring stage, the genomic DNA of all parents and 106 progenies was extracted from young fresh leaf tissue by the Plant Genomics DNA Kit (Tiangen, Beijing, China). Then, the quality of DNA was determined by electrophoresis on a 1% (*w/v*) agarose gel stained with ethidium bromide, and the concentration was quantified by an ND-1000 Spectrophotometer (Nano Drop, Wilmington, DE, USA) and adjusted to a concentration of 50 ng/μL.

### *2.3. SLAF Library Preparation and Sequencing*

According to the genome size and GC (guanine-cysteine) content of the tested materials, the potato genome ([http://solanaceae.plantbiology.msu.edu/pgsc\\_download.shtml](http://solanaceae.plantbiology.msu.edu/pgsc_download.shtml)) was selected as a reference genome to make predictions of the electronic enzyme, and finally determine the enzyme combination of Rsa I and Hae III to digest the genomic DNA of the 106 F<sub>1</sub> individuals and their two parents. The read length used for sequencing ranged from 264 to 394 bp. The SLAF labels (the length of fragments ranged from 314 to 364 bp) were selected for paired-end sequencing (125 bp per end) on an Illumina HiSeq 2500 sequencing platform, performed by the Beijing Biomarker Technologies Corporation (<http://Biomarker.com.cn/>). The SLAFs with a sequence depth of less than 10-fold were considered as low-depth SLAFs and filtered out. Several steps were defined to deal with SLAF-seq data: Samples were distinguished by barcodes and data grouping by sequence similarity; sequence error evaluation by control data; minor allele frequency (MAF) filtering and SLAF definition; correction of sequence errors; and definition and evaluation of genotypes. In addition, the quality score algorithm was developed to evaluate the quality of SNP discovery and genotyping, which can help researchers balance accuracy and cost during heterozygote detection using high-throughput sequencing technology. The Q30 (a quality score of 30; indicating 99.90% confidence) was used to evaluate the sequencing quality of reads, and examination of the base distribution was used to detect the GC content of the raw data for data quality control. The raw sequence reads were deposited in the NCBI-short read archive (SRA) database (accession: PRJNA597429).

### *2.4. SLAF Data Analysis and Development of SNP Markers*

The approach of clustering among reads was used to develop and search for polymorphic SLAF labels from 106 F<sub>1</sub> individuals and their parents. All paired-end reads generated from SLAF-seq raw reads were compared according to their sequence similarity as detected by the BLAST-like alignment

tool (BLAT) [30]. The F<sub>1</sub> individual sequence reads were aligned on the referenced potato genome using Burrows–Wheeler Aligner (BWA) software [31]. Identical reads from different samples were clustered, and the fragment with over 90% sequence identity was defined as an SLAF label. The SLAF labels with differences in high-depth fragments were also considered as SNP or indel markers. According to the differences among sequences or allele numbers, the SLAF labels were divided into three categories, including NoPoly (non-polymorphic), Poly (polymorphic), and Repeat (repetitive). After comparing the sequence differences on SLAFs from each sample, the polymorphic SLAF labels were screened for further analysis. Both Sequence Alignment/Map tools (SAMtools) [32] and Genome Analysis Toolkit (GATK) [33] were used to identify SNPs, and their intersection was identified as the candidate SNP dataset. Only biallelic SNPs were retained as the final SNPs. The SNP locus were confirmed from the polymorphic SLAF labels, with the screening criteria of MAF > 0.5.

### 2.5. Construction of High-Density Linkage Map

The HighMap software was used to construct a high-density genetic linkage map of tetraploid potato [34]. The single-linkage clustering algorithm was used to cluster the SNP markers, which were ordered into linkage groups. The high quality MLOD value among SLAF labels was calculated and used for linkage grouping. The genotyping errors were corrected using the module of error genotyping correction of HighMap software.

## 3. Results

### 3.1. SLAF Library Construction and SLAF Labels Development

The *in silico* restriction enzyme combination of *RsaI* and *HaeIII* was used for genome DNA digestion and the prediction of the potato reference genome. A total of 334,787 SLAF labels were predictably obtained, which were evenly distributed on the genome. The rice genome (*Oryza sativa*) was used as a control for the restriction enzyme digestion control trial, in order to indirectly monitor the progress of the potato SLAF library construction. Compared with the control, the ratio of paired-end mapping reads was 89.20%, and the digestion efficiency of the *RsaI* and *HaeIII* restriction enzymes was 90.91%, which indicated that the potato SLAF library was constructed normally and suitable for high-throughput sequencing.

After SLAF library construction and high-throughput sequencing, a total of 2269.98 million pair-end reads (556.71 Gb data) with a length of 100 bp were obtained. The Q30 ratio was 95.05%, and the average GC (guanine-cytosine) content was 35.51%. Of all the high-quality data, 48,849,737 reads were from the male parent MIN-21, 41,510,213 reads were from the female parent YSP-4, and the average 90,562,465 reads were from 106 offspring of the F<sub>1</sub> mapping population (Table 1). According to bioinformatics analysis, a total of 838,604 SLAF labels were developed, with an average depth of 26.14-fold and 15.36-fold for each SLAF of the parents and offspring, respectively. Of all the 838,604 high-quality SLAFs, 282,838 were polymorphic, of which 113,473 polymorphic SLAFs could be used for map construction.

**Table 1.** Basic statistic of the SLAF-seq data in tetraploid potato.

Sample	Total Reads	Total Base	Q30 Percentage (%)	GC Percentage (%)
MIN-21	48,849,737	11,944,501,020	95.18	40.28
YSP-4	41,510,213	10,126,475,890	95.66	40.18
offspring	20,562,465	5,043,758,955	95.04	39.49
Control (Rice)	9,873,113	2,422,915,844	94.69	39.49
Total	2,269,981,207	556,709,426,104	95.05	39.51

### 3.2. SNP Marker Detection

A total of 7638 SLAF labels were screened from 113,473 polymorphic SLAFs, which were successfully classified into four segregation patterns: Hk × hk, lm × ll, nn × np, and ef × eg (Table 2). The patterns, except aa × bb, were used for later genetic map construction which was suitable for the F<sub>1</sub> population, because the potato F<sub>1</sub> population was not obtained by a cross between two fully homozygous parents with genotype aa or bb. After filtering out the SNP markers with sequence depths no more than 4-fold, a total of 7329 SNP markers were detected from 7638 SLAFs for map construction.

**Table 2.** The number of SLAFs in different types of segregation patterns for map construction.

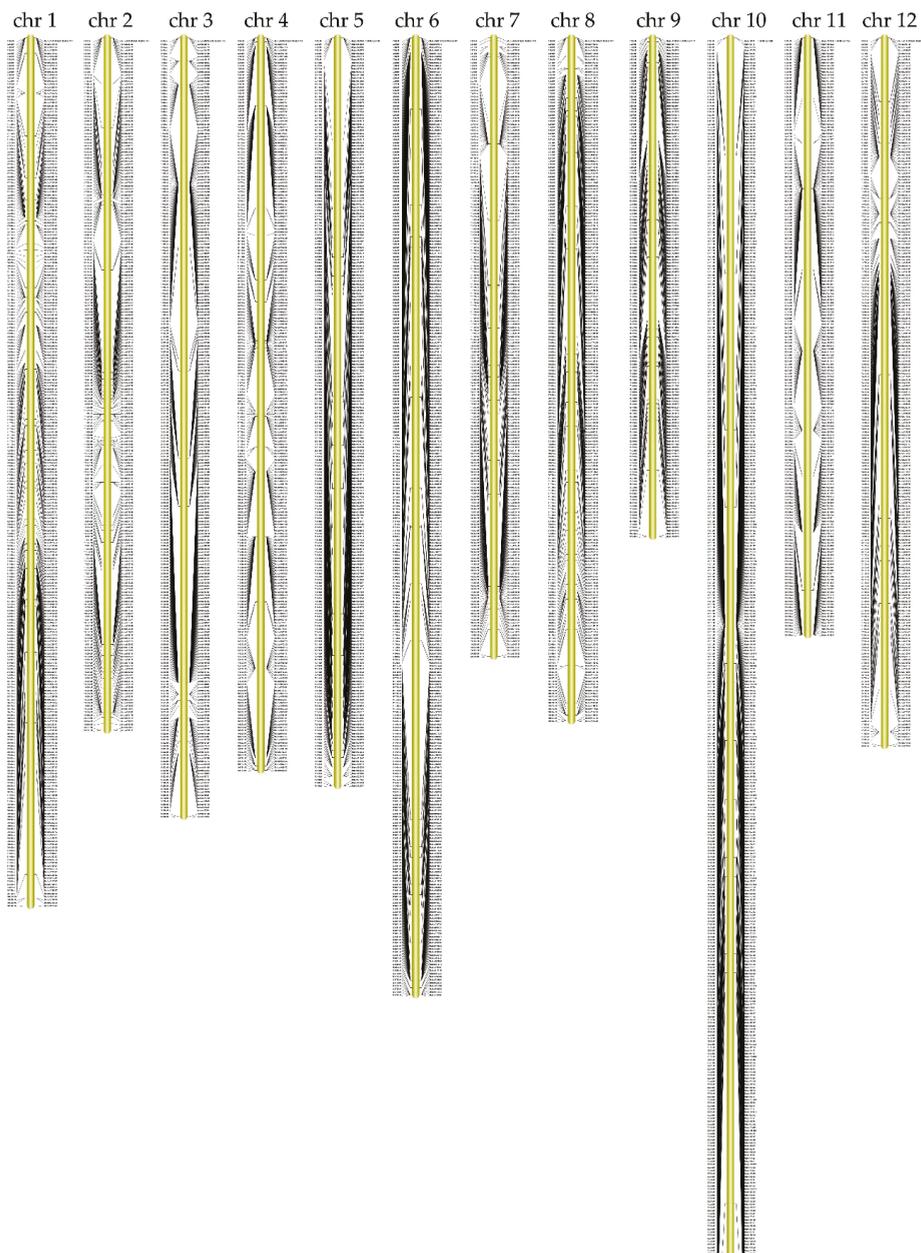
Type of Segregation Patterns	Number of SLAFs	Percentage (%)
Hk × hk	51	0.67
Lm × ll	3898	54.03
Nn × np	3638	47.63
ef × eg	51	0.67
Total	7638	100

### 3.3. Construction of the Genetic Linkage Map

After four quality control steps, the 7329 screened SNPs were used to calculate the modified logarithm of odds (MLOD) values between two markers [35]. The markers with an MLOD value of less than three were filtered, and the remaining markers were grouped into 12 linkage groups (LGs). The HighMap software was used to analyze the linear arrangements of all the grouped SNPs and the genetic distance between adjacent SNP markers within each LG. An integrated map as well as two separate linkage maps for the female and male parents were constructed, including 12 linkage groups.

In YSP-4, the maternal linkage map contained 1638 SNP markers, which covered a total length of 1383.86 cM, with an average marker distance of 0.83 cM. The number of markers in the linkage groups ranged from 43 to 341 markers, with an average of 137 markers. The length of LGs ranged from 32.82 to 282.89 cM, with an average size of 0.84 cM (Table A1). In MIN-021, the paternal linkage map consisted of 1402 SNP markers, and covered a total length of 1203.94 cM, with an average marker distance of 0.87 cM. The number of mapped markers in the LGs ranged from 542 to 243, with an average of 117 markers. The length of LGs ranged from 26.05 to 170.2 cM, with an average size of 100.33 cM (Table A2).

The integrated genetic map included 3001 SNP markers, which covered a total length of 1415.88 cM, and the average distance between adjacent markers was 0.47 cM. The number of markers in the linkage groups ranged from 43 to 341 markers, with an average of 137 markers. The length of LGs ranged from 45.02 to 282.89 cM, with an average size of 117.99 cM (Table 3, Figures 1 and A1). LG chr10 was not only the shortest but also the densest group, with 440 loci spanning 33.47 cM, which had an average marker density of 0.08 cM. LG chr2 was the longest group, with 225 loci spanning 205.09 cM. The largest gap on this map was 25.19 cM, located in LG chr7 (Table 3; Figure 1).



**Figure 1.** The high-density integrated genetic map of tetraploid potato. A total of 3001 SNP markers were distributed in 12 linkage groups, covering 1415.88 cM, with an average interval of 0.47 cM between markers.

**Table 3.** The integrated genetic map for tetraploid potato.

Linkage Group (ID)	Total Marker	Total DISTANCE (cM)	Average Distance (cM)	Max Gap (cM)	Gap $\geq 5$ cM (%)
chr1	282	107.42	0.38	12.84	99.29
chr2	225	205.09	0.92	24.43	95.98
chr3	253	63.48	0.25	14.38	98.02
chr4	238	169.78	0.72	23.76	94.51
chr5	243	76.4	0.32	10.46	98.35
chr6	311	125.8	0.41	22.86	97.1
chr7	201	203.86	1.02	25.19	95
chr8	222	109.69	0.5	14.07	97.74
chr9	162	115.41	0.72	15.34	96.89
chr10	440	33.47	0.08	5.23	99.77
chr11	194	156.97	0.81	21.35	94.82
chr12	230	48.51	0.21	7.54	99.13
Total	3001	1415.88	0.47	25.19	97.22

The average depth of the SNP markers on the integrated map was 85.63-fold in the paternal parent MIN-021 and 65.10-fold in the maternal parent YSP-4, as well as 40.34-fold in the offspring of the F<sub>1</sub> population. Segregation distortion is occurs when the segregation ratio deviates from the expected Mendelian ratio, which is considered as a powerful driving force for organic evolution [36]. The Chi-square ( $\chi^2$ ) test ( $\alpha = 0.05$ ) was used to analyze the goodness-of-fit to the expected segregation ratios for all the SNP markers. A total of 80 out of 3001 markers (2.7%) did not fit the expected segregation ratios at a level of  $\alpha \leq 0.05$ . The distorted SNP markers were mainly located on LG chr 3, chr 5, chr 7, chr 8, chr 11, and chr 12 (Table 4).

**Table 4.** The distorted SNP markers on integrated genetic map of tetraploid potato.

Linkage Group (ID)	The Number of Distorted SNP Markers
chr3	10
chr5	24
chr7	9
chr8	3
chr11	18
chr12	16
Total	80

### 3.4. Evaluation of the Genetic Map

The quality of this genetic map was evaluated by haplotype maps and heat maps, which directly revealed the recombination relationship among SNP markers in the 12 LGs. Haplotype maps were created to reflect the crossover events. The recombination events of the 12 LGs are shown on the haplotype maps (Figure A2). The haplotype maps from 12 LGs showed that all LGs had an extremely low double crossover rate, which indicated the genetic map had a high quality.

Heat maps were also constructed to evaluate the quality of the genetic map by using pair-wise recombination values for the 3001 SNP markers (Figure A3). It showed that most of the heat maps for 12 LGs performed well in visualization, which indicated that the markers were well-ordered, and the genetic distances of adjacent markers were accurate in each LG.

## 4. Discussion

### 4.1. The Development of SNP Markers Using SLAF-seq Technology

A genetic linkage map is the basis for QTL identification of important traits, map-based gene cloning, and molecular marker-assisted breeding of crops. Different types and numbers of polymorphic markers were used to construct genetic maps. For potato, most genetic linkage maps were mainly

based on several conventional low-throughput molecular markers, including RFLP [6–10], AFLP [9,11], as well as RAPD, ISSR, and SSR markers [10]. However, it is time-consuming and costly to construct a high-density genetic map for potato using conventional molecular marker technologies. SNP markers are the most frequent polymorphisms and are suitable for high-throughput genotyping. In addition, many SNP markers are located within transcribed regions, which can generate more links between the genetic and physical maps. To date, high-density polymorphic SNP markers have been used in potato for large-scale genotyping and high-density genetic map construction [19–21].

SNP markers can be rapidly developed on a large scale by different high-throughput sequencing technologies and genotyping methods, such as genotyping by sequencing (GBS) [22], restriction site-associated sequencing (RAD-seq) [24,25], and SLAF-seq. The SLAF-seq technology, a combination of locus-specific amplification and high-throughput sequencing, provides a high-resolution strategy with a shorter period of time and lower cost for large-scale genotyping and can be generally applicable to various species and populations [26].

In the present research, we first used SLAF-seq technology in potato to develop SNP markers and construct the high-density genetic map. A total of 2269.98 million pair-end reads were obtained based on high-throughput sequencing. According to bioinformatics analysis, a total of 838,604 SLAF labels were generated, of which 282,838 were polymorphic. Finally, a total of 7329 polymorphic SNP markers were developed for high-density genetic map construction. The present study extends the utility of SLAF-seq technology to potato. The results showed that SLAF-seq was an effective tool to rapidly develop large-scale SNP markers, which met the requirements for high-density genetic map construction of tetraploid potato.

#### 4.2. Mapping Population and Strategies

In general, the  $F_2$ , backcross (BC), doubled haploid (DH), or recombinant inbred lines (RIL) population are used as an appropriate mapping population to construct genetic linkage maps [5,37,38]. However, for most autopolyploid species, it is very difficult to obtain a typical family-based population in potato because of its high heterozygosity. To date, most of the reported potato linkage maps have been established by applying a double pseudo-testcross strategy on an  $F_1$  population. The double pseudo-testcross strategy was first put forward by Grattapaglia and Sederoff (1994) [37] to construct the genetic linkage map for genetically heterozygous species of forest trees. An  $F_1$  population was used as a mapping population by crossing between two irrelevant and highly heterozygous parents. The gene segregation patterns were assumed as a backcross. Afterwards, this strategy has been widely used to construct linkage maps for those heterozygous species, such as danshen [39], pineapple [40], rhodesgrass [41], and sweet potato [5].

In the present research, an  $F_1$  segregation population from a cross YSP-4  $\times$  MIN-021 was created, of which 106 individuals were randomly selected and used for SNP genotyping and map construction based on the double pseudo-testcross strategy. In the pseudo-testcross, a total of 7638 polymorphic SLAF markers were classified into four segregation patterns, which were  $hk \times hk$ ,  $lm \times ll$ ,  $nn \times np$ , and  $ef \times eg$ . The 7329 SNP markers screened and confirmed from 7638 polymorphic SLAFs were then used to construct a genetic linkage map. In our study, among the 838,604 high-quality SLAFs, 282,838 were polymorphic, with a polymorphic rate of 33.7%. It indicates that there is considerable genetic difference between YSP-4 and MIN-021. Therefore, it is suitable to use them as mapping parents, and the  $F_1$  population derived from the cross between them conforms to the requirements of the mapping population for high-density map construction.

#### 4.3. The High-Density Genetic Map of Tetraploid Potato

Segregation distortion is a common phenomenon that has been observed in many studies [42–44]. It may generate from cytological attributes, genetic drift, gametophyte selection, or some biological reasons [45,46]. Segregation distortion could alter the estimation of recombination and cause a spurious linkage [47]. Therefore, distorted markers may affect the accuracy of genetic maps. In our study,

only 2.7% SNP markers located on the integrated map were distorted markers, which indicated the high map accuracy.

To our knowledge, only one high-density SNP genetic linkage map for tetraploid potato was reported because of the high heterozygosity of autotetraploid potato [20]. In the present research, we first used the SLAF-seq method for genotyping and developing SNP markers, and constructed high-density genetic maps of tetraploid potato. The integrated map included 3001 SNP markers, and had a genetic length of 1415.88 cM, with an average distance between markers of 0.47 cM. Compared with the map obtained by Hackett et al. (2013) [20], the integrated map had more SNP markers (3001 vs. 1130), higher marker density (0.47 cM vs. 1.60 cM), and larger total length (1415.88 cM vs. 1087.5 cM). Thus, our map has better coverage of the potato genome and nearer marker density.

## 5. Conclusions

In the present study, the SLAF-seq technology was first successfully used for the development of large-scale SNP markers and the construction of high-density linkage maps in tetraploid potato. The integrated high genetic linkage map generated here has the best coverage of the potato genome and the nearest marker density reported for tetraploid potato until now. This work represents an important step forward in genomics and marker-assisted breeding of tetraploid potato. It also provides a foundation for QTL location and map-based gene cloning of important traits for potato, such as tuber yield, starch content, and protein content. In addition, the application of SLAF-seq strategy and the mapping population in our study will provide valuable references for other tetraploid plants.

**Author Contributions:** Z.Y. and X.Y. conceived and designed the research. Z.Y., X.Y. and M.Z. performed the experiments. X.Y. and M.Z. completed the writing of the article. D.Y., J.L. (Jingwei Li), G.W., and J.L. (Jiaqi Li) assisted in the completion of the experiments. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by grants from the Inner Mongolia Major Science and Technology Project (zdzx2018019), Inner Mongolia Youth Science and Technology Talents Project for Colleges and Universities (NJYT-18-B04), Inner Mongolia Natural Science Foundation (2018MS03040), Major Project for Inner Mongolia Science and Technology Program (201602048), and Animal and Plant Breeding Project for Transformation of Scientific Achievements in Inner Mongolia Agricultural University (YZGC2017006).

**Conflicts of Interest:** The authors declare no conflict of interest.

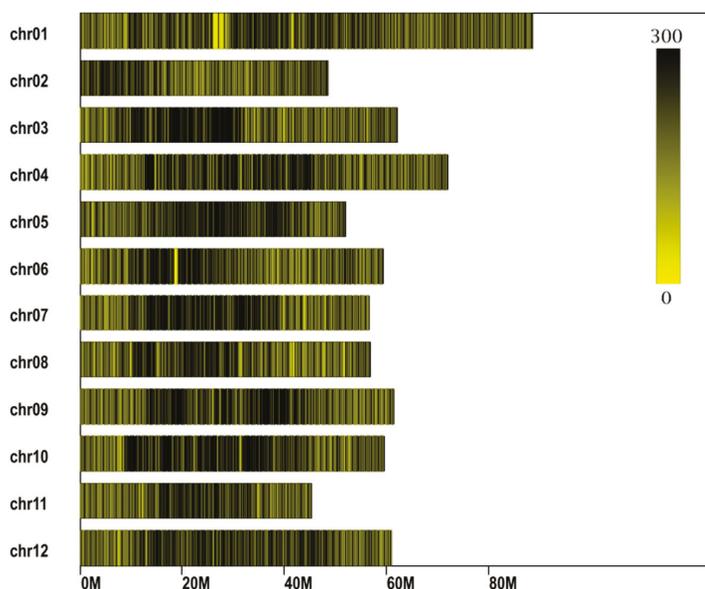
## Appendix A

**Table A1.** The maternal genetic map for tetraploid potato.

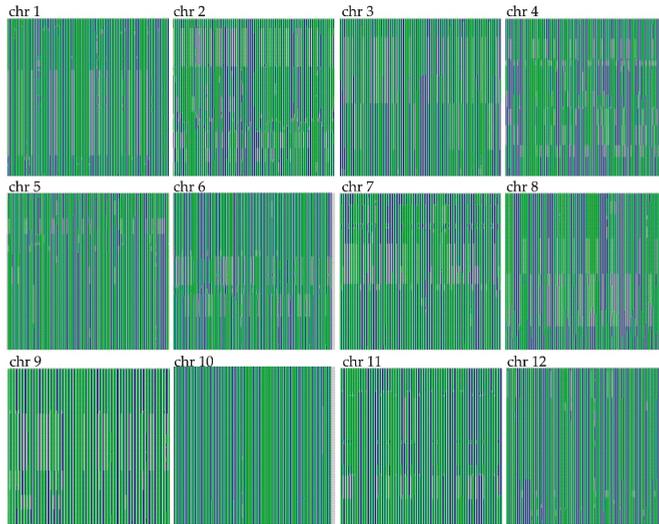
Linkage Group (ID)	Total Marker	Total Distance (cM)	Average Distance (cM)	Max Gap (Cm)	Gap $\geq$ 5 cM (%)
chr1	43	54.06	1.29	26.6	95.24
chr2	70	199.96	2.9	66.56	86.96
chr3	138	58.29	0.43	48.73	99.27
chr4	107	149.13	1.41	30.15	93.4
chr5	167	45.02	0.27	21.7	98.8
chr6	235	172.13	0.74	31.9	95.73
chr7	153	282.89	1.86	88.65	94.74
chr8	153	116.85	0.77	15.34	96.71
chr9	80	91.37	1.16	55.92	98.73
chr10	341	32.82	0.1	7.32	99.85
chr11	57	123.36	2.2	31.91	91.07
chr12	94	57.8	0.62	16.63	96.77
Total	1638	1383.68	0.84	88.65	95.62

**Table A2.** The paternal genetic map for tetraploid potato.

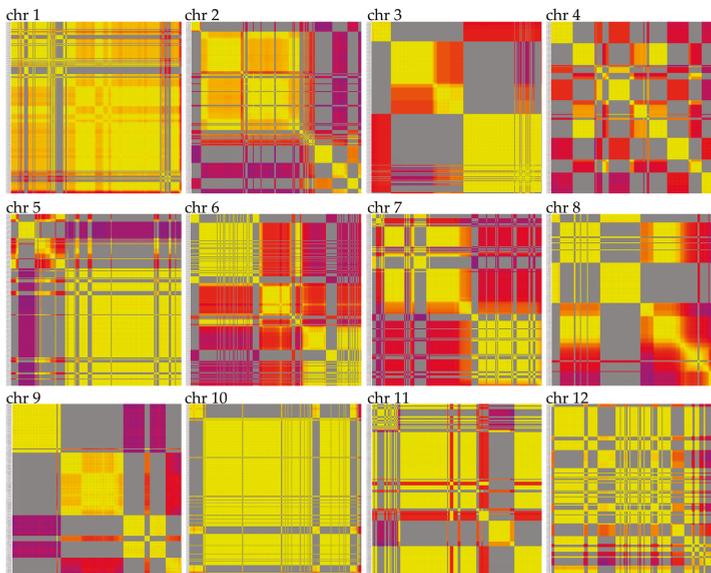
Linkage Group (ID)	Total Marker	Total Distance (cM)	Average Distance (cM)	Max Gap (cM)	Gap $\geq$ 5 cM (%)
chr1	243	117.29	0.48	12.87	99.17
chr2	156	167.27	1.08	28.45	96.77
chr3	125	51.51	0.42	15.34	96.77
chr4	132	151.04	1.15	51.29	95.42
chr5	76	83.47	1.11	10.46	93.33
chr6	83	77.56	0.95	37.57	96.34
chr7	52	101.5	1.99	71.52	96.08
chr8	71	88.68	1.27	70.27	98.57
chr9	82	108.18	1.34	16.63	93.83
chr10	102	52.42	0.52	10.46	98.02
chr11	138	178.97	1.31	41.74	96.35
chr12	142	26.05	0.18	3.92	100
Total	1402	1203.94	0.87	71.52	96.72



**Figure A1.** The SNP distribution on the potato genome. The *x*-axis represents the chromosome length and the *y*-axis indicates the chromosome code. Each band represents a chromosome, and the genome is divided according to the size of 1 M. The more SNPs in each band, the darker the color; the smaller the number of SNPs, the lighter the color. The darker areas in the figure are the areas where SNPs are concentrated.



**Figure A2.** Haplotype maps for 12 linkage groups of the integrated genetic map for tetraploid potato. The haplotype maps consist of 12 maps from LG chr 1 to LG chr 12. Each two columns represent the genotype of an individual. Blank columns are used between two individuals. The first and second columns represent the paternal and maternal chromosome, respectively. Rows correspond to genetic markers. Green and blue boxes indicate one chromatid from parents, and gray boxes indicate missing data.



**Figure A3.** Heat maps for 12 linkage groups of the integrated genetic map for tetraploid potato. The heat maps consist of 12 maps from LG chr 1 to LG chr 12. Markers of each row and column are ranked according to the map order. Each small square represents the rate of recombination between two markers. Yellow color represents highly tight linkage; red color represents relatively weak linkage, the darker the red color, the less tight linkage; and blue color represents no linkage.

## References

1. FAOSTAT. 2017. Available online: <http://faostat.fao.org/site/339/default.aspx> (accessed on 2 September 2019).
2. Hermesen, J.G.T. Nature, evolution and breeding of polyploids. *Iowa State J. Res.* **1984**, *58*, 411–420.
3. Sybenga, J. Chromosome pairing affinity and quadrivalent formation in polyploids: Do segmental allopolyploids exist? *Genome* **1996**, *39*, 1176–1184. [[CrossRef](#)]
4. Cervantes-Flores, J.C.; Yencho, G.C.; Kriegner, A.; Pecota, K.V.; Faulk, M.A.; Mwanga, R.O.M.; Sosinski, B.R. Development of a genetic linkage map and identification of homologous linkage groups in sweetpotato using multiple-dose AFLP markers. *Mol. Breed.* **2008**, *21*, 511–532. [[CrossRef](#)]
5. Zhao, N.; Yu, X.X.; Jie, Q.; Li, H.; Li, H.; Hu, J.; Zhai, H.; He, S.Z.; Liu, Q.C. A genetic linkage map based on AFLP and SSR markers and mapping of QTL for dry-matter content in sweetpotato. *Mol. Breed.* **2013**, *32*, 807–820. [[CrossRef](#)]
6. Gebhardt, C.; Ritter, E.; Barone, A.; Debener, T.; Walkemeier, B.; Schachtschabel, U.; Kaufmann, H.; Thompson, R.D.; Bonierbale, M.W.; Ganal, M.W. RFLP maps of potato and their alignment with the homoeologous tomato genome. *Theor. Appl. Genet.* **1991**, *83*, 49–57. [[CrossRef](#)] [[PubMed](#)]
7. Tanksley, S.D.; Ganal, M.W.; Prince, J.P.; De Vicente, M.C.; Bonierbale, M.W.; Broun, P.; Fulton, T.M.; Giovannoni, J.J.; Grandillo, S.; Martin, G.B. High density molecular linkage maps of the tomato and potato genomes. *Genetics* **1992**, *132*, 1141–1160. [[PubMed](#)]
8. Jacobs, J.M.E.; Van Eck, H.J.; Arens, P.; Verkerk-Bakker, B.; Te Lintel Hekkert, B.; Bastiaanssen, H.J.M.; El-Kharbotly, A.; Pereira, A.; Jacobsen, E.; Stiekema, W.J. A genetic map of potato (*Solanum tuberosum*) integrating molecular markers, including transposons, and classical markers. *Theor. Appl. Genet.* **1995**, *91*, 289–300. [[CrossRef](#)]
9. Caromel, B.; Mugniéry, D.; Lefebvre, V.; Andrzejewski, S.; Ellissèche, D.; Kerlan, M.C.; Rousselle, P.; Roussellebourgeois, F. Mapping QTLs for resistance against *Globodera pallida* (Stone) Pa2/3 in a diploid potato progeny originating from *Solanum spegazzinii*. *Theor. Appl. Genet.* **2003**, *106*, 1517–1523. [[CrossRef](#)]
10. Yamanaka, S.; Ikeda, S.; Imai, A.; Luan, Y.; Watanabe, J.A.; Watanabe, K.N. Construction of integrated genetic map between various existing DNA markers and newly developed P450-related PBA markers in diploid potato (*Solanum tuberosum*). *Breed. Sci.* **2005**, *55*, 223–230. [[CrossRef](#)]
11. Van Os, H.; Andrzejewski, S.; Bakker, E.; Barrena, I.; Glenn, J.B.; Caromel, B.; Ghareeb, B.; Isidore, E.; De Jong, W.; Van Koert, P. Construction of a 10,000-marker ultradense genetic recombination map of potato: Providing a framework for accelerated gene isolation and a genomewide physical map. *Genetics* **2006**, *173*, 1075–1087. [[CrossRef](#)]
12. Sa, K.J.; Park, J.Y.; Park, K.C.; Ju, K.L. Analysis of genetic mapping in a waxy/dent maize RIL population using SSR and SNP markers. *Genes. Genom.* **2012**, *34*, 157–164. [[CrossRef](#)]
13. Su, C.F.; Wang, W.; Gong, S.L.; Zuo, J.H.; Li, S.J.; Xu, S.Z. High density linkage map construction and mapping of yield trait QTLs in maize (*Zea mays*) using the genotyping-by-sequencing (GBS) technology. *Front. Plant Sci.* **2017**, *8*, 706–719. [[CrossRef](#)] [[PubMed](#)]
14. Zhai, H.J.; Feng, Z.Y.; Liu, X.Y.; Cheng, X.J.; Peng, H.R.; Yao, Y.Y.; Sun, Q.X.; Ni, Z.F. A genetic linkage map with 178 SSR and 1 901 SNP markers constructed using a RIL population in wheat (*Triticum aestivum* L.). *J. Integr. Agric.* **2015**, *14*, 1697–1705. [[CrossRef](#)]
15. Leon, T.B.D.; Linscombe, S.; Subudhi, P.K. Molecular dissection of seedling salinity tolerance in rice (*Oryza sativa* L.) using a high-density GBS-based SNP linkage map. *Rice* **2016**, *9*, 52–73. [[CrossRef](#)] [[PubMed](#)]
16. Yang, Q.; Yang, Z.J.; Tang, H.F.; Yu, Y.; Chen, Z.Y.; Wei, S.H.; Sun, Q.X.; Peng, Z.S. High-density genetic map construction and mapping of the homologous transformation sterility gene (*hts*) in wheat using *gbs* markers. *BMC Plant Biol.* **2018**, *18*, 301–309. [[CrossRef](#)] [[PubMed](#)]
17. Colasuonno, P.; Gadaleta, A.; Giancaspro, A.; Nigro, D.; Giove, S.; Incerti, O.; Mangini, G.; Signorile, A.; Simeone, R.; Blanco, A. Development of a high-density SNP-based linkage map and detection of yellow pigment content QTLs in durum wheat. *Mol. Breed.* **2014**, *34*, 1563–1578. [[CrossRef](#)]
18. Xu, X.; Pan, S.K.; Cheng, S.F.; Zhang, B.; Mu, D.S.; Ni, P.X.; Zhang, G.Y.; Yang, S.; Li, R.Q.; Wang, J. Genome sequence and analysis of the tuber crop potato. *Nature* **2011**, *475*, 189–195.
19. Felcher, K.J.; Coombs, J.J.; Massa, A.N.; Hansey, C.N.; Hamilton, J.P.; Veilleux, R.E.; Buell, C.R.; Douches, D.S. Integration of two diploid potato linkage maps with the potato genome sequence. *PLoS ONE* **2012**, *7*, e36347. [[CrossRef](#)]

20. Hackett, C.A.; Mclean, K.; Bryan, G.J. Linkage analysis and QTL mapping using SNP dosage data in a tetraploid potato mapping population. *PLoS ONE* **2013**, *8*, e63939. [[CrossRef](#)]
21. Endelman, J.B.; Jansky, S.H. Genetic mapping with an inbred line-derived F<sub>2</sub> population in potato. *Theor. Appl. Genet.* **2016**, *129*, 935–943. [[CrossRef](#)]
22. Elshire, R.J.; Glaubitz, J.C.; Sun, Q.; Poland, J.A.; Kawamoto, K.; Buckler, E.S.; Mitchell, S.E. A robust, simple genotyping by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **2011**, *6*, e19379. [[CrossRef](#)] [[PubMed](#)]
23. Van Orsouw, N.J.; Hogers, R.C.; Janssen, A.; Yalcin, F.; Snoeijers, S.; Verstege, E.; Schneiders, H.; van der Poel, H.; Van Oeveren, J.; Verstegen, H.; et al. Complexity reduction of polymorphic sequences (CRoPS): A novel approach for large-scale polymorphism discovery in complex genomes. *PLoS ONE* **2007**, *2*, e1172. [[CrossRef](#)] [[PubMed](#)]
24. Miller, M.R.; Dunham, J.P.; Amores, A.; Cresko, W.A.; Johnson, E.A. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res.* **2007**, *17*, 240–248. [[CrossRef](#)] [[PubMed](#)]
25. Baird, N.A.; Etter, P.D.; Atwood, T.S.; Currey, M.C.; Shiver, A.L.; Lewis, Z.A.; Selker, E.U.; Cresko, W.A.; Johnson, E.A. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* **2008**, *3*, e3376. [[CrossRef](#)]
26. Sun, X.W.; Liu, D.Y.; Zhang, X.F.; Li, W.B.; Liu, H.; Hong, W.G.; Jiang, C.B.; Guan, N.; Ma, C.X.; Zeng, H.P. SLAF-seq: An efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. *PLoS ONE* **2013**, *8*, e58700. [[CrossRef](#)] [[PubMed](#)]
27. Xu, X.W.; Lu, L.; Zhu, B.Y.; Xu, Q.; Qi, X.H.; Chen, X.H. QTL mapping of cucumber fruit flesh thickness by SLAF-seq. *Sci. Rep.* **2015**, *5*, 15829. [[CrossRef](#)]
28. Zhang, Y.; Zhang, J.P.; Huang, L.; Gao, A.N.; Yang, X.M.; Liu, W.H.; Li, X.Q.; Li, L.H. A high-density genetic map for P genome of *Agropyron Gaertn.* based on specific-locus amplified fragment sequencing (SLAF-seq). *Planta* **2015**, *242*, 1335–1347. [[CrossRef](#)]
29. Zhao, X.X.; Huang, L.K.; Zhang, X.Q.; Wang, J.P.; Yan, D.F.; Li, J.; Tang, L.; Li, X.L.; Shi, T.W. Construction of high-density genetic linkage map and identification of flowering-time QTLs in orchardgrass using SSRs and SLAF-seq. *Sci. Rep.* **2016**, *6*, 29345. [[CrossRef](#)]
30. Ken, W.J. BLAT—The BLAST-like alignment tool. *Genome Res.* **2002**, *12*, 656–664.
31. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)]
32. Depristo, M.A.; Banks, E.; Poplin, R.E.; Garimella, K.V.; Maguire, J.R.; Hartl, C. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **2011**, *43*, 491–498. [[CrossRef](#)] [[PubMed](#)]
33. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)] [[PubMed](#)]
34. Liu, D.Y.; Ma, C.X.; Hong, W.G.; Huang, L.; Liu, M.; Liu, H.; Zeng, H.P.; Deng, D.J.; Xin, H.G.; Song, J.; et al. Construction and analysis of high-density linkage map using high-throughput sequencing data. *PLoS ONE* **2014**, *9*, e98855. [[CrossRef](#)] [[PubMed](#)]
35. Vision, T.J.; Brown, D.G.; Shmoys, D.B.; Durrett, R.T.; Tanksley, S.D. Selective mapping: A strategy for optimizing the construction of high-density linkage maps. *Genetics* **2000**, *155*, 407–420.
36. Foisset, N.; Delourme, R.; Barret, P.; Hubert, N.; Landry, B.S.; Renard, M. Molecular-mapping analysis in *Brassica napus* using isozyme, RAPD and RFLP markers on a doubled-haploid progeny. *Theor. Appl. Genet.* **1996**, *93*, 1017–1025. [[CrossRef](#)]
37. Grattapaglia, D.; Sederoff, R. Genetic linkage maps of eucalyptus *grandis* and *eucalyptus urophylla* using a pseudo-testcross: Mapping strategy and RAPD markers. *Genetics* **1994**, *137*, 1121–1137.
38. Venkat, S.K.; Bommisetty, P.; Patil, M.S.; Reddy, L.; Chennareddy, A. The genetic linkage maps of anthurium species based on RAPD, ISSR and SRAP markers. *Sci. Hortic.* **2014**, *178*, 132–137. [[CrossRef](#)]
39. Liu, T.; Guo, L.L.; Pan, Y.L.; Zhao, Q.; Wang, J.H.; Song, Z.Q. Construction of the first high-density genetic linkage map of *salvia miltiorrhiza* using specific length amplified fragment (SLAF) sequencing. *Sci. Rep.* **2016**, *6*, 24070. [[CrossRef](#)]

40. Carlier, J.D.; Reis, A.; Duval, M.F.; Coppens d'Eeckenbrugge, G.; Leitao, J.M. Genetic maps of RAPD, AFLP and ISSR markers in *Ananas bracteatus* and *A. comosus* using the pseudo-testcross strategy. *Plant Breed.* **2010**, *123*, 186–192. [[CrossRef](#)]
41. Ubi, B.E.; Fujimori, M.; Mano, Y.; Komatsu, T. A genetic linkage map of rhodesgrass based on an F<sub>1</sub> pseudo-testcross population. *Plant Breed.* **2010**, *123*, 247–253. [[CrossRef](#)]
42. Cloutier, S.; Ragupathy, R.; Niu, Z.; Duguid, S. SSR-based linkage map of flax (*Linum usitatissimum* L.) and mapping of QTLs underlying fatty acid composition traits. *Mol. Breed.* **2011**, *28*, 437–451. [[CrossRef](#)]
43. Escudero, M.; Hahn, M.; Hipp, A.L. RAD-seq linkage mapping and patterns of segregation distortion in sedges: Meiosis as a driver of karyotypic evolution in organisms with holocentric chromosomes. *J. Evol. Biol.* **2018**, *31*, 833–843. [[CrossRef](#)] [[PubMed](#)]
44. Moreira, F.M.; Madini, A.; Marino, R.; Zulini, L.; Stefanini, M.; Velasco, R.; Kozma, P.; Grando, M.S. Genetic linkage maps of two interspecific grape crosses (*vitispp.*) used to localize quantitative trait loci for downy mildew resistance. *Tree Genet. Genomes* **2011**, *7*, 153–167. [[CrossRef](#)]
45. Shappley, Z.W.; Jenkins, J.N.; Meredith, W.R.; Mccarty, J.C. An RFLP linkage map of upland cotton, *Gossypium hirsutum* L. *Theor. Appl. Genet.* **1998**, *97*, 756–761. [[CrossRef](#)]
46. Lacape, J.M.; Nguyen, T.B.; Thibivilliers, S.; Bojinov, B.; Courtois, B.; Cantrell, R.G.; Burr, B.; Hau, B. A combined RFLP-SSR-AFLP map of tetraploid cotton based on a *Gossypium hirsutum* x *Gossypium barbadense* backcross population. *Genome* **2003**, *46*, 612–626. [[CrossRef](#)] [[PubMed](#)]
47. Cloutier, S.; Cappadocia, M.; Landry, B.S. Analysis of RFLP mapping inaccuracy in *Brassica napus* L. *Theor. Appl. Genet.* **1997**, *95*, 83–91. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



Article

# Mapping Agronomic and Quality Traits in Elite Durum Wheat Lines under Differing Water Regimes

Rosa Mérida-García <sup>1</sup>, Alison R. Bentley <sup>2</sup>, Sergio Gálvez <sup>3</sup>, Gabriel Dorado <sup>4</sup>, Ignacio Solís <sup>5</sup>, Karim Ammar <sup>6</sup> and Pilar Hernandez <sup>1,\*</sup>

<sup>1</sup> Institute for Sustainable Agriculture (IAS-CSIC), Consejo Superior de Investigaciones Científicas (CSIC), Alameda del Obispo s/n, 14004 Córdoba, Spain; rmerida@ias.csic.es

<sup>2</sup> The John Bingham Laboratory, NIAB, Huntingdon Road, Cambridge CB3 0LE, UK; alison.bentley@niab.com

<sup>3</sup> Departamento de Lenguajes y Ciencias de la Computación, Campus de Teatinos s/n, Universidad de Málaga, Andalucía Tech, ETSI Informática, 29071 Málaga, Spain; galvez@uma.es

<sup>4</sup> Departamento de Bioquímica y Biología Molecular, Campus Rabanales C6-1-E17, Campus de Excelencia Internacional Agroalimentario (ceiA3), Universidad de Córdoba, 14071 Córdoba, Spain; bb1dopeg@uco.es

<sup>5</sup> ETSIA (University of Seville), Ctra de Utrera km1, 41013 Seville, Spain; isolis@agrovegetal.es

<sup>6</sup> International Maize and Wheat Improvement Center (CIMMYT), C.A.P. Plaza Galerías, Col. Verónica Anzures, Ciudad de Mexico 11305, Mexico; k.ammar@cgiar.org

\* Correspondence: phernandez@ias.csic.es

Received: 17 December 2019; Accepted: 14 January 2020; Published: 19 January 2020

**Abstract:** Final grain production and quality in durum wheat are affected by biotic and abiotic stresses. The association mapping (AM) approach is useful for dissecting the genetic control of quantitative traits, with the aim of increasing final wheat production under stress conditions. In this study, we used AM analyses to detect quantitative trait loci (QTL) underlying agronomic and quality traits in a collection of 294 elite durum wheat lines from CIMMYT (International Maize and Wheat Improvement Center), grown under different water regimes over four growing seasons. Thirty-seven significant marker-trait associations (MTAs) were detected for sedimentation volume (SV) and thousand kernel weight (TKW), located on chromosomes 1B and 2A, respectively. The QTL loci found were then confirmed with several AM analyses, which revealed 12 sedimentation index (SDS) MTAs and two additional loci for SV (4A) and yellow rust (1B). A candidate gene analysis of the identified genomic regions detected a cluster of 25 genes encoding blue copper proteins in chromosome 1B, with homoeologs in the two durum wheat subgenomes, and an ubiquinone biosynthesis O-methyltransferase gene. On chromosome 2A, several genes related to photosynthetic processes and metabolic pathways were found in proximity to the markers associated with TKW. These results are of potential use for subsequent application in marker-assisted durum wheat-breeding programs.

**Keywords:** durum wheat; genome wide association study; GWAS water use; agronomic traits; MTAs; candidate genes; TKW; sedimentation volume; SDS; YR

## 1. Introduction

Wheat is one of the most widely grown crops worldwide (FAO, 2015), and is essential for the human diet [1]. Its importance and worldwide dominance are due, in part, to its agronomic adaptability. Durum wheat (*Triticum durum*) is a tetraploid wheat species (AABB genomes) mainly grown in the Mediterranean basin, in the Northern Plains (between the USA and Canada), in the arid areas of South Western USA and in Northern Mexico [2]. Durum wheat is well-adapted to a broad range of climatic conditions (including dry environments) and marginal soils, and has low water requirements [3,4]. Climatic conditions, as temperature and water availability, together with biotic stresses, can strongly affect durum wheat development and production [3–6]. Crop adaptation

is a central objective for breeding progress, driving improvement in final production, quantity and quality [7,8]. For over two decades, CIMMYT (International Maize and Wheat Improvement Center) has had an intensive breeding and improvement programme focused on the acceleration of durum productivity in developing countries.

Grain quality is an important breeding aim determining product end-use linked to financial returns. It is influenced by both genetic and environmental conditions [9], and biotic and abiotic stresses during growth and at key development stages [10]. Temperature, water availability and soil properties, especially nitrogen content, influence the final quality and protein content of wheat and its end-products [11–13].

There is a growing need to increase wheat yield without losing grain quality [14,15]. Key end-use grain quality traits include grain protein content (GPC), gluten strength, kernel size and vitreousness [7,16] and are all influenced by climatic conditions [17]. A number of agronomic components influence final productivity, including phenology (maturity) and plant architecture (plant height and lodging resistance). The majority of important agronomic traits, including yield, are controlled or influenced by multiple genes and are quantitatively inherited [18]. In addition, most are influenced by the environment and interactions between environmental and genetic (G×E) effects [19–23]. One of the most common methods currently used for dissection of quantitative agronomic and quality traits is the association mapping (AM) approach [24].

AM, originating in human genetics, was initially combined with linkage disequilibrium (LD) to identify the role of genes and linked markers for the determination of disease loci [25]. It is now widely used in plant and crop genetics. Some of the first studies based on LD mapping applied in plants were done in maize [26], rice [27] and oat [28]. AM has the main objective of determining, based on LD, correlations between genotypes and phenotypes in a panel of selected individuals [29]. It can support the development of new genetic markers for use in marker-assisted plant breeding [30]. It also facilitates the analysis of genetic variation underlying traits for further characterisation of the loci of interest [31].

Single nucleotide polymorphism (SNP) markers are commonly used in quantitative trait loci (QTL) mapping experiments [32,33] and genome-wide association studies (GWAS) for the detection of marker-trait associations (MTAs) in wheat [34–38]. DArTseq, a variant of the microarray-based DArT technology, has also been widely used in QTL mapping [39–41]. It reduces the complexity of the genome, using combinations of restriction enzymes [42] and next-generation sequencing. Several studies have assessed MTAs in durum wheat. The analyzed traits include grain yield, yield and yield components [6,37,43,44], heading date [6], and grain quality traits (thousand kernel weight, vitreousness, protein content, sedimentation index [17,45–47], yellow colour [48,49]).

In this study, three panels of elite durum wheat lines from CIMMYT were assessed in field trials conducted over multiple seasons and with differing water regimes. The AM approach was used to detect SNP and DArT markers associated with heterogeneous agronomic and quality trait data in order to test the approach as a tool for marker discovery within a live and ongoing breeding programme.

## **2. Material and Methods**

### *2.1. Plant Material, Phenotyping and Genotyping*

Panels of elite durum lines from CIMMYT wheat preliminary yield trials (PYT), comprising a total of 294 accessions (Supplementary Materials Table S1) were used for agronomic and quality assessment. PYT trials consisted of the best advanced breeding lines which were promoted to unreplicated trials, including one or two repeated checks. The trials were sown, assessed and analysed according to their specific statistical designs [50] and consisted of two blocks with different water treatments, one with full irrigation (FI) and the other with reduced irrigation (RI). In the FI treatment four to five irrigations were applied during the field season to maintain the optimal soil moisture conditions, whilst in the RI block a single irrigation was applied at planting, in order to ensure establishment. In both water treatments the irrigation was applied by gravity in furrows. The rainfall data ([https://www.meteoblue.com/en/weather/historyclimate/weatherarchive/ciudad-obreg%c3%b3n\\_mexico\\_4013704](https://www.meteoblue.com/en/weather/historyclimate/weatherarchive/ciudad-obreg%c3%b3n_mexico_4013704)) for the four

field seasons is included in Supplementary Materials Figure S1. The agronomic and quality assessment of the panels over seasons is summarised in Table 1.

**Table 1.** Agronomic and quality assessment of wheat field trials. The number of lines, year, location and water regime applied is shown.

Wheat Panel	No. of Assessed Lines	Field Season			
		2012	2013	2014	2015
1	98	YAQ-FI			
2	97		YAQ-FI	YAQ-FI YAQ-RI	
3	98		YAQ-FI	YAQ-FI YAQ-RI	YAQ-FI YAQ-RI

YAQ: Yaqui (Mexico); FI: Full irrigation; RI: Reduced irrigation.

Field experiments were conducted at CIMMYT's experimental station in the Yaqui Valley, Mexico (27.282848° N; 109.923878° W) over four field seasons (2012 to 2015 harvest years, inclusive). Wheat panels 2 and 3 were phenotypically assessed across years, while panel 1 was only grown in 2012. The experimental plots (1.6 × 3 m) were sown in November/December of each year and harvested in May of the following year. Data for yellow rust were assessed in semi-controlled conditions at CIMMYT's experimental station in Toluca (Mexico).

Plant material for genetic analysis was harvested for each line at the 4th leaf stage (growth stage 14 on the Zadoks scale [51]) and immediately frozen in dry-ice. Samples were stored at −80 °C until DNA extraction. Approximately 100 mg of frozen tissue was used for DNA isolation with a DNeasy Plant Mini Kit from Qiagen, following the manufacturer's protocol. DNA sample quality and concentration were assessed using electrophoresis on a 0.8% agarose gel and the restriction enzyme Tru1I (ThermoFisher) was used to check for the absence of nucleases in DNA prior to genotyping.

Samples were genotyped by Diversity Arrays Technology Pty Ltd. (Montana St, University of Canberra, Bruce ACT 2617, Australia) (DArT) using DartSeq™. A total of 35,509 polymorphic dominant DArT markers and 9142 biallelic SNP markers were generated. Both datasets were thinned by removing one marker from each pair with a correlation coefficient of >0.95. The final dataset consisted on 14,588 DArT markers (of which 8411 were mapped) and 5716 SNP markers (4142 mapped markers). DartSeq™ genotyping and mapping of the corresponding markers to the wheat reference genome sequence RefSeq v1 from the International Wheat Genome Sequencing Consortium (IWGSC, <http://www.wheatgenome.org/>) was performed by DArT (diversityarrays.com), as described by Sukumaran et al. [43]. The distribution of markers across the A and B subgenomes is given in Table 2.

**Table 2.** Molecular markers distribution across the wheat genome. The distribution of DArT (Diversity Arrays Technology, left) and single nucleotide polymorphism (SNP, right) markers across the durum wheat A and B subgenomes and the number of unmapped markers are shown.

Chr	DArT Markers				SNP Markers			
	A	B	Un	Total <sup>1</sup>	A	B	Un	Total <sup>1</sup>
1	377	866	37	1280	193	444	16	653
2	563	791	69	1423	320	375	19	714
3	496	818	33	1347	250	352	11	613
4	585	325	10	920	283	171	4	458
5	312	725	13	1050	162	376	2	540
6	449	690	21	1160	262	308	5	575
7	623	573	35	1231	293	287	9	589
<b>Total<sup>2</sup></b>	3405	4788	218	8411	1763	2313	66	4142
<b>Un</b>				6177				1574
<b>Total</b>				14,588				5716

Chr: chromosome; A: wheat A subgenome; B: wheat B subgenome; Un: unmapped; <sup>1</sup>: total markers by group; <sup>2</sup>: total markers by genome.

## 2.2. Phenotypic Data

Ten agronomic and quality traits were assessed for three durum wheat elite line panels: days to heading (days, DTH); plant height (cm, PH); lodging (%; LOD); yellow rust (%; YR); yellow colour (YC); sedimentation index (cm<sup>3</sup>, SDS); sedimentation volume (cm<sup>3</sup>, SV); grain protein content (%; GPC); thousand kernel weight (g, TKW); and grain yield (Kg/ha, GY). Agronomic traits (DTH, PH, LOD, YR, TKW and GY) were assessed under both water treatments (FI and RI), and quality traits (YC, SV, SDS and GPC) were only evaluated under FI conditions. Visual disease evaluation and phenology assessments were made in the field, while quality parameters were evaluated on grain samples post-harvest. DTH, PH and LOD and YR were visually assessed at the field trials in Yaqui, while YR was assessed at Toluca. To assess DTH, heading date was recorded as the time when 50% of the spikes have emerged from the flag leaf sheath (stage 59 in Zadoks scale [51]); PH was recorded by measuring the distance between the stem's base and the top of the spike (awns not included); LOD was assessed as the percentage of lodging plot; and YR was assessed as the percentage of leaves with rust pustules. YC was assessed by a rapid colorimetric method with a Minolta color meter following CEN/TS 15,465:2008 [52–54]; SDS was evaluated following UNE 34,903:2014 [55,56]; SV and GPC were assessed by Near-infrared spectroscopy (NIRS) [57]; TKW was measured by weighing 2 samples of 100 entire kernels randomly chosen previously dried at 70 °C for 48 h.

The correlation between the assessed traits was analysed using the 'cor' function in R [58–60]. Then, an analysis of variance (ANOVA) was undertaken, using the 'aov' function in R [61], to obtain the descriptive statistics for each trait.

Traits were analysed using a Q + K linear mixed-model [62,63] which follows the model equation:

$$y = X\beta + S\alpha + Qv + Z\mu + \varepsilon \quad (1)$$

where  $y$  is a vector of observed phenotypes;  $X$ ,  $S$  and  $Z$  are matrices related to  $\beta$ ,  $\alpha$  and  $\mu$ , respectively;  $\beta$  is a vector of fixed effects;  $\alpha$  is a vector of marker effects;  $Qv$  is a vector of population effect;  $\mu$  is a vector of polygenic effects (with covariance proportional to a kinship or relationship matrix); and  $\varepsilon$  is a vector of residuals.

These analyses were carried out using GenStat (14th Edition) to generate the best linear unbiased estimates (BLUEs) of variety performance in different ways: (i) across years and blocks; (ii) across years for each block (FI and RI); (iii) across a reduced dataset (years 2013 and 2014) and blocks; and (iv) across the reduced dataset for each block. The resulting datasets (available in Supplementary Material Table S2) were then used in different association mapping analyses.

## 2.3. Population Structure and Linkage Disequilibrium

Population structure was assessed using principal component analysis (PCA) based on the combined DArT and SNP genotyping datasets. Euclidean distances were calculated using the R package 'ggfortify' [64] and the PCA was visualised with 'ggplot2' [65].

The pattern of linkage disequilibrium (LD) was assessed between each pair of SNP markers on the same chromosome across the two constitutive genomes with the allele frequency correlation ( $r^2$ ) using the 'popgen' package in R [66]. A heatmap was obtained with the  $D'$  and  $r^2$  values for each chromosome and a scatterplot to determine LD decay (genetic distance in cM).

## 2.4. Association Mapping (AM)

The AM analyses were performed on the BLUEs obtained above using an additive model with 'rrBLUP' [67] and 'GWASpoly' [68] packages in R in different ways. Two marker-based kinship matrices (k-matrix), created from a subset of 14,588 DArT and 5716 SNP markers, respectively, were used for the adjustment based on relatedness of individuals (Supplementary Materials Tables S3 and S4). A minor allele frequency (maf) threshold of 0.05 was used. To establish a  $p$ -value detection threshold

for statistical significance of associations, the Bonferroni correction, which employs a threshold of  $\alpha/m$  to ensure the genome-wide type error I of 0.05, was applied with a total of 1000 permutations.

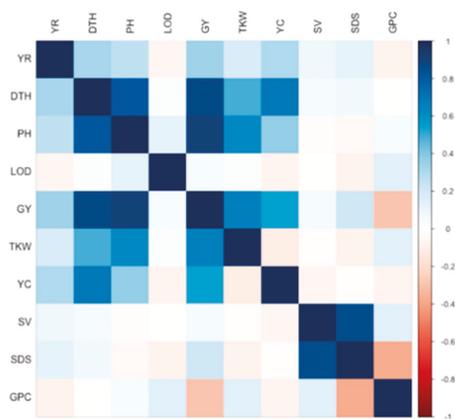
Associated DartSeq<sup>TM</sup> and SNP markers were blasted against the wheat reference assembly RefSeqv1 [69] with no indels or mismatches allowed, using an ad hoc Java program, to confirm their physical mapping location on the A or B genomes. The molecular markers were also mapped against the durum wheat genome (<https://www.interomics.eu/durum-wheat-genome>) to confirm their physical positions. In addition, to identify candidate genes, results were filtered, selecting the hits with best *e*-value for each molecular marker, and candidate genes were manually selected based on their annotations.

### 3. Results

#### 3.1. Phenotypic Assessment

Results from the ANOVA for all the traits across years and water treatments are shown in Table 3. The mean phenotypic values across years were calculated for each block and panel to evaluate the influence of water conditions on the assessed traits (Table 3). Days to heading during the field seasons assessed ranged from 63 to 94 days. In plots with lower water availability (RI block), the spike emergence from the flag leaf took place approximately 11 days earlier than in FI plots. Plant height ranged from 39 to 110 cm showing differences between water regimes, with a decrease of 25–30 cm under RI conditions. Likewise, and as result of the RI treatment, GY (ranging from 1.35 to 10.63 ton/ha) and TKW (from 29.6 to 63.2 g) also varied, being reduced by 4–5 tons/ha and 7–10 g, respectively, in the low water availability RI treatment. This strong RI treatment resulted in very low heritability values for DTH, PH and LOD.

Several significant phenotypic correlations were observed between the analysed traits (Figure 1 and Supplementary Materials Table S5). The most correlated traits were PH and GY ( $r = 0.90$ ,  $p$ -value =  $<2.2 \times 10^{-16}$ ), followed by DTH and GY ( $r = 0.87$ ,  $p$ -value =  $<2.2 \times 10^{-16}$ ), SDS and SV ( $r = 0.85$ ,  $p$ -value =  $<2.2 \times 10^{-16}$ ) and also DTH and PH ( $r = 0.82$ ,  $p$ -value =  $<2.2 \times 10^{-16}$ ). Other traits showed important positive correlations too, including YC and DTH ( $r = 0.69$ ,  $p$ -value =  $2.82 \times 10^{-09}$ ), GY and TKW ( $r = 0.66$ ,  $p$ -value =  $<2.2 \times 10^{-16}$ ), PH and TKW ( $r = 0.62$ ,  $p$ -value =  $<2.2 \times 10^{-16}$ ), GY and YC ( $r = 0.53$ ,  $p$ -value = 0.039) and DTH and TKW ( $r = 0.49$ ,  $p$ -value =  $<2.2 \times 10^{-16}$ ). Negative correlations were also observed for SDS and GPC ( $r = -0.38$ ,  $p$ -value =  $<2.2 \times 10^{-16}$ ), and for GPC and YC ( $r = -0.08$ ,  $p$ -value =  $<2.2 \times 10^{-16}$ ) (Figure 1).



**Figure 1.** Phenotypic correlations. YR: yellow rust; DTH: days to heading; PH: plant height; GY: grain yield; TKW: thousand kernel weight; YC: yellow colour; SV: sedimentation volume; SDS: sedimentation index; and GPC: grain protein content.

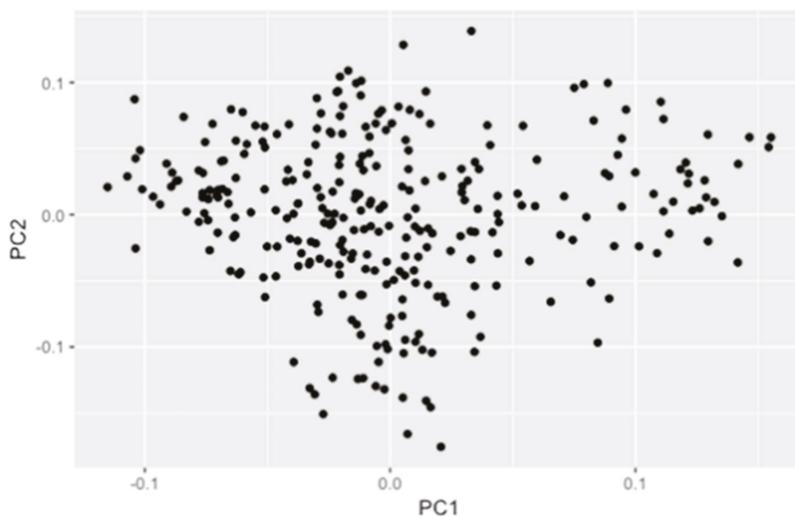
Table 3. Summary of overall phenotypic data including analysis of variance (ANOVA) results and broad sense heritability ( $h^2$ ) for agronomic and quality traits.

Panel	ANOVA												
	Min <sup>1</sup>	Mean <sup>2</sup>	Max <sup>3</sup>	Df	Sum Sq	Mean Sq	F Value	Pr (>F)	Significance	$h^2$	Mean Across Years		
											1	2	3
											FI	RI	FI
Water regimes													
DTH (days)	63	79.38	94	1	1.8	1.768	0.24	0.626	***	0.14	87.55	83.73	72.54
PH (cm)	39	81.97	110	1	167.5	167.51	8.025	$6.59 \times 10^{-03}$	**	0.16	92.98	93.31	62.58
GY (ton/ha)	1.35	5.68	10.63	1	24.52	24.516	121.6	$4.16 \times 10^{-15}$	***	0.44	10	7	2.03
TKW (g)	29.6	44.69	63.2	1	38.6	38.65	2.064	$1.57 \times 10^{-01}$	*	0.41	49.25	47.58	37.29
YR	0	2.286	40	1	338.5	338.5	6.702	$1.25 \times 10^{-02}$	*	0.00	4.77	3.98	0
LOD (%)	0	2.142	90	1	0	0	-	-	-	0.01	0.1	0	9.25
YC	14.6	16.59	20	1	0.008	0.0085	0.02	$8.89 \times 10^{-01}$	-	0.59	17.21	17.09	-
SV (ml)	7	10.51	14.5	1	2.54	2.536	2.388	0.128	-	0.57	10.87	10.14	-
SDS	0.54	0.854	1.19	1	0.0927	0.09268	14.86	$3.25 \times 10^{-04}$	***	0.55	1	1	-
GPC (%)	10.4	12.33	14.9	1	5.965	5.965	16.41	$1.74 \times 10^{-04}$	***	0.31	11.67	12.38	-

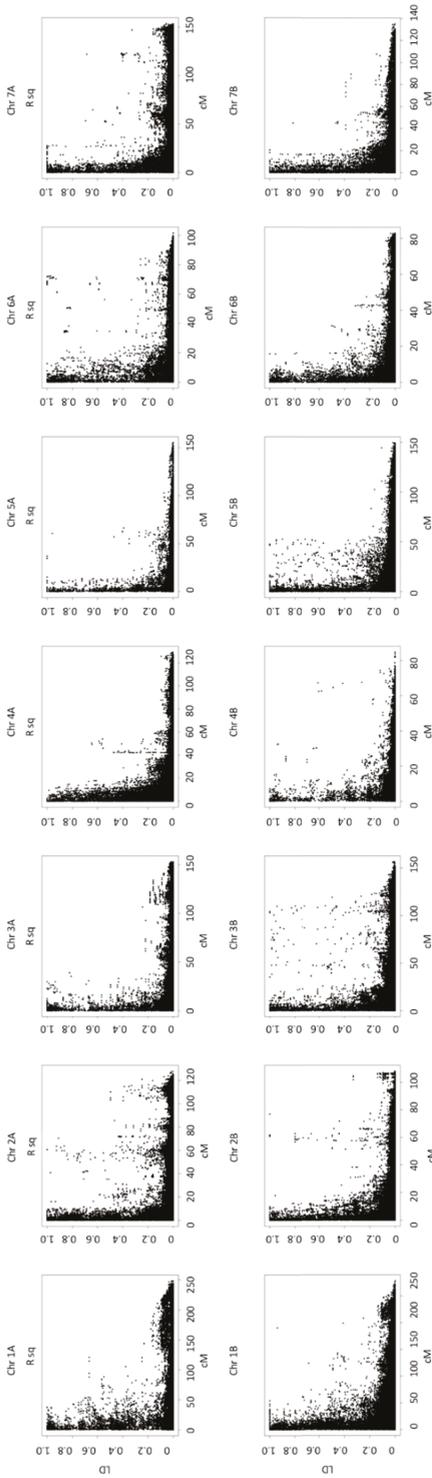
DTH: days to heading; PH: plant height; LOD: lodging; GY: grain yield; TKW: thousand kernel weight; YR: yellow rust; LOD: lodging; YC: yellow color; SV: sedimentation volume; SDS: sedimentation index; GPC: grain protein content; FI: full irrigation block; RI: reduced irrigation block; <sup>1</sup>: minimum value across years and water regimes; <sup>2</sup>: mean across years and water regimes; <sup>3</sup>: maximum value across years and water regimes; Df: degrees of freedom; Sq: sum square; F value: measure of significance in the F-test; Pr (>F): p-value associated with the F statistic; Significance levels: 0 = \*\*\*\*, 0.001 = \*\*\*, 0.01 = \*\*, 0.1 = \*; Df = degrees of freedom.

### 3.2. Population Structure and Linkage Disequilibrium

The PCA used a total of 14,588 DArT and 5716 SNP markers. The first and second principal components explained 3.91% of the genetic variation (Figure 2). No underlying genetic structure was detected within or between the panels assessed. LD was estimated using the mapped SNP markers dataset. LD decay was determined within 20–30 cM for all the chromosomes (Figure 3). Using the classification defined by Maccaferri et al. [70], the markers presented loose linkage (Class 2), showing a distance value between 21 to 50 cM.



**Figure 2.** Principal components analysis (PCA) of the genotypic data.



**Figure 3.** Linkage disequilibrium (LD) decay analysis using SNP data. Estimates  $r^2$  versus linkage distance on chromosome in centimorgan (cM) was represented. The LD decay was established between 20–30 cM.

### 3.3. AM Analysis

Thirty-seven significant marker-trait associations (MTAs) were detected for TKW and SV across all years and water treatments with most of the significant markers located on chromosome 2A (Table 4). Twenty DArT and seven SNP markers were found in association with TKW on chromosome 2A (with additive effects ranging from  $-3.41$  to  $3.46$ ). In addition, eight unmapped DArT and one SNP marker were also associated with TKW (additive effects ranged from  $-3.39$  to  $3.46$  g). Most of these MTAs showed a negative additive effect, reducing the final weight value (ranging from  $-2.84$  to  $-3.19$  g), and only two MTAs were found to increase TKW (values of  $2.97$  and  $3.09$  g). Finally, a single SNP associated with SV was located on chromosome 1B (showing a positive effect increasing the final value by  $1.26$  mL). The resulting Manhattan and QQ-plots from this AM analysis are included in Supplementary Materials Figures S2–S5.

The AM analyses on partitioned subsets of the data consistently detected the QTLs for TKW and SV. Nevertheless, the individual assessment of the water treatments significantly reduced the number of MTAs found, due in part to less available data for the RI block (Supplementary Material Table S6). The initial dataset of 294 durum wheat elite lines was reduced to 200 lines (assessed during the 2013 and 2014 seasons) to give a dataset balanced across assessment years. Using this reduced dataset for AM analysis, the results confirmed the QTLs previously found for the full dataset (Supplementary Materials Table S6). The analysis also detected an additional locus for SV on chromosome 4A, and a locus for YR on chromosome 1B (with additive effects of  $-0.84$  and  $2.79$ , respectively).

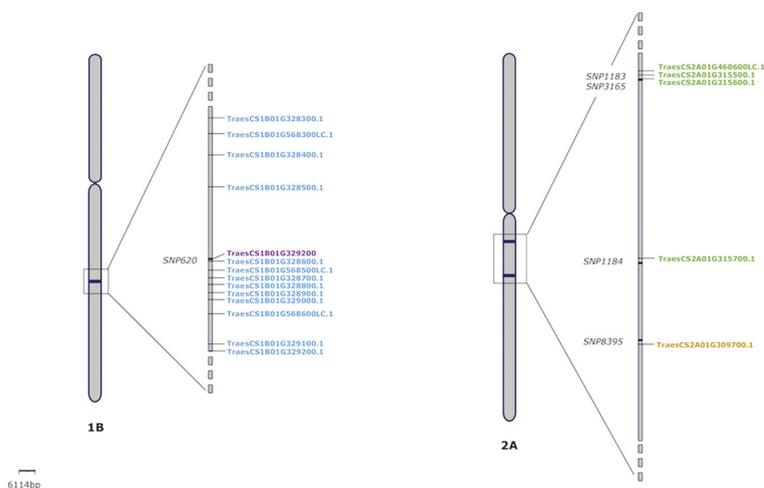


### 3.4. Candidate Genes Analysis

The marker *SNP620*, located on chromosome 1B and detected in association with SV, was found included into a cluster of 12 genes encoding blue copper proteins (BCP), with homoeologs in the two durum wheat subgenomes (Figure 4, Table 5 and Supplementary Material Figure S6). In the hexaploid wheat genome, this set of genes form a cluster of homeolog triads [72] with a total of 31 genes (Supplementary Materials Table S7 and Figure S7). Additionally, another interesting gene (*TraesCS1B01G568400LC.1*) was found closer this marker, coding for the ubiquinone biosynthesis *O*-methyltransferase.

There were several markers located in chromosome 2A, in close proximity to some interesting genes. Markers *SNP1183*, *SNP1184* and *DART3165* were found next to several genes encoding reductase-1 (Figure 4 and Table 5). In addition, the marker *SNP8395*, also located on the same chromosome, was found in proximity to the gene *TraesCS2A01G309700.1*, which encodes a type A response regulator 1 (Figure 4 and Table 5).

Significant MTAs from the partitioned analysis also allowed the identification of potentially interesting genes. On chromosome 1B, marker *DART1744*, previously described by Mérida-García et al. [73] related to high molecular-weight glutenin subunits, was found in proximity to genes encoding isocitrate dehydrogenase kinase/phosphatase G and leucine-rich repeat receptor-like protein kinase family protein. These genes participate in the carbohydrate metabolism during the Krebs cycle and play a crucial role in plant development and stress responses, respectively [74,75]. On this chromosome, another marker (*SNP809*) was found in proximity to some interesting genes encoding sugar transporter proteins. Additionally, some markers located on chromosome 2A were found in proximity to Acyl-CoA *N*-acyltransferase genes (*SNP1206*, *SNP8395* and *DART3180*) and chloroplastic zeaxanthin epoxidase (*SNP1189*) (Supplementary Materials Table S8).



**Figure 4.** Location of candidate genes on chromosomes 1B and 2A. Markers are indicated with the symbol '-'; blue copper proteins are shown in blue colour; ubiquinone biosynthesis *O*-methyltransferase in purple colour; the regulator response gene in brown colour; and for reductase 1 genes in green colour.

**Table 5.** Candidate genes for markers with significant marker-traits associations. Genes located in proximity of markers found in association with TKW and SV across years and water treatments (within a ±50 kb window). Values for physical position and distance are indicated in base pairs (bp). Chr: chromosome position. Blue copper proteins are shown in blue colour; ubiquinone biosynthesis O-methyltransferase in purple colour; the regulator response gene in brown colour; and for reductase 1 genes in green colour. Physical positions and gene annotations are based on the wheat reference assembly RefSeqv1 [69].

Marker	Identity	Transcript	Chr	Physical Position	Distance	Description
DA713156	88.525	TraesCS1B01G193400L.C.1	1B	134,604,832	-33,988	LINE-1 reverse transcriptase-like protein
		TraesCS1B01G193500L.C.1	1B	134,605,786	-33,034	Retrotransposon protein, putative, unclassified
		TraesCS1B01G193600L.C.1	1B	134,607,645	-31,175	Retrotransposon protein, putative, unclassified
		TraesCS1B01G193700L.C.1	1B	134,621,105	-17,715	Retrotransposon protein, putative, unclassified
		TraesCS1B01G193800L.C.1	1B	134,622,150	-16,670	Solute carrier organic anion transporter family member 2B1
		TraesCS1B01G193900L.C.1	1B	134,622,613	-16,207	Tetratricopeptide repeat (TPR)-like superfamily protein
		TraesCS1B01G194000L.C.1	1B	134,632,362	-6458	RNA-directed DNA polymerase (reverse transcriptase)-related family protein
		TraesCS1B01G194100L.C.1	1B	134,633,505	-5315	LINE-1 reverse transcriptase like
		TraesCS1B01G194200L.C.1	1B	134,639,730	910	Transposon Ty3-G Gag-Pol polyprotein
		TraesCS1B01G114900.1	1B	134,645,780	6960	F-box protein
		TraesCS1B01G194300L.C.1	1B	134,648,368	9548	Sister chromatid cohesion protein PD55 homolog B-B
		TraesCS1B01G194400L.C.1	1B	134,649,000	10,180	BTB/POZ domain containing protein, expressed
DA721834	100	TraesCS1B01G231000L.C.1	1B	176,756,965	-3171	Retrotransposon protein, putative, LINE subclass
		TraesCS1B01G137700.1	1B	176,783,830	23,694	Phototropic-responsive NPH3 family protein
		TraesCS1B01G137700.2	1B	176,783,926	23,790	Phototropic-responsive NPH3 family protein
		TraesCS1B01G231100L.C.1	1B	176,788,566	28,430	Disease resistance protein (TIR-NBS-LRR class) family
		TraesCS1B01G231200L.C.1	1B	176,790,108	29,972	Transposon protein, putative, Mutator sub-class
		TraesCS1B01G231300L.C.1	1B	176,790,698	30,562	Transposon protein, putative, mutator sub-class
		TraesCS1B01G231400L.C.1	1B	176,791,607	31,471	Sterile alpha motif (SAM) domain-containing protein
		TraesCS1B01G137800.1	1B	176,793,792	33,656	GRF zinc finger family protein, expressed
		TraesCS1B01G231500L.C.1	1B	176,803,822	43,686	Retrotransposon protein, putative, unclassified
		TraesCS1B01G568300L.C.1	1B	555,010,255	-46,132	Blue copper protein
		TraesCS1B01G328400.1	1B	555,018,059	-38,328	Blue copper protein
		TraesCS1B01G328500.1	1B	555,029,816	-26,571	Blue copper protein
SNP620	100	TraesCS1B01G568400L.C.1	1B	555,056,445	58	Ubiquinone biosynthesis O-methyltransferase
		TraesCS1B01G328600.1	1B	555,057,325	938	Blue copper protein
		TraesCS1B01G568500L.C.1	1B	555,060,627	4240	Blue copper protein
		TraesCS1B01G328700.1	1B	555,063,537	7150	Blue copper protein
		TraesCS1B01G328800.1	1B	555,065,987	9600	Blue copper protein
		TraesCS1B01G328900.1	1B	555,068,889	12,502	Blue copper protein
		TraesCS1B01G329000.1	1B	555,071,391	15,004	Blue copper protein
		TraesCS1B01G568600L.C.1	1B	555,076,734	20,347	Blue copper protein
		TraesCS1B01G329100.1	1B	555,088,083	31,696	Blue copper protein
		TraesCS1B01G329200.1	1B	555,090,695	34,308	Blue copper protein
		TraesCS1B01G568700L.C.1	1B	555,095,491	39,104	purple acid phosphatase 23
		TraesCS1B01G568800L.C.1	1B	555,096,179	39,792	Disease resistance protein (TIR-NBS-LRR class) family
	TraesCS1B01G568900L.C.1	1B	555,097,247	40,860	Retrotransposon protein, putative, Ty3-typsy subclass	
	TraesCS1B01G569000L.C.1	1B	555,098,272	41,885	Retrotransposon protein, putative, Ty3-typsy subclass	
	TraesCS1B01G569100L.C.1	1B	555,101,471	45,084	50S-ribosomal protein L2	
	TraesCS1B01G569200L.C.1	1B	555,103,053	46,666	LINE-1 reverse transcriptase like	

**Table 5.** *Contd.*

Marker	Identity	Transcript	Chr	Physical Position	Distance	Description
DA7T3155	95,939	TraesCS2A01G213400L.C.1	2A	174,011,203	-24,981	Retrotransposon protein, putative, unclassified, expressed
		TraesCS2A01G213500L.C.1	2A	174,024,497	-11,687	APOLLO
		TraesCS2A01G201000L.C.1	2A	174,026,194	-9990	Cytochrome P450-like
		TraesCS2A01G213600L.C.1	2A	174,034,039	-2145	Retrotransposon protein, putative, unclassified, expressed
SNP1153	98,551	TraesCS2A01G213700L.C.1	2A	174,039,127	2943	Cytochrome P450
		TraesCS2A01G493000L.C.1	2A	582,628,952	-7722	Retrotransposon protein, putative, Ty3-gypsy subclass
		TraesCS2A01G493100L.C.1	2A	582,629,900	-6774	Retrovirus-related Pol polyprotein from transposon gypsy
		TraesCS2A01G493200L.C.1	2A	582,630,406	-6268	Retrotransposon protein, putative, unclassified
		TraesCS2A01G493300L.C.1	2A	582,630,979	-5695	Retrotransposon protein, putative, Ty3-gypsy subclass
		TraesCS2A01G344800L.C.1	2A	582,634,003	-2671	RAN guanine nucleotide release factor
		TraesCS2A01G344900L.C.1	2A	582,637,903	1229	Nucleosome assembly protein 1-like 1
		TraesCS2A01G457700L.C.1	2A	536,796,624	-29,094	Retrovirus-related Pol polyprotein LINE-1
DA7T3146	100	TraesCS2A01G333200L.C.1	2A	566,207,430	-659	Kinesin-like protein
		TraesCS2A01G333200L.C.2	2A	566,209,291	1202	Kinesin-like protein
DA7T3150	100	TraesCS2A01G460800L.C.1	2A	541,301,644	-464	1-phosphatidylinositol-3-phosphate 5-kinase FAB1A
DA7T3162	100	TraesCS2A01G457200L.C.1	2A	535,217,367	-18,487	Acetylglutamate kinase-like protein
		TraesCS2A01G457300L.C.1	2A	535,221,879	-13,975	LINE-1 reverse transcriptase like
		TraesCS2A01G457400L.C.1	2A	535,222,237	-13,617	LINE-1 reverse transcriptase
		TraesCS2A01G311500L.C.1	2A	535,240,748	4894	NAC domain protein
SNP1183	100	TraesCS2A01G460600L.C.1	2A	541,195,856	-5055	Reductase 1
		TraesCS2A01G315500L.C.1	2A	541,197,465	-3446	Reductase 1
		TraesCS2A01G460700L.C.1	2A	541,198,542	-2369	NADH dehydrogenase [ubiquinone] iron-sulfur protein 3, mitochondrial
		TraesCS2A01G315600L.C.1	2A	541,200,447	-464	Reductase 1
SNP1184	100	TraesCS2A01G460900L.C.1	2A	541,386,879	-4975	Serine-type endopeptidase inhibitor, putative
		TraesCS2A01G461000L.C.1	2A	541,387,542	-4312	Aldose reductase
		TraesCS2A01G315700L.C.1	2A	541,391,385	-469	Reductase 1
		TraesCS2A01G460600L.C.1	2A	541,195,856	-5055	Reductase 1
DA7T3165	98,246	TraesCS2A01G315500L.C.1	2A	541,197,465	-3446	Reductase 1
		TraesCS2A01G460700L.C.1	2A	541,198,542	-2369	NADH dehydrogenase [ubiquinone] iron-sulfur protein 3, mitochondrial
		TraesCS2A01G315600L.C.1	2A	541,200,447	-464	Reductase 1
		TraesCS2A01G460900L.C.1	2A	541,386,879	-4972	Serine-type endopeptidase inhibitor, putative
		TraesCS2A01G461000L.C.1	2A	541,387,542	-4309	Aldose reductase
		TraesCS2A01G315700L.C.1	2A	541,391,385	-466	Reductase 1
		TraesCS2A01G316900L.C.1	2A	542,642,276	-44,928	Phosphate carrier protein, mitochondrial
		TraesCS2A01G317000L.C.1	2A	542,648,355	-38,849	Zeaxanthin epoxidase, chloroplastic
SNP1189	100	TraesCS2A01G317000L.C.2	2A	542,649,684	-37,520	Zeaxanthin epoxidase, chloroplastic
		TraesCS2A01G317000L.C.3	2A	542,650,147	-37,057	Zeaxanthin epoxidase, chloroplastic
		TraesCS2A01G462000L.C.1	2A	542,652,662	-34,542	AUGMIN subunit 6
		TraesCS2A01G317100L.C.1	2A	542,654,847	-32,357	Mitochondrial carrier protein
		TraesCS2A01G317200L.C.1	2A	542,658,154	-29,050	Phosphatase 2C family protein
		TraesCS2A01G317300L.C.1	2A	542,686,655	-549	transmembrane protein

Table 5. Contd.

Marker	Identity	Transcript	Chr	Physical Position	Distance	Description
DArT3172	100	TraesCS2A01G333900.1	2A	567,725,771	-8576	RNA-dependent RNA polymerase MLP protein
		TraesCS2A01G334000.1	2A	567,735,196	849	
DArT3174	100	TraesCS2A01G333300.1	2A	566,454,172	-2950	F-box/RN1-like superfamily protein Transposon Ty3-G Gag-Pol polyprotein Craniofacial development protein 2 Retrotransposon protein, putative, unclassified
		TraesCS2A01G481800L.C.1	2A	566,461,411	4289	
		TraesCS2A01G481900L.C.1	2A	566,462,897	5775	
		TraesCS2A01G482000L.C.1	2A	566,478,234	21,112	
DArT3175	100	TraesCS2A01G319300.1	2A	544,359,272	-32,496	target of AVR8 operation1 Retrotransposon protein, putative, unclassified
		TraesCS2A01G464000L.C.1	2A	544,395,881	4113	
DArT3176	98.182	TraesCS2A01G464500L.C.1	2A	546,477,078	31,281	Transposon Ty3-I Gag-Pol polyprotein Transposon Ty3-I Gag-Pol polyprotein Transposon Ty3-I Gag-Pol polyprotein
		TraesCS2A01G464600L.C.1	2A	546,478,437	32,640	
		TraesCS2A01G464700L.C.1	2A	546,479,163	33,366	
DArT3180	100	TraesCS2A01G333900.1	2A	567,725,771	-10,352	RNA-dependent RNA polymerase MLP protein
		TraesCS2A01G334000.1	2A	567,735,196	-927	
DArT3182	100	TraesCS2A01G483800L.C.1	2A	569,357,751	-46,773	autoinhibited Ca(2+) $\gamma$ -ATPase, isoform 4 Zinc finger family protein
		TraesCS2A01G335600.1	2A	569,363,189	-41,335	
DArT3187	98.551	TraesCS2A01G460800L.C.1	2A	541,301,644	-458	1-phosphatidylinositol-3-phosphate 5-kinase FABIA
DArT3198	98.305	TraesCS2A01G457200L.C.1	2A	535,217,367	-18,493	Acetylglutamate kinase-like protein LINE-1 reverse transcriptase like LINE-1 reverse transcriptase NAC domain protein
		TraesCS2A01G457300L.C.1	2A	535,221,879	-13,981	
		TraesCS2A01G457400L.C.1	2A	535,222,237	-13,623	
		TraesCS2A01G311500.1	2A	535,240,748	4888	
DArT3201	100	TraesCS2A01G483800L.C.1	2A	569,357,751	-46,711	autoinhibited Ca(2+) $\gamma$ -ATPase, isoform 4 Zinc finger family protein
		TraesCS2A01G335600.1	2A	569,363,189	-41,273	
DArT10906	98.551	TraesCS2A01G482500L.C.1	2A	566,976,225	12,025	RNA-directed DNA polymerase (Reverse transcriptase) Gibberellin-regulated protein 2
		TraesCS2A01G333600.1	2A	566,986,482	22,282	
SNP8395	96.296	TraesCS2A01G309400.1	2A	532,849,120	-4840	Pentatricopeptide repeat-containing protein Smr domain containing protein Acyl-CoA N-acyltransferase isoform 2 Acyl-CoA N-acyltransferase isoform 2 Response regulator
		TraesCS2A01G309500.1	2A	532,854,936	976	
		TraesCS2A01G309600.1	2A	532,859,077	5117	
		TraesCS2A01G309600.2	2A	532,859,077	5117	
		TraesCS2A01G309700.1	2A	532,865,797	11,837	
DArT20759	97.619	TraesCS2A01G460800L.C.1	2A	541,301,644	-573	1-phosphatidylinositol-3-phosphate 5-kinase FABIA
DArT20961	100	TraesCS2A01G308900.1	2A	532,040,483	-40,124	Translocase of chloroplast GTPase Der Protein NRT1/PTR FAMILY 1.1 Protein NRT1/PTR FAMILY 1.1
		TraesCS2A01G309000.1	2A	532,077,143	-3464	
		TraesCS2A01G309100.1	2A	532,085,578	4971	
		TraesCS2A01G309100.2	2A	532,085,909	5302	

#### 4. Discussion

The maintenance of crop production is a current and pressing need given growing populations and reduced availability of arable land [76]. There is an increasing need for breeding programs to improve yield potential and the adaptation of new varieties to different biotic and abiotic stresses [77]. Abiotic stresses, including drought and heat, are impacting productivity on the million hectares of wheat grown worldwide each year [78]. Detailed molecular and phenotypic characterization are valuable tools in the dissection of complex traits [79], and especially those that are influenced by water availability [14].

The improvement of key traits is essential for better end-use production quantity and quality in wheat [80]. In this study, we analysed a set of 10 agronomic and quality traits under full irrigation conditions (FI), with an additional six traits also assessed under low water availability (RI) in order to understand trait variation under contrasting water regimes in the CIMMYT durum wheat breeding programme. Irrigation conditions influenced some important yield and yield-related traits such as GY and TKW, as well as adaptive traits including DTH and PH (Table 3). The RI treatments had decreased final yields in line with previous observations [6,81]. Previous reports have also shown TKW to be reduced by high temperatures [17], most likely related to water availability. Groos et al. [82] assessed the genetic basis of grain yield and protein content in a segregating population of wheat RILs grown at six locations and also identified effects from GxE interactions involving protein content and yield. Our mean trait values corroborated this, with the highest values for GY recorded for FI blocks across panels (see Table 3). A similar trend was shown for DTH, PH and TKW, which decreased under low-water regimes.

Correlations between the assessed traits showed that GY was positively correlated with two different phenology traits (PH and DTH). This is in agreement with Maccaferri et al. [6], who showed important positive and negative correlations for GY and DTH, and also positive correlations for GY and PH in several environments with different water regimes. DTH and PH were also positively correlated (Supplementary Materials Table S5), with taller plants having a longer time period to the emergence of the tip of the spike stage.

Wheat TKW is an important yield component with a direct effect on grain yield [83,84]. In line with this, our results showed a significant and positive correlation between TKW and GY. However, the previously reported negative correlation between TKW and DTH [6] was not observed, potentially as result of temperatures and water availability from emergence to heading, and also from heading to harvest. Rharrabti et al. [17] previously highlighted a positive correlation between protein content and TKW, which is in agreement with the results obtained in the present study. They highlighted that warm temperatures during grain filling could negatively affect this correlation.

Significant associations between endosperm proteins (gliadin and glutenin subunits) and SV have been previously highlighted [85,86]. Here we found a positive correlation ( $r = 0.15$ ) between SV and GPC. This correlation is thought to be the result of grain protein content influencing the sedimentation volume value [87], which depends on the degree of protein hydration and oxidation [88]. Finally, for sedimentation index (SDS) analysis, we observed a negative correlation with protein content, in agreement with results presented by Rharrabti et al. [17,45]. This is also in agreement with Oelofse et al. [89] who highlighted the significant influence of protein content on the SDS sedimentation test [90–92].

The SNP and DArT markers used to analyze patterns of genetic structure (Figure 2) and LD (Figure 3) for the durum wheat lines revealed no detectable genetic structure and consistent patterns of LD across chromosomes (LD was estimated to extend ~20 cM). These results support the suitability of durum elite line sets currently in use in breeding programmes for the practical application of GWAs analysis. The rate of unmapped markers was lower for SNP than for DArT markers (27.5% vs. 42.0%, Table 2), indicating higher precision in genetically mapping SNP markers, probably as a result of co-dominance.

