

Article

Deciphering Early Movements and Domestication of *Coffea arabica* through a Comprehensive Genetic Diversity Study Covering Ethiopia and Yemen

Christophe Montagnon ^{1,*}, Faris Sheibani ², Tadesse Benti ³, Darrin Daniel ⁴ and Aduugna Debela Bote ⁵

¹ RD2 Vision, 60 rue du Carignan, 34270 Valflaunes, France

² Qima Coffee, 21 Warren Street, Fitzrovia, London W1T 5LT, UK

³ Department of Coffee Breeding and Genetics, Ethiopian Institute of Agricultural Research (EIAR), Jimma Research Center, Jimma, Ethiopia

⁴ Alliance for Cup of Excellence, 2250 NW 22nd Ave Suite 612, Portland, OR 97210, USA

⁵ Ethiopia Coffee and Tea Authority, Addis Ababa, Ethiopia

* Correspondence: christophe.montagnon@rd2vision.com

Citation: Montagnon, C.; Sheibani, F.; Benti, T.; Daniel, D.; Bote, A.D. Deciphering Early Movements and Domestication of *Coffea arabica* through a Comprehensive Genetic Diversity Study Covering Ethiopia and Yemen. *Agronomy* **2022**, *12*, 3203. <https://doi.org/10.3390/agronomy12123203>

Academic Editors: Evangelia V. Avramidou, Aliko Kapazoglou and Ermioni Malliarou

Received: 31 October 2022

Accepted: 13 December 2022

Published: 16 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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 (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The coffee species *Coffea arabica* is facing numerous challenges regarding climate change, pests and disease pressure. Improved varieties will be part of the solution. Making optimal use of the scarce genetic diversity of the species is hence essential. In this paper, we present the first study of *C. arabica* genetic diversity covering its complete native habitat in Ethiopia together with its main domestication centers: Yemen and Hararghe region in Ethiopia. All in all, 555 samples were analyzed with a set of Single Sequence Repeat markers. Through admixture genetic analysis, six clusters were identified. A total of two “Core Ethiopian” clusters did not participate in the domestication of the species. There were four clusters that were part of the “Domestication Pathway” of *C. arabica*. The first one was named “Ethiopian Legacy” as it represents the genetic link between “Core Ethiopia” and the “Domestication Pathway” in Yemen and Hararghe. The geographic origin of this cluster in Ethiopia was the south of Ethiopia, namely Gedio, Guji and Sidama, which hence appears as the source of coffee seeds that led to the domestication of *C. arabica*. In Yemen, in addition to the “Ethiopian Legacy” cluster, we confirmed the “Typica/Bourbon” and “New-Yemen” clusters. In Hararghe, the “Harrar” cluster, never described before, likely originates from a re-introduction of domesticated coffee from Yemen into this region of Ethiopia. Cultivated varieties around the world today originate from the “Ethiopian Legacy” and “Typica/Bourbon” clusters and but none are related to the “new-Yemen” and “Harrar” clusters. Implications for breeding strategies are discussed.

Keywords: breeding; gene flow; genetic bottleneck

1. Introduction

Coffea arabica is an amphidiploid formed by hybridization between *C. eugenoides* and *C. canephora*, or ecotypes related to these diploid species [1–3]. The “natural habitats” are defined as “areas composed of viable assemblages of plant and/or animal species of largely native origin and/or where human activity had not essentially modified an area’s primary ecological functions and species composition” [4]. The southwestern and southern mountains of Ethiopia on both sides of the Rift Valley have been accepted since the 1960’s as the main and potentially sole natural habitat of *C. arabica* [5,6]. Whether only the west of the Rift Valley or both western and eastern ridges of the Rift Valley are the natural habitat of *C. arabica* has been previously discussed by Montagnon and Bouharmont [7], and Lashermes et al. [8]. Davis et al. [9] concluded from a climate suitability study that there was no reason to disregard the southern mountains on the

eastern ridge of the Rift Valley as a natural habitat for *C. arabica*, as it is the case for numerous species that are indigenous to both parts of the Rift Valley [10]. South Sudan was also proposed [11] and recently confirmed [12] as a natural habitat for the species. Mount Marsabit in Kenya was also suggested as a possible natural habitat for *C. arabica* [13]. However, further genetic [8] and climatic suitability [9] studies support a recent introduction of *C. arabica* in Mount Marsabit.

The Hararghe region, further east from Ethiopia (Oromia zone), and Yemen's climate are clearly not compatible with being a natural habitat for *C. arabica* [9]. However, Yemen is acknowledged to be the place where *C. arabica* was first cultivated [5,6,14–19]. The Hararghe region is therefore an important place in which early coffee cultivation took place. Yet no study has ever established the origin of coffee cultivated to be Hararghe. As such, coffee in Hararghe either originated (i) from the south-west or south Ethiopian native habitat of *C. arabica* and/or (ii) from Yemen through a reintroduction of coffee plants and cultivation know-how [3,5,6,20,21].

The genetic diversity of *C. arabica* is one of the lowest found in crops [3]. In its Ethiopian native habitat, a slight structuration of the genetic diversity of *C. arabica*, often corresponding to a west–east geographical pattern [3,7,22,23]. However, other studies have shown that movement of coffee seeds inside Ethiopia and across the Rift Valley possibly blurred a possible and likely initial genetic structure [24–26]. Yet a consistent and clear genetic separation has been established between *C. arabica* plants found today in their native habitat as compared to the cultivated *C. arabica* varieties worldwide [3,12,21–23,27]. Montagnon et al. [21], demonstrated that most coffee varieties cultivated worldwide derived from two out of three mother populations found in Yemen: “Typica/Bourbon” and “SL-34”. The third mother population found in Yemen (“New-Yemen”) could not be genetically related to any populations in Ethiopia or to varieties cultivated outside Yemen.