In the assessment of MTAs for quality and yield-related traits, different AM analyses were performed on subsets of the dataset. Several MTAs for SV and TKW were detected across years and water regimes, located on chromosomes 1B and 2A, respectively. All GWAS analyses corroborated the major QTLs previously detected, and reported two new QTLs, one for YR in chromosome 1B, and another for SV in chromosome 4A.

Associations on chromosome 1B were significant for wheat quality. There are known loci including Gli-B1/Glu-B3 on this chromosome, which are of great importance for some gliadin and glutenin subunits [93]. In fact, the candidate gene analysis reported the presence of a high molecular-weight glutenin subunit (HMW-GS) gene in the proximity of marker *DArT1744* (found to be significantly associated with SV and SDS), which was previously described by Mérida-García et al. [73] in association with specific weight. In line with this, Pogna et al. [93] highlighted the importance of Glu-B3 for determining protein quality with these endosperm proteins showing significant effects on SV, which also showed a high and positive correlation with SDS in our study (Figure 1 and Supplementary Materials Table S5). Likewise, Blanco et al. [86] reported three QTLs on chromosomes 1B, 6A and 7B (based on the analysis of 259 polymorphic markers) associated with SDS and SV in a recombinant inbred line population. In the present study, we found a SNP marker (*SNP620*) associated with SV, showing a positive additive effect of 1.26 (see Table 4) and also with SDS (marker effect of 0.11) (Supplementary Materials Table S6). This marker was previously placed on chromosome 1B, in the same location as MTAs for gluten index and sedimentation index [73]. Other previous studies, such as Reif et al. [94] and Fiedler et al. [95], also reported markers associated with SV on chromosome 1B, but with differing genetic positions. The additional locus for SV found on chromosome 4A (marker *DArT9459*) has not been previously reported.

Marker *SNP620*, significantly associated with SV, is located within a cluster of homoeolog gene triads coding for blue copper proteins (Table 5 and Supplementary Materials Figure S6). These proteins have been described containing a copper atom, and participate in redox processes [96], with a crucial role in the electron shuttle in plants [97]. In addition, Yao et al. [98] described the blue copper protein genes as the targets of miR408 in wheat, which is involved in the regulation of gene transcription required for heading time [99]. In our study *SNP620* was also found in proximity to a gene coding for an ubiquinone biosynthesis O-methyltransferase. Liu et al. [100] highlighted its crucial role as an electron transporter in the electron transport chain of the aerobic respiratory chain. This ubiquinone gene is involved in plant growth and development, is implied in chemical compounds biosynthesis and metabolism which are involved in plant responses to stress, and also participates in gene expression regulation and cell signal transduction [100].

On chromosome 1B we also found a significant MTA for yellow rust, in agreement with previous studies in durum and bread wheat, which placed different markers significantly associated with this trait, but in differing genetic positions [101–104]. The candidate gene analysis revealed the proximity of this marker (*SNP809*) to sugar transporter protein genes. Sugars are formed during the photosynthetic process and are essential for plant nutrition. The sucrose transport has been considered of great importance for plant productivity [105]. In line with this, the sucrose is involved in the gene expression regulation of the supposedly sugar-sensing pathway [106,107].

The majority of MTAs for TKW were located on chromosome 2A, showing both positive and negative effects. Previous studies have reported different markers in association with this quality trait, including a number mapped on chromosome 2A [38,108–110]. One of the markers found by Yao et al. [38] (SSR marker *gwm445* on chromosome 2A (68.2 cM), belongs to the same QTL found in this study for the marker *SNP1153* (chromosome 2A, 68.6 cM), and also found by Juliana et al. [111] in bread wheat lines from CIMMYT's first year-yield trials. Sukumaran et al. [43] analysed a durum wheat panel of 208 lines under yield potential, heat and drought stress conditions, and identified markers on chromosome 2A with a similar position to those detected in this study (4 markers at 70 cM and 6 markers at 69 cM) under heat stress conditions. They highlighted that several SNP markers were related to transmembranes or were uncharacterized proteins. We found several candidate genes for this

important TKW QTL (Table 5) among which the most striking feature is the presence of four reductase 1 genes (NADPH-dependent 6'-deoxychalcone synthase) and a type A response regulator 1 (Figure 4). These genes are both related with photosynthesis. Hu et al. [112] highlighted that NADPH plays a crucial role in biological processes in plants, such as the regulation of the production of reactive oxygen species (ROS) for the stress tolerance [113,114]. Additional GWAS analyses using reduced datasets revealed other interesting genes for this QTL (chromosome 2A, Supplementary Materials Table S8), encoding for the Acyl-CoA *N*-acyltransferase and the chloroplastic zeaxanthin epoxidase. The first gene has several functions in signaling and metabolic pathways [115]. The zeaxanthin epoxidase plays an important role in the xanthophyll cycle and abscisic acid (ABA) biosynthesis. The xanthophyll cycle has a main function in the dissipation of light energy excess and also increasing the photosynthetic system stability [116].

The proposed approach has successfully detected genetic markers with significant associations with TKW, SV, SDS and YR. These are of potential use in durum wheat breeding programs, and can be further interrogated to the candidate gene level using the RefSeqv1 bread wheat genome reference [69] and the durum wheat genome reference [71].

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4395/10/1/144/s1>: Figure S1: Rainfall data for Ciudad Obregon (Mexico) for the growing seasons 2012–2015; Figure S2: Manhattan plots for durum wheat mapped DArT markers. DTH: days to heading; PH: plant height; GY: grain yield; TKW: thousand kernel weight; YC: yellow color; SV: sedimentation volume; SDS: sedimentation index; and GPC: grain protein content; Figure S3: Manhattan plots for durum wheat mapped SNP markers. DTH: days to heading; PH: plant height; GY: grain yield; TKW: thousand kernel weight; YC: yellow color; SV: sedimentation volume; SDS: sedimentation index; and GPC: grain protein content; Figure S4: Quantile quantile-plots from genome-wide association studies (GWAS) analysis for durum wheat DArT markers (mapped and unmapped). DTH: days to heading; PH: plant height; LOD: lodging; GY: grain yield; TKW: thousand kernel weight; YC: yellow color; SV: sedimentation volume; SDS: sedimentation index; and GPC: grain protein content; Figure S5: Quantile quantile-plots from GWAS analysis for durum wheat SNP markers (mapped and unmapped). DTH: days to heading; PH: plant height; LOD: lodging; GY: grain yield; TKW: thousand kernel weight; YC: yellow color; SV: sedimentation volume; SDS: sedimentation index; and GPC: grain protein content; Figure S6: Blue copper protein gene cluster on durum wheat chromosome 1B. High confidence genes are shown in green colour, low confidence genes are shown in yellow; Figure S7: Cluster tree of blue copper protein gene homoeologs in bread wheat (RefSeqv1 [69]). For chromosome 1A, high confidence (HC) and low confidence (LC) genes are shown in brown and orange colour, respectively; for chromosome 1B, HC and LC genes are shown in dark and light green colour, respectively; for chromosome 1D, HC and LC genes are shown in dark and light blue colour, respectively; Table S1: Durum wheat elite lines assessed; Table S2: Best Linear Unbiased Estimates (BLUEs) outputs for all assessed traits and the association mapping analyses performed in durum wheat: [i] across years and blocks; [ii] across years for each block (FI and RI); [iii] across years and blocks for a reduced dataset (years 2013 and 2014); and [iv] across the reduced dataset for each block. DTH: days to heading; GPC: grain protein content; GY: grain yield; PH: plant height; SDS: sedimentation index; SV: sedimentation volume; TKW: thousand kernel weight; and YC: yellow colour; YR: yellow rust; LOD: lodging; Table S3: Kinship matrix for durum wheat DArT markers; Table S4: Kinship matrix for durum wheat SNP markers; Table S5: Phenotypic correlations between the assessed traits in durum wheat and their corresponding p values. YR: yellow rust; DTH: days to heading; PH: plant height; LOD: lodging; GY: grain yield; TKW: thousand kernel weight; YC: yellow colour; SV: sedimentation volume; SDS: sedimentation index; and GPC: grain protein content; Table S6: Marker-trait associations found for the association mapping analyses performed in durum wheat: [i] across years and blocks; [ii] across years for each block (FI and RI); [iii] across years and blocks for a reduced dataset (years 2013 and 2014); and [iv] across the reduced dataset for each block. SV: sedimentation volume; TKW: thousand kernel weight; SDS: sedimentation index; YR: yellow rust; “-”: unmapped marker; Table S7: Homoeolog triads for blue copper protein genes mapped in the wheat reference assembly RefSeqv1 [69]; Table S8: Candidate genes for GWAS analyses performed in durum wheat: [i] across years for each block (FI and RI); [ii] across years and blocks for a reduced dataset (years 2013 and 2014); and [iii] across the reduced dataset for each block. Molecular markers mapping positions are shown both in the durum wheat genome [71] and the wheat reference assembly RefSeqv1 [69]; Supplementary Material S1. R script used to perform the GWAS analysis; Supplementary Material S2. R script used to perform the LD analysis.

**Author Contributions:** P.H., I.S. and K.A. conceived the experiment. K.A. and I.S. selected the plant materials and agronomic traits. K.A. managed the field trials and contributed the phenotypic data. I.S., P.H., G.D., S.G., R.M.-G. and A.R.B. analysed the genotypic and phenotypic data. R.M.-G. isolated the DNA. R.M.-G., A.R.B. and P.H. carried out the AM analyses. R.M.-G., A.R.B. and P.H. drafted the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by project P12-AGR-0482 from Junta de Andalucía (Andalusian Regional Government), Spain (Co-funded by FEDER). ARB is supported by the UK Biotechnology and Biological Sciences Research Council (BBSRC) ‘Designing Future Wheat’ cross-institute strategic programme

(BB/P016855/1). PH is supported by the Spanish Ministry of Economy, Industry and Competitiveness (MINECO) project AGL2016-77149-C2-1-P.

**Conflicts of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Curtis, B.C. Wheat in the world. In *Bread Wheat Improvement and Production*; Curtis, B.C., Rajaram, S., Gomez Macpherson, H., Eds.; Plant Production and Protection Series; Food and Agriculture Organization of the United Nations: Rome, Italy, 2002; Volume 30, pp. 1–17.
2. Ranieri, R. *Geography of Durum Wheat Crop*; Pastaria Open Fields: Collecchio, Italy, 2015.
3. Bozzini, A.; Fabriani, G.; Lintas, C. Origin, Distribution, and Production of Durum Wheat in the World. In *American Associate of Cereal Chemists International, Durum Wheat*, 2nd ed.; AACCI International Press: Saint Paul, MN, USA, 2012. [[CrossRef](#)]
4. Nachit, M.M. Durum wheat breeding for Mediterranean drylands of north Africa and west Asia. In *Durum Wheats: Challenges and Opportunities*; Rajaram, S., Saari, E.E., Hettel, G.P., Eds.; Wheat Special Report; CIMMYT: Ciudad Obregón, Mexico, 1992.
5. Singh, R.P.; Huerta-Espino, J.; Fuentes, G.; Duveiller, E.; Gilchrist, L.; Henry, M.; Nicol, M.J. Resistance to diseases. In *Durum Wheat Breeding: Current Approaches and Future Strategies*; Royo, C., Nachit, M.M., Fonzo, D., Araus, J.L., Er, W.H.P., Slafer, G.A., Eds.; Food Prod Press: Binghamton, NY, USA, 2005; pp. 291–315.
6. Maccaferri, M.; Sanguineti, M.C.; Demontis, A.; El-Ahmed, A.; Garcia del Moral, L.; Maalouf, F.; Nachit, M.; Nserallah, N.; Ouabbou, H.; Rhouma, S.; et al. Association mapping in durum wheat grown across a broad range of water regimes. *J. Exp. Bot.* **2011**, *62*, 409–438. [[CrossRef](#)]
7. Abdalla, O.; Dieseth, J.A.; Singh, R.P. Breeding durum wheat at CIMMYT. In *Durum Wheats: Challenges and Opportunities*; Rajaram, S., Saari, E.E., Hettel, G.P., Eds.; Wheat Special Report; CIMMYT: Ciudad Obregón, Mexico, 1992.
8. Wrigley, C.W. Wheat: An overview of the grain that provides our daily bread. In *Encyclopedia of Food Grains*; Colin, W., Wrigley, H.C., Koushik, S., Jonathan, F., Eds.; Elsevier: Oxford, UK, 2016; pp. 105–116. [[CrossRef](#)]
9. Triboi, E.; Abad, A.; Michelena, A.; Lloveras, J.; Ollier, J.L.; Daniel, C. Environmental effects on the quality of two wheat genotypes: 1. quantitative and qualitative variation of storage proteins. *Eur. J. Agron.* **2000**, *13*, 47–64. [[CrossRef](#)]
10. Halford, N.G.; Curtis, T.Y.; Chen, Z.; Huang, J. Effects of abiotic stress and crop management on cereal grain composition: Implications for food quality and safety. *J. Exp. Bot.* **2015**, *66*, 1145–1156. [[CrossRef](#)]
11. Blumenthal, F.; Batey, I.L.; Wrigley, C.W.; Moss, H.J.; Mares, D.J.; Barlow, E.W.C. Interpretation of grain quality results from wheat variety trials with reference to high temperatures stress. *Aust. J. Agric. Res.* **1991**, *42*, 325–334. [[CrossRef](#)]
12. Campbell, C.A.; Davidson, H.R.; Winkleman, G.E. Effect of nitrogen, temperature, growth stage and duration of moisture stress on yield components and protein content of manitou spring wheat. *Can. J. Plant Sci.* **1981**, *61*, 549–563. [[CrossRef](#)]
13. Uhlen, A.K.; Hafskjold, R.; Kalhovd, A.H.; Sahlström, S.; Longva, Å.; Magnus, E.M. Effects of cultivar and temperature during grain filling on wheat protein content, composition, and dough mixing properties. *Cereal Chem.* **1998**, *75*, 460–465. [[CrossRef](#)]
14. Araus, J.L.; Slafer, G.A.; Royo, C.; Serret, M.D. Breeding for Yield Potential and Stress Adaptation in Cereals. *Crit. Rev. Plant Sci.* **2008**, *27*, 377–412. [[CrossRef](#)]
15. Curtis, T.; Halford, N.G. Food security: The challenge of increasing wheat yield and the importance of not compromising food safety. *Ann. Appl. Biol.* **2014**, *164*, 354–372. [[CrossRef](#)]
16. Steiger, D.K.; Elias, E.M.; Joppa, L.R.; Cantrell, R.G. Quality Evaluation of Lines Derived from Crosses of Langdon (*Triticum dicoccoides*) Substitution Lines to a Common Durum Wheat. In *Durum Wheats: Challenges and Opportunities*; Rajaram, S., Saari, E.E., Hettel, G.P., Eds.; Wheat Special Report; CIMMYT: Ciudad Obregón, Mexico, 1992.
17. Rharrabti, Y.; Villegas, D.; Royo, C.; Martos-Núñez, V.; Garcia del Moral, L.F. Durum wheat quality in Mediterranean environments II. Influence of climatic variables and relationships between quality parameters. *Field Crops Res.* **2003**, *80*, 133–140. [[CrossRef](#)]

18. Falconer, D.S.; Mackay, T.F.C. *Introduction to Quantitative Genetics*, 4th ed.; Pearson Education Limited: Harlow, UK, 1997; p. 464.
19. Borghi, B.; Corbellino, M.; Ciuffi, M.; La fiandra, D.; Destefanis, E.; Sgrulletta, D.; Boggini, G.; Di Fonzo, N. Effect of heat-shock during grain filling on grain quality of bread and durum wheats. *Aust. J. Agric. Res.* **1995**, *46*, 1365–1380. [[CrossRef](#)]
20. Garrido-Lestache, E.; Lopez-Bellido, R.J.; Lopez-Bellido, L. Durum wheat quality under Mediterranean conditions as affected by N rate, timing and splitting, N form and S fertilization. *Eur. J. Agron.* **2005**, *23*, 265–278. [[CrossRef](#)]
21. Holland, J.B. Genetic architecture of complex traits in plants. *Plant Biol.* **2007**, *10*, 156–161. [[CrossRef](#)]
22. Jackson, P.; Robertson, M.; Cooper, M.; Hammer, G. The role of physiological understanding in plant breeding; from a breeding perspective. *Field Crops Res.* **1996**, *49*, 11–39. [[CrossRef](#)]
23. Taghouti, M.; Gaboun, F.; Nsarellah, N.; Rhrib, R.; El-Haila, M.; Kamar, M.; Abbad-Andaloussi, F.; Udapa, S.M. Genotype × Environment interactin for quality traits in durum wheat cultivars adapted to different environments. *Afr. J. Biotechnol.* **2010**, *9*, 3054–3062.
24. Flint-Garcia, S.A.; Thuillet, A.C.; Yu, J.; Pressoir, G.; Romero, S.M.; Mitchell, S.E.; Doebley, J.; Kresovich, S.; Goodman, M.M.; Buckler, E.S. Maize association population: A high-resolution platform for quantitative trait locus dissection. *Plant J.* **2005**, *44*, 1054–1064. [[CrossRef](#)]
25. Bodmer, W.F. Human genetics: The molecular challenge. *BioEssays: News and Reviews in Molecular. Cell. Dev. Biol.* **1987**, *7*, 41–45. [[CrossRef](#)]
26. Bar-Hen, A.; Charcosset, A.; Bourgoïn, M.; Guiard, J. Relationship between genetic markers and morphological traits in a Maize inbred lines collection. *Euphytica* **1995**, *84*, 145–154. [[CrossRef](#)]
27. Virk, P.S.; Ford-Lloyd, B.V.; Jackson, M.T.; Pooni, H.S.; Clemeno, T.P. Predicting quantitative variation within rice germplasm using molecular markers. *Heredity* **1996**, *76*, 296–304. [[CrossRef](#)]
28. Beer, M.U.; Wood, P.J.; Weisz, J.; Fillion, N. Effect of cooking and storage on the amount and molecular weight of (1→3) (1→4)-D-glucan extracted from oat products by an in vitro digestion system. *Cereal Chem.* **1997**, *74*, 705–709. [[CrossRef](#)]
29. Zondervan, K.T.; Cardon, L.R. The complex interplay among factors that influence allelic association. *Nat. Rev. Genet.* **2004**, *5*, 89–100. [[CrossRef](#)]
30. Breseghello, F.; Sorrells, M.E. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* **2006**, *172*, 1165–1177. [[CrossRef](#)]
31. Maccaferri, M.; Sanguineti, M.C.; Mantovani, P.; Demontis, A.; Massi, A.; Ammar, K.; Kolmer, J.A.; Czembor, J.H.; Ezrati, S.; Tuberosa, R. Association mapping of leaf rust response in durum wheat. *Mol. Breed.* **2010**, *26*, 189–228. [[CrossRef](#)]
32. Marcotuli, I.; Gadaleta, A.; Mangini, G.; Signorile, A.M.; Zacheo, S.A.; Blanco, A.; Simeone, R.; Colasuonno, P. Development of a High-Density SNP-Based Linkage Map and Detection of QTL for beta-Glucans, Protein Content, Grain Yield per Spike and Heading Time in Durum Wheat. *Int. J. Mol. Sci.* **2017**, *18*, 1329. [[CrossRef](#)]
33. Liu, J.; Luo, W.; Qin, N.; Ding, P.; Zhang, H.; Yang, C.; Mu, Y.; Tang, H.; Liu, Y.; Li, W.; et al. A 55 K SNP array-based genetic map and its utilization in QTL mapping for productive tiller number in common wheat. *Theor. Appl. Genet.* **2018**, *131*, 2439–2450. [[CrossRef](#)]
34. Marcotuli, I.; Houston, K.; Schwerdt, J.G.; Waugh, R.; Fincher, G.B.; Burton, R.A.; Blanco, A.; Gadalena, A. Genetic Diversity and Genome Wide Association Study of beta-Glucan Content in Tetraploid Wheat Grains. *PLoS ONE* **2016**, *11*, e0152590. [[CrossRef](#)]
35. Poland, J.A.; Brown, P.J.; Sorrells, M.E.; Jannink, J.L. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* **2012**, *7*, e32253. [[CrossRef](#)]
36. Somers, D.J.; Banks, T.; Depauw, R.; Fox, S.; Clarke, J.; Pozniak, C.; McCartney, C. Genome-wide linkage disequilibrium analysis in bread wheat and durum wheat. *Genome* **2007**, *50*, 557–567. [[CrossRef](#)]
37. Wang, S.X.; Zhu, Y.L.; Zhang, D.X.; Shao, H.; Liu, P.; Hu, J.B.; Zhang, H.; Zhang, H.P.; Chang, C.; Lu, J.; et al. Genome-wide association study for grain yield and related traits in elite wheat varieties and advanced lines using SNP markers. *PLoS ONE* **2017**, *12*, e0188662. [[CrossRef](#)]
38. Yao, J.; Wang, L.; Liu, L.; Zhao, C.; Zheng, Y. Association mapping of agronomic traits on chromosome 2A of wheat. *Genetica* **2009**, *137*, 67–75. [[CrossRef](#)]

39. Li, H.; Vikram, P.; Singh, R.P.; Kilian, A.; Carling, J.; Song, J.; Burgueno-Ferreira, J.A.; Bhavani, S.; Huerta-Espino, J.; Payne, T.; et al. A high density GBS map of bread wheat and its application for dissecting complex disease resistance traits. *BMC Genom.* **2015**, *16*, 216. [[CrossRef](#)]
40. Ren, R.; Ray, R.; Li, P.; Xu, J.; Zhang, M.; Liu, G.; Yao, X.; Kilian, A.; Yang, X. Construction of a high-density DArTseq SNP-based genetic map and identification of genomic regions with segregation distortion in a genetic population derived from a cross between feral and cultivated-type watermelon. *Mol. Genet. Genom.* **2015**, *290*, 1457–1470. [[CrossRef](#)] [[PubMed](#)]
41. Sanchez-Sevilla, J.F.; Horvath, A.; Botella, M.A.; Gaston, A.; Folta, K.; Kilian, A.; Denoyes, B.; Amaya, I. Diversity Arrays Technology (DArT) Marker Platforms for Diversity Analysis and Linkage Mapping in a Complex Crop, the Octoploid Cultivated Strawberry (*Fragaria x ananassa*). *PLoS ONE* **2015**, *10*, e0144960. [[CrossRef](#)] [[PubMed](#)]
42. Kilian, A.; Huttner, E.; Wenzl, P.; Jaccoud, D.; Carling, J.; Caig, V.; Evers, M.; Heller-Uszynska, K.; Cayla, C.; Patarapuwadol, S.; et al. The fast and the cheap, SNP and DArT-based whole genome profiling for crop improvement. In *The Wake of the Double Helix: From the Green Revolution to the Gene Revolution, Proceedings of the International Congress Avenue Media, Bologna, Italy, 27–31 May 2005*; Tuberosa, R., Phillips, R.L., Gale, M., Eds.; Avenue Media: Bologna, Italy, 2005; pp. 443–461.
43. Sukumaran, S.; Reynolds, M.P.; Sansaloni, C. Genome-Wide Association Analyses Identify QTL Hotspots for Yield and Component Traits in Durum Wheat Grown under Yield Potential, Drought, and Heat Stress Environments. *Front. Plant Sci.* **2018**, *9*, 81. [[CrossRef](#)]
44. Turuspekov, Y.; Baibulatova, A.; Yermekbayev, K.; Tokhetova, L.; Chudinov, V.; Sereda, G.; Ganal, M.; Griffiths, S.; Abugalieva, S. GWAS for plant growth stages and yield components in spring wheat (*Triticum aestivum* L.) harvested in three regions of Kazakhstan. *BMC Plant Biol.* **2017**, *17* (Suppl. 1), 190. [[CrossRef](#)]
45. Johnson, M.; Kumar, A.; Oladzad-Abbasabadi, A.; Salsman, E.; Aoun, M.; Manthey, F.A.; Elias, E.M. Association Mapping for 24 Traits Related to Protein Content, Gluten Strength, Color, Cooking, and Milling Quality Using Balanced and Unbalanced Data in Durum Wheat [*Triticum turgidum* L. var. durum (Desf)]. *Front. Genet.* **2019**, *10*. [[CrossRef](#)]
46. Mangini, G.; Gadaleta, A.; Colasuonno, P.; Marcotuli, I.; Signorile, A.M.; Simeone, R.; De Vita, P.; Mastrangelo, A.M.; Laido, G.; Pecchioni, N.; et al. Genetic dissection of the relationships between grain yield components by genome-wide association mapping in a collection of tetraploid wheats. *PLoS ONE* **2018**, *13*, e0190162. [[CrossRef](#)]
47. Mengistu, D.K.; Kidane, Y.G.; Catellani, M.; Frascaroli, E.; Fadda, C.; Pe, M.E.; Dell'Acua, M. High-density molecular characterization and association mapping in Ethiopian durum wheat landraces reveals high diversity and potential for wheat breeding. *Plant Biotechnol. J.* **2016**, *14*, 1800–1812. [[CrossRef](#)]
48. Pozniak, C.J.; Knox, R.E.; Clarke, F.R.; Clarke, J.M. Identification of QTL and association of a phytoene synthase gene with endosperm colour in durum wheat. *Theor. Appl. Genet.* **2007**, *114*, 525–537. [[CrossRef](#)]
49. Reimer, S.; Pozniak, C.J.; Clarke, F.R.; Clarke, J.M.; Somers, D.J.; Knox, R.E.; Singh, A.K. Association mapping of yellow pigment in an elite collection of durum wheat cultivars and breeding lines. *Genome* **2008**, *51*, 1016–1025. [[CrossRef](#)]
50. Van Ginkel, M.R.; Trethowan, R.; Cukadar, B. *A Guide to the CIMMYT Bread Wheat Program; Wheat Special Report No 5; CIMMYT: Ciudad Obregón, Mexico*, 1998.
51. Zadoks, J.C.; Chang, T.T.; Konzak, C.F. A decimal code for the growth stages of cereals. *Weed Res.* **1974**, *14*, 415–421. [[CrossRef](#)]
52. Henstchel, V.; Kranl, K.; Hollmann, J.; Lindhauer, M.G.; Bohm, V.; Bitsch, R. Spectrophotometric determination of yellow pigment content and evaluation of carotenoids by high-performance liquid chromatography in durum wheat grain. *J. Agric. Food Chem.* **2002**, *50*, 6663–6668.
53. Martinez, C.S.; Ribotta, P.D.; León, A.E.; Añon, M.C. Colour assessment on bread wheat and triticale fresh pasta. *Int. J. Food Prop.* **2010**, *15*. [[CrossRef](#)]
54. Beleggia, R.; Platani, C.; Nigro, F.; Papa, R. Yellow pigment determination for single kernels of durum wheat (*Triticum durum* Desf.). *Cereal Chem.* **2011**, *88*, 504–508. [[CrossRef](#)]
55. Axford, D.W.E.; McDermott, E.E.; Redman, D.G. Small-scale test for breadmaking quality of wheat. *Cereal Foods World* **1978**, *23*, 477–478.
56. Seabourn, B.W.; Xiao, Z.S.; Tilley, T.; Herald, T.J.; Park, S.H. A rapid, small-scale sedimentation method to predict bread-making quality of hard winter wheat. *Crop Sci.* **2012**, *52*, 1306–1315. [[CrossRef](#)]

57. Williams, P.C.; Norris, K. *Near Infrared Technology in the Agricultural and Food Industries*, 2nd ed.; American Association of Cereal Chemistry, Inc.: St Paul, MN, USA, 2001.
58. Becker, R.A.; Chambers, J.M.; Wilks, A.R. *The New S Language: A Programming Environment for Data Analysis and Graphics*; Wadsworth & Brooks/Cole: Pacific Grove, CA, USA, 1988.
59. Kendall, M.G. A new measure of rank correlation. *Biometrika* **1938**, *30*, 81–93. [[CrossRef](#)]
60. Kendall, M.G. The treatment of ties in rank problems. *Biometrika* **1945**, *33*, 239–251. [[CrossRef](#)]
61. Chambers, J.M.; Freeny, A.; Heiberger, R.M. Chapter 5: Analysis of Variance; Designed Experiment. In *Statistical Models in S*; Chambers, J.M., Hastie, T.J., Eds.; Wadsworth & Brooks/Cole: Pacific Grove, CA, USA, 1992; pp. 145–193.
62. Yu, J.; Pressoir, G.; Briggs, W.H.; Vroh Bi, I.; Yamasaki, M.; Doebley, J.F.; McMullen, M.D.; Gaut, B.S.; Nielsen, D.M.; Holland, J.B.; et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **2006**, *38*, 203–208. [[CrossRef](#)]
63. Kang, H.M.; Zaitlen, N.A.; Wade, C.M.; Kirby, A.; Heckerman, D.; Daly, M.J.; Eskin, E. Efficient control of population structure in model organism association mapping. *Genetics* **2008**, *178*, 1709–1723. [[CrossRef](#)]
64. Horikoshi, M.; Tang, Y. ggfortify: Data Visualization Tools for Statistical Analysis Results. *R J.* **2016**, *8*, 474–489.
65. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2009.
66. Marchini, J.L. Popgen: Statistical and Population Genetics. R Package Version 1.0-3. 2013. Available online: <http://CRANR-project.org/package=popgen/> (accessed on 14 January 2020).
67. Endelman, J.B. Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* **2011**, *4*, 250–255. [[CrossRef](#)]
68. Rosyara, U.R.; De Jong, W.S.; Douches, D.S.; Endelman, J.B. Software for Genome-Wide Association Studies in Autopolyploids and Its Application to Potato. *Plant Genome* **2016**, *9*. [[CrossRef](#)] [[PubMed](#)]
69. Appels, R.; Eversole, K.; Feuillet, C.; Keller, B.; Rogers, J.; Stein, N.; Pozniak, C.J.; Choulet, F.; Distelfeld, A.; Poland, J.; et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **2018**, *361*, eaar7191.
70. Maccaferri, M.; Sanguineti, M.C.; Noli, E.; Tuberosa, R. Population structure and long-range linkage disequilibrium in a durum wheat elite collection. *Mol. Breed.* **2005**, *15*, 271–290. [[CrossRef](#)]
71. Maccaferri, M.; Harris, N.S.; Twardziok, S.O.; Pasam, R.K.; Gundlach, H.; Spannagl, M.; Ormanbekova, D.; Lux, T.; Prade, V.M.; Milner, S.G.; et al. Durum wheat genome highlights past domestication signatures and future improvement targets. *Nat. Genet.* **2019**, *51*, 885–895. [[CrossRef](#)]
72. Galvez, S.; Merida-Garcia, R.; Camino, C.; Borrill, P.; Abrouk, M.; Ramirez-Gonzalez, R.H.; Biyikliglu, S.; Amil-Ruiz, F.; Dorado, G.; Budak, H.; et al. Hotspots in the genomic architecture of field drought responses in wheat as breeding targets. *Funct. Integr. Genom.* **2019**. [[CrossRef](#)]
73. Mérida-García, R.; Guozheng, L.; He, S.; González-Dugo, V.; Dorado, G.; Gálvez, S.; Solís, I.; Zarco-Tejada, P.; Reif, J.C.; Hernández, P. Genetic dissection of agronomic and quality traits based on association mapping and genomic selection approaches in durum wheat grown in Southern Spain. *PLoS ONE* **2019**. [[CrossRef](#)]
74. Ikeda, T.; Laporte, D.C. Isocitrate Dehydrogenase Kinase/Phosphatase: AceK alleles that express kinase but not phosphatase activity. *J. Bacteriol.* **1991**, 1801–1806. [[CrossRef](#)]
75. Liu, P.L.; Du, L.; Huang, Y.; Gao, S.M.; Yu, M. Origin and diversification of leucine-rich repeat receptor-like protein kinase (LRR-RLK) genes in plants. *BMC Evol. Biol.* **2017**, *17*, 47. [[CrossRef](#)]
76. Cassman, K.G.; Dobermann, A.; Walters, D.T.; Yang, H. Meeting cereal demand while protecting natural resources and improving environmental quality. *Annu. Rev. Environ. Resour.* **2003**, *28*, 315–358. [[CrossRef](#)]
77. Araus, J.L.; Slafer, G.A.; Reynolds, M.P.; Royo, C. Plant Breeding and Drought in C3 Cereals: What Should We Breed For? *Ann. Bot.* **2002**, *89*, 925–940. [[CrossRef](#)] [[PubMed](#)]
78. Ortiz, R.; Sayre, K.D.; Govaerts, B.; Gupta, R.K.; Subbarao, G.V.; Ban, T.; Hodson, D.; Dixon, J.; Ortiz-Monasterio, I.; Reynolds, M. Climate change: Can wheat beat the heat? *Agric. Ecosyst. Environ.* **2008**, *126*, 46–58. [[CrossRef](#)]
79. Alonso-Blanco, C.; Aarts, M.G.; Bentsink, L.; Keurentjes, J.J.; Reymond, M.; Vreugdenhil, D.; Koornneef, M. What has natural variation taught us about plant development, physiology, and adaptation? *Plant Cell* **2009**, *21*, 1877–1896. [[CrossRef](#)] [[PubMed](#)]
80. Amaya, A.; Peña, R.J. Utilization and quality of durum wheat. In *Durum Wheats: Challenges and Opportunities*; Rajaram, S., Saari, E.E., Hettel, G.P., Eds.; Wheat Special Report; CIMMYT: Ciudad Obregón, Mexico, 1992.

81. Maccaferri, M.; Sanguineti, M.C.; Corneti, S.; Ortega, J.L.A.; Salem, M.B.; Bort, J.; DeAmbrogio, E.; García del Moral, L.F.; Demontis, A.; El-Ahmed, A.; et al. Quantitative trait loci for grain yield and adaptation of durum wheat (*Triticum durum* Desf.) across a wide range of water availability. *Genetics* **2008**, *178*, 489–511. [[CrossRef](#)]
82. Groos, C.; Robert, N.; Bervas, E.; Charmet, G. Genetic analysis of grain protein-content, grain yield and thousand-kernel weight in bread wheat. *Theor. Appl. Genet.* **2003**, *106*, 1032–1040. [[CrossRef](#)]
83. Sen, C.; Toms, B. Character association and component analysis in wheat (*Triticum aestivum* L.). *Crop. Res.* **2007**, *34*, 166–170.
84. Shamsi, K.; Petrosyan, M.; Noor-mohammadi, G.; Haghparast, A.; Kobrace, S.; Rasekhi, B. Differential agronomic responses of bread wheat cultivars to drought stress in the west of Iran. *Afr. J. Biotechnol.* **2011**, *10*, 2708–2715.
85. Blanco, A.; de Giovanni, C.; Laddomada, B.; Sciancalepore, A.; Simeone, R.; Devos, K.M.; Gale, M.D. Quantitative trait loci influencing grain protein content in tetraploid wheats. *Plant Breed.* **1996**, *115*, 310–316. [[CrossRef](#)]
86. Blanco, A.; Bellomo, M.P.; Cenci, A.; De Giovanni, C.; D'Ovidio, R.; Iacono, E.; Laddomada, B.; Pagnotta, M.A.; Porceddu, E.; Sciancalepore, A.; et al. A genetic linkage map of durum wheat. *Theor. Appl. Genet.* **1998**, *97*, 721–728. [[CrossRef](#)]
87. Pasha, I.; Anjum, F.M.; Butt, M.S.; Sultan, J.I. Gluten quality prediction and correlation studies in spring wheats. *J. Food Qual.* **2007**, *30*, 438–449. [[CrossRef](#)]
88. Shewry, P.R.; Tatham, A.S. Wheat. In *The Royal Society of Chemistry*; Elsevier Science B.V.: Amsterdam, The Netherlands, 2000; pp. 335–339.
89. Oelofse, R.M.; Labuschagne, M.T.; van Deventer, C.S. Influencing factors of sodium dodecyl sulfate sedimentation in bread wheat. *J. Cereal Sci.* **2010**, *52*, 96–99. [[CrossRef](#)]
90. Cubadda, R.E.; Carcea, M.; Marconi, E.; Trivisonno, M.C. Influence of protein content on durum wheat gluten strength determined by the SDS sedimentation test and by other methods. *Cereal Foods World* **2007**, *52*, 273–277. [[CrossRef](#)]
91. Carter, B.P.; Morris, C.F.; Anderson, J.A. Optimizing the SDS sedimentation test for end-use quality selection in a soft white and club wheat-breeding program. *Cereal Chem.* **1999**, *76*, 907–911. [[CrossRef](#)]
92. De Villiers, O.T.; Laubscher, E.W. Use of the SDSS test to predict the protein content and bread volume of wheat cultivars. *S. Afr. J. Plant Soil* **1995**, *12*, 140–142. [[CrossRef](#)]
93. Pogna, N.E.; Autran, J.C.; Mellini, F.; Lafiandra, D.; Feillet, P. Chromosome 1B-encoded gliadins and glutenins subunits. *J. Cereal Sci.* **1990**, *11*, 15–34. [[CrossRef](#)]
94. Reif, J.C.; Gowda, M.; Maurer, H.P.; Longin, C.F.; Korzun, V.; Ebmeyer, E.; Bothe, R.; Pietsch, C.; Wurschum, T. Association mapping for quality traits in soft winter wheat. *Theor. Appl. Genet.* **2010**, *122*, 961–970. [[CrossRef](#)]
95. Fiedler, J.D.; Salsman, E.; Liu, Y.; Michalak de Jimenez, M.; Hegstad, J.B.; Chen, B.; Manthey, F.A.; Chao, S.; Xu, S.; Elias, E.M.; et al. Genome-Wide Association and Prediction of Grain and Semolina Quality Traits in Durum Wheat Breeding Populations. *Plant Genome* **2017**, *10*. [[CrossRef](#)]
96. Sykes, A.G. Plastocyanin and the Blue Copper Proteins. In *Long-Range Electron Transfer in Biology. Structure and Bonding*; Springer: Berlin/Heidelberg, Germany, 1990; pp. 175–224. [[CrossRef](#)]
97. Feng, H.; Zhang, Q.; Wang, Q.; Wang, X.; Liu, J.; Li, M.; Huang, L.; Kang, Z. Target of taе-miR408, a chemocyanin-like protein gene (TaCLP1), plays positive roles in wheat response to high-salinity, heavy cupric stress and stripe rust. *Plant Mol. Biol.* **2013**, *83*, 433–443. [[CrossRef](#)]
98. Yao, Z.J.; Lin, R.M.; Xu, S.C.; Li, Z.F.; Wan, A.M.; Ma, Z.Y. The molecular tagging of the yellow rust resistance gene Yr7 in wheat transferred from differential host Lee using microsatellite markers. *Sci. Agric. Sin.* **2006**, *39*, 1146–1152.
99. Zhao, X.Y.; Hong, P.; Wu, J.Y.; Chen, X.B.; Ye, X.G.; Pan, Y.Y.; Wang, J.; Zhang, X.S. The taе-miR408-Mediated Control of TaTOC1 Genes Transcription Is Required for the Regulation of Heading Time in Wheat. *Plant Physiol.* **2016**, *170*, 1578–1594. [[CrossRef](#)]
100. Liu, G.; Zhao, Y.; Gowda, M.; Longin, C.F.H.; Reif, J.C.; Mette, M.F. Predicting hybrid performances for quality traits through genomic-assisted approaches in central European wheat. *PLoS ONE* **2016**, *11*, e0158635. [[CrossRef](#)] [[PubMed](#)]

101. Maccaferri, M.; Zhang, J.; Bulli, P.; Abate, Z.; Chao, S.; Cantu, D.; Bossolini, E.; Chen, X.; Pumphrey, M.; Dubcovsky, J. A genome-wide association study of resistance to stripe rust (*Puccinia striiformis* f. sp. *tritici*) in a worldwide collection of hexaploid spring wheat (*Triticum aestivum* L.). *G3 Genes Genomes Genet.* **2015**, *5*, 449–465. [CrossRef]
102. Liu, W.; Maccaferri, M.; Rynearson, S.; Letta, T.; Zegeye, H.; Tuberosa, R.; Chen, X.; Pumphrey, M. Novel Sources of Stripe Rust Resistance Identified by Genome-Wide Association Mapping in Ethiopian Durum Wheat (*Triticum turgidum* ssp. *durum*). *Front. Plant Sci.* **2017**, *8*, 774. [CrossRef] [PubMed]
103. Godoy, J.G.; Rynearson, S.; Chen, X.; Pumphrey, M. Genome-Wide Association Mapping of Loci for Resistance to Stripe Rust in North American Elite Spring Wheat Germplasm. *Phytopathology* **2018**, *108*, 234–245. [CrossRef]
104. Liu, W.; Naruoka, Y.; Miller, K.; Garland-Campbell, K.A.; Carter, A.H. Characterizing and Validating Stripe Rust Resistance Loci in US Pacific Northwest Winter Wheat Accessions (*Triticum aestivum* L.) by Genome-wide Association and Linkage Mapping. *Plant Genome* **2018**, *11*. [CrossRef]
105. Lemoine, R. Sucrose transporters in plants: Update on function and structure. *Biochim. Biophys. Acta* **1999**, *1465*, 246–262. [CrossRef]
106. Koch, K.E. Carbohydrate-modulated gene expression in plants. *Annu. Rev. Plant Physiol. Plant Mole Biol.* **1996**, *47*, 509–540. [CrossRef]
107. Smeekens, S.; Rook, F. Sugar sensing and sugar-mediated signal transduction in plants. *Plant Physiol* **1997**, *115*, 7–13. [CrossRef]
108. Sun, X.-Y.; Wu, K.; Zhao, Y.; Kong, F.-M.; Han, G.-Z.; Jiang, H.-M.; Huang, X.-J.; Li, R.-J.; Wang, H.-G.; Li, S.-S. QTL analysis of kernel shape and weight using recombinant inbred lines in wheat. *Euphytica* **2008**, *165*, 615–624. [CrossRef]
109. Patil, R.M.; Tamhankar, S.A.; Oak, M.D.; Raut, A.L.; Honrao, B.K.; Rao, V.S.; Misra, S.C. Mapping of QTL for agronomic traits and kernel characters in durum wheat (*Triticum durum* Desf.). *Euphytica* **2013**, *190*, 117–129. [CrossRef]
110. Cui, F.; Fan, X.; Chen, M.; Zhang, N.; Zhao, C.; Zhang, W.; Han, J.; Ji, J.; Zhao, X.; Yang, L.; et al. QTL detection for wheat kernel size and quality and the responses of these traits to low nitrogen stress. *Theor. Appl. Genet.* **2016**, *129*, 469–484. [CrossRef] [PubMed]
111. Juliana, P.; Poland, J.A.; Huerta-Espino, J.; Shrestha, S.; Crossa, J.; Crespo-Herrera, L.; Henrique Toledo, F.; Govidan, V.; Mondal, S.; Kumar, U.; et al. Improving grain yield, stress resilience and quality of bread wheat using large-scale genomics. *Nat. Genet.* **2019**. [CrossRef] [PubMed]
112. Hu, C.H.; Wei, X.Y.; Yuan, B.; Yao, L.B.; Ma, T.T.; Zhang, P.P.; Wang, X.; Wang, P.Q.; Liu, W.T.; Li, W.Q.; et al. Genome-Wide Identification and Functional Analysis of NADPH Oxidase Family Genes in Wheat During Development and Environmental Stress Responses. *Front. Plant Sci.* **2018**, *9*, 906. [CrossRef] [PubMed]
113. Kaya, H.; Nakajima, R.; Iwano, M.; Kanaoka, M.M.; Kimura, S.; Takeda, S.; Kawarazaki, T.; Senzaki, E.; Hamamura, Y.; Higashiyama, T.; et al.  $Ca^{2+}$ -activated reactive oxygen species production by Arabidopsis RbohH and RbohJ is essential for proper pollen tube tip growth. *Plant Cell* **2014**, *26*, 1069–1080. [CrossRef]
114. Gupta, D.K.; Pena, L.B.; Romero-Puertas, M.C.; Hernández, A.; Inouhe, M.; Sandalio, L.M. NADPH oxidases differentially regulate ROS metabolism and nutrient uptake under cadmium toxicity. *Plant Cell Environ.* **2017**, *40*, 509–526. [CrossRef]
115. Fu, W.; Shen, Y.; Hao, J.; Wu, J.; Ke, L.; Wu, C.; Huang, K.; Luo, B.; Xu, M.; Cheng, X.; et al. Acyl-CoA N-acyltransferase influences fertility by regulating lipid metabolism and jasmonic acid biogenesis in cotton. *Sci. Rep.* **2015**, *5*, 11790. [CrossRef]
116. Sui, N.; Li, M.; Meng, Q.-W.; Tian, J.-C.; Zhao, S.-J. Photosynthetic Characteristics of a Super High Yield Cultivar of Winter Wheat During Late Growth Period. *Agric. Sci. China* **2010**, *9*, 346–354. [CrossRef]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).





Article

# Exploring the Genetic Architecture of Root-Related Traits in Mediterranean Bread Wheat Landraces by Genome-Wide Association Analysis

Rubén Rufo <sup>1</sup>, Silvio Salvi <sup>2</sup>, Conxita Royo <sup>1</sup> and Jose Miguel Soriano <sup>1,\*</sup>

<sup>1</sup> Sustainable Field Crops Programme, IRTA (Institute for Food and Agricultural Research and Technology), 25198 Lleida, Spain; ruben.rufo@irta.cat (R.R.); conxita.royo@irta.cat (C.R.)

<sup>2</sup> Department of Agricultural and Food Sciences, University of Bologna, Viale Fanin 44, 40127 Bologna, Italy; silvio.salvi@unibo.it

\* Correspondence: josemiguel.soriano@irta.cat

Received: 3 April 2020; Accepted: 23 April 2020; Published: 25 April 2020

**Abstract:** Background: Roots are essential for drought adaptation because of their involvement in water and nutrient uptake. As the study of the root system architecture (RSA) is costly and time-consuming, it is not generally considered in breeding programs. Thus, the identification of molecular markers linked to RSA traits is of special interest to the breeding community. The reported correlation between the RSA of seedlings and adult plants simplifies its assessment. Methods: In this study, a panel of 170 bread wheat landraces from 24 Mediterranean countries was used to identify molecular markers associated with the seminal RSA and related traits: seminal root angle, total root number, root dry weight, seed weight and shoot length, and grain yield (GY). Results: A genome-wide association study identified 135 marker-trait associations explaining 6% to 15% of the phenotypic variances for root related traits and 112 for GY. Fifteen QTL hotspots were identified as the most important for controlling root trait variation and were shown to include 31 candidate genes related to RSA traits, seed size, root development, and abiotic stress tolerance (mainly drought). Co-location for root related traits and GY was found in 17 genome regions. In addition, only four out of the fifteen QTL hotspots were reported previously. Conclusions: The variability found in the Mediterranean wheat landraces is a valuable source of root traits to introgress into adapted phenotypes through marker-assisted breeding. The study reveals new loci affecting root development in wheat.

**Keywords:** drought stress; association mapping; root system architecture; QTL hotspot; seminal root

## 1. Introduction

Wheat is the most widely cultivated crop in the world, covering around 219 million ha (Faostat 2017, <http://www.fao.org/faostat/>). It is a staple food for humans, as it provides 18% of daily human intake of calories and 20% of protein (<http://www.fao.org/faostat/>). Global wheat demand is estimated to increase by 60% by the year 2050 [1], so wheat production will need to rise by 1.7% per year until then. Achieving this objective is a great challenge under the current climate change scenario, as the prediction models estimate a precipitation decrease of 25% to 30% and a temperature increase of 4 °C to 5 °C for the Mediterranean region [2]. It is well known that wheat production is greatly affected by environmental stresses such as drought and heat [3] that negatively affect yield and grain quality [4]. Drought is considered the greatest environmental constraint to yield and yield stability in rainfed production systems [5]. Environmental effects on yield in the Mediterranean Basin have been estimated at 60% for bread wheat [6] and 98% for durum wheat [7]. The expected effects of climate change and the declining availability of water and chemical fertilizers will require the release of cultivars with an enhanced genetic capacity to maintain acceptable yield levels and yield stability under harmful

environmental conditions [8,9]. To cope with the challenges of climate change, breeders are particularly challenged to stretch the adaptability and performance stability of new cultivars, so many improvement programs are focussing on breeding for adaptation [10].

Plants respond and adapt to water deficit using various strategies that have evolved at several levels of function and are components of the conceptual framework developed by Reynolds et al. [11], which defines drought resistance in terms of dehydration escape, tolerance, and avoidance. Traits defining root system architecture (RSA) are critical for wheat adaptation to drought environments and non-optimal nutritional supply conditions [12]. Besides, water-use efficiency (WUE) can be significantly increased by optimizing the anatomy and growth features of roots [13]. Root traits are critical for drought tolerance due to its role in plant performance and the acquisition of nutrients and water from dry soils [14]. The wheat plant includes two types of roots: seminal (embryonal) and nodal (crown or adventitious or adult root system). The seminal roots are the first to penetrate the soil and remain functional during the whole plant cycle [9,15]. A correlation between seminal and adult roots in terms of size, dry-weight, or even specific architectural features have been reported [9,13]. Since the evaluation of RSA features in the field is very difficult, expensive, and time-consuming when a large number of genotypes need to be phenotyped, several studies have been carried out at early growth stages to allow an optimal screening of RSA traits [8,12,16–18]. Maccaferri et al. [9] observed that among RSA traits, those involving the root structure and related to the uptake of nutrients and water are root length, surface area and volume, and the number of roots, while root diameter is significantly associated with drought tolerance. Another RSA trait of interest in wheat is the seminal root angle (SRA), whose features suggest that narrow angles could lead to deeper root growth to obtain water from deeper soil layers and hence maintain higher yields [5,13].

Identifying quantitative trait loci (QTLs) and applying marker-assisted selection is of particular interest for RSA because the trait is important but difficult to phenotype. In the last few years, genome-wide association studies (GWAS) have become very popular because of their use of germplasm collections with wider variability than the classical bi-parental crosses. These collections allow many recombination events to be detected, making the association between genotype and phenotype more accurate. Collections of landraces are an ideal subject of GWAS [19] since they are genetically diverse repositories of unique traits that have evolved in local environments characterized by a wide range of biotic and abiotic conditions. Several studies have shown that Mediterranean wheat landraces possess a wide genetic background for root architecture, yield formation, stress tolerance, and quality traits [17–22]. In the current study, a GWAS for three RSA traits and two related traits was performed on a panel of 170 bread wheat (*Triticum aestivum* L.) landraces from 24 Mediterranean countries with the following goals: (1) to detect differences in RSA among genetic subpopulations previously distinguished in the panel, (2) to identify correlations among RSA and grain yield under rainfed conditions, and (3) to identify molecular markers and candidate genes linked to root-related traits and candidate gene models for the associations.

## 2. Materials and Methods

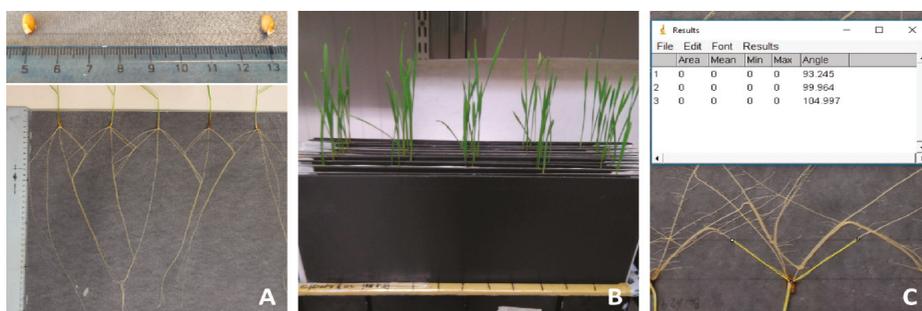
### 2.1. Plant Material

A germplasm collection of 170 bread wheat (*Triticum aestivum* L.) genotypes from the MED6WHEAT IRTA panel described by Rufo et al. [23] was used in this study. The panel was genotyped and characterized using the Illumina Infinium 15K Wheat SNP Chip at Trait Genetics GmbH (Gatersleben, Germany), and markers were ordered according to the SNP map developed by Wang et al. [24]. The collection was previously structured into three subpopulations (SPs) matching their geographical origin [23]: western (SP1, WM), northern (SP2, NM), and eastern Mediterranean (SP3, EM) (Supplementary Materials, Table S1). Additionally, the cultivars ‘Arthur Nick’, ‘Anza’, ‘Soissons’, and ‘Chinese Spring’ were included as checks.

## 2.2. Root Morphology and Statistical Analysis

Root analysis was performed following the protocol described by Canè et al. [8], which was slightly modified in the current study (Figure 1). Ten representative seeds were randomly chosen from each genotype, weighed, sterilized in a 10% sodium hypochlorite solution for 5–10 min, washed thoroughly in distilled water and placed on hydrated filter paper in a 140 mm Petri dish at 28 °C for 24 h. Subsequently, five seedlings were selected on the basis of a normal seminal root emergence and were spaced 8 cm from each other on a filter paper sheet placed on a vertical black rectangular polycarbonate plate (42.5 × 38.5 cm). Finally, each plate was covered with another wet sheet of filter paper. Distilled water was used for the plantlets' growth. The plantlets were grown in a growth chamber for 14 days at 22 °C under a 16-h light photoperiod. In addition to the ten seed weight (SW), four other traits were scored for each genotype: total root number (TRN), shoot length (SL) from the seed to the tip of the longest leaf and SRA, obtained using a digital camera following the methodology described in Canè et al. [8]. The images were processed with ImageJ software [25]. The angle between the two external roots of each plantlet was measured at a distance of 3.5 cm from the tip of the seed. Finally, the roots were desiccated at 70 °C for 24 h to obtain the root dry weight (RDW).

The experimental design followed a randomized complete block with two replications in time. Means of five observational units for each genotype were used for TRN, RDW, and SL, while only three observational units were used for SRA because the two external ones were considered as border plantlets for root angle.



**Figure 1.** Experimental setup for the analysis of seminal root traits. Seeds were placed 8 cm apart on moist filter paper (A) and kept in a box with distilled water in a growth chamber for 14 days at 22 °C under a 16-h light photoperiod (B). (C) Example of seminal root angle measurement, using ImageJ software.

## 2.3. Grain Yield

Field experiments were carried out in 2016, 2017, and 2018 harvesting seasons in Gimenezs, Lleida, north-east Spain (41°38' N and 0°22' E, 260 m a.s.l) under rainfed conditions. The experiments followed a non-replicated augmented design with two replicated checks (the cultivars 'Anza' and 'Soissons') and plots of 3.6 m<sup>2</sup>. The experimental design is shown in Supplementary Materials, Figure S1. Sowing density was adjusted to 250 germinable seeds m<sup>2</sup>. Weeds and diseases were controlled following standard practices at the site. The anthesis date was determined in each plot. Grain yield (GY, t ha<sup>-1</sup>) was determined by mechanically harvesting the plots at maturity and expressed on a 12% moisture level.

#### 2.4. Statistical Analysis

Phenotypic data for GY was fitted to a linear mixed model with the check cultivars as fixed effects and the row number, column number and cultivar as random effects following the SAS PROC MIXED procedure:

$$y = X\beta + Z\gamma + \varepsilon \quad (1)$$

where  $\beta$  is an unknown vector of fixed-effects parameters with known design matrix  $X$ ,  $\gamma$  is an unknown vector of random-effects parameters with known design matrix  $Z$ , and  $\varepsilon$  is an unknown random error vector whose elements are no longer required to be independent and homogeneous.

Restricted maximum likelihood was used to estimate the variance components and to produce the best linear unbiased predictors (BLUPs) for the traits of each cultivar and year with the SAS-STAT statistical package (SAS Institute Inc, Cary, NC, USA).

Analyses of variance (ANOVA) were performed for the root traits, considering the genotypes and the replication as random effects in the model. Additionally, for a subset of 55 of the 141 structured landraces, selected as having an SP membership  $q > 0.8$  (WM, 17; NM, 15; EM, 23), the sum of squares of the cultivar effect in the ANOVAs was partitioned into differences between SPs and differences within them. ANOVA for grain yield was performed for the complete collection, considering genotype, year, and the combination of genotype and year the sources of variation. Least squares means were calculated and compared using the Tukey HSD test at  $P < 0.05$ . Pearson correlation coefficients among root traits were computed. Repeatability (H) was calculated on a mean basis across two replications following the formula described by Harper [26]  $r = (B - W) / (B + (n - 1) W)$ , where  $n$  is the number of genotypes and  $B$  and  $W$  the two variances from the ANOVA table: between ( $B$ ) and within ( $W$ ). Frequency distributions, ANOVAs, the Tukey test, and the Pearson correlation coefficients were calculated using the JMP v13.1.0 statistical package (SAS Institute Inc, Cary, NC, USA).

#### 2.5. Genome-Wide Association Analysis

A GWAS was performed for the mean of measured root traits and from the BLUPs for GY per year and across years with TASSEL 5.0 software [27]. A mixed linear model (MLM) was conducted using the information of the genetic structure reported in Rufo et al. [23] as the fixed effect and a kinship ( $K$ ) matrix, calculated using Haploview [28], as the random effect ( $Q + K$  model) at the optimum compression level. In addition, the anthesis date was incorporated as a cofactor in the analysis. As reported in other studies [29–32], an adjusted  $-\log_{10} P > 3$  was established as a threshold for considering a marker-trait association (MTA) statistically significant. A moderate threshold at  $-\log_{10} P > 2.5$  was also established for GY. Confidence intervals (CI) for MTAs were calculated for each chromosome according to the linkage disequilibrium (LD) decay reported by Rufo et al. [23]. In order to simplify the MTA information, the associations were grouped into QTL hotspots when at least two MTAs belonging to different traits overlapped their CIs. Circular Manhattan plots were performed using the R package “CMplot” (<http://www.r-project.org>).

#### 2.6. Gene Annotation

Gene annotation within the CIs of the QTL hotspots was performed using the gene models for high-confidence genes reported for the wheat genome sequence [33], available at <https://wheat-urgi.versailles.inra.fr/Seq-Repository/Assemblies>. Markers flanking the CIs were used to estimate physical distances from genetic distances.

### 3. Results

#### 3.1. Phenotypic Data of Root Traits

A summary of the genetic variation of the root traits is shown in Table 1. The genotypes showed a low coefficient of variation (CV) with a narrow range of variation among traits, from 10.4 for SW to 18.8 for RDW and repeatability (H) ranging from 48.5% for RDW to 75.4% for SW.

**Table 1.** Statistics of the seminal root traits.

Table	TRN (N)	RDW (mg)	SRA (°)	SW (g)	SL (cm)
Min	3.2	43	53.1	0.27	14.9
Max	5.4	20.5	125.9	0.63	30.2
Mean	4.4	11.7	98.6	0.48	22.0
SD	0.5	2.2	13.3	0.05	2.8
CV (%)	10.9	18.8	13.5	10.4	12.7
H (%)	52.0	48.5	70.0	75.4	50.0

SD, standard deviation; CV, coefficient of variation; H, repeatability; TRN, total root number; RDW, root dry weight; SRA, seminal root angle; SW, seed weight; SL, seed length.

The ANOVA (Table 2) for cultivars with a high membership coefficient ( $q > 0.8$ ) showed that for all traits the total variability was mainly explained by the genotype effect, in a range from 63.5% for SL to 88.8% for SW. When the sum of squares of the genotype effect was partitioned into differences between and within SPs, the results revealed that the genetic variability was mainly explained by differences within SPs in a range from 47.8% for TRN to 71.8% for SRA (Table 2). Differences between SPs were statistically significant for SRA, TRN, SW, and SL, in a range from 6.0% of the genotype effect for SL to 25.3% for TRN (Table 2). The sum of squares within SPs was partitioned into western (WM), northern (NM), and eastern (EM) effects, being statistically significant for SRA (40.8%), TRN (28.3%), SW (38.3%), and SL (26.8%) in the western SP, SRA (34.4%) in the northern SP and RDW (53.6%) and SW (55.4%) in the eastern SP.

**Table 2.** Percentage of the sum of squares of the ANOVA in a set of 55 bread wheat landraces structured into three genetic subpopulations with membership coefficient  $q > 0.8$ .

Source of Variation	df	TRN	RDW	SRA	SW	SL
Replicate	1	0.1	0.7	0.0	0.1	0.3
Genotype	54	73.1 ***	64.8 *	82.1 ***	88.8 ***	63.5 *
Between SPs	2	25.3 ***	0.6	10.3 ***	18.9 ***	6.0 *
Within SPs	52	47.8 *	64.2 *	71.8 ***	69.9 ***	57.5 *
WM	16	28.3 ***	31.6	40.8 ***	38.3 ***	26.8 *
NM	14	16.1	14.8	34.4 ***	6.3	25.2
EM	22	55.6	53.6 *	24.8	55.4 ***	48.0
Replicate x Genotype	54	26.8	34.5	17.9	11.1	36.2
Total	108					

WM, western Mediterranean; NM, northern Mediterranean; EM, eastern Mediterranean; TRN, total root number; RDW, root dry weight; SRA, seminal root angle; SW, seed weight; SL, seed length. \*  $P < 0.05$ , \*\*\*  $P < 0.001$ .

The ANOVA for grain yield revealed that the genotype effect was the most important in the phenotypic expression of traits, accounting for 59% of the total phenotypic variation, whereas the year effect accounted only for 5%. The interaction accounted for almost 36% of the phenotypic variation although it was not significant (Table 3).

**Table 3.** Percentage of the sum of squares for grain yield of the ANOVA in the collection of 170 bread wheat landraces.

Source of Variation	df	Grain Yield	P
Genotype	169	59.2	< 0.001
Year	2	5.1	< 0.001
Genotype x Year	338	35.7	No significant
Total	509		

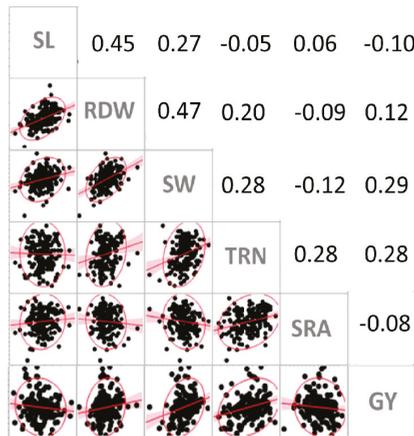
The landraces from northern Mediterranean countries showed the highest number of seminal roots with a root angle not statistically different from the western Mediterranean ones. On the other hand, eastern Mediterranean landraces showed the lowest number of roots but the widest angle. These landraces reported the lowest SW and the longest shoots. No differences were reported for RDW among the three SPs (Table 4).

**Table 4.** Means comparison of seminal root traits measured in a set of 55 Mediterranean wheat landraces structured into three genetic subpopulations [23] with  $q > 0.8$ . Means within columns with different letters are significantly different at  $P < 0.05$  following a Tukey test.

	TRN (N)	RDW (mg)	SRA (°)	SW (g)	SL (cm)
Northern Mediterranean	4.7 a	0.011 a	98.5 b	0.50 a	20.8 b
Western Mediterranean	4.3 b	0.011 a	96.2 b	0.49 a	21.4 ab
Eastern Mediterranean	4.0 c	0.011 a	106.5 a	0.45 b	22.5 a

TRN, total root number; RDW, root dry weight; SRA, seminal root angle; SW, seed weight; SL, seed length.

Correlation coefficients between root traits were calculated, showing highly significant correlation coefficients between RDW and SW and RDW and SL ( $r = 0.47$  and  $0.45$  respectively;  $P < 0.0001$ ). Moderate significant correlations were reported for TRN with RDW, SW and SRA ( $r = 0.20$ ,  $0.28$  and  $0.28$ , respectively), and for SW with SL ( $r = 0.27$ ). Finally, a negative correlation coefficient ( $r = -0.12$ ) was found between SRA and SW (Figure 2). GY showed a moderate significant correlation with TRN and SW ( $r = 0.28$  and  $0.29$ , respectively;  $P < 0.0005$ ).

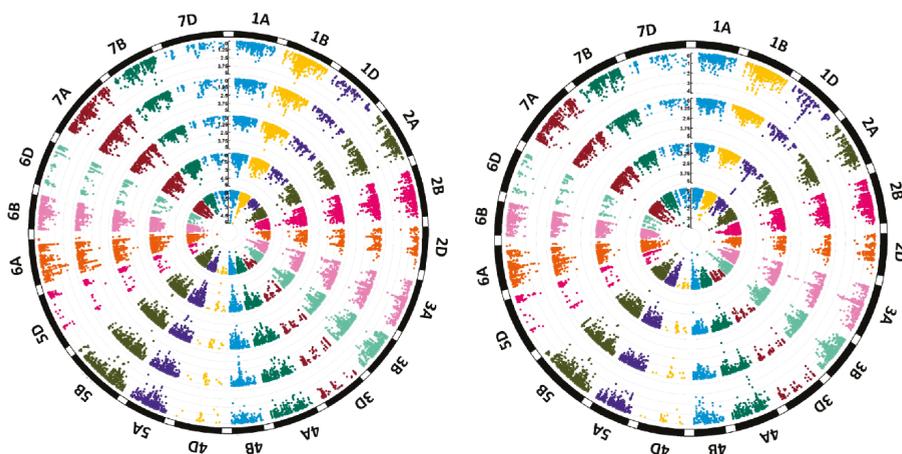


**Figure 2.** Correlations between seminal root traits and grain yield. On the right side are shown the values of the correlation coefficients ( $r$ ). SL, seed length; RDW, root dry weight; SW, seed weight; TRN, total root number; SRA, seminal root angle; GY, grain yield.

### 3.2. Marker-Trait Associations

After filtering for duplicated patterns, missing values, and minor frequency alleles, a total of 10,458 SNPs were used to genotype the panel of 170 wheat landraces [23].

The results of the GWAS for root related traits are reported in Figure 3 and Supplementary Materials, Table S2. Using a common threshold of  $-\log_{10} P > 3$ , as reported by other authors [29–32], a total of 135 MTAs were identified for the analyzed traits. Of these, 50 MTAs corresponded to SW, 39 to RDW, 18 to SL, 17 to SRA, and 11 to TRN. The A and B genomes harbored 46% and 48% of MTAs, respectively, whereas the D genome harbored only 6% of MTAs. The number of MTAs per chromosome ranged from 1 in chromosomes 4D, 5D, and 6D to 14 in chromosome 1B, with a mean of 7 MTAs per chromosome. Most of the MTAs (88%) showed a phenotypic variance explained (PVE) by each MTA in a range of 5% to 10%, and only 2% showed a PVE higher than 15%. Among traits, the PVE mean was stable in a range of 7% (SL) to 9% (RDW).



**Figure 3.** GWAS for root related traits (left circle) and grain yield for 3 years and across years (right circle). From the inside out, root traits correspond to RDW, SW, TRN, SRA, and SL, whereas for GY corresponds to 2016, 2017, 2018 harvesting seasons and the mean across years.

In order to identify and summarize the genomic regions most involved in trait variation, QTL hotspots were defined when two or more MTAs from different traits were grouped together within the same LD block. LD was previously estimated for locus pairs in each chromosome, and its decay was set to 1 to 10 cM depending on the chromosome [23]. Using this approach, 15 QTL hotspots grouping 43 MTAs were identified (Table 5), while 92 MTAs remained as singletons.

The results of the GWAS for GY are reported in Figure 3 and Supplementary Materials, Table S3. A common threshold of  $-\log_{10} P > 3$ , detected a total of 40 MTAs, thus a moderate threshold at  $-\log_{10} P > 2.5$  was applied, increasing the number of significant associations to 112. Of these, 32 MTAs corresponded to the year 2016, 30 to 2017, 18 to 2018, and 32 across years. The A and B genomes harbored 43% and 38% of MTAs, respectively, whereas the D genome harbored only 18% of MTAs. The number of MTAs per chromosome ranged from 1 in chromosomes 3D and 6B to 16 in chromosome 1D. Chromosomes 1A, 4D, 5D, and 7D did not show any association. All of MTAs showed a phenotypic variance explained (PVE) by each MTA in a range from 5% to 11%. Most of the MTAs with a PVE  $> 8\%$  were located on chromosome 1D (76%, 13 out of 17), whereas the percentage increased to 80% among MTAs with a PVE  $> 10\%$  (4 out of 5).

In order to identify and summarize the genomic regions with a pleiotropic effect for root traits and grain yield, QTL hotspots were defined as previously but including the MTAs for GY. Using

this approach, 17 QTL hotspots grouping 81 MTAs were identified (Table 6). From them, five were in common with those reported only with root traits (rootQTL1B.3, rootQTL2A.2, rootQTL3B.2, rootQTL6A.1, and rootQTL6A.2). GY shared 8 genomic regions with SW and 9 with RDW, 4 with SL, and 3 with SRA, whereas no regions were in common with TRN. In 59% of these genomic regions, GY co-localize with only one root trait, whereas the other 41% co-localize with two different root traits.

**Table 5.** Root QTL hotspots. Positions are indicated in cm.

QTL Hotspot	MTAs	Trait	Chromosome	Peak	CI Left	CI Right
rootQTL1B.1	2	RDW, SW	1B	70.6	69.6	71.6
rootQTL1B.2	2	RDW, TRN	1B	77.5	75.9	79.1
rootQTL1B.3	2	RDW, SL	1B	83.0	81.4	84.6
rootQTL2A.1	2	RDW, SW	2A	47.8	46.7	48.9
rootQTL2A.2	2	RDW, SW	2A	104.1	103.6	104.6
rootQTL2A.3	2	SW, SL	2A	177.5	176.9	178.2
rootQTL2B.1	2	SW, SL	2B	109.5	109.0	110.0
rootQTL3B.1	3	RDW, SW, TRN	3B	62.3	61.4	63.2
rootQTL3B.2	2	SRA, SW	3B	80.6	79.6	81.5
rootQTL5A.1	2	RDW, SL	5A	56.5	56.0	57.0
rootQTL5B.1	2	RDW, SL	5B	95.7	94.5	96.9
rootQTL6A.1	2	RDW, SL	6A	45.8	40.0	51.6
rootQTL6A.2	2	RDW, TRN	6A	76.7	70.7	82.6
rootQTL6A.3	2	RDW, SW	6A	138.4	132.3	144.6
rootQTL7A.1	3	RDW, SRA, TRN	7A	216.6	215.3	218.0

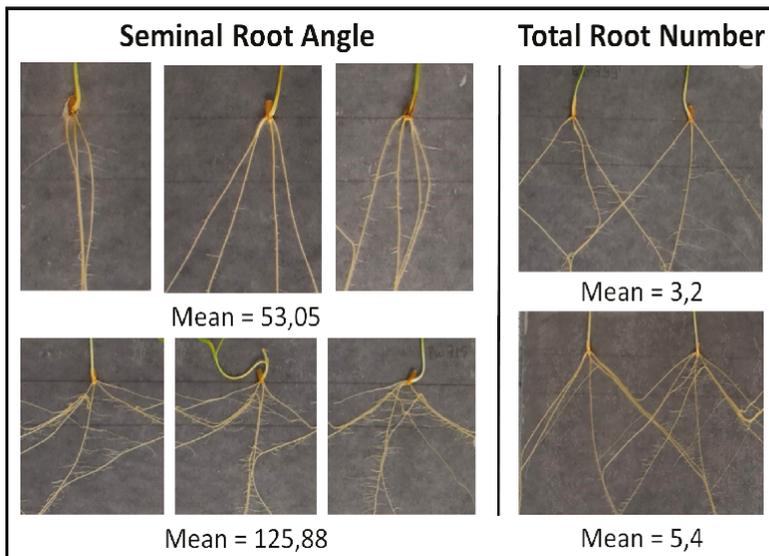
TRN, total root number; RDW, root dry weight; SRA, seminal root angle; SW, seed weight; SL, seed length.

**Table 6.** QTL hotspots including grain yield. Positions are indicated in cm.

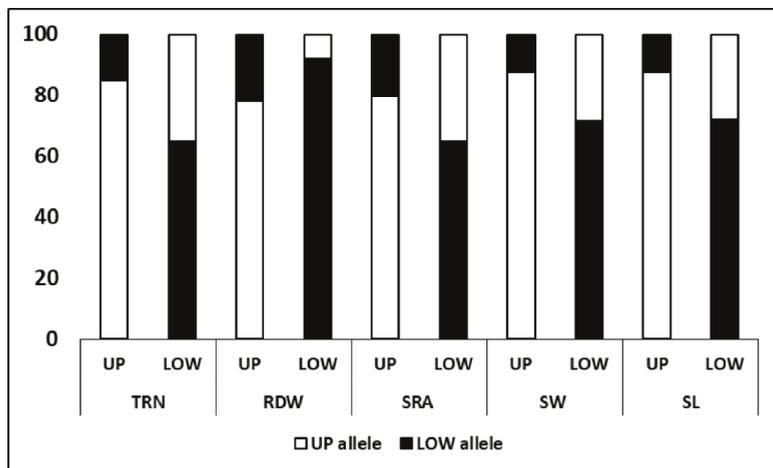
QTL Hotspot	MTAs	Trait	Chromosome	Peak	CI Left	CI Right
QTL yield/root_1B.1	3	GY, SRA	1B	8.4	7.4	9.4
QTL yield/root_1B.2	3	GY, SW	1B	43.9	42.9	44.9
QTL yield/root_1B.3	3	GY, SW	1B	63.5	61.5	65.5
QTL yield/root_1B.4	8	GY, RDW, SL	1B	83.3	82.3	84.3
QTL yield/root_2A.1	8	GY, RDW, SW	2A	104.1	103.6	104.6
QTL yield/root_2A.2	3	GY, SRA	2A	151.3	150.8	151.9
QTL yield/root_3A.1	8	GY, RDW, SL	3A	84.3	81.9	86.7
QTL yield/root_3B.1	5	GY, SW	3B	72.8	70.8	74.8
QTL yield/root_3B.2	3	GY, SRA, SW	3B	80.5	79.6	81.5
QTL yield/root_4B.1	4	GY, SW	4B	76.6	74.1	79.2
QTL yield/root_5B.1	3	GY, SL	5B	57.8	56.8	58.9
QTL yield/root_5B.2	7	GY, RDW	5B	77.3	75.9	78.8
QTL yield/root_5B.3	2	GY, RDW	5B	176.2	175.2	177.2
QTL yield/root_6A.1	5	GY, RDW, SL	6A	45.6	39.6	51.6
QTL yield/root_6A.2	10	GY, RDW, SW	6A	76.6	70.7	82.6
QTL yield/root_7A.1	3	GY, RDW, SW	7A	135	133.2	136.8
QTL yield/root_7B.1	3	GY, RDW	7B	70.0	67.8	72.3

TRN, total root number; RDW, root dry weight; SRA, seminal root angle; SW, seed weight; SL, seed length; GY, grain yield.

In order to identify the most useful markers for selecting for the root traits, extreme phenotypes were identified in the upper and lower 10th percentile of genotypes within the collection for each trait (Figure 4). Among the most significant MTAs for each trait, markers with different alleles between extreme genotypes were identified (Table 7, Figure 5). The frequency of the most common allele among genotypes from the upper 10th percentile ranged from 78% for RDW to 88% for SW, while for the lower 10th percentile it ranged from 65% for TRN and SRA to 92% for RDW (Figure 2).



**Figure 4.** Extreme phenotypes for SRA and TRN. The means correspond for 3 observational units of the genotype for SRA and 5 observational units of the genotype for TRN.



**Figure 5.** Marker allele frequency means from landraces within the upper and lower 10th percentile for the analyzed traits. All significant markers shown in Table 5 are included. TRN, total root number; RDW, root dry weight; SRA, seminal root angle; SW, seed weight; SL, seed length.

**Table 7.** Selected significant markers from the GWAS with different allele composition for the upper (UP) and lower (LOW) 10th percentile of genotypes. Different letters on the UP and LOW 10th phenotype indicate that means are significantly different at  $P < 0.01$  following a Tukey test.

Trait	Phenotype		Marker	Chr	Position	R <sup>2</sup>	Most Frequent Allele				
	Mean	UP 10th					LOW 10th	UP	LOW	Freq	
TRN (N)	4.4	5.2 <sup>a</sup>	3.4 <sup>b</sup>	1B	78.1	0.09	T	0.70	C	0.80	
RDW (g)	0.012	0.016 <sup>a</sup>	0.007 <sup>b</sup>	Excalibur_c8613_1266	5B	98.3	0.12	T	1.00	Y	0.50
				Tdurum_contig12995_792	1A	95.2	0.14	A	0.60	A	0.90
				RAC875_c8482_160	1A	95.5	0.11	T	0.60	C	0.70
				GENE-0392_97	1A	104.1	0.15	T	1.00	C	1.00
SRA (°)	98.65	119.98 <sup>a</sup>	69.88 <sup>b</sup>	RAC875_c39401_502	1B	95.5	0.15	G	1.00	A	1.00
				BS00023071_51	6A	79.1	0.14	C	0.70	T	1.00
				Kukri_c14877_303	4B	70.9	0.11	A	0.60	C	0.70
				Kukri_c34453_332	5B	51.0	0.11	T	1.00	C	0.60
SW (g)	0.487	0.59 <sup>a</sup>	0.37 <sup>b</sup>	BS00028183_51	2B	139.5	0.14	T	1.00	C	0.77
				RAC875_c37540_583	4B	77.7	0.11	Y	0.50	C	0.80
				Ku_c101046_1063	4D	80.4	0.12	C	1.00	A	0.60
				BS00094770_51	7D	161.1	0.12	T	1.00	T	0.70
SL (cm)	22.04	27.72 <sup>a</sup>	16.80 <sup>b</sup>	GENE-2220_165	1B	82.4	0.09	A	0.80	G	0.90
				Tdurum_contig8158_269	2A	177.6	0.09	C	1.00	C	0.60
				RFL_Contig5277_888	5A	130.9	0.09	T	0.80	C	0.90
				BobWhite_c6966_236	5B	48.3	0.08	T	0.90	Y	0.50
				BS00023161_51	6A	48.1	0.09	A	0.88	C	0.70
				RAC875_rep_c106439_1159							

Chr, chromosome. TRN, total root number. RDW, root dry weight. SRA, seminal root angle. SW, seed weight. SL, shoot length.

### 3.3. Gene Annotation

As reported in Supplementary Materials, Table S4, a total of 1489 gene models were identified within the 15 QTL hotspots using the high-confidence gene annotation from the wheat genome sequence [33]. Genetic distances were converted into physical distances using the position of common flanking markers on the genetic map [24] and the genome sequence. The number of gene models ranged from 224 in rootQTL\_2A.2 to 9 in rootQTL\_5B.1. Based on the high number of gene models, a selection was made according to gene families involved in root traits, growth and development, and abiotic stress resistance (Table 8). Thus, 31 gene families with a total of 96 gene models remained for subsequent analysis. Among them, F-box and zinc finger family proteins were identified in 12 of the 15 QTL hotspots, whereas 10 gene families were present in only one QTL hotspot. Among chromosomes with QTL hotspots, chromosome 2A had the highest number of gene models (22), whereas chromosomes 5A and 5B had the lowest number (4).

**Table 8.** Selected gene model families.

Description	N QTL Hotspots	Function
F-box family protein	12	Salt and drought stress responses
RING/FYVE/PHD zinc finger protein	12	Salt and drought stress responses
MYB-related transcription factor	8	Salt and drought stress responses
NAC domain-containing proteins	8	Induced by biotic and abiotic stresses
Cytochrome P450 family protein	5	Involved in seed size
BZIP transcription factor	5	Regulated by abiotic stress
Ethylene-responsive transcription factor	4	Induced by biotic and abiotic stresses
Calmodulin	4	Heat shock transduction pathway
Peroxidase	4	Root growth
ABC transporter	4	Control root development
Nucleoside triphosphate hydrolase	3	Associated with drought stress
E3 ubiquitin-protein ligase	3	Associated with drought stress
Glycine-rich protein	2	Enhance drought stress tolerance
Xyloglucan endotransglucosylase/hydrolase	2	Response dehydration, salinity, cold
Aquaporin	2	Drought stress tolerance
Expansin protein	2	Drought tolerance in wheat
Trihelix transcription factor	2	Stomatal development, drought
VQ motif family protein	2	Involved in seed size
Heat shock family protein	2	Induced by abiotic stress
Protein root UVB sensitive 6	2	Early seedling morphogenesis
SAUR-like auxin-responsive family protein	2	Maintain growth during abiotic stress
Bax inhibitor-1 family protein	1	Tolerance to abiotic stresses
Formin-like protein	1	Structure organization in drought-stressed plants
Late embryogenesis abundant protein	1	Participate in drought response
Cell wall invertase	1	Downregulated by drought
Senescence regulator	1	Related to drought stress
Plastid-lipid associated protein PAP/fibrillin	1	Induced by drought
Protein STAY-GREEN LIKE, chloroplastic	1	Improves drought resistance
PI-PLC X domain-containing protein	1	Induced by abiotic stresses
Histidine-containing phosphotransfer protein	1	Enhance tolerance to drought stress
Phospholipase D	1	Enhance drought stress tolerance

## 4. Discussion

Breeding for drought adaptation is one of the main challenges to be addressed in the coming years in order to increase wheat production and ensure sufficient food supply in the current scenario of climate change. Roots are crucial in this adaptation, as they are responsible for water and nutrient uptake. The wide morphological plasticity of the root system to different soil conditions and the role of root traits in drought environments are well known [34,35]. Wheat roots reduce their growth in water-limited conditions but increase the water uptake rate, extracting the water from deep soil

layers [36]. The shape and spatial arrangement of the RSA can provide a growth advantage and increasing yield performance during periods of water scarcity [37]. Thus, it is necessary to increase the knowledge of the genetics of root architecture in order to improve wheat yield stability under stress conditions by introgressing favorable alleles through breeding programs.

The current study evaluated root-related traits in a collection of Mediterranean bread wheat landraces representative of the variability existing for the species in the Mediterranean Basin [23] with the aim of providing QTL information for these traits regarding seminal roots. Seminal roots are important for early vigor and crop establishment in dryland areas because they explore the soil for nutrients and water [38]. Moreover, it has been reported that under drought stress, seminal roots activity is more important than that of nodal roots [39]. Additionally, field phenotyping of hundreds of genotypes is a complex and expensive task. As the root geometry of adult plants is strongly related to the SRA [5], it may be assumed that genotypes that differ in root architecture at an early developmental stage would also differ in the field at later growth stages, when nutrient and/or water capture become critical for yield performance [8].

The range of variation for the traits analyzed in the present study (from 10.9% for TRN to 18.8% for RDW) is in agreement with those reported for elite durum wheat cultivars by Canè et al. [8], who explained this variability as an adaptive value for the environmental conditions of the region of origin of the cultivars. Moreover, the high repeatability found for the traits supports the approach followed to analyze the seminal roots under controlled conditions.

Landraces from the eastern Mediterranean Basin showed the widest SRA, the lowest SW, the longest SL, and the lowest number of roots. According to previous studies in durum wheat [18,40], landraces from southeastern Mediterranean countries corresponding to the warmest and driest areas of the Mediterranean Basin, reported more grains per unit area and lighter grains than those developed in cooler and wetter zones of the region. Although it has been reported that in water-limited environments a vigorous root system could have benefits at the beginning of the growing season because it offers a more efficient water capture [41], no significant differences were observed for RDW among the SPs in the current study. Moreover, our results for SRA are in agreement with those reported by Roselló et al. [18], who found that durum wheat landraces from the eastern Mediterranean have the widest root angle, which probably allows them to cover a larger soil area and be more efficient in water uptake than landraces that originated in wetter areas.

Although not significant, probably due to the very early stage when the root traits were measured, the correlation between SRA and SW was negative. The same result was also reported by Canè et al. [8], who suggested that it could be due to the influence of the root angle on the distribution of the roots on soil layers and, therefore, the water uptake from deeper layers. On the other hand, the correlation between RDW and SW was positive, in agreement with the findings of Fang et al. [42], thus indicating the effectiveness of greater root mass for obtaining more soil water for plant growth and grain filling in drought. Seedling growth has also been related to SW in wheat [43]. The vertical distribution of the root system can have a strong effect on yield [44], so mass root concentrated in upper layers can be more effective for resource capture, while roots in deeper layers have more access to deep water.

The complexity of the genetic control of root traits was confirmed with 135 marker-trait associations identified in the current study. Their distribution across genomes was similar in the A and B genomes (46% and 48%, respectively), leaving only 6% of MTAs in the D genome. These results agree with the lower genetic diversity and higher LD found in the D genome, as reported previously [23]. According to Chao et al. [45], the different levels of diversity in wheat genomes could be due to different rates of gene flow from the ancestors of wheat, since polyploidy bottleneck resulting from speciation reduced diversity and increased the levels of LD in the D genome in comparison with the A and B genomes.

In order to simplify and to integrate closely linked MTAs in a consensus region, QTL hotspots were identified based on the results of LD decay reported in [23]. LD decay was used to define the CIs for the QTL hotspots. Following this approach, 43 MTAs were grouped in 15 QTL hotspots. The genomic position of QTL hotspots was compared with previous studies reporting meta-QTLs

for root traits [46] and MTAs from GWAS studies in order to detect previously identified regions controlling root traits. Among the 15 QTL hotspots, only rootQTL6A.3 was located in the same region of a previously mapped meta-QTL, RootMQTL74 [46]. When compared with MTA-QTLs reported by [18] in durum wheat Mediterranean landraces, the QTL hotspot rootQTL6A.3 corresponded to the MTA-QTLs mtaq-6A.3 and mtaq-6A.6. This hotspot was also in the same region of a major SRA QTL identified by Alahmad et al. [47] and by a QTL controlling root growth angle identified by Maccaferri et al. [9], who also found a QTL for grain weight that is located in a common region with the hotspot rootQTL2A.2, which includes an MTA for SW. rootQTL3B.1 shared a common position with an MTA reported by Ayalew et al. [48] on chromosome 3B under stress conditions. rootQTL7A.1, including an MTA for RDW, was located in a similar position as MLM-RDWB-10 reported by Li et al. [49] and associated with RDW at the booting stage. Finally, no genomic regions were shared with the study carried out by Beyer et al. [50]. Only four of the 15 QTL hotspots identified in this work had been detected previously, suggesting the importance of wheat Mediterranean landraces for the identification of new loci controlling root-related traits.

As reported in previous studies, at early developmental stages [8,18] the co-location of MTAs for grain yield and root related traits within the same QTL hotspot suggests their pleiotropic effect, however, deeper analyses should be necessary to confirm it. In durum wheat elite cultivars, Canè et al. [8] found that 30% of the QTLs affecting root system architecture were included within QTLs for agronomic traits. More recently, Roselló et al. [18] using a collection of Mediterranean durum wheat landraces found that 45% of QTL hotspots for root related traits were mapped in similar regions to yield-related traits reported for the same collection of landraces.

From a breeding standpoint, exploiting genetic diversity from local landraces is a valuable approach for recovering and broadening allelic variation for traits of interest [19]. Therefore, identifying the genotypes showing the extreme phenotypes within the pool of Mediterranean landraces and the associated markers provide the opportunity for introgressing suitable traits in elite cultivars by marker-assisted breeding using the most recent technologies to speed the process.

The availability of a high-quality reference wheat genome sequence [33] enabled us to quickly identify gene models corresponding to QTLs. Thus, the genetic position of the CIs of the QTL hotspots was projected into physical distances on the reference sequence to search for putative candidate gene models. To narrow the number of candidates, only gene models involved in the development and abiotic stress according to the literature were taken into consideration. Therefore, of 1489 gene models identified within the 15 QTL hotspots, only 31 gene families were selected.

F-box and zinc finger family proteins were the most represented, each one appearing in 12 hotspots. F-box proteins play important roles in plant development and abiotic stress responses via the ubiquitin pathway [51] and the ABA signaling pathway [52]. In wheat, the F-box protein *TaFBA1* is involved in plant hormone signaling and response to abiotic stresses and is expressed in all plant organs, including roots [53]. The overexpression of *TaFBA1* in transgenic tobacco reported by Li et al. [54] to improve heat tolerance resulted in increased root length in the transgenic plants. Zinc finger proteins are involved in several processes, such as regulation of plant growth and development, and response to abiotic stresses [46]. In *Arabidopsis* and rice, they play a role in tolerance to drought and salt stresses [55], while in wheat the overexpression of *TaZFP34* enhances root-to-shoot ratio during plant adaptation to drying soil [56].

Other kinds of gene models found in a high number of QTL hotspots were *MYB* transcription factors and *NAC* domain-containing proteins, each of them presents in 8 hotspots. *MYB* domain-containing transcription factors are involved in salt and drought stress adaptation in wheat. Some examples in wheat are the genes *TaMyb1*, *TaMYBsdu1*, and *TaMYB33*. The expression of *TaMyb1* in roots is strongly related to responses to abiotic stresses [57]. The gene *TaMYBsdu1* was found to be upregulated in leaves and roots of wheat under long-term drought stress [58]. Finally, the overexpression of *TaMYB33* in *Arabidopsis* enhances tolerance to drought and salt stresses [59]. *NAC* domain-containing proteins have been described to play many important roles in abiotic stress adaptation [46]. Xie et al. [60]

reported that *NAC1* promoted the development of lateral roots. Similarly, He et al. [61] found that the expression of *AtNAC2* in response to salt stress led to an increase in the development of lateral roots. Xia et al. [62] demonstrated that the gene *TaNAC4* is a transcriptional activator involved in wheat's response to biotic and abiotic stresses.

Proteins belonging to the cytochrome *P450* family and *bZIP* transcription factors were present in five QTL hotspots. The first class of proteins belongs to one of the largest families of plant proteins, with genes affecting important traits for crop improvement such as *TaCYP78A3*, which is involved in the control of seed size [63]. *bZIP* transcription factors are involved in abiotic stress response [64]. In *Arabidopsis*, it has been observed that the overexpression of *TabZIP14-B*, involved in salt and freezing tolerance, hindered root growth in transgenic plants in comparison with the control plants [65].

Other proteins involved in root growth and development are the peroxidases and *ABC* transporters that were identified in four QTL hotspots. Extracellular peroxidases are involved in plant defense reactions against biotic and abiotic stresses through the generation of reactive oxygen species in wounded root cells [66]. In *Arabidopsis*, the *ABC* transporter *AtPGP4* is expressed mainly during early root development, and its loss of function enhances lateral root initiation and root hair development [67]. Gaedeke et al. [68] reported a new member of the *ABC* transporter superfamily of *Arabidopsis thaliana*, *AtMRP5*. Using reverse genetics, these authors found that the recessive allele *mrp5* exhibited decreased root growth and increased lateral root formation. In addition to peroxidases and *ABC* transporters, other proteins identified in four QTLs were the ethylene-responsive transcription factors (ERFs), found to be involved in the response to abiotic stresses. In wheat, the *ERF TaERFL1a* is induced in wheat seedlings in response to salt, cold, and water deficiency [69].

Other family proteins involved in drought stress, seed size, or early development were represented in a lower number of QTL hotspots. Among them, aquaporins are known to affect drought tolerance influencing the capacity of roots to take up the soil water [70]. The expansins were suggested to be involved in root development, as the overexpression of the wheat expansin *TaEXPB23* improved drought tolerance by stimulating the growth of the root system in tobacco [71].

## 5. Conclusions

The exploitation of unexplored genetic variation present in local landraces can potentially contribute to breeding programs aimed at enhancing drought tolerance in wheat. Roots are crucial for adaptation to drought stress because they are the plant organ responsible for water and nutrient uptake and interaction with soil microbes. Thus, designing and developing novel root system ideotypes could be one of the targets of wheat breeding for the coming years. The variability found in the Mediterranean wheat landraces together with the newly identified QTL hotspots shows landraces as a valuable source of favorable root traits to introgress into adapted phenotypes through marker-assisted breeding. Among the different marker trait associations, those reported in extreme genotypes could result as a starting point to develop new mapping populations to fine map the corresponding traits.

**Supplementary Materials:** Supplementary materials can be found at <http://www.mdpi.com/2073-4395/10/5/613/s1>. Figure S1: Field scheme of the experimental design for grain yield, Table S1: List of accessions, Table S2: Significant GWAS results for root related traits, Table S3: Significant GWAS results for grain yield, Table S4: Gene models identified within the 15 root QTL hotspots.

**Author Contributions:** Conceptualization, S.S. and J.M.S.; Data curation, R.R.; Formal analysis, R.R.; Funding acquisition, C.R. and J.M.S.; Investigation, R.R., S.S. and J.M.S.; Methodology, R.R., S.S. and J.M.S.; Project administration, C.R. and J.M.S.; Supervision, S.S. and J.M.S.; Writing—original draft, R.R.; Writing—review & editing, S.S., C.R. and J.M.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was funded by projects AGL2015-65351-R and RTA2015-00072-C03-01 of the Spanish Ministry of Economy and Competitiveness. R.R. is a recipient of a PhD grant from the Spanish Ministry of Economy and Competitiveness.

**Acknowledgments:** The authors acknowledge the contribution of the CERCA programme (Generalitat de Catalunya). Thanks are given to the group of Agricultural Genetics of University of Bologna for technical support.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

EM	Eastern Mediterranean
GWAS	Genome wide association study
GY	Grain yield
MTA	Marker-trait association
NM	Northern Mediterranean
QTL	Quantitative trait locus
RDW	Root dry weight
RSA	Root system architecture
SL	Shoot length
SP	Sub-population
SRA	Seminal root angle
SW	Seed weight
TRN	Total root number
WM	Western Mediterranean

## References

1. Leegood, R.C.; Evans, J.R.; Furbank, R.T. Food security requires genetic advances to increase farm yields. *Nature* **2010**, *464*, 831. [[CrossRef](#)]
2. Giorgi, F.; Lionello, P. Climate change projections for the Mediterranean region. *Glob. Planet. Chang.* **2008**, *63*, 90–104. [[CrossRef](#)]
3. Uga, Y.; Kitomi, Y.; Ishikawa, S.; Yano, M. Genetic improvement for root growth angle to enhance crop production. *Breed. Sci.* **2015**, *65*, 111–119. [[CrossRef](#)] [[PubMed](#)]
4. Kulkarni, M.; Soolanayakanahally, R.; Ogawa, S.; Uga, Y. Drought Response in Wheat: Key Genes and Regulatory Mechanisms Controlling Root System Architecture and Transpiration Efficiency. *Front. Chem.* **2017**, *5*, 1–13. [[CrossRef](#)] [[PubMed](#)]
5. Manschadi, A.M.; Hammer, G.L.; Christopher, J.T.; de Voil, P. Genotypic variation in seedling root architectural traits and implications for drought adaptation in wheat (*Triticum aestivum* L.). *Plant Soil* **2008**, *303*, 115–129. [[CrossRef](#)]
6. Sanchez-Garcia, M.; Álvaro, F.; Martín-Sánchez, J.A.; Sillero, J.C.; Escribano, J.; Royo, C. Breeding effects on the genotype×environment interaction for yield of bread wheat grown in Spain during the 20th century. *Field Crop. Res.* **2012**, *126*, 79–86. [[CrossRef](#)]
7. Royo, C.; Maccaferri, M.; Álvaro, F.; Moragues, M.; Sanguineti, M.C.; Tuberosa, R.; Maalouf, F.; del Moral, L.F.G.; Demontis, A.; Rhouma, S.; et al. Understanding the relationships between genetic and phenotypic structures of a collection of elite durum wheat accessions. *Field Crop. Res.* **2010**, *119*, 91–105. [[CrossRef](#)]
8. Canè, M.A.; Maccaferri, M.; Nazemi, G.; Salvi, S.; Francia, R.; Colalongo, C.; Tuberosa, R. Association mapping for root architectural traits in durum wheat seedlings as related to agronomic performance. *Mol. Breed.* **2014**, *34*, 1629–1645. [[CrossRef](#)]
9. Maccaferri, M.; El-Feki, W.; Nazemi, G.; Salvi, S.; Canè, M.A.; Colalongo, M.C.; Stefanelli, S.; Tuberosa, R. Prioritizing quantitative trait loci for root system architecture in tetraploid wheat. *J. Exp. Bot.* **2016**, *67*, 1161–1178. [[CrossRef](#)]
10. Bhatta, M.; Morgounov, A.; Belamkar, V.; Baenziger, P. Genome-Wide Association Study Reveals Novel Genomic Regions for Grain Yield and Yield-Related Traits in Drought-Stressed Synthetic Hexaploid Wheat. *Int. J. Mol. Sci.* **2018**, *19*, 3011. [[CrossRef](#)]
11. Reynolds, M.P.; Mujeeb-Kazi, A.; Sawkins, M. Prospects for utilising plant-adaptive mechanisms to improve wheat and other crops in drought- and salinity-prone environments. *Ann. Appl. Biol.* **2005**, *146*, 239–259. [[CrossRef](#)]
12. Sanguineti, M.C.; Li, S.; Maccaferri, M.; Corneti, S.; Rotondo, F.; Chiari, T.; Tuberosa, R. Genetic dissection of seminal root architecture in elite durum wheat germplasm. *Ann. Appl. Biol.* **2007**, *151*, 291–305. [[CrossRef](#)]

13. Wasson, A.P.; Richards, R.A.; Chatrath, R.; Misra, S.C.; Prasad, S.V.S.; Rebetzke, G.J.; Kirkegaard, J.A.; Christopher, J.; Watt, M. Traits and selection strategies to improve root systems and water uptake in water-limited wheat crops. *J. Exp. Bot.* **2012**, *63*, 3485–3498. [[CrossRef](#)] [[PubMed](#)]
14. Liu, P.; Jin, Y.; Liu, J.; Liu, C.; Yao, H.; Luo, F.; Guo, Z.; Xia, X.; He, Z. Genome-wide association mapping of root system architecture traits in common wheat (*Triticum aestivum* L.). *Euphytica* **2019**, *215*, 1–12. [[CrossRef](#)]
15. Chochois, V.; Voge, J.P.; Rebetzke, G.J.; Watt, M. Variation in adult plant phenotypes and partitioning among seed and stem-borne roots across *Brachypodium distachyon* accessions to exploit in breeding cereals for well-watered and drought environments. *Plant Physiol.* **2015**, *168*, 953–967. [[CrossRef](#)]
16. Mace, E.S.; Singh, V.; van Oosterom, E.J.; Hammer, G.L.; Hunt, C.H.; Jordan, D.R. QTL for nodal root angle in sorghum (*Sorghum bicolor* L. Moench) co-locate with QTL for traits associated with drought adaptation. *Theor. Appl. Genet.* **2012**, *124*, 97–109. [[CrossRef](#)]
17. Ruiz, M.; Giraldo, P.; González, J.M. Phenotypic variation in root architecture traits and their relationship with eco-geographical and agronomic features in a core collection of tetraploid wheat landraces (*Triticum turgidum* L.). *Euphytica* **2018**, *214*, 54. [[CrossRef](#)]
18. Roselló, M.; Royo, C.; Sanchez-Garcia, M.; Soriano, J.M. Genetic Dissection of the Seminal Root System Architecture in Mediterranean Durum Wheat Landraces by Genome-Wide Association Study. *Agronomy* **2019**, *9*, 364. [[CrossRef](#)]
19. Lopes, M.S.; El-Basyoni, I.; Baenziger, P.S.; Singh, S.; Royo, C.; Ozbek, K.; Aktas, H.; Ozer, E.; Ozdemir, F.; Manickavelu, A.; et al. Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change. *J. Exp. Bot.* **2015**, *66*, 3477–3486. [[CrossRef](#)]
20. Nazco, R.; Peña, R.J.; Ammar, K.; Villegas, D.; Crossa, J.; Royo, C. Durum wheat (*Triticum durum* Desf.) Mediterranean landraces as sources of variability for allelic combinations at Glu-1/Glu-3 loci affecting gluten strength and pasta cooking quality. *Genet. Resour. Crop Evol.* **2014**, *61*, 1219–1236. [[CrossRef](#)]
21. Moragues, M.; Del Moral, L.F.G.; Moralejo, M.; Royo, C. Yield formation strategies of durum wheat landraces with distinct pattern of dispersal within the Mediterranean basin I: Yield components. *Field Crop. Res.* **2006**, *95*, 194–205. [[CrossRef](#)]
22. Roselló, M.; Villegas, D.; Álvaro, F.; Soriano, J.M.; Lopes, M.S.; Nazco, R.; Royo, C. Unravelling the relationship between adaptation pattern and yield formation strategies in Mediterranean durum wheat landraces. *Eur. J. Agron.* **2019**, *107*, 43–52. [[CrossRef](#)]
23. Rufo, R.; Alvaro, F.; Royo, C.; Soriano, J.M. From landraces to improved cultivars: Assessment of genetic diversity and population structure of Mediterranean wheat using SNP markers. *PLoS ONE* **2019**, *14*, e0219867. [[CrossRef](#)] [[PubMed](#)]
24. Wang, S.; Wong, D.; Forrest, K.; Allen, A.; Chao, S.; Huang, B.E.; Maccaferri, M.; Salvi, S.; Milner, S.G.; Cattivelli, L.; et al. Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnol. J.* **2014**, *12*, 787–796. [[CrossRef](#)] [[PubMed](#)]
25. Abramoff, M.D.; Magalhães, P.J.; Ram, S.J. *Image Processing with ImageJ*, 2nd ed.; Packt Publishing: Birmingham, UK, 2004; Volume 11, ISBN 9781785889837.
26. Harper, D.G.C. Some comments on the repeatability of measurements. *Ringing Migr.* **1994**, *15*, 84–90. [[CrossRef](#)]
27. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [[CrossRef](#)]
28. Barrett, J.C.; Fry, B.; Maller, J.; Daly, M.J. Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* **2005**, *21*, 263–265. [[CrossRef](#)]
29. Wang, S.-X.; Zhu, Y.-L.; Zhang, D.-X.; Shao, H.; Liu, P.; Hu, J.-B.; Zhang, H.; Zhang, H.-P.; Chang, C.; Lu, J.; et al. Genome-wide association study for grain yield and related traits in elite wheat varieties and advanced lines using SNP markers. *PLoS ONE* **2017**, *12*, e0188662. [[CrossRef](#)]
30. Mangini, G.; Gadaleta, A.; Colasuonno, P.; Marcotuli, I.; Signorile, A.M.; Simeone, R.; De Vita, P.; Mastrangelo, A.M.; Laidò, G.; Pecchioni, N.; et al. Genetic dissection of the relationships between grain yield components by genome-wide association mapping in a collection of tetraploid wheats. *PLoS ONE* **2018**, *13*, e0190162. [[CrossRef](#)]
31. Condorelli, G.E.; Maccaferri, M.; Newcomb, M.; Andrade-Sanchez, P.; White, J.W.; French, A.N.; Sciara, G.; Ward, R.; Tuberosa, R. Comparative Aerial and Ground Based High Throughput Phenotyping for the Genetic Dissection of NDVI as a Proxy for Drought Adaptive Traits in Durum Wheat. *Front. Plant Sci.* **2018**, *9*. [[CrossRef](#)] [[PubMed](#)]

32. Sukumaran, S.; Reynolds, M.P.; Sansaloni, C. Genome-Wide Association Analyses Identify QTL Hotspots for Yield and Component Traits in Durum Wheat Grown under Yield Potential, Drought, and Heat Stress Environments. *Front. Plant Sci.* **2018**, *9*, 81. [[CrossRef](#)] [[PubMed](#)]
33. The International Wheat Genome Sequencing Consortium (IWGSC); Appels, R.; Eversole, K.; Stein, N.; Feuillet, C.; Keller, B.; Rogers, J.; Pozniak, C.J.; Choulet, F.; Distelfeld, A.; et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **2018**, *361*, eaar7191. [[CrossRef](#)] [[PubMed](#)]
34. Christopher, J.; Christopher, M.; Jennings, R.; Jones, S.; Fletcher, S.; Borrell, A.; Manschadi, A.M.; Jordan, D.; Mace, E.; Hammer, G. QTL for root angle and number in a population developed from bread wheats (*Triticum aestivum*) with contrasting adaptation to water-limited environments. *Theor. Appl. Genet.* **2013**, *126*, 1563–1574. [[CrossRef](#)] [[PubMed](#)]
35. Paez-Garcia, A.; Motes, C.; Scheible, W.-R.; Chen, R.; Blancaflor, E.; Monteros, M. Root Traits and Phenotyping Strategies for Plant Improvement. *Plants* **2015**, *4*, 334–355. [[CrossRef](#)] [[PubMed](#)]
36. Asseng, S.; Ritchie, J.T.; Smucker, A.J.M.; Robertson, M.J. Root growth and water uptake during water deficit and recovering in wheat. *Plant Soil* **1998**, *201*, 265–273. [[CrossRef](#)]
37. Rogers, E.D.; Benfey, P.N. Regulation of plant root system architecture: Implications for crop advancement. *Curr. Opin. Biotechnol.* **2015**, *32*, 93–98. [[CrossRef](#)]
38. Reynolds, M.; Tuberosa, R. Translational research impacting on crop productivity in drought-prone environments. *Curr. Opin. Plant Biol.* **2008**, *11*, 171–179. [[CrossRef](#)]
39. Manschadi, A.M.; Christopher, J.T.; Hammer, G.L.; Devoil, P. Experimental and modelling studies of drought-adaptive root architectural traits in wheat (*Triticum aestivum* L.). *Plant Biosyst. Int. J. Deal. Asp. Plant Biol.* **2010**, *144*, 458–462.
40. Royo, C.; Nazco, R.; Villegas, D. The climate of the zone of origin of Mediterranean durum wheat (*Triticum durum* Desf.) landraces affects their agronomic performance. *Genet. Resour. Crop Evol.* **2014**, *61*, 1345–1358. [[CrossRef](#)]
41. Liao, M.; Palta, J.A.; Fillery, I.R.P. Root characteristics of vigorous wheat improve early nitrogen uptake. *Aust. J. Agric. Res.* **2006**, *57*, 1097. [[CrossRef](#)]
42. Fang, Y.; Du, Y.; Wang, J.; Wu, A.; Qiao, S.; Xu, B.; Zhang, S.; Siddique, K.H.M.; Chen, Y. Moderate Drought Stress Affected Root Growth and Grain Yield in Old, Modern and Newly Released Cultivars of Winter Wheat. *Front. Plant Sci.* **2017**, *8*, 672. [[CrossRef](#)] [[PubMed](#)]
43. Aparicio, N.; Villegas, D.; Araus, J.L.; Blanco, R.; Royo, C. Seedling development and biomass as affected by seed size and morphology in durum wheat. *J. Agric. Sci.* **2002**, *139*, 143–150. [[CrossRef](#)]
44. King, J.; Gay, A.; Sylvester-Bradley, R.; Bingham, I.; Foulkes, J.; Gregory, P.; Robinson, D. Modelling Cereal Root Systems for Water and Nitrogen Capture: Towards an Economic Optimum. *Ann. Bot.* **2003**, *91*, 383–390. [[CrossRef](#)] [[PubMed](#)]
45. Chao, S.; Dubcovsky, J.; Dvorak, J.; Luo, M.-C.; Baenziger, S.P.; Matnyazov, R.; Clark, D.R.; Talbert, L.E.; Anderson, J.A.; Dreisigacker, S.; et al. Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). *BMC Genomics* **2010**, *11*, 727. [[CrossRef](#)] [[PubMed](#)]
46. Soriano, J.M.; Alvaro, F. Discovering consensus genomic regions in wheat for root-related traits by QTL meta-analysis. *Sci. Rep.* **2019**, *9*, 10537. [[CrossRef](#)]
47. Alahmad, S.; El Hassouni, K.; Bassi, F.M.; Dinglasan, E.; Youssef, C.; Quarry, G.; Aksoy, A.; Mazzucotelli, E.; Juhász, A.; Able, J.A.; et al. A major root architecture QTL responding to water limitation in durum wheat. *Front. Plant Sci.* **2019**, *10*, 436. [[CrossRef](#)]
48. Ayalew, H.; Liu, H.; Börner, A.; Kobiljski, B.; Liu, C.; Yan, G. Genome-Wide Association Mapping of Major Root Length QTLs Under PEG Induced Water Stress in Wheat. *Front. Plant Sci.* **2018**, *9*, 1759. [[CrossRef](#)]
49. Li, L.; Peng, Z.; Mao, X.; Wang, J.; Chang, X.; Reynolds, M.; Jing, R. Genome-wide association study reveals genomic regions controlling root and shoot traits at late growth stages in wheat. *Ann. Bot.* **2019**, *124*, 993–1006. [[CrossRef](#)]
50. Beyer, S.; Daba, S.; Tyagi, P.; Bockelman, H.; Brown-Guedira, G.; Mohammadi, M. Loci and candidate genes controlling root traits in wheat seedlings—A wheat root GWAS. *Funct. Integr. Genomics* **2019**, *19*, 91–107. [[CrossRef](#)]

51. Jia, Y.; Gu, H.; Wang, X.; Chen, Q.; Shi, S.; Zhang, J.; Ma, L.; Zhang, H.; Ma, H. Molecular cloning and characterization of an F-box family gene CarF-box1 from chickpea (*Cicer arietinum* L.). *Mol. Biol. Rep.* **2012**, *39*, 2337–2345. [[CrossRef](#)] [[PubMed](#)]
52. Koops, P.; Pelsler, S.; Ignatz, M.; Klose, C.; Marrocco-Selden, K.; Kretsch, T. EDL3 is an F-box protein involved in the regulation of abscisic acid signalling in *Arabidopsis thaliana*. *J. Exp. Bot.* **2011**, *62*, 5547–5560. [[CrossRef](#)] [[PubMed](#)]
53. Zhou, S.; Sun, X.; Yin, S.; Kong, X.; Zhou, S.; Xu, Y.; Luo, Y.; Wang, W. The role of the F-box gene *TaFBA1* from wheat (*Triticum aestivum* L.) in drought tolerance. *Plant Physiol. Biochem.* **2014**, *84*, 213–223. [[CrossRef](#)] [[PubMed](#)]
54. Li, Q.; Wang, W.; Wang, W.; Zhang, G.; Liu, Y.; Wang, Y.; Wang, W. Wheat F-Box Protein Gene *TaFBA1* Is Involved in Plant Tolerance to Heat Stress. *Front. Plant Sci.* **2018**, *9*. [[CrossRef](#)] [[PubMed](#)]
55. Xu, D.-Q.; Huang, J.; Guo, S.-Q.; Yang, X.; Bao, Y.-M.; Tang, H.-J.; Zhang, H.-S. Overexpression of a TFIIIA-type zinc finger protein gene *ZFP252* enhances drought and salt tolerance in rice (*Oryza sativa* L.). *FEBS Lett.* **2008**, *582*, 1037–1043. [[CrossRef](#)]
56. Chang, H.; Chen, D.; Kam, J.; Richardson, T.; Drenth, J.; Guo, X.; McIntyre, C.L.; Chai, S.; Rae, A.L.; Xue, G.-P. Abiotic stress upregulated *TaZFP34* represses the expression of type-B response regulator and *SHY2* genes and enhances root to shoot ratio in wheat. *Plant Sci.* **2016**, *252*, 88–102. [[CrossRef](#)]
57. Lee, T.G.; Jang, C.S.; Kim, J.Y.; Kim, D.S.; Park, J.H.; Kim, D.Y.; Seo, Y.W. A Myb transcription factor (*TaMyb1*) from wheat roots is expressed during hypoxia: Roles in response to the oxygen concentration in root environment and abiotic stresses. *Physiol. Plant.* **2007**, *129*, 375–385. [[CrossRef](#)]
58. Rahaie, M.; Xue, G.-P.; Naghavi, M.R.; Alizadeh, H.; Schenk, P.M. A MYB gene from wheat (*Triticum aestivum* L.) is up-regulated during salt and drought stresses and differentially regulated between salt-tolerant and sensitive genotypes. *Plant Cell Rep.* **2010**, *29*, 835–844. [[CrossRef](#)]
59. Qin, Y.; Wang, M.; Tian, Y.; He, W.; Han, L.; Xia, G. Over-expression of *TaMYB33* encoding a novel wheat MYB transcription factor increases salt and drought tolerance in *Arabidopsis*. *Mol. Biol. Rep.* **2012**, *39*, 7183–7192. [[CrossRef](#)]
60. Xie, Q.; Frugis, G.; Colgan, D.; Chua, N.H. *Arabidopsis* NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. *Genes Dev.* **2000**, *14*, 3024–3036. [[CrossRef](#)]
61. He, X.-J.; Mu, R.-L.; Cao, W.-H.; Zhang, Z.-G.; Zhang, J.-S.; Chen, S.-Y. *AtNAC2*, a transcription factor downstream of ethylene and auxin signaling pathways, is involved in salt stress response and lateral root development. *Plant J.* **2005**, *44*, 903–916. [[CrossRef](#)] [[PubMed](#)]
62. Xia, N.; Zhang, G.; Liu, X.-Y.; Deng, L.; Cai, G.-L.; Zhang, Y.; Wang, X.-J.; Zhao, J.; Huang, L.-L.; Kang, Z.-S. Characterization of a novel wheat NAC transcription factor gene involved in defense response against stripe rust pathogen infection and abiotic stresses. *Mol. Biol. Rep.* **2010**, *37*, 3703–3712. [[CrossRef](#)] [[PubMed](#)]
63. Ma, M.; Wang, Q.; Li, Z.; Cheng, H.; Li, Z.; Liu, X.; Song, W.; Appels, R.; Zhao, H. Expression of *TaCYP78A3*, a gene encoding cytochrome P450 CYP78A3 protein in wheat (*Triticum aestivum* L.), affects seed size. *Plant J.* **2015**, *83*, 312–325. [[CrossRef](#)] [[PubMed](#)]
64. Sornaraj, P.; Luang, S.; Lopato, S.; Hrmova, M. Basic leucine zipper (bZIP) transcription factors involved in abiotic stresses: A molecular model of a wheat bZIP factor and implications of its structure in function. *Biochim. Biophys. Acta Gen. Subj.* **2016**, *1860*, 46–56. [[CrossRef](#)] [[PubMed](#)]
65. Zhang, L.; Zhang, L.; Xia, C.; Gao, L.; Hao, C.; Zhao, G.; Jia, J.; Kong, X. A Novel Wheat C-bZIP Gene, *TabZIP14-B*, Participates in Salt and Freezing Tolerance in Transgenic Plants. *Front. Plant Sci.* **2017**, *8*. [[CrossRef](#)] [[PubMed](#)]
66. Minibayeva, F.V.; Gordon, L.K.; Kolesnikov, O.P.; Chasov, A.V. Role of extracellular peroxidase in the superoxide production by wheat root cells. *Protoplasma* **2001**, *217*, 125–128. [[CrossRef](#)] [[PubMed](#)]
67. Santelia, D.; Vincenzetti, V.; Azzarello, E.; Bovet, L.; Fukao, Y.; Düchtig, P.; Mancuso, S.; Martinoia, E.; Geisler, M. MDR-like ABC transporter *AtPGP4* is involved in auxin-mediated lateral root and root hair development. *FEBS Lett.* **2005**, *579*, 5399–5406. [[CrossRef](#)]
68. Gaedeke, N.; Klein, M.; Kolukisaoglu, U.; Forestier, C.; Müller, A.; Ansoerge, M.; Becker, D.; Mammun, Y.; Kuchler, K.; Schulz, B.; et al. The *Arabidopsis thaliana* ABC transporter *AtMRP5* controls root development and stomata movement. *EMBO J.* **2001**, *20*, 1875–1887. [[CrossRef](#)]
69. Gao, T.; Li, G.-Z.; Wang, C.-R.; Dong, J.; Yuan, S.-S.; Wang, Y.-H.; Kang, G.-Z. Function of the ERFL1a Transcription Factor in Wheat Responses to Water Deficiency. *Int. J. Mol. Sci.* **2018**, *19*, 1465. [[CrossRef](#)]

70. Javot, H. The Role of Aquaporins in Root Water Uptake. *Ann. Bot.* **2002**, *90*, 301–313. [[CrossRef](#)]
71. Li, A.X.; Han, Y.Y.; Wang, X.; Chen, Y.H.; Zhao, M.R.; Zhou, S.M.; Wang, W. Root-specific expression of wheat expansin gene TaEXPB23 enhances root growth and water stress tolerance in tobacco. *Environ. Exp. Bot.* **2015**, *110*, 73–84. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



Article

# Improvement of a RD6 Rice Variety for Blast Resistance and Salt Tolerance through Marker-Assisted Backcrossing

Korachan Thanasilungura <sup>1</sup>, Sukanya Kranto <sup>2</sup>, Tidarat Monkham <sup>1</sup>, Sompong Chankaew <sup>1</sup> and Jirawat Sanitchon <sup>1,\*</sup>

<sup>1</sup> Department of Agronomy, Faculty of Agriculture, Khon Kaen University, Khon Kaen 40002, Thailand; Chun.sn@gmail.com (K.T.); tidamo@kku.ac.th (T.M.); somchan@kku.ac.th (S.C.)

<sup>2</sup> Ratchaburi Rice Research Center, Muang, Ratchaburi 70000, Thailand; sukanya.me@hotmail.com

\* Correspondence: jirawat@kku.ac.th; Tel.: +66-81567-4364

Received: 24 July 2020; Accepted: 30 July 2020; Published: 1 August 2020

**Abstract:** RD6 is one of the most favorable glutinous rice varieties consumed throughout the north and northeast of Thailand because of its aroma and softness. However, blast disease and salt stress cause decreases in both yield quantity and quality during cultivation. Here, gene pyramiding via marker-assisted backcrossing (MAB) using combined blast resistance QTLs (*qBl 1, 2, 11, and 12*) and *Saltol* QTL was employed in solving the problem. To pursue our goal, the RD6 introgression line (RGD07005-12-165-1), containing four blast-resistant QTLs, were crossed with the Pokkali salt tolerant variety. Blast resistance evaluation was thoroughly carried out in the fields, from BC<sub>2</sub>F<sub>2,3</sub> to BC<sub>4</sub>F<sub>4</sub>, using the upland short-row and natural field infection methods. Additionally, salt tolerance was validated in both greenhouse and field conditions. We found that the RD6 “BC<sub>4</sub>F<sub>4</sub> 132-12-61” resulting from our breeding programme successfully resisted blast disease and tolerated salt stress, while it maintained the desirable agronomic traits of the original RD6 variety. This finding may provide a new improved rice variety to overcome blast disease and salt stress in Northeast Thailand.

**Keywords:** gene pyramiding; aroma; QTL; chromosome; selection; introgression line

## 1. Introduction

Rice (*Oryza sativa* L.) is consumed as a staple food in Asia, especially in the southeast region. In Thailand, the indica rice variety RD6 developed from KDML105 through gamma irradiation is one of the most favorable glutinous rice consumed throughout the northeast of Thailand [1,2]. Because of its cooking quality, aroma, and softness, production demand has increased over time. However, its yield of 4.16 ton/ha fails to meet its potential, due to biotic and abiotic stress.

Rice blast disease caused by the fungus *Pyricularia grisea* (Cooke) Sacc. leads to crop losses up to 85% of total yield [3]. Disease symptoms occur in all stages of plant growth, beginning with blast discoloration and wilting of the foliage [4]. Neck blast can be found at the flowering stage, accelerating plant death [5]. Severe damage was also observed within areas of intensive planting with high doses of nitrogen application [6]. Development of new rice varieties resistant to blast fungus is an alternative approach to diminish or control the invasion of this pathogen. The resistance quantitative trait loci (QTL) have been investigated to achieve parental varieties, which are further used for gene pyramiding in breeding programmes. Currently, more than 100 blast-resistant genes have been identified, of which 22 genes structures have been cloned [7]. In Thailand, few studies of blast resistant genes have been conducted. Noenplab et al. [8] studied the relationship of leaf blast and neck blast of resistant genes in the Jao Hom Nil (JHN) variety, in which the resistant QTLs were detected on chromosomes 1 and 11. The resistant QTLs conferred resistance to both leaf blast and neck blast. Suwannual et al. [9]

pyramided four blast-resistant QTLs, individually, on chromosomes 2 and 12 within the P0489 variety, and on chromosomes 1 and 11 carried by the JHN variety, resulting in the creation of new RD6 introgression lines. Their results demonstrated that the RD6 introgression lines carrying a high number of QTLs (achieved through pyramiding) reached a broader spectrum of blast resistance to the blast pathogens prevalent in the region.

In addition to rice blast fungus, salt stress is a crucial constraint for RD6 production. Thailand's northeast region is the country's largest rice-producing area, and it is comprised of two basins: Sakon Nakhon and Nakhon Ratchasima. In those basins containing an understructure of accumulated salt rock, the salt-affected range covers approximately 1.84 Mha [10]. Evaporation during the dry season tends to raise salinity from the subsoil to the surface, thereby increasing salinity intensity and increasing salt stress from 2–4 dS/m to 8–16 dS/m [10–12]. Rice is a salt-sensitive crop, capable of tolerating salinity at moderate levels of electrical conductivity (4–8 dS/m) [13]. Therefore, rice produced under rain-fed, lowland conditions is usually exposed to high levels of soil salinity. The RD6 variety, which is well known, was identified as a geographical indication (GI) within the Tung Gula Rong Hai of the Northeast, Thailand. Specifically, RD6 requires optimal soil salinity to enhance rice seed aroma [14]. However, an abundance of salinity can reduce rice plant growth, tiller number, and seed set-up [15], and the stress caused by excessive salt can significantly reduce total crop yield and result in plant death [16].

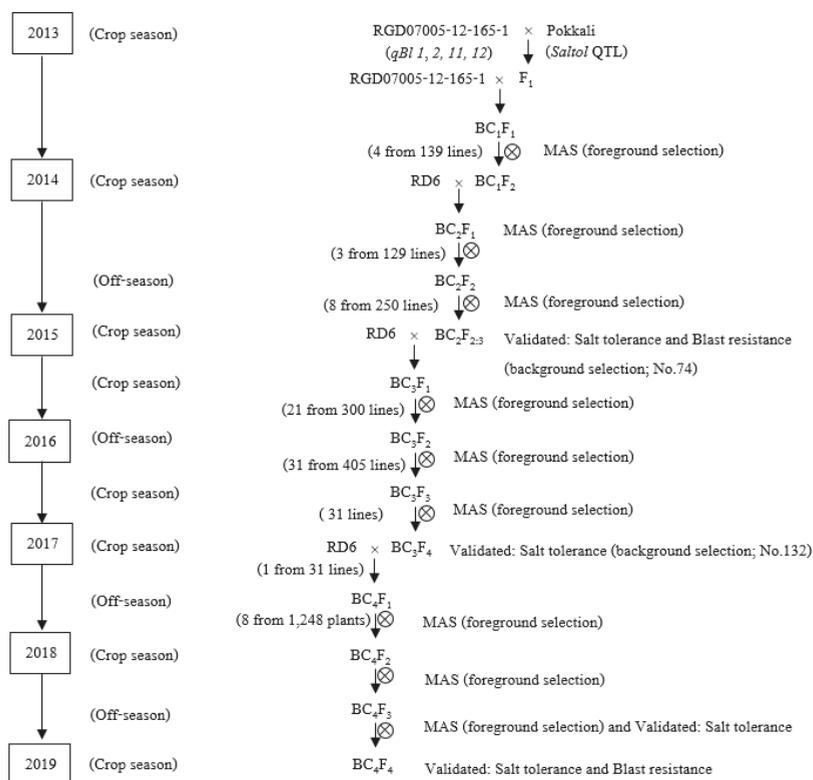
The Pokkali variety, derived from the International Rice Research Institute (IRRI), has become a well-known source of salinity tolerance worldwide, attributed to the salt-tolerant QTL located on rice chromosome 1 (*Saltol*) [17–20]. Therefore, several researchers have attempted to develop salt-tolerant rice varieties using the *Saltol* QTL [21–26].

The marker-assisted backcrossing (MAB) method has been employed to obtain beneficial QTLs from donor parents via introgression between the qualitative and quantitative traits from landraces and wild relatives [27] due to the precision method with shortened time frame in both foreground and background selection. MAB provides effective gene selection and/ or QTLs for pyramiding multi-genes/QTLs within the rice population. These benefits further support breeding practices for improved resistance and tolerance [28–32]. The objective of this study was to determine the blast resistance and salt tolerance levels within the RD6 introgression lines by pyramiding four blast-resistant and one salt-tolerant QTL into the RD6 rice variety in both greenhouse and field conditions.

## 2. Materials and Methods

### 2.1. Plant Materials and Marker-Assisted Backcrossing Selection (MABS)

Three parental varieties/lines were used to generate the BC<sub>4</sub>F<sub>4</sub> population, representing a pseudo-backcrossing approach to increasing the recurrent genetic background of a pyramiding population, comprised of the RD6 (recurrent parent), Pokkali (obtained for the *saltol* QTL present on chromosome 1), and RGD07005-12-165-1 (the RD6 near-isogenic line obtained from the Rice Gene Discovery Unit, Kasetsart University, Thailand). The RGD07005-12-165-1 obtained four blast-resistant QTLs from the JHN and P0489 varieties on chromosomes 1 and 11, and chromosomes 2 and 12, respectively. The breeding program was subsequently improved within the population through MAB, by crossing the RGD07005-12-165-1 with the Pokkali variety to improve salt tolerance. The F<sub>1</sub> was then backcrossed with RGD07005-12-165-1 through BC<sub>1</sub>F<sub>1</sub>, whereas BC<sub>1</sub>F<sub>2</sub> was utilized as a marker-assisted selection (MAS) in the blast-resistant and salt-tolerant QTLs. In this step, the flanking marker RM3412/RM10748 was used for the selected *Saltol* QTL [33], RM319/RM212 and RM114/RM224 were used for selected blast-resistant QTLs on chromosomes 1 and 11, respectively [8], and RM48/RM207 and RM313/RM277 were used for selected blast-resistant QTLs on chromosomes 2 and 12, respectively [9], as shown in Figure 1.



**Figure 1.** Breeding schematics for the development and validation of the RD6 NILs populations.

Total genomic DNA from young leaves of individual plants, lines, and their parents was extracted according to the method described by Dellaporta et al. [34] with slight modifications. The PCR reactions for SSR markers were carried out in a volume of 10  $\mu$ L, containing 25 ng of genomic DNA, 1  $\times$  PCR buffer, 1.8 mM MgCl<sub>2</sub>, 0.2 mM dNTP, 0.2  $\mu$ M forward and reverse primer, and 0.05 unit Taq DNA polymerase. DNA amplification was performed in a DNA thermal cycle for five minutes at 95  $^{\circ}$ C, followed by 35 cycles of 30 s at 95  $^{\circ}$ C, 30 s at 55  $^{\circ}$ C, and two min at 72  $^{\circ}$ C, with a final extension of seven minutes at 72  $^{\circ}$ C. The amplification products were separated via 4.5% polyacrylamide gel electrophoresis [32]. The selected lines in the BC<sub>1</sub>, BC<sub>2</sub>, and BC<sub>3</sub> generations were backcrossed with the RD6 variety, using MAS for *saltol* and blast-resistant QTL selection. Trait qualities; such as glutinous type, aromatic character, and gelatinization temperature (GT) in the BC<sub>2</sub>F<sub>2</sub> populations were fixed through MAS using glutinous 23 primer on chromosome 6, badh2 on chromosome 8, and RM190 on chromosome 6 (Table S1). Each backcross generation within the BC<sub>4</sub>F<sub>4</sub> populations was evaluated for salt tolerance and blast resistance (Figure 1).

## 2.2. Evaluation of Salt Tolerance and Blast Resistance in the BC<sub>2</sub>F<sub>2,3</sub> Populations (Exp. 1)

The evaluation of salt tolerance and blast resistance of the BC<sub>2</sub>F<sub>2,3</sub> lines, as well as the parental and check varieties, were conducted at the Department of Agronomy, Faculty of Agriculture, Khon Kaen University, Khon Kaen, Thailand. The salt tolerance evaluation was performed through two methods, salt solution and artificial soil salinity. The salt solution method was laid out in a completely randomized design (CRD), with four replications. Seedlings were transplanted at seven days to 50  $\times$  57 cm<sup>2</sup> Styrofoam sheets with 1.5 cm diameter holes [17]. Fertilizer was applied with

Yoshida nutrient solution [35] at three to twenty-one days of age. NaCl was then added to reach the electrical conductivity (EC) of 4 dS/m. EC was subsequently increased at three-day intervals, reaching EC 8 and 12 dS/m. When the susceptible checks (IR29) presented salt injury symptoms, salt tolerance data were recorded following the standard evaluation score (SES) [36]. The artificial soil salinity method was also laid out in CRD with four replications. Seedlings were transplanted at seven days and transferred from the spent soil trays to the water trays. At 14 days after transplanting, the experimental trays were treated with NaCl, which adjusted the water solution to EC 8dS/m, which increased to EC 12dS/m after two days. Salt tolerance data were recorded similarly in both methods.

Blast resistance was also evaluated via the upland short-row method at the Sakon Nakhon Rice Research Center, Sakon Nakhon, Thailand. The experiment was laid out in CRD with three replications. Seeds of each BC<sub>2</sub>F<sub>2,3</sub> line were sown in rows (approximately) 50 cm long and 10 cm apart. A susceptible KDML105 variety was planted alternately with every two testing varieties. Blast resistance scores were recorded following the SES method [36].

### *2.3. Evaluation of Salt Tolerance in the BC<sub>3</sub>F<sub>4</sub> Populations (Exp. 2)*

The BC<sub>3</sub>F<sub>4</sub> lines and parental varieties were evaluated for salt tolerance in field conditions. The experiment was conducted at the Ban Daeng Village, Ban Fhang, Khon Kaen, Thailand. The experiment was laid out in a randomized complete block design (RCBD) with three replications. Germinated seeds were sown on seedbeds; then, at thirty days, the seedlings were transplanted to the field. Plot sizes were 1 × 1.5 m<sup>2</sup>, in three rows, spaced 25 × 25 cm between and within rows. The RD6 variety was planted between every five plots within the test lines to ensure that salinity occurred uniformly. Fertilizer (23.44 kg/ha of N, P<sub>2</sub>O<sub>5</sub>, and K<sub>2</sub>O) was applied at four days after transplanting, and hand weeding and chemical application for disease and insect control were performed as needed. When the susceptible check (RD6) presented salt injury symptoms, salt tolerance data were recorded following SES [37]. Moreover, the agronomic traits including 1000/seed weight, seed length, seed width, and seed shape were recorded.

### *2.4. Evaluation of Salt Tolerance and Blast Resistance Evaluations in the BC<sub>4</sub>F<sub>3</sub> Populations (Exp. 3)*

The salt tolerance evaluation of the BC<sub>4</sub>F<sub>3</sub> lines was conducted in greenhouse conditions at the Department of Agronomy, Faculty of Agriculture, Khon Kaen University, Khon Kaen, Thailand. The experiment was laid out in CRD with three replications, as described in the salt solution method of Exp. 1. The experiment for blast resistance of the BC<sub>4</sub>F<sub>3</sub> lines was conducted at the Department of Agronomy, Faculty of Agriculture, Khon Kaen University, Khon Kaen, Thailand through field and upland short-row experiments. The upland short-row method was conducted similarly to Exp. 1. The field experiment was laid out in RCBD with three replications, in plots 0.5 × 1 m<sup>2</sup>, in three rows, spaced 25 × 25 cm between and within rows. Seedlings were transplanted at 30 days of age. Symptoms of natural blast infection were identified when seedlings began to show signs of infection, following the protocol of the SES [37], in which both leaf and neck blast symptoms were recorded.

Agronomic trait data, such as plant height (PH) and panicle length (PL), were recorded at pre-harvest; whereas post-harvest data recorded 4/panicle seed weight (SW4P), 1000/seed weight (1000SW), total dry weight (TDW), total seed weight (TSW), harvest index (HI), seed length (SL), seed width (SW), and seed shape (SS) (ratio of SL/SW). Additionally, seed qualities of the BC<sub>4</sub>F<sub>4</sub> seeds, such as seed morphology, including SL, SW, SS and seed color of brown and paddy rice and aromatic traits, were evaluated and compared with the RD6 variety. The seed aromatic evaluation of each line was achieved through the quantitative determination of 2-acetyl-1-pyrroline (2AP) content using automated headspace gas chromatography following the methods as prescribed by Sriseadka et al. [38]. In brief, polished seed (1.00 g) were ground and then placed in a 20 mL headspace vial. The headspace vials were immediately sealed with PTFE/silicone septa and aluminum caps prior to analysis by static headspace-gas chromatography. A static headspace (Model 7697A, Agilent Technologies, Santa Clara, CA, USA) coupled to an Agilent 7890B Series GC system equipped with an Agilent

5977B GC/MSD system was used. A series of 2AP standard solutions with concentrations of 1.22, 2.45, 4.90, 9.79, and 19.67 ppm in isopropanol were prepared, which was added to headspace vials containing 1.00 g of non-aromatic rice seed (cv. Chai Nat 1) used as the external standard. The optimum headspace operating conditions were oven temperature 110 °C, loop temperature 120 °C, transfer line temperature 130 °C, vial equilibration time 10 min with high speed shaking, pressurizing time 0.15 min, loop equilibration time 0.40 min, and inject time 0.50 min. The headspace volatiles were separated using an HP-5 (25 m × 250 µm × 0.25 µm film thickness) column (J&W Scientific, Folsom, CA, USA). The optimum GC conditions were achieved using an HP-5 column with a splitless injection at 210 °C. The column temperature programme began at 50 °C and increased to 200 °C at 10 °C/min. Purified helium was used as the GC carrier gas at a flow rate of 1.2 mL/min. A calibration curve for 2AP analysis by headspace was generated by spiking known concentrations of 2AP into a non-fragrant rice variety (Chai Nat 1). Samples were run in triplicate, and the concentration of 2AP was calculated based upon the relative peak area of external standard.

#### *2.5. Evaluation of Salt Tolerance in the BC<sub>4</sub>F<sub>4</sub> Populations (Exp. 4)*

The salt tolerance evaluations of the BC<sub>4</sub>F<sub>4</sub> lines were conducted in greenhouse conditions at the Department of Agronomy, Faculty of Agriculture, Khon Kaen University, Khon Kaen, Thailand, via the salt solution method. Laid out in CRD with four replications, the planting methods and experiment protocol were similar to those described in Exp1, except that the EC in Exp. 4 was adjusted to 18 dS/m. The leaf, stem, root, and total dry weight data were recorded. Dry weight of the seedlings was determined by oven-drying the seedlings at 80 °C for 3 days, and the percentages of Na<sup>+</sup> and K<sup>+</sup> in the leaves, stems, and roots also were determined, according to the flame photometric method [39]. In addition, the percentages of Na<sup>+</sup> and K<sup>+</sup> in the leaves, stems, and roots also were determined, according to the flame photometric method [39]. In brief, pounding sample volume 0.5 g of each tissues were digest by 10 mL nitric acid (HNO<sub>3</sub>) and 5 mL perchloric acid (HClO<sub>4</sub>), then incubated at 200 °C. The contents were covered to reflux acid fumes generated during digestion until digest appeared translucent. After cooling down, 100 mL deionized water was added to each digestion tube. Contents were then vortexed and passed through qualitative cellulose filter paper (Whatman No.1, Sigma-Aldrich®, St. Louis, MO, USA) and measurement the K<sup>+</sup> (768 nm wavelength) and Na<sup>+</sup> (589 nm wavelength) by flame photometer (Model 410 Flame Photometer, Sherwood Scientific Limited, Cambridge, UK). The K<sup>+</sup> and Na<sup>+</sup> concentrations of samples were compared with the known standard solutions of 0.0, 5.0, 10.0, 15, and 20 ppm from a calibration curve with a correlation coefficient ( $r^2$ ) = 0.999. Finally, the amount of K<sup>+</sup> and Na<sup>+</sup> were transformed to percentage when compared with dry weight of raw materials.

#### *2.6. Data Analysis*

Salt tolerance scores, blast resistance scores, and agronomic trait data were analyzed via the STATISTIC 10© program (1985–2013) (Analytical Software, Tallahassee, FL, USA). Means were compared by the least significant difference (LSD) at  $p < 0.05$ .

### **3. Results**

#### *3.1. Development of Populations Through MABS*

The F<sub>1</sub> population (RGD07005-12-165-1 × Pokkali) was backcrossed with RGD07005-12-165-1 using MAS through BC<sub>2</sub>F<sub>2</sub>, in which the MAS contained genes of the glutinous type, aromatic, and gelatinization temperature (GT); and the BC<sub>2</sub>F<sub>2,3</sub> populations were then evaluated for salt tolerance and blast resistance. One BC<sub>2</sub>F<sub>2,3</sub> line (no. 74) obtained all QTLs required to obtain the target traits necessary for generation advancement (Table S2). In developing the BC<sub>3</sub> generation, BC<sub>2</sub>F<sub>2,3</sub> (no. 74) was crossed back to RD6 to produce the BC<sub>3</sub>F<sub>1</sub>, in which the BC<sub>3</sub>F<sub>1</sub> was developed through the BC<sub>3</sub>F<sub>4</sub> population by MAS. Thirty-one BC<sub>3</sub>F<sub>4</sub> lines were evaluated for salt tolerance and agronomic traits

within the salted field. Among the BC<sub>3</sub>F<sub>4</sub> lines, the BC<sub>3</sub>F<sub>4</sub> (no. 132) demonstrated salt tolerance and agronomic characteristics similar to those of the RD6 variety and were subsequently selected for the development of the BC<sub>4</sub>F<sub>1</sub> through BC<sub>4</sub>F<sub>3</sub>. The BC<sub>4</sub>F<sub>3</sub>-132-12-61 line was successfully developed via the MABS, consisting of one QTL for salt tolerance and four QTLs for blast resistance. BC<sub>4</sub>F<sub>3</sub> lines with varied QTL combinations were also validated in the tests.

### 3.2. Evaluation of Salt Tolerance and Blast Resistance in the BC<sub>2</sub>F<sub>2.3</sub> Populations

Eight of the BC<sub>2</sub>F<sub>2.3</sub> lines were evaluated for salt tolerance in the seedling stage through the salt solution and artificial soil salinity methods. The results concluded that the BC<sub>2</sub>F<sub>2.3</sub> lines presented as highly significant at EC 12 dS/m. BC<sub>2</sub>F<sub>2.3</sub> (nos. 23, 67, and 74), demonstrated tolerance (T) in both methods, similar to that of the tolerant check (Pokkali); while the remaining five BC<sub>2</sub>F<sub>2.3</sub> lines showed moderate tolerance (MT) under both the salt solution and artificial soil salinity tests (Table 1). The BC<sub>2</sub>F<sub>2.3</sub> lines were evaluated for blast resistance via the upland short-row method, through which the blast resistance reaction (R) was similar to that of the resistance checks (JHN and P0489). Their resistance proved greater than that of the susceptible check (RD6), due to the presence of blast-resistant QTLs, and superior to that of the original RD6 (Table 1).

**Table 1.** Salt tolerance scores at EC 12 Ds/m in salt solution, artificial soil salinity, and blast resistance within the BC<sub>2</sub>F<sub>2.3</sub> populations.

BC <sub>2</sub> F <sub>2.3</sub> Populations	Salt Solution		Artificial Soil Salinity		Blast Resistance	
	Score	Reaction	Score	Reaction	Score	Reaction
No. 74	4.3 c	T	4.5 cd	T	2.0 c	R
No. 42	5.5 b	MT	4.8 cd	T	1.0 c	R
No. 44	5.5 b	MT	4.5 cd	T	2.0 c	R
No. 23	4.8 bc	T	4.5 cd	T	1.0 c	R
No. 56	5.0 bc	MT	4.5 cd	T	2.0 c	R
No. 67	4.0 c	T	4.0 d	T	2.0 c	R
No. 33	5.5 b	MT	5.0 cd	MT	2.0 c	R
No. 7	4.0 c	T	5.8 bc	MT	2.0 c	R
Pokkali (tolerant check)	4.0 c	T	4.0 d	T	6.0 a	MS
IR29 (susceptible check)	9.0 a	HS	9.0 a	HS	4.0 b	MS
RD6 (recurrent parent)	9.0 a	HS	9.0 a	HS	7.0 a	S
RGD07005-12-165-1	8.0 a	S	7.0 b	S	2.0 c	R
Jao Hom Nil (resistant check)	-	-	-	-	1.0 c	R
P0489 (resistant check)	-	-	-	-	2.0 c	R
F-Test	**		**		**	
C.V. (%)	11.22		13.23		24.12	

T = tolerance, MT = moderate tolerance, S = susceptible, HS = highly susceptible, R = resistant, MR = moderate resistance, MS = moderately susceptible. \*\* = significant different at  $p < 0.01$ . Different letters after the mean within a column showed a significant difference. CV = the coefficient of variation.

### 3.3. Evaluation of Salt Tolerance in the BC<sub>3</sub>F<sub>4</sub> Populations

BC<sub>2</sub>F<sub>2.3</sub> (No.74) was selected for the next backcross cycle with the RD6, and MAS was performed through the BC<sub>3</sub>F<sub>4</sub> population. Thirty-one BC<sub>3</sub>F<sub>4</sub> lines were screened for salt tolerance and agronomic performance. The results showed that the 21 BC<sub>3</sub>F<sub>4</sub> lines were salt-tolerant (T), whereas the 10 BC<sub>3</sub>F<sub>4</sub> lines proved only moderately tolerant (MT). Agronomic traits of the BC<sub>3</sub>F<sub>4</sub> lines were similar to those of the RD6 variety, which was in accordance with our objectives (Table 2). Five BC<sub>3</sub>F<sub>4</sub> lines (nos.22, 36, 115, 129, and 132) demonstrated superior tolerance (T) and agronomic traits (1000/SW, SL, SW, and SS), again, similar to those of the original RD6, and were selected as donor parents for the development of the BC<sub>4</sub>. The results showed that BC<sub>3</sub>F<sub>4</sub> (no. 132) produced the best performance for pollination, and was therefore selected for development of the BC<sub>4</sub> population.

Table 2. Salt tolerance scores and agronomic traits within the BC<sub>3</sub>F<sub>4</sub> populations.

Line	STS	1000SW (g)	SL (cm)	SW (cm)	SS	Line	STS	1000SW (g)	SL (cm)	SW (cm)	SS
BC <sub>3</sub> F <sub>4</sub> no.2	4.33 b-d	26.80 a-g	9.95 h-l	2.62 c-h	3.81 b-g	BC <sub>3</sub> F <sub>4</sub> no.124	3.67 cd	27.83 a-d	10.64 a-d	2.70 b-h	3.97 a-e
BC <sub>3</sub> F <sub>4</sub> no.14	4.33 b-d	23.93 hi	10.06 f-l	2.64 c-h	3.83 b-g	BC <sub>3</sub> F <sub>4</sub> no.126	3.67 cd	27.60 a-d	10.36 b-i	2.64 b-h	3.93 a-f
BC <sub>3</sub> F <sub>4</sub> no.22	3.67 cd	24.40 f-i	9.86 i-m	2.58 f-h	3.83 b-g	BC <sub>3</sub> F <sub>4</sub> no.129	3.00 d	26.23 b-h	10.18 d-k	2.68 b-h	3.81 b-g
BC <sub>3</sub> F <sub>4</sub> no.26	5.00 a-c	24.17 g-i	9.79 j-m	2.59 f-h	3.79 b-g	BC <sub>3</sub> F <sub>4</sub> no.132	3.67 cd	26.47 a-h	10.39 b-i	2.73 b-g	3.81 b-g
BC <sub>3</sub> F <sub>4</sub> no.35	4.33 b-d	24.47 e-i	10.14 d-k	2.72 b-h	3.74 c-h	BC <sub>3</sub> F <sub>4</sub> no.136	4.33 b-d	24.23 g-i	9.88 h-m	2.65 b-h	3.73 c-h
BC <sub>3</sub> F <sub>4</sub> no.36	3.67 cd	26.00 c-h	9.76 k-m	2.56 gh	3.82 b-g	BC <sub>3</sub> F <sub>4</sub> no.151	6.33 a	25.97 c-h	10.02 g-l	2.75 b-f	3.68 d-h
BC <sub>3</sub> F <sub>4</sub> no.40	5.00 a-c	24.30 f-i	9.93 h-l	2.64 b-h	3.47 gh	BC <sub>3</sub> F <sub>4</sub> no.152	5.67 ab	27.13 a-f	9.92 h-l	2.79 bc	3.56 f-h
BC <sub>3</sub> F <sub>4</sub> no.55	5.67 ab	24.40 f-i	9.35 mn	2.76 b-e	3.40 h	BC <sub>3</sub> F <sub>4</sub> no.153	5.00 a-c	25.07 d-i	10.05 g-l	2.73 b-g	3.69 c-h
BC <sub>3</sub> F <sub>4</sub> no.58	5.00 a-c	24.50 e-i	9.57 l-n	2.62 c-h	3.66 d-h	BC <sub>3</sub> F <sub>4</sub> no.195	5.00 a-c	27.40 a-d	10.59 a-f	2.55 h	4.15 ab
BC <sub>3</sub> F <sub>4</sub> no.67	6.33 a	22.90 i	9.04 n	2.68 b-h	3.38 h	BC <sub>3</sub> F <sub>4</sub> no.235	3.67 cd	25.80 d-h	10.32 b-j	2.61 d-h	4.00 a-d
BC <sub>3</sub> F <sub>4</sub> no.69	5.67 ab	27.67 a-d	10.76 a-c	2.73 b-g	3.97 a-e	BC <sub>3</sub> F <sub>4</sub> no.241	3.67 cd	27.27 a-e	10.37 b-i	2.74 b-g	3.81 b-g
BC <sub>3</sub> F <sub>4</sub> no.89	4.33 b-d	27.57 a-d	10.40 b-h	2.72 b-h	3.84 b-f	BC <sub>3</sub> F <sub>4</sub> no.254	3.00 d	29.07 ab	10.71 a-c	2.64 b-h	4.06 a-c
BC <sub>3</sub> F <sub>4</sub> no.93	3.67 cd	26.40 a-h	10.53 a-g	2.73 b-g	3.86 b-f	BC <sub>3</sub> F <sub>4</sub> no.265	3.67 cd	27.50 a-d	10.83 ab	2.58 f-h	4.25 a
BC <sub>3</sub> F <sub>4</sub> no.109	4.33 b-d	29.13 a	10.97 a	2.62 c-h	3.91 a-f	BC <sub>3</sub> F <sub>4</sub> no.298	3.67 cd	26.47 a-h	10.28 c-k	2.64 b-h	3.91 a-f
BC <sub>3</sub> F <sub>4</sub> no.113	4.33 b-d	27.63 a-d	10.73 a-c	2.68 b-h	4.01 a-d	Pokkali	4.33 b-d	13.77 j	8.47 o	3.42 a	2.48 i
BC <sub>3</sub> F <sub>4</sub> no.115	3.00 d	28.80 a-c	10.64 a-e	2.81 b	3.79 b-g	RD6	6.33 a	24.20 g-i	9.98 h-l	2.78 b-d	3.61 e-h
BC <sub>3</sub> F <sub>4</sub> no.118	3.67 cd	25.63 d-i	10.10 e-k	2.59 e-h	3.92 a-f	BC <sub>2</sub> F <sub>5</sub> no.74	5.67 ab	24.53 e-i	9.88 h-m	2.68 b-h	3.70 c-h
<b>F-test</b>	**	**	**	**	**	<b>F-test</b>	**	**	**	**	**
<b>C.V. (%)</b>	25.4	6.8	3.96	3.2	6.18	<b>C.V. (%)</b>	25.4	6.8	3.96	3.2	6.18

STS = salt tolerance score, 1000SW = 1000 seeds weight, SL = seed length, SW = seed width, SS = seed shape (ratio of SL/SW). \*\* = significantly different at  $p < 0.01$ . Different letters after the mean within a column showed a significant difference. CV = the coefficient of variation.

### 3.4. Evaluation of Salt Tolerance and Blast Resistance in the BC<sub>4</sub>F<sub>3</sub> Populations

This experiment evaluated eight BC<sub>4</sub>F<sub>3</sub> lines obtained via *Saltol* QTL from MAS, specifically, the combination lines, donor parents, and tolerance, and susceptible checks were screened for salt tolerance via the salt solution method (EC 12 dS/m). The results showed highly significant scores within the BC<sub>4</sub>F<sub>3</sub> population and the tolerant check (Pokkali), with salt tolerance scores less than 5.0, indicated as tolerant (T). The recurrent parent (RD6) presented moderate tolerance (MT) with a score of 6.5 (Table 3). The results indicated that the *Saltol* QTL within the BC<sub>4</sub>F<sub>3</sub> population demonstrated superior performance in salt tolerance to that of the recurrent parent (RD6).

The BC<sub>4</sub>F<sub>3</sub> populations were evaluated for blast resistance in field conditions through natural infection. Since blast infection in the field was not severe, leaves with severe symptoms from the surrounding trap plants were collected and stored in a bag, under dark conditions, for twelve hours to induce spore formation. The natural inoculum was added with water, and spore suspension was then sprayed over the field. The BC<sub>4</sub>F<sub>3</sub> populations and their parents showed a high level of blast resistance. A total of eight BC<sub>4</sub>F<sub>3</sub> lines demonstrated high resistance (HR), similar to those of both the donor parents, whereas the recurrent parent (RD6) presented moderate susceptibility (MS) (Table 3). Importantly, some introgression lines were greater in resistance than the recurrent parent (RD6). The second peak of bimodal rain, which occurred in the flowering stage, initiated significant signs of neck blast. The results found that eight BC<sub>4</sub>F<sub>3</sub> lines showed resistance (R) similar to that of the donor parent, whereas the RD6 was moderately susceptible (MS) to neck blast (Table 3).

Additionally, blast resistance evaluation was also conducted via the upland short-row method in the seven BC<sub>4</sub>F<sub>3</sub> lines, which presented highly significant blast-resistant levels. The BC<sub>4</sub>F<sub>3</sub> lines showed resistance abilities similar to both donor parents, whereas the RD6 recurrent parent presented moderate resistance (MR). Notably, the blast-resistant genes in the BC<sub>4</sub>F<sub>3</sub> populations provided blast resistance in the seedling stage similar to that of the tillering and grain filling stage, and evidenced greater resistance than that of the RD6 recurrent parent (Table 3).

The agronomic traits were also evaluated in the blast field experiments. The results were highly significant within the BC<sub>4</sub>F<sub>3</sub> populations and their parents for ten traits: PH; PL; SW4P; 1000/SW; TDW; TSW; HI; SL; SW; and SS (Table 4). The BC<sub>4</sub>F<sub>3</sub> 132-12 maintained agronomic traits similar to those of the recurrent parent (RD6) for nine traits, except for the 1000/SW (1000 seed weight), in which the RD6 presented moderate susceptibility (MS) for leaf and neck blast, resulting in low grain filling (Table 4). The results indicate that the newly developed MAB population had greater resistance than that of the original RD6 variety while maintaining its desirable agronomic traits and satisfying consumer demand. The BC<sub>4</sub>F<sub>3</sub> lines also showed improvements in seed length (SL), seed width (SW), and seed shape (SS) as long and slender type. Additionally, the color of paddy rice was straw yellow, similar to the original RD6 varieties (Figure 2).

**Table 3.** Blast resistance scores of leaf blast neck blast and salt tolerant evaluations in the field and upland short-row methods in the BC<sub>4</sub>F<sub>3</sub> populations.

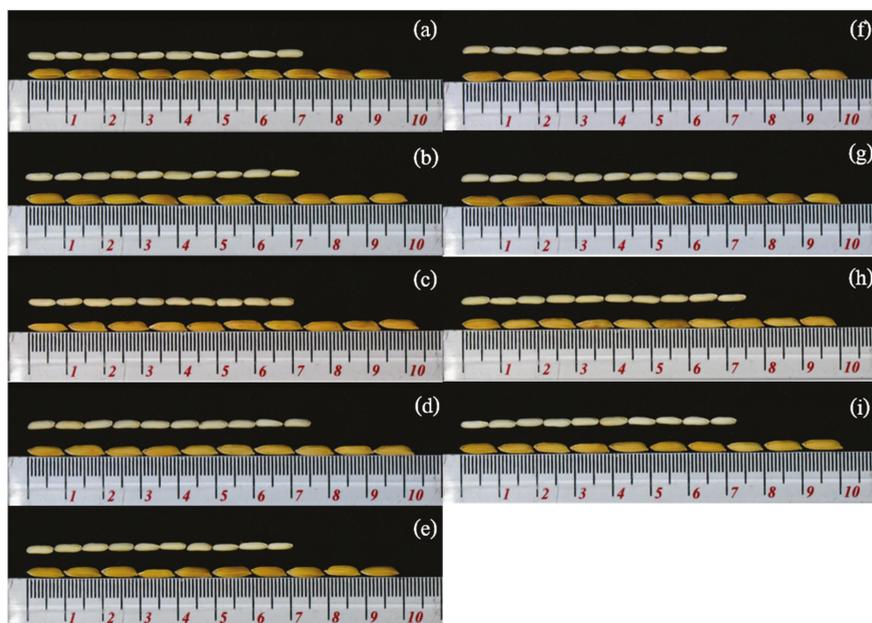
Line	QTLs	Leaf Blast		Neck Blast		Upland Short Row Method		Salt Score	
		Resistance	Mean	Resistance	Mean	Resistance	Mean	Tolerance	Mean
BC <sub>4</sub> F <sub>3</sub> 132-12-61	<i>Saltol</i> , <i>qBl</i> (1 2 11 12)	HR	0.00 c	R	1.53 b	HR	0.00 c	T	3.67 b
BC <sub>4</sub> F <sub>3</sub> 132-25-18	<i>Saltol</i> , <i>qBl</i> (1 11 12)	HR	0.00 c	R	1.27 b	HR	0.00 c	T	3.45 b
BC <sub>4</sub> F <sub>3</sub> 132-14-43	<i>Saltol</i> , <i>qBl</i> (1 11 12)	HR	0.00 c	R	1.67 b	HR	0.00 c	T	3.67 b
BC <sub>4</sub> F <sub>3</sub> 132-98-87	<i>Saltol</i> , <i>qBl</i> (1 11 12)	HR	0.00 c	R	1.67 b	HR	0.00 c	T	3.67 b
BC <sub>4</sub> F <sub>3</sub> 132-174-54	<i>Saltol</i> , <i>qBl</i> (1 11)	HR	0.00 c	R	1.93 b	-	-	T	3.89 b
BC <sub>4</sub> F <sub>3</sub> 132-51-17	<i>Saltol</i> , <i>qBl</i> (12)	HR	0.87 c	R	1.00 b	R	1.00 c	T	3.67 b
BC <sub>4</sub> F <sub>3</sub> 132-167-58	<i>Saltol</i> , <i>qBl</i> (11)	HR	0.23 c	R	2.53 b	HR	0.33 c	T	3.67 b
BC <sub>4</sub> F <sub>3</sub> 132-276-84	<i>Saltol</i> , <i>qBl</i> (2)	HR	0.43 c	R	1.13 b	HR	0.67 c	T	3.89 b
RD6 (recurrent)	-	MS	6.77 a	MS	5.80 a	MR	4.33 b	MS	6.56 a
JHN (resistant check)	<i>qBl</i> (1 11)	MR	3.67 b	R	1.00 b	HR	0.00 c	-	-
P0489 (resistant check)	<i>qBl</i> (2 12)	HR	0.00 c	R	1.00 b	HR	0.00 c	-	-
KDML105 (susceptible check)	-	-	-	-	-	S	7.67 a	-	-
IR 29 (susceptible check)	-	-	-	-	-	-	-	S	7.45 a
F-test		**		**		**		**	
C.V. (%)		48.2		60.73		31.03			17.20

HR = high resistance, R = resistance, MR = moderate resistance, MS = moderate susceptible, S = susceptible. \*\* = significantly different at  $p < 0.01$ . Different letters after the mean within a column show a significant difference. CV = the coefficient of variation.

**Table 4.** Evaluation of agronomic traits and 2-acetyl-1-pyrroline (2AP) content from the blast fields of the BC<sub>4</sub>F<sub>3</sub> populations.

Line	PH (cm)	PL (cm)	SW4P (g)	1000SW (g)	TDW (g)	TSW (g)	HI	SL (cm)	SW (cm)	SS	2AP Content (ppm)
BC <sub>4</sub> F <sub>3</sub> 132-12-61	161.33 ab	27.10 a-c	20.98 a-c	27.90 c-e	354.66 ab	146.05 a-c	0.42 a	1.09 b	0.26 bc	4.11 cd	3.63 b-d
BC <sub>4</sub> F <sub>3</sub> 132-25-18	165.90 a	27.67 a	20.33 a-c	29.07 ab	367.31 a	157.76 ab	0.43 a	1.09 b	0.26 c	4.20 bc	3.15 d-f
BC <sub>4</sub> F <sub>3</sub> 132-14-43	163.90 ab	27.30 ab	22.03 ab	29.10 ab	330.40 ab	146.54 a-c	0.44 a	1.07 b	0.27 ab	4.06 c-e	4.05 b
BC <sub>4</sub> F <sub>3</sub> 132-98-87	163.00 ab	25.60 de	20.81 a-c	26.93 e	325.53 ab	147.19 a-c	0.45 a	1.02 c	0.26 bc	3.88 ef	4.68 a
BC <sub>4</sub> F <sub>3</sub> 132-174-54	164.43 ab	25.70 c-e	20.80 a-c	28.50 bc	345.89 ab	151.41 a-c	0.44 a	1.05 bc	0.26 bc	4.06 c-f	3.44 c-e
BC <sub>4</sub> F <sub>3</sub> 132-51-17	166.10 a	27.00 a-d	19.82 a-c	28.17 bc	291.88 bc	122.30 bc	0.42 a	1.08 b	0.26 bc	4.08 cd	3.73 bc
BC <sub>4</sub> F <sub>3</sub> 132-167-58	162.90 ab	27.93 a	23.27 a	27.97 cd	334.45 ab	146.77 a-c	0.44 a	1.09 b	0.26 c	4.22 bc	3.84 bc
BC <sub>4</sub> F <sub>3</sub> 132-276-84	165.10 a	26.73 a-d	20.91 a-c	28.53 bc	368.80 a	164.07 a	0.44 a	1.05 bc	0.27 a	3.86 f	3.06 ef
RD6 (recurrent)	167.57 a	25.87 b-e	18.46 bc	25.10 f	295.79 a-c	115.33 cd	0.39 ab	1.02 c	0.26 bc	3.88 ef	4.62 a
JHN (resistant check)	109.80 d	25.03 e	12.99 d	21.80 h	224.05 c	60.30 e	0.26 c	1.07 b	0.24 d	4.43 a	-
P0489 (resistant check)	137.70 c	25.53 de	17.36 c	23.30 g	232.65 c	82.13 de	0.34 b	0.95 d	0.23 d	4.05 c-f	-
F-test	**	**	**	**	**	**	**	**	**	**	**
C.V. (%)	2.46	3.33	11.28	2.16	13.95	16.74	10.15	2.79	1.70	2.85	3.72

PH = plant height, PL = panicle length, SW4P = seed weight 4 panicles, 1000SW = 1000 seeds weight, TDW = total dry weight 4 plants, TSW = total seed weight 4 plants, HI = harvest index, SL = seed length, SW = seed, width, SS = seed shape (ratio of SL/SW), 2AP 2-acetyl-1-pyrroline. \*\* = significantly different at  $p < 0.01$ . Different letters after the mean within a column show a significant difference. CV = the coefficient of variation.



**Figure 2.** Seed quality, seed length, seed shape, and seed color of 10 seeds of the BC<sub>4</sub>F<sub>3</sub> populations compared with the RD6 variety. (a) RD6, (b) BC<sub>4</sub>F<sub>3</sub> 133-12-61, (c) BC<sub>4</sub>F<sub>3</sub> 133-25-18, (d) BC<sub>4</sub>F<sub>3</sub> 132-14, (e) BC<sub>4</sub>F<sub>3</sub> 132-98, (f) FC<sub>4</sub>F<sub>3</sub> 132-174, (g) BC<sub>4</sub>F<sub>3</sub> 132-51, (h) BC<sub>4</sub>F<sub>3</sub> 132-167, (i) BC<sub>4</sub>F<sub>3</sub> 132-276. The top row is brown rice and bottom row is paddy rice with straw color. The small scale is in millimeters.

Eight BC<sub>4</sub>F<sub>3</sub> lines and the RD6 (recurrent parent) were evaluated for seed aroma through the determination of 2AP content via the automated headspace gas chromatography method. The results showed a significant difference among the BC<sub>4</sub>F<sub>3</sub> lines and RD6 variety, with mean values of 2AP content of the BC<sub>4</sub>F<sub>3</sub> exceeding 3.00 ppm (Table 4). BC<sub>4</sub>F<sub>3</sub> 132-98-87 presented the highest 2AP content (4.68 ppm), similar to that of the RD6 (Table 4). The BC<sub>4</sub>F<sub>3</sub> lines and RD6 were determined to be similar in fragrance.

### 3.5. Evaluation of Salt Tolerance in the BC<sub>4</sub>F<sub>4</sub> Population

Two BC<sub>4</sub>F<sub>4</sub> lines, BC<sub>4</sub>F<sub>4</sub> 132-12 and BC<sub>4</sub>F<sub>4</sub> 132-167, presented as tolerant (T) in salt tolerance evaluations, similar to that of the Pokkali, whereas the remaining five BC<sub>4</sub>F<sub>4</sub> lines and the recurrent parent (RD6) showed moderate tolerance (MT) (Table 5). Statistically, seven of the BC<sub>4</sub>F<sub>4</sub> lines demonstrated EC values up to 18 dS/m. The dry weights of the seven BC<sub>4</sub>F<sub>4</sub> lines, their parent, and KDML105 check varieties presented as highly significant, in which the tolerant check (Pokkali) presented the highest LDW, SDW, RDW, and TDW evident in the weights of leaf stems and roots, whereas the BC<sub>4</sub>F<sub>4</sub> lines were similar to that of the recurrent parent (RD6) (Table 5). The results indicate that salinity significantly affected dry weight.

**Table 5.** Leaf, stem, root, and total dry weights in the experiment 4.

Line	LDW (g)		SDW (g)		RDW (g)		TDW (g)		Salt Score
BC4F4 132-12-61	0.56	bc	0.47	b	0.27	b	1.3	bc	5.2 c-e
BC4F4 132-25-18	0.47	b-d	0.33	bc	0.22	b	1.02	bc	5.9 b-d
BC4F4 132-14-43	0.55	bc	0.41	bc	0.25	b	1.21	bc	6.0 b-e
BC4F4 132-51-17	0.5	bc	0.35	bc	0.22	b	1.07	bc	5.2 c-e
BC4F4 132-167-58	0.49	bc	0.42	bc	0.21	bc	1.12	bc	4.9 de
BC4F4 132-276-84	0.44	b-d	0.39	bc	0.23	b	1.06	bc	5.5 b-e
RD6 (recurrent)	0.39	cd	0.27	cd	0.2	bc	0.85	cd	6.2 a-c
Pokkali (tolerant check)	1.21	a	1.18	a	0.47	a	2.86	a	3.6 f
IR29 (susceptible check)	0.23	d	0.11	d	0.13	c	0.47	d	7.4 a
KDML105	0.63	b	0.51	b	0.25	b	1.39	b	6.5 ab
F-test	**		**		**		**		**
CV (%)	22.24		21.68		17.1		18.88		11.35

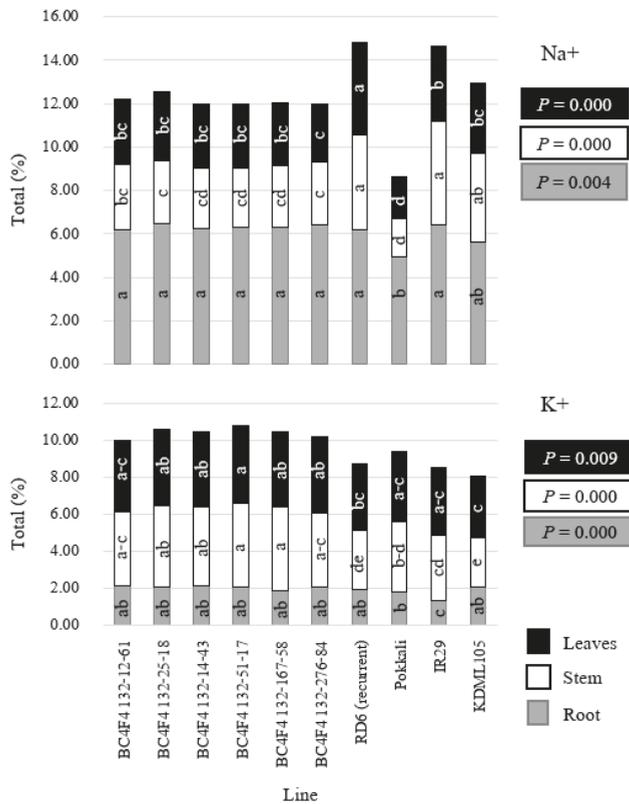
LDW = leaf dry weight, SDW = stem dry weight, RDW = root dry weight, TDW = total dry weight. \*\* = significantly different at  $p < 0.01$ . Different letters after the mean within a column show a significant difference. CV = the coefficient of variation.

Additionally, the percentages of Na<sup>+</sup> and K<sup>+</sup> in leaf stems and roots were also recorded and proved highly significant in all traits. The tolerant checks (Pokkali) presented the lowest Na<sup>+</sup> in leaves (1.94), stems (1.75), and roots (4.94), whereas K<sup>+</sup> in leaf stems were non-significant within the BC<sub>4</sub>F<sub>4</sub> lines. Moreover, Pokkali also displayed low Na<sup>+</sup>-to-K<sup>+</sup> ratios in rice shoots. The BC<sub>4</sub>F<sub>4</sub> 132-12 also showed low levels of Na<sup>+</sup> in both stems and leaves, as well as low Na<sup>+</sup>-to-K<sup>+</sup> ratios in rice shoots comparable to those of the Pokkali variety, yet lower than those of the RD6. We may infer that transferring the saltol QTL from the Pokkali line created greater tolerance in subsequent breeding lines that that of RD6 recurrent parent (Figure 3). Interestingly, the BC<sub>4</sub>F<sub>4</sub> lines presented Na<sup>+</sup> levels in leaves and stems similar to those in Pokkali, and significantly different in RD6. This confirmed the Pokkali saltol QTL's ability to exclude Na<sup>+</sup> from the leaf blade, expressing salt tolerance through the salt-excluder method. Moreover, the salt tolerance scores were negatively correlated to the leaf, stem, root, and total dry weights ( $r = -0.7899$  \*\*,  $r = -0.8136$  \*\*,  $r = -0.8140$  \*\*,  $r = -0.8065$  \*\*, respectively) and positively correlated with leaf and stem Na<sup>+</sup> ( $r = 0.7670$  \*\* and  $r = 0.8917$  \*\*, respectively) (Table 6). The results indicated that salt stress decreased plant growth in susceptible varieties, while growth was maintained in salt-tolerant varieties and that the accumulation of Na<sup>+</sup> in leaves and stems was related to salt susceptibility (Table 6).

**Table 6.** The correlation between leaf dry weight, stem dry weight, root dry weight, total dry weight, leaf Na<sup>+</sup>, leaf K<sup>+</sup>, stem Na<sup>+</sup>, stem K<sup>+</sup>, root Na<sup>+</sup>, root K<sup>+</sup>, and the salt score of seeds in the BC<sub>4</sub>F<sub>4</sub> populations and check varieties in the experiment 4.

	LDW	SDW	RDW	TDW	Leaves Na <sup>+</sup>	Leaves K <sup>+</sup>	Stem Na <sup>+</sup>	Stem K <sup>+</sup>	Root Na <sup>+</sup>	Root K <sup>+</sup>
SDW	0.9926 **									
RDW	0.9863 **	0.9864 **								
TDW	0.9978 **	0.9981 **	0.9911 **							
Leaves Na <sup>+</sup>	-0.7490 *	-0.7762 **	-0.7513 *	-0.7665 **						
Leaves K <sup>+</sup>	-0.0157	-0.0096	0.0313	-0.0044	-0.4369					
Stem Na <sup>+</sup>	-0.6957 *	-0.6999 *	-0.7268 *	-0.7056 *	0.8532 **	-0.6924 *				
Stem K <sup>+</sup>	-0.0578	-0.0566	-0.0188	-0.0500	-0.3733	0.9680 **	-0.6326 *			
Root Na <sup>+</sup>	-0.9140 **	-0.9037 **	-0.8716 **	-0.9054 **	0.5259	0.3967	0.3537	0.4257		
Root K <sup>+</sup>	0.1762	0.1411	0.2168	0.1666	-0.1362	0.3735	-0.4801	0.2748	0.0495	
Salt score	-0.7899 **	-0.8136 **	-0.8140 **	-0.8065 **	0.7670 **	-0.4516	0.8917 **	-0.4183	0.5386	-0.3405

LDW = leaf dry weight, SDW = stem dry weight, RDW = root dry weight, TDW = total dry weight. \* = significantly different at  $p < 0.05$ . \*\* = significantly different at  $p < 0.01$ .



**Figure 3.** Na<sup>+</sup> and K<sup>+</sup> on leaves, stems, and roots in the experiment 4. Different letters within a color bar show a mean significant difference of each line.

#### 4. Discussion

Since its release in 1977, the RD6 glutinous rice variety has remained a staple food crop for domestic consumption in Thailand’s north and northeast regions. Comprising 83% of total glutinous rice production in these areas, consumers have developed a preference for its superior characteristics. However, the RD6 variety suffers from several production constraints, including biotic stress responsible for both rice blast [40] and bacterial blight disease [41]. Current research has attempted to eliminate sustainable infection-resistant production practices by pyramiding multiple resistant genes [42]. To date, an RD6 introgression line capable of resisting both biotic and abiotic stress has yet to be developed. Thailand’s salt rock basins of Sakon Nakhon and Nakhon Ratchasima have demonstrated that consistent levels of salinity can enhance the fragrance of the RD6 rice variety [14] and increase production.

This study proposes the successful introgression of blast-resistant QTLs (*qBL 1, 2, 11, and 12*) from RGD07005-12-165-1 and the *Saltol* QTL (Pokkali, chromosome 1) to improve the RD6 rice variety through the MAB method within BC<sub>4</sub>F<sub>4</sub> populations. Trait evaluations were completed for the validation of progenies with desirable traits in each advanced population, based on the introgression of the genetic foregrounds and maintenance of the genetic backgrounds, respectively.

Salt salinity was absent in several areas of northeast Thailand, due to high levels of NaCl [43], and such factors as precipitation, soil type, and field management. In past research, the evaluation of salt tolerance was typically conducted through salt screening, hydroponic culture, and soil culture, as well as through pot and field methods [33]. The current study assessed salt tolerance within the breeding

populations studied through salt solution, artificial salt culture, and field condition evaluations. Based on the results, salt evaluation under field conditions produced the lowest capability among the tested rice lines (Tables 1–3 and 5), due to the inherent difficulties and uncertainties present under field conditions. Kranto et al. [33] reported that effective alternative screening approaches must be proven to correlate with results produced within the early phases of growth in both greenhouse and field conditions. Within the present study, visual symptom scores of salt stress generated through the salt solution method proved to be the most appropriate method with which to confirm tolerance abilities within a breeding population (Table 5), suggesting that the salt solution method could, therefore, substitute salt tolerance score analysis in field conditions (Table 5). However, we acknowledge the necessity to evaluate RD6 plant types, yield performances, and agronomic traits within the field. The introgression lines developed within our study were evaluated for similarity with the original RD6 agronomic traits, namely plant height, panicle length, 4/panicle seed weight, 1000/seed weight, total dry weight, total seed weight, harvest index, seed length, seed width, and seed shape, as well as seed qualities, such as seed morphology (Tables 2 and 4).

The *Saltol* QTL on chromosome 1 from the Pokkali rice variety has been commonly used for rice improvement in several studies [21–26]. In our results, the *Saltol* QTLs from the Pokkali variety produced the greatest salt tolerance within the RD6 introgression lines (Table 1, Table 2, Table 3, and Table 5). This *saltol* QTL also contributed to the maintenance of low Na<sup>+</sup>, high K<sup>+</sup>, and low Na<sup>+</sup>/K<sup>+</sup> homeostasis levels in rice stems, further resulting in increased salt tolerance [24,44] (Figure 3). The Pokkali variety was classified to balance the influx of Na<sup>+</sup> and K<sup>+</sup> for dilution in the mechanism, creating the ability to exclude Na<sup>+</sup> from leaf blades and stems [45,46]. As the water up-take mechanisms in rice accept both nutrients and salt together, the Pokkali variety thereby demonstrated the highest and most significant differences in leaf, stem, root, and total dry weights when compared with other breeding lines (Table 5). However, the BC4F4 lines presented the agronomic traits (above) more closely matched to the RD6 than to the Pokkali (Table 5), due to the advance generation and visual selection of the trait performances (Tables 2 and 4). RD6 performance is very important for farmer acceptance and crop adaptation in our test areas. For example, excessively tall RD6 rice plants present problems in the grain filling stages as a result of heavy wind or rain [47]. Visual selection may explain the differences in percentages of Na<sup>+</sup> of the RD6 introgression lines with those of Pokkali (Table 5, Figure 3).

As a photosensitive rice variety, the RD6 grows once a year, during Thailand's rainy season from late May to November [48]. These bimodal rain patterns produce favorable conditions for the occurrence of blast disease, causing damage in all stages of growth. Leaf blast generally occurs during the seedling and tillage stages, whereas neck blast usually occurs during the reproductive phase [4]. In our study, introgression lines were evaluated for blast disease in both the field and upland short-row evaluations.

In this study, the upland short-row method displayed greater incidences of blast disease, due to the favorable microclimate and moisture contents around the experimental plots (Tables 1 and 3) [49]. The experimental field was influenced by bimodal rain, capable of inducing leaf and neck blast symptoms (Table 3), further indicating the resistance of the QTLs [42,50]. Noenplab et al. [8] also reported that the blast QTL on chromosome 11 in the JHN variety successfully contributed to leaf and neck blast resistance. Pyramiding of four blast-resistant QTLs through MAS achieved high levels of blast resistance and broad-spectrum resistance to pathogens prevalent in the region [9]. Moreover, the testing of RD6 introgression lines for durable blast resistance and no-yield penalties were observed [42]. The results, herein, further demonstrated that neck blast disease caused direct yield loss during the grain filling phase [51], as well as lower 1000/SW within the original RD6 variety compared with those of the RD6 introgression lines (Table 4).

The resistance/tolerance abilities of the RD6 introgression lines represent the foreground genetics capable of enhancing plant breeding programs. However, maintaining the background of the original RD6 variety is also desirable; therefore, the quality and performance of the RD6 within the QTL introgression was also a consideration. The BC<sub>4</sub>F<sub>4</sub> populations, herein, were achieved through the

introgression of blast-resistant QTLs (*qBL 1, 2, 11*, and *12*) from RGD07005-12-165-1 and *Saltol* QTL (Pokkali) and improved the RD6 rice variety through MAB. Consequently, the performance of the RD6 introgression lines was similar to that of the original RD6 variety (Table 4, Figure 2). The results indicate that foreground and background selection, together with visual selection, accurately depicts the efficiency of MAB.

## 5. Conclusions

Improvement of the RD6 rice variety for salt tolerance and blast resistance was successfully achieved utilizing the *Saltol* QTL and *qBl* (*1, 2, 11*, and *12*) through marker-assisted backcrossing, together with phenotypic selection. The resulting BC<sub>4</sub>F<sub>4</sub> 132-12 introgression line exhibited superior salt tolerance, blast resistance, and reduced neck blast and was capable of maintaining higher qualities and agronomic performances than that of the original RD6 variety.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4395/10/8/1118/s1>, Table S1: QTL traits and primer sequences of the SSR markers for blast resistance and salt tolerance, Table S2: Genotype of the BC<sub>2</sub>F<sub>2</sub> populations derived from MAB.

**Author Contributions:** K.T. and S.K. conceived the study. T.M., S.C., and J.S. designed the experiments. K.T. and S.K. performed the experiments. J.S. and S.C. supervised the study. K.T., T.M., S.C., and J.S. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** This research was supported by The Plant Breeding Research Center for Sustainable Agriculture and The Salt-Tolerance Rice Research Group, Khon Kaen University, Khon Kaen, Thailand. Our gratitude is also extended to the Sakon Nakhon Rice Research Center for their support in our field experiments and Ubon Ratchathani Rice Research Center for their support in 2AP analysis.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Keeratipibul, S.; Luangsakul, N.; Lertsachayarn, T. The effect of Thai glutinous rice cultivars, grain length and cultivating locations on the quality of rice cracker (arare). *Food Sci. Technol.* **2008**, *41*, 1934–1943. [CrossRef]
2. Chairote, E.; Jannoey, P.; Chairote, G. Improvement of Monacolin K and minimizing of citrinin content in Korkor 6 (RD 6) red yeast rice. *World Acad. Sci. Eng. Technol.* **2015**, *9*, 43–46.
3. Raj, S.V.; Saranya, R.S.; Kumar, D.S.; Chinnadurai, M. Farm-level economic impact of rice blast: A Bayesian approach. *Agric. Econ. Res. Rev.* **2018**, *31*, 141. [CrossRef]
4. Mentlak, T.A.; Kombrink, A.; Shinya, T.; Ryder, L.S.; Otomo, I.; Saitoh, H.; Terauchi, R.; Nishizawa, Y.; Shibuya, N.; Thomma, B.P.H.J.; et al. Effector-mediated suppression of chitin-triggered immunity by *magnaporthe oryzae* is necessary for rice blast disease. *Plant Cell* **2012**, *24*, 322–335. [CrossRef]
5. Chumpol, A.; Chankaew, S.; Saepaisan, S.; Monkham, T.; Sanitchon, J. New sources of rice blast resistance obtained from Thai indigenous upland rice germplasm. *Euphytica* **2018**, *214*, 183. [CrossRef]
6. Greer, C.A.; Webster, R.K. Occurrence, distribution, epidemiology, cultivar reaction, and management of rice blast disease in California. *Plant Dis.* **2001**, *85*, 1096–1102. [CrossRef]
7. Wang, D.; Guo, C.; Huang, J.; Yang, S.; Tian, D.; Zhang, X. Allele-mining of rice blast resistance genes at AC134922 locus. *Biochem. Biophys. Res. Commun.* **2014**, *446*, 1085–1090. [CrossRef]
8. Noenplab, A.; Vanavichit, A.; Toojinda, T.; Sirithunya, P.; Tragoonrun, S.; Sriprakhon, S.; Vongsaprom, C. QTL mapping for leaf and neck blast resistance in Khao Dawk Mali105 and Jao Hom Nin recombinant inbred lines. *Sci. Asia* **2006**, *32*, 133–142. [CrossRef]
9. Suwannal, T.; Chankaew, S.; Monkham, T.; Saksirirat, W.; Sanitchon, J. Pyramiding of four blast resistance QTLs into Thai rice cultivar RD6 through marker-assisted selection. *Czech J. Genet. Plant Breed.* **2017**, *53*, 1–8. [CrossRef]
10. Arunin, S.; Pongwichian, P. Salt-affected soils and management in Thailand. *Bull. Soc. Sea Water Sci.* **2015**, *69*, 319–325.
11. Wongsomsak, S. Salinization in Northeast Thailand. *Southeast Asian Stud.* **1986**, *24*, 133–153.