Most scientific publications on *C. arabica* genetic diversity in Ethiopia are based on the accessions surveyed in Ethiopia by the FAO [28] and French Orstom [29]. Both surveys fell short in covering all of the potential natural habitats of coffee in Ethiopia. Furthermore, the surveys took place in the 1960's, almost 60 years ago and there is a need for a more up to date survey. Genetic diversity might have decreased in its natural habitat since the previous surveys. The loss of genetic diversity in natural habitat is well documented in crops in general [30]. In Côte d'Ivoire, the natural habitat for *Coffea canephora* and the risk of losing local genetic diversity was recently highlighted [31]. Moreover, there are some serious concerns about the negative effect of climate change on wild Arabica coffee trees in Ethiopia [9].

The genetic background of coffee grown in Hararghe is unknown. Scalabrin et al. [3] indicated that Yemen and Hararghe coffee were genetically related. However, while this statement makes sense from a historic perspective and would correspond to some unwritten farmers' memories [6,17], the study did not include enough accessions from Hararghe and Yemen to be conclusive. While Montagnon et al. [21] revealed for the first time the intra Yemen coffee genetic diversity, the exact genetic link between Ethiopia and Yemen is yet to be understood.

In this article, we represent for the first time in one single study (i) more than 350 samples from Ethiopia covering a wide geographical range, including Hararghe and (ii) close to 200 samples from various coffee areas in Yemen. The objectives of the study are (i) to precisely identify the genetic diversity of cultivated *C. arabica* in Ethiopia, (ii) to assess the genetic identity of *C. arabica* cultivated in Hararghe and (iii) to decipher the genetic relation between Ethiopian south–west and south, Hararghe and Yemeni coffee germplasm.

2. Materials and Methods

A total of 555 samples were included in the study. There were 356 samples from Ethiopia covering the main coffee growing areas of the country (Table 1 and Figure 1). We

took the opportunity of including samples from the 2020 and 2021 Cup of Excellence competition in Ethiopia organized by the Coffee and Tea Authority of Ethiopia and the Alliance for Coffee Excellence, in which farmers from all over the country send samples of their production to participate in the competition. A total of two hundred samples were randomly picked in 2020 and 158 in 2021. It was ensured that the main regions, including Hararghe, were represented (Figure 1) in the sample sets. In our study, we term the “southwest of Ethiopia” to refer to the coffee region in the west of the Rift Valley One, ‘south of Ethiopia’ to refer to the coffee regions in the immediate east of the Rift and “east” or “Hararghe” to refer to the coffee region of east and west Hararghe which included samples from the northeastern part of Arsi. The southwest and south regions are corresponding to the native habitat of *C. arabica* in Ethiopia [9]. The east region is out of the native habitat area. One single bean was selected at random from each sample and analyzed in line with general recommendations for the observation of the genetic diversity [32]. As the cost of reaching distinct locations was low, which is the case when the samples are sent to a central place, it was optimal to increase the number of locations and decrease the number of individuals per location to one. All samples were anonymized.

Table 1. Geographical origin of coffee samples from Ethiopia and Yemen included in the study.

High Level Areas	Ethiopia	Yemen	
	Woreda (District)	# Samples	Governorate # Samples
Southwest	Awri/Agew	1	Sadaa 2
	Bench Maji	4	Sanaa 84
	Hadiya	2	Mahwit 25
	Horo guduru	1	Dhamar 31
	Ilubabor	9	Ibb 44
	Jimma	66	Total 186
	Keffa	9	
	Kelem Wellega	4	
	KT	1	
	Majang	1	
	Sheka	5	
	South Omo	1	
	West Shewa	3	
	West Wellega	6	
Southwest	Sub Total	113	
South	Borena	1	
	Gedio	51	
	Guji	54	
	Sidama	41	
	West Arsi	65	
South	Sub Total	212	
East (Hararghe)	Arsi	2	
	East Hararghe	20	
	West Hararghe	9	
East (Hararghe)	Sub Total	31	
Total		356	

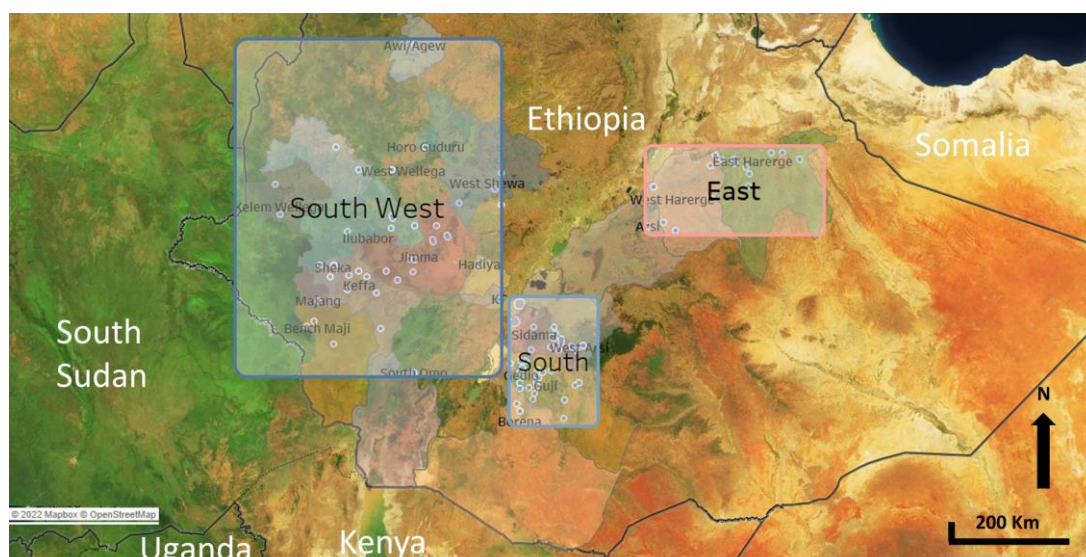


Figure 1. Geographical origin of Ethiopian coffee samples included in the study. Samples originate from several Woreda (districts) located in three regions: Southwest, south and east. Southwest and south regions correspond to the natural habitat of *C. arabica* in Ethiopia.

Main Yemeni coffee producing governorates are highlighted. Samples in the study originate from governorates highlighted in red. The main historical ports for entry and exit of coffee beans in and from Yemen are indicated: Mocha and Aden.

From Yemen, 186 green coffee samples were originating from several coffee governorates and growing areas (Figure 2, Table 1). Again, one single bean was analyzed for each sample.



Figure 2. Geographical origin of Yemeni coffee samples included in the study.

A total of nine important varieties representing the genetic diversity of cultivated *C. arabica* around the world [21] were added: Bourbon, Typica, K-7, Kent, SL-09, SL-14, SL-17, SL-28 and SL-34. Finally, four varieties with a clear Ethiopian genetic background [21] were also added: Chiroso, Gesha, Pink Bourbon and SL-06.

DNA extraction and SSR marker analysis was performed by the ADNid laboratory of the Qualtech company located in the South of France (<http://www.qualtech-groupe.com/en/> accessed on 9 November 2022). Genomic DNA was extracted from 20 mg of dried tissue using 1 mL of SDS buffer. Then, DNA was purified with magnetic beads

(Agencourt AMPure XP, Beckman Coulter, Brea, California, USA) and then eluted in Tris EDTA (TE) buffer. The DNA concentration was checked with an Enspire spectrofluorimeter (Perkin Elmer) with a bisbenzimidazole DNA intercalator (Hoechst 33258) and using a known standard of DNA for comparison.

Ten SSR (Single Sequence Repeat) primer pairs were used (Table 2). Eight were selected after their wide discrimination power was confirmed by [33,34]. Two other SSR primer pairs were included (Sat-207 and Sat-244) [21,35]. PCR was run in a final volume of 15 µL including 30 ng genomic DNA and 7.5 µL of 2× PCR buffer (Type-it Microsatellite PCR Kit, Qiagen, Hilden, Germany), 1.0 µM each of forward and reverse primer (10 µM). Amplifications were performed in thermal cycler (Eppendorf, Hamburg, Germany) programmed at 94 °C for 5 min for initial denaturation, followed by 94 °C for 30 s, annealing temperature depending on the primer used for 30 s and 72 °C for 1 min for 35 cycles followed by a final step of extension at 72 °C for 5 min. Final holding temperature was 4 °C. PCR samples were run on a capillary electrophoresis, ABI 3130XL with an internal standard: GeneScan 500 LIZ size standard (Applied Biosystems, Waltham, MA, USA). Alleles were scored with GeneMapper v.4.1 software (Applied Biosystems) and then visually inspected.

Table 2. List of microsatellite markers with their locus code, primer sequences, and product size.

Code SSR	Primer Sequence Forward	Primer Sequence Reverse	Size Product (Base Pairs)
Sat-11	ACCCGAAAGAAAGAACCAA	CCACACAACCTCTCCTCATTC	143–145
Sat-207	CAATCTCTTTCCGATGCTCT	GAAGCCGTTTCAAGCC	83–93
Sat-225	CATGCCATCATCAATTCCAT	TTACTGCTCATCATTCGCCA	283–317
Sat-235	TCGTTCTGTCATTAAATCGTCAA	GCAAATCATGAAAATAGTT GGTG	245–278
Sat-24	GGCTCGAGATATCTGTTTAG	TTTAATGGGCATAGGGTCC	167–181
Sat-244	GCATACTAAGGAATTATCTGAC TGCT	GCATGTGCTTTTGTATGTGC T	178–306
Sat-254	ATGTTCTTCGCTTCGCTAAC	AAGTGTGGGAGTGTCTGCAT	221–237
Sat-29	GACCATTACATTTCACACAC	GCATTTTGTTCACACTGTA	137–154
Sat-32	AACTCTCCATTCGCCGATTC	CTGGGTTTCTGTGTTCTCG	119–125
Sat-47	TGATGGACAGGAGTTGATGG	TGCCAATCTACCTACCCCTT	135–169

The structure of the population and the number of clusters (K) was estimated through Structure v2.3.4 software [35,36] with burnin period length of 50,000 and 50,000 replicates of Markov Chain Monte Carlo after burnin. The geographical location of samples was not used in the model. The number of clusters (K) was allowed to vary from 2 to 8 and the log likelihood was estimated as its average value for 10 runs per K.

In parallel, DARwin6 software [37] was used to produce a dissimilarity matrix using Dice Index and then perform a Principal Coordinates Analysis (PCoA). A Hierarchical Ascending Classification (HAC) or cluster analysis was performed using Xlstat software [38] based on the first five components of the PCoA using the Ward method [39].

As advised by Paterson et al. [40], the number of clusters (K) was chosen based on both the log likelihood values and the genetic structure as observed after PCoA and HAC. Once K was fixed, Structure software was used to produce the admixture model based on the ancestry coefficients of each single individual [35,36].

In line with previous studies [21,34,41], *C. arabica* being tetraploid, SSR allelic phenotype rather than genotype is observed. In this case, SSRs are to be treated as dominant markers. Indeed, for instance, the genotypes AABB, ABAB, AAAB or ABBB are all corresponding to the observed AB phenotype. Hence, for Structure software ploidy was set to 4 and RECESSIVEALLELES code to MISSING. A single data file with allele scoring 0 (absence) and 1 (present) was used for DARwin6 software.

Tableau software [42] was used to produce all figures except Figures 3 and S1.