12. Arunin, S. Salt effect soil in Southeast Asia. In Proceedings of the International Symposium on Salt Affected Lagoon Ecosystem, Valencia, Spain, 18–25 September 1995.
13. Akbar, M. Breeding for salinity resistance in rice. In *Prospects for Bio-Saline Research*; Ahmed, R., Pietro, A.S., Eds.; Department of Botany, University of Karachi: Karachi, Sindh, Pakistan, 1986; pp. 37–55.
14. Summart, J.; Thanonkeo, P.; Panichajakul, S.; Prathepha, P.; McManus, M.T. Effect of salt stress on growth, inorganic ion and proline accumulation in Thai aromatic rice, KhaoDawk Mali 105, callus culture. *Afr. J. Biotechnol.* **2010**, *9*, 145–152.
15. Kumar, K.; Kumar, M.; Kim, S.-R.; Ryu, H.; Cho, Y.-G. Insights into genomics of salt stress response in rice. *Rice* **2013**, *6*, 27. [[CrossRef](#)] [[PubMed](#)]
16. Kobayashi, N.I.; Yamaji, N.; Yamamoto, H.; Okubo, K.; Ueno, H.; Costa, A.; Tanoi, K.; Matsumura, H.; Fujii-Kashino, M.; Horiuchi, T.; et al. OsHKT1;5 mediates Na<sup>+</sup> exclusion in the vasculature to protect leaf blades and reproductive tissues from salt toxicity in rice. *Plant J.* **2017**, *91*, 657–670. [[CrossRef](#)] [[PubMed](#)]
17. Gregorio, G.B.; Senadhira, D.; Mendoza, R.D. Screening rice for salinity tolerance. *IRRI Discuss. Pap. Ser.* **1997**, *22*, 2–23.
18. Bhowmik, S.K.; Islam, M.M.; Emon, R.M.; Begum, S.N.; Siddika, A.; Sultana, S. Identification of salt tolerant rice cultivars via phenotypic and marker-assisted procedures. *Pak. J. Boil. Sci.* **2007**, *10*, 4449–4454.
19. Kavitha, P.G.; Miller, A.J.; Mathew, M.K.; Maathuis, F.J.M. Rice cultivars with differing salt tolerance contain similar cation channels in their root cells. *J. Exp. Bot.* **2012**, *63*, 3289–3296. [[CrossRef](#)]
20. Ferreira, L.J.; Azevedo, V.S.; Marôco, J.P.; Oliveira, M.M.; Santos, A.P. Salt tolerant and sensitive rice varieties display differential methylome flexibility under salt stress. *PLoS ONE* **2015**, *10*, e0124060. [[CrossRef](#)]
21. Mohammadi-Nejad, G.; Arzani, A.; Rezai, A.M.; Singh, R.K.; Gregorio, G.B. Assessment of rice genotypes for salt tolerance using microsatellite markers associated with the saltol QTL. *Afr. J. Biotechnol.* **2008**, *7*, 730–736.
22. Huyen, L.T.N.; Cuc, L.M.; Ham, L.H.; Khanh, T.D. Introgression the Saltol QTL into Q5BD, the elite variety of Vietnam using marker assisted selection (MAS). *Am. J. Biosci.* **2013**, *1*, 80–84. [[CrossRef](#)]
23. Ali, S.; Gautam, R.K.; Mahajan, R.; Krishnamurthy, S.L.; Sharma, S.K.; Singh, R.K. Stress indices and selectable traits in Saltol QTL introgressed rice genotypes for reproductive stage tolerance to sodicity and salinity stresses. *Field Crops Res.* **2013**, *154*, 65–73. [[CrossRef](#)]
24. Waziri, A.; Kumar, P.; Purty, R.S. Saltol QTL and their role in salinity tolerance in rice. *Austin. J. Biotechnol. Bioeng.* **2016**, *3*, 1067.
25. Ganie, S.A.; Borgohain, M.J.; Kritika, K.; Talukdar, A.; Pani, D.R.; Mondal, T.K. Assessment of genetic diversity of Saltol QTL among the rice (*Oryza sativa* L.) genotypes. *Physiol. Mol. Boil. Plants* **2016**, *22*, 107–114. [[CrossRef](#)] [[PubMed](#)]
26. Aala, W.F.; Gregorio, G.B. Morphological and molecular characterization of novel salt-tolerant rice germplasms from the philippines and bangladesh. *Rice Sci.* **2019**, *26*, 178–188. [[CrossRef](#)]
27. Tanksley, S.; Nelson, J. Advanced backcross QTL analysis: A method for simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theor. Appl. Genet.* **1996**, *92*, 191–203. [[CrossRef](#)]
28. Steele, K.; Price, A.H.; Shashidhar, H.E.; Witcombe, J.R. Marker-assisted selection to introgress rice QTLs controlling root traits into an Indian upland rice variety. *Theor. Appl. Genet.* **2005**, *112*, 208–221. [[CrossRef](#)]
29. Neeraja, C.N.; Maghirang-Rodriguez, R.; Pamplona, A.; Heuer, S.; Collard, B.C.; Septiningsih, E.M.; Vergara, G.; Sanchez, D.; Xu, K.; Ismail, A.M.; et al. A marker-assisted backcross approach for developing submergence-tolerant rice cultivars. *Theor. Appl. Genet.* **2007**, *115*, 767–776. [[CrossRef](#)]
30. Sundaram, R.M.; Vishnupriya, M.R.; Biradar, S.K.; Laha, G.S.; Reddy, G.A.; Rani, N.S.; Sarma, N.P.; Sonti, R.V. Marker assisted introgression of bacterial blight resistance in Samba Mahsuri, an elite indica rice variety. *Euphytica* **2007**, *160*, 411–422. [[CrossRef](#)]
31. Iftekharuddaula, K.M.; Newaz, M.A.; Salam, M.A.; Ahmed, H.U.; Mahbub, M.A.A.; Septiningsih, E.M.; Collard, B.C.; Sanchez, D.L.; Pamplona, A.M.; Mackill, D. Rapid and high-precision marker assisted backcrossing to introgress the SUB1 QTL into BR11, the rainfed lowland rice mega variety of Bangladesh. *Euphytica* **2010**, *178*, 83–97. [[CrossRef](#)]
32. Srichant, N.; Chankaew, S.; Monkham, T.; Thammabenjapone, P.; Sanitchon, J. Development of Sakon Nakhon rice variety for blast resistance through marker assisted backcross breeding. *Agronomy* **2019**, *9*, 67. [[CrossRef](#)]
33. Kranto, S.; Chankaew, S.; Monkham, T.; Theerakulpisuta, P.; Sanitchon, J. Evaluation for salt tolerance in rice using multiple screening methods. *J. Agric. Sci. Technol.* **2016**, *18*, 1921–1931.

34. Dellaporta, S.L.; Wood, J.; Hicks, J.B. A plant DNA miniprep: Version II. *Plant Mol. Boil. Rep.* **1983**, *1*, 19–21. [[CrossRef](#)]
35. Yoshida, S.; Forno, D.A.; Cock, J.H.; Gomez, K.A. Routine procedure for growing rice plants in cultures solution. *Lab. Man. Physiol. Stud. Rice* **1976**, *17*, 60–65.
36. International Rice Research Institute. *Standard Evaluation System Manual*; International Rice Research Institute: Manila, Philippines, 1996.
37. International Rice Research Institute. *Standard Evaluation System for Rice (SES). Leaf blast*; International Rice Research Institute: Manila, Philippines, 2002; p. 15.
38. Sriseadka, T.; Wongpornchai, S.; Kitsawatpaiboon, P. Rapid method for quantitative analysis of the aroma impact compound, 2-Acetyl-1-pyrroline, in fragrant rice using automated headspace Gas chromatography. *J. Agric. Food Chem.* **2006**, *54*, 8183–8189. [[CrossRef](#)] [[PubMed](#)]
39. Havre, G.N. The flame photometric determination of sodium, potassium and calcium in plant extracts with special reference to interference effects. *Anal. Chim. Acta* **1961**, *25*, 557. [[CrossRef](#)]
40. Wongsaprom, C.; Sirithunya, P.; Vanavichit, A.; Pantuwan, G.; Jongdee, B.; Sidhiwong, N.; Lanceras-Siangliw, J.; Toojinda, T. Two introgressed quantitative trait loci confer a broad-spectrum resistance to blast disease in the genetic background of the cultivar RD6 a Thai glutinous jasmine rice. *Field Crop. Res.* **2010**, *119*, 245–251. [[CrossRef](#)]
41. Pinta, W.; Toojinda, T.; Thummabenjapone, P.; Sanitchon, J. Pyramiding of blast and bacterial leaf blight resistance genes into rice cultivar RD6 using marker assisted selection. *Afr. J. Biotechnol.* **2013**, *12*, 4432–4438. [[CrossRef](#)]
42. Nan, M.S.A.; Janto, J.; Sribunrueang, A.; Monkham, T.; Sanitchon, J.; Chankaew, S. Field evaluation of RD6 introgression lines for yield performance, blast, bacterial blight resistance, and cooking and eating Qualities. *Agronomy* **2019**, *9*, 825. [[CrossRef](#)]
43. Suwanich, K. Geology and geological structure of potash and rock salt deposits in Chalerm Phrakiat District, Nakhon Ratchasima Province in Northeastern Thailand. *Kasetsart J. (Nat. Sci.)* **2010**, *44*, 1058–1068.
44. Chunthaburee, S.; Dongsansuk, A.; Sanitchon, J.; Pattanagul, W.; Theerakulpisut, P. Physiological and biochemical parameters for evaluation and clustering of rice cultivars differing in salt tolerance at seedling stage. *Saudi J. Boil. Sci.* **2015**, *23*, 467–477. [[CrossRef](#)]
45. Suriya-arunroj, D.; Supapoj, N.; Vanavichit, A.; Toojinda, T. Screening and selection for physiological characters contributing to salinity tolerance in rice. *Kasetsart J. (Nat. Sci.)* **2005**, *39*, 174–185.
46. Kao, C.H. Mechanisms of salt tolerance in rice plants: Na<sup>+</sup> transporters. *Crop Environ. Bioinform.* **2015**, *12*, 113–119.
47. Miyakawa, S. Expansion of an improved variety into rain-fed rice cultivation in Northeast Thailand. *Southeast Asian Stud.* **1995**, *33*, 187–203.
48. Jongdee, B.; Pantuwan, G.; Fukai, S.; Fischer, K. Improving drought tolerance in rainfed lowland rice: An example from Thailand. *Agric. Water Manag.* **2006**, *80*, 225–240. [[CrossRef](#)]
49. Vasudevan, K.; Cruz, C.M.V.; Gruissem, W.; Bhullar, N.K. Large scale germplasm screening for identification of novel rice blast resistance sources. *Front. Plant Sci.* **2014**, *5*, 505. [[CrossRef](#)]
50. Wang, R.; Fang, N.; Guan, C.; He, W.; Bao, Y.-M.; Zhang, H. Characterization and fine mapping of a blast resistant gene Pi-jnw1 from the japonica rice landrace Jianganwan. *PLoS ONE* **2016**, *11*, e0169417. [[CrossRef](#)]
51. Goto, K. Estimating losses from rice blast in Japan. In *The Rice Blast Disease*; Johns Hopkins Press: Baltimore, MD, USA, 1965; pp. 195–202.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).





Article

# Identification of QTLs Controlling Resistance/Tolerance to *Striga hermonthica* in an Extra-Early Maturing Yellow Maize Population

Baffour Badu-Apraku \*, Samuel Adewale, Agre Patern, Melaku Gedil and Robert Asiedu

International Institute of Tropical Agriculture (IITA), PMB 5320, Ibadan 200001, Nigeria; S.Adewale@cgiar.org (S.A.); P.Agre@cgiar.org (A.P.); M.Gedil@cgiar.org (M.G.); R.Asiedu@cgiar.org (R.A.)

\* Correspondence: b.badu-apraku@cgiar.org; Tel.: +234-8108482590

Received: 12 July 2020; Accepted: 6 August 2020; Published: 10 August 2020

**Abstract:** *Striga hermonthica* parasitism is a major constraint to maize production in sub-Saharan Africa with yield losses reaching 100% under severe infestation. The application of marker-assisted selection is highly promising for accelerating breeding for *Striga* resistance/tolerance in maize but requires the identification of quantitative trait loci (QTLs) linked to *Striga* resistance/tolerance traits. In the present study, 194 F<sub>2:3</sub> families of TZEI 79 × TZDEI 11 were screened at two *Striga*-endemic locations in Nigeria, to identify QTLs associated with *S. hermonthica* resistance/tolerance and underlying putative candidate genes. A genetic map was constructed using 1139 filtered DARSeq markers distributed across the 10 maize chromosomes, covering 2016 cM, with mean genetic distance of 1.70 cM. Twelve minor and major QTLs were identified for four *Striga* resistance/tolerance adaptive traits, explaining 19.4%, 34.9%, 14.2% and 3.2% of observed phenotypic variation for grain yield, ears per plant, *Striga* damage and emerged *Striga* plants, respectively. The QTLs were found to be linked to candidate genes which may be associated with plant defense mechanisms in *S. hermonthica* infested environments. The results of this study provide insights into the genetic architecture of *S. hermonthica* resistance/tolerance indicator traits which could be employed for marker-assisted selection to accelerate efficient transfer host plant resistance genes to susceptible genotypes.

**Keywords:** maize (*Zea mays* L.); *Striga* resistance/tolerance; QTL mapping; F<sub>2:3</sub> biparental mapping; Marker-assisted selection

## 1. Introduction

Maize (*Zea mays* L.) is the most widely grown staple food crop in sub-Saharan Africa (SSA), and accounts for a large proportion of carbohydrates, proteins, lipids and vitamins for millions of people in the sub-region [1,2]. The root hemi-parasitic plant *Striga hermonthica* is an important biotic constraint limiting maize production in SSA. The *S. hermonthica* problem in SSA is the result of the shift from traditional cereal based farming system which facilitated longer fallow periods which ensured that the soil *Striga* seed bank was maintained at levels that plants could tolerate [1]. However, the pressure on agricultural land has necessitated land use intensification, cereal mono-cropping and reduced fallow periods resulting in increased *Striga* seed bank and infestation levels that threaten the livelihood of millions of farmers [3,4]. De Groote et al. [5] reported that over six million hectares of agricultural land in Western, Eastern and Southern Africa are seriously affected by *Striga*. Reduction in grain yield due to *S. hermonthica* parasitism may be up to 100% under severe infestation and unfavorable environmental conditions such as low soil fertility, erratic rainfall patterns and low-input conditions [6–8]. The subsistence farmers are usually the most severely affected. The extent to which *S. hermonthica* affects the growth of its host varies tremendously, depending on the level of host plant resistance/tolerance, extent of infestation, and the prevailing environmental

conditions [9,10]. Resistance to *Striga* denotes the capability of the host plant to induce the germination of *Striga* seeds but prevents the parasite from attaching to the roots of the maize plants or kills the attached parasitic plants. Under *S. hermonthica* infestation, the resistant genotype supports considerably fewer *Striga* plants and produces a greater yield than the susceptible genotype [11–13]. Contrarily, a *Striga* tolerant genotype supports as many *Striga* plants as the sensitive or susceptible genotype [14] but produces more dry matter and shows fewer damage symptoms [15]. *Striga* damage in maize is used as the indicator of tolerance while emerged *Striga* plants is the indicator of resistance. The identification of maize genotypes that combine outstanding levels of resistance and tolerance is a promising breeding strategy and has been recommended for *Striga* resistance breeding in several studies [12–14,16,17]. In selecting for tolerance or resistance and high grain yield under *Striga* infestation, the primary traits of interest are the *Striga* damage and number of emerged *Striga* plants. Presently, maize genotypes with combined resistance and tolerance to *S. hermonthica* (possessing both low *Striga* damage and few *Striga* emergence counts) as well as high grain yield have been identified in the International Institute of Tropical Agriculture - Maize Improvement Program (IITA-MIP). *Striga* damage rating score is positively associated with *Striga* emergence counts, and the two traits are negatively associated with yield under *S. hermonthica* infested conditions. Similarly, large positive additive genetic correlation was recorded between grain yield and ears per plant as well as moderately large negative genetic correlations between grain yield and flowering traits [18]. Comparable results were reported by earlier workers [19,20]. Nevertheless, the genotypic correlation between *S. hermonthica* damage rating and *S. hermonthica* emergence counts have been found to be low, implying that different genes control the inheritance of the two traits [18,21].

*Striga* infestation is dependent on *Striga* seedbank in the soil resulting from continuous cropping of host plants, leading to the accumulation of *Striga* seeds which can remain dormant in the soil for more than a decade [22]. The germination of *Striga* seed is induced by the production of plant hormones called strigolactones produced by the maize plant in the roots. The hormones are released when the plant is under stress [23]. For the germinated *Striga* seed to survive as an obligate parasite, it must produce haustoria that attach to the roots of the maize plant through which it draws water and photosynthates [24]. Even though *Striga* possesses chlorophyll for photosynthesis, it still depends on its host for survival [25], as a result of its inability to accumulate enough photosynthates for autotrophic growth. Therefore, *Striga* establishes direct xylem links with the root system of the host [26] to obtain nutrients from its xylem sap [27]. Furthermore, a higher rate of transpiration in *Striga* compared to the host plant creates a water potential gradient from the host to the *Striga* plant [28]. *Striga* phytotoxic effects on the maize plant direct the partitioning of assimilates into the roots rather than the shoots for grain filling, thereby resulting in plant biomass and yield reduction [29].

Presently, management strategies of *S. hermonthica* include cultural, chemical, and biological approaches, which are non-economical and/or knowledge-intensive for subsistence farmers [6]. Planting of *Striga*-resistant maize varieties is presently considered the best control strategy and easy to adopt or deploy, particularly when combined with other management practices [30–33]. Resistance to *S. hermonthica* parasitism is mainly attributed to low production of *Striga* germination stimulants by the host plant, attachment of few *Striga* plants to the roots of the host plant as well as fewer *Striga* emergence [13,34]. When breeding for *S. hermonthica* resistance in maize, a combination of these resistance mechanisms is desirable in achieving effective and durable resistance [33]. The slow rate of development and deployment of *Striga* resistant genotypes is largely attributable to the complex genetics of resistance as well as limited knowledge of the specific mechanisms associated with resistance to *Striga* [35]. The resistance to *S. hermonthica* in maize is regulated by many genes or quantitative trait loci (QTL) with small additive effects and it is significantly influenced by the environment [13,36]. Therefore, breeding for *Striga*-resistant cultivars using conventional approaches by selecting maize cultivars with enhanced resistance which requires evaluation in multi-locations and years, has been less effective and time-consuming [27].

Marker-assisted breeding makes use of genotypic data in the identification of genotypes possessing desirable alleles, using linked genetic markers. Breeders employ marker-assisted selection (MAS) when an important trait that is difficult to assess phenotypically, is tightly linked to a molecular DNA marker that can be scored quickly and precisely [37]. QTL mapping approaches are important genomic tools employed in dissecting the genetic architecture of complex traits [38,39] as well as identification of genetic linkage through wide genotyping of a panel of germplasm displaying contrasting phenotypes across different environments [40]. For QTL identification, the development of next-generation sequencing technology has become a practicable technique to rapidly identify large number of single nucleotide polymorphisms (SNPs) throughout the genome [41]. Unlike the use of second-generation molecular markers which result in low-quality mapping, SNP markers provide new insights to rapidly identify QTL of interest.

Information on map positions of genes and linked markers on a chromosome are crucial for efficient determination of genetic architecture of polygenic traits in crop plants [42]. Several QTLs and candidate genes controlling resistance to *Striga* have been reported in cereals. For instance, Swarbrick et al. [32] identified three *S. hermonthica*-resistant QTLs in Kasalath–Koshihikari rice backcross inbreds, two of these QTLs originated from the Kasalath allele and one from the Koshihikari allele. The largest-effect QTL (Kasalath-derived allele) explained 16% phenotypic variance in the mapping population and was located on linkage group 4. Haussmann et al. [43] detected molecular markers associated with *S. hermonthica*-resistant QTLs, with the most significant QTL corresponding to the major-gene locus low germination stimulant (LGS) in linkage group I. Five genomic regions (QTLs) linked to stable *S. hermonthica*-resistant alleles from resistant variety N13 were detected through evaluation across a large number of field trials in Mali and Kenya. However, limited reports are available on the QTLs and genes controlling *Striga* resistance in maize. In a recent study by Adewale et al. [44] to identify molecular markers linked to *S. hermonthica* resistance in maize, 24 SNPs significantly associated with *S. hermonthica* resistance indicator traits under artificial *S. hermonthica* infestation were detected. The authors also identified four candidate genes on chromosomes 3, 5, 9 and 10, with functions closely associated with maize plant defense mechanisms against *S. hermonthica* parasitism. Identification of QTLs linked to *S. hermonthica* resistance followed by gene introgression into elite genetic backgrounds, has the potential to reduce yield losses due to *Striga* and will ultimately provide solid foundation for improving *Striga* resistance [45].

The objectives of this study were to identify QTL and underlying candidate genes conferring resistance/tolerance to *S. hermonthica* in maize using F<sub>2,3</sub> biparental mapping population derived from a cross between a *Striga* resistant inbred line, TZEEI 79 and *Striga* susceptible inbred TZdEEI 11.

## 2. Materials and Methods

### 2.1. Germplasm and Phenotyping

Based on the reports of previous studies, two extra-early maturing yellow inbred lines, TZEEI 79 (*Striga* resistant/tolerant) and TZdEEI 11 (*Striga* susceptible) were selected as parents to generate the F<sub>2,3</sub> progenies used in the present study [1]. TZEEI 79 is an outstanding *S. hermonthica* resistant/tolerant, drought and low-soil N tolerant inbred line developed in the IITA-MIP from the broad-based *S. hermonthica* resistant/tolerant as well as drought and low-soil N tolerant population, TZEE-Y Pop STR C<sub>0</sub>. TZEEI 79 has significant positive GCA (general combining ability) for grain yield as well as significant negative GCA effects for *Striga* damage and *Striga* emergence counts under *Striga* infestation and has been extensively used in the IITA hybrid program as an important resource for developing high-yielding, multiple stress tolerant hybrids as well as an efficient tester for classifying other inbreds into heterotic groups [46]. Crosses were made between TZEEI 79 and TZdEEI 11 designated as P<sub>1</sub> and P<sub>2</sub> respectively, to generate 220 F<sub>1</sub> progenies. The F<sub>1</sub> progenies and the parental lines were planted, and leaf samples were collected at 3 weeks after planting. Verification of the parental type alleles (quality control analysis) was carried out on the F<sub>1</sub> progenies prior to advancement to F<sub>2</sub>.

The F<sub>1</sub> progenies were screened using two SSR primers (bnlg 182 and umc 1568) which were found to be polymorphic between the two parents. The analysis identified 170 true-to-type F<sub>1</sub> hybrids which were advanced to F<sub>2</sub>. The 170 true-to-type F<sub>2</sub> ears were planted ear-to-row and 194 F<sub>2</sub> individuals which were randomly selected and selfed were used in the present study.

The F<sub>2,3</sub> progenies and the two parental lines were screened under artificial *S. hermonthica* infestation at Mokwa (9°18' N, 5°4' E, 210 m above sea level, 1100 mm yearly rainfall, luvisol soil) and Abuja (9°16' N, 7°20' E, 445 m above sea level, 1500 mm yearly rainfall, ferric-luvisol soil) in the Southern Guinea savanna of Nigeria in 2018. At each experimental site, the trial was laid out using randomized incomplete block design (14 × 14 lattice) with two replicates. The experimental units were 3 m long single-row plots, with an inter-row spacing of 0.75 m and within-row spacing of 0.4 m, to achieve a target population density of 66,666 plants/ha. The fields for artificial *S. hermonthica* infestation at Mokwa and Abuja were treated with ethylene gas at 2 weeks before planting to eliminate any potential *Striga* seeds present in the soil. The *S. hermonthica* seeds used for the experiment were obtained from sorghum farms around the test locations at Abuja and Mokwa in 2017. The artificial *S. hermonthica* infestation was carried out as proposed by the IITA Maize Program [16]. Briefly, about a week before inoculation, the *S. hermonthica* seeds were carefully mixed with finely sieved sand at the ratio 1:99 by weight to ensure rapid and uniform infestation. A standard scoop calibrated to deliver approximately 5000 germinable seeds per hill was utilized for the artificial infestation. Three maize seeds were planted per infested hill and the seedlings were later thinned to two plants per stand at 2 weeks after emergence. Fertilizer application on the maize plots was delayed till about 30 days after planting, in order to subject the maize plants to stress, a condition that was expected to enhance strigolactone production. This ensured good germination of *Striga* seeds and attachment of *Striga* plants to the roots of host plants. At this plant growth stage, 20–30 kg Nha<sup>-1</sup>, 30 kg each of P and K were applied as NPK 15-15-15, taking into consideration the fertility status of the soil. Reduction in the fertilizer application rate was important because *Striga* emergence decreases at high N rate [16]. At 10 weeks after planting, typical symptoms of *Striga* infestation on the host plants were observed, such as chlorosis, leaf scorching (firing) and blotching, stunting, decrease in ear and tassel size, brown necrotic spots, leaf wilting and rolling, stalk lodging, open-tip of ears at late growing stage, and premature death of host plants. Host plant *Striga* damage severity was scored using a scale of 1 to 9. Rating scales 1–5 indicated resistance while 6–9 indicated susceptibility, where 1 = normal plant growth, no obvious symptoms, and 9 = all leaves completely scorched, collapse of host plants and no ear formation [16]. In addition, data were collected on *Striga* emergence count at 10 WAP as the number of *Striga* plants thriving on the maize root system as well as ears per plant (EPP) by dividing the total number of ears harvested per plot by the number of plants in a plot at harvest. Grain moisture was determined using Kett moisture tester PM-450 and grain yield (kg/ha) was calculated using the field weight of harvested ears per plot, adopting a shelling percentage of 80, adjusted to 15% moisture content [47].

## 2.2. DArTseq Genotyping and SNP Data Filtering

Young and healthy leaf samples from single plants of the F<sub>2</sub> individuals and bulk samples from the parental lines were collected and frozen immediately after harvesting using liquid nitrogen and thereafter stored at –80 °C. Genomic DNA extraction was carried out following the DArT protocol ([www.diversityarrays.com/files/DArT\\_DNA\\_isolation/](http://www.diversityarrays.com/files/DArT_DNA_isolation/)). The extracted DNA was assessed for quality by visualization on agarose gel (2% w/v) and the quantity was estimated on NanoDrop-1000 spectrophotometer (NanoDrop, Wilmington, DE, USA) using the absorbance ratio A<sub>260</sub>/A<sub>280</sub> to determine the concentration (ng/μL) and purity level of the DNA.

Genotyping of the 194 F<sub>2</sub> individuals plus the two parents was carried out using DArTseq technology [48,49]. Genome complexity reduction which involved the use of a combination of two restriction enzymes (*Pst*I–*Mse*I) was used to create a genome representation of the analyzed samples. All fragments generated were amplified and sequenced to identify the single nucleotide polymorphisms (SNPs) using a proprietary analytical pipeline developed by DArT P/L. After a strict quality control process, which included parameters such as call rate, data reproducibility (~20% of samples replicated), and rate of monomorphism to eliminate monomorphic markers, 9951 SNPs were extracted from the evaluated germplasm. The 9951 SNPs were filtered for unmapped markers, duplicate markers and markers segregating between the *Striga* resistant and *Striga* susceptible parents. A total of 1139 high-quality DArTseq markers distributed across the 10 maize chromosomes were retained for the construction of genetic linkage map as well as the QTL mapping.

### 2.3. Data Analysis

The data recorded on emerged *Striga* counts and *Striga* damage severity scores were subjected to natural logarithm transformation. Thereafter, data collected on grain yield, ears per plant, *Striga* damage as well as *Striga* emergence counts were tested for normality using Shapiro–Wilk’s (*W*) test [39,50] before analysis of variance. Box plots were made to visualize the distributions of grain yield and other traits under each research environment using ggplot2 library [51]. Analysis of variance was conducted across research environments using the general linear model procedure (PROC GLM) implemented in the Statistical Analytical System (SAS), version 9.3 [52]. In the analysis, environment, replications (environments), blocks (replications × environments) were considered as random and the F<sub>2:3</sub> families (genotypes) as fixed effects. Estimates of broad sense heritability of the traits (*H*<sup>2</sup>) across research environments were computed on a family-mean basis as proposed by Holland et al. [53], using the following formula:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{ge}^2}{e} + \frac{\sigma_e}{re}}$$

where  $\sigma_g^2$  = variance component due to the genotypes,  $\sigma_{ge}^2$  = genotype × environment variance,  $\sigma_e$  = experimental error variance; *e* = number of environments, and *r* = number of replications within environment.

Correlation coefficients were estimated among the traits with the adjusted means of the F<sub>2:3</sub> families using the Ggally function implemented in GGally package [54]. Furthermore, the mixed linear model (MLM) established in META-R software [55] was used to compute the best linear unbiased estimates (BLUES) for each genotype in each and across environments which were for the QTL analysis. The R/qtl was used to construct a linkage map [56]. Markers that were identical across all genotypes were identified and eliminated as duplicates. Furthermore,  $\chi^2$ -test for goodness-of-fit (*p* ≤ 0.0001) was used to identify markers with distorted segregation patterns [50,57,58]. Markers with significant deviation from the expected Mendelian segregation ratio (1:2:1) for F<sub>2:3</sub> population were excluded from the analysis, resulting in a total of 1139 SNP markers used for the genetic map construction and QTL analysis.

### 2.4. QTL Analysis and Candidate Gene Identification

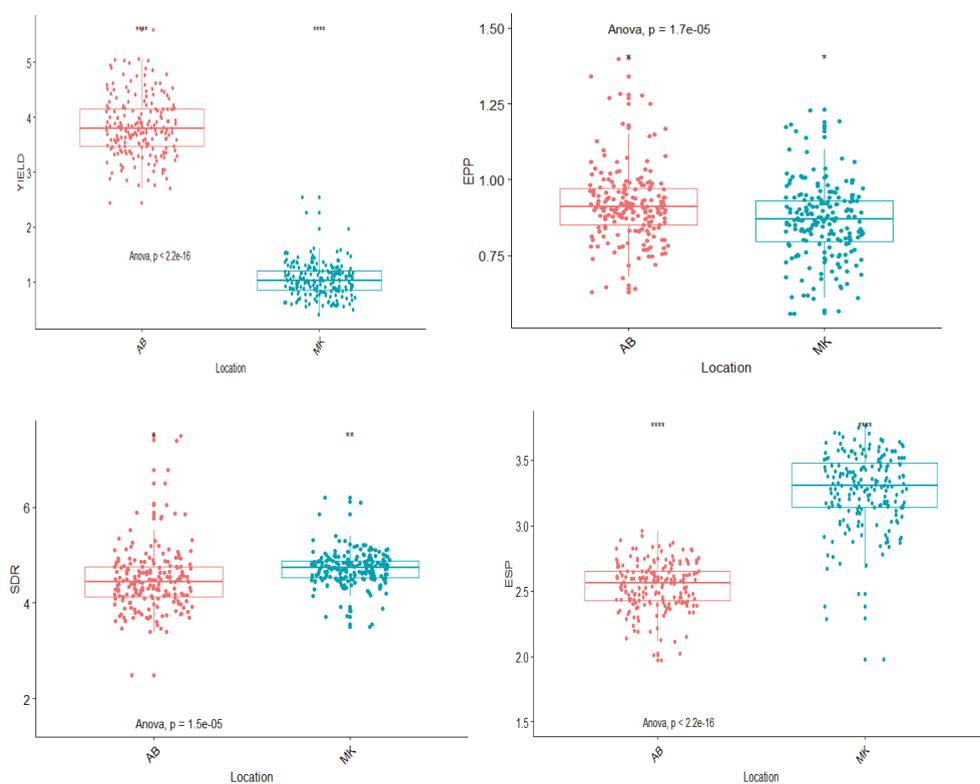
Quantitative trait loci (QTL) mapping for each and across environments was carried out for four *Striga* resistance/tolerance adaptive traits (grain yield, *Striga* damage at 10 WAP, emerged *Striga* counts at 10 WAP and ears per plant (EPP)), using R/qtl package with the composite interval mapping (CIM) algorithms as proposed by Wang et al. [50]. The statistical significance of the QTL was assessed using permutation tests (1000 replications) for all traits. A logarithm of odds (LOD) of 3.0 was set through the permutation test to identify significant QTLs for the traits [59]. The additive effects and proportion of phenotypic variance explained (PVE) by each QTL were estimated using the “fitqtl” function of R version 3.3.4. The sign of the effect of each QTL was used to identify the origin

of the favorable alleles [60]. The potential locations of the QTLs were described according to their LOD peaks and their surrounding regions. Identified QTLs were named based on conventions method described by Bo et al. [61]. For example, *qepp-2* represented the QTL identified for number of ears per plant on chromosome 2. Putative candidate genes were searched within a 2.0 Mb interval downstream and upstream of the significant associated SNPs using the MaizeGDB database version (RefGen\_v4).

### 3. Results

#### 3.1. Phenotypic Analysis of Grain Yield and Other *Striga* Resistance Adaptive Traits

The 194  $F_{2,3}$  families and the two parental lines were evaluated under artificial *Striga* infestation to assess variation in their level of resistance. The distributions of grain yield, *Striga* damage, number of emerged *Striga* plants as well as ears per plant in the  $F_{2,3}$  population are displayed in Figure 1. Significant variation was detected among the genotypes under each and across research environments (Figure 1, Table S1). The performance of the genotypes ( $F_{2,3}$  families and the parental lines) for *Striga* emergence count and ears per plant were not significantly influenced by the environment whereas grain yield and *Striga* damage displayed significant genotype  $\times$  environment interactions. The two parental lines TZEEI 79 and TZdEEI 11 differed significantly and consistently in their performance under artificial *Striga* infestation, and phenotypic values for each trait of segregating population displayed wide ranges (Table 1). Transgressive segregation was observed for all traits in that some of the  $F_{2,3}$  families showed higher and lower levels of grain yield, *Striga* damage, number of emerged *Striga* plants and ears per plant compared to the parental lines (Table 1). The *Striga* resistant inbred line TZEEI 79 exhibited high grain yield and ears per plant as well as reduced *Striga* damage and *Striga* emergence count whereas the *Striga* susceptible line TZdEEI 11 showed significantly lower grain yield and ears per plant as well as increased *Striga* damage and emergence count. Grain yield (kg/ha) across the  $F_{2,3}$  population varied from 1070.1 to 4113.9, with a mean of 2439.2 (Table 1). In addition, individual means varied from 0.5 to 1.4, 2.4 to 7.5, 0.4 to 3.7 for ears per plant, *Striga* damage and *Striga* emergence count, respectively. Broad-sense heritability estimates of the traits derived from the variance components varied from 0.47 for *Striga* damage to 0.70 for ears per plant. The normality tests by Shapiro–Wilk ( $W$ ) revealed that the distributions of grain yield and *Striga* damage phenotypic data were normally distributed while those of *Striga* emergence counts and ears per plant were not (Table S2). High  $W$ -test values were obtained for all studied traits ranging from 0.97–0.99 (Table S2). Correlation analysis revealed significant and positive correlations between number of ears per plant and grain yield whereas negative and significant correlations were observed between the ears per plant and *Striga* damage as well as between *Striga* damage and grain yield (Figure 2).

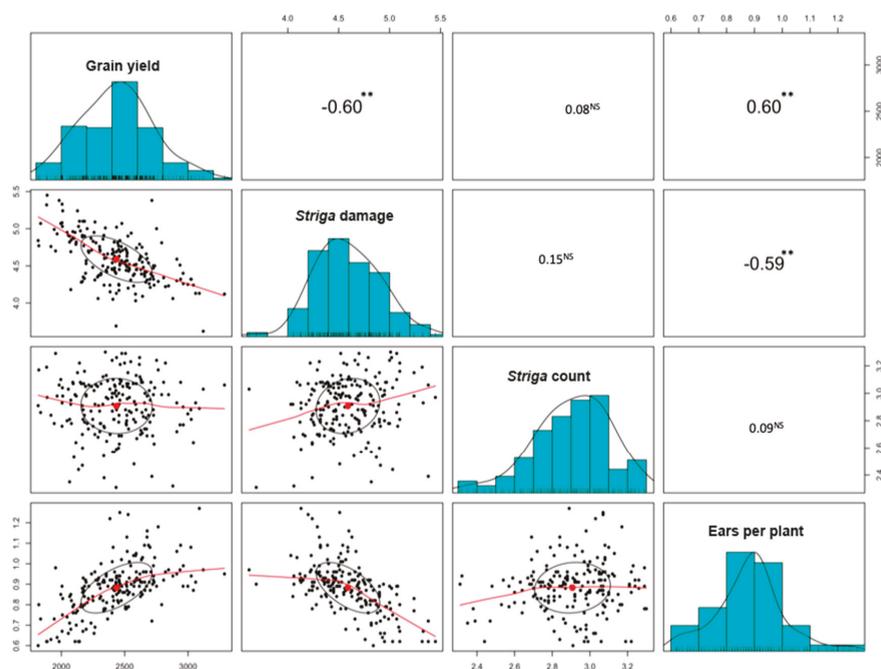


**Figure 1.** Box plots showing the distribution of (A) grain yield (YIELD, t/ha), (B) emerged *Striga* plants (ESP), (C) *Striga* damage rating (SDR) and (D) number of ears per plant (EPP) under artificial *Striga* infestation at Mokwa (MK) and Abuja (AB) in 2018. The points represent the F<sub>2,3</sub> families and the parental genotypes.

**Table 1.** Descriptive statistics of *Striga* resistance indicator traits of parents and F<sub>2,3</sub> population derived from the cross between TZEEI 79 × TZdEEI 11 across *Striga*-infested environments.

	Grain Yield, kg/ha	Ears per Plant	<i>Striga</i> Damage	<i>Striga</i> Count
Parents				
TZEEI 79	2517.5	0.9	2.5	0.5
TZdEEI 11	1485.9	0.8	7.0	3.2
F <sub>2,3</sub> population				
Range	1070.1–4113.9	0.5–1.4	2.4–7.5	0.4–3.7
Mean ± SE	2439.2 ± 578.9	0.9 ± 0.14	4.6 ± 0.7	2.5 ± 0.5
H <sup>2</sup>	0.49	0.70	0.47	0.48
CV (%)	25.2	21.3	18.2	25.4

H<sup>2</sup>—broad sense heritability, CV—coefficient of variation.



**Figure 2.** Correlation between grain yield and other *Striga* resistance indicator traits in an  $F_{2:3}$  mapping population derived from TZEEI 79  $\times$  TZdEEI 11 under artificial *Striga* infestation. The axis displayed the range of values obtained for each trait; black circles represent the most predominant values for each trait among the genotypes while the red lines represent the direction of the relationship between two traits. \*\* = 0.01 and indicates highly significant.

### 3.2. Linkage Map

A genetic linkage map containing 1139 SNPs mapped on the 10 maize chromosomes was constructed (Figure S1, Table 2). The resulting map spanned a total genetic distance of 2016 cM, with mean interlocus distance of 1.70 cM. The average genetic distances between successive markers ranged from 0.86 cM to 11.86 cM for chromosomes 1 and 7 respectively.

**Table 2.** Summary statistics of the linkage map.

Linkage Group	Number of Markers	Genetic Length (cM)	Average Marker Interval (cM)
1	349	304.3	0.86
2	188	229.3	1.22
3	148	224.5	1.52
4	73	243.0	3.36
5	167	221.7	1.30
6	15	167.4	7.69
7	16	178.4	11.86
8	46	160.1	3.37
9	28	139.2	4.74
10	109	148.1	1.36
Total	1139	2016.0	1.70

### 3.3. QTL Detection and Identification of Potential Candidate Genes

Through the QTL analysis, a total of 12 QTLs with significant LOD score  $\geq 3.0$ , were identified for the four *Striga* resistance indicator traits using the integrated genetic map and mean phenotypic data across research environments (Table 3). The 12 QTLs identified included three for grain yield, five for ears per plant, three for *Striga* damage and one for *Striga* emergence counts. The proportion of phenotypic variation explained by the QTLs varied from 2.0% for *qepp-2.1* to 13.5% for *qepp-1*. Three QTLs *qgy-1.1*, *qgy-2.1* and *qgy-7* detected for grain yield explained 5.6, 10.3 and 2.3% phenotypic variation, respectively. Furthermore, five QTLs *qepp-1*, *qepp-2.1*, *qepp-3*, *qepp-7*, and *qepp-8.1* were identified for ears per plant, explaining phenotypic variation ranging from 2.0 to 13.5%. Similarly, QTLs *qsd-2*, *qsd-5.1* and *qsd-7* detected for *Striga* damage displayed phenotypic variance of 8.0, 3.0 and 3.2 respectively. The only QTL (*qsc-3.1*) identified for *Striga* emergence count had PVE of 3.1. The QTL *qsd-7* and *qepp-7* were detected at the same position for *Striga* damage and ears per plant. Similarly, *qgy-2.1* and *qsd-2* detected for grain yield and number of emerged *Striga* plants were consistently identified at the same position in each of the two locations. Three major QTL genomic regions were detected on chromosomes 1, 2, and 8 with flanking marker intervals 216–226 cM, 134–156 cM and 35–38 cM, respectively (Figure S2). In all cases, favorable alleles for *Striga* resistance/tolerance were contributed by the *Striga* resistant inbred line TZEI 79.

A total of 116 protein coding genes were identified within 2.0 Mb interval downstream and upstream of the significantly associated SNPs (Table S3). Of the 116 candidate genes, 17 key candidate genes associated with the identified QTL for *Striga* resistance/tolerance indicator traits under artificial *Striga* infested environments are presented in Table 4. For grain yield, the *qgy-1.1* was found associated with GRMZM2G408305 which encodes ARM repeat superfamily protein as well as GRMZM2G072376 which encodes bHLH-transcription factor 56. The QTL *qgy-7*, *qepp-7* and *qsd-7* were linked to GRMZM6G199466 (*hsp3*—heat shock protein3), GRMZM2G008234 (*ereb114*—AP2-EREBP-transcription factor 114) as well as GRMZM2G044194 (*phytosulfokine peptide precursor1*). Similarly, for ears per plant, *qepp-1* was found associated with GRMZM2G324999 which encodes the WRKY-transcription factor 25; *qepp-2.1* was associated with GRMZM2G174784 (*ERE197*—putative AP2-EREBP transcription factor superfamily protein), GRMZM2G174917 (*ereb47*—AP2-EREBP-transcription factor 47) and GRMZM2G131961 (*bzip27*—bZIP-transcription factor 27); *qepp-3* was linked to *Zma-MIR167g* which promotes lateral root development in plants. On chromosome 8, QTL *qepp-8.1* detected for ears per plant was linked to the genes GRMZM2G051528 which encodes myb transcription factor95 and GRMZM2G053503 which encodes ethylene-responsive factor-like protein. For *Striga* damage, QTL *qsd-5.1* was associated with genes GRMZM2G059851 which encodes the heat shock factor protein as well as GRMZM2G099334 which encodes myb3—WD40 repeat protein. The QTL *qsc-3.1* detected for *Striga* emergence count was found associated with the genes GRMZM2G054050 which encodes multicopper oxidase protein, GRMZM2G162709 (*MYB*-transcription factor 137), and GRMZM2G340342 which encodes the ARM repeat superfamily protein.

**Table 3.** Summary of quantitative trait loci (QTLs) mapped in the F<sub>2,3</sub> population derived from TZEEI 79 × TZdEEI 11 under artificial *Striga* infestation.

Trait	Location	QTL	Chr	Position (cM)	Flanking Markers	LOD	Add	Dom	PVE (%)
Grain yield	Across	<i>qgy-1.1</i>	1	37.6	S1_21989679, S1_39100143, S2_133867986, S2_15644237	4.4	88.34	30.58	5.6
		<i>qgy-2.1</i>	2	141.0	S7_28123354, S7_39973949	5.5	116.80	0.49	10.3
		<i>qgy-7</i>	7	28.4	S1_19567556, S1_26003993, S2_133867986, S2_15644237	3.8	36.60	14.29	2.3
	AB	<i>qgy-1.2</i>	1	23.0	S2_133867986, S2_15644237	3.2	154.78	−52.26	0.8
		<i>qgy-2.1</i>	2	141.0	S2_133867986, S2_149550704	4.8	196.49	−43.08	1.0
	MK	<i>qgy-2.2</i>	2	141.0	S7_2840233, S7_39973949	4.5	2.80	−1.94	6.4
		<i>qgy-7</i>	7	28.4	S8_89593967, S8_96110467	3.8	1.73	−0.49	4.1
		<i>qgy-8</i>	8	93.6		4.5	1.93	4.43	9.6
	Ears per plant	Across	<i>qepp-1</i>	1	219.2	S1_216043878, S1_225511262, S2_5979544, S2_63035407	5.8	0.03	0.03
<i>qepp-2.1</i>			2	62.7	S3_121468077, S3_122362644	5.8	0.04	0.02	2.0
<i>qepp-3</i>			3	122.0	S7_28123354, S7_39973949	4.3	−0.04	0.03	6.0
<i>qepp-7</i>			7	28.4	S8_34609716, S8_37717543	6.1	0.04	−0.01	3.4
AB		<i>qepp-8.1</i>	8	37.5	S2_147705224, S2_154942934	4.5	0.01	0.01	10.0
		<i>qepp-2.2</i>	2	151.9	S2_57731037, S2_68592918	3.8	0.04	0.02	0.3
MK		<i>qepp-2.1</i>	2	62.7	S7_28123354, S7_39973949	3.4	3.63	0.61	1.2
		<i>qepp-7</i>	7	28.4	S8_30333465, S8_37493871	3.2	1.75	−1.31	3.2
		<i>qepp-8.2</i>	8	34.6		3.5	2.07	−2.31	5.5
<i>Striga</i> damage	Across	<i>qsd-2</i>	2	141.0	S2_133867986, S2_147705224	5.0	−0.13	0.05	8.0
		<i>qsd-5.1</i>	5	172.0	S5_171268215, S5_188880765	4.3	0.11	−0.03	3.0
		<i>qsd-7</i>	7	28.4	S7_28123354, S7_39973949	3.0	−0.04	0.04	3.2
	AB	<i>qsd-2</i>	2	141.0	S2_140980014, S2_149550704	3.0	−0.18	0.08	0.3
		<i>qsd-5.2</i>	5	63.5	S5_5997392, S5_67117772	4.4	−0.03	−0.04	3.0
	MK	<i>qsd-2</i>	2	141.0	S2_133867986, S2_149550704	3.0	0.23	0.11	5.2
<i>Striga</i> emergence count	Across	<i>qsc-3.1</i>	3	23.9	S3_21323318, S3_28235766	5.8	−0.09	0.03	3.2
		<i>qsc-3.2</i>	3	156.7	S3_152189313, S3_158981913	3.0	−0.02	0.03	2.0

AB—Abuja, MK—Mokwa, Across—across the two *Striga* infested environments; Add—Additive effect, LOD—Logarithm of odds, PVE—proportion of phenotypic variance explained by single QTL. Grain yield (kg/ha), ears per plant (number of ears per plant), *Striga* damage (based on rating scale 1–9) and *Striga* emergence count (number of emerged *Striga* plants).

**Table 4.** Key putative candidate genes associated with the identified QTLs for key *Striga* resistance/tolerance indicator traits under artificial *Striga* infested environments.

Trait	QTL	LG: Start-End Position *	Gene_ID	Sequence Description
Grain yield	<i>qgy-1.1</i>	1:35591728–39591728	GRMZM2G408305	ARM repeat superfamily protein
			GRMZM2G072376	bHLH—transcription factor 56
	<i>qep-1</i>	1:217230073–221230073	GRMZM2G324999	wrky25—WRKY-transcription factor 25
Ears per plant	<i>qep-2</i>	2:4272353–8272353	GRMZM2G174784	EREB197—putative AP2-EREBP transcription factor superfamily protein
			GRMZM2G174917	ereb47—AP2-EREBP-transcription factor 47
<i>Striga</i> damage			GRMZM2G131961	bzip27—bZIP-transcription factor 27
	<i>qep-3</i>	3:119393084–123393084	Zma-MIR167_5	miR167—lateral root development
	<i>qep-8.1</i>	8:35493871–39493871	GRMZM2G053503	ERF1—ethylene-responsive factor-like protein
			GRMZM2G051528	myb95—myb transcription factor95
	<i>qsd-5.1</i>	5:170001287–174001287	GRMZM2G059851	HSF-6—Heat shock factor protein
			GRMZM2G099334	myb3—WD40 repeat protein
Grain yield, ears per plant and <i>Striga</i> damage		7:840233–4840233	GRMZM6G199466	hsp3—Heat shock protein3
	<i>qgy-7, qep-7, qsd-7</i>		GRMZM2G044194	psk1—phytosulfolkine peptide precursor1
<i>Striga</i> emergence count			GRMZM2G008234	ereb114—AP2-EREBP-transcription factor 114
	<i>qsc-3.1</i>	3:21951408–25951408	GRMZM2G054050	Multicopper oxidase
			GRMZM2G340342	ARM repeat superfamily protein
		GRMZM2G162709	myb137—MYB-transcription factor 137	

\* Linkage group start and end positions within 2.0 Mb interval downstream and upstream of the significant associated SNPs.

#### 4. Discussion

Marker-assisted selection (MAS) is an efficient approach for increasing the accuracy and efficiency of selection using markers tightly linked to genes or QTLs of interest, to complement phenotypic selection [40,62]. The identification of QTLs associated with *Striga* resistance/tolerance would facilitate rapid development of *Striga* resistant/tolerant maize genotypes using MAS, due to the polygenic nature of host–parasite relationship and its interaction with environmental factors [63]. The normal distribution observed for grain yield and *Striga* damage in the present study is a result of the highly diverse genotypes segregating in the mapping population [64]. The selection of parental lines with varying levels of resistance to *Striga* allowed sufficient segregation of the traits in the population. The distribution of measured traits in the F<sub>2,3</sub> population indicated the existence of transgressive segregation (i.e., progenies performing outside the range of the parental genotypes). Transgressive segregation has been observed in populations screened under low N [65] and *Striga* infestation [66,67]. This phenomenon results from the accumulation of favorable and unfavorable alleles resulting from both parents. Moderate-to-high broad sense heritability estimates (0.47–0.70) observed for grain yield and other *Striga* resistance/tolerance indicator traits confirmed that high-quality phenotypic data were used for the genetic analysis. The moderate-to-high heritability estimates obtained in the present study implied that the observed genetic variation among the genotypes was strongly influenced by genetic factors, and that the *Striga* resistance indicator traits could be effectively improved in *Striga* resistance breeding programs. Previous studies reported moderate-to-high heritability values, ranging from 0.53–0.84 for *Striga* resistance/tolerance adaptive traits [68,69].

Linkage map density and resolution largely depend on population size and type, marker density, as well as the accuracy of genotyping [70,71]. The F<sub>2,3</sub> mapping population, developed from the cross between inbred TZEEI 79 (*Striga* resistant) and inbred TZdEEI 11 (*Striga* susceptible) was used to investigate the inheritance of *Striga* resistance/tolerance. The QTL mapping in the TZEEI 79 x TZdEEI 11 F<sub>2,3</sub> mapping population identified twelve QTLs for the *Striga* resistance indicator traits across the two research environments. These QTLs explained moderate variation of the phenotype, with values ranging from 2.0% for *qepp-2* to 13.5% for *qepp-1*. This finding confirmed the complexity of the genetic basis of *S. hermonthica* resistance [13,44]. The 12 QTLs identified included three for grain yield, five for ears per plant, three for *Striga* damage as well as one for *Striga* emergence counts. The identified QTLs were located on chromosomes 1, 2, 3, 5, 7 and 8. Similarly, Adewale et al. [44] identified markers linked to *Striga* resistance indicator traits in maize on chromosomes 1, 3, 5, 7 and 8. Samayoa et al. [72] found QTLs associated with Mediterranean corn borer resistance in maize on chromosomes 1, 5 and 6 using Recombinant Inbred Lines (RIL) population obtained from the cross B73 × CML103. In a study by Haussmann et al. [43], five genomic regions (QTLs) linked to *Striga* resistance in sorghum were reported on chromosomes 1, 2, 5 and 6. Generally, the QTLs mapped in the present study provided more information on the genetic basis of *Striga* resistance/tolerance in maize, indicating that the resistance/tolerance to *Striga* is quantitatively inherited. The additive effects of the identified QTLs indicated that favorable alleles for each QTL were contributed by either the resistant or susceptible parent, depending on the signs of the QTL additive effects. The resistant parental inbred TZEEI 79 contributed favorable alleles for resistance/tolerance to *Striga* for most of the identified QTLs. Estimates of genetic effects of the QTL indicated that additive gene action was preponderant in most cases for *Striga* resistance indicator traits.

The QTL analysis in the F<sub>2,3</sub> mapping population identified three major QTL (*qepp-1*, *qgy-2.1* and *qepp-8.1*) genomic regions on chromosomes 1, 2, and 8 with flanking marker intervals of 216–226 cM, –156 cM and 35–38 cM, respectively. Interestingly, the QTLs detected on chromosome 2 were found to be consistent across environments for grain yield and *Striga* damage. The QTL *qgy-2.1* identified for grain yield on chromosome 2 at 141.0 cM was found to be pleiotropic with QTL *qsd-2* detected for *Striga* damage. Similarly, QTL *qgy-7* at 28.4 cM detected for grain yield on chromosome 7 was pleiotropic with the QTL for *Striga* damage and number of ears per plant. The co-localization of QTLs for these traits may reflect the high correlation coefficients observed among the different *Striga*

resistance indicator traits. These two QTLs would be invaluable genomic resources for fine mapping and candidate gene discovery. The validation of a common QTL region in different environments and/or genetic backgrounds is important for application in MAS to improve breeding efficiency [73]. The QTL *qgy-2.1* and *qsd-2* from TZEEI 79 located on chromosome 2 (133.9–156.4) and identified in the two test locations in this study have not been previously reported. This QTL could be a hot spot for genes for genetic improvement of *Striga* resistance in maize. Overlapping regions of QTL on chromosome 7 (28.1–39.9) for *Striga* damage and ears per plant were identified in the present study. This common region could also provide better prospects for breeders to enhance resistance to *Striga* parasitism in maize using MAS.

Putative candidate genes associated with some of the identified QTLs for *Striga* resistance indicator traits are presented in Table 4. The gene model GRMZM2G054050 (*qsc-3.1*), associated with *Striga* emergence count encodes a multicopper oxidase Lpr-2 (low phosphate root 2) protein, whose homologous gene, Lpr-1 has been found [74,75] to regulate primary roots length under Pi (inorganic phosphate) deficient conditions. Lpr-1 and Lpr-2 play important roles in Pi sensing at root tips [75]. Similarly, the gene model GRMZM2G044194 linked to QTLs for grain yield (*qgy-7*), ears per plant (*qepp-7*), and *Striga* damage (*qsd-7*) were associated with the *psk1* (phytosulfokine peptide precursor1) gene. PSK genes have been reported to promote cell growth especially in the quiescent centre cells of the root apical meristem [76]. Similarly, the gene model GRMZM2G408305 (*qgy-1.1*) associated with grain yield encodes the ARM family proteins which promote lateral root growth in plants [77]. QTL *qepp-3* located on chromosome 3 was found to be associated with MIR167g. In Arabidopsis, soybean and maize, miR167 has been reported to play important roles in lateral root growth and architecture [78]. Under plant nutrient deprivation conditions such as *Striga* parasitism, plants alter their root systems to discover heterogeneous soil regions for nutrients. The branching of secondary roots from primary roots in plants is one of the processes through which plants efficiently obtain nutrients from the soil. The QTL *qsd-5.1* was associated with the gene GRMZM2G059851 encoding the heat-shock factor protein, HSF 6. Heat-shock proteins are ubiquitous proteins responsible for protein folding, assembly, translocation as well as degradation in response to biotic stresses, depending on the nature of the causal organisms and plant genotypes (either susceptible or resistant), as well as plant's growth stage [64,79]. In addition, Ng et al. [80] identified AP2/ERF, MYB, bHLH, WRKY as well as bZIP as major transcription factor families involved in plant defense signalling. In a recent study, Adewale et al. [44] identified four putative candidate genes GRMZM2G060216, GRMZM2G103085, GRMZM2G057243 and GRMZM2G164743 located on chromosomes 3, 5, 9 and 10, having functions related to plant defense mechanisms under *Striga* infested conditions. The identified candidate genes in the present study differ from those earlier reported. The candidate genes identified from the dissection of *qsc-3.1*, *qgy-1.1*, *qepp-3*, and *qsd-5.1* are suggestive of *Striga* resistance response mechanisms. The QTL identified in the present study would be validated in different genetic backgrounds and in different environments to verify the reproducibility for effective use in MAS breeding for resistance to *Striga*. Overall, the QTL/markers with significant association to *S. hermonthica*-resistant adaptive traits would be useful as potential candidate loci for the enhancement of *Striga* resistance in maize. The application of these markers for selection would lead to the elimination of the bulk of *Striga* susceptible genotypes, which in turn may significantly reduce the number and cost of screening required to improve maize for *Striga* resistance. Based on our results, we initiated a program aimed at developing extra-early mapping populations from different genetic backgrounds so that putative markers identified in our studies could be validated and deployed in maize breeding programs through MAS.

## 5. Conclusions

A total of 12 QTLs associated with *S. hermonthica* resistance/tolerance traits in maize were identified across *Striga* infested environments in the present study. The identified QTLs displayed varying contributions to phenotypic expression and are in regions that play roles which may be associated

with plant defense response under *Striga* infestation in maize. The co-localization of QTL for grain yield and other traits indicated strong associations between the traits. The QTLs mapped in this study could be candidates for marker-assisted introgression of *Striga* resistance/tolerance genes in maize, after validation in different genetic backgrounds and in different environments.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4395/10/8/1168/s1>, Table S1: Mean squares of  $F_{2:3}$  mapping population evaluated under artificial *Striga* infestation at both Abuja and Mokwa in 2018 growing season. Table S2: Shapiro-Wilk's normality tests for *Striga* resistance/tolerance indicator traits for  $F_{2:3}$  population derived from the cross between TZEEL 79 (*Striga* resistant) and TZdEEI 11 (*Striga* susceptible). Table S3: Candidate genes associated with the identified QTL for key *Striga* resistance/tolerance indicator traits under artificial *Striga* infestation. Figure S1: Linkage map of  $F_{2:3}$  mapping population based on 1139 DArTseq markers. Left bar of the linkage map indicates cM distance while right bar of linkage map displayed the marker names. Red bars and letters indicate QTL identified across *Striga* infested environments. Figure S2. Major QTL identified for *Striga* resistance in the extra-early yellow mapping population. A likelihood of odds (LOD) scan showing the QTL identified on chromosomes 1, 2, and 8 explaining  $\geq 10\%$  phenotypic variation.

**Author Contributions:** Conceptualization, B.B.-A.; Resource, B.B.-A.; Methodology, B.B.-A.; Formal Analysis, S.A., A.P.; Data Curation, A.P., S.A.; Writing—Original Draft Preparation, S.A., B.B.-A.; Writing—Review & Editing, B.B.-A., M.G., A.P., and R.A.; Funding Acquisition, B.B.-A., M.G. and R.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Bill & Melinda Gates Foundation [OPP1134248] as well as the Integrated Genotyping Service and Support (IGSS) platform grant (ref. number PJ-002507) of BeCA-ILRI, Kenya.

**Acknowledgments:** The authors are grateful to Ana Luisa Garcia-Oliveira and Clay Sneller for their roles in funding acquisition and technical contributions, as well as the IITA Maize Improvement Program, (particularly A. Talabi and V. Oladipo) and the Bioscience Center staff (particularly N. Unachukwu, Q. Obi and Y. Ilesanmi) for technical assistance during the evaluation of field trials and DNA extraction, respectively.

**Conflicts of Interest:** The authors declare no conflicts of interest.

**Data Availability:** The DArTseq datasets used/analyzed in the manuscript have been deposited at the IITA-CKAN repository. doi:10.25502/aabs-rc02/d; doi:10.25502/8dkd-0h42/d.

## References

1. Badu-Apraku, B.; Fakorede, M.A.B. *Advances in Genetic Enhancement of Early and Extra-Early Maize for Sub-Saharan Africa*; Springer: Cham, Switzerland, 2017.
2. Cuello, C.; Baldy, A.; Brunaud, V.; Joets, J.; Delannoy, E.; Jacquemot, M.-P.; Botran, L.; Griveau, Y.; Guichard, C.; Soubigou-Taconnat, L.; et al. A systems biology approach uncovers a gene co-expression network associated with cell wall degradability in maize. *PLoS ONE* **2019**, *14*, e0227011. [[CrossRef](#)] [[PubMed](#)]
3. Schulz, S.; Hussaini, M.A.; Kling, J.G.; Berner, D.K.; Ikie, F.O. Evaluation of integrated *Striga hermonthica* control technologies under farmers' management. *Expl. Agric.* **2003**, *39*, 99–108. [[CrossRef](#)]
4. Oswald, A. *Striga* control—Technologies and their dissemination. *Crop. Prot.* **2005**, *24*, 333–342. [[CrossRef](#)]
5. De Groote, H.; Wangare, L.; Kanampiu, F.; Odendo, M.; Diallo, A.; Karaya, H.; Friesen, D. The potential of a herbicide resistant maize technology for *Striga* control in Africa. *Agric. Syst.* **2008**, *97*, 83–94. [[CrossRef](#)]
6. Babiker, A.G.T. *Striga*: The spreading scourge in Africa. *Regul. Plant Growth Dev.* **2007**, *42*, 74–87.
7. Samejima, H.; Babiker, A.G.; Mustafa, A.; Sugimoto, Y. Identification of *Striga hermonthica*-resistant upland rice varieties in Sudan and their resistance phenotypes. *Front. Plant Sci.* **2016**, *7*, 634. [[CrossRef](#)]
8. Badu-Apraku, B.; Talabi, A.O.; Fakorede, M.A.B.; Fasanmade, Y.; Gedil, M.; Magorokosho, C.; Asiedu, R. Yield gains and associated changes in an early yellow bi-parental maize population following genomic selection for *Striga* resistance and drought tolerance. *BMC Plant Biol.* **2019**, *19*, 129. [[CrossRef](#)]
9. Van Ast, A.; Bastiaans, A.L.; Katile, S. Cultural control measures to diminish sorghum yield loss and parasite success under *Striga hermonthica* infestation. *Crop. Prot.* **2005**, *24*, 1023–1034. [[CrossRef](#)]
10. Kamara, A.Y.; Menkir, A.; Chikoye, D.; Solomon, R.; Tofa, A.I.; Omoigui, L.O. Seed dressing maize with imazapyr to control *Striga hermonthica* in farmers' fields in the savannas of Nigeria. *Agriculture* **2020**, *10*, 83. [[CrossRef](#)]
11. Ejeta, G.; Butler, L.G.; Babiker, A.G. *New Approaches to the Control of Striga. Striga Research at Purdue University, Research Bulletin*; Agricultural Experiment Station, Purdue University: West Lafayette, IN, USA, 1992.

12. Haussmann, B.I.G.; Hess, D.E.; Reddy, B.V.S.; Mukuru, S.Z.; Kayentao, M.; Welz, H.G.; Geiger, H.H. Pattern analysis of genotype x environment interaction for *Striga* resistance and grain yield in African sorghum trials. *Euphytica* **2001**, *122*, 297–308. [[CrossRef](#)]
13. Rodenburg, J.; Bastiaans, L.; Kropff, M.J.; van Ast, A. Effects of host plant genotype and seed bank density on *Striga* reproduction. *Weed Res.* **2006**, *46*, 251–263. [[CrossRef](#)]
14. DeVries, J. The inheritance of *Striga* reactions in maize. In *Proceedings of a Workshop, Breeding for Striga Resistance in Cereals*; Haussmann, B.I.G., Hess, D.E., Koyama, M.L., Grivet, L., Rattunde, H.F.W., Geiger, H.H., Eds.; IITA: Ibadan, Nigeria; Margraf: Weikersheim, Germany, 2000; pp. 73–84.
15. Kim, S.K. Genetics of maize tolerance of *Striga hermonthica*. *Crop. Sci.* **1994**, *34*, 900907. [[CrossRef](#)]
16. Kim, S.K. (Ed.) Breeding maize for *Striga* tolerance and the development of a field infestation technique. In *Combating Striga in Africa, Proceedings of the Workshop by IITA, ICRISAT and IDRC, 22–24 August 1988*; IITA: Ibadan, Nigeria, 1991; pp. 96–108.
17. Pierce, S.; Mbwa, A.M.; Press, M.C.; Scholes, J.D. Xenognosin production and tolerance to *Striga asiatica* infection of high-yielding maize cultivars. *Weed Res.* **2003**, *43*, 139–145. [[CrossRef](#)]
18. Badu-Apraku, B.; Menkir, A.; Lum, A.F. Genetic variability for grain yield and components in an early tropical yellow maize population under *Striga hermonthica* infestation. *Crop. Improv.* **2007**, *20*, 107–122. [[CrossRef](#)]
19. Kim, S.K.; Adetimirin, V.O. Overview of tolerance and resistance maize hybrids to *Striga hermonthica* and *Striga asiatica*. In *Maize Research for Stress Environments, Proceedings of the Fourth Eastern and Southern Africa Regional Maize Conference, 28 March–1 April 1994*; Jewell, D.C., Waddington, S.R., Ransom, J.K., Pixey, K.V., Eds.; CIMMYT: Harare, Zimbabwe, 1995; pp. 255–262.
20. Menkir, A.; Kling, J.G. Response to recurrent selection for resistance to *Striga hermonthica* (Del.) Benth in a tropical maize population. *Crop. Sci.* **2007**, *47*, 674–684. [[CrossRef](#)]
21. Akanvou, L.; Doku, E.V.; Kling, J. Estimates of genetic variances and interrelationships of traits associated with *Striga* resistance in maize. *Afr. Crop. Sci. J.* **1997**, *5*, 1–8. [[CrossRef](#)]
22. Gbehounou, G.; Pieterse, A.H.; Verkleij, J.A.C. Longevity of *Striga* seeds reconsidered: Results of a field study on purple witchweed (*Striga hermonthica*) in Bénin. *Weed Sci.* **2003**, *51*, 940–946. [[CrossRef](#)]
23. Bouwmeester, H.J.; Matusova, R.; Zhongkui, S.; Beale, M.H. Secondary metabolite signalling in host-parasitic plant interactions. *Curr. Opin. Plant Biol.* **2003**, *6*, 358–364. [[CrossRef](#)]
24. Yallou, C.G.; Menkir, A.; Adetimirin, V.O.; Kling, J.G. Combining ability of maize inbred lines containing genes from *Zea diploperennis* for resistance to *Striga hermonthica* (Del.) Benth. *Plant Breed.* **2009**, *128*, 143–148. [[CrossRef](#)]
25. Graves, J.D.; Wylde, A.; Press, M.C.; Stewart, G.R. A carbon balance model of the sorghum—*Striga hermonthica* host–parasite association. *Plant Cell Environ.* **1989**, *12*, 101–107. [[CrossRef](#)]
26. Dörr, I. How *Striga* parasitizes its host: A TEM and SEM study. *Ann. Bot.* **1997**, *79*, 463–472. [[CrossRef](#)]
27. Ejeta, G.; Gressel, J. (Eds.) *Integrating New Technologies for Striga Control: Towards Ending the Witch-Hunt*; World Scientific Publishing Co. Pte Ltd, 5 Tol Tuck Link: Singapore, 2007; pp. 3–16.
28. Ackroyd, R.D.; Graves, J.D. The regulation of the water potential gradient in the host and parasite relationship between *Sorghum bicolor* and *Striga hermonthica*. *Ann. Bot.* **1997**, *80*, 649–656. [[CrossRef](#)]
29. Ransom, J.K.; Odhiambo, G.D.; Eplee, R.E.; Diallo, A.O. Estimates from field studies of the phytotoxic effects of *Striga* spp. on maize. In *Advances in Parasitic Plant Research—Proceedings of the Sixth Parasitic Weed Symposium*; Moreno, M.T., Cubero, J.L., Berner, D., Joel, D., Musselman, L.J., Parker, C., Eds.; Junta de Andalucía: Cor-doba, Spain, 1996; pp. 327–333.
30. Gurney, A.L.; Slate, J.; Press, M.C.; Scholes, J.D. A novel form of resistance in rice to the angiosperm parasite *Striga hermonthica*. *New Phytol.* **2006**, *169*, 199–208. [[CrossRef](#)] [[PubMed](#)]
31. Badu-Apraku, B.; Fakorede, M.A.B.; Menkir, A.; Kamara, A.Y.; Akanvou, L.; Chaby, Y. Response of early maturing maize to multiple-stresses in the Guinea savanna of West and Central Africa (*Zea mays* L.). *J. Gen. Breed.* **2004**, *58*, 119–130.
32. Swarbrick, P.J.; Scholes, J.D.; Press, M.C.; Slate, J. A major QTL for resistance of rice to the parasitic plant *Striga hermonthica* is not dependent on genetic background. *Pest. Manag. Sci.* **2009**, *65*, 528–532. [[CrossRef](#)]
33. Gasura, E.; Setimela, P.; Mabasa, S.; Rwafa, R.; Kageler, S.; Nyakurwa, C. Response of IITA maize inbred lines bred for *Striga hermonthica* resistance to *Striga asiatica* and associated resistance mechanisms in southern Africa. *Euphytica* **2019**, *215*, 151. [[CrossRef](#)]

34. Haussmann, B.I.G.; Hess, D.E.; Reddy, B.V.S.; Welz, H.G.; Geiger, H.H. Improved methodologies for breeding *Striga*-resistant sorghum. *Field Crops Res.* **2000**, *66*, 195–211. [[CrossRef](#)]
35. Amusan, I.O.; Rich, P.J.; Menkir, A.; Housley, T.; Ejeta, G. Resistance to *Striga hermonthica* in a maize inbred line derived from *Zea diploperennis*. *New Phytol.* **2008**, *178*, 157–166. [[CrossRef](#)]
36. Shayanowako, A.I.T.; Shimelis, H.; Laing, M.D.; Mwadzingeni, L. *Striga* resistance and compatibility of maize genotypes to a biocontrol agent, *Fusarium oxysporum* f.sp.*strigea*. *J. Crop. Improv.* **2020**, *34*, 437–454. [[CrossRef](#)]
37. Van Bueren, E.T.L.; Backes, G.; De Vriend, H.; Østergård, H. The role of molecular markers and marker assisted selection in breeding for organic agriculture. *Euphytica* **2010**, *175*, 51–64. [[CrossRef](#)]
38. Redinbaugh, M.; Pratt, R. Virus resistance. In *Handbook of Maize: It's Biology*; Bennetzen, J.L., Hake, S.C., Eds.; Springer: New York, NY, USA, 2009; pp. 251–268.
39. Awata, L.A.O.; Beyene, Y.; Gowda, M.; Suresh, L.M.; Jumbo, M.B.; Tongoona, P.; Danquah, E.; Ifie, B.E.; Marchelo-Dragna, P.W.; Olsen, M.; et al. Genetic analysis of QTL for resistance to Maize Lethal Necrosis in multiple mapping populations. *Genes* **2020**, *11*, 32. [[CrossRef](#)] [[PubMed](#)]
40. Srivastava, R.K.; Singh, R.B.; Pujarula, V.L.; Bollam, S.; Pusuluri, M.; Chellapilla, T.S.; Yadav, R.S.; Gupta, R. Genome-wide association studies and genomic selection in Pearl Millet: Advances and prospects. *Front. Genet.* **2020**, *10*, 1389. [[CrossRef](#)] [[PubMed](#)]
41. Wang, H.; Zaman, Q.U.; Huang, W.; Mei, D.; Liu, J.; Wang, W.; Ding, B.; Hao, M.; Fu, L.; Cheng, H.; et al. QTL and candidate gene identification for silique length based on high-dense genetic map in *Brassica napus* L. *Front. Plant Sci.* **2019**, *10*, 1579. [[CrossRef](#)] [[PubMed](#)]
42. Kotla, A.; Phuke, R.; Hariprasanna, K.; Mehtrec, S.P.; Rathore, A.; Gorthy, S.; Srivastava, R.K.; Das, R.; Prakash, A.B.; Radhika, K.; et al. Identification of QTLs and candidate genes for high grain Fe and Zn concentration in sorghum (*Sorghum bicolor* (L.) Moench). *J. Cereal Sci.* **2019**, *90*, 102850. [[CrossRef](#)]
43. Haussmann, B.I.G.; Hess, D.E.; Omany, G.; Folkertsma, R.T.; Reddy, B.V.S.; Kayentao, M.; Welz, G.; Geiger, H.H. Genomic region influencing resistance to parasitic weed *Striga hermonthica* in two recombinant inbred populations of sorghum. *Theor. Appl. Genet.* **2004**, *109*, 1005–1016. [[CrossRef](#)] [[PubMed](#)]
44. Adewale, S.A.; Badu-Apraku, B.; Akinwale, R.O.; Paterne, A.A.; Gedil, M.; Garcia-Oliveira, A.L. Genome-wide association study of *Striga* resistance in early maturing white tropical maize inbred lines. *BMC Plant Biol.* **2020**, *20*, 1–16. [[CrossRef](#)]
45. Ali, R.; Hash, C.T.; Damris, O.; Elhoussein, A.; Mohamed, A.H. Introgression of *Striga* resistance into popular Sudanese sorghum varieties using marker assisted selection. *World J. Biot.* **2016**, *1*, 48–55. [[CrossRef](#)]
46. Badu-Apraku, B.; Oyekunle, M. Genetic analysis of grain yield and other traits of extra-early yellow maize inbreds and hybrid performance under contrasting environments. *Field Crops Res.* **2011**, *129*, 99–110. [[CrossRef](#)]
47. Akinwale, R.O.; Badu-Apraku, B.; Fakorede, M.A.B.; Vroh-Bi, I. Heterotic grouping of tropical early-maturing maize inbred lines based on combining ability in *Striga*-infested and *Striga*-free environments and the use of SSR markers for genotyping. *Field Crops Res.* **2014**, *156*, 48–62. [[CrossRef](#)]
48. Sansaloni, C.P.; Petroli, C.D.; Jaccoud, D.; Carling, J.; Detering, F.; Grattapaglia, D.; Kilian, A. Diversity Arrays Technology (DArT) and next-generation sequencing combined: Genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. *BMC Proc.* **2011**, *5*, 54. [[CrossRef](#)]
49. Chen, J.; Zavala, C.; Ortega, N.G.; Petroli, C.D.; Franco, J.; Burgüeno, J.A.; Costich, D.E.; Hearne, S.J. The development of quality control genotyping approaches: A case study using elite maize lines. *PLoS ONE* **2016**, *11*, e0157236. [[CrossRef](#)] [[PubMed](#)]
50. Wang, R.; Liu, Y.; Isham, K.; Zhao, W.; Wheeler, J.; Klassen, N.; Hu, Y.; Bonman, J.M.; Chen, J. QTL identification and KASP marker development for productive tiller and fertile spikelet numbers in two high-yielding hard white spring wheat cultivars. *Mol. Breed.* **2018**, *38*, 135. [[CrossRef](#)] [[PubMed](#)]
51. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016.
52. SAS Institute Inc. *Statistical Analysis Software (SAS) User's Guide*; SAS Inst: Cary, NC, USA, 2013.
53. Holland, J.B.; Nyquist, W.E.; Cervantes-Martinez, C.T. Estimating and interpreting heritability for plant breeding: An update. In *Plant Breeding Reviews*; Janick, J., Ed.; John Wiley & Sons: New York, NY, USA, 2003; Volume 22, pp. 9–111.

54. Schloerke, B.; Crowley, J.; Cook, D.; Briatte, F.; Marbach, M.; Thoen, E.; Elberg, A.; Larmanange, J. GGally: Extension to 'ggplot2'. R Package Version 1.5.0. 2020. Available online: <https://CRAN.R-project.org/package=GGally> (accessed on 10 February 2020).
55. Alvarado, G.; López, M.; Vargas, M.; Pacheco, A.; Rodríguez, F.; Burgueño, J.; Crossa, J. META-R (Multi Environment Trial Analysis with R for Windows). Version 6.0—CIMMYT Research Software Dataverse-CIMMYT Dataverse Network. Available online: <http://hdl.handle.net/11529/10201> (accessed on 2 December 2019).
56. Broman, K.W.; Wu, H.; Sen, S.; Churchill, G.A. R/qtl: QTL mapping in experimental crosses. *J. Bioinform.* **2003**, *19*, 889–890. [[CrossRef](#)]
57. Zuo, J.; Niu, Y.; Cheng, P.; Feng, J.; Han, S.; Zhang, Y.-H.; Shu, G.; Wang, Y.; Zhang, Y.-M. Effect of marker segregation distortion on high density linkage map construction and QTL mapping in soybean (*Glycine max* L.). *Heredity* **2019**, *123*, 579–592. [[CrossRef](#)] [[PubMed](#)]
58. Wessinger, C.A.; Hileman, L.C.; Rauscher, M.D. Identification of major quantitative trait loci underlying floral pollination syndrome divergence in *Penstemon*. *Phil. Trans. R. Soc. B Biol. Sci.* **2014**, *369*, 20130349. [[CrossRef](#)] [[PubMed](#)]
59. Du, B.; Wang, Q.; Sun, G.; Ren, X.; Cheng, Y.; Wang, Y.; Gao, S.; Li, C.; Sun, D. Mapping dynamic QTL dissects the genetic architecture of grain size and grain filling rate at different grain-filling stages in barley. *Sci. Rep.* **2019**, *9*, 18823. [[CrossRef](#)]
60. Lübberstedt, T.; Melchinger, A.E.; Schön, C.C.; Utz, H.F.; Klein, D. QTL mapping in testcrosses of European flint lines of maize: I. Comparison of different testers for forage yield traits. *Crop. Sci.* **1997**, *37*, 921–931. [[CrossRef](#)]
61. Bo, K.L.; Ma, Z.; Chen, J.F.; Weng, Y. Molecular mapping reveals structural rearrangements and quantitative trait loci underlying traits with local adaptation in semi-wild Xishuangbana cucumber (*Cucumis sativus* L. var. *xishuangbannanensis* Qi et Yuan). *Theor. Appl. Genet.* **2015**, *128*, 25–39. [[CrossRef](#)]
62. Wang, X.; Liu, H.; Pang, M.; Fu, B.; Yu, X.; He, S.; Tong, J. Construction of a high-density genetic linkage map and mapping of quantitative trait loci for growth-related traits in silver carp (*Hypophthalmichthys molitrix*). *Sci Rep.* **2019**, *9*, 17506. [[CrossRef](#)]
63. Gedil, M.; Menkir, A. An integrated molecular and conventional breeding scheme for enhancing genetic gain in maize in Africa. *Front. Plant Sci.* **2019**, *10*, 1430. [[CrossRef](#)]
64. Septiani, P.; Lanubile, A.; Stagnati, L.; Busconi, M.; Nelissen, H.; Pè, M.E.; Dell'Acqua, M.; Marocco, A. Unravelling the genetic basis of *Fusarium* seedling rot resistance in the MAGIC maize population: Novel targets for breeding. *Sci Rep.* **2019**, *9*, 5665. [[CrossRef](#)] [[PubMed](#)]
65. Ribeiro, P.F.; Badu-Apraku, B.; Gracen, V.E.; Danquah, E.Y.; Garcia-Oliveira, A.L.; Asante, M.D.; Afriyie-Debrah, C.; Gedil, M. Identification of Quantitative Trait Loci for grain yield and other traits in tropical maize under high and low soil-nitrogen environments. *Crop. Sci.* **2018**, *58*, 321–331. [[CrossRef](#)]
66. Mbogo, P.O.; Dida, M.M.; Owuor, B. Generation means analysis for estimation of genetic parameters for *Striga hermonthica* resistance in maize (*Zea mays* L.). *J. Agric. Sci.* **2015**, *7*, 143. [[CrossRef](#)]
67. Mrema, E.; Shimelis, H.; Laing, M.; Mwandzingeni, L. Genetic analysis of the maximum germination distance of *Striga* under *Fusarium oxysporum* f. sp. *strigae* biocontrol in sorghum. *J. Integr. Agric.* **2018**, *17*, 1585–1593. [[CrossRef](#)]
68. Makumbi, D.; Diallo, A.; Kanampiu, F.; Mugo, S.; Karaya, H. Agronomic performance and genotype x environment interaction of herbicide-resistant maize varieties in Eastern Africa. *Crop. Sci.* **2015**, *55*, 540–555. [[CrossRef](#)]
69. Shayanowako, A.I.T.; Shimelis, H.; Laing, M.D.; Mwandzingeni, L. Variance components and heritability of traits related to *Striga asiatica* resistance and compatibility to *Fusarium oxysporum* F.Sp. *Strigae* in maize. *Maydica* **2018**, *63*, 8.
70. Qiu, C.; Han, Z.; Li, W.; Ye, K.; Xie, Y.; Wang, Z. A high-density genetic linkage map and QTL mapping for growth and sex of yellow drum (*Nibeia albiflora*). *Sci. Rep.* **2018**, *8*, 1–12. [[CrossRef](#)]
71. Womack, E.D.; Warburton, M.L.; Williams, W.P. Mapping of quantitative trait loci for resistance to fall armyworm and Southwestern Corn Borer leaf-feeding damage in maize. *Crop. Sci.* **2018**, *58*, 529–539. [[CrossRef](#)]
72. Samayoa, L.F.; Malvar, R.A.; McMullen, M.D.; Butrón, A. Identification of QTL for resistance to Mediterranean corn borer in a maize tropical line to improve temperate germplasm. *BMC Plant Biol.* **2015**, *15*, 265. [[CrossRef](#)]

73. Langridge, P.; Lagudah, E.; Holton, T.; Appels, R.; Sharp, P.; Chalmers, K. Trends in genetic and genome analyses in wheat: A review. *Aust. J. Agric. Res.* **2001**, *52*, 1043–1077. [[CrossRef](#)]
74. Ma, L.; Qing, C.; Frei, U.; Shen, Y.; Lubberstedt, T. Association mapping for root system architecture traits under two nitrogen conditions in germplasm enhancement of maize doubled haploid lines. *Crop. J.* **2019**, *8*, 213–226. [[CrossRef](#)]
75. Ramaekers, L.; Remans, R.; Rao, I.M.; Blair, M.W.; Vanderleyden, J. Strategies for improving phosphorus acquisition efficiency of crop plants. *Field Crops Res.* **2010**, *117*, 169176. [[CrossRef](#)]
76. Sauter, M. Phytosulfokine peptide signalling. *J. Exp. Bot.* **2015**, *66*, 5161–5169. [[CrossRef](#)] [[PubMed](#)]
77. Sharma, M.; Pandey, G.K. Expansion and function of repeat domain proteins during stress and development in plants. *Front. Plant Sci.* **2016**, *6*, 1218. [[CrossRef](#)]
78. Liu, X.; Liu, S.; Wang, R.; Chen, X.; Fan, Z.; Wu, B.; Zhou, T. Analyses of MiRNA functions in maize using a newly developed ZMBJ-CMV-2bN81-STTM vector. *Front. Plant Sci.* **2019**, *10*, 1277. [[CrossRef](#)]
79. Park, C.-J.; Seo, Y.-S. Heat shock proteins: A review of the molecular chaperones for plant immunity. *Plant Pathol. J.* **2015**, *31*, 323–333. [[CrossRef](#)]
80. Ng, D.W.-K.; Abeysinghe, J.K.; Kamali, M. Regulating the regulators: The control of transcription factors in plant defense signaling. *Int. J. Mol. Sci.* **2018**, *19*, 3737. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Article

# Molecular Assisted Selection for Pollination-Constant and Non-Astringent Type without Male Flowers in Spanish Germplasm for Persimmon Breeding

Manuel Blasco <sup>1</sup>, Francisco Gil-Muñoz <sup>2</sup>, María del Mar Naval <sup>1</sup> and María Luisa Badenes <sup>2,\*</sup>

<sup>1</sup> CANSO, Mestre Serrano, 1, 46250 L'Alcúdia, Valencia, Spain; blasco\_manvil@externos.gva.es (M.B.); naval\_marmer@gva.es (M.d.M.N.)

<sup>2</sup> Instituto Valenciano de Investigaciones Agrarias, CV 315 km 10,5., 46113 Moncada, Valencia, Spain; gil\_framuna@externos.gva.es

\* Correspondence: badenes\_mlu@gva.es

Received: 14 July 2020; Accepted: 7 August 2020; Published: 11 August 2020

**Abstract:** Persimmon (*Diospyros kaki* Thunb) species is a hexaploid genotype that has a morphologically polygamous gynodioecious sexual system. *D. kaki* bears unisexual flowers. The presence of male flowers resulted in the presence of seeds in the varieties. The fruits of persimmon are classified according to their astringency and the pollination events that produced seeds and modify the levels of astringency in the fruit. The presence of seeds in astringent varieties as pollination variant astringent (PVA), pollination variant non-astringent (PVNA) and pollination constant astringent (PCA) resulted in fruits not marketable. Molecular markers that allow selection of the varieties according to the type of flowers at the plantlet stage would allow selection of seedless varieties. In this study, a marker developed in *D. lotus* by bulk segregant analysis (BSA) and amplified fragment length polymorphism (AFLP) markers, named DLSx-AF4, has been validated in a germplasm collection of persimmon, results obtained agree with the phenotype data. A second important trait in persimmon is the presence of astringency in ripened fruits. Fruits non-astringent at the ripen stage named pollination constant non-astringent (PCNA) are the objective of many breeding programs as they do not need removal of the astringency by a postharvest treatment. Astringency in the hexaploid persimmon is a dominant trait. The presence of at least one astringent allele confers astringency to the fruit. In this paper we checked the marker developed linked to the AST gene. Our goal has been to validate both markers in germplasm from different origins and to test the usefulness in a breeding program.

**Keywords:** persimmon; sex determination; fruit astringency; molecular markers

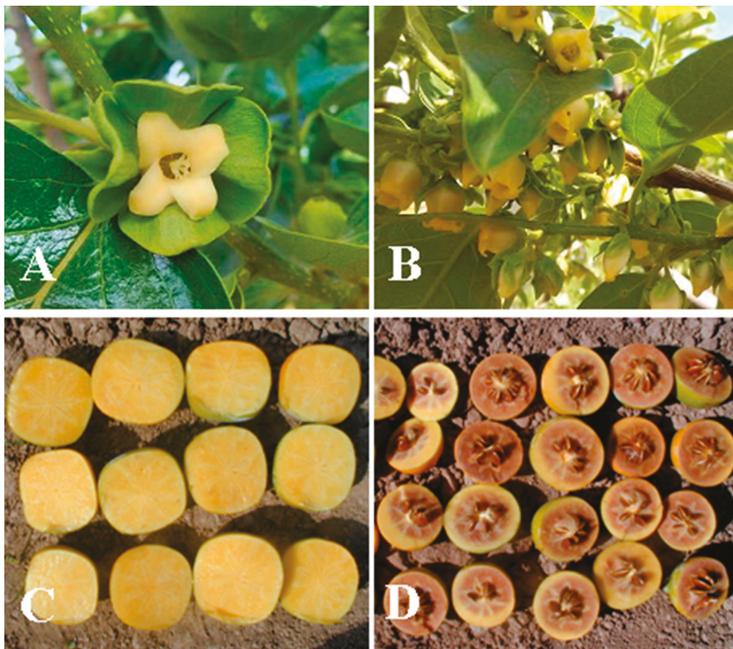
## 1. Introduction

Persimmon (*Diospyros kaki* Thunb) species is a hexaploid genotype that has a morphologically polygamous gynodioecious sexual system [1]. *D. kaki* bears unisexual flowers, as do other *Diospyros* species. There are genotypes that bear only female flowers and genotypes bearing male and female flowers [2]. Furthermore, varieties bearing only male flowers were described in China [3] and occasional male flower formation was reported in varieties that usually bear only female flowers [4]. In addition to the unisexual flowers, some varieties or genotypes bear hermaphrodite flowers; however, these flowers do not function fully as female flowers [5]. Most of the commercial varieties present only female flowers [6], however the presence of the male flowers type is important in two scenarios: first when the production of seeded fruits is convenient and second in breeding activities in which crosses are requested.

The fruits of persimmon are classified according to their astringency and the pollination events that resulted in different types of fruits. The PCNA (pollination constant non-astringent) varieties

are always not astringent at maturity regardless of pollination events and the presence or absence of seeds in the fruit. The presence of male flowers in these varieties is most convenient since it allows pollination and produces seeded fruits that increase the fruit size and weight. In Japan, the presence of seeds in the fruit does not affect the consumers demand [7]. However, in Europe and western countries consumers prefer seedless fruits.

Three additional variety types can be distinguished: PVNA-type (pollination variant non-astringent), which are non-astringent varieties when seeds are present; PVA-type (pollination variant astringent), which are astringent varieties in most parts of the fruit and non-astringent around the seeds if they are present, and PCA-type (pollination constant astringent), which are astringent varieties regardless the presence of seeds. The loss of astringency in these types of persimmons is associated to the ability of the seeds to produce acetaldehyde. This production resulted in browning of the flesh around the seeds (Figure 1), which interfere with the postharvest treatment for removing the astringency in the fruit, all together the PVA and PCA type fruits are unmarketable if they are pollinated and the fruits present seeds [8,9]. In PVA and PCA varieties it is crucial to avoid pollination, hence the presence of male flowers in the variety and in the vicinity of the crop should be avoided. The ability of the production of male flowers is a genetic trait that should be determined in the varieties for avoiding seeds in astringent varieties or improving the presence of them in non-astringent varieties.



**Figure 1.** Phenotypes of the traits selected. (A,B) flowers from the variety 'Cal Fuyu', female and male respectively and (C,D) results of pollination on astringent fruits from 'Rojo Brillante' a pollination variant astringent (PVA) variety: (C) parthenocarpic non-pollinated fruits and (D) pollinated fruits in which presence of seeds resulted in no marketable fruits.

Elucidation of the genetic and molecular basis of sex expression in *D. kaki* leading to the development of molecular markers would allow selection of the varieties according to the type of flowers, being a great contribution for persimmon production and *Diospyros* breeding. The hexaploidy of *D. kaki* made elucidation of this question more difficult than in diploid genotypes. Since the genus includes more than 700 species with different levels of polyploidy, the diploid *Diospyros lotus* was

used for investigation of the sex expression into the genus [10]. These authors described the model of inheritance and developed molecular markers associated to sex expression. Later small RNA acting as a sex determinant was identified [11]. Development of markers used the bulk segregant analysis (BSA) and amplified fragment length polymorphism (AFLP). An AFLP marker identified as DISx-AF4 was sequence-characterized and converted into the sequence characterized amplified region (SCAR) [10]. In this study the marker has been tested in a germplasm selection of varieties phenotyped for sex expression and a backcross population obtained at Instituto Valenciano de Investigaciones Agrarias (IVIA). The results provide evidence of the usefulness of molecular marker assistance in identifying the genetic potential of production of male flowers in persimmon, an important trait in breeding.

PCNA varieties are highly desired because their mature fruits are not astringent, as they stop accumulating tannins at early steps of fruit development [12]. In Japanese varieties, the PCNA trait is recessive to the non-PCNA trait [13] and is controlled by a single locus, AST [14]. Due to persimmon being a hexaploid, the PCNA type should contain six recessive ast alleles [15]. In breeding programs aimed at obtaining PCNA cultivars, the hexaploidy of persimmon along with the recessive inheritance of the non-astringency trait led to breeders to develop crosses that involved only PCNA genotypes. Consequently, several generations of crosses between PCNA genotypes along with the low genetic diversity of this group of persimmons resulted in families with a high rate of inbreeding and plenty of the problems derived from this fact. To avoid inbreeding, the programs need to use non-astringent cultivars in the crosses, but the rate of PCNA obtained could be very low depending on the number of dominant AST alleles carried by the parents selected. In a backcross BC<sub>1</sub>, the expected proportion of PCNA offspring from a non-PCNA F<sub>1</sub> parent with one dominant AST, two or three is 50, 20 or 5%, respectively, under an autohexaploid model. In this context, it is of high interest to be able of selecting PCNA types and non-PCNA types in the families obtained at the plantlet stage. The alternative is to select the type of the fruits in the fields after a juvenile period of four years minimum, which is extremely costly and has a low efficiency.

Many efforts have been made to target the region linked to AST [14,16–18]. The most promising results were obtained in a study that identified a region tightly linked to the AST gene [19]. These authors developed a multiplex PCR method based on primers developed from the region identified, highly reliable that allowed detecting recessive and dominant alleles. These primers have been used to test a group of varieties [15]. The region contains microsatellites that allow distinguishing 12 different alleles from 14 non PCNA genotypes. More than 200 accessions and several crosses between PCNA and non-PCNA genotypes were analyzed [20]. Based on the number of fragments detected per individual these authors were able to determine the dominant (AST) and recessive (ast) alleles in the hexaploid persimmon germplasm.

Using this methodology, in this paper we applied molecular assisted selection for discriminate PCNA cultivars and seedlings from different segregated populations obtained in the frame of the IVIA breeding program. The markers for both traits were developed from Japanese varieties, our goal is to validate the markers in a set of germplasm from different origins and different type of astringency and applied them to the IVIA breeding program, in which the involvement of varieties from Mediterranean origin is relevant.

## **2. Materials and Methods**

### *2.1. Plant Materials*

#### *2.1.1. Validation of AST and DISx-AF4S Markers*

Molecular markers developed for sex expression and type of astringency in persimmon were studied in a set of 42 accessions (Table 1) from the persimmon germplasm collection maintained at IVIA, Moncada, Spain (39.588741, -0.394848). The accessions were phenotyped regarding the presence of male flowers. The phenotype of astringency type was known from previous germplasm characterization [21,22].

**Table 1.** Plant material studied, origin, astringency type, genotype of AST marker, flower type and genotype of the DISx-AF4S marker.

Variety	Origin	Type <sup>1</sup>	AST <sup>2</sup>	Flowers <sup>1</sup>	DISx-AF4S <sup>2</sup>	Variety	Origin	Type <sup>1</sup>	AST <sup>2</sup>	Flower <sup>1</sup>	DISx-AF4S <sup>2</sup>
Anheca	Spain	PCA	+	♀	-	Kwabata	Japan	PCNA	-	♀	-
Ferrán-12	Spain	PCA	+	♀	-	Koda Goshō	Japan	PCNA	-	♀	-
Reus-6	Spain	PCA	+	♀	-	Maekawa jiro	Japan	PCNA	-	♀	-
Tomatero	Spain	PCA	+	♀	-	Mukaku jiro	Japan	PCNA	-	♀	-
Costata	Italy	PCA	+	♀	-	O'Goshō	Japan	PCNA	-	♀	-
Lycopersicon	Italy	PCA	+	♀	-	Suruga	Japan	PCNA	-	♀	-
Aizumishirazu-A	Japan	PCA	+	♀	-	Yamato Goshō	Japan	PCNA	-	♀	-
Fuji	Japan	PCA	+	♀	-	Bétera-2	Spain	PVA	+	♀	-
Korea Kaki	Japan	PCA	+	♀	-	Reus-15	Spain	PVA	+	♀	-
Takura	Japan	PCA	+	♀	-	Rojo Brillante	Spain	PVA	+	♀	-
Yokono	Japan	PCA	+	♀	-	Xato de Bonrepós	Spain	PVA	+	♀	-
Cal Fuyu	Japan	PCNA	-	♀/♂	+	Aizumishirazu-B	Japan	PVA	+	♀	-
Fau Fau	Japan	PCNA	-	♀/♂	+	Atago	Japan	PVA	+	♀	-
Fukuro Goshō	Japan	PCNA	-	♀/♂	+	Hiratanekaki	Japan	PVA	+	♀	-
Fuyu	Japan	PCNA	-	♀	-	Hiratanenashi	Japan	PVA	+	♀	-
Giant Fuyu	Japan	PCNA	-	♀	-	Maru	Japan	PVA	+	♀/♂	+
Hana fuyu	Japan	PCNA	-	♀	-	Tone Wase	Japan	PVA	+	♀	-
Hana Goshō	Japan	PCNA	-	♀/♂	+	Pakistan Seedless	Pakistan	PVNA	+	♀	-
Ichikikei Jiro	Japan	PCNA	-	♀	-	Agakaki	Japan	PVNA	+	♀/♂	+
Isahaya	Japan	PCNA	-	♀	-	Castellani	Italy	PVNA	+	♀	-
Jiro	Japan	PCNA	-	♀	-	Edoichi	Italy	PVNA	+	♀	-

<sup>1</sup> Data from phenotyping. <sup>2</sup> Data from genotyping.

### 2.1.2. Marker Assisted Selection

Marker assisted selection was made on 12 segregated populations obtained from ('Rojo Brillante' × 'Cal Fuyu') × 'Cal Fuyu' in 2016. The backcross was made using 'Rojo Brillante' a high-quality variety astringent (PVA) and 'Cal Fuyu' a PCNA variety with male flowers. Both parents were selected based on agronomic characteristics and adaptability to the Mediterranean environment [21].

Segregated populations screened and individuals per population are described in Table 2. All progenies and seedlings obtained were maintained in orchards at CANSO's Experimental Station, L'Alcudia, Valencia, Spain (39.189086, -0.542067).

**Table 2.** Results of genotypes analyzed by molecular markers. Number of offspring with the astringent allele (AST), with the D1Sx-AF4S allele, with AST + D1Sx-AF4S (not selected) and number of offspring with the absence of both markers (genotypes selected).

Progeny	Total	Number of Offspring			
		AST +	D1Sx-AF4S +	AST + D1Sx-AF4S +	AST – D1Sx-AF4S –
F-1.34	99	61 (61.6)	39 (39.4)	22 (22.2)	21 (21.2)
F-1.50	65	39 (60.0)	43 (66.2)	28 (43.1)	11 (16.9)
F-1.52	47	13 (27.7)	24 (51.1)	4 (8.5)	14 (29.8)
F-2.27	5	4 (80.0)	3 (60.0)	3 (60.0)	1 (20.0)
F-4.19	11	8 (72.7)	9 (81.8)	7 (63.6)	1 (9.1)
F-4.24	61	20 (32.8)	31 (50.8)	10 (16.4)	20 (32.8)
F-4.35	38	14 (36.8)	9 (23.7)	4 (10.5)	19 (50.0)
F-4.49	52	14 (26.9)	17 (32.7)	3 (5.8)	24 (46.2)
F-5.32	12	10 (83.3)	6 (50.0)	4 (33.3)	0 (0.0)
F-5.34	27	11 (40.7)	16 (59.3)	5 (18.5)	5 (18.5)
F-5.36	17	12 (70.6)	9 (52.9)	4 (23.5)	0 (0.0)
F-5.41	7	5 (71.4)	3 (42.9)	3 (42.9)	2 (28.6)
total	441	211 (47.8)	209 (47.4)	97 (22.0)	118 (26.8)

Numbers in parentheses are the rate (%) of corresponding offspring in each progeny.

## 2.2. Methods

### 2.2.1. DNA Isolation

Young fully expanded leaves were collected from trees and kept at  $-20\text{ }^{\circ}\text{C}$  until DNA isolation. DNA was isolated according to the CTAB method described in [23] with minor modifications [24].

### 2.2.2. Molecular Markers Analysis

The capacity of producing male flowers was checked with the sequence characterized amplified region (SCAR) marker 'D1Sx-AF4S' [10], primers used were: forward (D1Sx-AF4-3F; 5'-ACA TCC AAA GTT CTG GAG AAT CA-3') and reverse (D1Sx-AF4-3R; 5'-ATT GGT GCT TGG TCA AAC ATA TC-3').

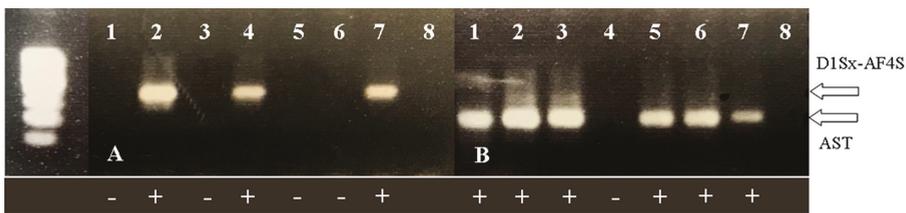
Determination of PCNA genotypes used the primers described in [19] PCNA-F (CCCCTCAGTGGCAGTGCTGC) and 5R3R (GAAACACTCATCCGGAGACTTC).

Polymerase chain reactions (PCRs) were performed in a final volume of 20  $\mu\text{L}$  containing 1 $\times$  of DreamTaq Buffer (Thermo Fisher Scientific, Vilnius, Lithuania), 0.1 mM of each dNTPs (Promega, Madison, WI, USA), 20 ng of genomic DNA and 1 U of DreamTaq polymerase (Thermo Fisher Scientific, Vilnius, Lithuania). The PCR program consisted of pre-denaturation at  $94\text{ }^{\circ}\text{C}$  for 2 min; 35 cycles at  $98\text{ }^{\circ}\text{C}$  for 15 s,  $60\text{ }^{\circ}\text{C}$  for 20 s and  $72\text{ }^{\circ}\text{C}$  for 1 min; followed by a final extension at  $72\text{ }^{\circ}\text{C}$  for 10 min. PCR products were separated by electrophoresis on 1.5% agarose gels in 0.5 $\times$  TAE buffer and visualized with GelRED<sup>®</sup> (Sigma-Aldrich, St. Louis, MI, USA).

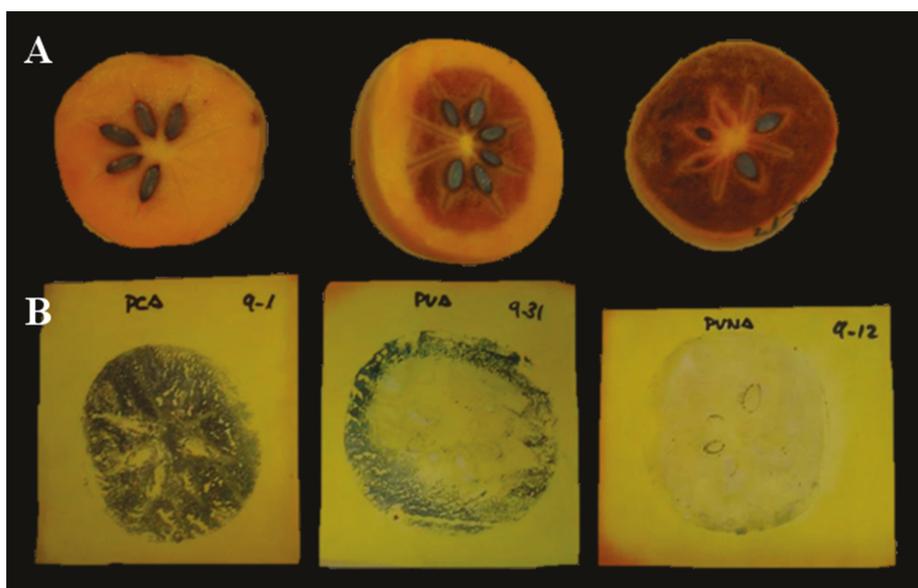
### 3. Results and Discussion

#### 3.1. Marker Assisted Selection Validation: Production of Male Flowers

A set of accessions belonging to the persimmon germplasm bank were phenotyped for the presence of male flowers and later genotyped with the marker DISx-AF4S [10] to test the accuracy of the marker for Molecular Assisted Selection (MAS). Results of the genotype agreed with the results of the phenotype (Table 1), non-discrepancies were observed. The capacity of developing male flowers was clearly stated by the presence of the amplified band (320 bp). Figure 2a shows the results on an agarose gel of the presence of male flowers in the genotypes ‘Agakaki’, ‘Cal Fuyu’ and the selection ‘F-1.34’ from the IVIA breeding program. Total correlation between the phenotype and the presence/absence of the band was obtained for all the genotypes studied (Table 1). This marker is a great advantage in breeding programs in which astringent and non-astringent genotypes are involved. The presence of male flowers in astringent varieties (PVNA, PVA and PCA) resulted in the presence of seeds in the fruit. The loss of astringency in these types of persimmons is associated to the ability of the seeds to produce acetaldehyde (Figure 3). Production of acetaldehyde is a quantitative trait in which less production by the seed resulted in higher astringency on the pulp (PCA), and high production resulted in low/none astringent flesh (PVNA) being PVA intermediate. This acetaldehyde production resulted in browning of the flesh around the seeds, which interferes with the postharvest treatment for removing the astringency in the fruit [25]. All together the PVA and PCA type fruits pollinated are unmarketable. In the case of PVNA types in which the presence of seeds browned completely the flesh (Figure 3), there are specific markets in which these varieties are accepted. However, in most of the markets, the PVNA fruits are accepted with no seeds and after removing the astringency by postharvest treatment. In all breeding programs that use astringent varieties, MAS for discriminating male flowers is very important for avoiding self-pollination and/or mix of cultivars that can cross pollinated among them and produced seeds. In breeding programs that involve non-astringent varieties the discrimination of the presence of male flowers is necessary too. Some Japanese programs look for varieties with male flowers and seeds that increase the size and setting of fruits, but in western countries, where the presence of seeded fruits is not acceptable, the presence of male flowers is discarded similarly to astringent varieties. Selection of this trait in persimmon species that have a four-year juvenile period resulted in great interest to avoid plants in the fields that will be eliminated in the future and, additionally, to avoid undesired pollination in the breeding plots.



**Figure 2.** PCR results by electrophoresis in an agarose gel (1.5%) vs. phenotype data; (A) DISx-AF4S PCR results. The (+) presence and (-) absence of male flowers from phenotype data; the DISx-AF4S marker is present in varieties ‘Agakaki’, ‘Cal Fuyu’ and F.1-34 in agreement with the phenotype; (B) AST PCR results; (+) astringent fruits according to phenotype data and (-) non-astringent fruits (pollination constant non-astringent (PCNA)) according to phenotype data. The AST marker was present in all astringent varieties and absent in ‘Cal Fuyu’, a PCNA variety.



**Figure 3.** Three types of astringent persimmon according to the amount of acetaldehyde produced by the seed; (A) fruits of pollination constant astringent (PCA), PVA and pollination variant non-astringent (PVNA; from right to left) and (B) distribution of condensed tannins, visualized by precipitation of blue ferric chloride impregnated in a paper [25,26]. Fruits of PCA, PVA and PVNA (from right to left).

### 3.2. Marker Assisted Selection Validation: Selection of PCNA

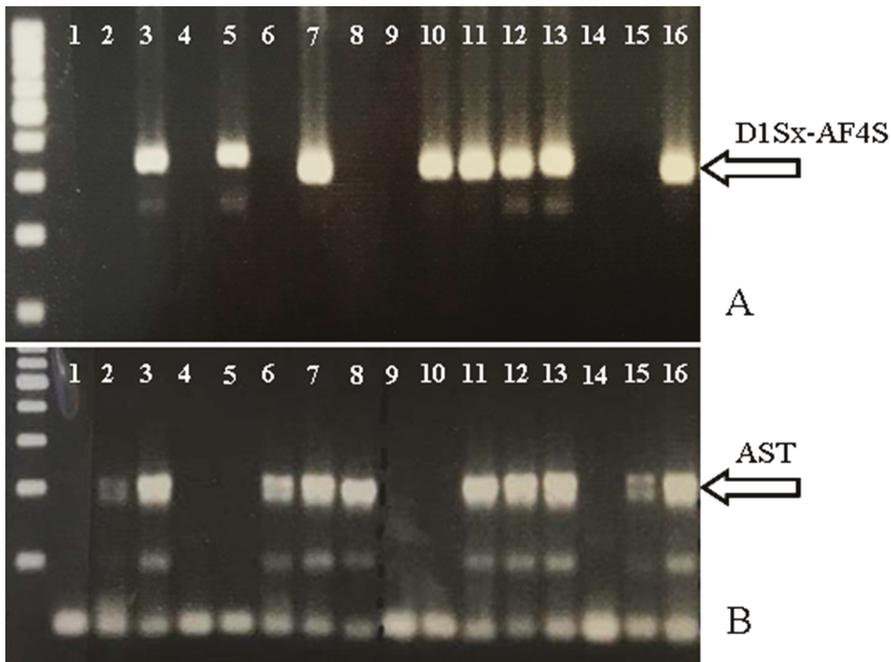
In Japanese cultivars, the PCNA trait is recessive to the non-PCNA trait [13] and is controlled by a single locus, AST [14]. Due to persimmon being a hexaploid, the PCNA type should contain six recessive ast alleles [15]. Detection of at least one AST allele determines the astringency of fruit. Selection of non-astringent fruits or PCNA are the objective in most of the persimmon breeding programs currently active in the world [27–32]. In this study we validated the AST marker developed in [19] for discrimination between PCNA genotypes and the different astringent types.

Validation of AST marker was carried out in a set of cultivars from the germplasm collection with known astringency (Table 1). A total correlation between the phenotypic data of astringency and the markers obtained in the genotypes analyzed was obtained. Figure 2b shows the PCR products of a set of accessions. Two PVA cultivars ‘Rojo Brillante’ and ‘Tone Wase’, three PVNA cultivars ‘Agakaki’, ‘Castellani’, ‘Edoichi’ and ‘F-1.34’ showed a clear band for AST marker. The PCNA cultivar ‘Cal Fuyu’ showed no amplified product.

### 3.3. Marker Assisted Selection of Both Traits in the IVIA Breeding Program

After validation of the D1Sx-AF4S and AST markers in a set of accessions phenotyped, we applied both markers in the breeding program for selecting the individuals of several segregated populations (Figure 4).

A total of 441 individuals belonging to 12 segregated populations obtained by a backcross that consisted in (PVA × PCNA × PCNA) were evaluated (Table 2). The cross (PVA × PCNA) was made using ‘Rojo Brillante’ a high-quality variety astringent and ‘Cal Fuyu’ a PCNA variety with male flowers.



**Figure 4.** PCR results from agarose gel electrophoresis of 16 backcross (BC) individuals genotyped for AST and D1Sx-AF4S markers. (A) Presence of the marker (D1Sx-AF4S) indicates ability for developing male flowers; (B) Absence of the marker means absence of any astringent allele (AST), which corresponds to a PCNA genotype. The BC progenies are selected based on the absence of both markers, which corresponds to PCNA types without capability of the development of male flowers. Individuals 1, 4, 9 and 14 have been selected.

The astringency trait is a dominant marker and taking into account the hexaploidy of persimmon, the number of PCNA genotypes obtained in crosses that involved astringent types depends on the number of AST alleles present in the astringent parental. The validation of the AST marker was made based on different segregated families by [33]. Identification of different AST alleles was made in a set of cultivars by means of crosses with PCNA varieties, analysis of the segregation obtained and sequence of the genomic region [15]. If a non-PCNA has a single A allele (Aaaaaa) and is crossed with a PCNA individual (aaaaaa), 50% of the offspring will be astringent. In the program the F<sub>1</sub> seedlings obtained and crossed with the PCNA ‘Cal Fuyu’ resulted in a different percentage of astringent genotypes. It has been demonstrated that the number of PCNA obtained depends on the allelic dose of the F<sub>1</sub> backcrossed [17]. In this study, the number of individuals per progeny was very low for studying segregation ratios and inferred the number of A alleles in the F<sub>1</sub> mothers. However, taking all the tested BC<sub>1</sub> seedlings together the rate of astringent genotypes was around 50%, which indicates that the F<sub>1</sub> group of maternal genitors might contain one AST allele on average.

Flower gender analysis revealed that around half of the genotypes analyzed have the capacity to generate the male flower (47.4%). This proportion is as expected, since the crosses need always a parent bearing male flowers. In the IVIA breeding program the presence of male flowers is a discarded trait for all types of fruit. In astringent types as PCA, PVA and PVNA, the presence of male flowers resulted in fruits pollinated and the production of seeds that brown the flesh and difficult the postharvest treatment for removing the astringency. PCNA types are discarded as well because

consumers do not accept seeds in the fruits and the presence of male flowers can pollinate astringent fruits, affecting negatively the quality and marketability of them.

Combined results of both markers, AST positive (astringency of the fruit) and DISx-AF4S positive (presence of male flowers) resulted in a high number of discarded genotypes. In column (AST- and DISx-AF4S) from Table 2 we indicated the genotypes that will be selected according to our breeding objectives. Only genotypes not astringent (PCNA type) and without male flowers will be selected (absence of both markers), a total of 118 from 441 (26.8%). It is important to point out that the markers segregated independently. According to the published genome of *Diospyros oleifera* [34], identified as the diploid *D. kaki* ancestor, the DNA fragments from which the markers were derived are located in different chromosomes. Therefore, the AST and DISx-AF4 markers must segregate independently.

The low rate of genotypes selected from the populations generated indicates the usefulness of the MAS applied in persimmon breeding. We could select at a plantlet stage the genotypes that will be planted in the fields for further agronomic selection. This MAS avoids keeping the future rejected plants during 4 years in the experimental fields. In our case near to 75% of the genotypes obtained can be discarded at the seedling stage in the greenhouse, indicating a high effectiveness of MAS in persimmon breeding.

#### 4. Conclusions

The markers DISx-AF4, linked to the production of male flowers and AST linked to astringency of the fruits, have been validated in a germplasm collection of persimmon. Although the markers were developed from Japanese cultivars, the correlation between the phenotype and genotype was 100% in germplasm from a different origin, which demonstrated the usefulness of the markers for selecting these important traits. Both markers have been screened in different progenies from a backcross that includes an astringent parent from non-Japanese origin. Results demonstrated that selection of both traits combined resulted in a very low rate of selection. In a context of breeding programs that involve astringent cultivars the MAS applied to discriminate PCNA genotypes is highly valuable.

**Author Contributions:** Conceptualization, M.B. and M.L.B.; Methodology, M.B. and F.G.-M.; Investigation, M.B. and M.d.M.N.; Writing—Original Draft, M.L.B.; Writing—Review and Editing, M.B., F.G.-M. and M.d.M.N.; Funding Acquisition, M.L.B. and M.B.; Resources, M.L.B.; Supervision, M.L.B. and M.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by IVIA grant 51914 and partially funded by FEDER. FGM was funded by European Social Fund and Generalitat Valenciana grant ACIF/2016/115.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### References

1. Dellaporta, S.L.; Calderon-Urrea, A. Sex determination in flowering plants. *Plant Cell* **1993**, *5*, 1241–1251. [[PubMed](#)]
2. Yonemori, K.; Sugiura, A.; Tanaka, K.; Kameda, K. Floral Ontogeny and Sex Determination in Monoecious-type Persimmons. *J. Am. Soc. Hortic. Sci.* **1993**, *118*, 293–297. [[CrossRef](#)]
3. Xu, L.Q.; Zhang, Q.L.; Luo, Z.R. Occurrence and cytological mechanism of 2n pollen formation in Chinese *Diospyros* spp. (Ebenaceae) staminate germplasm. *J. Hortic. Sci. Biotechnol.* **2008**, *83*, 668–672. [[CrossRef](#)]
4. Yakushiji, H.; Yamada, M.; Yonemori, K.; Sato, A.; Kimura, N. Staminate flower production on shoots of Fuyu' and "Jiro" persimmon (*Diospyros kaki* Thunb.). *J. Jpn. Soc. Hortic. Sci.* **1995**, *64*, 41–46. [[CrossRef](#)]
5. Giordani, E.; Picardi, E.; Radice, S. Morfologia y Fisiologia. In *El Cultivo del Caqui*; Badenes, M.L., Intrigliolo, D.S., Salvador, A., Vicent, A., Eds.; Generalitat Valenciana: Valencia, Spain, 2015; pp. 17–33; ISBN 9788448260187.
6. Kajiuira, M.; Blumenfeld, A. *Diospyros kaki*. In *CRC Handbook of Flowering*; Halevy, A.H., Ed.; CRC Press: Boca Raton, FL, USA, 1989; Volume 6, pp. 298–306; ISBN 9781315893464.
7. Yamada, M.; Giordani, E.; Yonemori, K. Persimmon. In *Fruit Breeding*; Badanes, M.L., Byrne, D.H., Eds.; Springer: New York, NY, USA, 2012; Volume 8, pp. 663–693; ISBN 9781441907639.

8. Sugiura, A.; Yonemori, K.; Harada, H.; Tomana, T. Changes of ethanol and acetaldehyde contents in Japanese persimmon fruits and their relation to natural deastringency. *Stud. Inst. Hortic. Kyoto Univ.* **1979**, *9*, 41–47.
9. Sugiura, A.; Tomana, T. Relationships of ethanol production by seeds of different types of Japanese persimmons and their tannin content. *Hort. Sci.* **1983**, *18*, 319–321.
10. Akagi, T.; Kajita, K.; Kibe, T.; Morimura, H.; Tsujimoto, T.; Nishiyama, S.; Kawai, T.; Yamane, H.; Tao, R. Development of molecular markers associated with sexuality in *Diospyros lotus* L. and their application in *D. kaki* Thunb. *J. Jpn. Soc. Hortic. Sci.* **2014**, *83*, 214–221. [[CrossRef](#)]
11. Akagi, T.; Henry, I.M.; Tao, R.; Comai, L. A Y-chromosome-encoded small RNA acts as a sex determinant in persimmons. *Science* **2014**, *346*, 646–650. [[CrossRef](#)]
12. Yonemori, K.; Matsushima, J. Property of Development of the Tannin Cells in Non-Astringent Type Fruits of Japanese Persimmon (*Diospyros kaki*) and Its Relationship to Natural Deastringency. *J. Jpn. Soc. Hortic. Sci.* **1985**, *54*, 201–208. [[CrossRef](#)]
13. Ikeda, I.; Yamada, M.; Kurihara, A.; Nishida, T. Inheritance of Astringency in Japanese Persimmon. *J. Jpn. Soc. Hortic. Sci.* **1985**, *54*, 39–45. [[CrossRef](#)]
14. Kanzaki, S.; Yamada, M.; Sato, A.; Mitani, N.; Ustunomiya, N.; Yonemori, K. Conversion of RFLP markers for the selection of pollination-constant and non-astringent type persimmons (*Diospyros kaki* Thunb.) into PCR-based markers. *J. Jpn. Soc. Hortic. Sci.* **2009**, *78*, 68–73. [[CrossRef](#)]
15. Kono, A.; Kobayashi, S.; Onoue, N.; Sato, A. Characterization of a highly polymorphic region closely linked to the AST locus and its potential use in breeding of hexaploid persimmon (*Diospyros kaki* Thunb.). *Mol. Breed.* **2016**, *36*, 1–13. [[CrossRef](#)]
16. Kanzaki, S.; Yonemori, K.; Sugiura, A.; Sato, A.; Yamada, M. Identification of molecular markers linked to the trait of natural astringency loss of Japanese persimmon (*Diospyros kaki*) fruit. *J. Jpn. Soc. Hortic. Sci.* **2001**, *126*, 51–55. [[CrossRef](#)]
17. Ikegami, A.; Yonemori, K.; Sugiura, A.; Sato, A.; Yamada, M. Segregation of astringency in F1 progenies derived from crosses between pollination-constant, nonastringent persimmon cultivars. *HortScience* **2004**, *39*, 371–374. [[CrossRef](#)]
18. Akagi, T.; Takeda, Y.; Yonemori, K.; Ikegami, A.; Kono, A.; Yamada, M.; Kanzaki, S. Quantitative genotyping for the astringency locus in hexaploid persimmon cultivars using quantitative real-time PCR. *J. Jpn. Soc. Hortic. Sci.* **2010**, *135*, 59–66. [[CrossRef](#)]
19. Kanzaki, S.; Akagi, T.; Masuko, T.; Kimura, M.; Yamada, M.; Sato, A.; Mitani, N.; Ustunomiya, N.; Yonemori, K. SCAR markers for practical application of marker-assisted selection in persimmon (*Diospyros kaki* Thunb.) breeding. *J. Jpn. Soc. Hortic. Sci.* **2010**, *79*, 150–155. [[CrossRef](#)]
20. Onoue, N.; Kobayashi, S.; Kono, A.; Sato, A. SSR-based molecular profiling of 237 persimmon (*Diospyros kaki* Thunb.) germplasms using an ASTRINGENCY-linked marker. *Tree Genet. Genomes* **2018**, *14*. [[CrossRef](#)]
21. Martínez-Calvo, J.; Naval, M.; Zuriaga, E.; Llácer, G.; Badenes, M.L. Morphological characterization of the IVIA persimmon (*Diospyros kaki* Thunb.) germplasm collection by multivariate analysis. *Genet. Resour. Crop Evol.* **2013**, *60*, 233–241. [[CrossRef](#)]
22. Del Naval, M.M.; Zuriaga, E.; Pecchioli, S.; Llácer, G.; Giordani, E.; Badenes, M.L. Analysis of genetic diversity among persimmon cultivars using microsatellite markers. *Tree Genet. Genomes* **2010**, *6*, 677–687. [[CrossRef](#)]
23. Doyle, J.J.; Doyle, J.L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytolog. Bull* **1987**, *19*, 11–15. [[CrossRef](#)]
24. Soriano, J.M.; Pecchioli, S.; Romero, C.; Vilanova, S.; Llácer, G.; Giordani, E.; Badenes, M.L. Development of microsatellite markers in polyploid persimmon (*Diospyros kaki* Thunb) from an enriched genomic library. *Mol. Ecol. Notes* **2006**, *6*, 368–370. [[CrossRef](#)]
25. Besada, C.; Novillo, P.; Navarro, P.; Salvador, A. Causes of flesh browning in persimmon—A review. *Acta Hortic.* **2018**, *1195*, 203–210. [[CrossRef](#)]
26. Munera, S.; Besada, C.; Blasco, J.; Cubero, S.; Salvador, A.; Talens, P.; Aleixos, N. Astringency assessment of persimmon by hyperspectral imaging. *Postharvest Biol. Technol.* **2017**, *125*, 35–41. [[CrossRef](#)]
27. Yamada, M. Persimmon Breeding in Japan. *Jpn. Agric. Res. Q.* **1993**, *27*, 33–37.
28. Yamada, M.; Sato, A.; Yakushiji, H.; Yoshinaga, K.; Yamane, H.; Endo, M. Characteristics of “Luo Tian Tian Shi”, a non-astringent cultivar of oriental persimmon (*Diospyros kaki* Thunb.) of Chinese origin in relation to non-astringent cultivars of Japanese origin. *Bull. Fruit Tree Res. Stn.* **1993**, *25*, 19–32.

29. Bellini, E.; Giordani, E. Germplasm and breeding of persimmon in Europe. *Acta Hortic.* **2005**, *685*, 65–74. [[CrossRef](#)]
30. Badenes, M.L.; Martinez-Calvo, J.; Naval, M.M. The persimmon breeding program at IVIA: Alternatives to conventional breeding of persimmon. *Acta Hortic.* **2013**, *996*, 71–76. [[CrossRef](#)]
31. Luo, Z.; Zhang, Q.; Luo, C.; Xie, F. Recent advances of persimmon research and industry in China. *Acta Hortic.* **2013**, *996*, 43–48. [[CrossRef](#)]
32. Ma, K.B.; Lee, I.B.; Kim, Y.K.; Won, K.H.; Cho, K.S.; Choi, J.J.; Lee, B.H.N.; Kim, M.S. ‘Jowan’, an early maturing PCNA (pollination constant non-astringent) persimmon (*Diospyros kaki* Thunb.). *Acta Hortic.* **2018**, *1195*, 61–64. [[CrossRef](#)]
33. Mitani, N.; Kono, A.; Yamada, M.; Sato, A.; Kobayashi, S.; Ban, Y.; Ueno, T.; Shiraishi, M.; Kanzaki, S.; Tsujimoto, T.; et al. Application of marker-assisted selection in persimmon breeding of PCNA offspring using SCAR markers among the population from the cross between Non-PCNA ‘taigetsu’ and PCNA ‘kanshu’. *HortScience* **2014**, *49*, 1132–1135. [[CrossRef](#)]
34. Zhu, Q.G.; Xu, Y.; Yang, Y.; Guan, C.F.; Zhang, Q.Y.; Huang, J.W.; Grierson, D.; Chen, K.S.; Gong, B.C.; Yin, X.R. The persimmon (*Diospyros oleifera* Cheng) genome provides new insights into the inheritance of astringency and ancestral evolution. *Hortic. Res.* **2019**. [[CrossRef](#)] [[PubMed](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).





Article

# Multi-Trait Regressor Stacking Increased Genomic Prediction Accuracy of Sorghum Grain Composition

Sirjan Sapkota <sup>1,2,\*</sup>, J. Lucas Boatwright <sup>1,2</sup>, Kathleen Jordan <sup>1</sup>, Richard Boyles <sup>2,3</sup> and Stephen Kresovich <sup>1,2</sup>

<sup>1</sup> Advanced Plant Technology Program, Clemson University, Clemson, SC 29634, USA; jboatw2@clemson.edu (J.L.B.); kjorda7@clemson.edu (K.J.); skresov@clemson.edu (S.K.)

<sup>2</sup> Department of Plant and Environmental Sciences, Clemson University, Clemson, SC 29634, USA; rboyles@clemson.edu

<sup>3</sup> Pee Dee Research and Education Center, Clemson University, Florence, SC 29506, USA

\* Correspondence: ssapkot@clemson.edu

Received: 4 July 2020; Accepted: 14 August 2020; Published: 19 August 2020

**Abstract:** Genomic prediction has enabled plant breeders to estimate breeding values of unobserved genotypes and environments. The use of genomic prediction will be extremely valuable for compositional traits for which phenotyping is labor-intensive and destructive for most accurate results. We studied the potential of Bayesian multi-output regressor stacking (BMORS) model in improving prediction performance over single trait single environment (STSE) models using a grain sorghum diversity panel (GSDP) and a biparental recombinant inbred lines (RILs) population. A total of five highly correlated grain composition traits—amylose, fat, gross energy, protein and starch, with genomic heritability ranging from 0.24 to 0.59 in the GSDP and 0.69 to 0.83 in the RILs were studied. Average prediction accuracies from the STSE model were within a range of 0.4 to 0.6 for all traits across both populations except amylose (0.25) in the GSDP. Prediction accuracy for BMORS increased by 41% and 32% on average over STSE in the GSDP and RILs, respectively. Prediction of whole environments by training with remaining environments in BMORS resulted in moderate to high prediction accuracy. Our results show regression stacking methods such as BMORS have potential to accurately predict unobserved individuals and environments, and implementation of such models can accelerate genetic gain.

**Keywords:** genomics; genomic selection; genomic prediction; marker-assisted selection; whole genome regression; grain quality; near infra-red spectroscopy; cereal crop; sorghum; multi-trait

## 1. Introduction

Cereal grains provide more than half of the total human caloric consumption globally and amount to over 80% in some of the poorest nations of the world [1]. Sorghum [*Sorghum bicolor* (L.) Moench], a drought-tolerant cereal crop, is a dietary staple for over half a billion people of semi-arid tropics which is inhabited by some of the most food insecure and malnourished populations [2]. In industrialized countries, such as the United States and Australia, grain sorghum is primarily grown for animal feed. But in recent years the uses of sorghum grain have expanded to baking, malting, brewing, and biofortification [3–5]. Therefore, genetic improvement of sorghum grain composition is crucial to mitigate the global malnutrition crisis, to increase efficiency of feed grains used in animal production, and to serve evolving niche markets for gluten-free grains.

In the last two decades, the use of genome-wide markers in prediction of genetic merit of individuals has revolutionized plant and animal breeding. Genomic prediction (GP) uses statistical models to estimate marker effects in a training population with phenotypic and genotypic data which is then used to predict breeding values of individuals solely from genetic markers [6,7].

Training population size, genetic relatedness between individuals in training and testing population, marker density, span of linkage disequilibrium and genetic architecture of traits are some of the factors that can affect the predictive ability of the models [8–10]. Genomic prediction models are routinely studied and applied by breeding programs around the world in several crops. Novel statistical methods that are capable of incorporating pedigree, genomic, and environmental covariates into statistical-genetic prediction models have emerged as a result of extensive computational research [11].

One of the main advantages of GP is that breeders can use phenotypic values from some lines in some environments to make predictions of new lines and environments. Genomic best linear unbiased prediction (GBLUP) proposed by VanRaden [12] is probably the most widely used genomic prediction model in both plant and animal breeding. Since then GBLUP model has been extended to include  $G \times E$  interactions resulting in improved prediction accuracy of unobserved lines in environments [13,14]. Burgueño et al. [13] found an increase in prediction ability of unobserved wheat genotypes by about 20% in multi-environment GBLUP model compared to single environment model. Also an extension of the GBLUP model, Jarquín et al. [14] introduced a reaction norm model which introduces the main and interaction effects of markers and environmental covariates by using high-dimensional random variance-covariance structures of markers and environmental covariates. While most of the genomic prediction studies have been on individual traits, breeding programs use selection indices based on several traits to make breeding decisions. To address those challenges, expanded genomic prediction models that perform joint analysis of multiple traits have been studied using empirical and simulated data [15,16]. Subsequent improvement in prediction accuracy from multi-trait model over single-trait model depends on trait heritability and correlation between the traits involved [15,17].

Data generated in breeding programs span multiple environment and are recorded for multiple traits for each individual. While multi-environment models and multi-trait models are implemented separately, a single model to account for complexity of variance-covariance structure in a combined multi-trait multi-environment (MTME) model was lacking until Montesinos-López et al. [18] developed a Bayesian whole genome prediction model to incorporate and analyze multiple traits and multiple environments simultaneously. Montesinos-López et al. [18] also developed a computationally efficient Markov Chain Monte Carlo (MCMC) method that produces a full conditional distribution of the parameters leading to an exact Gibbs sampling for the posterior distribution. Another MTME model that employs a completely different method was proposed by Montesinos-López et al. [19]. This method, called the Bayesian multi-output regression stacking (BMORS), is a Bayesian version of multi-target regressor stacking (MTRS) originally proposed by Spyromitros-Xioufif et al. [20,21]. This method consists of training in two stages: first training multiple learning algorithms for the same dataset and then subsequently combining the predictions to obtain the final predictions.

Genomic prediction for grain quality traits has previously been reported in crops such as wheat [22–24], rye [25], maize [26], and soybean [27]. Hayes et al. [28] and Battenfield et al. [23] used near-infrared derived phenotypes in genomic prediction of protein content and end-use quality in wheat. Multi-trait genomic prediction models can simultaneously improve grain yield and protein content despite being negatively correlated [24,29]. In sorghum, grain macronutrients have shown to be inter-correlated among one another [30], which suggests the multi-trait models may increase predictive ability of individual grain quality traits. The ability to assess genetic merit of unobserved selection candidates across environments is promising for reducing evaluation cost and generation interval in the sorghum breeding pipeline where parental lines of commercial hybrids are currently selected on the basis of extensive progeny testing [31]. In order to extend capacities to performance index selection for multiple environments, we need to study and effectively implement MTME genomic prediction models in our breeding programs. In this study, we report the first implementation of genomic prediction for grain composition in sorghum, and the objective was to assess potential for improvement in prediction accuracy of multi-trait regressor stacking model over single trait model for five grain composition traits: amylose, fat, gross energy, protein and starch.

## 2. Results

### 2.1. Phenotypic Variation

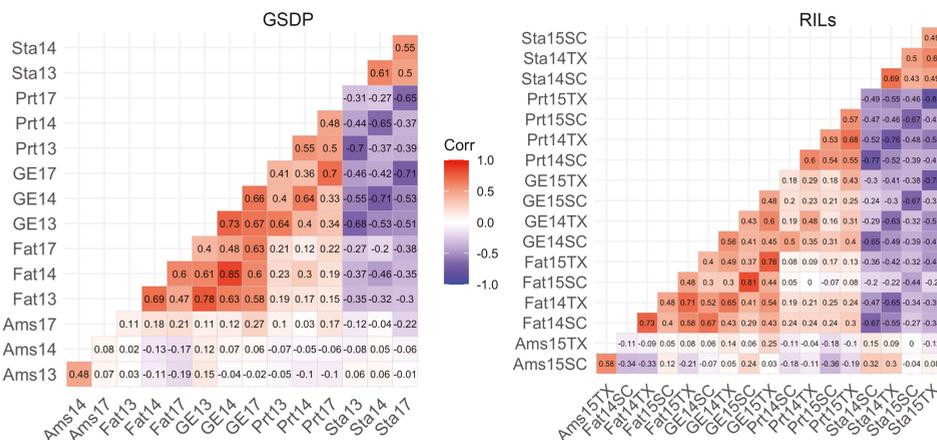
A single calibration curve for near infra-red spectroscopy (NIRS) was used for the two populations studied. Table 1 outlines the summary statistics of NIRS predictions and phenotypic distribution and heritability of the grain composition traits. The cross validation accuracy ( $R^2$ ) of the NIRS calibration curve was moderately high to high, except for fat which had a moderate  $R^2$  value (0.41). We had a total of three environments (three years in one location) for the GSDP and four environments (two years in two locations) for the RILs. Traits were normally distributed except amylose in two 2014 environments in the RILs which had bimodal distribution (Figures S1 and S2). All traits showed significant variation in distribution across the environments, except starch in GSDP.

**Table 1. Summary statistics of near infra-red spectroscopy (NIRS) calibration and phenotypic distribution.**  $R^2$  is the prediction accuracy and SECV is the standard error of cross validation for the NIRS calibration curve. Mean represents the phenotypic mean of the trait with its standard deviation (SD).  $h^2$  is the estimate of genomic heritability.

Trait	NIRS		GSDP		RILs	
	$R^2$	SECV	Mean $\pm$ SD	$h^2$	Mean $\pm$ SD	$h^2$
Amylose	0.60	2.24	13.87 $\pm$ 2.98	0.24	11.49 $\pm$ 4.32	0.77
Fat	0.41	0.53	2.53 $\pm$ 0.57	0.54	3.07 $\pm$ 0.67	0.76
Gross energy	0.71	25.80	4108.33 $\pm$ 55.15	0.59	4124.56 $\pm$ 41.74	0.69
Protein	0.96	0.27	12.02 $\pm$ 1.45	0.39	11.43 $\pm$ 1.03	0.83
Starch	0.89	0.75	68.30 $\pm$ 2.44	0.44	68.37 $\pm$ 1.87	0.79

The genomic heritabilities of all traits except gross energy were significantly higher ( $p < 0.05$ ) in the RILs than in the GSDP (Table 1). Trait heritabilities were high in the RILs, with protein and gross energy having the highest and lowest heritabilities, respectively. In the GSDP, genomic heritability was moderately high for fat and gross energy, moderate for protein and starch, and low for amylose. The poor genomic heritability (0.24) of amylose in the GSDP was expected because only a very small proportion (1%) of accessions have low amylose as a result of *waxy* gene (Mendelian trait).

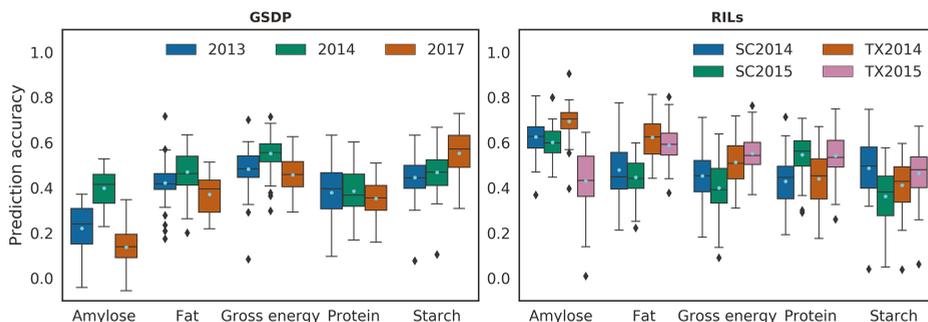
Figure 1 shows correlation between the adjusted phenotypic means for trait and environment combination. Starch was negatively correlated ( $p < 0.001$ ) with all other traits in both populations except for amylose in the RILs. Fat, protein and gross energy were significantly positively ( $p < 0.001$ ) correlated to each other across environments in both populations. The strongest positive correlation was between gross energy and fat, whereas the strongest negative correlations were found between starch gross energy and starch protein. Moderate (0.4) to high (0.73) positive correlation was observed between years for all traits except for amylose ( $r = 0.08$ ) between 2014 and 2017 in GSDP (Figure 1). We conducted a principal component analysis (PCA) of correlation matrix for the traits in each environment. In both populations, the first component separated amylose and starch from the other three traits, whereas, the second component separated amylose from starch and gross energy from protein and fat (Figure S3). The first component explained 78.8% and 75.9% of variation, and second component explained 6.3% and 9.8% of variation in the GSDP and RILs, respectively. The third principal component in the RILs separated proteins from fat and explained about 7.6% of the variation.



**Figure 1.** Correlation between traits across year and location combination for the two populations. Ams: amylose, GE: gross energy, Prt: protein, Sta: starch, SC: South Carolina, TX: Texas, and numbers in x and y-axes represent the year.

2.2. Prediction Performance in Single and Multiple Environment

We first implemented GBLUP prediction model for single-trait single-environment (STSE). Prediction accuracies of the STSE model varied across environments in both populations (Figure 2). The environments 2014 in the GSDP and TX2014 in the RILs had highest average prediction accuracy but were not always the best predicted environment for all traits. While poorly predicted for amylose, the environments 2017 in the GSDP and TX2015 in the RILs had higher prediction accuracy for starch compared to all or most environments. Despite variation across environments and populations, the average prediction accuracies from the STSE were within the range of 0.4 to 0.6 for all traits except amylose (0.25) in the GSDP. The average prediction accuracy of the STSE model in the GSDP was positively correlated ( $r = 0.86$ ) with the genomic heritability of the traits. In the RILs, there was a positive correlation ( $r = 0.77$ ) between average prediction accuracy and genomic heritability for amylose, fat and gross energy, but the traits (protein and starch) with the highest heritabilities had relatively lower average prediction accuracies.



**Figure 2.** Prediction accuracy for single-trait single-environment (STSE) model. The y-axis shows prediction accuracy calculated as Pearson’s correlation between observed values and predicted values of phenotypes. Legend represents the environment/years. SC: South Carolina, TX: Texas, GSDP: Grain sorghum diversity panel, RILs: recombinant inbred lines. Pale blue dots represent the mean of prediction accuracy.

We did not see substantial improvement in multi-environment (BME) model over the STSE prediction accuracies (Figure S4). In the GSDP, the multi-environment models resulted in a decline in average prediction accuracy compared to the STSE model for fat (21%), amylose (10%) and protein (4%), however, no significant change was observed for gross energy and starch (Table 2, Figure S5). The prediction accuracy in the RILs increased by an average of 3% in the BME compared to the STSE, however, the overall trend of prediction accuracy for traits and environments remained unchanged (Table 3, Figure S5). The environment SC2014 showed consistent increase in accuracy for BME over STSE model across all traits with about 10% increase for protein (Table 3). Amylose in TX2015 environment had the single greatest increase (12%) in average prediction accuracy in the BME among all trait-environment combinations for the RILs (Table 3).

**Table 2.** Percent change in mean prediction accuracy ( $r$ ) over the single trait single environment (STSE) model in the diversity panel (GSDP). BME: Bayesian multi-environment, and BMORS: Bayesian multi-output regressor stacking. Values were rounded to the nearest whole number.

Trait	2013		2014		2017	
	BME	BMORS	BME	BMORS	BME	BMORS
Amylose	−11	66	−5	11	−13	−13
Fat	−24	47	−12	47	−27	58
Gross energy	3	54	−2	40	1	57
Protein	−3	56	−1	55	−8	52
Starch	4	37	−2	38	1	17
<b>Average</b>	<b>−6</b>	<b>52</b>	<b>−4</b>	<b>38</b>	<b>−9</b>	<b>34</b>

**Table 3.** Percent change in mean prediction accuracy ( $r$ ) over the single trait single environment (STSE) model in the recombinant inbred lines (RILs). BME: Bayesian multi-environment, BMORS: Bayesian multi-output regressor stacking, SC: South Carolina, and TX: Texas. Values were rounded to the nearest whole number.

Trait	SC2014		SC2015		TX2014		TX2015	
	BME	BMORS	BME	BMORS	BME	BMORS	BME	BMORS
Amylose	2	28	0	28	−1	25	12	43
Fat	5	33	1	15	2	18	2	20
Gross energy	7	28	−3	27	1	18	3	17
Protein	10	51	1	23	5	60	−4	33
Starch	5	36	−4	40	7	54	4	37
<b>Average</b>	<b>6</b>	<b>35</b>	<b>−1</b>	<b>27</b>	<b>3</b>	<b>35</b>	<b>3</b>	<b>30</b>

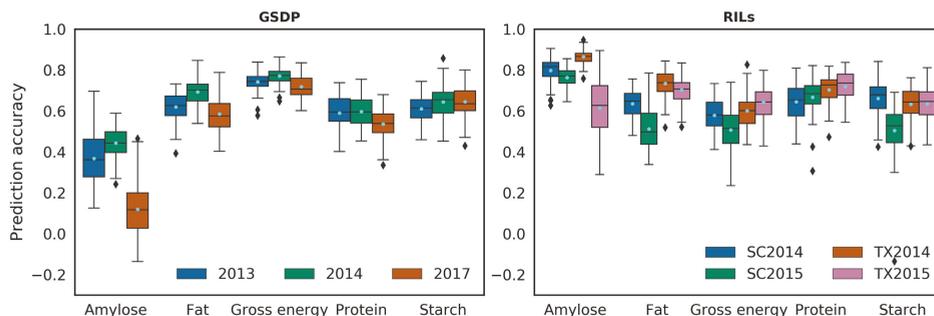
### 2.3. Bayesian Multi-Output Regression Stacking

We tested two different prediction schemes in the BMORS prediction model using the two functions  $BMORS()$  and  $BMORS\_Env()$  as described in Montesinos-López et al. [19]. While the  $BMORS()$  function was used for a five-fold CV as described in the methods section, the  $BMORS\_Env()$  was used to assess the prediction performance of whole environments while using the remaining environments as training. So in  $BMORS\_Env$ , an environment was left out during training and correlation between the predicted values (obtained from training with remaining environments) and observed values for the test environment was measured as prediction accuracy for that environment in  $BMORS\_Env$  model.

#### 2.3.1. Five-Fold CV

The prediction accuracy from five-fold CV in BMORS increased by 41% and 32% on average over the STSE model in GSDP and RILs, respectively. Figure 3 shows the prediction accuracy of BMORS for each trait and environment combination across the two populations. While the percent change in

accuracy varied across environments, the BMORS model nonetheless had higher average prediction accuracy than the STSE and BME models for all traits (Figure S4). The increase in average accuracy in BMORS over STSE ranged from 11% (amylose, 2014) to 66% (amylose, 2013) in the GSDP with exception of amylose in 2017 (13% decrease), and 15% (fat, SC2015) to 60% (protein, TX2014) in the RILs (Tables 2 and 3). The increase in average prediction accuracy was higher (35%) for both locations in 2014 for the RILs, whereas, the year 2013 in the GSDP increased the most. Among the traits, protein (54%) had the greatest average increase in prediction accuracy in the GSDP, whereas in the RILs, protein and starch (42%) both showed the greatest increase.

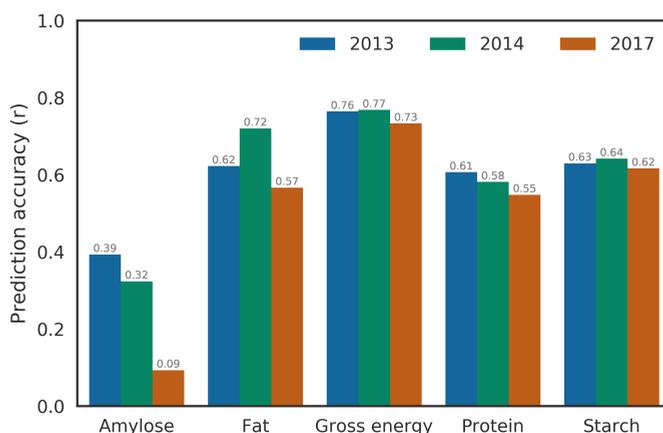


**Figure 3. Prediction accuracy of Bayesian multi-output regressor stacking (BMORS) model using five-fold cross validation.** Legend represents the years/environment. The y-axis shows prediction accuracy calculated as Pearson's correlation between observed values and predicted values of phenotypes. SC: South Carolina, TX: Texas, GSDP: Grain sorghum diversity panel, and RILs: Recombinant inbred lines. Pale blue dots show mean of the prediction accuracy.

### 2.3.2. Prediction of Whole Environment

Predicting a whole environment using the BMORS model usually yielded higher accuracy than the mean prediction accuracy from the STSE or BME model where only portions of each environment was tested instead of whole environment as in *BMORS\_Env* model (Figures 2 and 4, Figure S5, Table 4). This shows that BMORS model can be reliably used in predicting unobserved environment with the same accuracy as from STSE or BME models from training within the environments. The distribution of prediction accuracy across trait and environment combination were, however, similar to the results from the STSE model. In the GSDP, little variation in prediction accuracies was observed across environments for gross energy, starch and protein, whereas, amylose and fat showed greater variability in prediction accuracy between environments. In the RILs, prediction accuracy for all traits except protein had high variability across the environments (Table 4).

In order to assess predictability by location or year in the RILs, we tested one location or year by training the BMORS model using the other location or year, respectively (Table 4). The Texas location had higher accuracy of prediction for fat (+0.11) and gross energy (+0.1) compared to South Carolina, but rest of the traits had similar prediction accuracy (difference < 0.02). Prediction accuracy of whole years varied across traits, amylose (+0.09) and fat(+0.04) were higher in 2014, protein was higher (+0.05) in 2015, and starch and gross energy were similar.



**Figure 4.** Prediction accuracy of whole environment predicted using the Bayesian multi-output regressor stacking (BMORS\_Env) in the diversity panel (GSDP). The y-axis shows prediction accuracy calculated as Pearson's correlation between observed values and predicted values of phenotypes. Values on top of the bar represent the height of the bar.

**Table 4.** Prediction accuracy of the test environments predicted using the Bayesian multi-output regressor stacking (BMORS\_Env) in the recombinant inbred lines (RILs). SC: South Carolina, TX, Texas. Prediction accuracy was calculated as Pearson's correlation between observed values and predicted values of phenotypes.

Trait	Year × Location				Location		Year	
	SC2014	SC2015	TX2014	TX2015	SC	TX	2014	2015
Amylose	0.79	0.80	0.88	0.60	0.76	0.74	0.74	0.65
Fat	0.69	0.49	0.78	0.74	0.60	0.71	0.64	0.60
Gross energy	0.56	0.49	0.62	0.66	0.48	0.58	0.56	0.56
Protein	0.65	0.66	0.66	0.70	0.59	0.58	0.61	0.66
Starch	0.64	0.52	0.68	0.60	0.55	0.56	0.56	0.55

### 3. Discussion

Phenotyping for grain compositional traits is—(1) challenging and labor-intensive, (2) destructive for most accurate results, and (3) only performed after plants reach physiological maturity and are harvested. The use of genomic prediction for compositional traits will be extremely valuable because it increases selection intensity and decreases generational interval by overcoming the phenotyping challenges. Moreover, these traits are complex and quantitatively inherited so will benefit from genomic prediction's ability to account for many small effect QTLs in estimating breeding values.

#### 3.1. Trait Architecture and Prediction Accuracy

While the accuracy of NIRS calibration for traits in this study ranged from moderate to high, there was prediction error associated with NIRS prediction. However, it is unclear if and what effects NIRS prediction error had on genomic prediction. No direct correlation was observed between the genomic prediction accuracy and NIRS statistics for the traits studied. The trait with the lowest NIRS  $R^2$ , fat, was predicted as well as or better than starch, protein and gross energy, which had NIRS  $R^2 > 0.7$ . Despite varying strength of correlations between traits across the two populations studied,

the nature of relationship was similar for a given pair of traits, which is also in agreement with previous studies [30,32,33]. The strong negative relationship of starch and amylose to protein, fat and gross energy was further elucidated by the PCA analysis of phenotypic correlation matrix (Figure S3). Since starch, protein and fat were measured on a percent dry matter basis, the strong correlation between them is expected.

Genetic relatedness and trait architecture are known to affect the accuracy of genomic prediction [8,34]. The genetic relatedness between individuals and heritability of the traits were higher in the RILs than in GSDP (Figure S6, Table 1). Those factors could be contributing to higher average prediction accuracy in the RILs. However, the average prediction accuracies for gross energy and starch were comparable between GSDP and RILs (Figure S4). Prediction accuracy in the GSDP could have been boosted by greater genetic diversity despite lower genetic relatedness [35]. Heffner et al. [22] observed a prediction accuracy of 0.5–0.6 for wheat flour protein in two biparental populations. Guo et al. [26] reported prediction accuracies of 0.44 and 0.8 for protein and amylose in rice diversity panel. Similar results were observed in our STSE models for protein content (Figure 2). Whereas, lower prediction accuracy of amylose in our diversity panel is probably due to the lack of sufficient low-amylose lines with the *waxy* gene [30]. While genomic prediction study for starch, fat and gross energy has not been reported in sorghum, these traits are nutritionally one of the most important traits for any cereal grain. The moderate to high prediction accuracy observed suggests implementation of genomic selection can improve genetic gain for these grain quality traits.

### 3.2. Multi-Trait Regressor Stacking

One of the daunting tasks of genomic prediction is estimating the effects of unobserved individuals and environments. As multiple traits are analyzed across several environments, the ability to combine information from multiple traits and environments can be crucial in increasing accuracy of prediction [13,15,16]. When the correlations among traits are high, prediction accuracies of complex traits can be increased by using multivariate model that takes this correlation into account [15,18]. We fit a Bayesian multi-environment (BME) model (2) that takes the genotype  $\times$  environment effects into consideration. In the GSDP, where environments were three years at the same location, the BME model showed a slight decline (7%) in average prediction accuracy which was mostly due to the two traits, amylose and fat (Table 2). The RILs showed slight increase (2–3%) in prediction accuracy of traits when averaged over the environments, but there was variability across the environments (Table 3).

We implemented two functions [*BMORS()* and *BMORS\_Env()*] which are not only used to evaluate prediction accuracy but are also computationally efficient [19]. The *BMORS* model (3) performs two-stage training by stacking the multi-environment models from all the traits. The five-fold cross validation conducted for *BMORS* was similar to the CV1 strategy of Montesinos-López et al. [18]. The use of multi-trait models has been consistently shown to increase prediction accuracy over single-trait models across different crops and traits [15–17,36]. The multi-target regressor stacking increased average prediction accuracy by 41% and 32% in the GSDP and RILs, respectively, as compared to the STSE prediction accuracy. Average prediction accuracy of all traits improved in *BMORS* over STSE and BME across both the populations (Figure S4). Consistent improvement in accuracy of *BMORS* is a result of the ability to use not only correlation between traits but also between environment in the model training [18,19]. The ability to accurately estimate genetic merit of lines in unobserved environments is of tremendous value in plant breeding. Our results show potential of *BMORS\_Env()* function for predicting the whole environment. Testing a whole environment by training *BMORS* model using all other environments resulted in higher prediction accuracy for that trait-environment combination than using STSE or BME model. Prediction accuracy of all environments were 0.5 or higher with exception of amylose in GSDP, the reason for which we have discussed above (Figure 4, Table 4).

### 3.3. Application for Breeding

Grain quality traits such as starch and protein content have been under selection since the inception of phenotypic selection in modern breeding practices. More recently, total energy supplement of grain has gained attention for increasing feed efficiency in animal production, and a need exists for increasing total calories for human nutrition in the wake of global malnutrition crisis. Despite high correlations among these traits, the genetic variation underlying starch, protein and fat can be decoupled. Boyles et al. [30] showed major and minor effect QTLs underlying the three traits are distributed across the genome and are segregating in biparental populations. However, in practice, selection would be conducted simultaneously for these traits using a selection index rather than for individual traits. Velazco et al. [31] observed an increase in predictive ability by using a multi-trait model for grain yield and stay green in sorghum, and argue that such an exercise would allow for using selection index for implementation of genomic selection for correlated traits. Increased prediction accuracy, improved selection index, and estimation of precise genetic, environmental and residual co-variances makes multi-trait multi-environment models preferable over univariate models [18]. The multi-trait regression stacking model we tested shows large scale improvement in model prediction and can be used in tandem with Bayesian multi-trait multi-environment (BMTME) model for parameter estimation and assessing prediction accuracy. The ability to estimate genetic effects and breeding values of unobserved environments will be of great advantage to predict performance in diverse environments and for implementation of selection theory.

## 4. Materials and Methods

### 4.1. Plant Material

#### 4.1.1. Grain Sorghum Diversity Panel

A grain sorghum diversity panel (GSDP) of 389 diverse sorghum accessions was planted in randomized complete block design with two replications in 2013, 2014, and 2017 field seasons at the Clemson University Pee Dee Research and Education Center in Florence, SC. The GSDP included a total of 332 accessions from the original United States sorghum association panel (SAP) developed by Casa et al. [37]. The details on experimental field design and agronomic practices are described in Boyles et al. [38] and Sapkota et al. [35]. Briefly, the experiments were planted in a two row plots each 6.1 m long, separated by row spacing of 0.762 m with an approximate planting density of 130,000 plants ha<sup>-1</sup>. Fields were irrigated only when signs of drought stress was seen across the field.

#### 4.1.2. Recombinant Inbred Population

A biparental population of 191 recombinant inbred lines (RILs) segregating for grain quality traits was studied along with the GSDP. The parents of the RIL population were BTx642, a yellow-pericarp drought tolerant line, and BTxARG-1, a white pericarp waxy endosperm (low amylose) line. The population was planted in two replicated plots in randomized complete block design across two years (2014 and 2015) in Blackville, SC and College Station, TX. Field design and agronomic practices have previously been described in detail in Boyles et al. [30].

### 4.2. Phenotyping

The primary panicle of three plants selected from each plot were harvested at physiological maturity. The plants from beginning and end of the row were excluded to account for border effect. Panicles were air dried to a constant moisture (10–12%) and threshed. A 25 g subsample of cleaned and homogenized grain ground to 1-mm particle size with a CT 193 Cyclotec Sample Mill (FOSS North America) was used in near-infrared spectroscopy (NIRS) for compositional analysis.

Grain composition traits such as total fat, gross energy, crude protein, and starch content can be measured using NIRS. Previous studies have shown high NIRS predictability of the traits used in

feed analysis [39,40]. We used a DA 7250™ NIR analyzer (Perten Instruments). The ground sample was packed in a gradually rotating Teflon dish positioned under the instrument's light source and predicted phenotypic values was generated based on calibration curve for spectral measurements. The calibration curve was built using wet chemistry values from a subset of samples. The wet chemistry was performed by Dairyland Laboratories, Inc. (Arcadia, WI, USA) and the Quality Assurance Laboratory at Murphy-Brown, LLC (Warsaw, NC, USA). The details on the prediction curves and wet chemistry can be found in Boyles et al. [30].

#### 4.3. Genotypic Data

Genotyping-by-sequencing (GBS) was used for genetic characterization of the GSDP and RILs populations [30,38,41]. Sequenced reads were aligned to the BTx623 v3.1 reference assembly (phytozome) using Burrows-Wheeler aligner [42]. SNP calling, imputation and filtering was done using TASSEL 5.0 pipeline [43]. The TASSEL plugin FILLIN for GSDP and FSHap for RILs population were used to impute for missing genotypes. Following imputation SNPs with minor allele frequency (MAF) < 0.01, and sites missing in more than 10% and 30% of the genotypes in GSDP and RILs, respectively, were filtered. The number of genotypes studied for each population represent those with at least 70% of SNP sites. The genotype matrix with 224,007 SNPs from GSDP and 56,142 SNPs from RILs population was used for whole genome regression.

#### 4.4. Statistical Analysis

The statistical software environment 'R' was used for model building and analysis [44]. The phenotypic values of the traits were adjusted for random effects of replications within environment using 'lme4' package in R [45]. Principal component analysis was done using the R package 'factoextra' [46]. Marker-based estimates of narrow sense (genomic) heritabilities were calculated using the SNP genotype matrix and phenotypic values using the R package 'heritability' [47]. A matrix with dummy variables '1' and '0' representing combinations of environmental variables (replication and year for GSDP, and replication, year and location for RILs) was used as co-variate in heritability estimation.

##### 4.4.1. Single-Trait Single-Environment (STSE) Model:

The following genomic best linear unbiased prediction (GBLUP) model was used to assess prediction performance of an individual trait from a single environment:

$$y_j = \mu + g_j + e_j, \quad (1)$$

where  $y_j$  is a vector of adjusted phenotypic mean of the  $j$ th line ( $j = 1, 2, \dots, J$ ).  $\mu$  is the overall mean which is assigned a flat prior,  $g_j$  is a vector of random genomic effect of the  $j$ th line, with  $g = (g_1, \dots, g_j)^T \sim N(0, G\sigma_g^2)$ ,  $\sigma_g^2$  is a genomic variance,  $G$  is the genomic relationship matrix in the order  $J \times J$  and is calculated [12] as  $G = \frac{ZZ^T}{2\sum p_j q_j}$ , where  $q_j$  and  $p_j$  denote major and minor allele frequency of  $j$ th line respectively, and  $Z$  is the design matrix for markers of order  $J \times p$  ( $p$  is total number of markers). Further,  $e_j$  is residual error assigned the normal distribution  $e \sim N(0, I\sigma_e^2)$  where  $I$  is identity matrix and  $\sigma_e^2$  is the residual variance with a scaled-inverse Chi-square density.

##### 4.4.2. Bayesian Multi-Environment (BME) GBLUP Model

Considering genotype  $\times$  environment interaction can contribute to substantial amount of phenotypic variance in complex traits, we fit the following univariate linear mixed model to account for environmental effects in prediction performance:

$$y_{ij} = E_i + g_j + gE_{ij} + e_{ij}, \quad (2)$$

where  $y_j$  is a vector of adjusted phenotypic mean of the  $j$ th line in the  $i$ th environment ( $i = 1, 2, \dots, I$ ,  $j = 1, 2, \dots, J$ ).  $E_i$  represents the effect of  $i$ th environment and  $g_j$  represents the genomic effect of the  $j$ th line as described in Equation (1). The term  $gE_{ij}$  represents random interaction between the genomic effect of  $j$ th line and the  $i$ th environment with  $gE = (gE_{11}, \dots, gE_{IJ})^T \sim N(0, \sigma_2^2 \mathbf{I}_I \otimes \mathbf{G})$ , where  $\sigma_2^2$  is an interaction variance, and  $e_{ij}$  is a random residual associated with the  $j$ th line in the  $i$ th environment distributed as  $N(0, \sigma_e^2)$  where  $\sigma_e^2$  is the residual variance.

#### 4.4.3. Bayesian Multi-Output Regressor Stacking (BMORS)

BMORS is the Bayesian version of multi-trait (or multi-target) regressor stacking method [48]. The multi-target regressor stacking (MTRS) was proposed by Spyromitros-Xioufis et al. [20,21] based on multi-labeled classification approach of Godbole and Sarawagi [49]. In BMORS or MTRS, the training is done in two stages. First,  $L$  univariate models are implemented using the multi-environment GBLUP model given in Equation (2), then instead of using these models for prediction, MTRS performs the second stage of training using a second set of  $L$  meta-models for each of the  $L$  traits. The following model is used to implement each meta-model:

$$y_{ij} = \beta_1 \hat{Z}_{1ij} + \beta_2 \hat{Z}_{2ij} + \dots + \beta_L \hat{Z}_{Lij} + e_{ij}, \quad (3)$$

where the covariates  $\hat{Z}_{1ij}, \hat{Z}_{2ij}, \dots, \hat{Z}_{Lij}$  represent the scaled prediction from the first stage training with the GBLUP model for  $L$  traits, and  $\beta_1, \dots, \beta_L$  are the regression coefficients for each covariate in the model. The scaling of each prediction was performed by subtracting its mean ( $\mu_{lij}$ ) and dividing by its corresponding standard deviation ( $\sigma_{lij}$ ), that is,  $\hat{Z}_{lij} = (\hat{y}_{lij} - \mu_{lij})\sigma_{lij}^{-1}$ , for each  $l = 1, \dots, L$ . The scaled predictions of its response variables yielded by the first-stage models as predictor information by the BMORS model. Simply put, the multi-trait regression stacking model is based on the idea that a second stage model is able to correct the predictions of a first-stage model using information about the predictions of other first-stage models [20,21].

#### 4.4.4. Performance of Prediction Model:

All prediction models were fit using Bayesian approach in statistical program 'R'. The STSE model (1) was fit using the R package 'BGLR' [50], BME model (2) and BMORS model (3) were fit using the R package 'BMTME' [19]. A minimum of 20,000 iterations with 10,000 burn-in steps was used for each Bayesian run.

The evaluation of prediction performance of models was done using a five-fold cross validation (CV), which means 80% of the samples were used as training set and testing was done on the remaining 20% for each cross-validation fold. The individuals were randomly assigned into five mutually exclusive folds. Four folds were used to train prediction models and to predict the genomic estimated breeding values (GEBVs) of the individuals in fifth fold (validation/test set). The accuracy of prediction for each fold was calculated as Pearson's correlation coefficient ( $r$ ) between predicted values and adjusted phenotypic means for the individuals in validation set. Each cross validation run, therefore, resulted in five estimates of prediction accuracy. The same set of individuals were assigned to training and validation across different traits and models tested by using *set.seed()* function in R. In order to avoid bias due to sampling, we performed 10 different cross-validation runs to calculate the mean and dispersion of the prediction accuracies.

## 5. Conclusions

Phenotyping of grain compositional traits using near-infrared spectroscopy is labor-intensive, generally destructive, and time limiting. Therefore, the use of genomic selection for these traits will be extremely valuable. This study establishes the potential to improve genomics-assisted selection of grain composition traits by using multi-trait multi-environment model. The phenotypic measurements obtained from NIRS prediction were amenable to genomic selection as shown by moderate to high

prediction accuracy for single trait prediction. While multi-environment model alone did not lead to much improvement over single environment model, stacking of regression from multiple traits showed substantial improvement in prediction accuracy. The prediction accuracy increased by 32% and 41% in the RILs and GSDP, respectively, when using the Bayesian multi-output regressor stacking (BMORS) model compared to a single trait single environment model. The ability to predict line performance in an unobserved environment is of great importance to breeding programs, and results show high accuracy for predicting whole environments using BMORS.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4395/10/9/1221/s1>. The supplementary file contains six figures: Figure S1. Phenotypic distribution of grain composition traits in the RILs. In the x-axes, SC: South Carolina, TX: Texas, numbers represent years. Values are percentage dry basis for protein, fat and starch; gross energy is in KCal/lb; and amylose is in percent of starch. Figure S2. Phenotypic distribution of grain composition traits in the GSDP. Numbers in x-axes represent years. Values are percentage dry basis for protein, fat and starch; gross energy is in Cal/g; and amylose is in percent of starch. Figure S3. PCA analysis of correlation matrix between traits. a. GSDP, and b. RILs. Ams: amylose, GE: gross energy, Prt: protein, Sta: starch, SC: South Carolina, TX: Texas. The numbers in the text represent years of the environment. Figure S4. Overall prediction accuracy of traits across all the environment for the three prediction methods in the two populations. The y-axis shows prediction accuracy calculated as Pearson's correlation between observed values and predicted values of phenotypes. Legend represents the environment/years. SC: South Carolina, TX: Texas, GSDP: Grain sorghum diversity panel, RILs: recombinant inbred lines. Figure S5. Prediction accuracy using five-fold CV in Bayesian multi-environment (BME) model. a. GSDP, and b. RILs. Legend represents the environment/years. SC: South Carolina, TX: Texas. Pale blue dots represent the mean of prediction accuracy. Figure S6. Heatmap for genomic relationship matrix calculated using vanRaden (2008). a. GSDP, b. RILs. Trees show hierarchical clustering using Euclidean distance.

**Author Contributions:** S.S. conceptualized the study, performed data analysis, and wrote the manuscript; R.B. helped with experimental design and field phenotyping; J.L.B. helped in computation and data analysis; K.J. helped in near infra-red phenotyping; S.K. helped with conception of the study, acquisition of fund, and management of the study. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was funded by United States Department of Energy TERRA grant: DE-AR0001134.

**Acknowledgments:** The authors would like to thank William L. Rooney and Brian K. Pfeiffer for their contributions to phenotyping of the recombinant inbred population at College Station, TX. Our appreciation goes to the Wade Stackhouse Fellowship, and Robert and Lois Coker Endowment for their support during the study.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

**Data Availability:** The codes and phenotypic data used in the study can be accessed through github at [sirjansapkota/GrainComp\\_GS](https://github.com/sirjansapkota/GrainComp_GS).

## Abbreviations

The following abbreviations are used in this manuscript:

GSDP	Grain sorghum diversity panel
RIL	Recombinant inbred line
NIRS	Near infra-red spectroscopy
GP	Genomic prediction
GBLUP	Genomic best linear unbiased prediction
STSE	Single trait single environment
BME	Bayesian multi-environment
BMORS	Bayesian multi-output regressor stacking
MTME	Multi-trait multi-environment
CV	Cross validation
SC	South Carolina
TX	Texas
QTL	Quantitative trait loci
SNP	Single nucleotide polymorphism

## References

- Awika, J.M. Major cereal grains production and use around the world. In *Advances in Cereal Science: Implications to Food Processing and Health Promotion*; ACS Publications: Washington, DC, USA, 2011; pp. 1–13.
- Mace, E.S.; Tai, S.; Gilding, E.K.; Li, Y.; Prentis, P.J.; Bian, L.; Cruickshank, A. Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nat. Commun.* **2013**, *4*, 2320. [[CrossRef](#)]
- Taylor, J.R.; Schober, T.J.; Bean, S.R. Novel food and non-food uses for sorghum and millets. *J. Cereal Sci.* **2006**, *44*, 252–271. [[CrossRef](#)]
- Taylor, J. Food product development using sorghum and millets: opportunities and challenges. *Qual. Assur. Saf. Crop. Foods* **2012**, *4*, 151. [[CrossRef](#)]
- Zhu, F. Structure, physicochemical properties, modifications, and uses of sorghum starch. *Compr. Rev. Food Sci. Food Saf.* **2014**, *13*, 597–610. [[CrossRef](#)]
- Meuwissen, T.H.E.; Hayes, B.J.; Goddard, M.E. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* **2001**, *157*, 1819–1829.
- Bernardo, R.; Yu, J. Prospects for genomewide selection for quantitative traits in maize. *Crop. Sci.* **2007**, *47*, 1082–1090. [[CrossRef](#)]
- Habier, D.; Fernando, R.L.; Dekkers, J.C. The impact of genetic relationship information on genome-assisted breeding values. *Genetics* **2007**, *177*, 2389–2397. [[CrossRef](#)] [[PubMed](#)]
- Zhong, S.; Dekkers, J.C.; Fernando, R.L.; Jannink, J.L. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: A barley case study. *Genetics* **2009**, *182*, 355–364. [[CrossRef](#)] [[PubMed](#)]
- Combs, E.; Bernardo, R. Accuracy of genomewide selection for different traits with constant population size, heritability, and number of markers. *Plant Genome* **2013**, *6*, 1–7. [[CrossRef](#)]
- Crossa, J.; Pérez-Rodríguez, P.; Cuevas, J.; Montesinos-López, O.; Jarquín, D.; de los Campos, G.; Dreisigacker, S. Genomic selection in plant breeding: Methods, models, and perspectives. *Trends Plant Sci.* **2017**, *22*, 961–975. [[CrossRef](#)]
- VanRaden, P.M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **2008**, *91*, 4414–4423. [[CrossRef](#)] [[PubMed](#)]
- Burgueño, J.; de los Campos, G.; Weigel, K.; Crossa, J. Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. *Crop. Sci.* **2012**, *52*, 707–719. [[CrossRef](#)]
- Jarquín, D.; Crossa, J.; Lacaze, X.; Du Cheyron, P.; Daucourt, J.; Lorgeou, J.; Burgueño, J. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor. Appl. Genet.* **2014**, *127*, 595–607. [[CrossRef](#)] [[PubMed](#)]
- Jia, Y.; Jannink, J.L. Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics* **2012**, *192*, 1513–1522. [[CrossRef](#)]
- Guo, G.; Zhao, F.; Wang, Y.; Zhang, Y.; Du, L.; Su, G. Comparison of single-trait and multiple-trait genomic prediction models. *BMC Genet.* **2014**, *15*, 30. [[CrossRef](#)]
- Lado, B.; Vázquez, D.; Quincke, M.; Silva, P.; Aguilar, I.; Gutiérrez, L. Resource allocation optimization with multi-trait genomic prediction for bread wheat (*Triticum aestivum* L.) baking quality. *Theor. Appl. Genet.* **2018**, *131*, 2719–2731. [[CrossRef](#)]
- Montesinos-López, O.A.; Montesinos-López, A.; Crossa, J.; Toledo, F.H.; Pérez-Hernández, O.; Eskridge, K.M.; Rutkoski, J. A genomic Bayesian multi-trait and multi-environment model. *G3 Genes Genomes Genet.* **2016**, *6*, 2725–2744. [[CrossRef](#)]
- Montesinos-López, O.A.; Montesinos-López, A.; Luna-Vázquez, F.J.; Toledo, F.H.; Pérez-Rodríguez, P.; Lillemo, M.; Crossa, J. An R package for Bayesian analysis of multi-environment and multi-trait multi-environment data for genome-based prediction. *G3 Genes Genomes Genet.* **2019**, *9*, 1355–1369. [[CrossRef](#)]
- Spyromitros-Xioufis, E.; Tsoumakas, G.; Groves, W.; Vlahavas, I. Multi-target regression via input space expansion: Treating targets as inputs. *Mach. Learn.* **2016**, *104*, 55–98.
- Spyromitros-Xioufis, E.; Tsoumakas, G.; Groves, W.; Vlahavas, I. Multi-label classification methods for multi-target regression. *arXiv* **2012**, arXiv:12116581.

22. Heffner, E.L.; Jannink, J.L.; Iwata, H.; Souza, E.; Sorrells, M.E. Genomic selection accuracy for grain quality traits in biparental wheat populations. *Crop. Sci.* **2011**, *51*, 2597–2606. [[CrossRef](#)]
23. Battenfield, S.D.; Guzmán, C.; Gaynor, R.C.; Singh, R.P.; Peña, R.J.; Dreisigacker, S.; Poland, J.A. Genomic selection for processing and end-use quality traits in the CIMMYT spring bread wheat breeding program. *Plant Genome* **2016**, *9*, 1–12. [[CrossRef](#)] [[PubMed](#)]
24. Haile, J.K.; N'Diaye, A.; Clarke, F.; Clarke, J.; Knox, R.; Rutkoski, J.; Pozniak, C.J. Genomic selection for grain yield and quality traits in durum wheat. *Mol. Breed.* **2018**, *38*, 75. [[CrossRef](#)]
25. Schulthess, A.W.; Wang, Y.; Miedaner, T.; Wilde, P.; Reif, J.C.; Zhao, Y. Multiple-trait-and selection indices-genomic predictions for grain yield and protein content in rye for feeding purposes. *Theor. Appl. Genet.* **2016**, *129*, 273–287. [[CrossRef](#)] [[PubMed](#)]
26. Guo, Z.; Tucker, D.M.; Basten, C.J.; Gandhi, H.; Ersoz, E.; Guo, B.; Gay, G. The impact of population structure on genomic prediction in stratified populations. *Theor. Appl. Genet.* **2014**, *127*, 749–762. [[CrossRef](#)] [[PubMed](#)]
27. Duhnen, A.; Gras, A.; Teyssède, S.; Romestant, M.; Claustres, B.; Daydé, J.; Mangin, B. Genomic selection for yield and seed protein content in Soybean: A study of breeding program data and assessment of prediction accuracy. *Crop. Sci.* **2017**, *57*, 1325–1337. [[CrossRef](#)]
28. Hayes, B.; Panozzo, J.; Walker, C.; Choy, A.; Kant, S.; Wong, D.; Spangenberg, G.C. Accelerating wheat breeding for end-use quality with multi-trait genomic predictions incorporating near infrared and nuclear magnetic resonance-derived phenotypes. *Theor. Appl. Genet.* **2017**, *130*, 2505–2519. [[CrossRef](#)]
29. Rapp, M.; Lein, V.; Lacoudre, F.; Lafferty, J.; Müller, E.; Vida, G.; Leiser, W.L. Simultaneous improvement of grain yield and protein content in durum wheat by different phenotypic indices and genomic selection. *Theor. Appl. Genet.* **2018**, *131*, 1315–1329. [[CrossRef](#)]
30. Boyles, R.E.; Pfeiffer, B.K.; Cooper, E.A.; Rauh, B.L.; Zielinski, K.J.; Myers, M.T.; Kresovich, S. Genetic dissection of sorghum grain quality traits using diverse and segregating populations. *Theor. Appl. Genet.* **2017**, *130*, 697–716. [[CrossRef](#)]
31. Velazco, J.G.; Jordan, D.R.; Mace, E.S.; Hunt, C.H.; Malosetti, M.; Van Eeuwijk, F.A. Genomic prediction of grain yield and drought-adaptation capacity in sorghum is enhanced by multi-trait analysis. *Front. Plant Sci.* **2019**, *10*, 997. [[CrossRef](#)]
32. Murray, S.C.; Sharma, A.; Rooney, W.L.; Klein, P.E.; Mullet, J.E.; Mitchell, S.E.; Kresovich, S. Genetic improvement of sorghum as a biofuel feedstock: I. QTL for stem sugar and grain nonstructural carbohydrates. *Crop. Sci.* **2008**, *48*, 2165–2179. [[CrossRef](#)]
33. Sukumaran, S.; Xiang, W.; Bean, S.R.; Pedersen, J.F.; Kresovich, S.; Tuinstra, M.R.; Yu, J. Association mapping for grain quality in a diverse sorghum collection. *Plant Genome* **2012**, *5*, 126–135. [[CrossRef](#)]
34. Jannink, J.L.; Lorenz, A.J.; Iwata, H. Genomic selection in plant breeding: From theory to practice. *Briefings Funct. Genom.* **2010**, *9*, 166–177. [[CrossRef](#)] [[PubMed](#)]
35. Sapkota, S.; Boyles, R.; Cooper, E.; Brenton, Z.; Myers, M.; Kresovich, S. Impact of sorghum racial structure and diversity on genomic prediction of grain yield components. *Crop. Sci.* **2020**, *60*, 132–148. [[CrossRef](#)]
36. Bhatta, M.; Gutierrez, L.; Cammarota, L.; Cardozo, F.; Germán, S.; Gómez-Guerrero, B.; Castro, A.J. Multi-trait Genomic Prediction Model Increased the Predictive Ability for Agronomic and Malting Quality Traits in Barley (*Hordeum vulgare* L.). *G3 Genes Genomes Genet.* **2020**, *10*, 1113–1124. [[CrossRef](#)] [[PubMed](#)]
37. Casa, A.M.; Pressoir, G.; Brown, P.J.; Mitchell, S.E.; Rooney, W.L.; Tuinstra, M.R.; Kresovich, S. Community resources and strategies for association mapping in sorghum. *Crop. Sci.* **2008**, *48*, 30–40. [[CrossRef](#)]
38. Boyles, R.E.; Cooper, E.A.; Myers, M.T.; Brenton, Z.; Rauh, B.L.; Morris, G.P.; Kresovich, S. Genome-wide association studies of grain yield components in diverse sorghum germplasm. *Plant Genome* **2016**, *9*, 1–17. [[CrossRef](#)]
39. Kays, S.E.; Barton, F.E. Rapid prediction of gross energy and utilizable energy in cereal food products using near-infrared reflectance spectroscopy. *J. Agric. Food Chem.* **2002**, *50*, 1284–1289. [[CrossRef](#)]
40. De Alencar Figueiredo, L.F.; Sine, B.; Chantereau, J.; Mestres, C.; Fliedel, G.; Rami, J.F.; Courtois, B. Variability of grain quality in sorghum: Association with polymorphism in Sh2, Bt2, SssI, Ae1, Wx and O2. *Theor. Appl. Genet.* **2010**, *121*, 1171–1185. [[CrossRef](#)]
41. Morris, G.P.; Ramu, P.; Deshpande, S.P.; Hash, C.T.; Shah, T.; Upadhyaya, H.D.; Riera-Lizarazu, O.; Brown, P.J.; Acharya, C.B.; Mitchell, S.E.; et al. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 453–458. [[CrossRef](#)]

42. Li, H.; Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **2010**, *26*, 589–595. [[CrossRef](#)]
43. Glaubitz, J.C.; Casstevens, T.M.; Lu, F.; Harriman, J.; Elshire, R.J.; Sun, Q.; Buckler, E.S. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS ONE* **2014**, *9*, e90346. [[CrossRef](#)] [[PubMed](#)]
44. R Core Team. R: A Language and Environment for Statistical Computing. 2019. Available online: <https://www.R-project.org/> (accessed on 26 April 2019).
45. Bates, D.; Mächler, M.; Bolker, B.; Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* **2015**, *67*, 1–48. [[CrossRef](#)]
46. Kassambara, A.; Mundt, F. Factoextra: extract and visualize the results of multivariate data analyses. *R Package Version* **2017**, *1*, 337–354.
47. Kruijer, W.; Boer, M.P.; Malosetti, M.; Flood, P.J.; Engel, B.; Kooke, R.; van Eeuwijk, F.A. Marker-based estimation of heritability in immortal populations. *Genetics* **2015**, *199*, 379–398. [[CrossRef](#)] [[PubMed](#)]
48. Montesinos-López, O.A.; Montesinos-López, A.; Crossa, J.; Cuevas, J.; Montesinos-López, J.C.; Gutiérrez, Z.S.; Singh, R. A Bayesian genomic multi-output regressor stacking model for predicting multi-trait multi-environment plant breeding data. *G3 Genes Genomes Genet.* **2019**, *9*, 3381–3393. [[CrossRef](#)] [[PubMed](#)]
49. Godbole, S.; Sarawagi, S. Discriminative methods for multi-labeled classification. In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Sydney, Australia, 26–28 May 2004; Springer: Berlin/Heidelberg, Germany, 2004; p. 22–30.
50. Pérez, P.; de Los Campos, G. Genome-wide regression and prediction with the BGLR statistical package. *Genetics* **2014**, *198*, 483–495. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).





Article

# QTL Mapping for Seedling and Adult Plant Resistance to Leaf and Stem Rusts in Pamyati Azieva × Paragon Mapping Population of Bread Wheat

Yuliya Genievszkaya <sup>1</sup>, Saule Abugalieva <sup>1,2</sup>, Aralbek Rsaliyev <sup>3</sup>, Gulbahar Yskakova <sup>3</sup> and Yerlan Turuspekov <sup>1,4,\*</sup>

<sup>1</sup> Laboratory of Molecular Genetics, Institute of Plant Biology and Biotechnology, Almaty 050040, Kazakhstan; julia.genievszkaya@gmail.com (Y.G.); absaule@yahoo.com (S.A.)

<sup>2</sup> Department of Biodiversity and Bioresources, Faculty of Biology and Biotechnology, al-Farabi Kazakh National University, Almaty 050040, Kazakhstan

<sup>3</sup> Laboratory of Phytosanitary Safety, Research Institute of Biological Safety Problems, Gvardeisky 080409, Zhambyl Region, Kazakhstan; aralbek@mail.ru (A.R.); y\_gulbahar@mail.ru (G.Y.)