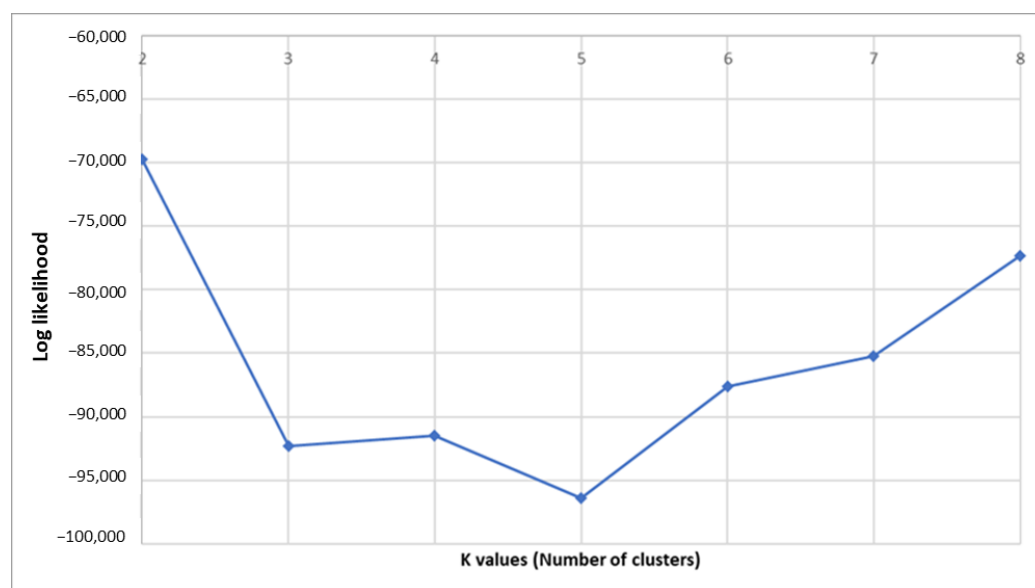


Figure 3. Log likelihood (average of 10 runs) for various K values from the admixture model run on the whole population of 555 *C. arabica* samples.

3. Results

The different log likelihood values of the Admixture model for K ranging from 2 to 8 (Figure 3) show a drop after K = 2 and rising again for K = 6 to K = 8.

The cluster analysis run on the first five components of the PCoA including all the samples, revealed two main clusters which could be further split in two and four sub-clusters, respectively (Figure S1). Combining the information from the log likelihood values and the cluster analysis, the K value was set to K = 6.

Based on the ancestry coefficients—or probability of each individual to derive from each of the six clusters—the admixture model was produced (Figure 4).

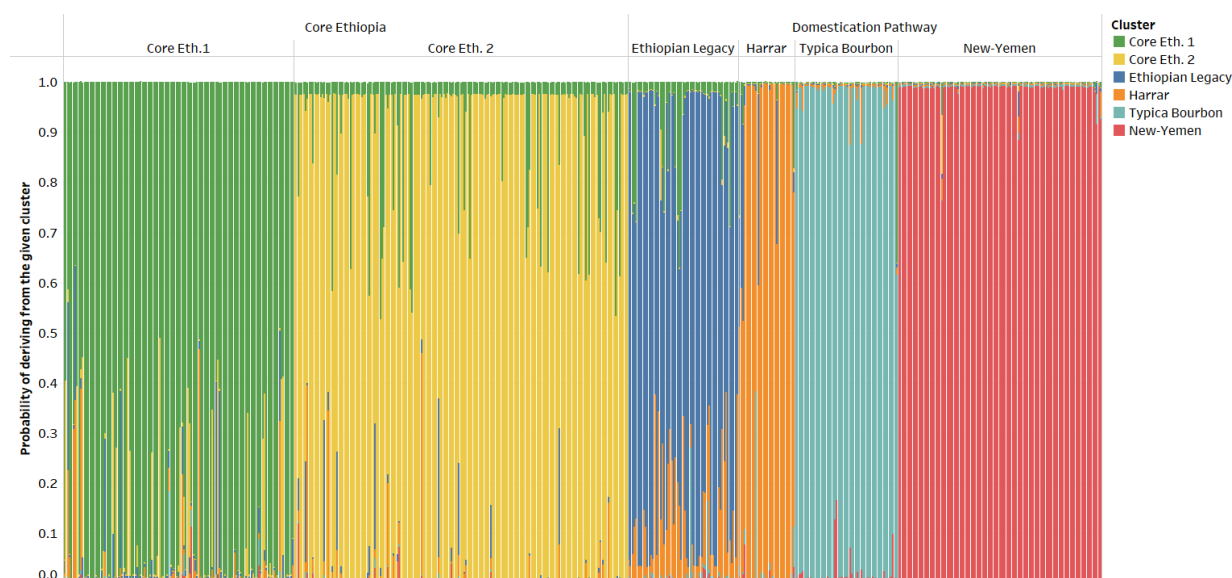


Figure 4. Ancestry coefficient for each individual obtained from the admixture model for the population of 555 *C. arabica* samples with K = 6.

A total of two clusters included only samples from Ethiopia, the vast majority of which (99%) were from the southwest and south of Ethiopia—corresponding to the native habitat of *C. arabica*—while the remaining 1% were from the Hararghe region (Table 3). As expected, the four samples corresponding to Ethiopian landraces cultivated out of

Ethiopia were part of these two clusters. We identified these clusters as “Core Eth. 1” and “Core Eth. 2” and represented together “Core Ethiopian” samples. There was no clear geographical pattern for these two clusters even if “Core Eth. 2” was more represented than “Core Eth. 1” in the south (Table 3).

Table 3. Repartition of studied coffee samples according to their geographical origin and genetic cluster.