<sup>4</sup> Agrobiology Faculty, Kazakh National Agrarian University, Almaty 050010, Kazakhstan

\* Correspondence: yerlant@yahoo.com; Tel.: +7-727-394-8006

Received: 6 August 2020; Accepted: 22 August 2020; Published: 30 August 2020

**Abstract:** Leaf rust (LR) and stem rust (SR) pose serious challenges to wheat production in Kazakhstan. In recent years, the susceptibility of local wheat cultivars has substantially decreased grain yield and quality. Therefore, local breeding projects must be adjusted toward the improvement of LR and SR disease resistances, including genetic approaches. In this study, a spring wheat segregating population of Pamyati Azieva (PA) × Paragon (Par), consisting of 98 recombinant inbred lines (RILs), was analyzed for the resistance to LR and SR at the seedling and adult plant-growth stages. In total, 24 quantitative trait loci (QTLs) for resistance to rust diseases at the seedling and adult plant stages were identified, including 11 QTLs for LR and 13 QTLs for SR resistances. Fourteen QTLs were in similar locations to QTLs and major genes detected in previous linkage mapping and genome-wide association studies. The remaining 10 QTLs are potentially new genetic factors for LR and SR resistance in wheat. Overall, the QTLs revealed in this study may play an important role in the improvement of wheat resistance to LR and SR per the marker-assisted selection approach.

**Keywords:** *Triticum aestivum*; QTL; mapping population; leaf rust; stem rust; pathogen races; disease resistance

## 1. Introduction

Bread wheat (*Triticum aestivum* L.) is one of the major cereal crops in the world. In 2018/2019, the global production of wheat was 734.7 million metric tons, ranking second place amongst the grains after maize [1]. It is used mostly as flour for the production of a large variety of leavened and flat breads and the manufacturing of a wide range of other baked products [2]. In 2018/2019, Kazakhstan was ranked the 12th largest wheat producer in the world [3]. In Kazakhstan, wheat is cultivated on about 13 million hectares annually. The country produces up to 20–25 million tons of bread wheat per year and exports up to 5–7 million tons of the grain [4]. The primary goals of modern wheat breeding programs worldwide include enhancing grain yield and quality and increasing resistance to biotic and abiotic stresses to ensure global food security [5]. Biotic stresses include dangerous fungal diseases and particularly the most common representatives of the *Puccinia* genus: *Puccinia recondita* Rob. ex Desm. f. sp. *tritici*, causing leaf rust (LR), and *Puccinia graminis* Pers. f. sp. *tritici* Eriks. & Henn., which is responsible for stem rust (SR) of wheat.

LR generally causes light to moderate yield losses ranging from 1% to 20% over a large area, but when the disease is severe prior to heading time, it may destroy up to 90% of the wheat crop [6]. For example, in Kazakhstan, epiphytotic development of the pathogen on spring wheat during 2000–2001 resulted in 50–100% LR severity on commercial cultivars in the Akmola region (northern Kazakhstan), which is the main wheat-growing region in the country [7].

SR is another important rust disease that is often considered the most devastating of the wheat rust diseases because it can cause complete crop loss over a large area within a short period of time [8]. In 2016, northern Kazakhstan was subjected to an epiphytotic outbreak of SR, resulting in 50% disease development severity in the field, decreasing wheat yield and grain quality [7]. Nowadays, local farmers prefer the usage of fungicides to protect wheat fields from LR and SR; however, this method is harmful to the environment and more expensive than breeding and growing genetically resistant wheat cultivars [9].

LR and SR resistances are controlled by a diverse group of genes, designated as *Lr* and *Sr*, respectively [10]. In the last 100 years, approximately 80 *Lr* resistance genes have been identified and described in bread wheat, durum wheat, and diploid wheat species [10], and the list is still growing. For SR, nearly 60 *Sr* genes have been identified to date in wheat and its wild relatives [10]. Generally, resistance to rust diseases can be broadly categorized into two types. The first is resistance at all growth stages (called seedling resistance), detected at the seedling stage and expressed until the plant dies. This type of resistance is controlled by the R type of genes, and the majority of *Lr* and *Sr* genes belong to this group. The efficacy of the R gene is pathogen-strain-dependent [11]. The second type of resistance is adult plant resistance (APR), where genes are ineffective during the seedling stage but provide robust resistance at maturity [11]. For example, LR resistance genes *Lr12*, *Lr13*, *Lr22a*, *Lr22b*, *Lr34*, *Lr35*, *Lr46*, *Lr48*, *Lr49*, and *Lr67*, and SR resistance genes *Sr2* and *Sr57* are well-characterized APR genes [10]. Durable rust resistance is more likely to be the APR type rather than the seedling type [12]; both types are important for wheat breeding [11].

Two of the most effective methods of quantitative trait locus (QTL) mapping are based on association panels and biparental segregating populations [13]. Both of these methods provide the means to investigate the genome and describe the etiology of complex quantitative traits, including disease resistance [14–17]. Genetic maps are a key tool enabling genetic linkage studies and searches for novel loci responsible for traits. Modern high-throughput sequencing technologies allow for the high-accuracy genotyping of large collections with genetically diverse germplasm [18,19] and segregating mapping populations, such as doubled haploids (DHs), recombinant inbred lines (RILs),  $F_2$ , and backcross (BC) populations [20]. Linkage maps were successfully used for the QTL analyses of wheat yield components [21], grain quality traits [22], abiotic [23], and biotic stress factors, including pests [24].

The primary goal of this study was to identify QTLs involved in seedling and adult plant resistance of bread wheat to LR and SR under environmental conditions in southern and southeastern Kazakhstan. To meet this goal, the Pamyati Azieva  $\times$  Paragon (PA  $\times$  Par) RILs mapping population (MP) was studied in field and greenhouse (GH) conditions. Previously, this population was successfully used for the analysis of yield-related traits [25] and adult plant resistance to LR and SR in south-east and northern Kazakhstan in 2018 [26,27]. Hence, the current study adds the investigation of seedling resistance in the MP to LR and SR races. In addition to one-year studies in south-east and north Kazakhstan, this work covers the analysis of LR and SR resistance in the MP in southern Kazakhstan in 2018 and 2019.

## 2. Materials and Methods

### 2.1. Plant Material and Genotyping

The biparental mapping population PA  $\times$  Par composed of 98 RILs was developed in greenhouse conditions of the John Innes Centre (Norwich, UK) during 2011–2015 under the ADAPTAWHEAT

project [28]. The RIL population was obtained via a single-seed descent method using two parental cultivars: Paragon (elite spring cultivar originated from the U.K.) and Pamyati Azieva (a commercial spring cultivar originating from Russia and registered in Kazakhstan) [29]. Both cultivars were chosen due to their diverse genetic backgrounds and different manifestations of yield traits as well as resistance to diseases. The RIL population was further developed for F<sub>8</sub> generation in the fields of southeastern Kazakhstan.

The RILs and two parental cultivars were genotyped using the Illumina's iSelect 20K single nucleotide polymorphism (SNP) array at the TraitGenetics Company (TraitGenetics GmbH, Gatersleben, Germany). The genotypic data were filtered from markers with >10% missing data and with <0.1 minor allele frequency and consisted of 4595 polymorphic SNP markers.

## 2.2. Phenotyping of Seedling Resistance in Greenhouse

For the comprehensive study of the PA × Par MP response to LR and SR pathogens, the resistance was evaluated at the seedling and adult plant-growth stages. Race-specific resistance at the seedling stage was assessed in a greenhouse of the Research Institute of Biological Safety Problems (RIBSP, Gvardeisky, Zhambyl region, southern Kazakhstan). For the inoculation of RILs seedlings (7–10 days after sowing) in greenhouse conditions, three races of *P. graminis* and three races of *P. recondita* with different levels of virulence to *Sr* and *Lr* genes, respectively, were used (Table 1). Inoculated plants were placed in the boxes of the greenhouse with appropriate temperature conditions (22 ± 2 °C for SR, 18 ± 2 °C for LR) and illumination (10,000–15,000 lux, 16 h' light period) [30–32]. RIL reaction was assessed on the 14th day after inoculation, according to the scale reported by Stakman [33]. The experiment was performed in two independent replicates.

**Table 1.** Virulence/avirulence pattern of pathogen races used in the study.

Disease (Pathogen)	Race	Avirulent (Effective) Genes	Virulent (Ineffective) Genes
LR ( <i>Puccinia recondita</i> Rob. ex Desm f. sp. <i>tritici</i> )	TQKHT	<i>Lr24, 26, 3ka, 19, 25</i>	<i>Lr1, 2a, 2c, 3, 9, 16, 11, 17, 30, 20, 29, 2b, 3bg, 14a, 15</i>
	TRTHT	<i>Lr24, 19, 25</i>	<i>Lr1, 2a, 2c, 3, 9, 16, 26, 3ka, 11, 17, 30, 20, 29, 2b, 3bg, 14a, 15</i>
	TQTMQ	<i>Lr24, 26, 20, 25, 14a, 15</i>	<i>Lr1, 2a, 2c, 3, 9, 16, 3ka, 11, 17, 30, 19, 29, 2b, 3bg</i>
SR ( <i>Puccinia graminis</i> Pers. f. sp. <i>tritici</i> Eriks. & E. Henn.)	TKRTF	<i>Sr11, 30, 24, 31</i>	<i>Sr5, 21, 9e, 7b, 6, 8a, 9g, 36, 9b, 17, 9a, 9d, 10, 38, Tmp, McN</i>
	PKCTC	<i>Sr21, 11, 36, 9b, 30, 24, 31, 38</i>	<i>Sr5, 9e, 7b, 6, 8a, 9g, 17, 9a, 9d, 10, Tmp, McN</i>
	RKRTF	<i>Sr9e, 11, 30, 24, 31</i>	<i>Sr5, 21, 7b, 6, 8a, 9g, 36, 9b, 17, 9a, 9d, 10, 38, Tmp, McN</i>

LR, leaf rust; SR, stem rust.

Races of *P. graminis* were differentiated in 2018 [34] using the North American nomenclature [35] with the assistance of five sets of SR-differentiating wheat cultivars. Races of *P. recondita* were also identified in 2018 using 20 Thatcher near-isogenic lines (NILs) sets of *Lr* genes [36–38]. For the nomenclature of *P. recondita* races, Virulence Analysis Tools [39] were used.

## 2.3. Adult Plant Resistance and Yield Components in Field Conditions

APR in the field was tested in two environments: the RIBSP and the Kazakh Research Institute of Agriculture and Plant Industry (KRIAPI, Almalybak, Almaty region, southeastern Kazakhstan) (Table 2).

**Table 2.** Meteorological data on average temperature and precipitations during the vegetation period in the fields of Research Institute of Biological Safety Problems (RIBSP, southern Kazakhstan) and Kazakh Research Institute of Agriculture and Plant Industry (KRIAPI, southeastern Kazakhstan).

	KRIAPI				
	March	April	May	June	July
Temperature (°C)	8.2	12.8	17.0	22.3	27.0
Precipitation (mm)	27.3	168.4	39.3	72.7	22.6
	RIBSP				
	March	April	May	June	July
Temperature (°C)	13.0	17.0	24.0	29.0	33.0
Precipitation (mm)	1.0	2.3	1.8	0.9	0.1

In RIBSP fields, mixed races of LR and SR urediniospores common in Kazakhstan were applied as inoculum. The inoculum was activated at a temperature of 37–40 °C for 30 min, followed by watering in a humid chamber at a temperature of 18–22 °C for 2 h. At the booting stage, individual plants were treated with an aqueous suspension of leaf and stem rust urediniospores dissolved in Tween 80 detergent. After inoculation, the plots were covered with plastic wrap for 16–18 h. In KRIAPI fields, inoculation occurred with local LR and SR pathogen populations in uncontrolled natural conditions.

Thus, experiments were conducted in three independent environments, including the study of seedling resistance in greenhouse conditions, the study of APR at RIBSP (controlled inoculation), and APR at KRIAPI (uncontrolled inoculation). In both field conditions, phenotyping of APR to LR and SR was performed in two independent replicates at the stage of grain ripening with the maximum level of disease manifestation. Disease assessment was performed using the scale of Stakman for SR [33] and the scale of Mains and Jackson for LR [40]. The severity of rust infection on leaf and stem surfaces was evaluated using the modified Cobb scale [41,42]. To meet the data format required for linkage analysis, the results of LR and SR evaluations at both seedling and adult plant-growth stages were converted to the 0–9 linear disease scale as described by Zhang et al. [43].

To identify the influence of LR and SR severity on the productivity of the studied population, two important yield-related components, thousand kernel weight (TKW, g) and kernel yield per plot (YP, g/m<sup>2</sup>), were also evaluated.

#### 2.4. Statistical Analysis of Phenotypic Data and QTL Mapping

Phenotypic data processing, descriptive statistics, and one-tailed correlation tests were performed with SPSS Statistics v. 22 (SPSS Inc., Chicago, IL, USA).

The composite interval mapping (CIM) method with the Kosambi mapping function was used for the detection of QTLs by Windows QTL Cartographer v2.5 [44]. The threshold value for the logarithm of odds (LOD) score was calculated based on 1000 permutations and was 3.0 for all experiments with a walking step of 1 cM. QTLs were detected for each environment and replication separately (seedling resistance in GH, APR at KRIAPI, and APR at RIBSP). QTLs identified in individual environments and/or replications overlapping in 20 cM intervals and associated with the same trait were considered as identical [45]. Genetic maps with QTLs were drawn using MapChart v. 2.32 software [46]. For the markers with the same positions, only one single nucleotide polymorphism (SNP) maker was selected for the map.

All genes present within the interval of 500 kb to the left and 500 kb to the right (1 Mb in total) from the peak marker were identified using the Ensembl Plant database [47]. As a reference, the genome of *T. aestivum* RefSeq v1 was used. The exact position of the peak SNP in the genome was determined using a BLAST tool [48]. Proteins and RNA gene products were identified using the UniProt database [49] via cross-reference from Ensembl Plant.

### 3. Results

#### 3.1. Phenotyping Variations of Seedling and Adult Plant Resistance in Mapping Population

The values of the resistance to target diseases in parents and 98 RILs are summarized in Table 3.

**Table 3.** Descriptive statistics for leaf rust (LR) and stem rust (SR) resistance at two plant-growth stages in the Pamyati Azieva × Paragon mapping population.

Env.	Plant-Growth Stage	Race (Disease)	Parents IT <sup>1</sup>				RILs IT			
			PA	Par	Mean	Range	R (%)	MR (%)	MS (%)	S (%)
GH	Seedling	TQTMQ (LR)	6 (3)	5 (3-)	6.1 ± 0.9	5–8	0	0	89.8	10.2
		TQKHT (LR)	8 (4-)	6 (3)	5.9 ± 1.0	4–9	0	3.1	91.8	5.1
		TRTHT (LR)	6 (3)	6 (3)	5.9 ± 0.8	4–8	0	1.0	96.9	2.1
	Seedling	TKRTF (SR)	8 (4-)	8 (4-)	7.0 ± 1.4	5–9	0	0	45.9	54.1
		PKCTC (SR)	5 (3-)	8 (4-)	5.4 ± 1.8	1–9	1.0	22.4	58.2	18.4
		RKRTF (SR)	6 (3)	5 (3-)	6.5 ± 1.4	5–9	0	0	64.3	35.7
RIBSP	Adult	LR	8 (30S)	6 (50MS)	3.8 ± 1.4	0–9	38.8	11.2	29.6	20.4
		SR	9 (80S)	8 (40S)	6.6 ± 2.9	1–9	1.0	23.5	18.4	57.1
KRIAPI	Adult	LR	6 (40MS)	6 (40MS)	6.6 ± 1.7	2–9	0	9.2	64.3	26.5
		SR	0 (0)	1 (10R)	1.8 ± 2.3	0–8	56.1	26.5	18.4	1.0

<sup>1</sup>—parents IT scores are given in 0–9 numeric scale, traditional IT scores are given in parentheses. Env., environment; GH, greenhouse conditions; PA, Pamyati Azieva; Par, Paragon; IT, infection type; R, percentage of resistant lines (0–1 on 9-point scale); MR, percentage of moderately resistant lines (2–4 on 9-point scale); MS, percentage of moderately susceptible lines (5–7 on 9-point scale); S, percentage of susceptible lines (8–9 on 9-point scale).

The average seedling resistance of RILs to LR races was between 5.9 and 6.1 points, corresponding to the moderately susceptible (MS) level. The major part of the population belonged to the MS group, with only several lines observed in the susceptible (S) group. Several lines were also in the resistant (R) group to the races TQKHT and TRTHT. Parental cultivars demonstrated an MS level of resistance to studied LR races, except for PA, which was susceptible to the race TQKHT. As for seedling resistance to SR, the average level in the RILs population was MS to all three SR races. However, unlike in the case of LR, the distribution of lines among resistance groups was different. Races TKRTF and RKRTF were divided between MS and S groups with a dominance of the S reaction to TKRTF and MS reaction to RKRTF. R and moderately resistant (MR) levels were detected only in the race PKCTC. Levels of resistance in parental cultivars were similar to races TKRTF (S) and RKRTF (MS), but PA demonstrated higher resistance to the race PKCTC than Par.

At the adult plant stage, the reactions of parents and RILs to LR and SR were significantly different between the studied environments. At RIBSP, the average reaction of RILs to LR was MR, with almost even distribution among all possible reactions observed in the population. At KRIAPI, the average level of resistance was MS with a dominance of MS and S reactions in the population. The parental cultivars demonstrated the same reaction to LR at KRIAPI, but at RIBSP, Par was more resistant than PA. For SR at the adult plant stage at RIBSP, the majority of RILs were in the S group, and the average level was MS. At KRIAPI, the largest part of the population was in the R group and the average level was MR. Parental cultivars also were in the S group at RIBSP and the R group at KRIAPI.

The analysis of variance showed that the resistance of RILs to LR at the seedling growth stage was significantly affected by the RIL genotype ( $p < 0.01$ ) and the race of LR pathogen ( $p < 0.05$ ), but not by genotype × race interaction (Table 4). For SR resistance, all factors had a significant influence ( $p < 0.001$ ) on the resistance at the seedling stage (Table 4).

**Table 4.** ANOVA of plant genotype (Geno), pathogen race (Race), and plant genotype × pathogen race (Geno: Race) effects on seedling resistance to leaf rust (LR) and stem rust (SR) in the Pamyati Azieva × Paragon mapping population.

Factor	df	LR			SR		
		MeanS	F	p	MeanS	F	p
Geno	97	2.506	1.630	0.001	5.380	2.679	$8.39 \times 10^{-11}$
Race	2	5.420	3.525	0.031	136.940	68.180	$<2.00 \times 10^{-16}$
Geno: Race	194	1.396	0.908	0.766	4.540	2.262	$1.20 \times 10^{-10}$
Residuals	294	1.537			2.010		

df, degree of freedom; MeanS, mean square.

### 3.2. Correlations among SR and LR Seedling and Adult Plant Resistance and Influence of APR on the Yield-Related Traits

Significant positive correlations were found among the reactions to all three LR races at the seedling stage, as well as between APR to LR at KRIAPI and seedling resistance to LR races TQTMQ and TRTHT (Table 5). APR to LR at KRIAPI was also positively correlated with APR to SR at KRIAPI and seedling resistance to SR race TKRTF. For the other SR races, race PKCTC had a positive correlation with APR to SR at KRIAPI, and race RKRTF was negatively correlated with LR race TQTMQ. The only significant correlation of APR to SR at RIBSP was associated with LR race TRTHT.

**Table 5.** Correlations among race-specific seedling resistance and adult plant resistance (APR) to leaf rust (LR) and stem rust (SR) in the Pamyati Azieva × Paragon mapping population.

	LR TQTMQ	LR TQKHT	LR TRTHT	APR LR RIBSP	APR LR KRIAPI	SR TKRTF	SR PKCTC	SR RKRTF	APR SR RIBSP
LR TQKHT	0.193 *								
LR TRTHT	0.207 *	0.283 **							
APR LR RIBSP	0.008 <sup>ns</sup>	0.122 <sup>ns</sup>	0.154 <sup>ns</sup>						
APR LR KRIAPI	0.251 **	−0.028 <sup>ns</sup>	0.192 *	−0.124 <sup>ns</sup>					
SR TKRTF	−0.152 <sup>ns</sup>	0.000 <sup>ns</sup>	0.141 <sup>ns</sup>	0.088 <sup>ns</sup>	0.228 *				
SR PKCTC	−0.150 <sup>ns</sup>	0.021 <sup>ns</sup>	0.058 <sup>ns</sup>	0.081 <sup>ns</sup>	−0.009 <sup>ns</sup>	−0.019 <sup>ns</sup>			
SR RKRTF	−0.202 *	−0.006 <sup>ns</sup>	0.091 <sup>ns</sup>	−0.125 <sup>ns</sup>	0.069 <sup>ns</sup>	0.050 <sup>ns</sup>	0.110 <sup>ns</sup>		
APR SR RIBSP	−0.071 <sup>ns</sup>	0.096 <sup>ns</sup>	0.197 *	0.130 <sup>ns</sup>	0.038 <sup>ns</sup>	−0.075 <sup>ns</sup>	−0.040 <sup>ns</sup>	−0.029 <sup>ns</sup>	
APR SR KRIAPI	0.127 <sup>ns</sup>	0.040 <sup>ns</sup>	0.117 <sup>ns</sup>	0.018 <sup>ns</sup>	0.245 **	−0.006 <sup>ns</sup>	0.182 *	0.015 <sup>ns</sup>	−0.043 <sup>ns</sup>

APR, adult plant resistance; <sup>ns</sup> not significant; \*  $p < 0.05$ ; \*\*  $p < 0.01$ .

The negative influence of LR and SR severities at the adult plant stage on the wheat YP and TKW was confirmed by significant negative correlations ( $p < 0.01$ ) between these traits at RIBSP (Table 6). At KRIAPI, the severity of LR at the adult plant stage was negatively correlated with YP only.

**Table 6.** Correlations between leaf rust (LR) and stem rust (SR) resistance at the adult plant stage and yield-related traits in the Pamyati Azieva × Paragon mapping population.

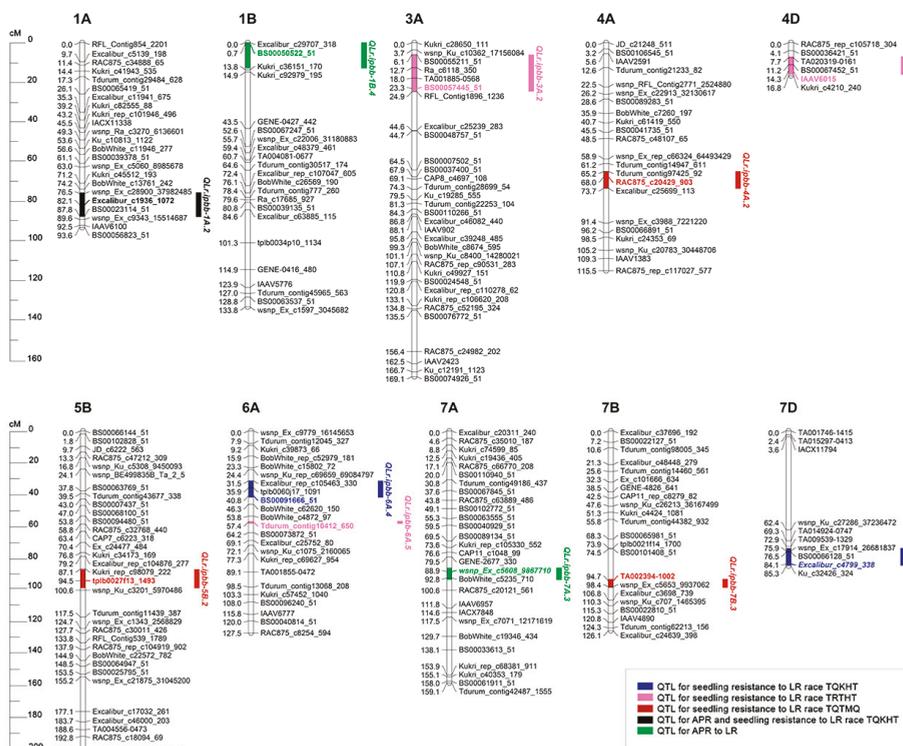
	TKW RIBSP	YP RIBSP	TKW KRIAPI	YP KRIAPI
LR RIBSP	−0.175 **	−0.252 **	–	–
SR RIBSP	−0.490 **	−0.474 **	–	–
LR KRIAPI	–	–	−0.055 <sup>ns</sup>	−0.200 *
SR KRIAPI	–	–	0.134 <sup>ns</sup>	−0.151 <sup>ns</sup>

TKW, thousand kernel weight (g); YP, yield per plot (g/m<sup>2</sup>); <sup>ns</sup> not significant; \*  $p < 0.05$ ; \*\*  $p < 0.01$ .

### 3.3. Identification of QTLs for Seedling and Adult Plant Resistance to LR in the RIL Population

A total of 11 QTLs for resistance to LR were identified at the seedling and adult plant-growth stages. Out of these 11 QTLs, eight QTLs were detected for different LR races at the seedling stage, two QTLs were for APR, and one QTL was observed for both seedling and adult plant resistance (Table 7, Figure 1). QTLs for LR resistance were located on 10 chromosomes of the A, B, and D genomes. The phenotypic variations explained by an individual QTL ranged from 11.6% to 25.7%. Because all QTLs for LR resistance identified in this study explained more than 10% of the phenotypic variation, they were considered major QTLs [50]. The LOD score of QTLs for LR resistance was in the range of 3.2–8.6.

For the LR race TQTMQ, three QTLs identified on chromosomes 4A, 5B, and 7B were revealed. They explained 12.1–15.7% of the phenotypic variations. The alleles of all three QTLs associated with the increase in resistance to LR originated from PA. For the second LR race TQKHT, two QTLs on chromosomes 6A and 7D explained 25.7% and 11.6% of the variations in phenotype, respectively. Both QTLs associated with higher resistance to race TQKHT originated from Paragon. For the third LR race TRHT, three QTLs on chromosomes 3A, 4D, and 6A were observed. Identified QTLs explained 12.3–16.2% of the variation in resistance to race TQKHT. The alleles of QTLs *Qlr.ipbb-3A.2* and *Qlr.ipbb-6A.5* increasing resistance were from Paragon, and the allele of *Qlr.ipbb-4D.1* was from PA.



**Figure 1.** Pamyati Azieva × Paragon genetic map with quantitative trait loci (QTLs) for adult plant resistance (APR) to leaf rust (LR) in two regions and seedling resistance to three LR races. The region containing the QTL is indicated by a vertical bar on the right and followed by the name of the QTL. Single nucleotide polymorphism (SNP) markers are shown on the right and their genetic positions (cM) on the left. The peak marker for each QTL is highlighted in color and bolded. Colors of QTL indicate APR or race-specific seedling resistance.

**Table 7.** The list of quantitative trait loci (QTLs) for adult plant resistance (APR) and race-specific seedling resistance to leaf rust (LR) identified in the Pamyati Azieva × Paragon recombinant inbred lines (RILs) population.

Race	Env.	QTL	Flanking Markers (Left–Right)	Chr.	CI (cM)	Peak (cM)	Max. LOD	R <sup>2</sup> (%)	Add. Effect <sup>1</sup>	Allele <sup>1</sup>
TQTMQ	GH	<i>QL:ipbb-4A.2</i>	IAAV7104–Excalibur_c25699_113	4A	65.2–73.9	68.0	3.3	12.1	−0.47	PA
	GH	<i>QL:ipbb-5B.2</i>	Kukri_rep_c98079_222–BS00075815_51	5B	87.4–99.2	94.5	4.3	15.2	−0.57	PA
	GH	<i>QL:ipbb-7B.3</i>	BS00023023_51–wsnp_Ex_c5653_9937062	7B	93.5–98.9	94.7	4.9	15.7	−0.79	PA
TQKHT	GH	<i>QL:ipbb-6A.4</i>	Excalibur_rep_c105463_330–Ku_c37893_495	6A	31.5–42.0	40.8	8.6	25.7	0.63	Paragon
	GH	<i>QL:ipbb-7D.1</i>	Kukri_c16416_647–BS00062644_51	7D	74.0–84.9	84.1	4.2	11.6	0.41	Paragon
TRTHI	GH	<i>QL:ipbb-3A.2</i>	Excalibur_c74666_291–RFL_Contig1896_1236	3A	6.2–24.7	23.3	5.1	16.2	0.46	Paragon
	GH	<i>QL:ipbb-4D.1</i>	TA020319-0161–BS00022436_51	4D	7.1–16.2	14.3	3.7	12.3	−0.39	PA
	GH	<i>QL:ipbb-6A.5</i>	BobWhite_c30930_192–BS00022992_51	6A	56.9–58.5	57.4	4.8	15.1	0.44	Paragon
APR	RIBSP	<i>QL:ipbb-1B.4</i>	Excalibur_c29707_318–Kukri_c36151_170	1B	0–13.0	0.7	3.2	13.6	1.29	Paragon
APR	RIBSP	<i>QL:ipbb-7A.3</i>	Ra_c4601_2417–CAP7_c7296_88	7A	86.4–94.0	88.9	3.4	11.7	1.18	Paragon
TQKHT/APR	GH/KRIAP1	<i>QL:ipbb-1A.2</i>	RAC875_c12348_720–BS00022824_51	1A	75.9–88.2	82.1	4.6	16.0	0.86	Paragon

CI, confidence interval; LOD, logarithm of odds; R<sup>2</sup>, phenotypic variance explained by the QTL; Chr., chromosome; Add., additive effect. <sup>1</sup>—Additive effect of QTL indicates increasing of the trait expression explained by the allele of one parent (positive for PA and negative for Paragon). In the case of disease resistance, increased expression of the trait is undesired, and the effective allele is taken from the other parent (negative for PA and positive for Paragon).

Two QTLs for APR to LR at RIBSP were identified on chromosomes 1B and 7A and explained 13.6% and 11.7% of phenotypic variation, respectively. For both APR QTLs, alleles associated with increased LR resistance were from Paragon. One QTL for APR to LR at KRIAPI was also detected at the seedling stage for the resistance to LR race TQKHT on chromosome 1A. It explained 16.0% of LR resistance variation. In both cases, alleles associated with higher resistance originated from Paragon.

### 3.4. QTLs for SR Resistance at Seedling and Adult Plant Stages Identified in PA × Par Mapping Population

A total of 13 QTLs were detected in this study for SR resistance at the seedling and adult plant-growth stages. Among them, seven race-specific QTLs were identified at the seedling stage (three QTLs for race TKRTF, three QTLs for race PKCTC, and one QTL for race RKRTF), three QTLs were observed for APR (two QTLs at KRIAPI and one QTL at RIBSP), and three QTLs were revealed in both the seedling and adult stages (Table 8, Figure 2). The identified QTLs for SR resistance were distributed among nine chromosomes of A, B, and D genomes and explained from 8.9% to 39.1% of the variation in the resistance to SR. In total, 11 out of 13 QTLs for SR resistance had  $R^2 > 10\%$  and could be considered major QTLs. The LOD score for the detected QTL varied from 3.0 to 6.8.



**Figure 2.** Pamyati Azieva × Paragon genetic map with quantitative trait loci (QTLs) for adult plant resistance (APR) to stem rust (SR) in two regions and seedling resistance to three SR races. The region containing the QTL is indicated by a vertical bar on the right and followed by the name of the QTL. Single nucleotide polymorphism (SNP) markers are positioned on the right and their genetic positions (cM) are shown on the left. The peak marker for each QTL is highlighted in color and bold. Colors of QTL indicate APR or race-specific seedling resistance.

**Table 8.** The list of quantitative trait loci (QTLs) for APR and race-specific seedling resistance to stem rust (SR) identified in Pamyati Azieva × Paragon recombinant inbred lines (RILs) population.

Race	Env.	QTL	Flanking Markers (Left-Right)	Chr.	CI (cM)	Peak (cM)	Max. LOD	R <sup>2</sup> (%)	Add. Effect *	Allele *
TKRTF	GH	<i>QSt.ipbb-2D.2</i>	w SNP_Ex_rep_c80588_75758453-TA010191-1338	2D	0–3.7	0.1	3.0	10.0	0.43	Paragon
	GH	<i>QSt.ipbb-6B.6</i>	w SNP_Ex_c12618_20079758-w SNP_Ex_c1249_2399894	6B	46.0–54.3	47.8	4.2	14.1	–0.57	PA
	GH	<i>QSt.ipbb-6B.7</i>	w SNP_Ku_c43368_50890819-w SNP_Ku_c4910_8793327	6B	75.5–87.3	79.9	4.7	19.2	–0.66	PA
PKCTC	GH	<i>QSt.ipbb-2B.3</i>	RAC875_rep_c71112_400-Kukri_c1175_1577	2B	6.3–10.4	9.2	3.9	12.2	–0.63	PA
	GH	<i>QSt.ipbb-2B.4</i>	BS00041323_51-RAC875_c32503_134	2B	75.9–78.3	78.0	2.9	8.9	0.54	Paragon
	GH	<i>QSt.ipbb-5B.1</i>	RAC875_rep_c114200_428-w SNP_RFL_Contig1548_762547	5B	137.6–146.3	141.1	4.9	14.7	0.96	Paragon
RKRTF	GH	<i>QSt.ipbb-5A.3</i>	BS00036851_51-Excalibur_c27357_146	5A	184.7–194.4	190.3	6.8	24.7	0.72	Paragon
	KRIAPI	<i>QSt.ipbb-ID.1</i>	BS00051826_51-BobWhite_rep_c65565_359	1D	61.7–65.5	64.3	5.0	18.0	2.24	Paragon
APR	RIBSP	<i>QSt.ipbb-3D.1</i>	Ra_c10284_405-Kukri_c5252_107	3D	0–11.4	1.9	3.1	11.3	–0.99	PA
	KRIAPI	<i>QSt.ipbb-5A.2</i>	BS00089076_51-CAP11_c2623_196	5A	168.2–178.5	172.7	3.0	10.8	0.72	Paragon
TKRTF/APR	GH/KRIAPI	<i>QSt.ipbb-1B.4</i>	Excalibur_c29707_318-Kukri_c92979_195	1B	0–15.2	12.6	4.2	15.7	0.59/–0.77	Paragon/PA
RKRTF/APR	GH/RIBSP	<i>QSt.ipbb-2A.2</i>	Kukri_c33374_1048-Tdurum_contig42153_5854	2A	0.5–37.2	30.8	5.2	39.1	–1.45	PA
TKRTF/APR	GH/RIBSP	<i>QSt.ipbb-6B.5</i>	RAC875_rep_c105224_352-Kukri_rep_c107077_360	6B	2.7–15.1	6.7	4.6	17.3	–0.56/1.26	PA/Paragon

\* Additive effect of QTL indicates increasing of the trait expression explained by the allele of one parent (positive for PA and negative for Paragon). In the case of disease resistance, increased expression of the trait is undesired, and the effective allele is taken from the other parent (negative for PA and positive for Paragon).

Three QTLs for resistance to SR race TKRTF were identified, including two QTLs mapped on chromosome 6B and one QTL on 2D. These QTLs explained 10.0–19.2% of the variation in SR resistance to race TKRTF. QTLs *Q<sub>Sr</sub>.ipbb-6B.6* and *Q<sub>Sr</sub>.ipbb-6B.7* had alleles increasing SR resistance to race TKRTF carried by PA, whereas resistance allele *Q<sub>Sr</sub>.ipbb-2D.2* originated from Paragon. For the second SR race PKCTC, two QTLs for resistance to this race were located on chromosome 2B and one QTL was on 5B. The phenotypic variance conditioned by these QTLs varied from 8.9% to 14.7%. The third SR race RKRTF allowed the identification of one race-specific QTL on chromosome 5A explaining 24.7% of the phenotypic variation. Its allele, associated with an increase in SR resistance, originated from Paragon.

Three QTLs for APR to SR were identified on chromosomes 1D, 3D, and 5A, and explained from 10.8% to 18.0% of SR resistance variation. Two QTLs identified at KRIAPI had alleles increasing resistance to SR originating from Paragon, and the allele of QTL at RIBSP was from PA. The last three QTLs for LR resistance occurred multiple times in the experiment and are located on chromosomes 1B, 2A, and 6B. The QTL *Q<sub>Sr</sub>.ipbb-1B.4* was detected as race-specific to TKRTF at the seedling stage and as APR QTL at KRIAPI. It explained 15.7% of SR resistance variation and had alleles increasing resistance originating from Paragon at the seedling stage and PA at the adult plant stage. The QTL *Q<sub>Sr</sub>.ipbb-2A.2* was identified as effective against SR race RKRTF at the seedling stage and as APR QTL at RIBSP. This QTL explained 39.1% of the phenotypic variation, and alleles increasing SR resistance at both growth stages were inherited from PA. The QTL *Q<sub>Sr</sub>.ipbb-6B.5* was discovered at the seedling stage to race TKRTF and at the adult plant stage at RIBSP. The QTL explained 17.3% of SR resistance variations. Its alleles increasing resistance originated from PA in the case of seedling resistance and from Paragon at the adult growth stage.

### 3.5. Comparison of Identified QTLs with Previous Works and Gene Identification

The QTLs identified in this study were analyzed in comparison with previously reported QTLs for LR and SR resistance in the PA × Par RILs population [27] and with QTLs for LR and SR resistance at RIBSP identified using genome-wide association study (GWAS) [51]. The location of each identified QTL was compared to the genetic positions of known *Lr* and *Sr* genes (Table 9). In total, four candidate *Lr* genes and four QTLs were found for five QTLs associated with LR resistance in this study. In the analysis of QTLs for SR resistance, we found similarities with the genetic locations of eight previously identified QTLs and/or candidate *Sr* genes.

**Table 9.** Comparison of quantitative trait loci (QTLs) for leaf rust (LR) and stem rust (SR) resistance identified in this study in a Pamyati Azieva × Paragon mapping population with previously described QTLs and candidate *Lr* and *Sr* genes.

#	Trait	Type of Resistance (Race)	QTL	Reference QTL	Candidate Genes
1	LR	TQKHT/APR	<i>Q<sub>Lr</sub>.ipbb-1A.2</i>	<i>Q<sub>Lr</sub>.ipbb-1A.1</i> [51]	-
2		APR	<i>Q<sub>Lr</sub>.ipbb-1B.4</i>	-	<i>Lr21</i> [47]
3		TRTHT	<i>Q<sub>Lr</sub>.ipbb-3A.2</i>	-	<i>Lr63, Lr66</i> [52]
4		TQKHT	<i>Q<sub>Lr</sub>.ipbb-6A.4</i>	<i>Q<sub>Lr</sub>.ipbb-6A.1</i> [51]	-
5		TRTHT	<i>Q<sub>Lr</sub>.ipbb-6A.5</i>	<i>Q<sub>Lr</sub>.ipbb-6A.2</i> [51]	-
6		TQTMQ	<i>Q<sub>Lr</sub>.ipbb-7B.3</i>	<i>Q<sub>Lr</sub>.ipbb-7B.1</i> [51]	<i>Lr14</i> [52]
1	SR	TKRTF/APR	<i>Q<sub>Sr</sub>.ipbb-1B.4</i>	-	<i>Sr31</i> [53]
2		APR	<i>Q<sub>Sr</sub>.ipbb-1D.1</i>	-	<i>Sr18</i> [54]
3		RKRTF/APR	<i>Q<sub>Sr</sub>.ipbb-2A.2</i>	-	<i>Sr32-2A</i> [53]
4		PKCTC	<i>Q<sub>Sr</sub>.ipbb-2B.3</i>	-	<i>Sr32-2B, Sr39</i> [53]
5		PKCTC	<i>Q<sub>Sr</sub>.ipbb-2B.4</i>	<i>Q<sub>SR</sub>.IPBB-2B</i> [26]	<i>Sr36</i> [53]
6		TKRTF	<i>Q<sub>Sr</sub>.ipbb-2D.2</i>	-	<i>Sr32-2D</i> [53]
7		TKRTF/APR	<i>Q<sub>Sr</sub>.ipbb-6B.5</i>	<i>Q<sub>SR</sub>.IPBB-6B.1</i> [27], <i>Q<sub>Sr</sub>.ipbb-6B.3</i> [51]	-
8		TKRTF	<i>Q<sub>Sr</sub>.ipbb-6B.7</i>	<i>Q<sub>SR</sub>.IPBB-6B.2</i> [27], <i>Q<sub>Sr</sub>.ipbb-6B.4</i> [51]	<i>Sr11</i> [54]

The region of each QTL was analyzed for the presence of protein-coding genes in the interval 500 kb upstream and 500 kb downstream from the most significant SNP (Table S1). The analysis of LR-associated QTL regions suggested the presence of 158 genes ranging from 6 (*QLr.ipbb-7B.3* and *QLr.ipbb-7D.1*) to 22 (*QLr.ipbb-1B.4* and *QLr.ipbb-6A.4*) genes per interval. A similar search for SR-associated QTL regions indicated the presence of 226 genes ranging from 5 (*Qsr.ipbb-2B.4*) to 29 (*Qsr.ipbb-1B.4*) genes per interval. Among these 158 genes identified for QTLs associated with LR, 48.9% coded for proteins with functions known in *T. aestivum*, 48.6% for uncharacterized proteins, and 2.5% for RNAs. For QTLs associated with SR, 56.6% of genes coded for proteins uncharacterized in *T. aestivum*, 41.2% described protein-coding genes, and 2.2% coded for RNAs. Among genes coding for uncharacterized proteins, sequences similar to the 24 QTL regions for LR and 39 QTL regions for SR were identified in other grass species (Table S1). Orthologous genes with their sequence similarity level higher than 70% were selected and are listed.

## 4. Discussion

### 4.1. General Resistance of RILs in Studied Environments

At the seedling stage, the majority of RILs and parental cultivars showed MS and S levels of resistance to all races of LR and SR, except for the SR race PKCTC, where several lines were identified as R and MR (Table 3). The ANOVA test showed a more significant influence of pathogen genotype (race) on the resistance of RILs rather than the genotype of wheat lines (Table 4). This result indicated that these genetic factors associated with resistance are race-specific. In the world and in Kazakhstan, breeding programs are mostly focused on the combination of seedling resistance and APR in new cultivars. Pyramiding of seedling gene(s) with slow rusting APR gene(s) usually results in higher resistance of the crop. This agrees with wheat R genes conferring resistance to LR (*Lr1*, *Lr10*, *Lr21*) and SR (*Sr22*, *Sr33*, *Sr35*, *Sr45*, *Sr50*) being cloned and widely used in wheat breeding [55]. However, the significant positive correlations among LR races observed in this study (Table 5) suggested the involvement of genetic factors that are effective against all three races. The presence of strong positive correlations between APR to LR and SR at KRIAPI also indicated that genes conferring LR resistance are either closely linked or may have a pleiotropic effect on genes that control SR resistance [26,56]. Positive correlations were simultaneously observed between seedling resistance to LR races TQTHQ and TKTHT and APR to LR at KRIAPI, as well as between seedling resistance to SR race PKCTC and APR to SR at KRIAPI (Table 5). The relationship between race-specific seedling and broad adult plant resistances could be influenced by the presence of LR and SR races in the fields at KRIAPI. This also suggested that the wheat germplasm growing in this region could be effectively and rapidly screened for resistance to LR and SR at the seedling stage in a greenhouse [57].

LR and SR resistances are complex traits [58]; this was confirmed by the range of reactions to pathogens and the presence of transgressive segregations. Even when parents demonstrated the same level of resistance, such as APR to SR, RILs still showed transgressive phenotypes in the direction of either resistance (RIBSP) or susceptibility (KRIAPI) (Table 3). This phenomenon is not rare; it was previously described for many other quantitatively inherited wheat traits; for example, in studies of grain quality traits [22], grain Zn and Fe concentrations [59], grain yield and plant height [60], and rust diseases [61,62].

### 4.2. QTL Mapping for Leaf Rust Resistance

Alleles conferring increased resistance of QTLs for LR race TQKHT and APR at RIBSP originated from Par (Table 7). The higher LR resistance of Par in comparison with PA indicated that the U.K. cultivar is a promising source for wheat breeding programs in Kazakhstan. PA was simultaneously found to be a source for QTLs with increased LR resistance to race TQTMQ.

The 11 QTLs for the resistance to LR at the seedling and adult plant-growth stages can be divided into two categories: (1) similar to QTLs previously detected for LR resistance and (2) presumably

novel QTLs. The first category consisted of 6 out of 11 QTLs for LR resistance (Table 9). Four of the QTLs for LR resistance with similar genetic positions (*QLr.ipbb-1A.2* (APR at KRIAPI), *QLr.ipbb-6A.4* (seedling resistance to TQKHT), *QLr.ipbb-6A.5* (seedling resistance to TRTHT), and *QLr.ipbb-7B.3* (seedling resistance to TQTMQ) were previously identified in a GWAS study performed at RIBSP in 2018/2019 at the adult plant stage [50]. Hence, multiple occurrences of QTLs associated with the resistance to LR in different conditions and environments indicated the broad stability of these loci. *QLr.ipbb-7B.3* may be associated with the gene *Lr14* located in a similar region of the genome (Table 9). The effectiveness of allele *Lr14a* was described for northern Kazakhstan and *Lr14b* for eastern and western Kazakhstan [7]. *Lr14* was also described as an effective resistance factor to TQTMQ (Table 1). The APR QTL *QLr.ipbb-1B.4* is associated with the gene *Lr21*, positioned in close proximity to the peak of the QTL (Table S1). This gene was described as effective in southeastern Kazakhstan [7]. The last QTL from the first group, *QLr.ipbb-3A.2*, is probably associated with genes *Lr63* and *Lr66* (Table 9). Unfortunately, information is lacking about the role of these genes in the wheat-growing areas of Kazakhstan. However, *Lr63* and *Lr66* are known to condition low to intermediate infection types to most of *P. recondita* isolates [63]. The remaining five QTLs identified for LR resistance are presumably novel genetic factors, since there were no reliable matches between their positions in the genome and previously identified QTLs or genes.

#### 4.3. QTLs for Stem Rust

In 13 QTLs for the resistance to SR identified in this study, alleles presumably increasing resistance originated from both PA and Par (Table 8). Similar to LR resistance, SR-resistance-associated QTLs could be divided into two loci groups, where the first group has similar genetic positions with previously reported QTLs for SR resistance (Table 9), and the second group has none of those matches. The first group includes 8 out of 13 QTLs identified for SR resistance. For three of them—*Qsr.ipbb-2B.4* (seedling resistance to PKCTC), *Qsr.ipbb-6B.5* (seedling resistance to TKRTF and APR in RIBSP), and *Qsr.ipbb-6B.7* (seedling resistance to TKRTF)—QTLs for SR resistance with similar positions in the genome were identified in a previous work involving the PA × Par mapping population [27] and in a GWAS study using resistance data obtained from RIBSP [51]. Similar to the LR study, these findings may indicate the stability of identified QTLs. In addition to the information with QTL similarities, several specific *Sr* genes seem to be associated with QTLs from this study (Table 9). One of the most interesting findings was the identification of three QTLs on distal ends of chromosomes 2A, 2B, and 2D responsible for seedling resistance to SR races RKRTF, PKCTC, and TKRTF, respectively. These QTLs could be associated with the gene *Sr32*, which was mapped in these regions of chromosomes 2A [64], 2B [65], and 2D [66]. The gene was previously reported as effective against Ug99 and related SR races [66]. The other *Sr* genes involved in resistance to SR races in the Ug99 lineage and possibly associated with QTL from this study are *Sr31* (resistant to TTKSF and TTKSP), *Sr36* (all Ug99 lineage races, except TTTSK), and *Sr39* (all Ug99 lineage races) [67]. Among the SR races used in this study, *Sr36* was described as effective against PKCTC (Table 1). The resistance pattern is similar to *Qsr.ipbb-2B.4*, which is located in a nearby region of the chromosome. The second group of the genetic factors consisted of the remaining five QTLs that could be novel QTLs associated with resistance to SR.

#### 4.4. QTLs Cluster on Chromosome 1B

The QTLs associated with several traits are common in wheat. It may occur due to pleiotropic effect or their tight linkage. For the resistance to wheat fungal diseases, pleiotropic APR genes *Lr34/Yr18/Pm38/Sr57* [68], *Lr46/Yr29/Pm39/Sr58* [69], and *Lr67/Yr46/Pm46/Sr55* [70] were previously described. Among the QTLs identified for LR and SR resistance in this study, two QTLs (*QLr.ipbb-1B.4* and *Qsr.ipbb-1B.4*) occupy the same interval on chromosome 1B (Tables 7 and 8, Figures 1 and 2). In addition, the *QLr.ipbb-1B.4* interval contains the resistance gene *Lr21* less than 500 kb from the significant peak, whereas the interval of *Qsr.ipbb-1B.4* has genes for disease resistance proteins and resistance-related kinases next to the peak marker (Table S1). *Lr21* was described as effective

for southeastern Kazakhstan [7]. Common markers in these intervals suggest the usefulness for marker-assisted breeding of these QTLs to develop wheat cultivars with durable rust resistance for gene pyramiding [11].

## 5. Conclusions

Overall, 24 QTLs for the resistance to rust diseases at the seedling and adult plant stages were identified in this study, including 11 QTLs for LR and 13 QTLs for SR. Among the QTLs associated with LR, eight QTLs were race-specific and detected at the seedling stage, two QTLs were at the stage of the adult plant, and one QTL was identified in both stages. The QTLs for LR-resistance explained from 11.6% (*Q<sub>Lr.ipbb-7D.1</sub>*) to 25.7% (*Q<sub>Lr.ipbb-6A.4</sub>*) of the phenotypic variation and were detected on 10 chromosomes. The increased resistance to LR in TQTMQ race-specific QTLs originated from PA; in QTLs specific for the race TQKHT and APR, alleles were from Par. For TRTHT, the origin of resistance alleles in identified QTLs was both parental cultivars. For SR resistance, seven QTLs were race-specific and detected at the seedling stage, three QTLs were identified at the adult plant stage, and three QTLs were identified at both growth stages. SR-associated QTLs explained from 8.9% (*Q<sub>Sr.ipbb-2B.4</sub>*) to 39.1% (*Q<sub>Sr.ipbb-2A.2</sub>*) of variation in SR resistance and were mapped on nine chromosomes. The alleles increasing resistance to SR originated from both parents: effective alleles in six QTLs were from Par, in five QTLs from PA, and two QTLs had a different origin of resistance at the seedling and adult plant stages. Among the QTLs from this study, 10 QTLs were putative and 14 matching QTLs were found in previous works involving the PA × Par population, a GWAS study at RIBSP, and possible candidate resistance genes. The cluster of QTLs associated with both LR and SR resistances was identified on chromosome 1B. Thus, the QTLs revealed in this study may play an essential role in the improvement of wheat resistance to LR and SR via marker-assisted selection.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4395/10/9/1285/s1>, Table S1. The list of protein- and RNA-coding genes 500 kb upstream and 500 kb downstream from the most significant SNP of the QTL.

**Author Contributions:** Conceptualization, S.A. and Y.T.; methodology, A.R. and Y.T.; formal analysis, Y.G. and G.Y.; investigation, Y.G., A.R., and S.A.; resources, Y.T. and A.R.; data curation, Y.G., G.Y., and Y.T.; writing—original draft preparation, Y.G.; writing—review and editing, Y.G., S.A., A.R., G.A., and Y.T.; supervision, Y.T.; project administration, S.A.; funding acquisition, A.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Ministry of Agriculture of the Republic of Kazakhstan, grant number BR06249329. The APC was also funded by grant number BR06249329.

**Acknowledgments:** This work was conducted within the framework of the project “Development of new DNA markers associated with the resistance of bread wheat to the most dangerous fungal diseases in Kazakhstan” in Program “Development of the innovative systems for increasing the resistance of wheat varieties to especially dangerous diseases in the Republic of Kazakhstan”: BR06249329 supported by the Ministry of Agriculture of the Republic of Kazakhstan.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Statista. Available online: <https://www.statista.com/> (accessed on 12 March 2020).
2. Giraldo-Carbajo, P.; Barzana, M.E.B.; Manzano-Agugliaro, F.; Giménez, E. Worldwide Research Trends on Wheat and Barley: A Bibliometric Comparative Analysis. *Agronomy* **2019**, *9*, 352. [[CrossRef](#)]
3. USDA. Available online: <https://www.usda.gov/> (accessed on 17 March 2020).
4. Statistics Committee. Ministry of National Economy of the Republic of Kazakhstan. Available online: <https://stat.gov.kz/> (accessed on 15 March 2020).
5. Tadesse, W.; Sanchez-Garcia, M.; Assefa, S.G.; Amri, A.; Bishaw, Z.; Ogonnaya, F.C.; Baum, M. Genetic Gains in Wheat Breeding and Its Role in Feeding the World. *Crop Breed. Genet. Genom.* **2019**, *1*, e190005.
6. Samborski, D.J. Wheat leaf rust. In *The Cereal Rusts Vol. II: Diseases, Distribution, Epidemiology and Control*; Roelfs, A.P., Bushnell, W.R., Eds.; Academic Press: Orlando, FL, USA, 1985; pp. 39–59.

7. Koyshybaev, M. *Wheat Diseases*, 1st ed.; FAO: Ankara, Turkey, 2018; p. 365.
8. Eversmeyer, M.G.; Kramer, C.L. Epidemiology of Wheat Leaf and Stem Rust in the Central Great Plains of the USA. *Annu. Rev. Phytopathol.* **2000**, *38*, 491–513. [[CrossRef](#)] [[PubMed](#)]
9. Summers, R.W.; Brown, J.K.M. Constraints on breeding for disease resistance in commercially competitive wheat cultivars. *Plant Pathol.* **2013**, *62*, 115–121. [[CrossRef](#)]
10. McIntosh, R.A.; Dubcovsky, J.; Rogers, W.J.; Morris, C.F.; Xia, X.C. Catalogue of Gene Symbols for Wheat: 2017 Supplement. Available online: <https://shigen.nig.ac.jp/wheat/komugi/genes/macgene/supplement2017.pdf> (accessed on 27 July 2020).
11. Ellis, J.G.; Lagudah, E.; Spielmeyer, W.; Dodds, P.N. The past, present and future of breeding rust resistant wheat. *Front. Plant Sci.* **2014**, *5*, 641. [[CrossRef](#)]
12. Bariana, H.S.; Hayden, M.J.; Ahmed, N.U.; Bell, J.A.; Sharp, P.J.; McIntosh, R.A. Mapping of durable adult plant and seedling resistances to stripe rust and stem rust diseases in wheat. *Aust. J. Agric. Res.* **2001**, *52*, 1247. [[CrossRef](#)]
13. Xu, Y.; Li, P.; Yang, Z.; Yang, Z. Genetic mapping of quantitative trait loci in crops. *Crop J.* **2017**, *5*, 175–184. [[CrossRef](#)]
14. Arruda, M.P.; Brown, P.; Brown-Guedira, G.; Krill, A.M.; Thurber, C.; Merrill, K.R.; Foresman, B.J.; Kolb, F.L. Genome-Wide Association Mapping of Fusarium Head Blight Resistance in Wheat using Genotyping-by-Sequencing. *Plant Genome* **2016**, *9*, 1–14. [[CrossRef](#)]
15. Li, G.; Xu, X.; Tan, C.; Carver, B.F.; Bai, G.; Wang, X.; Bonman, J.M.; Wu, Y.; Hunger, R.; Cowger, C. Identification of powdery mildew resistance loci in wheat by integrating genome-wide association study (GWAS) and linkage mapping. *Crop J.* **2019**, *7*, 294–306. [[CrossRef](#)]
16. Cheng, B.; Gao, X.; Cao, N.; Ding, Y.; Gao, Y.; Chen, T.; Xin, Z.; Zhang, L.Y. Genome-wide association analysis of stripe rust resistance loci in wheat accessions from southwestern China. *J. Appl. Genet.* **2020**, *61*, 37–50. [[CrossRef](#)]
17. Ollier, M.; Talle, V.; Brisset, A.-L.; Le Bihan, Z.; Duerr, S.; Lemmens, M.; Goudemand, E.; Robert, O.; Hilbert, J.-L.; Buerstmayr, H. QTL mapping and successful introgression of the spring wheat-derived QTL Fhb1 for Fusarium head blight resistance in three European triticale populations. *Theor. Appl. Genet.* **2020**, *133*, 457–477. [[CrossRef](#)] [[PubMed](#)]
18. Unamba, C.I.N.; Nag, A.; Sharma, R.K. Next Generation Sequencing Technologies: The Doorway to the Unexplored Genomics of Non-Model Plants. *Front. Plant Sci.* **2015**, *6*, 6. [[CrossRef](#)]
19. Turuspekov, Y.; Plieske, J.; Ganal, M.; Akhunov, E.; Abugalieva, S. Phylogenetic analysis of wheat cultivars in Kazakhstan based on the wheat 90 K single nucleotide polymorphism array. *Plant Genet. Resour.* **2015**, *15*, 29–35. [[CrossRef](#)]
20. Singh, M.; Upadhyaya, H.D. *Genetic and Genomic Resources for Grain Cereals Improvement*, 1st ed.; Academic Press: Cambridge, MA, USA, 2015; pp. 58–59.
21. Hu, J.; Wang, X.; Zhang, G.; Jiang, P.; Chen, W.; Hao, Y.; Ma, X.; Xu, S.; Jia, J.; Kong, L.; et al. QTL mapping for yield-related traits in wheat based on four RIL populations. *Theor. Appl. Genet.* **2020**, *133*, 917–933. [[CrossRef](#)] [[PubMed](#)]
22. Goel, S.; Singh, K.; Singh, B.; Grewal, S.; Dwivedi, N.; Alqarawi, A.A.; Allah, E.A.; Ahmad, P.; Singh, N.K. Analysis of genetic control and QTL mapping of essential wheat grain quality traits in a recombinant inbred population. *PLoS ONE* **2019**, *14*, e0200669. [[CrossRef](#)] [[PubMed](#)]
23. Li, L.; Mao, X.; Wang, J.; Chang, X.; Reynolds, M.; Jing, R. Genetic dissection of drought and heat-responsive agronomic traits in wheat. *Plant Cell Environ.* **2019**, *42*, 2540–2553. [[CrossRef](#)]
24. Emebiri, L.C.; Tan, M.K.; El-Bouhssini, M.; Wildman, O.; Jighly, A.; Tadesse, W.; Ogbonnaya, F.C. QTL mapping identifies a major locus for resistance in wheat to Sunn pest (*Eurygaster integriceps*) feeding at the vegetative growth stage. *Theor. Appl. Genet.* **2017**, *130*, 309–318. [[CrossRef](#)] [[PubMed](#)]
25. Amalova, A.Y.; Yermekbayev, K.A.; Griffiths, S.; Abugalieva, S.I.; Turuspekov, Y.K.; Centre, U.K.J.I. Phenotypic variation of common wheat mapping population Pamyati Azieva x Paragon in south-east of Kazakhstan. *Int. J. Biol. Chem.* **2019**, *12*, 11–17. [[CrossRef](#)]
26. Genievskaya, Y.; Amalova, A.; Ydyrys, A.; Sarbayev, A.; Griffiths, S.; Abugalieva, S.; Turuspekov, Y. Resistance of common wheat (*Triticum aestivum* L.) mapping population Pamyati Azieva x Paragon to leaf and stem rusts in conditions of south-east Kazakhstan. *Eurasian J. Ecol.* **2019**, *61*, 14–23.

27. Genievskaya, Y.; Fedorenko, Y.; Sarbayev, A.; Amalova, A.; Abugalieva, S.; Griffiths, S.; Turuspekov, Y. Identification of QTLs for resistance to leaf and stem rusts in bread wheat (*Triticum aestivum* L.) using a mapping population of 'Pamyati Azieva × Paragon'. *Vavilov J. Genet. Breed.* **2019**, *23*, 887–895. [CrossRef]
28. ADAPTAWHEAT. Available online: <https://www.jic.ac.uk/adaptawheat/> (accessed on 12 April 2020).
29. Yermekbayev, K.; Turuspekov, Y.; Ganal, M.; Plieske, J.; Griffiths, S. Construction and utilization of the hexaploid map Pamyati Azieva × Paragon. In *Proceedings of PlantGen (Plant Genetics, Genomics, Bioinformatics and Biotechnology)*, Almaty, Kazakhstan, 29 May–2 June 2017; Turuspekov, Y., Abugalieva, S., Eds.; Institute of Plant Biology and Biotechnology: Almaty, Kazakhstan, 2017.
30. Chu, C.-G.; Friesen, T.L.; Xu, S.S.; Faris, J.D.; Kolmer, J.A. Identification of novel QTLs for seedling and adult plant leaf rust resistance in a wheat doubled haploid population. *Theor. Appl. Genet.* **2009**, *119*, 263–269. [CrossRef] [PubMed]
31. Pretorius, Z.A.; Jin, Y.; Bender, C.M.; Herselman, L.; Prins, R. Seedling resistance to stem rust race Ug99 and marker analysis for *Sr2*, *Sr24* and *Sr31* in South African wheat cultivars and lines. *Euphytica* **2011**, *186*, 15–23. [CrossRef]
32. Spanic, V.; Rouse, M.N.; Kolmer, J.A.; Anderson, J.A. Leaf and stem seedling rust resistance in wheat cultivars grown in Croatia. *Euphytica* **2014**, *203*, 437–448. [CrossRef]
33. Stakman, E.C.; Stewart, D.M.; Loegering, W.Q. Identification of physiologic races of *Puccinia graminis* var. *tritici*. *US Agric. Res. Serv.* **1962**, *617*, 1–53.
34. Rsaliyev, A.; Rsaliyev, S.S. Principal approaches and achievements in studying race composition of wheat stem rust. *Vavilov J. Genet. Breed.* **2019**, *22*, 967–977. [CrossRef]
35. Roelfs, A.P. An International System of Nomenclature for *Puccinia graminis* f. sp. *tritici*. *Phytopathology* **1988**, *78*, 526. [CrossRef]
36. Long, D.L.; Kolmer, J.A. A North American system of nomenclature for *Puccinia recondita* f. sp. *tritici*. *Phytopathology* **1989**, *79*, 525–529. [CrossRef]
37. Kolmer, J.A.; Ordoñez, M.E. Genetic Differentiation of *Puccinia triticina* Populations in Central Asia and the Caucasus. *Phytopathology* **2007**, *97*, 1141–1149. [CrossRef]
38. Kolmer, J.A.; Kabdulova, M.G.; Mustafina, M.A.; Zhemchuzhina, N.S.; Dubovoy, V. Russian populations of *Puccinia triticina* in distant regions are not differentiated for virulence and molecular genotype. *Plant Pathol.* **2014**, *64*, 328–336. [CrossRef]
39. Schachtel, G.A.; Dinooor, A.; Herrmann, A.; Kosman, E. Comprehensive Evaluation of Virulence and Resistance Data: A New Analysis Tool. *Plant Dis.* **2012**, *96*, 1060–1063. [CrossRef]
40. Mains, E.B.; Jackson, H.S. Physiologic specialization in the leaf rust of wheat *Puccinia triticiana* Erikss. *Phytopathology* **1926**, *16*, 89–120.
41. Peterson, R.F.; Campbell, A.B.; Hannah, A.E. A Diagrammatic Scale for Estimating Rust Intensity on Leaves and Stems of Cereals. *Can. J. Res.* **1948**, *26*, 496–500. [CrossRef]
42. Roelfs, A.; Singh, R.; Saari, E.E. *Rust Diseases of Wheat: Concepts and Methods of Disease Management*; CIMMYT: Mexico City, Mexico, 1992; p. 45.
43. Zhang, D.; Bowden, R.; Bai, G. A method to linearize Stakman infection type ratings for statistical analysis. In *Proceedings of the Borlaug Global Rust Initiative 2011 Technical Workshop*, Saint Paul, MN, USA, 13–16 June 2011.
44. Windows QTL Cartographer V2.5. Available online: <https://brcwebportal.cos.ncsu.edu/qtcart/WQTLCart.htm> (accessed on 9 February 2020).
45. Rong, J.; Feltus, F.A.; Waghmare, V.N.; Pierce, G.J.; Chee, P.W.; Draye, X.; Saranga, Y.; Wright, R.J.; Wilkins, T.A.; May, O.L.; et al. Meta-analysis of Polyploid Cotton QTL Shows Unequal Contributions of Subgenomes to a Complex Network of Genes and Gene Clusters Implicated in Lint Fiber Development. *Genetics* **2007**, *176*, 2577–2588. [CrossRef]
46. Voorrips, R.E. MapChart: Software for the graphical presentation of linkage maps and QTLs. *J. Hered.* **2002**, *93*, 77–78. [CrossRef] [PubMed]
47. EnsemblPlant. *Triticum Aestivum*. Available online: [https://plants.ensembl.org/Triticum\\_aestivum/Info/Index](https://plants.ensembl.org/Triticum_aestivum/Info/Index) (accessed on 27 February 2020).
48. BLAST Tool. *Triticum Aestivum*. Available online: [https://plants.ensembl.org/Triticum\\_aestivum/Tools/Blast](https://plants.ensembl.org/Triticum_aestivum/Tools/Blast) (accessed on 2 March 2020).
49. UniProt. Available online: <https://www.uniprot.org/> (accessed on 2 March 2020).

50. Collard, B.C.; Jahufer, M.Z.Z.; Brouwer, J.B.; Pang, E.C.K. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica* **2005**, *142*, 169–196. [[CrossRef](#)]
51. Genievskaya, Y.; Abugaliyeva, S.; Rsaliyev, A.; Turuspekov, Y. Genome-wide association mapping for resistance to leaf, stem, and yellow rusts of bread wheat in conditions of South Kazakhstan. *PeerJ* **2020**. submitted work under review.
52. Aoun, M.; Breiland, M.; Turner, M.K.; Loladze, A.; Chao, S.; Xu, S.S.; Ammar, K.; Anderson, J.A.; Kolmer, J.A.; Acevedo, M. Genome-Wide Association Mapping of Leaf Rust Response in a Durum Wheat Worldwide Germplasm Collection. *Plant Genome* **2016**, *9*, 1–24. [[CrossRef](#)] [[PubMed](#)]
53. Yu, L.-X.; Barbier, H.; Rouse, M.N.; Singh, S.; Singh, R.P.; Bhavani, S.; Huerta-Espino, J.; Sorrells, M.E. A consensus map for Ug99 stem rust resistance loci in wheat. *Theor. Appl. Genet.* **2014**, *127*, 1561–1581. [[CrossRef](#)]
54. Hart, G.E.; Gale, M.D.; McIntosh, R.A. Linkage maps of *Triticum aestivum* (hexaploid wheat, 2n = 42, genomes A, B & D) and *T. tauschii* (2n = 14, genome D). In *Progress in Genome Mapping of Wheat and Related Species, Proceedings of the 3rd Publ Wkshp Int Triticeae Mapping Initiative, Mexico City, Mexico, 22–26 September 1992*; Hoisington, D., McNab, A., Eds.; CIMMYT: Mexico City, Mexico, 1993; pp. 32–46.
55. Periyannan, S.; Milne, R.J.; Figueroa, M.; Lagudah, E.; Dodds, P.N. An overview of genetic rust resistance: From broad to specific mechanisms. *PLoS Pathog.* **2017**, *13*, e1006380. [[CrossRef](#)]
56. Figlan, S.; Terefe, T.G.; Shimelis, H.; Tsilo, T.J. Adult plant resistance to leaf rust and stem rust of wheat in a newly developed recombinant inbred line population. *S. Afr. J. Plant Soil* **2017**, *35*, 111–119. [[CrossRef](#)]
57. Tadesse, W.; Reents, H.J.; Hsam, S.L.K.; Zeller, F.J. Relationship of seedling and adult plant resistance and evaluation of wheat germplasm against tan spot (*Pyrenophora tritici-repentis*). *Genet. Resour. Crop Evol.* **2010**, *58*, 339–346. [[CrossRef](#)]
58. Niks, R.E.; Qi, X.; Marcel, T.C. Quantitative Resistance to Biotrophic Filamentous Plant Pathogens: Concepts, Misconceptions, and Mechanisms. *Annu. Rev. Phytopathol.* **2015**, *53*, 445–470. [[CrossRef](#)] [[PubMed](#)]
59. Crespo-Herrera, L.A.; Govindan, V.; Stangoulis, J.C.R.; Hao, Y.; Singh, R.P. QTL Mapping of Grain Zn and Fe Concentrations in Two Hexaploid Wheat RIL Populations with Ample Transgressive Segregation. *Front. Plant Sci.* **2017**, *8*, 1800. [[CrossRef](#)]
60. Asif, M.; Yang, R.C.; Navabi, A.; Iqbal, M.; Kamran, A.; Lara, E.P.; Randhawa, H.; Pozniak, C.; Spaner, D. Mapping QTL, selection differentials, and the effect of *Rht-B1* under organic and conventionally managed systems in the Attila × CDC Go spring wheat mapping population. *Crop Sci.* **2015**, *55*, 1129–1142. [[CrossRef](#)]
61. Tian, Z.M.; Zhang, L.P.; Sun, Q.L.; Shan, F.H.; Tian, Z.H. Research review on wheat QTLs. *Inn. Mong. Agric. Sci. Technol.* **2007**, *3*, 68–71.
62. Bokore, F.E.; Knox, R.E.; Cuthbert, R.D.; Pozniak, C.J.; McCallum, B.D.; N'Diaye, A.; Depauw, R.M.; Campbell, H.L.; Munro, C.; Singh, A.; et al. Mapping quantitative trait loci associated with leaf rust resistance in five spring wheat populations using single nucleotide polymorphism markers. *PLoS ONE* **2020**, *15*, e0230855. [[CrossRef](#)]
63. Kolmer, J.A.; Anderson, J.A.; Flor, J.M. Chromosome Location, Linkage with Simple Sequence Repeat Markers, and Leaf Rust Resistance Conditioned by Gene *Lr63* in Wheat. *Crop Sci.* **2010**, *50*, 2392–2395. [[CrossRef](#)]
64. McIntosh, R.A.; Wellings, C.R.; Park, R.F. *Wheat Rusts, An Atlas of Resistance Genes*; CSIRO Publications: Melbourne, Australia, 1995; pp. 93–99.
65. McIntosh, R.A.; Dubcovsky, J.; Rogers, W.J.; Morris, C.F.; Appels, R.; Xia, X.C. Catalogue of gene symbols for wheat: 2011 supplement. *Annu. Wheat Newsl.* **2011**, *57*, 303–321.
66. Mago, R.; Verlin, D.; Zhang, P.; Bansal, U.; Bariana, H.; Jin, Y.; Ellis, J.; Hoxha, S.; Dundas, I. Development of wheat-*Aegilops speltoides* recombinants and simple PCR-based markers for *Sr32* and a new stem rust resistance gene on the 2S#1 chromosome. *Theor. Appl. Genet.* **2013**, *126*, 2943–2955. [[CrossRef](#)]
67. Singh, R. Current status, likely migration and strategies to mitigate the threat to wheat production from race Ug99 (TTKS) of stem rust pathogen. *CAB Rev. Perspect. Agric. Veter-Sci. Nutr. Resour.* **2006**, *1*, 1–13. [[CrossRef](#)]
68. Singh, R.P.; Herrera-Foessel, S.A.; Huerta-Espino, J.; Bariana, H.; Bansal, U.; McCallum, B.; Hiebert, C.; Bhavani, S.; Singh, S.; Lan, C.X.; et al. *Lr34/Yr18/Sr57/Pm38/Bdv1/Ltn1* confers slow rusting, adult plant resistance to *Puccinia graminis tritici*. In Proceedings of the 13th International Cereal Rusts and Powdery Mildews Conference, Beijing, China, 28 August–1 September 2012.

69. Singh, R.P.; Herrera-Foessel, S.A.; Huerta-Espino, J.; Lan, C.X.; Basnet, B.R.; Bhavani, S.; Lagudah, E.S. Pleiotropic gene *Lr46/Yr29/Pm39/Ltn2* confers slow rusting, adult plant resistance to wheat stem rust fungus. In Proceedings of the Borlaug Global Rust Initiative, New Delhi, India, 19–22 August 2013.
70. Herrera-Foessel, S.A.; Singh, R.P.; Lillemo, M.; Huerta-Espino, J.; Bhavani, S.; Singh, S.; Lan, C.; Calvo-Salazar, V.; Lagudah, E.S. *Lr67/Yr46* confers adult plant resistance to stem rust and powdery mildew in wheat. *Theor. Appl. Genet.* **2014**, *127*, 781–789. [[CrossRef](#)] [[PubMed](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



Article

# Cost-Effective and Time-Efficient Molecular Assisted Selection for Ppv Resistance in Apricot Based on *ParPMC2* Allele-Specific PCR

Ángela Polo-Oltra <sup>1</sup>, Carlos Romero <sup>2</sup>, Inmaculada López <sup>1</sup>, María Luisa Badenes <sup>1</sup>

<sup>1</sup> Citriculture and Plant Production Center, Instituto Valenciano de Investigaciones Agrarias (IVIA), CV-315, km 10.7, 46113 Moncada, Valencia, Spain; polo\_ang@externos.gva.es (Á.P.-O.); lopez\_inmcap@gva.es (I.L.); badenes\_mlu@gva.es (M.L.B.)

<sup>2</sup> Instituto de Biología Molecular y Celular de Plantas (IBMCP), Consejo Superior de Investigaciones Científicas (CSIC)—Universidad Politécnica de Valencia (UPV), Ingeniero Fausto Elio s/n, 46022 Valencia, Spain; cromero@ibmcp.upv.es

\* Correspondence: garcia\_zur@gva.es; Tel.: +34-963424040

Received: 31 July 2020; Accepted: 25 August 2020; Published: 31 August 2020

**Abstract:** Plum pox virus (PPV) is the most important limiting factor for apricot (*Prunus armeniaca* L.) production worldwide, and development of resistant cultivars has been proven to be the best solution in the long-term. However, just like in other woody species, apricot breeding is highly time and space demanding, and this is particularly true for PPV resistance phenotyping. Therefore, marker-assisted selection (MAS) may be very helpful to speed up breeding programs. Tightly linked *ParPMC1* and *ParPMC2*, meprin and TRAF-C homology (MATH)-domain-containing genes have been proposed as host susceptibility genes required for PPV infection. Contribution of additional genes to PPV resistance cannot be discarded, but all available studies undoubtedly show a strong correlation between *ParPMC2*-resistant alleles (*ParPMC2res*) and PPV resistance. The *ParPMC2res* allele was shown to carry a 5-bp deletion (*ParPMC2-del*) within the second exon that has been characterized as a molecular marker suitable for MAS (PMC2). Based on this finding, we propose here a method for PPV resistance selection in apricot by combining high-throughput DNA extraction of 384 samples in 2 working days and the allele-specific genotyping of PMC2 on agarose gel. Moreover, the PMC2 genotype has been determined by PCR or by using whole-genome sequences (WGS) in 175 apricot accessions. These results were complemented with phenotypic and/or genotypic data available in the literature to reach a total of 325 apricot accessions. As a whole, we conclude that this is a time-efficient, cost-effective and straightforward method for PPV resistance screening that can be highly useful for apricot breeding programs.

**Keywords:** apricot; MAS; breeding; MATH; PPV resistance; agarose; *ParPMC*; *ParPMC2-del*

## 1. Introduction

Most cultivated apricots belong to the *Prunus armeniaca* L. species, a member of the Rosaceae family, *Prunus* genus and section *Armeniaca* (Lam.) Koch [1]. World apricot production reached 3.84 million tonnes in 2018, with Turkey, Uzbekistan and Iran as the main producers (<http://www.fao.org/faostat/>). This means an increase of about 45% since 1998 mainly due to Asian countries. By contrast, European production in this period has just increased slightly while the cultivated area declined up to 19%. Despite its wide geographical spread, apricot has very specific ecological requirements. Consequently, each region usually grows locally adapted cultivars. For this reason, significant breeding efforts have been undertaken since the first apricot breeding program started in 1925 at the Nikita Botanical Garden in Yalta (Crimea, Ukraine) [2]. However, apricot breeding based on biparental controlled crosses

and subsequent selection of the best new allelic combinations is hardly limited by the capacity to evaluate trees in the field [3]. On one side, fruit trees show high space requirements to be grown. On the other, their juvenile phase is quite long and reliable pomological phenotyping requires several cropping seasons, which means that at least ten years are needed to release a new variety. Therefore, the implementation of marker-assisted selection (MAS) has a great potential to improve breeding efficiency in fruit trees, including apricot.

Sharka disease, caused by *Plum pox virus* (PPV), is currently the most important viral disease affecting stone fruit trees (*Prunus* spp.) [4]. To date, nine PPV strains (D, M, C, EA, W, Rec, T, CR and An) are identified [5]. However, PPV genetic diversity may be even bigger, as observed by Chirkov et al. [6], who recently described the new Tat isolates affecting sour cherry (*Prunus cerasus*). PPV-D and M are the most widespread and economically important strains [5,7]. A clear host preference is observed: PPV-D/plum/apricot and PPV-M/peach. However, underlying genetic determinants are still unknown [8].

Particularly in apricot, PPV-D has severely hindered production in the last three decades, especially in endemic areas. In this context, development of PPV-resistant varieties is the main objective of apricot breeding programs. However, resistant sources are scarce. Just a handful of North American PPV-resistant cultivars have been identified to date, and they are commonly used as donors in all apricot resistance breeding programs currently in progress [9]. Several independent works aimed at dissecting the genetic control of PPV resistance in apricot have identified the major dominant *PPVres* locus in the upper part of linkage group 1 [10–17]. According to the pedigree and fine mapping data, a single common ancestor carrying *PPVres* has been suggested for all PPV-resistant cultivars [16,18–20]. Moreover, other minor loci contributing to PPV resistance have been suggested [13–16], but their role has not yet been well defined. More recently, transcriptomic and genomic analyses of *PPVres* locus have pointed out *ParPMC1* and *ParPMC2*, two members of a cluster of meprin and TRAF-C homology domain (MATHd)-containing genes, as host susceptibility paralogous genes required for PPV infection [21]. The *ParPMC2* allele linked in coupling with PPV resistance (*ParPMC2res*) accumulates 15 variants, including a 5 nt deletion (*ParPMC2-del*) that results in a premature stop codon. Moreover, cultivars carrying the *ParPMC2res* allele show that *ParPMC2* and especially *ParPMC1* genes are downregulated. As a result, this *ParPMC2res* was proposed to be a pseudogene that confers PPV resistance by silencing functional homologs, the non-mutated *ParPMC2* allele and/or *ParPMC1*. Another plausible scenario involves epigenetic modifications to explain *ParPMC* silencing in the resistant cultivars [22].

In spite of evidence supporting linkage with the *PPVres* locus, some genotype-phenotype incongruencies (GPIs) have been detected in biparental populations segregating for PPV resistance [17,23,24]. In other words, some phenotypically susceptible individuals carrying *ParPMC2res* were classified as genetically resistant. Possible causes underlying these discrepancies, including other loci contributing to PPV resistance, are still unresolved. However, the potential benefit of using a *ParPMC2* allele-specific marker (PMC2) for MAS is still very high since sharka resistance phenotyping is a major bottleneck in apricot breeding programs. The most reliable method for apricot PPV resistance phenotyping is based on a biological test that uses GF-305 peach rootstocks as woody indicators and graft-inoculation with PPV [25]. This procedure is time-consuming and requires visual inspection during two to four growing seasons in several replicates per genotype followed by ELISA [26] and RT-PCR tests [27]. It should be noted that the plant to be tested must be of a significant size in order to have enough buds for grafting replicates, so it takes a couple of years from the time of crossing. As a result of a genetic mapping approach, Soriano et al. [18] reported the first successful MAS application for PPV resistance using 3 SSRs within the *PPVres* locus resolved by capillary electrophoresis. Afterwards, these SSRs were combined with a single sequence length polymorphism marker (ZP002) interrogating the *ParPMC2-del* resolved by capillary or acrylamide electrophoresis [24] and by high resolution melting [28]. However, specialized DNA testing services are needed to adopt these MAS approaches, and together with the economic costs, this could be a challenge [29].

Here, we report a method combining high-throughput DNA extraction of 384 samples in 2 days and PMC2 genotyping by allele-specific PCR amplification and agarose gel electrophoresis. This method is proven to be an easily implemented tool for MAS of PPV-resistant seedlings in almost any apricot breeding program. Therefore, bioassays for PPV resistance evaluation will be needed to confirm the phenotype in selected materials. Moreover, PMC2 genotype has been determined and/or revised for 325 worldwide cultivated apricot accessions providing useful information for breeders to select parental genotypes.

## **2. Materials and Methods**

### *2.1. High-Throughput DNA Isolation in 96-Well Plate*

The genomic DNA extraction protocol was optimized from the original Doyle and Doyle method [30] to manage 384 samples per isolation using 8-well 1.2-mL strip tubes (VWR International). For each accession, 2 leaf discs were collected and placed into a tube with 3 glass beads (VWR International). The strips were frozen in liquid N<sub>2</sub> and stored at −20 °C before DNA isolation. Frozen tissue was ground for 1 min with a frequency of 26/s using a Qiagen TissueLyser 85210 (Qiagen, Hilden, Germany). Then, 340 µL of preheated CTAB isolation buffer (with 0.2% 2-mercaptoethanol) was added to the ground tissue and incubated at 65 °C for 40 min, shaking gently every 10 min. After a short spin, 340 µL of chloroform-isoamyl alcohol (24:1) was added and mixed inverting the plates. Tubes were centrifuged for 10 min at 3000 rpm and 4 °C. The clean aqueous phase was transferred to new strip tubes, and 1.5 vol of 100% ethanol and 15 mM ammonium acetate were added and mixed gently. After overnight incubation at −20 °C, tubes were centrifuged for 10 min at 3000 rpm at 4 °C. The supernatant was discarded inverting the tubes, and 300 µL of 70% ethanol was added. After centrifugation for 10 min at 3000 rpm at 4 °C, the supernatant was discarded and finally 75 µL of TE was added. DNA at 1:10 dilution was used for PCR. Some random DNA samples from each plate were subjected to quality control. DNA integrity was checked on an agarose gel, and quantification was performed using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, Wilmington, DE, USA).

### *2.2. PMC2 Genotype by Allele-Specific PCR Assay*

PMC2 marker genotyping was performed using the allele-specific forward primer (PMC2-F-alleleR: 5'-GTCATTTTCATTGATGTCATTC A-3' or PMC2-F-alleleS: 5'-GTCATTTTCATTGATGTCATTC A-3') and one common reverse primer (PMC2-R: 5'- GTCATTTTCATTGATGTCATTC A-3'), as described by Zuriaga et al. [21]. PCRs were performed in a final volume of 20 µL containing 1 × DreamTaq buffer, 0.2 mM of each dNTP, 5 µM of each primer, 1 U of DreamTaq DNA polymerase (Thermo Fisher) and 2 µL of DNA extraction (diluted 1:10). Cycling conditions were as follows: an initial denaturing of 95 °C for 5 min; 35 cycles of 95 °C for 30 s, 55 °C for 45 s and 72 °C for 45 s; and a final extension of 72 °C for 10 min. PCR products were electrophoresed in 1% (w/v) agarose gels.

Available DNA samples from 120 apricot cultivars and accessions were PCR screened in this work. Part of this collection is currently kept at the collection of the Instituto Valenciano de Investigaciones Agrarias (IVIA) in Valencia (Spain), while other samples were provided by the Departamento de Mejora y Patología Vegetal del CEBAS-CSIC in Murcia (Spain), the University of St. Istvan (Budapest, Hungary) or by SharCo project (FP7-KBBE-2007-1) partners.

### *2.3. WGS Mapping and PMC2 Screening*

WGSs of 73 cultivars were used in this study. Twenty-four of these WGSs and the 454 sequenced BAC clones belonging to the “Goldrich” *PPVres* locus R-haplotype were already screened in our previous works [20–31]. The other 49 WGSs were downloaded from the SRA repository (<https://www.ncbi.nlm.nih.gov/sra>). All raw reads were processed using the “run\_trimmomatic\_qual\_trimming.pl” script from the Trinity software [32]. After removing the low-quality regions as well as vector and

adaptor contaminants, cleaned reads were aligned to the peach genome v.2.0.a1 [33] using Bowtie2 v.2.2.4 software [34]. The presence/absence of the *ParPMC2-del* was visually inspected using IGV v.2.4.16 [35].

### 3. Results and Discussion

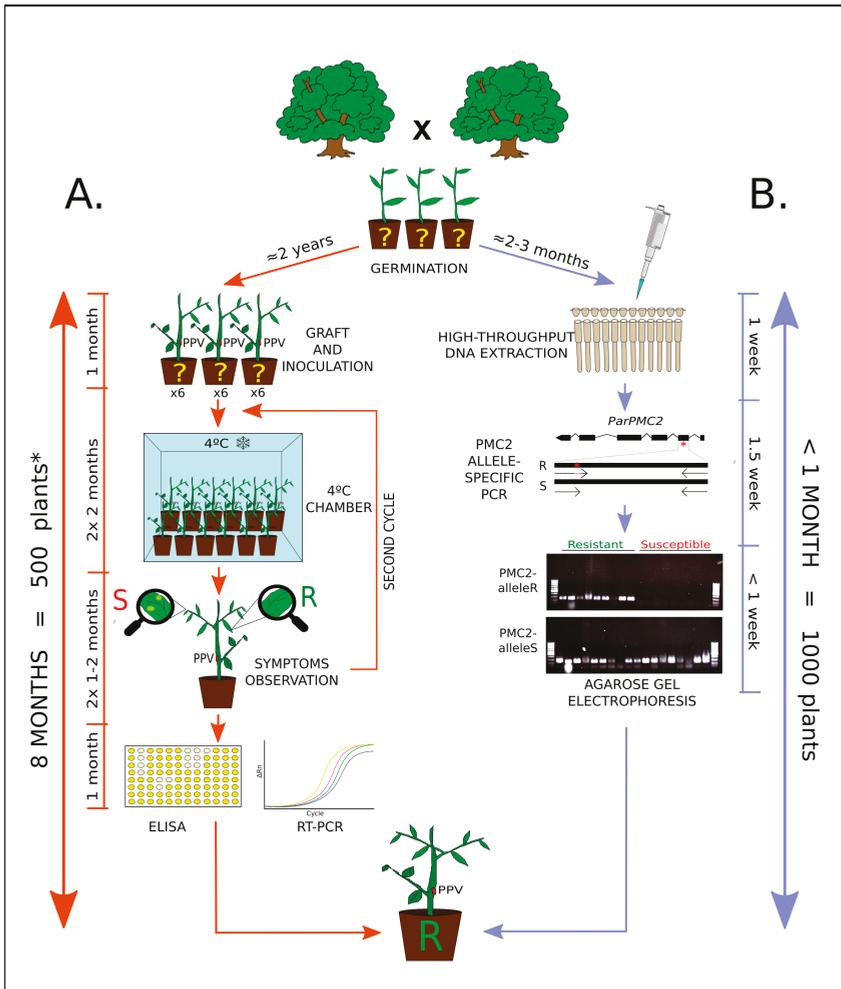
#### 3.1. High-Throughput DNA Extraction and *ParPMC2-del* Genotyping for MAS

MAS offers great advantages over traditional seedling selection based just on phenotypic evaluations in fruit breeding [36]. DNA tests in segregating populations can improve the cost efficiency and/or the genetic gain for each seedling selection cycle [29], allowing to identify a few seedlings from among many thousands that have the genetic potential for desired performance levels [37]. As a result, agronomical evaluation in field trials is restricted to the promising selected materials. Implementation of MAS is especially valuable for traits that are difficult and/or expensive to phenotype as PPV resistance. As previously explained, the most reliable PPV resistance phenotyping is based on a biological test that uses graft-inoculated GF-305 peach seedlings [25] (Figure 1A). This protocol requires several replicates per genotype and visual symptoms inspection during 2–4 growing seasons, which entails the main bottleneck in apricot breeding programs. For instance, following this method at the IVIA's greenhouse and cold chamber facilities, we can phenotype no more than 3000 plants per year, which equals 500 seedlings (i.e., 6 replicates are needed for each seedling).

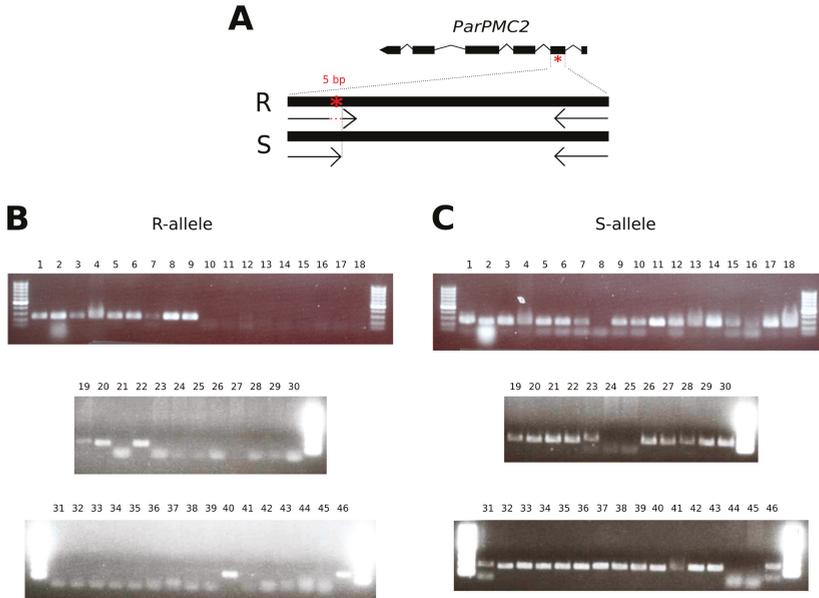
In this work, we present a new strategy to speed up while reducing costs of the current application of MAS for PPV resistance in apricot [18,24,28]. Here, we combine a high-throughput DNA extraction protocol that does not need sophisticated robotic systems and can be implemented in any regular laboratory, with PMC2 allele-specific PCR amplification using previously described primers [21] and agarose electrophoresis (Figure 1B). Both forwards primers differ at the 3'-end, allowing to easily discriminate the presence/absence of the 5-bp *ParPMC2-del* (Figure 2). With this DNA extraction method, one person can easily process up to 384 samples (four 96-well sample plates) in 2 working days, enabling high throughput sample preparation. This is 4 times more samples than a standard CTAB method using individual tubes, while the cost of reagents and consumables is similar in both cases (around 0.29–0.30 € per sample) (Table S2). DNA obtained has enough quantity and quality to ensure subsequent regular PCRs. A 1:10 dilution of the DNA obtained was directly used for PCR amplification, without any additional purification step. In contrast, commercial kits are much more expensive in terms of reagents and consumables with costs around 4€ per sample. Then, using this DNA, 3 different methods could be applied for PPV MAS in apricot: the fluorescent labelling of PCR fragments that are resolved using capillary electrophoresis [18], the high-resolution melting (HRM) approach [28], and the use of standard PCR resolved by agarose gel electrophoresis [21]. It should be noted that the first two methods require the use of special equipment that could not be available for some laboratories and that also make the protocol more expensive. For instance, just the capillary electrophoresis costs around 1.5–2€ per sample (PCR not included) and the fluorescently labelled primers needed for PCR (136€ 10 nm) are much more expensive than the non-labelled ones (4€ 20 nm). On the other hand, commercial kits for HRM are not very expensive (around 1€ per sample) but requires the use of real-time PCR machines specially calibrated for this type of experiments and the analysis software. As a resume, although prices differ between laboratories or countries, our rough estimate of the cost points to first and second approaches as 13 and 8 times more expensive, respectively, in terms of reagents and consumables than the protocol proposed in this work (Table S2).