Samples Origin	Areas	Core Ethiopia		Domestication Pathway				Harrar	Total
		Core Eth. 1	Core Eth. 2	Ethiopian Legacy	Typica Bourbon	New-Yemen			
Ethiopia COE Survey	Southwest	62	46	5					113
	South	54	133	17	1	1	6		212
	East/Hararghe	3		3		2	23		31
Yemen Survey				29	50	106	1		186
Varieties cultivated worldwide				5	4				9
Ethiopian landraces cultivated worldwide		4							4
Total		123	179	59	55	109	30		555
Total allele number		65	45	39	30	30	29		77

The other four clusters consisted of all the Yemeni samples, all the varieties cultivated around the world, most samples from Hararghe as well as some few samples from the southwest and south Ethiopian region with more samples from the south. These clusters represent the “Domestication Pathway” of *C. arabica*. One cluster consisted mostly of Yemeni samples together with four cultivated varieties worldwide—Typica, Bourbon, Kent and SL28- together with only one sample from the south of Ethiopia. This cluster was the equivalent of the “Typica/Bourbon” mother population of Montagnon et al. [21]. Another cluster was composed of 109 samples, 106 of which were from Yemen, 2 from Hararghe and one from the south of Ethiopia. This cluster was the equivalent of the “New-Yemen” mother population of Montagnon et al. [21]. A third cluster was composed of 30 samples, 23 of which were from Hararghe, 6 from the south of Ethiopia and 1 from Yemen. We named this cluster “Harrar”. A fourth cluster was made up of samples covering the different geographical areas of the study with the highest number in the south of Ethiopia and in Yemen. Five varieties cultivated worldwide were part of this cluster: SL-09, SL-14, SL-17, SL-34 and K-7. This cluster included both the SL-34 and SL-17 mother populations of Montagnon et al. [21]. Because this cluster appears to be a link between Ethiopia and Yemen, we named it “Ethiopian Legacy”.

Out of the 17 samples of the “Ethiopian Legacy” cluster in south Ethiopia, 8 and 6 were from Gedio—which includes Yirgachefe—and Guji zones, respectively, while 3 other samples were from Sidama (Table S1). In Yemen, the vast majority of the “Ethiopian Legacy” cluster was found in Ibb (Table S1). The “Typica/Bourbon” cluster was found mainly in Ibb and Dhamar with only a few samples in Saada and Sanaa. A total of 88% of the “New-Yemen” samples were from Sanaa (64%) and Mawhit (23%).

Admixture was observed namely between Core Eth. 1 and 2 (Figure 4). Admixture with Harrar genetic background was observed on few occasions in Core Eth. 1 and 2. And to a higher degree in Ethiopian legacy. However, only little admixture was found in Harrar cluster itself and only from Ethiopian Legacy. “Typica/Bourbon” and “New-Yemen” presented almost no admixture. Core Ethiopian genetic background was not found in “Harrar”, “Typica/Bourbon” or “New-Yemen” clusters. Plotting the ancestry coefficients for each pair of clusters confirms the absence of admixture of “Typica/Bourbon” and “New-Yemen” with any other clusters (Figure S2).

Figure 5 summarizes the relationship between genetic clusters and geographical origins. Both “core Eth.1 and 2” clusters are present only on both the southwest and south

of Ethiopia with few representatives in Hararghe. The “Ethiopian Legacy” cluster is the genetic cluster with more geographical diversity as it is present in all the regions: it is mostly represented in the south of Ethiopia and Yemen, but also to a lesser extent in the southwest of Ethiopia and Hararghe. Most of the “Harrar” cluster is in Hararghe with few representatives in the south of Ethiopia and in Yemen. “New-Yemen” and “Typica/Bourbon” are almost exclusively present in Yemen, with only two representatives of “New-Yemen” from Hararghe and one representative of “Typica/Bourbon” from the south of Ethiopia.



Figure 5. Schematic relationship between genetic clusters and geographical origins of *C. arabica* samples in the study. For each geographical origin (up), the colored area is proportional to the % of each cluster in the given geographical origin. The sum of % in each geographical origin is equal to 100. For each K cluster (bottom), the colored area is proportional to the corresponding % of geographical origin. The sum of % in each cluster is equal to 100.

Reciprocally (Figure 5), the southwest of Ethiopia is made of almost only “Core Eth. 1 and 2” clusters (96%) with only 4% of “Ethiopian Legacy” cluster. The south of Ethiopia is more diverse with 63% of “Core Eth. 2”, 25% of “Core Eth. 1”, 8% of “Ethiopian Legacy” and few “Harrar” and “Typica/Bourbon” (less than 3% each). Hararghe also demonstrates some genetic diversity. The vast majority (74%) of samples of Hararghe are from the “Harrar” genetic cluster, with 10% from the “Ethiopian Legacy” cluster, and less than 10% from both Core Ethiopian clusters. Interestingly, two “New-Yemen” samples were found in Hararghe. In Yemen, “New-Yemen”, “Typica/Bourbon” and “Ethiopian Legacy” represent 57%, 27% and 16% of the samples, respectively. Interestingly, one sample in Yemen was found to represent the “Harrar” genetic cluster.

Figure 6 shows the PCoA result with each individual sample and the barycenter of each genetic cluster. The first component explains the greatest part (40%) of the overall

variation and discriminates the two “Core Ethiopia” clusters from the “Domestication Pathway” clusters with “Ethiopian Legacy” being the closest to the “Core Ethiopia” clusters. The second axis explains much less variation (7%) and is discriminating between the “Domestication Pathway” clusters: “Typica/Bourbon”, “New-Yemen” and “Harrar” clusters.

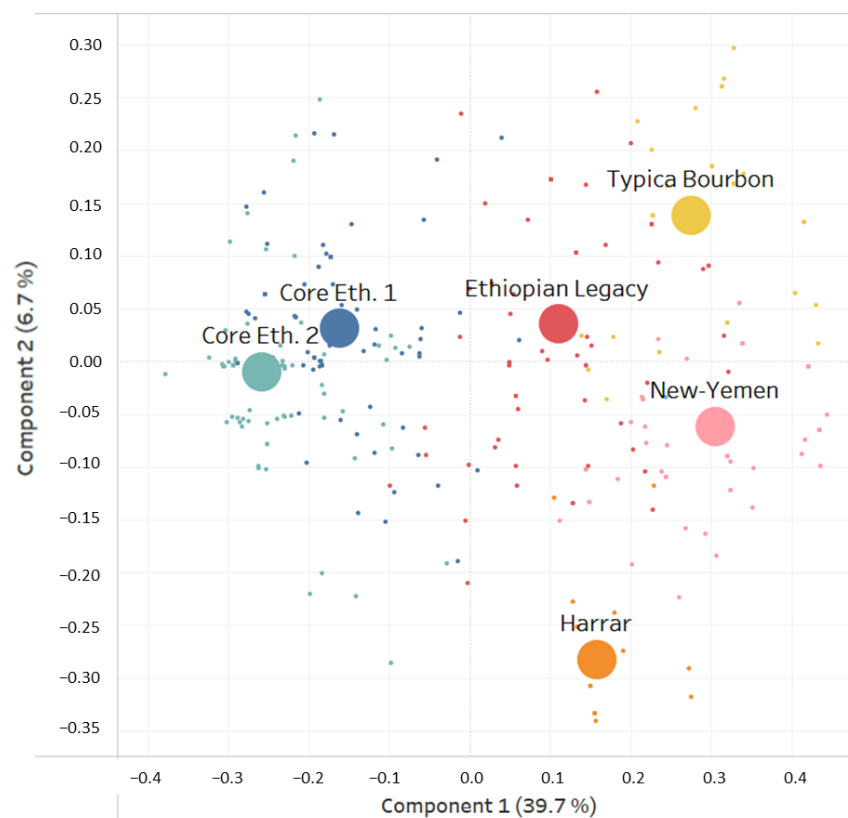


Figure 6. Representation of the first two components of the PCoA performed on Dice index dissimilarity matrix produced from 10 SSR markers on 555 *C. arabica* samples. Small circles are individual samples and large circles are the barycenter of each genetic cluster.

The total number of alleles for the 10 markers was 77 in total (Table 3). Both “Core Eth. 1 and 2” had the highest number of alleles with 65 and 45, respectively. “Ethiopian Legacy” cluster had an intermediate number of 39 alleles, while “Typica/Bourbon”, “New-Yemen” and “Harrar” had less alleles with 30, 30 and 29 alleles, respectively. While not strictly private alleles, allele 135 of marker 29 was born by 96% and 99% of samples of Core Eth. 1 and 2, respectively, while almost absent in Typica/Bourbon, New-Yemen and Harrar. The reverse was true for allele 133 of the same marker, with “Ethiopian Legacy” being intermediate. Other alleles were specific to one of several clusters. For instance, allele 302 of marker 225 for New-Yemen and Harrar. The repartition of the main discriminating alleles amongst the genetic clusters is shown in Table S2.

4. Discussion

To our knowledge, this study is the first that represents more than 550 samples covering most of the coffee growing areas of both Ethiopia and Yemen. The study provides an unprecedented opportunity to precisely identify and map the “Domestication Pathway” of *C. arabica*. Based on the genetic clusters found in our study, we propose the following scenario:

1. *C. arabica* was introduced from Ethiopia to Yemen with seeds from wild populations located in the southern part of Ethiopia, namely in the regions of Gedio (including Yirgacheffe) and Guji.

2. The descendants of these populations formed the “Ethiopian Legacy” cluster found in these regions of Ethiopia, Hararghe and in some regions of Yemen.
3. In Yemen, the “Ethiopian Legacy” cluster is preponderant in Ibb, which was probably a key location for the introduction of coffee into Yemen.
4. From this original population, two genetic clusters derived in Yemen, most likely through seed movements and isolation. This gave rise to the “Typica/Bourbon” and “New-Yemen” genetic clusters. The former is spreading from Ibb to Dhamar and the latter prospering up in the northern coffee growing areas of Sanaa and Mawhit.
5. The varieties that conquered the world from Yemen are derived from two clusters or mother populations: “Ethiopian Legacy” (SL34, SL09, K-7 or SL17 for instance) and “Typica/Bourbon” (Typica, Bourbon, SL-28 or Kent for instance).
6. In Hararghe, a specific genetic cluster—“Harrar”—is identified for the first time. This cluster likely originates from the same “Domestication Pathway” and is derived from the same “Ethiopian Legacy” cluster. In that sense, “Harrar”, “Typica/Bourbon”, and “New-Yemen” share a common genetic origin: The “Ethiopian Legacy” cluster.

This scenario is to be considered a theory that matches both previous knowledge on *C. arabica* genetic diversity, and the novel results from the present study. Further studies will serve to validate and refine it or correct it: The timeline of this scenario is likely the one generally accepted to propose the introduction of coffee seeds in Yemen in the 15th century and the exit of coffee seeds from Yemen to the world in the early 18th century [6,14,43,44]. Some authors indicate the 16th century for the introduction from Ethiopia to Yemen [19,45] but it does not change the bigger picture. This scenario is unlikely to be strictly linear. It is possible that coffee seeds were introduced on several occasions into Yemen, but either from different genetic backgrounds of which only “Ethiopian Legacy” remained through domestication or consistently from the same “Ethiopian Legacy” genetic source.