Practical advantages of PMC2 genotyping over classical phenotyping may be illustrated by the following example (Figure 1). The estimated time needed for evaluating 1000 samples at the IVIA's facilities using bioassays is about 16 months (500 samples/8 months), taking into account that plants should be big enough to be ready-to-graft (approximately 2 years old). In contrast, just about 4 weeks are needed to conduct PMC2 genotyping just after seed germination. This estimated time was calculated assuming a 40-h workweek. As 1000 samples could be distributed into 10.4 96-well plates, ideally the

DNA extraction would need 5.2 days (4 plates each 2 days), the 2 allele-specific PCRs would need 7.8 days (3 h each plate) and the agarose electrophoresis would last 2.6 days (2 PCR 96-well plates and 2 h per gel). In total, we would need 15.6 working days to genotype 1000 samples. This improvement removes the phenotyping bottleneck since all seedlings obtained from a particular cross can be PCR screened that same year. Hence, this quick and high-throughput method for DNA testing is expected to have an important effect on the cost efficiency of MAS, as suggested by Edge-Garza et al. [37].



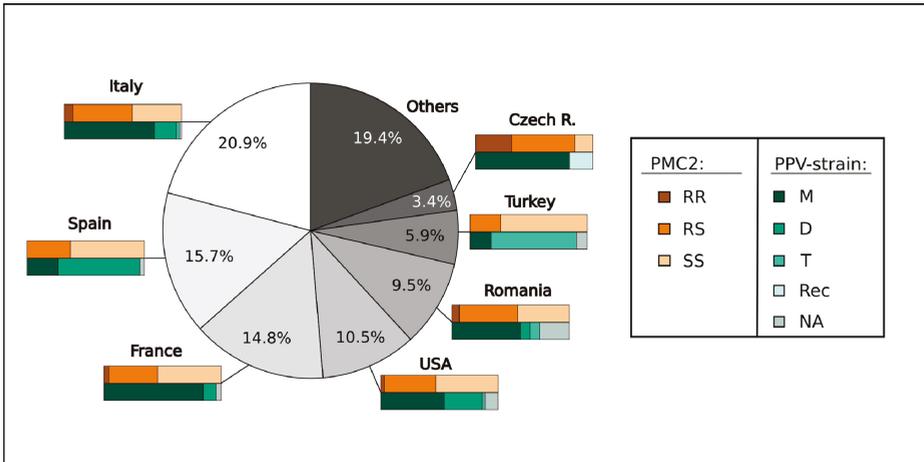
**Figure 1.** Comparison between traditional Plum pox virus (PPV) resistance phenotyping (A) and high-throughput marker-assisted selection (MAS) based on PMC2 allele-specific PCR (B). (\*) Estimated duration based on Instituto Valenciano de Investigaciones Agrarias (IVIA) facilities.



**Figure 2.** *PMC2* genotyping by allele-specific PCR using forward primers differing at the 3'-end (A): R-allele (B) and S-allele (C) amplifications in 1% agarose gel electrophoresis for 46 apricot accessions (1: Goldrich, 2: Harlayne, 3: Henderson, 4: Lito, 5: Orange Red, 6: Pandora, 7: SEO, 8: Stella, 9: Veecot, 10: Bebeco, 11: Bergeron, 12: Canino, 13: Currot, 14: Ginesta, 15: Katy, 16: Mitger, 17: Palau, 18: Tyrinthos, 19: Piera, 20: Selene, 21: Colorao, 22: Moixent, 23: Perla, 24: Dama Vermella, 25: Maravilla, 26: Ninfa, 27: Palabras, 28: Sublime, 29: Dorada, 30: Castlebrite, 31: Martinet, 32: Corbató, 33: Gandia, 34: Cristali, 35: Manri, 36: Gavatxet, 37: Pisana, 38: Xirivello, 39: Velazquez, 40: Mirlo Rojo, 41: Rojo Carlet, 42: Bulida, 43: ASP, 44: Silvercot, 45: Bora and 46: Roxana).

### 3.2. *ParPMC2-del* Highly Correlates with PPV Resistance in Apricot Germplasm

One of the main pillars of plant breeding relies on skilful parental selection to create new genetic variation by controlled crossing. Usually, breeders just connect the concept of DNA-informed breeding with the use of molecular markers for seedling selection, but it also can be very helpful for parental selection [36]. This is the case in apricot breeding for PPV resistance. Two decades ago, Martínez-Gómez et al. [9] reviewed phenotypic information regarding apricot cultivar behaviour against PPV. Similarly, here, we compile the *PMC2* genotype of a wide set of apricot accessions to facilitate parental selection tasks incorporating also their resistance phenotype, pedigree and origin data from the literature when available. The PPV strain used for phenotyping was also included because differences in severity of the induced symptoms have been observed [10,16]. As a result, after screening 120 accessions by PCR and other 49 by WGS and reviewing the available literature, *PMC2* genotype was determined in a total of 325 apricot cultivars or accessions that represent a wide range of geographic origins (Figure 3). A significant part of the materials come from European countries directly involved in PPV resistance research during the last decades, such as Italy (20.9%), Spain (15.7%) or France (14.8%) [38–42]. Regarding viral strain, PPV-M was more frequently used for phenotyping except for PPV-D in Spain and PPV-T in Turkey (Figure 3), in agreement with the prevalence of these two strains in every country [5,43].



**Figure 3.** Geographic distribution of apricot accessions: PMC2 genotypes (RR: homozygous for the resistant allele; SS: homozygous for the susceptible allele; and RS: heterozygous) and PPV strain used for phenotyping are also indicated.

In total, 110 accessions were considered phenotypically resistant (Table 1), 108 were susceptible (Table 2) and 11 showed uncertain phenotype against the same or different PPV strains (Table 3). *ParPMC2-del* highly correlates with PPV resistance, as evidenced by its presence in 92.8% of the resistant accessions (Table 1) and its absence in 92.6% of the susceptible accessions (Table 2). Only 16 out of 219 (7.3%) accessions phenotypically classified as resistant or susceptible showed genotype-phenotype incongruences (GPIs). GPIs were previously reported mainly when using segregating populations [18,23,24,28,44], but clarifying reasons underlying GPIs was found difficult, as quite different factors may be involved. These factors include complex phenotyping protocols, loci other than *PPVres* contributing to PPV resistance, environmental conditions and/or gene–environment interactions. Additionally, putative misclassifications could also explain some genotypic discrepancies observed in this work. For instance, Sunglo, the resistant donor parent of Goldrich, has been phenotyped as resistant by several authors using PPV-M [15,45,46] and PPV-D [47] and genotypically showed the SSR-resistant alleles targeting the *PPVres* locus [18]. However, WGS data (SRR2153157) supposedly corresponding to this accession do not have the *ParPMC2-del*. Something similar occurs with Mirlo Naranja, classified as resistant [48], that was found to carry one copy of the *ParPMC2-del* by PCR in this work but not in that of Passaro [49]. Detailed accession documentation may be helpful to resolve these discrepancies, but 13 of the 16 identified GPIs have no pedigree data available. This information would be very valuable to increase the efficiency of apricot breeding programs and germplasm management.

Table 1. Apricot PPV-resistant accessions genotyped for PMC2.

Name	Country <sup>a</sup>	Origin	Pedigree	PPV Resistance Phenotype <sup>b</sup>	PPV Strain Used	First Phenotype Ref	PMC2 Genotype <sup>c</sup>	PMC2 Genotype Ref
A4316	IT			R	M	[15]	RS	WGS
A4804	IT			R	M	[15]	RS	WGS
Adriana (= Le-3241)	CR	Horticulture Faculty, Lednice	Vestiar × SEO [50]	R	M Rec	[51] [52]	RR	[24]
Alfred (= NY345)	USA	Geneva, NY State Expt Sta, by Robert C. Lamb	OP seedling of selection from (Doty × Geneva)	R	M	[53]	RS	WGS
Andswee	IR			R	M	[15]	RS	WGS
Anegat	FR	INRA, CEP Innovation		R	M/D	[54]	RS	[49]
Bergarouge (= Avirine A2914)	FR	INRA	Bergeron × Orange Red [55]	R	D M	[23] [56]	RS	[49]
Bergeval (= Avicelo, A3950)	FR	INRA		R	M	[56]	RS	[49]
BO03615011	IT		Goldrich × Harlayne [28]	R	M*	[49]	RS	[49]
BO03615025	IT		Goldrich × Harlayne [28]	R	M*	[49]	RR	[49]
BO03615034	IT		Goldrich × Harlayne [28]	R	M*	[28]	RR	[28]
BO03615049	IT		Goldrich × Harlayne [28]	R	M*	[28]	RR	[28]
BO03615053	IT		Goldrich × Harlayne [28]	R	M*	[28]	RS	[28]
BO03615070	IT		Goldrich × Harlayne [28]	R	M*	[49]	RR	[49]
BO04624031	IT		Portici × Goldrich [28]	R	M*	[28]	RS	[28]
BO04624039	IT		Portici × Goldrich [28]	R	M*	[49]	SS	[49]
BO05636034	IT		Kyoto × Priscilla [28]	R	M*	[28]	RS	[28]
BO06609012	IT		Silvercot × Bora [28]	R	M*	[49]	RS	[49]
BO06609013	IT		Silvercot × Bora [28]	R	M*	[49]	RS	[49]
BO06609024	IT		Silvercot × Bora [28]	R	M*	[49]	RS	[49]
BO06609033	IT		Silvercot × Bora [28]	R	M*	[49]	RS	[49]
BO06609036	IT		Silvercot × Bora [28]	R	M*	[49]	RS	[49]
BO06609037	IT		Silvercot × Bora [28]	R	M*	[49]	RS	[49]
BO06609039	IT		Silvercot × Bora [28]	R	M*	[49]	RS	[49]
BO06609045	IT		Silvercot × Bora [28]	R	M*	[49]	RS	[49]

Table 1. *Conti.*

Name	Country <sup>a</sup>	Origin	Pedigree	PPV Resistance Phenotype <sup>b</sup>	PPV Strain Used	First Phenotype Ref	PMC2 Genotype <sup>c</sup>	PMC2 Genotype Ref
BO06609048	IT		Silvercot × Bora [28]	R	M*	[28]	RS	[28]
BO06609055	IT		Silvercot × Bora [28]	R	M*	[28]	RS	[28]
BO06609060	IT		Silvercot × Bora [28]	R	M*	[49]	RS	[49]
BO06609068	IT		Silvercot × Bora [28]	R	M*	[49]	RS	[49]
BO06609074	IT		Silvercot × Bora [28]	R	M*	[49]	RS	[49]
BO06609079	IT		Silvercot × Bora [28]	R	M*	[49]	RS	[49]
BO06609083	IT		Silvercot × Bora [28]	R	M*	[49]	RS	[49]
BO06609087	IT		Silvercot × Bora [28]	R	M*	[49]	RS	[49]
BO06609099	IT		Silvercot × Bora [28]	R	M*	[49]	RS	[49]
BO06609104	IT		Silvercot × Bora [28]	R	M*	[49]	RS	[49]
BO06609113	IT		Silvercot × Bora [28]	R	M*	[49]	RS	[49]
BO06609129	IT		Silvercot × Bora [28]	R	M*	[49]	RS	[49]
BO06609133	IT		Silvercot × Bora [28]	R	M*	[49]	RS	[49]
BO06609136	IT		Silvercot × Bora [28]	R	M*	[49]	RS	[49]
BO96621002	IT		Goldrich × Lito [28]	R	M	[57]	RR	[28]
BO96621030	IT		Goldrich × Lito [28]	R	M	[57]	RS	[28]
Bora (BO90610010)	IT	University of Bologna and Milan, by D. Bassi	Early Blush × PA 7005-2 [58]	R	M/D	[58]	RS	[21,28]
Candela (= LE-2927)	CR	Horticulture Faculty, Lednice	Hungarian Best × SEO [59]	R	M	[60]	RR	[49]
Cebir	TU			R	T	[61]	RS	[61]
Congat	FR	INRA, CEP Innovation		R	-	[62]	RS	[49]
Early Blush (= RUTBHART, NJA53, Aurora46)	US	Rutgers Horticultural Research Farm, New Brunswick, N.J.	RR17-62 × NJA-13 [63]	R	D M T	[64] [65] [61]	RS	PCR [21,28,61]
Fartis	FR	Marie-France BOIS, France (IFS)		R	M*	[28,49]	RS	[28]
Farmingdale (= NY346)	USA	Geneva, NY State Expt Sta, by Robert C. Lamb	OP seedling of selection from (Doty × Geneva) [66]	R	M	[53]	RS	[28]

Table 1. *Cont.*

Name	Country <sup>a</sup>	Origin	Pedigree	PPV Resistance Phenotype <sup>b</sup>	PPV Strain Used	First Phenotype Ref	PMC2 Genotype <sup>c</sup>	PMC2 Genotype Ref
Flavor cot (=Bayoto)	USA	Washington State University Research, by Tom Toyama		R	M	[57]	RS	[28]
Floppia	FR	PSB Producción Vegetal S.L.		R	M*	[28]	RS	PCR; [28]
GG9310	SP	IVIA, Moncada, Valencia	Goldrich × Ginesta [18]	R	D	IVIA	RS	PCR
GG9318	SP	IVIA, Moncada, Valencia	Goldrich × Ginesta [18]	R	D	IVIA	RS	PCR
GG937	SP	IVIA, Moncada, Valencia	Goldrich × Ginesta [18]	R	D	IVIA	RS	PCR
GG941	SP	IVIA, Moncada, Valencia	Goldrich × Ginesta [18]	R	D	IVIA	RS	PCR
GG979	SP	IVIA, Moncada, Valencia	Goldrich × Ginesta [18]	R	D	IVIA	RS	PCR
GG9869	SP	IVIA, Moncada, Valencia	Goldrich × Ginesta [18]	R	D	IVIA	RS	PCR
Gilgat	FR	INRA / CEP INNOVATION		R	M*	[28]; [49]	RS	[28]
GP9817	SP	IVIA, Moncada, Valencia	Goldrich × Palau [18]	R	D	IVIA	RS	PCR
Dama Rosa (GG9871)	SP	IVIA, Moncada, Valencia	Goldrich × Ginesta [18]	R	D	IVIA	RS	PCR; [49]
Dama Taronja (GK988)	SP	IVIA, Moncada, Valencia	Goldrich × Katy [18]	R	D	IVIA	RS	PCR; [49]
Dulcinea	IT	Pisa University	Moniqui OP [67]	R	D	[64]	SS	PCR; [49]
Fracasso	IT			Tolerant	T	[61]	SS	[61]
Harlayne	C	Agr. Canada, Res. Station, Harrow, Ontario, by REC Layne	V51092 ((Reliable × OP) × OP) × Sun Glo [66]	R	D	[68]	RS	PCR; [20,21,24,28,61]
Harval (=HW437)	C	Agr. Canada, Res. Station, Harrow, Ontario, by REC Layne	Veecot × HW435 (Rouge du Roussillon × NJA2 (Morden604 OP)) [66]	R	M	[69]	RS	[28]
Henderson	USA	Geneva, NY, by GW Henderson	Unknown [66]	R	M	[46]	RS	PCR; [21]
Kanış (=M2252)	TU			R	D	[70]	SS	[61]
Karum	TU			R	T	[61]	RS	[61]
Lady cot (=HYB 3-3)	FR	COT International		R	M*	[28]	RS	[28]
Laycot	C		V51092 ((Reliable o.p.) o.p.) × NJA1 [71]	R	M	[15]	RR	WGS

Table 1. *Contd.*

Name	Country <sup>a</sup>	Origin	Pedigree	PPV Resistance Phenotype <sup>b</sup>	PPV Strain Used	First Phenotype Ref	PMC2 Genotype <sup>c</sup>	PMC2 Genotype Ref
LE-2904	CR	Horticulture Faculty, Lednice	Velkopavlovická × SEO [19]	R	M	[72]	RS	[49]
LE-3205	CR	Horticulture Faculty, Lednice		R	M*	[49]	RR	[49]
Le-3246	CR	Horticulture Faculty, Lednice	Vestár × SEO [51]	R	M	[51]	RS	[24]
LE-3662	CR	Horticulture Faculty, Lednice		R	M	[72]	RR	[49]
Lifos	TU			R	T	[61]	RS	[61]
Lillycot	FR	SDR Fruit Lic (US)	Unknown [73]	R	M*	[28]	RS	[28]
Lito	GR		SEO × Tirynthos [18]	R	M D	[74] IVIA	RS	PCR; [24,28]
Mediabel (=Mediabel)	FR	Newcot and IPS		R	M*	[28]	RS	[28]
Mirlo Naranja (= Mirlo anaranjado)	SP	CEBAS-CSIC, Murcia	Rojo Pasión × Búvida Precoz [48]	R	D	[48]	RS	PCR
Mirlo Blanco	SP	CEBAS-CSIC, Murcia	Rojo Pasión × Búvida Precoz [48]	R	D	[48]	RS	[28]
Mirlo Rojo	SP	CEBAS-CSIC, Murcia	Rojo Pasión × Búvida Precoz [48]	R	D	[48]	RS	PCR; [49]
Mogador	SP	PSB Producción Vegetal S.L.		R	M*	[28,49]	RS	PCR; [28]
Moixent (=GM961)	SP	IVIA, Valencia	Goldrich × Mitiger [18]	R	D	IVIA	RS	PCR; [49]
Murciana	SP	CEBAS-CSIC, Murcia	Orange Red × Currot [73]	R	D M	[75] [15]	RS	WGS; PCR; [49]
Nikitskii	UKR			R	M	[15]	RS	WGS
NJA42	USA	New Jersey	NJA12 × NJA13 [76]	R	?	[77]	RS	PCR
Orange Red (=Barth; NJA-32)	USA	New Jersey	Lasgerdi Mashhad × NJA2 (= Morden 604 OP) [78]	R	D M	[68] [79]	RS	PCR; [21]
Pandora	GR		SEO × Tirynthos [18]	R	M D	[74] [47]	RS	PCR

Table 1. *Cont.*

Name	Country <sup>a</sup>	Origin	Pedigree	PPV Resistance Phenotype <sup>b</sup>	PPV Strain Used	First Phenotype Ref	PMC2 Genotype <sup>c</sup>	PMC2 Genotype Ref
Pelese di Giovanniello	IT	Murcia		Tolerant	D	[64]	SS	[49]
Perla	SP			R	D	[64]	SS	PCR
Petra (BO88617102)	IT	University of Bologna and Milan, Italy, by D Bassi	Goldrich × Pelese di Giovanniello [73]	R	M*	[28]	RS	[28]
Precoce d'Imola	IT			tolerant	D	[64]	SS	WGS
Priboto (=Zebra)	FR		bud mutation of Goldrich [80]	R	M	[15]	RS	WGS; [49]
Pricia	FR	Marie-France BOIS, France (IFS)		R	M*	[28,49]	RS	[28]
Pseudo Royal	USA			R	M	[15]	RS	WGS
Robada (= K106-2)	USA		Orange Red × K113-40 (ancestry includes Blenheim, Blush and Perfection) [81]	R	M	[82]	RS	WGS
Rojo Pasión	SP	CEBAS-CSIC	Orange Red × Currot [83]	R	D	[83]	RS	PCR; [49]
Rosa	SP	CEBAS-CSIC, Murcia	Orange Red × Palsteyn [73]	R	D	[41]	RS	[49]
Rubista	FR	Marie-France BOIS, France (IFS)		Tolerant	M*	[28,49]	RS	[28]
Sabbatani (= Selezione Sabbatani?)	IT			R	D	[64]	SS	[49]
Selene	SP	CEBAS-CSIC	Goldrich × A2564 (=Screara × SEO) [18]	R	D	[84]	RS	PCR; [49]
SEO9/34	SP	IVIA	SEO × Palau [18]	R	D	IVIA	RS	PCR
Spring Blush (= EA3126TH)	FR	Escande EARL		R	M*	[57]	RR	[49]
Stark Early Orange (= SEO, Earle Orange)	USA	Grandview, Washington, by WL Roberts	Unknown [66]	R	M	[85]	RS	PCR; [20,21,24,28,61]
Stella	USA		Unknown [18]	R	M	[85]	RR	PCR; [21]
Sunglo (= Sun Glo)	USA	Columbia & Okanogan Nursery Co.	Unknown [66]	R	M	[45]	RS	PCR
					D	[47]	SS	WGS

Table 1. *Contt.*

Name	Country <sup>a</sup>	Origin	Pedigree	PPV Resistance Phenotype <sup>b</sup>	PPV Strain Used	First Phenotype Ref	PMC2 Genotype <sup>c</sup>	PMC2 Genotype Ref
Sunnycot (= 97-3-203)	USA	SDR FRUIT LLC – USA		R		[62]	RS	[49]
Traian	RO			R	D	[86]	RS	PCR; [87]
Tsumami (= EA 5016)	FR	Escande EARL		R	M*	[28]	RS	[28]
Wonder Cot (= RM 7)	USA	SDR FRUIT LLC – USA		R	M*	[28]	RS	[28]
Zard	CA			R	T	[61]	RS	[61]

M\*: strain likely used for phenotyping by the Phytosanitary Service, Emilia-Romagna (Italy).<sup>a</sup> Countries: C: Canada, CA: Central Asia, CR: Czech Republic, FR: France, GR: Greece, IR: Iran, IT: Italy, RO: Romania, SP: Spain, TU: Tunisia, TR: Turkey, UKR: Ukraine, US: United States of America.<sup>b</sup> Phenotype: R: Resistant, S: Susceptible; <sup>c</sup> Genotype: RR: homozygous for PMC2 resistant allele, SS: homozygous for PMC2 susceptible allele, RS: heterozygous.

Table 2. Apricot PPV susceptible accessions genotyped for *ParrPMC2-del*.

Cultivar	Country <sup>a</sup>	Origin	Pedigree	PPV Resistance Phenotype <sup>b</sup>	PPV Strain Used	First Phenotype Ref	PMC2 Genotype <sup>c</sup>	PMC2 Genotype Ref
A 3521	IR			S	M	[15]	SS	WGS
A 3522	IR			S	M	[15]	SS	WGS
Amabile Vecchioni	IT	Seedling by Prof. F. Scaramuzzi	Unknown [67]	S	M	[45]	SS	[49]
Aprikoz	TR			S	M	[88]	SS	PCR
Arrogante	SP	Murcia		S	D	[89]	SS	[21]
Avikaline	FR			S	M	[15]	SS	WGS
Bebecou (Bébeco)	GR		Unknown [18]	S	M/D	[90]	SS	PCR; [21,28]
Bella Di Imola	IT		Spontaneous seedling [23]	S	D	[64]	SS	[28]
Bergeron	FR	Saint-Cyr-au-Mont-d'Or, Lyon	Spontaneous seedling [23]	S	M	[90]	SS	PCR; [21]
Big Red (EA4006)	FR	Escande EARL, France		S	M	[57]	RS	[28]
BO0462/042	IT		Portici × Goldrich [28]	S	M*	[28]	SS	[28]
BO0462/043	IT		Portici × Goldrich [28]	S	M*	[28]	SS	[28]
BO0660/003	IT		Silvercot × Bora [28]	S	M*	[49]	RS	[49]
BO81604311	IT		San Castrese × Reale di Imola [73]	S	D	[91]	SS	[24]
BO9662/1021	IT		Goldrich × Lito [28]	S	M*	[28]	RS	[28]
Boucharan Boutard	FR			S	M	[15]	SS	WGS
Búlida	SP	Murcia	Unknown [73]	S	D	[92]	SS	PCR; [21]
					M	[93]		

Table 2. *Conti.*

Cultivar	Country <sup>a</sup>	Origin	Pedigree	PPV Resistance Phenotype <sup>b</sup>	PPV Strain Used	First Phenotype Ref	PMC2 Genotype <sup>c</sup>	PMC2 Genotype Ref
Cařona	IT	Vesuvian area		S	M	[94]	SS	WGS
CAID AGDZ n2	MO			S	M	[15]	SS	WGS
Canino	SP	Valencia	Unknown [18]	S	D	[95]	SS	PCR; [20,21]
Castlebrite (=K111-6)	USA	USDA, Fresno, California	OP seedling of B60-12 (= Perfection x Castleton) [66]	S	M	[45]	SS	PCR
Cegledı Bıbor	HU	Cegledı Horticultural Research Institute	Chance seedling [96]	S	M	[46]	SS	[28]
Colorado (Colorado 43-15)	SP	PSB Produccıon Vegetal SL	Unknown	S	M*	[49]	SS	PCR; [28]
Corbatı	SP	Valencia		S	D	[95]	SS	PCR
Currot	SP	Valencia	Unknown [18]	S	D	[95]	SS	PCR
Estrella	SP	CEBAS-CSIC	Orange Red x ZZ11-18 (= Goldrich x Pepto del Rubio) [23]	S	D	[23]	SS	PCR; [49]
Faralia	FR	Marie-France BOIS, IPS		S	M*	[28]	SS	[28]
Fardo	FR	Marie-France BOIS, IPS		S	M	[57]	SS	[28]
Favorit	RO			S	M	[94,97]	SS	[49]
Geç Abıgıoz	TR			S	T	[61]	SS	[61]
Ginesta	SP	Valencia	Unknown [18]	S	D	[95]	SS	PCR
Dama Vermella (HG9869)	SP	IVIA	Harrot x Ginesta [18]	S	D	[61]	SS	PCR; [49]
Hacchalıođlu	TR			S	T	[61]	SS	[61]
Hangrand (= HW410)	C	Richard EC Layne, Agr. Canada, Res. Station	V51092 ((Reliable x OP) x OP) x NJAI (Phelps x Perfection) [66]	S	M	[45]	SS	[21]
Hasanbey	TR			S	M	[45] <sup>1</sup>	SS	PCR
Hungarian Best = (Best of Hungary?)	HU/RO			S	T	[61]	SS	[61]
Katy	USA	Zaiger's Genetics <sup>4</sup>		S	M	[94]	SS	[61]
Krasnoshchekı	UKR	Advanced/Improved cultivar		S	D	[18]	SS	PCR; [21]
				S	D	[20]	SS	[20,21]

Table 2. Contd.

Cultivar	Country <sup>a</sup>	Origin	Pedigree	PPV Resistance Phenotype <sup>b</sup>	PPV Strain Used	First Phenotype Ref	PMC2 Genotype <sup>c</sup>	PMC2 Genotype Ref
Kyoto (= Kioto)	FR	Escande	Unknown [73]	S	M*	[28]	SS	[28]
Lambertin-1	USA	USDA, Fresno, California	A95-45 × B69-85 (=Perfection × Royal) [98]	S	M	[45]	SS	[21]
Larclyd (= F168 cv; Jenny Cot)	NZ	Central Otago	Sundrop × Moorpark [99]	S	M	[15]	SS	WGS
Le-3218	CR	Faculty of Horticulture in Lednice	Vestar × SEO [51]	S	M	[51]	SS	[24]
Luziet (= Suchett; Hatif du clos; Abricot du Clos)	FR		Spontaneous seedling [71]	S	M	[93]	SS	WGS
Luna	IT			S	M*	[28]	RS	[28]
Madarska Narijlepaša	SL			S	M	[15]	SS	WGS
Magic cot (= RM 22)	USA	SDR FRUIT LLC - USA	Unknown [23]	S	D	[23]	SS	[49]
Manicot	FR			S	D	[97,100]	SS	WGS
Maravilla	SP	CEBAS-CSIC, Murcia	Orange Red × Z211-18 (= Goldrich × Pepito) [23]	S	D	[23]	SS	PCR; [49]
Mari de Cenaad	RO		Unknown	S		[86]	RS	PCR
Markalești	TR			S	T	[61]	SS	[61]
Marlén	CR	Horticulture Faculty, Lednice	clone of Hungarian Best [59]	S	Rec	[14]	SS	PCR; [24]
Marouch 14	MO		Local landrace	S	M	[15]	SS	WGS
Marouch 4	MO			S	M	[15]	SS	WGS
Mei Hwang	CH		Traditional cultivar/landrace	S	M	[15]	SS	WGS
Mektep	TR			S	T	[61]	SS	[61]
Mektep 8	TR			S	T	[61]	SS	[61]
Mitiger	SP	Castellón [30]	Unknown [18]	S	D	[95]	SS	PCR
Monaco Bello	IT			S	M	[46]	SS	WGS
Moniqui	SP	Murcia	Unknown	S	M	[97]	SS	WGS
Mono	USA	Le Grand, California, by FW Anderson	Perfection OP [66]	S	M	[90]	SS	PCR; [21,24]
Moongold (= Moongola?)	USA	University of Minnesota		S	-	[77]	SS	PCR
Moorpark (=Moor Park)	USA			S	M	[46]	SS	WGS
Morden 604	C	Morden, Manitoba, by Canada Dept. Agr. Res. Sta.	Scout × McClure [66]	S	M	[15]	SS	WGS
Ninfa (BO81/60/2075)	IT	University of Bologna and Milan, by D. Bassi	Ouardy × Tyrinthos [55]	S	M*	[28]	SS	PCR; [28,61]

Table 2. *Cont.*

Cultivar	Country <sup>a</sup>	Origin	Pedigree	PPV Resistance Phenotype <sup>b</sup>	PPV Strain Used	First Phenotype Ref	PMC2 Genotype <sup>c</sup>	PMC2 Genotype Ref
Olimp	RO			S	M	[45]	SS	WGS; [49]
Orange Rubis (=Couloumine)	FR	Mallard		S	M	[57]	SS	[28]
Ordubat B.	TR			S	T	[61]	SS	[61]
Ouardi	TU	INRAT, Ariana	Camino × Hamidi [101]	S	M	[46]	SS	[49]
Palstejn (Palstein)	SA		Blenheim × Canino [73]	S	M	[102]	SS	WGS
Palabras	SP			S	D	[95]	SS	PCR
Palau	SP		Unknown [18]	S	D	[46]	SS	PCR
Pavlot	FR			S	M	[93]	SS	WGS
Pêche De Nancy	FR			S	M	[15]	SS	WGS
Perfection	USA	Waterville, Washington	Unknown [66]	S	M	[46]	SS	[21]
Piera				S	M	[65]	RS	PCR
Poizat	FR			S	M	[15]	SS	WGS
Polonais	FR		Spontaneous seedling [23]	S	M	[93]	SS	[24]
Poppy	USA	Zaiger Genetics, Inc., Modesto, CA	78EB575 × 123CD161 [58]	S	D	[23]	SS	[49]
Portici (= Pertini)	IT	Vesuvian area	Unknown; Local selection [23]	S	D	[46]	SS	PCR; [28]
Precoce Ampuis	FR			S	M	[15]	SS	WGS
Reale d'Imola	IT		Luizet OP [23]	S	D	[64]	SS	[21,24,49]
Rojo de Carlet	SP	Valencia		S	D	[95]	SS	PCR
Rouge Du Roussillon	FR			S	M	[45]	SS	WGS
Rouge De Fournes	FR			S	M	[15]	SS	WGS
Saturn	RO			S	M	[45]	SS	WGS
Screara	FR			S	D	[70]	SS	WGS
Skerpare B.	TR			S	M	[45]	SS	WGS
Shalakh (= Yerevani, Erevani)	AR		Local selection [23]	S	T	[61]	SS	[61]
Silistra × Anamas (Marculesti 43/1)	RO			S	M	[93]	SS	WGS; [20,21]
				S	M	[15]	SS	WGS

Table 2. Contd.

Cultivar	Country <sup>a</sup>	Origin	Pedigree	PPV Resistance Phenotype <sup>b</sup>	PPV Strain Used	First Phenotype Ref	PMC2 Genotype <sup>c</sup>	PMC2 Genotype Ref
Sucre De Holub	HU	Bohème, by M. Holub		S	M	[15]	SS	WGS
Sublime	SP	CEBAS-CSIC	Orange Red × Z211-18 (= Goldrich × Pepto del Rubio) [103]	S	D	[103]	SS	PCR; [49]
Super Rouge	FR			S	M	[15]	SS	WGS
Sweet Red	FR			S	M	[57]	SS	[49]
Szegedi narnut (=Szegedi Mamut?)	HU	Foki István and Kovács Imre	Hybrid of Cegledi orias, "Giant" group [96]	S	M	[94]	SS	[49]
Tabriz	TR			S		[86]	SS	PCR
Tadeo (= Taddeo)	SP	Valencia		S	D	[95]	SS	PCR
Tardif De Bordaueil	FR		Unknown [23]	S	M	[46]	SS	WGS
Tardif De Jain	FR			S	D	[64]	SS	WGS
Tonda di costigliole	IT	Piedmont		S	M	[104]	SS	[49]
Trevatt	AU			S	M	[45]	SS	PCR
Tyrinthos	GR		Unknown [18]	S	D	[70]	SS	PCR; WGS; [49]
Uleamos	SP	Ulea, Murcia		S	D	[97]	SS	[49]
Velázquez	SP	Murcia		S	D	[89]	SS	PCR; [21]
Venus (= Venus 1414?)	RO		(Umberto × Ananas) × (Luizet × Umberto) [96]	S	M	[46]	SS	[49]
Vestar	CR		Hungarian Best × mixture of pollen from Chinese cultivars [55]	S	M	[105]	RS	WGS; [54]
Vivagold	C	Vineland Station, Ontario	Veccot × V49024 (= Geneva × Gibb) [66]	S	M	[15]	SS	WGS
Xirivello (=Chirivello)	SP	Valencia	Unknown	S	M	[46]	SS	PCR
Yilbat (=M2243)	TR			S	T	[61]	RS	[61]

M\*: strain likely used for phenotyping by the Phytosanitary Service, Emilia-Romagna (Italy). <sup>a</sup> Countries: AR: Armenia, AU: Australia, C: Canada, CH: China, CR: Czech Republic, FR: France, GR: Greece, HU: Hungary, IR: Iran, IT: Italy, MO: Morocco, NZ: New Zealand, RO: Romania, SA: South Africa, SL: Slovakia, SP: Spain, TR: Turkey, UKR: Ukraine and US: United States of America, <sup>b</sup> Phenotype: R: Resistant, S: Susceptible; <sup>c</sup> Genotype: RR: homozygous for PMC2 resistant allele, SS: homozygous for PMC2 susceptible allele and RS: heterozygous.

**Table 3.** Apricot accessions with uncertain PPV resistance phenotype genotyped for *ParrPMC2*-del.

Cultivar	Country <sup>a</sup>	Origin	Pedigree	PPV Resistance Phenotype <sup>b</sup>	PPV Strain Used	First Phenotype Ref	PMC2 Genotype <sup>c</sup>	PMC2 Genotype Ref
Bachmi	IR			S T	M D	[102]	SS	WGS
Farbaly	FR	Marie-France BOIS, IPS		S R	M* M	[28] [56]	RS	[28]
Goldrich	USA	USDA and Washington State University, Prosser, Washington	Sun Glo × Perfection [73]	R uncertain	D D	[68] [45] [106] [68]	RS	PCR; [20,21,24,28]
Harcot	C	Agr. Canada, Res. Station, Harrow, Ontario, by REC Layne	(T2 (Geneva × Narmata) × Morden 604 (Scout × McClure)) × NJAI (Phelps × Perfection) [66]	T?	-	[28]	RS	PCR; [21,24,28]
Incomparable de Malissard (= Valsard)	FR	Malissard, Valence		S R S	T M M	[61] [15] [57]	SS	[61] WGS
Pisana	IT		ICAPI 26/5 OP [55]	S R R	M* M D	[28] [65] [64]	SS	PCR; [28]
Pieve (BO96/08015)	IT	University of Bologna and Milan, by D. Bassi	Harcot × Reale di Imola [73]	S R	M* M	[28] [65]	SS	[28]
San Caestre	IT	Naples	Unknown [73]	T S	D M	[64] [46]	SS	WGS; [49]
Sulmona	RO		(Luizet × Re Umberto) × (Ananas × Ananas) [71]	S R	M -	[45] [77]	SS	[49]
Veecot	C	Ontario Dept Agr Res Inst, Vineland Station, Ontario, by OA Bradt	Reliable OP [18]	R S	M	[45] [105]	RS	PCR; [21]
Viceroy (=Viceroy_605_G?)	RO			R S	- -	[77] [86]	SS	PCR

M\*: strain likely used for phenotyping by the Phytosanitary Service, Emilia-Romagna (Italy).<sup>a</sup> Countries: C: Canada, FR: France, IR: Iran, IT: Italy and RO: Romania; <sup>b</sup> Phenotype: R: Resistant and S: Susceptible; and <sup>c</sup> Genotype: RR: homozygous for PMC2 resistant allele, SS: homozygous for PMC2 susceptible allele and RS: heterozygous.

Accurate evaluation of PPV resistance is a complex process, and results obtained by different researchers sometimes are contradictory, as exemplified by Farbaly and Pieve (Table 3), which may lead to GPIs. This problem is also observed in well-known accessions. For instance, Goldrich, usually classified as resistant against both PPV-D and M strains, has also been classified as uncertain or even as susceptible at least once (Table 3). Moreover, the effect of the PPV strain used [9,24] has also been observed, as at least 5 accessions showed different behaviour against PPV-M, D or T infection (Table 3). In addition, the environmental effect on symptoms and the different PPV detection techniques employed could also be involved in GPIs [9].

On the other hand, PPV resistance has been related with the downregulation of both *ParPMC2* and, especially, *ParPMC1*, putatively due to an RNA silencing mechanism triggered by the pseudogenization of *ParPMC2res* [21]. Notwithstanding, the presence of epigenetic changes has also been suggested as a possible cause [22]. In any case, resistant cultivars show residual expression levels that could somehow be influenced by environmental conditions. This might explain sporadic symptoms that eventually lead to GPI classification. Moreover, the role of additional PPV resistance loci or genes may also contribute to GPIs. In this sense, Gallois et al. [105] pointed out that a large part of a resistant phenotype conferred by a given QTL depends on the genetic background due to frequent epistatic effects between resistance genes. In fact, other minor loci, linked or not to *PPVres*, have been suggested to underlie PPV resistance in apricot [13–16]. Altogether, the identification and/or confirmation of GPIs in this work pave the way for future studies to unravel the PPV resistance mechanism.

The handful of North American cultivars originally described as PPV resistant [9] have been extensively used as donors in all breeding programs currently in progress. As a result, the *PPVres* locus has been introduced in different genetic backgrounds. In order to complete our survey, genotypic information was compiled from other 96 accessions without available PPV phenotype data (Table S1, [107–113]). In summary, 152 accessions (46.8%) have at least one copy of the *ParPMC2-del* (Figure 3) and 15 out of them are homozygous for *ParPMC2-del*, including the North American PPV-resistant cultivar Stella [114]. Those materials derived from crosses with North American PPV-resistant cultivars represent an opportunity to accelerate the development of new varieties better adapted to the Mediterranean basin conditions [9]. In this context, it should be highlighted that MAS allows to improve cost efficiency and/or genetic gain in apricot breeding programs aimed to select PPV-resistant seedlings. This improvement is highly significant even if some PPV susceptible individuals among those with *ParPMC2-del* are dragged, since they will be later identified by PPV phenotyping. Similarly, Tartarini et al. [115] underlined the advantage of the identification of homozygous *Rvi6* scab-resistant plants using MAS, despite segregating progenies showing at least 5% of GPIs.

#### 4. Conclusions

Here, we present a high-throughput method to quickly perform DNA testing for PPV resistance that may greatly improve the efficiency of apricot breeding programs. The long-lasting PPV phenotyping process will only be performed with those advanced selections showing promising agronomic behaviour in advanced stages to guarantee the selection of PPV-resistant individuals. Additionally, a wide survey over 300 accessions has been made to identify PPV-resistant sources that could also be useful in apricot breeding programs.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4395/10/9/1292/s1>, Table S1. PMC2 genotyped apricot accessions without phenotypic data against PPV infection; Table S2. Estimation cost of DNA extraction and PMC genotyping for PPV MAS in apricot.

**Author Contributions:** Conceptualization: C.R., M.L.B. and E.Z.; experimental procedures: Á.P.-O., I.L. and E.Z.; bioinformatics: E.Z.; funding acquisition: M.L.B.; writing—original draft, E.Z.; writing—review and editing, Á.P.-O., C.R., I.L., M.L.B. and E.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA)-FEDER (RTA2017-00011-C03-01). Á.P.O. was funded by a fellowship cofinanced by the Generalitat Valenciana and European Social Fund (2014–2020) (DOCV 8426/19.11.2018).

**Acknowledgments:** The authors would like to express their gratitude to Bassi (University of Milan, Italy) for providing pedigree information from their apricot breeding program.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Rehder, A. *Manual of Cultivated Trees and Shrubs Hardy in North America*, 2nd ed.; The Macmillan Company: New York, NY, USA, 1940.
2. Zhebentyayeva, T.N.; Ledbetter, C.; Burgos, L.; Llácer, G. Apricots. In *Fruit Breeding*, 1st ed.; Badenes, M.L., Byrne, D.H., Eds.; Springer: New York, NY, USA, 2012; Volume 3, pp. 415–458. [[CrossRef](#)]
3. Moreno, M.A. Breeding and selection of *Prunus* rootstocks at the Aula Dei experimental station, Zaragoza, Spain. *Acta Hort.* **2004**, *658*, 519–528. [[CrossRef](#)]
4. García, J.A.; Cambra, M. Plum pox virus and sharka disease. *Plant. Viruses* **2007**, *1*, 69–79.
5. García, J.A.; Glasa, M.; Cambra, M.; Candresse, T. Plum pox virus and sharka: A model potyvirus and a major disease. *Mol. Plant. Pathol.* **2014**, *15*, 226–241. [[CrossRef](#)]
6. Chirkov, S.; Ivanov, P.; Sheveleva, A.; Zakubanskiy, A.; Osipov, G. New highly divergent Plum pox virus isolates infecting sour cherry in Russia. *Virology* **2017**, *502*, 56–62. [[CrossRef](#)]
7. James, D.; Varga, A.; Sanderson, D. Genetic diversity of Plum pox virus: Strains, disease and related challenges for control. *Can. J. Plant. Pathol.* **2013**, *35*, 431–441. [[CrossRef](#)]
8. Sihelská, N.; Glasa, M.; Šubr, Z.W. Host preference of the major strains of Plum pox virus—Opinions based on regional and world-wide sequence data. *J. Integr. Agric.* **2017**, *16*, 510–515. [[CrossRef](#)]
9. Martínez-Gómez, P.; Dicenta, F.; Audergon, J.M. Behaviour of apricot (*Prunus armeniaca* L.) cultivars in the presence of Sharka (Plum pox potyvirus): A review. *Agronomie* **2000**, *20*, 407–422. [[CrossRef](#)]
10. Dondini, L.; Lain, O.; Vendramin, V.; Rizzo, M.; Vivoli, D.; Adami, M.; Guidarelli, M.; Gaiotti, F.; Palmisano, F.; Bazzoni, A.; et al. Identification of QTL for resistance to Plum pox virus strains M and D in Lito and Harcot apricot cultivars. *Mol. Breed.* **2011**, *27*, 289–299. [[CrossRef](#)]
11. Hurtado, M.A.; Romero, C.; Vilanova, S.; Abbott, A.G.; Llácer, G.; Badenes, M.L. Genetic linkage maps of two apricot cultivars (*Prunus armeniaca* L.) and mapping of PPV (sharka) resistance. *Theor. Appl. Genet.* **2002**, *105*, 182–191. [[CrossRef](#)]
12. Lalli, D.A.; Abbott, A.G.; Zhebentyayeva, T.N.; Badenes, M.L.; Damsteegt, V.; Polák, J.; Krška, B.; Salava, J. A genetic linkage map for an apricot (*Prunus armeniaca* L.) BC1 population mapping Plum pox virus resistance. *Tree Genet. Genomes* **2008**, *4*, 481–493. [[CrossRef](#)]
13. Lambert, P.; Dicenta, F.; Rubio, M.; Audergon, J.M. QTL analysis of resistance to sharka disease in the apricot (*Prunus armeniaca* L.) ‘Polonais’ × ‘Stark Early Orange’ F1 progeny. *Tree Genet. Genomes* **2007**, *3*, 299–309. [[CrossRef](#)]
14. Marandel, G.; Pascal, T.; Candresse, T.; Decroocq, V. Quantitative resistance to Plum pox virus in *Prunus davidiana* P1908 linked to components of the eukaryotic translation initiation complex. *Plant. Pathol.* **2009**, *58*, 425–435. [[CrossRef](#)]
15. Mariette, S.; Wong Jun Tai, F.; Roch, G.; Barre, A.; Chague, A.; Decroocq, S.; Groppi, A.; Laizet, Y.; Lambert, P.; Tricon, D.; et al. Genome-wide association links candidate genes to resistance to Plum pox virus in apricot (*Prunus armeniaca*). *New Phytol.* **2016**, *209*, 773–784. [[CrossRef](#)]
16. Pilarova, P.; Marandel, G.; Decroocq, V.; Salava, J.; Krška, B.; Abbott, A.G. Quantitative trait analysis of resistance to Plum pox virus in the apricot F1 progeny ‘Harlayne’ × ‘Vestar’. *Tree Genet. Genomes* **2010**, *6*, 467–475. [[CrossRef](#)]
17. Soriano, J.M.; Vera-Ruiz, E.; Vilanova, S.; Martínez-Calvo, J.; Llácer, G.; Badenes, M.L.; Romero, C. Identification and mapping of a locus conferring Plum pox virus resistance in two apricot-improved linkage maps. *Tree Genet. Genomes* **2008**, *4*, 391–402. [[CrossRef](#)]

18. Soriano, J.M.; Domingo, M.L.; Zuriaga, E.; Romero, C.; Zhebentyayeva, T.; Abbott, A.G.; Badenes, M.L. Identification of simple sequence repeat markers tightly linked to Plum pox virus resistance in apricot. *Mol. Breed.* **2012**, *30*, 1017–1026. [[CrossRef](#)]
19. Zhebentyayeva, T.N.; Reighard, G.L.; Lalli, D.; Gorina, V.M.; Krška, B.; Abbott, A.G. Origin of resistance to *Plum pox virus* in apricot: What new AFLP and targeted SSR data analyses tell. *Tree Genet. Genomes* **2008**, *4*, 403–417. [[CrossRef](#)]
20. Zuriaga, E.; Soriano, J.M.; Zhebentyayeva, T.; Romero, C.; Dardick, C.; Cañizares, J.; Badenes, M.L. Genomic analysis reveals MATH gene(s) as candidate(s) for Plum pox virus (PPV) resistance in apricot (*Prunus armeniaca* L.). *Mol. Plant Pathol.* **2013**, *14*, 663–677. [[CrossRef](#)]
21. Zuriaga, E.; Romero, C.; Blanca, J.M.; Badenes, M.L. Resistance to Plum pox virus (PPV) in apricot (*Prunus armeniaca* L.) is associated with down-regulation of two MATHd genes. *BMC Plant Biol.* **2018**, *18*, 25. [[CrossRef](#)]
22. Rodamilans, B.; Valli, A.; García, J.A. Molecular plant-Plum pox virus interactions. *Mol. Plant. Microbe Interact.* **2020**, *33*, 6–17. [[CrossRef](#)]
23. Rubio, M.; Ruiz, D.; Egea, J.; Martínez-Gómez, P.; Dicenta, F. Opportunities of marker assisted selection for Plum pox virus resistance in apricot breeding programs. *Tree Genet. Genomes* **2014**, *10*, 513–525. [[CrossRef](#)]
24. Decroocq, S.; Chague, A.; Lambert, P.; Roch, G.; Audergon, J.M.; Geuna, F.; Chiozzotto, R.; Bassi, D.; Dondini, L.; Tartarini, S.; et al. Selecting with markers linked to the *PPVres* major QTL is not sufficient to predict resistance to Plum pox virus (PPV) in apricot. *Tree Genet. Genomes* **2014**, *10*, 1161–1170. [[CrossRef](#)]
25. Moustafa, T.A.; Badenes, M.L.; Martínez-Calvo, J.; Llácer, G. Determination of resistance to sharka (*plum pox*) virus in apricot. *Sci. Hort.* **2001**, *91*, 59–70. [[CrossRef](#)]
26. Lommel, S.A.; McCain, A.H.; Morris, T.J. Evaluation of indirect-linked immunosorbent assay for the detection of plant viruses. *Phytopathology* **1982**, *72*, 1018–1022. [[CrossRef](#)]
27. Wetzel, T.; Candresse, T.; Ravelonandro, M.; Dunez, J. A polymerase chain reaction assay adapted to plum pox potyvirus detection. *J. Virol. Methods* **1991**, *33*, 355–365. [[CrossRef](#)]
28. Passaro, M.; Geuna, F.; Bassi, D.; Cirilli, M. Development of a high-resolution melting approach for reliable and cost-effective genotyping of *PPVres* locus in apricot (*P. armeniaca*). *Mol. Breed.* **2017**, *37*, 74. [[CrossRef](#)]
29. Ru, S.; Main, D.; Evans, K.; Peace, C. Current applications, challenges, and perspectives of marker-assisted seedling selection in Rosaceae tree fruit breeding. *Tree Genet. Genomes* **2015**, *11*, 8. [[CrossRef](#)]
30. Doyle, J.J.; Doyle, J.L. A rapid isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **1987**, *19*, 11–15.
31. Muñoz-Sanz, J.V.; Zuriaga, E.; Badenes, M.L.; Romero, C. A disulfide bond A-like oxidoreductase is a strong candidate gene for self-incompatibility in apricot (*Prunus armeniaca*) pollen. *J. Exp. Bot.* **2017**, *68*, 5069–5078. [[CrossRef](#)]
32. Haas, B.J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P.D.; Bowden, J.; Couger, M.B.; Eccles, D.; Li, B.; Lieber, M.; et al. De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nat. Protoc.* **2013**, *8*, 1494–1512. [[CrossRef](#)]
33. Verde, I.; Jenkins, J.; Dondini, L.; Micali, S.; Pagliarani, G.; Vendramin, E.; Paris, R.; Aramini, V.; Gazza, L.; Rossini, L.; et al. The Peach v2.0 release: High-resolution linkage mapping and deep resequencing improve chromosome-scale assembly and contiguity. *BMC Genom.* **2017**, *18*, 225. [[CrossRef](#)] [[PubMed](#)]
34. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with bowtie 2. *Nat. Meth.* **2012**, *9*, 357–359. [[CrossRef](#)] [[PubMed](#)]
35. Thorvaldsdóttir, H.; Robinson, J.T.; Mesirov, J.P. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* **2013**, *14*, 178–192. [[CrossRef](#)] [[PubMed](#)]
36. Peace, C. DNA-informed breeding of rosaceous crops: Promises, progress and prospects. *Hortic. Res.* **2017**, *4*, 17006. [[CrossRef](#)]
37. Edge-Garza, D.A.; Luby, J.J.; Peace, C. Decision support for cost-efficient and logistically feasible marker-assisted seedling selection in fruit breeding. *Mol. Breed.* **2015**, *35*, 223. [[CrossRef](#)]
38. Audergon, J.M.; Blanc, A.; Gilles, F.; Broquaire, J.M.; Clauzel, G.; Gouble, B.; Grotte, M.; Reich, M.; Bureau, S.; Pitiot, C. New recent selections issued from INRA’s apricot breeding programme. *Acta Hort.* **2010**, *862*, 179–182. [[CrossRef](#)]

39. Bassi, D.; Audergon, J.M. Apricot breeding: Update and perspectives. *Acta Hortic.* **2006**, *701*, 279–294. [CrossRef]
40. Bassi, D.; Bellini, E.; Guerriero, R.; Monastra, F.; Pennone, F. Apricot breeding in Italy. *Acta Hortic.* **1995**, *384*, 47–54. [CrossRef]
41. Egea, J.; Dicenta, F.; Burgos, L.; Martínez-Gómez, P.; Rubio, M.; Campoy, J.A.; Ortega, E.; Patiño, J.L.; Nortes, L.; Molina, A.; et al. New apricot cultivars from CEBAS-CSIC (Murcia, Spain) breeding programme. *Acta Hortic.* **2010**, *862*, 113–118. [CrossRef]
42. Martínez-Calvo, J.; Font, A.; Llácer, G.; Badenes, M.L. Apricot and Peach breeding programs from the IVIA. *Acta Hortic.* **2009**, *814*, 185–188. [CrossRef]
43. Gürçan, K.; Ceylan, A. Strain identification and sequence variability of Plum pox virus in Turkey. *Turkish J. Agric.* **2016**, *40*, 746–760. [CrossRef]
44. Decroocq, S.; Cornille, A.; Tricon, D.; Babayeva, S.; Chague, A.; Eyquard, J.P.; Karychev, R.; Dolgikh, S.; Kostriysna, T.; Li, S.; et al. New insights into the history of domesticated and wild apricots and its contribution to Plum pox virus resistance. *Mol. Ecol.* **2016**, *25*, 4712–4729. [CrossRef] [PubMed]
45. Dosba, F.; Orliac, S.; Dutranoy, F.; Maison, P.; Massonie, G.; Audergon, J.M. Evaluation of resistance to Plum pox virus in apricot trees. *Acta Hortic.* **1992**, *309*, 211–219. [CrossRef]
46. Karayiannis, I.; Mainou, A. Resistance to Plum pox virus in apricots. *EPPO Bull.* **1994**, *24*, 761–766. [CrossRef]
47. Martínez-Gómez, P.; Rubio, M.; Dicenta, F. Evaluation of resistance to Plum pox virus of North American and European apricot cultivars. *HortScience* **2003**, *38*, 568–569. [CrossRef]
48. Egea, J.; Rubio, M.; Campoy, J.A.; Dicenta, F.; Ortega, E.; Nortes, M.D.; Martínez-Gómez, P.; Molina, A.; Molina, A., Jr.; Ruiz, D. ‘Mirlo Blanco’, ‘Mirlo Anaranjado’, and ‘Mirlo Rojo’: Three new very early-season apricots for the fresh market. *HortScience* **2010**, *45*, 1893–1894. [CrossRef]
49. Passaro, M. Cost-Effective Use of Molecular Markers in the Practical Resolution of Common Horticultural Challenges. Ph.D. Thesis, Agriculture, Environment and Bioenergy in Università Degli Studi Di Milano, Milan, Italy, 2016.
50. Krška, B.; Salava, J.; Polák, J. Breeding for resistance: Breeding for Plum pox virus resistant apricots (*Prunus armeniaca* L.) in the Czech Republic. *EPPO Bull.* **2006**, *36*, 330–331. [CrossRef]
51. Krška, B.; Salava, J.; Polák, J.; Kominek, P. Genetics of resistance to Plum pox virus in apricot. *Plant Protect. Sci.* **2002**, *38*, 180–182. [CrossRef]
52. Krška, B.; Vachun, Z.; Nečas, T.; Ondrášek, I. New Sharka resistant apricots at the Horticultural Faculty in Lednice. *Acta Hortic.* **2015**, *1063*, 105–110. [CrossRef]
53. Rankovic, M.; Duli-Markovic, I.; Paunovic, S. Sharka virus in apricot and its diagnosis. *Acta Hortic.* **1999**, *488*, 783–786. [CrossRef]
54. CEP INNOVATION Website. Available online: <https://cepinnovation-novadi.com/variete/anegat/> (accessed on 29 July 2020).
55. Milatović, D.; Nikolić, D.; Krška, B. Testing of self-(in)compatibility in apricot cultivars from European breeding programmes. *HortScience* **2013**, *40*, 65–71. [CrossRef]
56. Brans, Y. Evaluation de la sensibilité de cultivars d’abricotier à la Sharka en zone confinée. Présentation de l’essai Ctifl 2012–2015. In Proceedings of the Rencontres Phytosanitaires Ctifl/DGAL—SDQPV Fruits à Noyau Ctifl Balandran, Bellegarde, France, 16 October 2016.
57. Babini, A.R.; Vicchi, V.; Missere, D. L’esame delle cultivar tolleranti alla sharka. *Ermes Agricoltura* **2010**. Available online: <http://www.crpv.it/doc/549738/DLFE-9612.pdf> (accessed on 29 August 2020).
58. Finn, C.E.; Clark, J.R. Register of New Fruit and Nut Cultivars List 44. *HortScience* **2008**, *43*, 1321–1343. [CrossRef]
59. Krška, B.; Vachun, Z. Apricot Breeding at the Faculty of Horticulture in Lednice. *Agronomy* **2016**, *6*, 27. [CrossRef]
60. Krška, B. Genetic resources of apricot for adaptability improvement and breeding. *Acta Hortic.* **2010**, *862*, 203–208. [CrossRef]
61. Gürçan, K.; Çetinsığ, N.; Pınar, H.; Macit, T. Molecular and biological assessment reveals sources of resistance to Plum pox virus—Turkey strain in Turkish apricot (*Prunus armeniaca*) germplasm. *Sci. Hortic.* **2019**, *252*, 348–353. [CrossRef]

62. Drogoudi, P. ΕΥΠΑΘΕΙΑ ΠΟΙΚΙΛΙΩΝ ΒΕΡΙΚΟΚΙΑΣ ΣΤΟΝ ΙΟ ΤΗΣ ΕΥΛΟΓΙΑΣ ΤΗΣ ΔΑΜΑΣΚΗΝΙΑΣ (PPV) (Sensitivity of apricot varieties against PPV). II. Δρογούδη 28ο Συνέδριο Ελληνικής Εταιρείας πιστήμης Οπωροκηπευτικών '50 χρόνια από την ίδρυση της ΕΕΕΟ'. In Proceedings of the 28th Conference of the Hellenic Fruit and Vegetable Credit Society '50 Years Since the Founding of ΕΕΕΟ', Thessaloniki, Greece, 16–20 October 2017.
63. Goffreda, J.C.; Voordeckers, A.; Butenis-Vorsa, L.; Cowgill, W.P., Jr.; Maletta, M.H.; Frecon, J.L. NJA53 apricot. *HortScience* **1995**, *30*, 389–390. [[CrossRef](#)]
64. Faggioli, F.; Barba, M. Screening of stone fruit germplasm for resistance to Plum Pox Potyvirus. *Adv. Hortic. Sci.* **1996**, *10*, 91–94.
65. Poggi Pollini, C.; Bianchi, L.; Babini, A.; Vicchi, V.; Liverani, A.; Brandi, F.; Giunchedi, L.; Autonell, C.; Ratti, C. Evaluation of Plum pox virus infection on different stone fruit tree varieties. *J. Plant. Pathol.* **2008**, *90*, S27–S31.
66. Brooks, R.M.; Olmo, H.P. *The Brooks and Olmo Register of Fruit and Nut Varieties*, 3rd ed.; ASHS Press: Alexandria, VA, USA, 1997.
67. Leccese, A.; Bartolini, S.; Viti, R. Genotype harvest season, and cold storage influence on fruit quality and antioxidant properties of apricot. *Int. J. Food Prop.* **2012**, *15*, 864–879. [[CrossRef](#)]
68. Fuchs, E.; Grünzig, M.; Kegler, H. Investigation on the Plum pox virus resistance in different apricot genotypes. *Acta Virol.* **1998**, *42*, 222–225.
69. Polák, J.; Oukropec, I.; Krška, B.; Pívalová, J.; Miller, W. Difference in reactions of apricot and peach cultivars to Plum pox virus: Serological and symptomatological evaluation. *HortScience* **2003**, *30*, 129–134. [[CrossRef](#)]
70. Audergon, J.M.; Dosba, F.; Karayiannis, I.; Dicenta, F. Amélioration de l'abricotier pour la résistance à la sharka. *EPPO Bull.* **1994**, *24*, 741–748. [[CrossRef](#)]
71. Mesarović, J.; Trifković, J.; Tosti, T.; Akšić, M.F.; Milatović, D.; Ličina, V.; Milojković-Opšena, D. Relationship between ripening time and sugar content of apricot (*Prunus armeniaca* L.) kernels. *Acta Physiol. Plant.* **2018**, *40*, 157. [[CrossRef](#)]
72. Krška, B.; Oukropec, I.; Polak, J.; Kominek, P. The evaluation of apricot (*Prunus armeniaca* L.) cultivars and hybrids resistant to Sharka. *Acta Hortic.* **2000**, *538*, 143–146. [[CrossRef](#)]
73. Salazar, J.A.; Rubio, M.; Ruiz, D.; Tartarini, S.; Martínez-Gómez, P.; Dondini, L. SNP development for genetic diversity analysis in apricot. *Tree Genet. Genomes* **2015**, *11*, 15. [[CrossRef](#)]
74. Syrgiannidis, G.; Mainou, A. Two new apricot varieties resistant to Sharka (Plum pox virus) disease created by crossing. In *Agriculture, Proceedings of the Programme de recherche Agrimed. Deuxiemes Rencontres sur L'abricotier, Avignon, France, 27–31 May 1991*; CEC Commission of the European Communities: Luxembourg, 1993.
75. Egea, J.; Ruiz, D.; Dicenta, F.; Burgos, L. Murciana apricot. *HortScience* **2005**, *40*, 254–255. [[CrossRef](#)]
76. Ledbetter, C.A.; Peterson, S.J. 'Apache' and 'Kettleman': Two early season apricots for the fresh market. *HortScience* **2005**, *40*, 2202–2203. [[CrossRef](#)]
77. Trandafirescu, M.; Dumitru, L.M.; Trandafirescu, I. Evaluating the Resistance to the Plum pox virus of Some Apricot Tree Cultivars and Hybrids in South-Eastern Romania. *Proc. Latvian Acad. Sci.* **2013**, *67*, 203–206. [[CrossRef](#)]
78. Egea, J.; Ruiz, D.; Martínez-Gómez, P. Influence of rootstock on the productive behaviour of 'Orange Red' apricot under Mediterranean conditions. *Fruits* **2004**, *59*, 367–373. [[CrossRef](#)]
79. Karayiannis, I.; Di Terlizzi, B.; Audergon, J.M. Susceptibility of apricot cultivars to Plum pox virus. *Acta Hortic.* **1999**, *488*, 753–760. [[CrossRef](#)]
80. Halász, J. Molecular Background of the S-locus Controlled Self-Incompatibility in Apricot. Ph.D. Thesis, Department of Genetics and Plant Breeding, Corvinus University, Budapest, Hungary, 2007.
81. Ledbetter, C.A.; Ramming, D.W. Apricot cv. Robada. United. States Patent USPP9890P, 13 May 1997.
82. Karayiannis, I.; Ledbetter, C.A. Susceptibility of certain apricot and plumcot cultivars to Plum pox virus infection. *Acta Hortic.* **2009**, *825*, 153–156. [[CrossRef](#)]
83. Egea, J.; Dicenta, F.; Burgos, L. 'Rojo Pasi6n' Apricot. *HortScience* **2004**, *39*, 1490–1491. [[CrossRef](#)]
84. Egea, J.; Dicenta, F.; Martínez-Gómez, P.; Burgos, L. Selene apricot. *HortScience* **2004**, *39*, 192–1493. [[CrossRef](#)]
85. Syrgiannidis, G. Selection of two apricot varieties resistant to Sharka virus. *Acta Phytopathol. Acad. Sci. Hung.* **1980**, *15*, 85–87. [[CrossRef](#)]
86. Adascalului, M.; Hoza, D.; Ion, L. Behaviour study for pollination a Romanian apricot varieties using different source of resistance to Sharka. *J. Hortic. For. Biotechnol.* **2014**, *18*, 13–17.

87. Ion, L.; Asănică, A.; Moale, C. Studies of resistance to Sharka in several Romanian apricot progenies. In Proceedings of the International Conference on Chemical, Agricultural and Biological Sciences (CABS-2015), Istanbul, Turkey, 4–5 September 2015.
88. Elibüyük, S.; Erdiller, G. The susceptibility of some apricot and plum varieties to Plum pox (sharka) virus. *Acta Hort.* **1995**, *384*, 549–552. [[CrossRef](#)]
89. Rodríguez, J.; Andrés, V.; Gil, L.; Martínez, J.; Hita, I. *Sensibilidad a Sharka en Variedades de Albaricoquero de Murcia*; Frutales Hueso Fund. La Caixa: Barcelona, Spain, 1995; pp. 56–64.
90. Dosba, F.; Lansac, M.; Maison, P.; Massonie, G.; Audergon, J.M. Tolerance to Plum pox virus in apricot. *Acta Hort.* **1988**, *235*, 275–281. [[CrossRef](#)]
91. Dondini, L.; Lain, O.; Geuna, F.; Banfi, R.; Gaiotti, F.; Tartarini, S.; Bassi, D.; Testolin, R. Development of a new SSR-based linkage map in apricot and analysis of synteny with existing *Prunus* maps. *Tree Genet. Genomes* **2007**, *3*, 239–249. [[CrossRef](#)]
92. Martínez-Gómez, P.; Dicenta, F. Evaluation of resistance of apricot cultivars to a Spanish isolate of plum pox potyvirus (PPV). *Plant. Breed.* **2000**, *119*, 179–181. [[CrossRef](#)]
93. Karayiannis, I. Susceptibility of apricots cultivars to Plum pox virus in Greece. *Acta Hort.* **1989**, *235*, 271–274. [[CrossRef](#)]
94. Balan, V.; Stoian, E. Susceptibility of certain apricot-tree to the Plum pox virus pathogenic action. *Acta Hort.* **1995**, *384*, 565–569. [[CrossRef](#)]
95. Avinent, L.; Hermoso de Mendoza, A.; Llácer, G.; García, S. Transmisión del virus de la sharka y sensibilidad varietal en albaricoquero. In Proceedings of the II Congreso Ibérico Ciencias Hortícolas, Zaragoza, Spain, 27–30 April 1993.
96. Maghuly, F.; Borroto-Fernandez, E.; Ruthner, S.; Pedryc, A.; Laimer, M. Microsatellite variability in apricots (*Prunus armeniaca* L.) reflects their geographic origin and breeding history. *Tree Genet. Genomes* **2005**, *1*, 151–165. [[CrossRef](#)]
97. Audergon, J.M.; Morvan, G.; Dicenta, F.; Chastelliere, G.; Karayiannis, I. A method to determine the susceptibility of apricot to Plum pox virus. *Acta Hort.* **1995**, *384*, 575–579. [[CrossRef](#)]
98. Egea, J.; Burgos, L. Detecting cross-incompatibility of three North American apricot cultivars and establishing the first incompatibility group in apricot. *J. Am. Soc. Hort. Sci.* **1996**, *121*, 1002–1005. [[CrossRef](#)]
99. McLaren, J. Apricot tree, ‘F168 cv’. United. States Patent USPP16071P2, 25 October 2005.
100. Dicenta, F.; Audergon, J.M. Localization of Plum pox virus (PPV) in tissues of susceptible and resistant apricot cultivars. *Phytopathol. Med.* **1995**, *34*, 83–87.
101. Lachkar, A.; Mlika, M. New apricot varieties selected from the Tunisian breeding programme. *Acta Hort.* **2006**, *717*, 189–192. [[CrossRef](#)]
102. Dosba, F.; Denise, F.; Audergon, J.M.; Maison, P.; Massonie, G. Plum pox virus resistance of apricot. *Acta Hort.* **1991**, *293*, 569–580. [[CrossRef](#)]
103. Egea, J.; Campoy, J.A.; Dicenta, F.; Burgos, L.; Patiño, J.L.; Ruiz, D. ‘Estrella’ and ‘Sublime’ apricot cultivars. *HortScience* **2009**, *44*, 469–470. [[CrossRef](#)]
104. Eynard, A.; Roggero, P.; Lenzi, R.; Conti, M.; Milne, R.G. Test for pollen and seed transmission on Plum pox virus (Sharka) in two apricot cultivars. *Adv. Hort. Sci.* **1991**, *5*, 104–106.
105. Gallois, J.L.; Moury, B.; German-Retana, S. Role of the Genetic Background in Resistance to Plant Viruses. *Int. J. Mol. Sci.* **2018**, *19*, 2856. [[CrossRef](#)]
106. Polák, J.; Kominek, P.; Jokes, M.; Oukropec, I.; Krška, B. The evaluation of resistance of apricots to Plum pox virus by ELISA and ISEM. *Acta Hort.* **1995**, *386*, 285–289. [[CrossRef](#)]
107. Layne, R.E.C.; Hunter, D.M. ‘AC Harostar’ apricot. *HortScience* **2003**, *38*, 140–141. [[CrossRef](#)]
108. Hegedús, A.; Lénárt, J.; Halász, J. Sexual incompatibility in *Rosaceae* fruit tree species: Molecular interactions and evolutionary dynamics. *Biol. Plant* **2012**, *56*, 201–209. [[CrossRef](#)]
109. Austin, P.T. Pollination of Sundrop Apricot. Ph.D. Thesis, Massey University, Auckland, New Zealand, 1995.
110. Egea, J.; Ruiz, D.; Burgos, L. “Dorada” apricot. *HortScience* **2005**, *40*, 1919–1920. [[CrossRef](#)]
111. Muñoz-Sanz, J.V.; Zuriaga, E.; López, I.; Badenes, M.L.; Romero, C. Self-(in)compatibility in apricot germplasm is controlled by two major loci, *S* and *M*. *BMC Plant. Biol.* **2017**, *17*, 82. [[CrossRef](#)]
112. Corrin, A.A. “Ruby” Apricot Tree. United. States Patent USPP8177, 16 March 1993.
113. Zaiger, C.F. Apricot Tree (Spring Giant). United. States Patent USPP5138, 15 November 1983.

114. Dicenta, F.; Audergon, J.M. Inheritance of resistance to plum pox potyvirus (PPV) in 'Stella' apricot seedlings. *Plant. Breed.* **1998**, *117*, 579–581. [[CrossRef](#)]
115. Tartarini, S.; Sansavini, S.; Vinatzer, B.; Gennari, F.; Domizi, C. Efficiency of marker assisted selection (MAS) for the Vf scab resistance gene. *Acta Hortic.* **2000**, *538*, 549–552. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).





Article

# High Resolution Melting and Insertion Site-Based Polymorphism Markers for Wheat Variability Analysis and Candidate Genes Selection at Drought and Heat MQTL Loci

Rosa Mérida-García <sup>1,\*</sup>, Sergio Gálvez <sup>2</sup>, Etienne Paux <sup>3</sup>, Gabriel Dorado <sup>4</sup>, Laura Pascual <sup>5</sup>, Patricia Giraldo <sup>5</sup> and Pilar Hernandez <sup>1,\*</sup>

<sup>1</sup> Institute for Sustainable Agriculture (IAS-CSIC), Consejo Superior de Investigaciones Científicas (CSIC), Alameda del Obispo s/n, 14004 Córdoba, Spain

<sup>2</sup> Department of Languages and Computer Science, ETSI Informática, Campus de Teatinos, Universidad de Málaga, Andalucía Tech, 29071 Málaga, Spain; galvez@uma.es

<sup>3</sup> Université Clermont Auvergne, INRAE, GDEC, 5 Chemin de Beaulieu, 63000 Clermont-Ferrand, France; etienne.paux@inrae.fr

<sup>4</sup> Departamento de Bioquímica y Biología Molecular, Campus Rabanales C6-1-E17, Campus de Excelencia Internacional Agroalimentario (ceiA3), Universidad de Córdoba, 14071 Córdoba, Spain; bb1dopeg@uco.es

<sup>5</sup> Department of Biotechnology-Plant Biology, School of Agricultural, Food and Biosystems Engineering, Universidad Politécnica de Madrid, 28040 Madrid, Spain; laura.pascual@upm.es (L.P.); patricia.giraldo@upm.es (P.G.)

\* Correspondence: rmerida@ias.csic.es (R.M.-G.); phernandez@ias.csic.es (P.H.)

Received: 30 July 2020; Accepted: 27 August 2020; Published: 1 September 2020

**Abstract:** The practical use of molecular markers is facilitated by cost-effective detection techniques. In this work, wheat insertion site-based polymorphisms (ISBP) markers were set up for genotyping using high-resolution melting analysis (HRM). Polymorphic HRM-ISBP assays were developed for wheat chromosomes 4A and 3B and used for wheat variability assessment. The marker sequences were mapped against the wheat genome reference sequence, targeting interesting genes. Those genes were located within or in proximity to previously described quantitative trait loci (QTL) or meta-quantitative trait loci (MQTL) for drought and heat stress tolerance, and also yield and yield related traits. Eighteen of the markers used tagged drought related genes and, interestingly, eight of the genes were differentially expressed under different abiotic stress conditions. These results confirmed HRM as a cost-effective and efficient tool for wheat breeding programs.

**Keywords:** high resolution melting; ISBP markers; drought; candidate genes; QTL; MQTL; wheat variability

## 1. Introduction

Wheat is among the most important and widely grown crops worldwide [1] and one of the most important grain food crops in the human diet (<https://www.fao.org>). Wheat development and yield can be affected by abiotic stresses as drought [2–5] and heat [6,7], whose frequencies would be increased by the strong effects of the predicted climate change and global warming [8–10]. In fact, drought is considered one of the most limiting environmental factors [3,11,12], strongly affecting the growth [13,14] and production of crops, with significant reductions in the final yield of cereals, including wheat [13]. Heat stress usually affects crops during the post-anthesis period, with negative effects on final production [15] and end-use quality products [16]. These abiotic stresses are important challenges for plant research and breeders, and plant breeding efforts have been focused on the improvement of final crops production under these limiting conditions [17].

Drought tolerance is considered a complex [18,19] and quantitative trait [14], which interacts with the environment, and possesses an additive polygenic nature [3]. Even though most of these genes have minor contributions, they are of great importance in the genetic improvement for drought tolerance [5]. Drought can also affect gene expression [20,21], both under controlled [22–25] and field conditions [26,27]. Heat stress tolerance includes plant mechanisms which occur at several levels, where the plant acquires thermotolerance to cope with high temperatures [28]. Plants heat shock response results from the reprogramming of gene expression [29] and is of great interest for studies focused on plants stress tolerance and gene expression regulation [28]. Wheat breeding programmes are necessary to ensure an improved selection of favorable alleles focused on interesting agronomic traits, as yield and quality, and also biotic and abiotic stresses tolerance [30,31].

Molecular markers can be developed and successfully applied to identify important genomic regions and major genes [32] closely related to target traits as drought tolerance [3]. In addition, public resources as Wheat Expression (<http://www.wheat-expression.com>) facilitate the identification of interesting candidate genes, as well as their validation [22,27]. Recent advances in genomics, and the available fully annotated wheat reference genome (IWGSC RefSeq v1) [33], allow the accurate identification of marker positions and their chromosome locations [5]. The available gene models have been used, through appropriate bioinformatic pipelines, for the identification of differentially expressed genes during drought and heat stress treatments (i.e., [22,27]).

The insertion site-based polymorphism markers (ISBP) are PCR markers designed based on the knowledge of the sequence flanking transposable element (TE) sequences, to design one primer in the transposable element and the other in the flanking DNA sequence [34]. TEs are very abundant and nested in the wheat genome, with unique (genome-specific) insertion sites that are highly polymorphic [30]. ISBP were developed for wheat genomic and genetic studies [34,35], and later used in marker-assisted selection (MAS), and as a selecting tool for new varieties in plant breeding programs [30]. The ISBP technique is a rapid and efficient way to develop single copy chromosome-specific markers from incomplete genomic sequences [30,34]. ISBP markers represent a valuable source of polymorphism, which is mostly genome specific in wheat [35], and therefore very convenient for wheat mapping applications [36]. These markers have been used to improve the wheat genome saturation [30,33,37], for genotyping and single-nucleotide polymorphism (SNP) detection [35,38], genetic diversity assessment [39], micro RNA (miRNA) coding sequences identification [40], or sequence composition analysis [41]. They were also successfully applied to develop physical or genetic maps, and to locate important agronomic traits [42,43] or resistance genes [36,44].

There are different techniques to detect the ISBP markers, as fluorescent polymerase chain reaction (PCR) and capillary electrophoresis, allele-specific PCR or melting curve analysis [35]. High resolution melting analysis has been described as a versatile and powerful analytical tool in molecular biology, characterized by its easy use, simplicity, flexibility, low cost, sensitivity, and specificity [45–47]. Briefly, this technique is based on the analysis of the PCR product's melting, by analyzing the fluorescence (due to an intercalated dye) broadcast level as response of a specific increasing temperature ramp [48]. ISBP was firstly carried out by Paux et al. [30] using wheat chromosome 3B markers and melting curve analysis as an alternative to agarose gel electrophoresis. ISBP marker detection using HRM analysis was later performed for resistance loci assessment in bread wheat [49]. The HRM technique has also been used to detect other molecular markers (i.e., SNP, expressed-sequence tag (EST), simple-sequence repeat (SSR), or insertion-deletion (InDeL)). HRM applications in wheat and closely related species as barley and *Aegilops*, include the characterization of InDeL and SNP markers involved in drought and salt tolerance [50]; the detection SNP markers [51,52] and mutations [53–56]; and the mapping of markers linked to resistance genes [57–59].

ISBP markers have been developed for all wheat chromosomes [33,35,40,41,43]. Chromosome 3B [30] contains loci for grain yield, kernel length, plant height, and related traits [60]. Chromosome 4A [33,61] harbors several QTLs related to biotic and abiotic stresses tolerance, agronomic traits as grain yield and quality, and regulation of physiological traits as plant height, maturity, or dormancy [62–70].

This chromosome represents an important target in plant breeding, marker design for variability analysis and candidate genes assessment. In this study, ISBP markers from wheat chromosome 3B [30] and 4A were used to develop and validate HRM assays, and to assess the genetic variability in a wheat collection. These markers were also used to target meta-quantitative trait loci (MQTL) [71,72] related to drought and heat stresses, as well as yield and yield related traits. Candidate gene analyses were performed for the ISBP markers and the genes were validated by gene expression analyses carried out among different drought and heat stress conditions.

## 2. Material and methods

### 2.1. Plant Material and DNA Isolation

Two wheat panels were used to perform the analyses: (i) panel 1: a collection of 62 wheat lines (37 *Triticum aestivum* L., 11 *Triticum turgidum* ssp. *durum* (Desf.) Husn., 11 *Triticum monococcum* L., 2 *Triticum turgidum* ssp. *turgidum* L. and one *Triticum urartu* Thumanian ex Gandilyan) from different sources (Supplementary Materials Table S1); (ii) panel 2: a collection of 76 durum wheat (*Triticum turgidum* L) landraces, provided by the Spanish National Plant Genetic Resources Center (CRF-INIA) (Supplementary Materials Table S2). This panel comprised genotypes of three subspecies: 8 *dicoccon* (Schrank) Thell., 21 *T. turgidum*, and 45 *durum* (Desf.) Husn.

Genomic DNA was isolated from young leaf tissue according to the cetyl trimethyl ammonium bromide (CTAB) method of Murray and Thompson [73], as optimized by Hernández et al. [74]. The quality and concentration of samples were assessed by electrophoresis in a 0.8% agarose gel.

### 2.2. Insertion Site-Based Polymorphism Markers Development

ISBP markers were initially developed based on the wheat chromosome 4A survey sequencing [61] and confirmed in the bread wheat reference genome sequence RefSeq v1 [33].

The assemblies corresponding to the 4A wheat chromosome survey sequencing were generated using the “Newbler v2.7” software package (Roche Diagnostics Corporation, Basel, Switzerland) using default parameters. IsbpFinder [30] was run on the assemblies obtained, and ISBP markers were located on the 4AS and 4AL chromosome arms assemblies. The corresponding 45 ISBP primers were designed using Primer3 (<http://primer3.sourceforge.net>) and mapped to the bread wheat reference genome RefSeq v1 [33]. Marker set up was carried out using 6 durum and bread wheat lines representative of variability (Supplementary Materials Table S1).

ISBP amplicons obtained by standard PCR (55 °C annealing, [30]) using 5 *T. aestivum* varieties (Supplementary Materials Table S1) were purified by Exonuclease I (Exo I, New England Biolabs, Inc., Ipswich, MA, USA) and SAP treatment (5 µL DNA, 1U ExoI, 1xSAP buffer, 1U SAP in 9 µL at 37 °C for 1 h). The purified fragments were then sequenced on an ABI PRISM® 3730XL (Applied Biosystems, Foster City, CA, USA) genetic analyzer using the forward and reverse ISBP primers [30] and using the ABIPRISM BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA). HRM analyses were carried out for the same 5 *T. aestivum* varieties (Supplementary Materials Table S1) using 6 ISBP primer pairs previously developed for wheat chromosome 3B [30]: *HRM3B\_273339424*, *HRM3B\_609364064*, *HRM3B\_124761338*, *HRM3B\_203288704*, *HRM3B\_465802537*, and *HRM3B\_331497483*. This analysis was carried out in a Rotorgene™ 6000, model 5-Plex real time PCR (Corbett Research, Mortlake, NSW, Australia). The PCR reaction volume was of 15 µL and the mixture composition was: Master Mix of Type-it HRM PCR Kit (Qiagen, CA, USA); 0.7 µM of each primer and 2 µL of genomic DNA (30 ng). The PCR protocol consisted on an initial denaturation step at 95 °C for 10 min; 45 amplification cycles of denaturation at 95 °C for 10 s, annealing at 55 °C for 15 s and a final extension at 72 °C for 20 s; the high-resolution melting was set out by ramping from 65 to 95 °C, with fluorescence data acquisition at 0.1 °C increments (waiting for 2 s every acquisition). HRM results were compared with the amplicon sequencing for the same five samples.

The high-resolution melting analysis was later performed to assign HRM pattern types to the ISBP markers, using 19 durum and bread wheat lines (Supplementary Materials Table S1). The PCR protocol consisted of an initial denaturation step at 95 °C for 5 min; 50 cycles of denaturation at 95 °C for 20 s, annealing at 60 °C for 20 s, and a final extension at 72 °C for 20 s. HRM analysis was undertaken once amplification was completed by ramping from 65 to 95 °C, with fluorescence data acquisition at 0.1 °C increments (waiting for 2 s every acquisition). Results were analyzed by using the Rotorgene software version 1.7 (Qiagen, the Netherlands), and HRM curves were normalized according to the manufacturer's instructions.

### 2.3. Candidate Genes and Gene Expression Analyses

ISBP markers were mapped in wheat chromosomes 4A and 3B, and then were compared to the wheat MQTLs described in Acuña-Galindo et al. [71]. To obtain the MQTLs positions in wheat chromosomes, marker sequences [71] were extracted from "Graingenes" (graingenes.org) and "Wheat SSR DB markers" (wheatssr.nig.ac.jp), and then located by mapping flanking markers (using BLAST) against RefSeq v1 [33]. Only the markers with a corresponding amplicon shorter than 500 bp and a perfect BLAST match (no gap, no mismatch) were considered.

Markers sequences in chromosomes 4A and 3B were blasted against the RefSeq1 v1 [33] using the parameters "-task", "blastn-short" and "-ungapped". The resulting hits were then processed to pair forward and reverse sequence hits with an amplicon <1000 base pairs (bp). For subsequent analysis, paired sequences were ordered by number of mismatches, so markers position was inferred from the position of pairs with lower number of mismatches (0 in most cases). To identify candidate genes associated to each marker, the results were filtered and the hits with best e-value were selected for each molecular marker. The candidate genes were manually selected within a window of +/-20 kb of the marker's hit in the pseudomolecule [33] gene model annotation. Due to the reduced gene density found for wheat chromosome 3B ISBP markers (only three genes were found), that window was extended to +/-300 kb for this chromosome. Genes described as "uncharacterized protein" were then manually annotated. Sequences were obtained in Ensembl Plants (*T. aestivum* RefSeq v1.1) (<https://plants.ensembl.org/>), and then searched in UniProt (<https://www.uniprot.org/>). The annotated hits with e-value 0.0 and a score >2000 were selected, except for the gene *TraesCS4A01G410700*, which possesses a short length (207aa).

To overview the results from gene expression analyses, heatmaps were drawn using the data retrieved from Wheat Expression ([www.wheat-expression.com/](http://www.wheat-expression.com/)) and the 'NMF 0.21.0' R package [75]. The information used was generated by Liu et al. [22], Ma et al. [26], and Galvez et al. [27]. Liu et al. [22] experimental seedling samples grown in controlled conditions were associated to NCBI SRA ID SRP045409 (control (IS), heat and drought (PEG induced drought) stress for 1 and 6 h (PEG1 and PEG6), respectively). Ma et al. [26] experimental samples grown in a shelter corresponded to NCBI SRA ID SRP102636 (anther stage irrigated leaf phenotype (AD\_C), anther stage drought-stressed leaf phenotype (AD\_S), tetrad stage irrigated developing spike phenotype (T\_C), and tetrad stage drought-stressed developing spike phenotype (T\_S)). Galvez et al. [27] flag leaf samples from field grown plants used have NCBI SRA ID SRP119300 (irrigated (IF), mild stress (MS), and severe stress (SS) flag leaf samples). Transcripts Per Kilobase Millions (TPMs) of genes under every condition were calculated as mean value of TPMs of its constitutive experiments. A differential gene expression (DGE) analysis was performed using RevSeqv1 [33] gene models through two bioinformatic pipelines: Kallisto (version 0.43.0) with the R library "sleuth" (version 0.28.1), and STAR with the R library DESeq2 (version 1.14.1). A consensus threshold for the two pipelines ( $|\log_2FC$ ,  $\beta$ ) > 1 and *p*-adjust, *Q*-value < 0.05 [27]) was used.

### 2.4. Wheat Variability Assessment

The PCR amplification protocol used was the previously described for HRM pattern type assessment. Samples genotyping was performed using Melt and HRM analysis options of the Rotor-Gene™ 6000 software. PCR was repeated three times to ensure the amplifications. Results

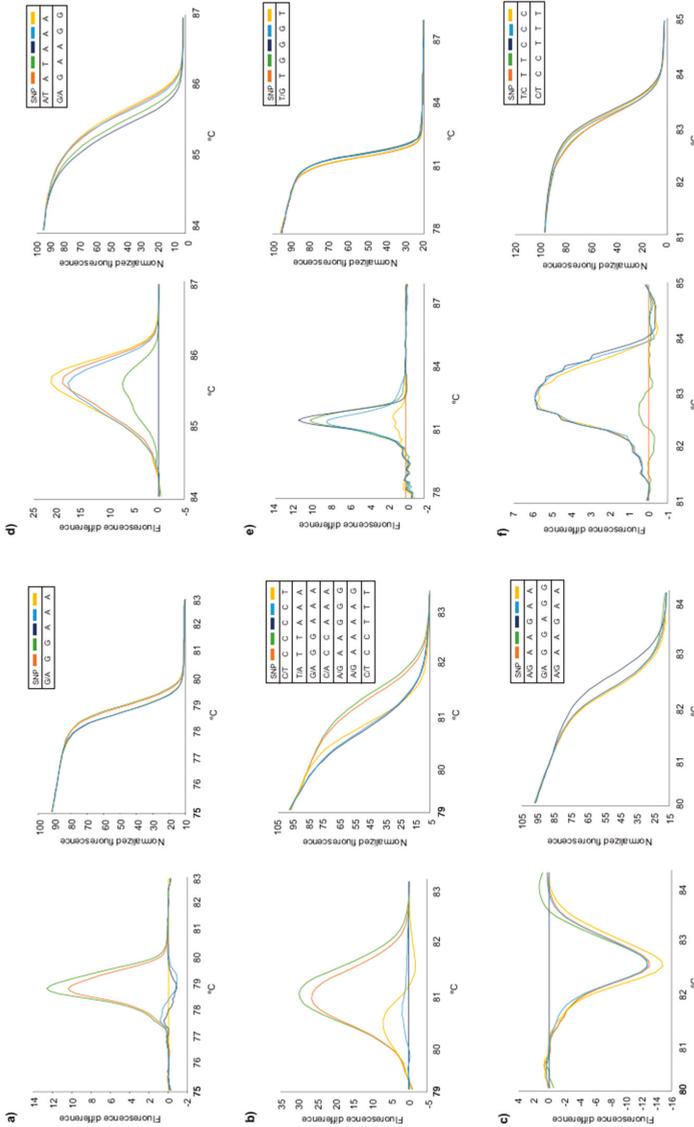
were then corroborated with Rotor-Gene™ ScreenClust HRM™ Software. In those cases where the number of genotypes assigned was unclear, ScreenClust HRM™ Software was also used for the final decision. A binary matrix for genotyping results was created and then analyzed using two different software, PhylTools v.1.32 (Wageningen Agricultural University, The Netherlands) and PowerMarker v.3.25 [76]. The first one was used to calculate the genetic distances for haploid data with “individuals” as hierarchy level and the Nei index [77]. PowerMarker was used to obtain the statistics mean allele number, mean gene diversity, and Polymorphism Index Content (PIC). The unweighted pair group method with arithmetic mean (UPGMA) clustering was carried out using the NEIGHBOR module in the Phylip 3.695 package [78] with default parameters. The selected tree method was UPGMA. The final dendrogram was drawn using MEGA v.6.0 [79] software with the results from the genotyping.