Our results suggest that the “New-Yemen” cluster became isolated from the “Ethiopian Legacy” cluster in the northern part of the coffee growing area in Yemen. This is consistent with the study on vernacular names from Montagnon et al. [35]. Because these regions were further away from the ports of Mokha and Aden, this would explain why the genetic cluster “New-Yemen” was not part of the coffee seeds exported from Yemen to the world, as shown in a previous study [21]. On the contrary, the “Typica/Bourbon” and “Ethiopian Legacy” clusters are grown in regions closer to the ports of Mokha and Aden (Figure 2) from which coffee seeds were smuggled out of Yemen in the 18th century. Hence, it would be reasonable to assume that these two clusters formed the genetic basis of Arabica coffee varieties cultivated worldwide.

For the first time, we identified the “Harrar” genetic cluster specific to the Hararghe region. A total of 23 out of the 30 samples of the “Harrar” genetic cluster were found in Hararghe. Only six were found in the eastern part of the rift valley (Table 3), more precisely in Gedio, Guji, Sidama and west Arsi (Table S1). Another one was found in the Ibb governorate of Yemen. No cultivated variety around the world is related to this “Harrar” cluster, not even the varieties whose name is somehow related to Hararghe (data not shown). However, while 74% (23/31) of Hararghe samples belonged to the “Harrar” cluster, three were from “Core Ethiopia” clusters, four from the “Ethiopian Legacy” and two from the “New-Yemen”. The number of samples from Hararghe in our study—31—is still limited and further studies are needed to confirm and/or narrow the genetic landscape of that region.

Most authors suggest that coffee growing in Hararghe started with seeds and knowledge from Yemen [5,6], hence after seeds were introduced from Ethiopia to Yemen in the early 15th century. This hypothesis is considered as a fact by Sylvain [46] and Schaefer [45]. However, another hypothesis is that coffee in Hararghe comes directly from seeds from the neighboring Bale forest. Bale forest borders the region where “Ethiopian Legacy” cluster is found in our study. Ahmed [47] indicates that coffee growing in Hararghe might date back to medieval times, hence at the same period as the start of coffee

growing in Yemen. Hararghe has experienced several political ups and down since the 15th century. The only clear information on coffee growing is that it was revived and almost imposed by the Egyptians in the 19th century [47]. It is certainly challenging to trace back various possible introductions of coffee seeds into Hararghe. However, our results indicate that 74% of the coffee sampled in the Hararghe region form a genetic cluster—“Harrar”—specific to this region. The “Harrar” genetic cluster is part of the “Domestication Pathway” of *C. arabica* and is most likely deriving from the same “Ethiopian Legacy” cluster. While some “Harrar” genetic background is found in the “Ethiopian legacy” cluster, the reverse is not true. Furthermore, the “Harrar” cluster has been isolated in Hararghe for enough time to be genetically different from other clusters of the “Domestication Pathway”. Our study shows that: (i) the “Harrar” genetic cluster forms most coffee samples in Hararghe, (ii) this cluster is clearly part of the “Domestication Pathway” and (iii) it is most likely deriving from the same “Ethiopian Legacy” genetic cluster as Yemeni clusters “Typica/Bourbon” and “New-Yemen”. This genetic pattern combined with strong historical hypothesis [5,6,45,46] supports the notion that Yemen was an important source of seeds used for early coffee growing in Hararghe. Furthermore, as expected from the geographical proximity and confirmed by the admixture analysis, there must have been at least some movement of material between southwestern Ethiopia and Hararghe. Hence, coffee seeds have been exchanged between the south of Ethiopia and Hararghe, geographically close, but not enough to prevent the isolation of the “Harrar” cluster. Inversely, no sign of the Core Ethiopian clusters is found in Yemen, confirming that the isolation of Yemen from the southwest and south of Ethiopia was stronger.

Just as for Hararghe, the history of early coffee growing in Yemen is not well referenced: “it is really unfortunate that the original Yemeni sources do not mention anything of value about the areas where coffee was grown” [43]. The only reliable referenced sources are from the first European travelers describing coffee growing in Yemen [14,48], but that was no less than two to three centuries after coffee growing started in Yemen. From these sources, Arwa [43] can list several locations where coffee was ascertained to be grown. These locations cover all the present coffee growing regions in Yemen. Hence, in the early 18th century, coffee had already been introduced in all present main coffee growing regions of Yemen. However, our genetic results suggest that it started in Ibb and then spread from there to different regions leading to the isolation of “Typica/Bourbon” and “New-Yemen” from early “Ethiopian Legacy”. Interestingly, this indicates that while the introduction of coffee seeds from Ethiopia into Yemen represented a first severe bottleneck in the 15th century, the further isolation of some populations in Yemen itself has led to a second bottleneck before coffee material was smuggled out from Yemen in the 18th century. In our study, we had no sample from Taiz and Lahij governorates. However, the geographical pattern in Yemen shown in our study with “New-Yemen” mostly in the northern coffee regions, and “Typica/Bourbon” and “Ethiopian Legacy” mostly in the southern regions, closer to the ports of Mokha and Aden, is coherent with our observation of only the two latter clusters exiting Yemen to form the basis of worldwide varieties [21]. Further sampling in Taiz and Lahij, closest governorates to the ports of Mocha and Aden, respectively (Figure 2), will be important to confirm this hypothesis.

Our study did not include samples from South Sudan, recently proven to host a specific *C. arabica* genetic diversity, characterized namely by the private allele 24-154 [12]. We did not find this allele in our study, hence (i) confirming its South Sudan private allele status and (ii) suggesting that South Sudan material did not participate to the domestication of *C. arabica*.