The goodness of fit of the UPGMA tree was calculated by the Cophenetic Correlation Coefficient (CCC) using a visual basic program for Microsoft Excel 2000 [80]. The CCC calculated from the linear regression between the corresponding values of the original distance matrix and the cophenetic matrix derived from the calculation of the UPGMA tree.

### **3. Results**

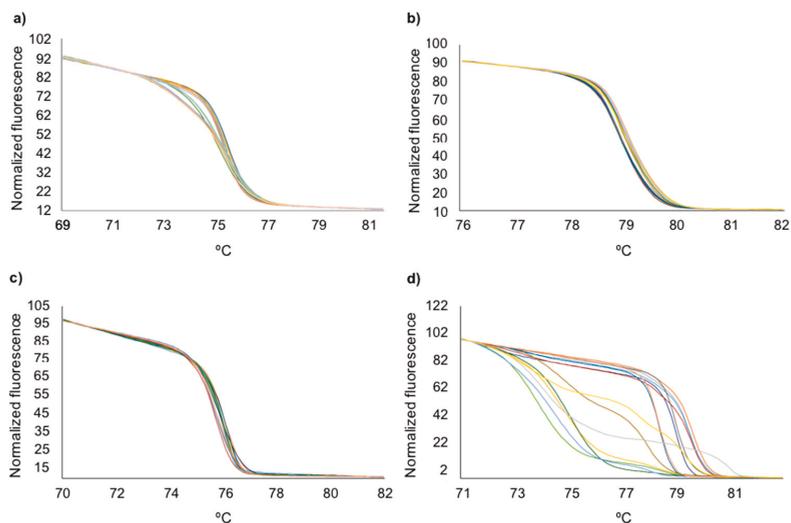
#### *3.1. Markers Sequence Validation and HRM Pattern Types Assignment*

To validate markers sequences, difference plots and HRM normalized curves from the high-resolution melting analyses were obtained for each of the six ISBP markers for wheat chromosome 3B and compared with their corresponding amplicon sequences. The HRM profiles were successfully validated. Sequence polymorphism from one to five nucleotides were detected by HRM analyses. Those included both transitions (examples are shown in Figure 1a,c,f), transversions (Figure 1e) and both (Figure 1b,d).



**Figure 1.** Sequence validation for insertion site-based polymorphisms (ISBP) markers using high-resolution melting (HRM) analysis. Normalized curves (on the right) and difference plots (on the left) are shown for 6 ISBP markers for wheat chromosome 3B, with different nucleotide variations: (a) marker *HRM3B\_273339424*—single nucleotide transition (A/G); (b) marker *HRM3B\_124761338*—five nucleotide variations (C/T, G/A, and A/G) and two transitions (T/A and C/A); (c) marker *HRM3B\_203288704*—three nucleotide transitions (A/G, G/A, and A/G); (d) marker *HRM3B\_465802537*—two nucleotide variations, one transition (A/T) and one transition (G/A); (e) marker *HRM3B\_609364064*—one transition (T/G); and (f) marker *HRM3B\_331497483*—two transitions (T/C and C/T).

The HRM pattern type assignment was based on the pattern of normalized high-resolution melting curves obtained, and their potential to genotype a high number of varieties. The ISBP HRM patterns were classified into four different types (Table 1, Figure 2): (i) pattern type A, excellent markers to genotype a high number of wheat varieties simultaneously. The HRM curves are very different and easily distinguishable into classes (Figure 2a); (ii) pattern type B, good markers to differentiate several groups of varieties (genotypes) in the same run. HRM curves can be differentiated in an easy way (Figure 2b); (iii) pattern type C, good HRM markers, but not recommended for genotyping a broad variety of samples. The differentiation between HRM curves is not clear in all cases (Figure 2c); and (iv) pattern type D, assigned to markers which are not recommended for HRM genotyping, due to a low amplification efficiency or to a gradient of HRM curves that hinders precise classification into classes (Figure 2d). Twelve ISBP markers were classified as HRM pattern type A; 4 markers showed HRM pattern type B; 10 markers had HRM pattern type C; and 13 were considered with pattern type D (Table 1). Four of the 45 wheat chromosome 4A ISBP primer pairs designed (Table 2), were discarded for the rest of analyses due to non-reliable PCR amplifications.



**Figure 2.** High resolution melting pattern types assessment for wheat chromosome 4A ISBP markers. Normalized HRM curves for 19 samples were shown for (a) marker *HRM4A\_67413676* (pattern type A); (b) marker *HRM4A\_618105078* (pattern type B); (c) marker *HRM4A\_317085557* (pattern type C); and (d) marker *HRM4A\_291420130* (pattern type D).

**Table 1.** Insertion site-based polymorphism markers developed for wheat chromosome 4A. ISBP primers selected for the wheat collection variability assessment are shown in bold.

Marker ID	Chr	Pattern	Forward Primer (5' → 3')	Reverse Primer (5' → 3')	Amp. Size
<b>HRM4A_2791416</b> *	4AS	A	TCCTACAAAAACGCTTATATTTTGG	GATCACITGCCAGTTGCATT	103
HRM4A_9618320	4AS	C	CGTCAGCTCAAAGGAAAAACC	GGGAGAAATTTGGCAGTC	151
HRM4A_73613394	4AS	D	GGTCTTGATGATGATGTG	CTTTGTAGCCCCATCTGAA	136
HRM4A_36371442	4AS	D	GCAITGGCTCATGTTCTTGG	TCAAAAACGCTTTATATATATGGGA	193
<b>HRM4A_3865455</b> *	4AS	A	TCTGAAAGAAGCTCAGCTTATTACTT	TTTGTAAAATTTGAGGGACCA	101
HRM4A_47108251	4AS	n/s	ACACGTGGGAGATATAGCCG	AAACCCTAGTCCCACTGCT	168
HRM4A_60660613	4AS	C	CATCTCTCAGCCCATGAAT	CAAAAATCAAACTTTCCGGTTTC	141
HRM4A_U1	-	n/s	AGAGCCCTTAAGTTTGGCTG	TCACTCATTTAGTCCCGGAAG	161
HRM4A_56502024	4AS	D	CAAACGATGCTCTCTGTCA	TCCATCTATCTGATTTCTGATATGAA	119
<b>HRM4A_67413676</b>	4AS	A	CGAGTCTAGCCAGTTCCTG	AAAAACATTCGATAAATGGATGG	109
HRM4A_59280888	4AS	B	CATCCACATGGATTTTCCAG	CGATTTGCTAGCCTAGGAGG	207
<b>HRM4A_141912346</b>	4AS	A	AGCAAAGCATGCAATGAAG	TTATCTCTGTAGGACCCGG	143
HRM4A_99034796	4AS	n/s	CCTCCGTTCCGAAATTACTTG	TTATGCAAGCAGCAGTTACC	157
<b>HRM4A_109848074</b> *	4AS	A	ATGAGACTTTTGGACGCCG	CAACGCTTTTTCAGACGGAGG	139
HRM4A_14111880	4AS	D	AGATCTGCTGGACATAAGCACA	CCCTCCGTCCTCCAAATAAGT	117
HRM4A_162423177	4AS	C	AAGCAAGAAGCGA A A A C A G C	AATCATCTAGTCGGTTGCGG	180
HRM4A_176868209	4AS	C	TTCCGAATTACTTGCTCGGAT	CACAGGCTCGGATAGGCTAC	191
HRM4A_183985221	4AS	C	TCCAGAAAATCTGATGGCACTG	AGATGGACGCGATAAGATGG	108
HRM4A_149049302	4AS	D	CCTCCGTTCCGAAATTACTTG	ATGAAAAGGCAGGCTAGACCA	223
HRM4A_617938526	4AL	C	CTTCGCTTTCCTCCGATAG	GTAAGGTAGTGAITCAAACGGTGT	113
<b>HRM4A_618105078</b>	4AL	C	AGTCATGGCAACCAACAACA	AGAGTTGCCGTGCCACTTAT	162
HRM4A_U2	-	n/s	TTTCTAAGGGTAAGGGCGT	TAGAGGGTTGTCTGGTTCC	100
<b>HRM4A_716986193</b> *	4AL	A	CTCCACCATAGATCGAAGCA	ACAGGACAATTTGGAGACCCA	159
HRM4A_U3	-	A	TTGAACCTGCCAAAACGTCICA	CTCCCTCTATAACCACCAATG	100
HRM4A_548541053	4AL	D	CGGTGCTAGATACATCCGTTT	AACCAAAGTAAAGCTACTAGAGAAA	112
<b>HRM4A_714743756</b>	4AL	A	CGCTAGTATAGTGTCAAAAACCG	ATGTAGGATGTCCTCCGCTC	117
HRM4A_U4 *	-	A	CAGACAATGTCAAAAACAACC	AAAAGATTCATGTACAAGGGGA	116
HRM4A_U5	-	n/s	TCCACCTATAAACACCCCG	TGTATTTCCAGGACGGGAAGG	202
HRM4A_460238681	4AL	C	GGTCTGATTGAAATCTTAAAAGTGC	CCTTTTGTCAGCCTGTGT	119
<b>HRM4A_583704598</b> *	4AL	A	CCCTGTAAAACATAATAGAAATGG	CTTGTTCCCTGTCTCTTG	165
HRM4A_317085557	4AS	C	ACATGGGTGACCCATATCCA	CGGACTGGTCCATTAGGTT	162
HRM4A_291420130	4AS	D	AAGTGGAAAAGGCACAATGC	CAAAACATTTGCCAACATGG	183
HRM4A_681664894	4AL	D	TATGCTTGATGCTCTGGCTG	CATCCATTTGAGCGGCAACT	121

Table 1. *Contd.*

Marker ID	Chr	Pattern	Forward Primer (5'→3')	Reverse Primer (5'→3')	Amp. Size
HRM4A_683608822	4A	C	TCAGTTTGGTCCCGTTGC	TGACACTACTAAAGTACTCCCCA	196
HRM4A_Ul-6	-	D	TTCCAAGAAAATGTTCCGAA	TCCTTCGTTCAAGACTCGCT	177
HRM4A_Ul-7	-	B	TCCCTAGCTGATGATTTGGG	ATAATAGCTCCATACCGGCC	107
<b>HRM4A_660524139 *</b>	4A	A	GATAAATCTAAGATAAGCTTTTGGG	GGGACACAATCTGATGGTGA	116
HRM4A_Ul-8	-	C	GGCCCTAGAAATGCAAAATGA	TTCCCACTCAATAACTGGG	127
HRM4A_Ul-9	-	B	TCITCACTGTTTCAGTCCG	AGACGAGCACACCGCATAC	131
<b>HRM4A_702156718</b>	4A	A	CCHTTGGCAACAACAATG	ATTGGCAGATTTCTAAGCG	133
HRM4A_Ul-10	-	D	AGCCGAGGAAGGTTACATA	TCCATTATAAACAATAAAGAGCG	119
HRM4A_Ul-11	-	A	GGTCAATGTATACTACCGTGC	ATCGGGAAACCAATGTTA	164
HRM4A_Ul-12	-	D	TGTTTGCTGAGGACCAACAG	CCGGGGTAACTCAATTTT	111
HRM4A_Ul-13	-	D	ATTGTTGTTCCGTTTTTGG	ACTCCCTCGTCCCAATAA	114
HRM4A_Ul-14	-	D	CCCTCTGTAAGAATAAAGACCGT	TGGATGCAGTAACTCGAAA	107

Chr: chromosome location; Pattern: high resolution melting pattern type (A: excellent marker to genotype different wheat varieties simultaneously; B: good marker to genotype several wheat varieties simultaneously; C: good marker but not recommended for a broad variety of wheat samples; D: marker not recommended for HRM genotyping; and *n/s* for markers non-suitable for HRM); Amp. size: amplicon size in base pairs (bp); \*: ISBP markers used in the durum wheat collection variability assessment.

Table 2. Wheat chromosome 4A ISBP markers linked to genes which alter their expression in response to water stress treatments. DE genes are shown in bold. The positive and negative values in "Dist (bp)" column indicate if the corresponding gene is downstream or upstream of the marker.

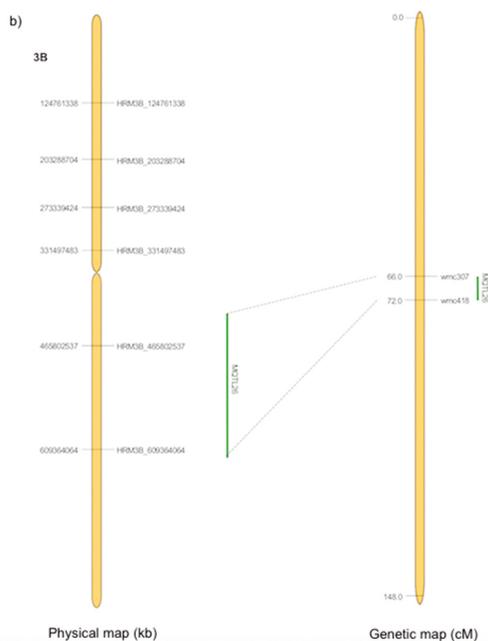
Marker ID	Chr	Dist (bp)	Gene ID	Description	Sum of TPMs
HRM4A_2791416	4A	-814	TraesCS4A01G003600	Alpha/beta-Hydrolases superfamily protein	8.79
		-5141	<b>TraesCS4A01G003500</b>	Thionin-like protein	124.0
HRM4A_36371442	4A	-2736	TraesCS4A01G043500	STAS domain-containing protein	47.32
HRM4A_38654555	4A	-9961	TraesCS4A01G047100	Activating signal cointegrator 1 complex subunit 2	12.70
		-14,885	<b>TraesCS4A01G047000</b>	Formin-like protein	4.87
HRM4A_56502024	4A	-417	TraesCS4A01G060200	BHLH domain-containing protein	10.54
		-5039	TraesCS4A01G060100	Uncharacterized protein	50.64
HRM4A_67413676	4A	975	<b>TraesCS4A01G069200</b>	Armaddillo repeat only	21.26
HRM4A_99034796	4A	-5609	TraesCS4A01G092400	SH3 domain-containing protein	43.46
HRM4A_109848074	4A	97	TraesCS4A01G098300	Xylosyltransferase 1	21.59
HRM4A_141111880	4A	-3308	TraesCS4A01G116200	MYND-type domain-containing protein	50.62
HRM4A_162423177	4A	9	TraesCS4A01G126300	Transcription factor TGA2.2 *	83.14
HRM4A_176868209	4A	-406	TraesCS4A01G132000	Uncharacterized protein	17.27
HRM4A_183985221	4A	-5474	TraesCS4A01G134900	Lung seven transmembrane receptor family protein, expressed *	44.02

**Table 2.** *Cont.*

Marker ID	Chr	Dist (bp)	Gene ID	Description	Sum of TPMs
HRM4A_317085557	4A	-14,884	TraesCS4A01G157000	SAP domain-containing protein	127.49
HRM4A_460238681	4A	-9041	TraesCS4A01G182900	SPRING type domain-containing protein *	11.42
HRM4A_548541053	4A	-11,850	TraesCS4A01G239500	ATP-dependent RNA helicase dhx8 *	19.65
HRM4A_583704598	4A	3305	TraesCS4A01G271900	Histone-lysine N-methyltransferase	11.28
HRM4A_617938526	4A	-5243	TraesCS4A01G335200	Protein kinase domain-containing protein *	10.09
HRM4A_681664894	4A	-3025	TraesCS4A01C408900	Protein PIR *	33.28
HRM4A_683608822	4A	-2685	<b>TraesCS4A01G410700</b>	Ras-related protein RABC2a *	82.16
HRM4A_716986193	4A	-3576	TraesCS4A01G671200L.C	Peptidase M20/M25/M40 family protein	104.94
HRM4A_U-2	4A	-9642	TraesCS4A01G287200	C2 calcium/lipid-binding and GRAM domain containing protein	2.72

Chr: chromosome location; Dist (bp): distance from the gene to the marker in base pairs (bp); TPMs: Transcripts Per Kilobase Millions; and \*: genes manually annotated.

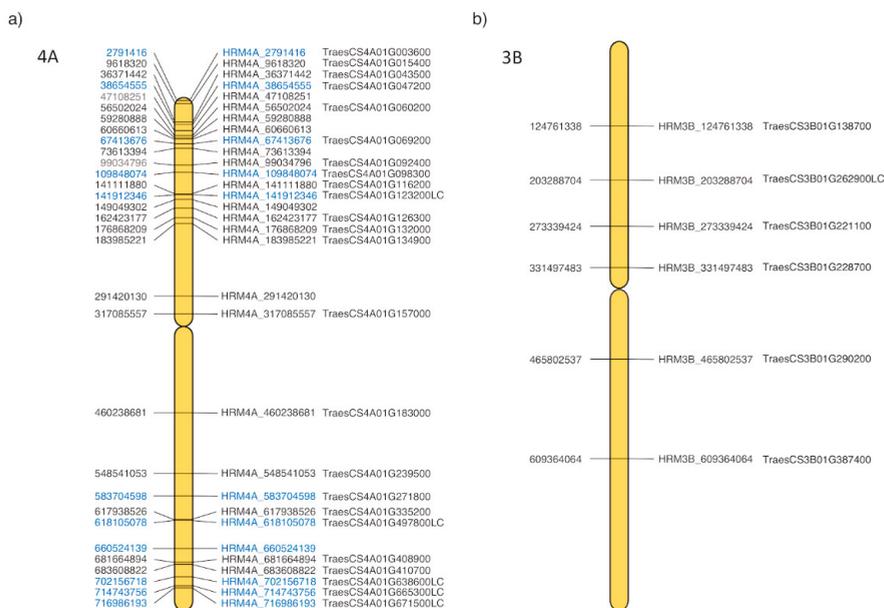




**Figure 3.** Physical and genetic maps for markers located in chromosome 4A (a) and 3B (b) including location for meta-quantitative trait loci (MQTLs) [71] associated to heat and drought stresses tolerance. Markers used in variability assessment are shown in blue color. MQTLs are indicated using green lines.

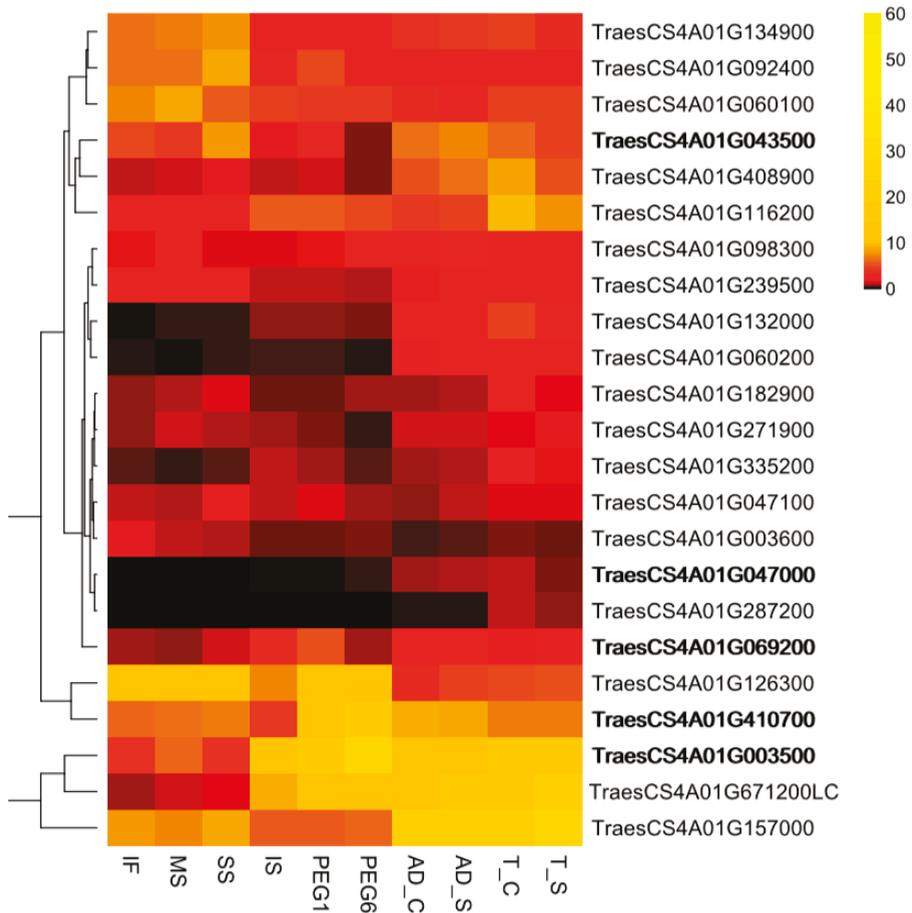
Two of the wheat chromosome 3B ISBP markers (*HRM3B\_465802537* and *HRM3B\_609364064*) were mapped within MQTL26 [71], in proximity to QTLs controlling yield component and biomass traits, and also related to heat stress (Figure 3b).

Additionally, as result of the candidate gene analysis, the developed ISBP markers for wheat chromosome 4A mapped in the proximity of 61 genes (Supplementary Materials Table S3 and Figure S1). The chromosome position for the ISBP and their closest genes are shown in Figure 4a.



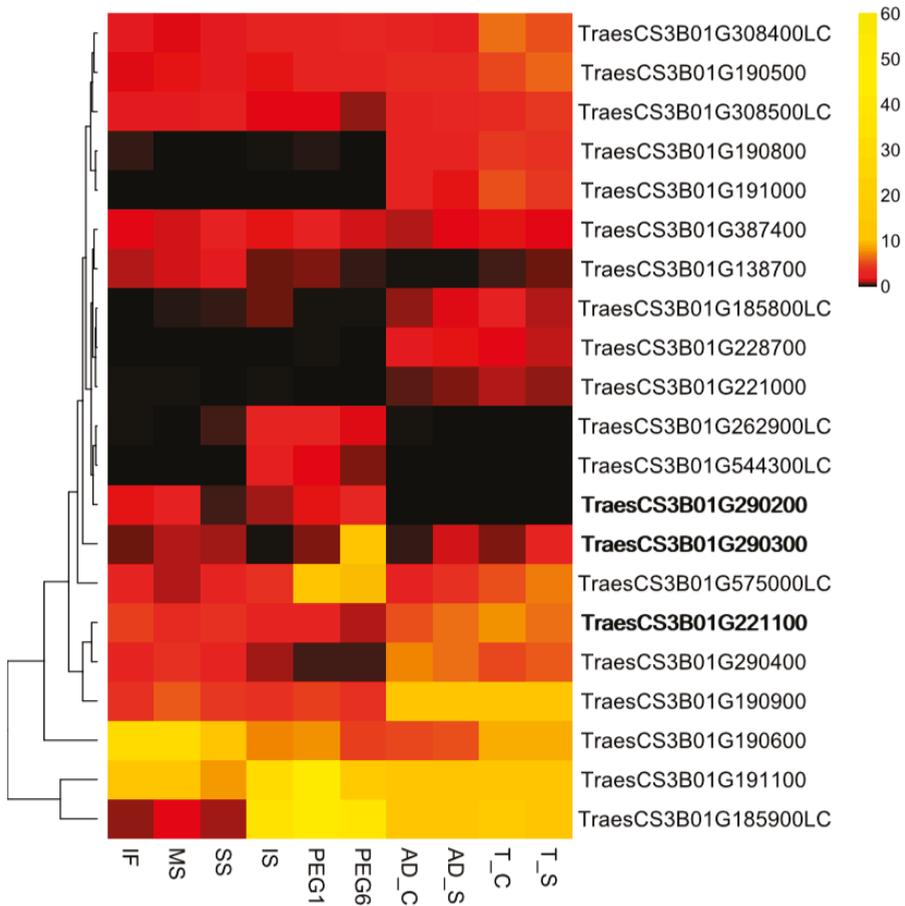
**Figure 4.** Physical maps for the wheat chromosomes 4A (a) and 3B (b) showing the molecular markers location and their nearest genes found within a window of  $\pm 20$  kb and  $\pm 300$  kb, respectively. ISBP markers used in wheat variability assessment are highlighted in blue.

Based on gene expression differences under different drought stress conditions, we filtered 23 genes (37.7% of the total genes) with a TPM value above 2.5 (Table 2 and Figure 5). An expression heatmap, using all available public studies RNASeq data in wheat drought responses [22,26,27] is shown in Figure 5.



**Figure 5.** Gene expression analysis under different water stress conditions for candidate genes located within a window of +/-20 kb to the wheat chromosome 4A ISBP markers. Differentially expressed genes (DEGs) are shown in bold. IF: irrigated field conditions; MS: mild stress field condition; SS: severe stress field condition [27]; IS: seedling control; PEG1: seedling 1 h PEG stress; PEG6: seedling 6 h PEG stress [22]; AD\_C: anther stage irrigated shelter phenotype; AD\_S: anther stage drought stressed shelter phenotype; T\_C: tetrad stage irrigated shelter phenotype; and T\_S: tetrad stage drought shelter phenotype [26].

The wheat chromosome 3B ISBP markers mapped in the proximity of 49 genes (Supplementary Materials Table S4 and Figure S2). The closest genes are shown next to the corresponding marker in Figure 4b. Seventeen of these genes (34.69% of the total) showed TPM values above 2.5 (Table 3 and Figure 6). The drought responsive genes mapped by ISBP markers located in proximity or within QTLs or MQTLs are shown in Table 4.



**Figure 6.** Gene expression analysis under different stress conditions for candidate genes located within a +/-300 kb window and in proximity to wheat chromosome 3B ISBP markers. Differentially expressed genes (DEGs) are shown in bold. IF: irrigated field conditions; MS: mild stress field conditions; SS: severe stress field conditions [27]; IS: seedling PEG shock control; PEG1: seedling 1 h PEG stress; PEG6: seedling 6 h PEG stress [22]; AD\_C: anther stage irrigated shelter phenotype; AD\_S: anther stage drought stressed shelter phenotype; T\_C: tetrad stage irrigated shelter phenotype; and T\_S: tetrad stage drought shelter phenotype [26].

**Table 3.** Wheat chromosome 3B ISBP markers linked to genes which alter their expression in response to water stress treatments. DE genes are shown in bold. The positive and negative values in “Dist” column indicate if the corresponding gene is downstream or upstream of the corresponding marker.

Marker ID	Chr	Dist (bp)	Gene ID	Description	Sum of TPMs
HRM3B_124761338 3B		6356	TraesCS3B01G138700	Ribonucleoside-diphosphate reductase	7.48
		8268	TraesCS3B01G185800LC	Serine/threonine-protein phosphatase 6 regulatory subunit 1	7.25
HRM3B_203288704 3B		14,040	TraesCS3B01G185900LC	LEM3 (Ligand-effect modulator 3)-like	180.07
		227,619	TraesCS3B01G262900LC	Retrovirus-related Pol polyprotein LINE-1	7.05
		260,457	TraesCS3B01G190500	Polynucleotidyl transferase ribonuclease H-like superfamily protein	28.66
		266,554	TraesCS3B01G190600	Beta-amylase	118.13
HRM3B_273339424 3B		-89,674	<b>TraesCS3B01G221100</b>	Protein kinase superfamily protein	43.92
		-90,463	TraesCS3B01G308500LC	translation initiation factor 3 subunit H1	22.30
HRM3B_331497483 3B		-101,322	TraesCS3B01G221000	Kelch repeat protein, putative	4.08
		-286,319	TraesCS3B01G308400LC	Ribonuclease H-like superfamily protein	29.01
		55,223	TraesCS3B01G228700	carboxyl-terminal peptidase, putative (DUF239)	6.57
		83,072	<b>TraesCS3B01G290200</b>	Glycosyltransferase	9.98
HRM3B_465802537 3B		125,805	<b>TraesCS3B01G290300</b>	ABC transporter B family protein	19.57
		296,991	TraesCS3B01G290400	zein-binding protein (Protein of unknown function, DUF593)	34.58
		-24,580	TraesCS3B01G387400	SAC3/GANP/Nim1/mts3/eIF-3 p25 family protein	15.46
HRM3B_609364064 3B		-31,936	TraesCS3B01G575000LC	Myosin-1	48.25

Chr: chromosome location; Dist (bp): distance from the gene to the marker in base pairs (bp); and TPMs: Transcripts Per Kilobase Millions.

**Table 4.** Drought responsive candidate genes tagged by the developed HRM-ISBP markers. DE genes are shown in bold and differentially expressed genes are indicated with “\*\*\*”. Gene expression responses to drought treatments are shown in Figures 5 and 6, and Tables 3 and 4. The positive and negative “Dist” column values indicate if the corresponding gene is downstream or upstream of the marker.

Marker ID	Chr	MQTL	Dist (bp)	Gene ID	Description
HRM4A_716986193 4A		32	-3576	<b>TraesCS4A01G671200LC</b>	Peptidase M20/M25/M40 family protein
HRM4A_714743756 4A		close to QTL in MQTL32	-5397	TraesCS4A01G665300LC	LINE-1 reverse transcriptase-like protein
F			-6886	TraesCS4A01G665200LC	C4-dicarboxylate transport protein
			-12,058	TraesCS4A01G665100LC	NBS-LRR disease resistance protein
			-14,064	TraesCS4A01G665000LC	Disease resistance protein (NBS-LRR class) family
			14,217	TraesCS4A01G665400LC	NBS-LRR disease resistance protein
			17,933	TraesCS4A01G665500LC	Transposase

**Table 4.** *Cont.*

Marker ID	Chr	MQTL	Dist (bp)	Gene ID	Description
HRM4A_583704598	4A	between MQTL30 and 31	-2715	TraesCS4A01G271800	Kinase family protein
			3305	<b>TraesCS4A01G271900</b>	Histone-lysine N-methyltransferase
			-6092	TraesCS4A01G431700LC	Serine/threonine-protein kinase mph1
HRM4A_660524139	4A	close to QTL in MQTL31			
HRM4A_702156718	4A	close to QTL in MQTL32	-1999	TraesCS4A01G638600LC	Retrotransposon protein, putative, unclassified
HRM4A_618105078	4A	31	14,448	TraesCS4A01G638700LC	Kinase, putative
			-2442	TraesCS4A01G497800LC	Receptor-like protein kinase
			-11,813	TraesCS4A01G497700LC	Protein FARI-RELATED SEQUENCE 5
			16,466	TraesCS4A01G335600	NBS-LRR-like resistance protein
			-19,471	TraesCS4A01G335500	NBS-LRR-like resistance protein
HRM4A_617938526	4A	31	-5243	<b>TraesCS4A01G335200</b>	Protein kinase domain-containing protein *
HRM4A_460238681	4A	30	132	TraesCS4A01G183000	Protein kinase domain-containing protein
			-9041	<b>TraesCS4A01G182900</b>	SP-RING-type domain-containing protein *
HRM4A_317085557	4A	30	-14,884	<b>TraesCS4A01G157000</b>	SAP domain-containing protein
HRM4A_683608822	4A	32	-2685	<b>TraesCS4A01G410700</b>	Ras-related protein RABC2a *
HRM3B_609364064	3B	26	52,100	TraesCS3B01G575100LC	Transposon protein, putative, CACTA, En/Spm sub-class
			-24,580	<b>TraesCS3B01G387400</b>	SAC3/GANP/Nim1/mts3/eIF-3 p25 family protein
			-31,936	<b>TraesCS3B01G575000LC</b>	Myosin-1
			-53,463	TraesCS3B01G574900LC	Endonuclease/exonuclease/phosphatase family protein
			-59,628	TraesCS3B01G387300	PP2A regulatory subunit TAP46
HRM3B_465802537	3B	26	83,072	<b>TraesCS3B01G290200</b>	Glycosyltransferase
			-88,899	TraesCS3B01G451200LC	phospholipase-like protein (PEARL1 4) family protein
			-95,341	TraesCS3B01G451100LC	Retrotransposon protein, putative, unclassified
			-98,315	TraesCS3B01G451000LC	Ubiquitin-1
			-106,901	TraesCS3B01G290100	Ribose-5-phosphate isomerase A
			125,805	<b>TraesCS3B01G290300</b>	ABC transporter B family protein
			138,676	TraesCS3B01G451300LC	Polynucleotidyl transferase, ribonuclease H-like superfamily protein
			-144,473	TraesCS3B01G450900LC	Protein regulator of cytokinesis 1
			-147,938	TraesCS3B01G450800LC	Beta-hexosaminidase
			161,237	TraesCS3B01G451400LC	Endonuclease/exonuclease/phosphatase family protein
			296,991	<b>TraesCS3B01G290400</b>	zein-binding protein (Protein of unknown function, DUF593)
			297,913	TraesCS3B01G290500	Potassium channel

Chr: chromosome location; MQTL: meta-quantitative trait loci where the marker was located; and Dist (bp): distance from the gene to the marker in base pairs (bp).

After differential gene expression analysis we obtained 5 DE genes in chromosome 4A and 3 in chromosome 3B, which were up and down-regulated by the PEG drought treatment [22]. The 4A chromosome DE genes *TraesCS4A01G003500* and *TraesCS4A01G043500*, were up and down regulated under PEG6 drought treatment, respectively. Genes *TraesCS4A01G047000*, *TraesCS4A01G410700* were up-regulated, while the gene *TraesCS4A01G069200* was down-regulated under PEG6 treatment. Chromosome 3B gene *TraesCS3B01G221100* was down-regulated under PEG6 treatment, while genes *TraesCS3B01G290200* and *TraesCS3b01G290300* were up-regulated (Supplementary Materials Table S5).

### 3.3. Wheat Variability Assessment by High Resolution Melting Analysis

To assess the polymorphism levels in a wheat collection, the 45 ISBP markers developed for the wheat chromosome 4A (Table 1), were evaluated. Thirteen of them were selected based on their reproducibility and polymorphism (Table 1) and were used in HRM analyses to study the genetic diversity among durum and bread wheat lines in panel 1 (Supplementary Materials Table S1).

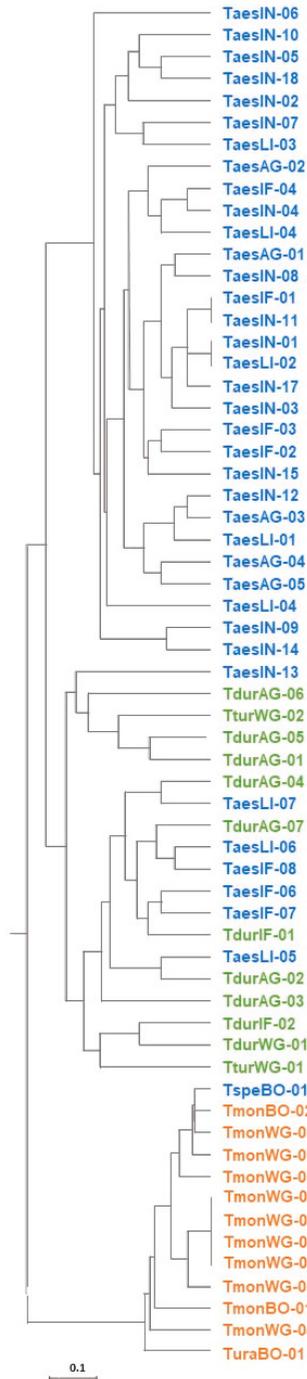
The number of alleles detected for these markers ranged from 2 to 7 (mean = 3.38), and the polymorphism index content varied between 0.24 and 0.68 (mean = 0.52) (Table 5).

**Table 5.** Genetic parameters for the wheat chromosome 4A ISBP markers used in the variability assessment.

Marker ID	HRM Type	No. of Alleles	PIC
HRM4A_2791416	A	2	0.336
HRM4A_38654555	A	2	0.566
HRM4A_67413676	A	3	0.584
HRM4A_141912346	A	5 *	0.595
HRM4A_109848074	A	2	0.544
HRM4A_618105078	B	4 *	0.677
HRM4A_716986193	C	4	0.602
HRM4A_U-3	C	2 *	0.242
HRM4A_714743756	A	2 *	0.368
HRM4A_U-4	A	3 *	0.474
HRM4A_583704598	A	7 *	0.571
HRM4A_660524139	A	4 *	0.57
HRM4A_702156718	A	4 *	0.563
<b>Mean</b>		3.38	0.515

HRM type: high resolution melting pattern type; PIC: polymorphism index content; \*: one of the alleles was described as "null genotype".

The cluster analysis shows three clearly differentiated clusters (Figure 7). The first one contains 30 *T. aestivum* lines, while the remaining 8 bread wheat lines (TaesIN-13, TaesLI-06, TaesLI-07, TaesIF-06, TaesIF-08, TaesIF-07, TaesLI-05, and TspeBO-01) were placed within the second cluster. This cluster also contains *Triticum durum* and *Triticum dicoccoides* accessions, while *Triticum monococcum* and *Triticum urartu* were placed in the third cluster (Figure 7). The cophenetic correlation coefficient obtained for the UPGMA tree was 0.86.



**Figure 7.** Phylogenetic unweighted pair group method with arithmetic mean (UPGMA) tree showing the relationships among 62 wheat lines genotyped with 13 ISBP markers. *Triticum aestivum* lines are shown in blue, *T. durum* and *T. dicoccoides* in green, and *T. monococcum* in orange.

Seven of these ISBP markers (Table 1) selected by their efficiency and polymorphism for durum wheat, were used for the variability assessment of durum wheat lines in panel 2 (Supplementary Materials Table S2). The UPGMA tree for the wheat panel 2 is shown in Supplementary Materials Figure S3a,b. This cluster analysis resulted in 9 differentiated clusters. Five lines (BGE002866, BGE013055, BGE020464, BGE013652, and BGE013722) were not placed in any of these clusters. The observed distribution of durum wheat lines across the clusters could be associated in some cases to the geographical location and agroclimatic areas (Supplementary Materials Figure S3c). There are two clear clusters (clusters 4 and 6) where species from southern Spain are presented in a larger proportion. Furthermore, wheat lines placed in cluster 1 are mainly located in norther temperate zones without dry season and temperate summer; while cluster 4 contains landraces mainly located in southern template areas with dry and cool summer. The CCC obtained for the UPGMA tree was 0.67. The number of alleles detected in this analysis ranged from 3 to 6 (mean allele number = 3.86), and the PIC mean value was 0.486 (Supplementary Materials Table S6).

#### 4. Discussion

Insertion site-based polymorphism markers have been described as a useful tool for wheat genomic studies [35] and attractive alternative to markers as SSR or SNP, due to the high repetitive content of some cereal genomes [81,82]. Due to their straightforward design and high polymorphism, there are previous studies which developed and used specific ISBP markers for several wheat chromosomes and different purposes (i.e., Barabaschi et al. [83] for the bread wheat chromosome 5A, focused on polymorphism assessment; Lucas et al. [84], who used markers derived from the wheat chromosome 1A to map this chromosome, and also with marker assisted breeding purposes; Li et al. [36] applied wheat chromosome 3B ISBP linked to mildew resistance genes in durum wheat; or Sehgal et al. [41], who used chromosome 3A ISBP markers for gene discovery and physical mapping). In this work, new ISBP markers were developed for the wheat chromosome 4A, which contains interesting genes related to biotic and abiotic stresses, as drought tolerance [65,68,69]. It also harbors loci related to essential agronomic traits as yield and grain quality [62,64,66]. The ISBP markers resulted highly polymorphic (Table 5). Thirteen of the developed ISBP markers were used for wheat variability assessment and showed melting curve polymorphisms, seven of them with a PIC value higher than 0.50. Thus, they resulted highly resolutive tools for wheat variability assessment using a cost-effective technique as high resolution melting analysis (HRM). This technique is described as an optimized methodology for melting curve assessment, which allows the determination of melting temperature and profile of an amplicon [85]. Some studies have highlighted some advantages for the HRM technique, as its reduced cost per sample in comparison with other techniques used for SNP detection [86–88]. It is worth noting that the required system consists of standard and affordable RT-PCR equipment, which is suitable for in-house genotyping and adequate for small/medium breeders. Other advantages for HRM are the excellent results for the detection of homozygous and heterozygous variants [46,89–91]; its use for gene mapping, SNPs and mutations [46,90,92]; and its efficiency for the identification of species and closely related varieties [88,93–97]. In this regard, our results from ISBP markers sequence validation, as well as HRM genotyping, support the efficiency of HRM analysis in wheat varieties differentiation. Our results are in agreement with Dong et al. [53], who highlighted that HRM does not require any digestion or gel electrophoresis, so it provides a worthwhile approach for SNP/indel genotyping of different varieties without prior sequence knowledge, as required by other methods. Results from the wheat genetic diversity assessment confirmed that HRM is a convenient way for a first screening to determine variability groups, prior to resequencing only representative varieties as the basis to develop other SNP platforms. Nevertheless, some HRM limitations have been pointed out. Distefano et al. [88] highlighted that sometimes, the HRM profiles could be similar preventing the differentiation of some of the genotypes. Regardless of this, Wu et al. [98] proposed that this issue can be solved using mixed strategies. Accordingly, our results show that a combined use of different ISBP markers can differentiate all the wheat lines studied.

Twenty ISBP markers for wheat chromosomes 4A and six for chromosome 3B mapped to interesting candidate genes, mainly related to drought and heat stresses and yield components (Tables 2 and 3). They were validated using data from available RNASeq public studies in wheat drought responses [22,26,27], as they showed differences on their expression under different stress conditions.

In chromosome 4A, the ISBP marker *HRM4A\_109848074* mapped next (97 bp) to the gene *TraesCS4A01G098300*, which encodes a xyloxytransferase 1, and participates in carbohydrates metabolism in the development of cellular walls [99]. This process is markedly affected by water stress [100,101] and this can be observed in Figure 5, where this gene decreases its expression under severe stress field conditions. This result is also in agreement with results found by Abebe et al. [102], who analyzed spikes of barley grown under controlled drought conditions, and also found that this gene was down-regulated. The marker *HRM4A\_2791416* mapped 814 bp upstream to the gene *TraesCS4A01G003600*, which encodes an alpha/beta-hydrolases superfamily protein with functional adaptability in plants [103]. This gene reduces its expression as drought stress increases under field conditions (Figure 5). Marker *HMR4A\_67413676*, mapped 975 bp to the gene *TraesCS4A01G069200*, which encodes an armadillo repeat-only protein. These kind of repeat proteins participate in the coordination of protein interactions during stress and hormonal signalling in plants [104]. In agreement with this, the gene was downregulated under PEG6 drought treatment (Supplementary Materials Table S5). The marker *HMR4A\_683608822*, which was located within the drought stress tolerance MQTL32 [71] (Table 5 and Figure 3a), mapped 2685 bp upstream to the gene *TraesCS4A01G410700*, which encodes a ras-related protein RABC2a. The function of this gene has been related to ABA induced stress tolerance in barley [105]. Accordingly to a drought stress response role, this gene was upregulated under PEG6 drought treatment (Supplementary Materials Table S5). The marker *HMR4A\_36371442* mapped 2736 bp upstream to the gene *TraesCS4A01G043500* and was downregulated under PEG6 drought treatment (Supplementary Materials Table S5). Contrary to this, this gene increases its expression under severe stress field conditions (Figure 5). This gene encodes a STAS domain containing-protein, which plays a role to membrane attachment of many anion transporters in transport activity and regulation in plants [106]. In fact, it has been demonstrated its crucial role in the activity of sulfate transporter in *Arabidopsis thaliana* [107], providing key amino acids in the sulfate transport activity [108]. The importance of this activity should be noted, since sulfate is an element that has been described as an essential component in the structure of plant enzymes and reserve proteins in grain [109]. Further, marker *HRM4A\_716986193*, which was located close to a QTL within MQTL32 [71] (Table 5 and Figure 3a), mapped in proximity (3576 bp upstream) to the gene *TraesCS4A01G671200LC*. This gene encodes a peptidase M20/M25/M40 family protein, which is involved in drought stress responses [110]. In fact, proteolysis under drought conditions allows a reorganization in the plant's metabolism, and also increases plants drought tolerance [20,111,112]. According to this, the results showed increased expression of this gene under different drought stress treatments (Figure 6). This is also in agreement with the results showed by Simova-Stoilova et al. [113], who assessed wheat leaves under severe soil drought and found an increase in peptidase activity. Thus, this drought responsive genes represent an interesting candidate for a known drought and heat stress tolerance MQTL.

Additionally, within the window of +/-20 kb used for the candidate gene analysis in chromosome 4A, there were two interesting genes: *TraesCS4A01G003500* (5141 bp upstream from marker *HRM4A\_2791416*) and *TraesCS4A01G047000* (14,885 bp upstream from marker *HRM4A\_38654555*). The first gene encodes a thionin like-protein gene, which plays an important role in the growth and development of the plant and its defense against pathogens [114]. It was found differentially expressed under PEG drought treatment, being upregulated under PEG6 drought treatment (Supplementary Materials Table S5). The gene *TraesCS4A01G047000* encodes a formin-like protein, which plays a primary role in the organization of plant's structure [115]. In agreement with our results where this gene was found upregulated under PEG6 drought treatment (Supplementary Materials Table S5), formin-like proteins showed variations in their expression under drought conditions in wheat [115].

Some HRM/ISBP markers mapped to previously described MQTL loci [71] (Table 5 and Figure 3a), which were mainly associated to drought and heat stresses tolerance. Within these markers, *HRM4A\_618105078*, which tags the MQTL31 [71] and mapped close (2442 bp upstream) to the gene *TraesCS4A01G497800LC* can be highlighted. This gene encodes a receptor-like protein kinase, which is involved in abiotic stress responses [116], matching the description assigned to the MQTL.

Additionally, within the window of +/-300 kb, four wheat chromosome 3B ISBP markers can be highlighted. The marker *HRM3B\_124761338* mapped 6356 bp downstream to the gene *TraesCS3B01G138700*, which encodes a ribonucleoside-diphosphate reductase, an essential enzyme for DNA synthesis [117,118]. This gene increases its expression under severe field stress conditions, and it decreases in response to a PEG drought treatment (Figure 6). Marker *HRM3B\_609364064* mapped 31,936 bp downstream to the gene *TraesCS3B01G575000LC*, encoding myosin-1. Plant myosins have a functional role in organelle movement in response to biotic and abiotic stresses [119]. This response is shown in Figure 6, where this gene significantly increments its expression under PEG1 and PEG6 drought treatments. Marker *HRM3B\_465802537* mapped to two interesting genes, the gene *TraesCS3B01G290200* (83,072 bp downstream) and the gene *TraesCS3B01G290300* (125,805 bp downstream), which were both found upregulated under PEG6 drought treatment (Supplementary Materials Table S5), and contrary to this, decreased their expression under severe stress field conditions (Figure 6). The gene *TraesCS3B01G290200* encodes a glycosyltransferase, an enzyme which possesses a main role in plant's stress tolerance and defense [120], and in agreement with our results, it has been previously found upregulated in wheat leaf under drought conditions [121]; and the gene *TraesCS3B01G290300* encodes an ABC transporter B family protein, which is significantly involved in organs growth, plant nutrition and development and plant responses to abiotic stresses [122]. In fact, as Rampino et al. [123] highlighted, and in agreement with our results (Supplementary Materials Table S6), the up-regulation of this gene in wheat under heat and drought conditions confirms the implication of this gene family in drought responses. Therefore, this family protein has been related to grain formation in wheat [124], which is consistent with the location of marker *HRM3B\_465802537* within MQTL26 [71], mainly related to yield components. Thus, this marker can be useful in wheat breeding, for the marker-assisted selection of this interesting gene and MQTL. Finally, the marker *HRM4A\_273339424* mapped 89,674 bp upstream to the gene *TraesCS3B01G221100*, which encodes a protein kinase superfamily protein. This protein's family is involved in plants' responses to abiotic stresses and plants' development [125]. According to this, this gene was found downregulated under PEG6 drought treatment (Supplementary Materials Table S5), and it also shows differences in its expression across different stress conditions (Figure 6). Therefore, these results agree with Wei et al. [126], who highlighted that kinase proteins are involved in various responses with exposure time of drought.

According to our results, the developed HRM-ISBP markers can be used in wheat breeding programs to genotype interesting genomic regions in a cost-effective manner. These markers can useful resources for marker-assisted selection (MAS) focused on abiotic stress responses and yield components, to tag interesting known QTLs and MQTLs related to drought and heat stresses tolerance, and also yield-related traits.

## 5. Conclusions

In this work, highly polymorphic ISBP markers for wheat chromosome 4A were successfully developed and applied in a genetic variability assessment of a collection of durum and bread wheats, using the high-resolution melting analysis technique. These HRM-ISBP markers represent cost-effective and efficient tools for wheat breeding programs focused on variability assessments. The obtained results provide an interesting framework for wheat genetic studies and varieties selection. These HRM-ISBP markers have also been shown useful for tagging interesting genes associated to drought and heat stresses tolerance, some of which showed differential expression patterns under stress conditions. In addition, some of these markers can be applied in breeding through marker-assisted selection of

QTL and MQTL related to abiotic stresses as drought and heat, and also yield and yield related traits. The resources and results presented here can also facilitate the understanding of important traits in other species with large genomes.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4395/10/9/1294/s1>; Table S1. List of durum and bread wheat lines used for variability analysis by high resolution melting. <sup>1</sup>: marker set up analysis; <sup>2</sup>: marker's sequence validation by HRM analysis; <sup>3</sup>: HRM pattern types obtained; IFAPA: Instituto Andaluz de Investigación y Formación Agraria, Pesquera, Alimentaria y de la Producción ecológica; INRA: Institut National de la Recherche Agronomique; and WGRC: Wheat Genetics Resource Center; Table S2. Durum wheat lines used for variability analysis by high resolution melting (HRM); Table S3. Candidate genes tagged by the developed HRM-ISBP markers for wheat chromosome 4A. Genes with expression differences are shown in bold and DE genes are indicated with “\*”. The positive and negative values in the “Dist” column indicate if the corresponding gene is downstream or upstream of the respective marker. Chr: chromosome location. Dist (bp): distance from the gene to the marker in base pairs; Table S4. Candidate genes tagged by the HRM-ISBP markers for wheat chromosome 3B. Genes with expression differences are shown in bold and DE genes are indicated with “\*”. The positive and negative values in the “Dist” column indicate if the corresponding gene is downstream or upstream of the respective marker. Chr: chromosome location. Dist (bp): distance from the gene to the marker in base pairs; Table S5. Differential expression (DE) analysis significance parameters for DE genes. Significant values ( $|\log_2FC$ ,  $\beta$ ) > 1 and  $p$ -adjust,  $Q$ -value < 0.05) are shown in bold; Table S6. Genetic parameters for wheat chromosome 4A ISBP markers used in durum wheat variability assessment. HRM type: high resolution melting pattern type; No. alleles: number of alleles found with the marker; PIC: polymorphism index content; and \*: one of the genotypes was described as “null genotype”; Figure S1. Gene expression analysis under different water stress conditions for all candidate genes located within a  $\pm 20$  kb window to the wheat chromosome 4A ISBP markers. Genes with differences on their expression are shown in bold and DE genes are indicated with “\*”. IF: irrigated field conditions; MS: mild stress field conditions; SS: severe stress field conditions; IS: seedling PEG shock control; PEG1: seedling 1 h PEG stress; PEG6: seedling 6 h PEG stress; AD\_C: anther stage irrigated shelter phenotype; AD\_S: anther stage drought stressed shelter phenotype; T\_C: tetrad stage irrigated shelter phenotype; and T\_S: tetrad stage drought shelter phenotype; Figure S2. Gene expression analysis under different water stress conditions for all candidate genes located within a  $\pm 300$  kb window to the wheat chromosome 3B ISBP markers. Genes with differences on their expression are shown in bold and DE genes are indicated with “\*”. IF: irrigated field conditions; MS: mild stress field conditions; SS: severe stress field conditions; IS: seedling PEG shock control; PEG1: seedling 1 h PEG stress; PEG6: seedling 6 h PEG stress; AD\_C: anther stage irrigated shelter phenotype; AD\_S: anther stage drought stressed shelter phenotype; T\_C: tetrad stage irrigated shelter phenotype; and T\_S: tetrad stage drought shelter phenotype; Figure S3. Phylogenetic UPGMA tree showing the relationships between 76 durum wheat lines genotyped with 7 wheat chromosome 4A ISBP markers. a) wheat lines are colored based on their geographic zone (Supplementary Materials Table S2) (Center: green; North: blue; North-east: light blue; North-west: dark blue; South: red; South-east: orange; South-west: maroon; and East: purple); b) wheat lines are colored for species (*T. turgidum* subsp *durum* appears in green, *T. turgidum* subsp *turgidum* in orange and *T. turgidum* subsp. *dicocon* in purple). The colored and vertical lines indicate the differentiated clusters (cluster 1—yellow; cluster 2—blue; cluster 3—dark blue; cluster 4—pink; cluster 5—purple; cluster 6—red; cluster 7—green; cluster 8—grey; cluster 9—light blue; and black lines show the wheat lines that have not been included within any cluster); and c) each dot represents a wheat line, the colors are the same in “a”).

**Author Contributions:** P.H., P.G., and E.P. conceived the experiments. R.M.-G. and L.P. isolated the DNA. E.P. sequenced ISBP amplicons and designed ISBP markers. R.M.-G. performed the PCR and HRM analyses. S.G. performed the bioinformatics analyses. R.M.-G., S.G., E.P., G.D., L.P., P.G., and P.H. analyzed the results. R.M.-G. and P.H. drafted the manuscript. R.M.-G., S.G., E.P., G.D., L.P., P.G., and P.H. have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by project P18-RT-992 from Junta de Andalucía (Andalusian Regional Government), Spain (Co-funded by FEDER), and by the Spanish Ministry of Science and Innovation project PID2019-109089RB-C32.

**Acknowledgments:** The authors gratefully acknowledge Agrovegetal S. A. (Spain), Instituto Andaluz de Investigación y Formación Agraria (IFAPA, Spain), Royal Botanic Gardens of Cordoba (Spain), INRA-Clermont-Ferrand and INRA-Versailles (Institut National de la Recherche Agronomique, France), Limagrain (France and Spain), the Wheat Genetics Resource Center (WGRC Kansas State University, USA), and CFR-INIA (Spanish National Plant Genetic Resources Center) for providing the seeds resources.

**Conflicts of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Lawlor, D.W. Wheat and Wheat Improvement. *Soil Sci.* **1988**, *146*, 292–293. [CrossRef]

2. Rampino, P.; Pataleo, S.; Gerardi, C.; Mita, G.; Perrotta, C. Drought stress response in wheat: Physiological and molecular analysis of resistant and sensitive genotypes. *Plant Cell Environ.* **2006**, *29*, 2143–2152. [CrossRef] [PubMed]
3. Nezhadahmadi, A.; Prodhani, Z.H.; Faruq, G. Drought tolerance in wheat. *Sci. World J.* **2013**, *2013*, 610721. [CrossRef] [PubMed]
4. Farooq, M.; Hussain, M.; Siddique, K.H.M. Drought stress in wheat during flowering and grain-filling periods. *Crit. Rev. Plant Sci.* **2014**, *33*, 331–349. [CrossRef]
5. Sallam, A.; Alqudah, A.M.; Dawood, M.F.A.; Baenziger, P.S.; Börner, A. Drought stress tolerance in wheat and barley: Advances in physiology, breeding and genetics research. *Int. J. Mol. Sci.* **2019**, *20*, 3137. [CrossRef]
6. Wang, W.; Vinocur, B.; Altman, A. Plant responses to drought, salinity and extreme temperatures: Towards genetic engineering for stress tolerance. *Planta* **2003**, *218*, 1–14. [CrossRef]
7. Wang, W.; Vinocur, B.; Shoseyov, O.; Altman, A. Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends Plant Sci.* **2004**, *9*, 244–252. [CrossRef]
8. Gooding, M.J.; Ellis, R.H.; Shewry, P.R.; Schofield, J.D. Effects of restricted water availability and increased temperature on the grain filling, drying and quality of winter wheat. *J. Cereal Sci.* **2003**, *37*, 295–309. [CrossRef]
9. IPCC Assessment Report. Published Online. 2020. Available online: <https://www.ipccch/srcl/> (accessed on 3 March 2020).
10. Lesk, C.; Rowhani, P.; Ramankutty, N. Influence of extreme weather disasters on global crop production. *Nature* **2016**, *529*, 84–87. [CrossRef]
11. Fathi, A.; Tari, D.B. Effect of drought stress and its mechanism in plants. *Int. J. Life Sci.* **2016**, *10*, 1–6. [CrossRef]
12. Gupta, A.; Balyan, H.; Gahlaut, V. QTL analysis for drought tolerance in wheat: Present status and future possibilities. *Agronomy* **2017**, *7*, 5. [CrossRef]
13. Reynolds, M.P.; Ortiz, R. Adapting crops to climate change: A summary. In *Climate Change and Crop Production*; CAB: Oxfordshire, UK, 2010; pp. 1–8. [CrossRef]
14. Kantar, M.; Lucas, S.J.; Budak, H. *Drought Stress: Molecular Genetics and Genomics Approaches*, 1st ed.; Academic Press: Cambridge, MA, USA, 2011; Volume 57, ISBN 9780123876928.
15. Li, J.; Zhou, R.; Endo, T.R.; Stein, N. High-throughput development of SSR marker candidates and their chromosomal assignment in rye (*Secale cereale* L.). *Plant Breed.* **2018**, *137*, 561–572. [CrossRef]
16. Farooq, M.; Bramley, H.; Palta, J.A.; Siddique, K.H.M. Heat stress in wheat during reproductive and grain-filling phases. *Crit. Rev. Plant Sci.* **2011**, *30*, 491–507. [CrossRef]
17. Cattivelli, L.; Rizza, F.; Badeck, F.W.; Mazzucotelli, E.; Mastrangelo, A.M.; Francia, E.; Marè, C.; Tondelli, A.; Stanca, A.M. Drought tolerance improvement in crop plants: An integrated view from breeding to genomics. *Field Crop. Res.* **2008**, *105*, 1–14. [CrossRef]
18. McWilliam, J.R. The Dimensions of Drought. In *Drought Resistant Cereals*; Bak, F., Ed.; CAB Int.: Wallingford, UK, 1989; pp. 1–11.
19. Blum, A. Drought resistance is it really a complex trait? *Funct. Plant Biol.* **2011**, *38*, 753–757. [CrossRef]
20. Ingram, J.; Bartels, D. The molecular basis of dehydration tolerance in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1996**, *47*, 377–403. [CrossRef]
21. Chaves, M.M.; Maroco, J.P.; Pereira, J.S. Understanding plant responses to drought—From genes to the whole plant. *Funct. Plant Biol.* **2003**, *30*, 239–264. [CrossRef]
22. Liu, Z.; Xin, M.; Qin, J.; Peng, H.; Ni, Z.; Yao, Y.; Sun, Q. Temporal transcriptome profiling reveals expression partitioning of homeologous genes contributing to heat and drought acclimation in wheat (*Triticum aestivum* L.). *BMC Plant Biol.* **2015**, *15*, 1–20. [CrossRef]
23. Li, H.; Li, M.; Wei, X.; Zhang, X.; Xue, R.; Zhao, Y.; Zhao, H. Transcriptome analysis of drought-responsive genes regulated by hydrogen sulfide in wheat (*Triticum aestivum* L.) leaves. *Mol. Genet. Genom.* **2017**, *292*, 1091–1110. [CrossRef]
24. Chaichi, M.; Sanjarian, F.; Razavi, K.; Gonzalez-Hernandez, J.L. Analysis of transcriptional responses in root tissue of bread wheat landrace (*Triticum aestivum* L.) reveals drought avoidance mechanisms under water scarcity. *PLoS ONE* **2019**, *14*, e0212671. [CrossRef]

25. Iqbal, M.A.; Sharma, P.; Jasrotia, R.S.; Jaiswal, S.; Kaur, A.; Saroha, M.; Angadi, U.B.; Sheoran, S.; Singh, R.; Singh, G.P.; et al. RNAseq analysis reveals drought-responsive molecular pathways with candidate genes and putative molecular markers in root tissue of wheat. *Sci. Rep.* **2019**, *9*, 1–18. [[CrossRef](#)] [[PubMed](#)]
26. Ma, J.; Li, R.; Wang, H.; Li, D.; Wang, X.; Zhang, Y.; Zhen, W.; Duan, H.; Yan, G.; Li, Y. Transcriptomics analyses reveal wheat responses to drought stress during reproductive stages under field conditions. *Front. Plant Sci.* **2017**, *8*, 592. [[CrossRef](#)] [[PubMed](#)]
27. Gálvez, S.; Mérida-García, R.; Camino, C.; Borrill, P.; Abrouk, M.; Ramírez-González, R.H.; Biyikliglu, S.; Amil-Ruiz, F.; Dorado, G.; Budak, H.; et al. Hotspots in the genomic architecture of field drought responses in wheat as breeding targets. *Funct. Integr. Genom.* **2019**, *19*, 295–309. [[CrossRef](#)] [[PubMed](#)]
28. Wahid, A.; Gelani, S.; Ashraf, M.; Foolad, M.R. Heat tolerance in plants: An overview. *Environ. Exp. Bot.* **2007**, *61*, 199–223. [[CrossRef](#)]
29. Schöffl, F.; Prandl, R.; Reindl, A. Molecular Responses to Heat Stress. In *Molecular Responses to Cold, Drought, Heat Salt Stress in Higher Plants*; Shinozaki, K., Yamaguchi-Shinozaki, K., Eds.; R.G. Landes Co.: Austin, TX, USA, 1999; pp. 81–98.
30. Paux, E.; Faure, S.; Choulet, F.; Roger, D.; Gauthier, V.; Martinant, J.P.; Sourdille, P.; Balfourier, F.; Le Paslier, M.C.; Chauveau, A.; et al. Insertion site-based polymorphism markers open new perspectives for genome saturation and marker-assisted selection in wheat. *Plant Biotechnol. J.* **2010**, *8*, 196–210. [[CrossRef](#)]
31. Mourad, A.M.I.; Alomari, D.Z.; Alqudah, A.M.; Sallam, A.; Salem, K.F.M. *Recent Advances in Wheat (Triticum spp.) Breeding*; Springer: Cham, Switzerland, 2019; Volume 5, ISBN 9783030231088.
32. Landjeva, S.; Korzun, V.; Börner, A. Molecular markers: Actual and potential contributions to wheat genome characterization and breeding. *Euphytica* **2007**, *156*, 271–296. [[CrossRef](#)]
33. IWGSC. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **2018**, *361*, eaar7191. [[CrossRef](#)]
34. Paux, E.; Roger, D.; Badaeva, E.; Gay, G.; Bernard, M.; Sourdille, P.; Feuillet, C. Characterizing the composition and evolution of homoeologous genomes in hexaploid wheat through BAC-end sequencing on chromosome 3B. *Plant J.* **2006**, *48*, 463–474. [[CrossRef](#)]
35. Paux, E.; Gao, L.; Faure, S.; Choulet, F.; Roger, D.; Chevalier, K.; Saintenac, C.; Balfourier, F.; Paux, K.; Cakir, M.; et al. Insertion site-based polymorphism: A Swiss army knife for wheat genomics. In Proceedings of the 11th International Wheat Genetics Symposium, Brisbane, QLT, Australia, 24–19 August 2008; pp. 4–6.
36. Li, G.; Fang, T.; Zhang, H.; Xie, C.; Li, H.; Yang, T.; Nevo, E.; Fahima, T.; Sun, Q.; Liu, Z. Molecular identification of a new powdery mildew resistance gene Pm41 on chromosome 3BL derived from wild emmer (*Triticum turgidum* var. *dicoccoides*). *Theor. Appl. Genet.* **2009**, *119*, 531–539. [[CrossRef](#)]
37. Cubizolles, N.; Rey, E.; Choulet, F.; Rimbart, H.; Laugier, C.; Balfourier, F.; Bordes, J.; Poncet, C.; Jack, P.; James, C.; et al. Exploiting the Repetitive Fraction of the Wheat Genome for High-Throughput Single-Nucleotide Polymorphism Discovery and Genotyping. *Plant Genom.* **2016**, *9*, 1–11. [[CrossRef](#)]
38. Rimbart, H.; Darrier, B.; Navarro, J.; Kitt, J.; Choulet, F.; Leveugle, M.; Duarte, J.; Rivière, N.; Eversole, K.; Le Gouis, J.; et al. High throughput SNP discovery and genotyping in hexaploid wheat. *PLoS ONE* **2018**, *13*, e0186329. [[CrossRef](#)]
39. Oliveira, H.R.; Campana, M.G.; Jones, H.; Hunt, H.V.; Leigh, F.; Redhouse, D.I.; Lister, D.L.; Jones, M.K. Tetraploid wheat landraces in the Mediterranean basin: Taxonomy, evolution and genetic diversity. *PLoS ONE* **2012**, *7*, e37063. [[CrossRef](#)] [[PubMed](#)]
40. Lucas, S.J.; Budak, H. Sorting the wheat from the chaff: Identifying miRNAs in genomic survey sequences of *Triticum aestivum* chromosome 1AL. *PLoS ONE* **2012**, *7*, e40859. [[CrossRef](#)] [[PubMed](#)]
41. Sehgal, S.K.; Li, W.; Rabinowicz, P.D.; Chan, A.; Šimková, H.; Doležel, J.; Gill, B.S. Chromosome arm-specific BAC end sequences permit comparative analysis of homoeologous chromosomes and genomes of polyploid wheat. *BMC Plant Biol.* **2012**, *12*, 64. [[CrossRef](#)]
42. Vaissayre, L.; Ardisson, M.; Borries, C.; Santoni, S.; David, J.; Roumet, P. Elite durum wheat genetic map and recombination rate variation in a multiparental connected design. *Euphytica* **2012**, *185*, 61–75. [[CrossRef](#)]
43. Salina, E.A.; Nesterov, M.A.; Frenkel, Z.; Kiseleva, A.A.; Timonova, E.M.; Magni, F.; Vrána, J.; Šafář, J.; Šimková, H.; Doležel, J.; et al. Features of the organization of bread wheat chromosome 5BS based on physical mapping. *BMC Genom.* **2018**, *19*, 129–141. [[CrossRef](#)] [[PubMed](#)]

44. McNeil, M.D.; Kota, R.; Paux, E.; Dunn, D.; McLean, R.; Feuillet, C.; Li, D.; Kong, X.; Lagudah, E.; Zhang, J.C.; et al. BAC-derived markers for assaying the stem rust resistance gene, Sr2, in wheat breeding programs. *Mol. Breed.* **2008**, *22*, 15–24. [[CrossRef](#)]
45. Vossen, R.H.A.M.; Aten, E.; Roos, A.; Den Dunnen, J.T. High-resolution melting analysis (HRMA)—More than just sequence variant screening. *Hum. Mutat.* **2009**, *30*, 860–866. [[CrossRef](#)]
46. Montgomery, J.L.; Sanford, L.N.; Wittwer, C.T. High-resolution DNA melting analysis in clinical research and diagnostics. *Expert Rev. Mol. Diagn.* **2010**, *10*, 219–240. [[CrossRef](#)]
47. Arias Aguirre, A.; Studer, B.; Do Canto, J.; Frei, U.; Lübberstedt, T.; Roggli, O.A. Validation of two models for self-incompatibility in autotetraploid perennial ryegrass using high resolution melting-based markers. *Plant Breed.* **2014**, *133*, 765–770. [[CrossRef](#)]
48. Reed, G.H.; Wittwer, C.T. Sensitivity and specificity of single-nucleotide polymorphism scanning by high-resolution melting analysis. *Clin. Chem.* **2004**, *50*, 1748–1754. [[CrossRef](#)] [[PubMed](#)]
49. Shatalina, M.; Messmer, M.; Feuillet, C.; Mascher, F.; Paux, E.; Choulet, F.; Wicker, T.; Keller, B. High-resolution analysis of a QTL for resistance to *Stagonospora nodorum* glume blotch in wheat reveals presence of two distinct resistance loci in the target interval. *Theor. Appl. Genet.* **2014**, *127*, 573–586. [[CrossRef](#)] [[PubMed](#)]
50. Mondini, L.; Nachit, M.M.; Porceddu, E.; Pagnotta, M.A. HRM technology for the identification and characterization of INDEL and SNPs mutations in genes involved in drought and salt tolerance of durum wheat. *Plant Genet. Resour. Charact. Util.* **2011**, *9*, 166–169. [[CrossRef](#)]
51. Iehisa, J.C.M.; Shimizu, A.; Sato, K.; Nasuda, S.; Takumi, S. Discovery of high-confidence single nucleotide polymorphisms from large-scale de novo analysis of leaf transcripts of *aegilops tauschii*, a wild wheat progenitor. *DNA Res.* **2012**, *19*, 487–497. [[CrossRef](#)]
52. Matsuda, R.; Iehisa, J.C.M.; Takumi, S. Application of real-time PCR-based SNP detection for mapping of Net2, a causal D-genome gene for hybrid necrosis in interspecific crosses between tetraploid wheat and *Aegilops tauschii*. *Genes Genet. Syst.* **2012**, *87*, 137–143. [[CrossRef](#)]
53. Dong, C.; Vincent, K.; Sharp, P. Simultaneous mutation detection of three homoeologous genes in wheat by high resolution melting analysis and mutation Surveyor®. *BMC Plant Biol.* **2009**, *9*, 143. [[CrossRef](#)]
54. Botticella, E.; Sestili, F.; Hernandez-Lopez, A.; Phillips, A.; Lafiandra, D. High resolution melting analysis for the detection of EMS induced mutations in wheat SbeIIa genes. *BMC Plant Biol.* **2011**, *11*, 156. [[CrossRef](#)]
55. Mondini, L.; Nachit, M.; Porceddu, E.; Pagnotta, M.A. Identification of SNP mutations in DREB1, HKT1, and WRKY1 genes involved in drought and salt stress tolerance in durum wheat (*Triticum turgidum* L. var *durum*). *Omics J. Integr. Biol.* **2012**, *16*, 178–187. [[CrossRef](#)]
56. Sestili, F.; Palombieri, S.; Botticella, E.; Mantovani, P.; Bovina, R.; Lafiandra, D. TILLING mutants of durum wheat result in a high amylose phenotype and provide information on alternative splicing mechanisms. *Plant Sci.* **2015**, *233*, 127–133. [[CrossRef](#)]
57. Lehmensiek, A.; Sutherland, M.W.; McNamara, R.B. The use of high resolution melting (HRM) to map single nucleotide polymorphism markers linked to a covered smut resistance gene in barley. *Theor. Appl. Genet.* **2008**, *117*, 721–728. [[CrossRef](#)]
58. Terracciano, I.; Maccaferri, M.; Bassi, F.; Mantovani, P.; Sanguineti, M.C.; Salvi, S.; Šimková, H.; Doležel, J.; Massi, A.; Ammar, K.; et al. Development of COS-SNP and HRM markers for high-throughput and reliable haplotype-based detection of Lr14a in durum wheat (*Triticum durum* Desf.). *Theor. Appl. Genet.* **2013**, *126*, 1077–1101. [[CrossRef](#)] [[PubMed](#)]
59. Curvers, K.; Pycke, B.; Kyndt, T.; Vanrompay, D.; Haesaert, G.; Gheysen, G. A high-resolution melt (HRM) assay to characterize CYP51 haplotypes of the wheat pathogen *Mycosphaerella graminicola*. *Crop. Prot.* **2015**, *71*, 12–18. [[CrossRef](#)]
60. Li, F.; Wen, W.; Zhang, Y.; Cao, S.; He, Z. Genetic architecture of grain yield in bread wheat based on genome-wide association studies. *BMC Plant Biol.* **2019**, *19*, 168. [[CrossRef](#)] [[PubMed](#)]
61. Hernandez, P.; Martis, M.; Dorado, G.; Pfeifer, M.; Gálvez, S.; Schaaf, S.; Jouve, N.; Šimková, H.; Valárik, M.; Doležel, J.; et al. Next-generation sequencing and syntenic integration of flow-sorted arms of wheat chromosome 4A exposes the chromosome structure and gene content. *Plant J.* **2012**, *69*, 377–386. [[CrossRef](#)] [[PubMed](#)]
62. Araki, E.; Miura, H.; Sawada, S. Identification of genetic loci affecting amylose content and agronomic traits on chromosome 4A of wheat. *Theor. Appl. Genet.* **1999**, *98*, 977–984. [[CrossRef](#)]

63. Borner, A.; Schumann, E.; Furst, A.; Coster, H.; Leithold, B.; Roder, S. Mapping of quantitative trait loci determining agronomic important characters in hexaploid wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* **2002**, *105*, 921–936. [[CrossRef](#)]
64. McCartney, C.A.; Somers, D.J.; Humphreys, D.G.; Lukow, O.; Ames, N.; Noll, J. Mapping quantitative trait loci controlling agronomic traits in the spring wheat cross RL445x'AC Domain'. *Genome* **2005**, *48*, 870–883. [[CrossRef](#)]
65. McDonald, G.K.; Gene, Y.; Nurzhanuly, B.; Trethowan, R.; Reynolds, M.; Yaqub Mujahid, M. Quantifying the value to grain yield of QTL for adaptation and tolerance to abiotic stress in bread wheat. In Proceedings of the 11th International Wheat Genetics Symposium, Brisbane Convention Exhibition Centre, Brisbane, QLD, Australia, 24–29 August 2008.
66. Liu, L.; Wang, L.; Yao, J.; Zheng, Y.; Zhao, C. Association mapping of six agronomic traits on chromosome 4A of wheat (*Triticum aestivum* L.). *Mol. Plant Breed.* **2010**, *1*. [[CrossRef](#)]
67. Jakobson, I.; Peusha, H.; Timofejeva, L.; Järve, K. Adult plant and seedling resistance to powdery mildew in a *Triticum aestivum* x *Triticum miitinae* hybrid line. *Theor. Appl. Genet.* **2006**, *112*, 760–766. [[CrossRef](#)]
68. Kirigwi, F.M.; Van Ginkel, M.; Brown-Guedira, G.; Gill, B.S.; Paulsen, G.M.; Fritz, A.K. Markers associated with a QTL for grain yield in wheat under drought. *Mol. Breed.* **2007**, *20*, 401–413. [[CrossRef](#)]
69. Rana, R.M.; Rehman, S.U.; Ahmed, J.; Bilal, M. A comprehensive overview of recent advances in drought stress tolerance research in wheat (*Triticum aestivum* L.). *Asian J. Agric. Biol.* **2013**, *1*, 29–37.
70. Shorinola, O.; Bird, N.; Simmonds, J.; Berry, S.; Henriksson, T.; Jack, P. The wheat Phs-A1 pre-harvest sprouting resistance locus delays the rate of seed dormancy loss and maps 0.3cM to the PM19 genes in UK germplasm. *J. Exp. Bot.* **2016**, *67*, 4169–4178. [[CrossRef](#)] [[PubMed](#)]
71. Acuña-Galindo, M.A.; Mason, R.E.; Subramanian, N.K.; Hays, D.B. Meta-analysis of wheat QTL regions associated with adaptation to drought and heat stress. *Crop. Sci.* **2015**, *55*, 477–492. [[CrossRef](#)]
72. Zhang, J.; Zhang, S.; Cheng, M.; Jiang, H.; Zhang, X.; Peng, C.; Lu, X.; Zhang, M.; Jin, J. Effect of drought on agronomic traits of rice and wheat: A meta-analysis. *Int. J. Environ. Res. Public Health* **2018**, *15*, 839. [[CrossRef](#)]
73. Murray, M.; Thompson, W. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* **1980**, *8*, 4321–4325. [[CrossRef](#)]
74. Hernández, P.; De la Rosa, R.; Rallo, L.; Martín, A.; Dorado, G. First evidence of a retrotransposon-like element in olive (*Olea europaea*): Implications in plant variety identification by SCAR-marker development. *Theor. Appl. Genet.* **2001**, *102*, 1082–1087. [[CrossRef](#)]
75. Gaujoux, R.; Seoighe, C. A flexible R package for nonnegative matrix factorization. *BMC Inform.* **2010**, *11*, 367. [[CrossRef](#)]
76. Liu, K.; Muse, S.V. PowerMaker: An integrated analysis environment for genetic marker analysis. *Bioinformatics* **2005**, *21*, 2128–2129. [[CrossRef](#)]
77. Nei, M. Genetic distance between populations. *Am. Nat.* **1972**, *106*, 283–292. [[CrossRef](#)]
78. Felsenstein, J. *Phylogeny Inference Package*; Version 3.6; Department of Genetics, University of Washington: Seattle, WA, USA, 2005.
79. Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **2013**, *30*, 2725–2729. [[CrossRef](#)]
80. Dighe, A.S.; Jangid, K.; González, J.M.; Pidiyar, V.J.; Patole, M.S.; Ranade, D.R.; Shouche, Y.S. Comparison of 16S rRNA gene sequences of genus *Methanobrevibacter*. *BMC Microbiol.* **2004**, *4*, 20. [[CrossRef](#)]
81. Kumar, A.; Hirochika, H. Applications of retrotransposons as genetic tools in plant biology. *Trends Plant Sci.* **2001**, *6*, 127–134. [[CrossRef](#)]
82. Schulman, A.H.; Flavell, A.J.; Paux, E.; Noel Ellis, T.H. The Application of LTR Retrotransposons as Molecular Markers in Plants. *Methods Mol. Biol.* **2012**, *859*, 115–153. [[CrossRef](#)] [[PubMed](#)]
83. Barabaschi, D.; Orru, L.; Lacrima, K.; Faccioli, P.; Colaiacovo, M.; Bagnaresi, P. On the road to a high density genetic linkage map of wheat chromosome 5A. *J. Med. Plants Res.* **2010**, *150*, 473. [[CrossRef](#)]
84. Lucas, S.J.; Šimková, H.; Šafář, J.; Jurman, I.; Cattonaro, F.; Vautrin, S.; Bellec, A.; Berges, H.; Doležel, J.; Budak, H. Functional features of a single chromosome arm in wheat (1AL) determined from its structure. *Funct. Integr. Genom.* **2012**, *12*, 173–182. [[CrossRef](#)] [[PubMed](#)]
85. Druml, B.; Cichna-Marx, M. High resolution melting (HRM) analysis of DNA—Its role and potential in food analysis. *Food Chem.* **2014**, *158*, 245–254. [[CrossRef](#)]

86. Hoffmann, M.; Hurlbaeus, J.; Weilke, C. High-resolution melting curve analysis on the LightCycler® 480 PCR system. *Nat. Methods Suppl.* **2007**, *S*, AN17–AN18.
87. Wittwer, C.T. High-resolution DNA melting analysis: Advancements and limitations. *Hum. Mutat.* **2009**, *30*, 857–859. [[CrossRef](#)]
88. Distefano, G.; La Malfa, S.; Gentile, A.; Wu, S.B. EST-SNP genotyping of citrus species using high-resolution melting curve analysis. *Tree Genet. Genom.* **2013**, *9*, 1271–1281. [[CrossRef](#)]
89. Liew, M.; Pryor, R.; Palais, R.; Meadows, C.; Erali, M.; Lyon, E.; Wittwer, C. Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. *Clin. Chem.* **2004**, *50*, 1156–1164. [[CrossRef](#)]
90. Erali, M.; Palais, R.; Wittwer, C. SNP Genotyping by Unlabeled Probe Melting Analysis. In *Molecular Beacons: Signalling Nucleic Acid Probes, Methods, and Protocols*; Marx, A., Seiz, O., Eds.; Humana Press: Totowa, NJ, USA, 2008; p. 429. [[CrossRef](#)]
91. Tindall, E.A.; Petersen, D.C.; Woodbridge, P.; Schipany, K.; Hayes, V.M. Assessing high-resolution melt curve analysis for accurate detection of gene variants in complex DNA fragments—Tindall—2009—Human Mutation—Wiley Online Library. *Hum. Mutat.* **2009**, *30*, 876–883. [[CrossRef](#)] [[PubMed](#)]
92. Reed, G.H.; Kent, J.O.; Wittwer, C.T. High-resolution DNA melting analysis for simple and efficient molecular diagnostics. *Pharmacogenomics* **2007**, *8*, 597–608. [[CrossRef](#)] [[PubMed](#)]
93. Ganopoulos, I.; Argiriou, A.; Tsaftaris, A. Microsatellite high resolution melting (SSR-HRM) analysis for authenticity testing of protected designation of origin (PDO) sweet cherry products. *Food Control* **2011**, *22*, 532–541. [[CrossRef](#)]
94. Ganopoulos, I.; Madesis, P.; Zambounis, A.; Tsaftaris, A. High-resolution melting analysis allowed fast and accurate closed-tube genotyping of *Fusarium oxysporum* formae speciales complex. *FEMS Microbiol. Lett.* **2012**, *334*, 16–21. [[CrossRef](#)] [[PubMed](#)]
95. Ganopoulos, I.; Xanthopoulou, A.; Mastrogrianni, A.; Drouzas, A.; Kalivas, A.; Bletsos, F.; Krommydas, S.K.; Ralli, P.; Tsaftaris, A.; Madesis, P. High Resolution Melting (HRM) analysis in eggplant (*Solanum melongena* L.): A tool for microsatellite genotyping and molecular characterization of a Greek Genebank collection. *Biochem. Syst. Ecol.* **2015**, *58*, 64–71. [[CrossRef](#)]
96. MacKay, J.F.; Wright, C.D.; Bonfiglioli, R.G. A new approach to varietal identification in plants by microsatellite high resolution melting analysis: Application to the verification of grapevine and olive cultivars. *Plant Methods* **2008**, *4*, 1–10. [[CrossRef](#)]
97. Madesis, P.; Ganopoulos, I.; Anagnostis, A.; Tsaftaris, A. The application of Bar-HRM (Barcode DNA-High Resolution Melting) analysis for authenticity testing and quantitative detection of bean crops (Leguminosae) without prior DNA purification. *Food Control* **2012**, *25*, 576–582. [[CrossRef](#)]
98. Wu, S.B.; Wirthensohn, M.G.; Hunt, P.; Gibson, J.P.; Sedgley, M. High resolution melting analysis of almond SNPs derived from ESTs. *Theor. Appl. Genet.* **2008**, *118*, 1–14. [[CrossRef](#)]
99. Culbertson, A.T.; Ehrlich, J.J.; Choe, J.Y.; Honzatko, R.B.; Zobotina, O.A. Structure of xyloglucan xylosyltransferase 1 reveals simple steric rules that define biological patterns of xyloglucan polymers. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 6064–6069. [[CrossRef](#)]
100. Kaur, K.; Gupta, A.K.; Kaur, N. Effect of water deficit on carbohydrate status and enzymes of carbohydrate metabolism in seedlings of wheat cultivars. *Indian J. Biochem. Biophys.* **2007**, *44*, 223–230.
101. Guo, R.; Shi, L.; Jiao, Y.; Li, M.; Zhong, X.; Gu, F.; Liu, Q.; Xia, X.; Li, H. Metabolic responses to drought stress in the tissues of drought-tolerant and drought-sensitive wheat genotype seedlings. *AoB Plants* **2018**, *10*, 1–13. [[CrossRef](#)] [[PubMed](#)]
102. Abebe, T.; Melmaiee, K.; Berg, V.; Wise, R.P. Drought response in the spikes of barley: Gene expression in the lemma, palea, awn, and seed. *Funct. Integr. Genom.* **2010**, *10*, 191–205. [[CrossRef](#)] [[PubMed](#)]
103. Mindrebo, J.T.; Nartey, C.M.; Seto, Y.; Burkart, M.D.; Noel, J.P. Unveiling the functional diversity of the alpha/beta hydrolase superfamily in the plant kingdom. *Curr. Opin. Struct. Biol.* **2016**, *41*, 233–246. [[CrossRef](#)] [[PubMed](#)]
104. Sharma, M.; Pandey, G.K. Expansion and function of repeat domain proteins during stress and development in plants. *Front. Plant Sci.* **2016**, *6*, 1218. [[CrossRef](#)] [[PubMed](#)]
105. Hunt, M.R. Biological roles of small RNAs expressed during infection of barley by the obligate fungal biotroph, *Blumeria graminis* f. sp. *hordei*. Master's Thesis, Iowa State University, Ames, IA, USA, 2018.
106. Shibagaki, N.; Grossman, A.R. The role of the STAS domain in the function and biogenesis of a sulfate transporter as probed by random mutagenesis. *J. Biol. Chem.* **2006**, *281*, 22964–22973. [[CrossRef](#)] [[PubMed](#)]

107. Shibagaki, N.; Grossman, A.R. Probing the function of STAS domains of the Arabidopsis sulfate transporters. *J. Biol. Chem.* **2004**, *279*, 30791–30799. [[CrossRef](#)]
108. Rouached, H.; Berthomieu, P.; El Kassis, E.; Cathala, N.; Catherinot, V.; Labesse, G. Structural and Functional Analysis of the C-terminal STAS (Sulfate Transporter and Anti-sigma Antagonist) domain of the Arabidopsis thaliana sulfate transporter SULTR1.1. *J. Biol. Chem.* **2005**, *280*, 15976–15983. [[CrossRef](#)]
109. Tabe, L.; Hagan, N.; Higgins, T.J.V. Plasticity of seed protein composition in response to nitrogen and sulfur availability. *Curr. Opin. Plant Biol.* **2002**, *5*, 212–217. [[CrossRef](#)]
110. Miazek, A.; Zagdańska, B. Involvement of exopeptidases in dehydration tolerance of spring wheat seedlings. *Biol. Plant.* **2008**, *52*, 687–694. [[CrossRef](#)]
111. Bray, E.A. Plant responses to water deficit. *Trends Plant Sci.* **1997**, *2*, 48–54. [[CrossRef](#)]
112. Vierstra, R.D. Proteolysis in plants: Mechanism and functions. *Plant Mol. Biol.* **2004**, *32*, 275–302. [[CrossRef](#)] [[PubMed](#)]
113. Simova-Stoilova, L.; Vaseva, I.; Grigorova, B.; Demirevska, K.; Feller, U. Proteolytic activity and cysteine protease expression in wheat leaves under severe soil drought and recovery. *Plant Physiol. Biochem.* **2010**, *48*, 200–206. [[CrossRef](#)] [[PubMed](#)]
114. Bohlmann, H.; Apel, K. Thionins. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1991**, *42*, 227–240. [[CrossRef](#)]
115. Begcy, K.; Walia, H. Drought stress delays endosperm development and misregulates genes associated with cytoskeleton organization and grain quality proteins in developing wheat seeds. *Plant Sci.* **2015**, *240*, 109–119. [[CrossRef](#)] [[PubMed](#)]
116. Ye, Y.; Ding, Y.; Jiang, Q.; Wang, F.; Sun, J.; Zhu, C. The role of receptor-like protein kinases (RLKs) in abiotic stress response in plants. *Plant Cell Rep.* **2017**, *36*, 235–242. [[CrossRef](#)]
117. Cabouté, M.-E.; Clément, B.; Sekine, M.; Philipps, G.; Chaubet-Gigot, N. Cell cycle regulation of the tobacco ribonucleotide reductase small subunit gene is mediated by E2F-like elements. *Plant Cell* **2000**, *12*, 1987–1999. [[CrossRef](#)]
118. Sauge-Merle, S.; Laulhère, J.P.; Covès, J.; Le Pape, L.; Ménage, S.; Fontecave, M. Ribonucleotide reductase from the higher plant Arabidopsis thaliana: Expression of the R2 component and characterization of its iron-radical center. *J. Biol. Inorg. Chem.* **1997**, *2*, 586–594. [[CrossRef](#)]
119. Sparkes, I. Recent advances in understanding plant myosin function: Life in the fast lane. *Mol. Plant* **2011**, *4*, 805–812. [[CrossRef](#)]
120. Vogt, T.; Jones, P. Glycosyltransferases in plant-natural product synthesis: Characterization of a supergene family. *Trends Plant Sci.* **2000**, *5*, 380–386. [[CrossRef](#)]
121. Akdogan, G.; Tufekci, E.D.; Uranbey, S.; Unver, T. miRNA-based drought regulation in wheat. *Funct. Integr. Genom.* **2016**, *16*, 221–233. [[CrossRef](#)]
122. Kang, J.; Park, J.; Choi, H.; Burla, B.; Kretschmar, T.; Lee, Y.; Martinoia, E. Plant ABC Transporters. *Arabidopsis Book* **2011**, *9*, e0153. [[CrossRef](#)] [[PubMed](#)]
123. Rampino, P.; Mita, G.; Fasano, P.; Borrelli, G.M.; Aprile, A.; Dalessandro, G.; De Bellis, L.; Perrotta, C. Novel durum wheat genes up-regulated in response to a combination of heat and drought stress. *Plant Physiol. Biochem.* **2012**, *56*, 72–78. [[CrossRef](#)]
124. Walter, S.; Kahla, A.; Arunachalam, C.; Perochon, A.; Khan, M.R.; Scofield, S.R.; Doohan, F.M. A wheat ABC transporter contributes to both grain formation and mycotoxin tolerance. *J. Exp. Bot.* **2015**, *66*, 2583–2593. [[CrossRef](#)]
125. Kulik, A.; Wawer, I.; Krzywińska, E.; Bucholc, M.; Dobrowolska, G. SnRK2 protein Kinases—Key regulators of plant response to abiotic stresses. *Omics J. Integr. Biol.* **2011**, *15*, 859–872. [[CrossRef](#)] [[PubMed](#)]
126. Wei, K.; Li, Y. Functional genomics of the protein kinase superfamily from wheat. *Mol. Breed.* **2019**, *39*, 141. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



MDPI  
St. Alban-Anlage 66  
4052 Basel  
Switzerland  
Tel. +41 61 683 77 34  
Fax +41 61 302 89 18  
[www.mdpi.com](http://www.mdpi.com)

*Agronomy* Editorial Office  
E-mail: [agronomy@mdpi.com](mailto:agronomy@mdpi.com)  
[www.mdpi.com/journal/agronomy](http://www.mdpi.com/journal/agronomy)





MDPI  
St. Alban-Anlage 66  
4052 Basel  
Switzerland

Tel: +41 61 683 77 34  
Fax: +41 61 302 89 18

[www.mdpi.com](http://www.mdpi.com)



ISBN 978-3-03943-864-8