In previous studies on *C. arabica* genetic diversity, Ethiopian material has often been the accessions from the FAO [28] and Orstom [29] surveys, maintained namely at the Catie germplasm collection of Turrialba in Costa Rica. A total of 100% and 90% of the Orstom and FAO surveys took place in the southwest of Ethiopia, respectively [28,29]. Only 12

accessions of the FAO survey were collected in the Sidamo region and seven in Hararghe. Anthony et al. [22] had included two Hararghe and eight Sidamo samples in their study where they found (i) a clear separation between *Typica* and *Bourbon* varieties on one hand and the Ethiopian material in general on the other hand and (ii) a distinction between one major Ethiopian cluster and three smaller ones. The three smaller ones included one of the two Hararghe and five of the eight Sidamo. Only three out of 69 samples of the west side of the Rift valley were part of these minor clusters. These results were likely early signals of our “Core Ethiopia” vs. “Domestication Pathway” clusters. However, because there were no Yemeni samples in this early study and only few coffee varieties cultivated worldwide, no more dots could be connected, and a wider domestication picture could not be formed. The Ethiopian material of Scalabrin et al. [3] was also the FAO and Orstom surveys, including two Hararghe and seven Sidamo samples. This study had a better representation of cultivated varieties as well as 88 Yemeni samples. A total of two main genetic clusters were identified: G1 was made of only Ethiopian accessions while G2 included all the cultivated varieties worldwide and all the Yemeni accessions together with a small share (9%) of the Ethiopian accessions. G1 and G2 were most likely genetically equivalent to our “Core Ethiopia” and “Domestication Pathway”, respectively. Amongst the few Ethiopian accessions part of G2 were the two Hararghe accessions and six of the seven Sidamo accessions. The remaining Ethiopian accessions of G2 were representing less than 7% of the accessions from the southwest of Ethiopia. Based on these results, the authors acknowledged the genetic similarity of Hararghe and Yemen, naming G2 the “Harar-Yemen” cluster. However, there were not enough samples from Hararghe and Sidamo, nor a sufficient focus on the genetic diversity of the Yemeni samples to go further in details. Moreover, the focus of the study was more on the genome sequencing and the deciphering of the early polyploidization event of *C. arabica* than on the detailed genetic diversity [3]. Montagnon et al. [21] included the samples of the FAO and Orstom Surveys that formed the core collection designed by World Coffee Research and the Centro Agronómico Tropical de Investigación y Enseñanza (CATIE) in 2014 (Solano, personal communication). There was only one Sidamo sample and no Hararghe sample in the study. However, there were 45 Yemeni samples together with a wide representation of varieties cultivated around the world. In this study, three mother populations or genetic clusters were identified in Yemen: “SL-34”, “*Typica/Bourbon*” and “New-Yemen”. Another genetic cluster was made only of Ethiopian accessions (Ethiopian only). A fifth genetic cluster (SL-17) was made of four varieties cultivated worldwide and four Ethiopian accessions. While this study of Montagnon et al. [21] has unveiled the until then unknown genetic diversity in Yemen and its relation to cultivated varieties around the world, it still lacked sufficient representation from the south of Ethiopia and of the Hararghe region.

The results of the present study develop a clearer picture of the *C. arabica* genetic diversity because it includes for the first time Ethiopian samples covering the southwest, the south and the east (Hararghe) regions and Yemeni samples covering most important coffee regions in Yemen. However, while our Ethiopian samples cover a wide geography, they are also different from former studies for two main reasons. First, our samples were not surveyed by experts focusing on spontaneous or sub-spontaneous trees (avoiding as much as possible cultivated samples). Our Ethiopian samples are clearly from cultivated coffee trees sent by coffee farmers participating to the Cup of Excellence of Ethiopia. Ethiopian Coffee research has been active in coffee breeding and has released new varieties [49–51]. Hence, there might be a bias in our samples due to (i) the fact that these are cultivated materials and (ii) they are cultivated materials that are intended to achieve success at the Cup of Excellence competition. Our results do not support the existence of such a bias. Indeed, the genetic diversity is large and continuous (Figures 3 and 6). In our study, the average number of alleles per marker for the “Core Ethiopia” cluster is 6.7 (6.5 and 4.5 for “Core Eth. 1” and “Core Eth. 2”, respectively), hence comparable to the 7.0 alleles per marker for the ‘Ethiopian only’ cluster of Montagnon [21], corresponding to the

core collection representative of the genetic diversity from the FAO and Orstom surveys. It is also comparable to the 7.5 alleles per marker observed in commercial arabica coffee varieties released in Ethiopia [51]. These results indicate that coffee trees cultivated in Ethiopia still represent high genetic diversity originating from local landraces and/or released varieties [51] exchanging genes with local forest coffee trees [24–26,52]. The second reason our samples are different from previous studies based on the FAO and Orstom survey is due to the fact that the latter were surveyed some 60 years ago. During the past 60 years, deforestation has been severe in Ethiopia [52,53] and climate change has impacted coffee genetic resources in Ethiopia [9]. According to FAO [54], the cultivated coffee area in Ethiopia has increased four-fold between 1990 and 2020, reaching 800,000 hectares, suggesting that coffee itself was partly responsible for the deforestation. Hence, the genetic diversity of coffee might have decreased over the last 60 years. However, this decrease is not observed in our study, at least for the cultivated coffee. During these past 60 years, there has been some coffee gene flow between regions [24–26,55], namely for large scale new planting efforts. Hence, the eventual geographical pattern related to the structure of the genetic diversity might have been blurred, as supported by the amount of admixture between Core Eth. 1 and 2 clusters. Interestingly, Dida et al. [26] studied 86 samples from different regions of Ethiopia found two major genetic clusters: one was more represented in the east and the other one in the west of the Rift Valley. All regions hosted both clusters, but east Hararghe hosted only the more “eastern” cluster. Scalabrini et al. [3] found two clear genetic clusters related to a geographical pattern but that was only in the western part of the Rift between two clusters called “Jimma-Bonga” and “Sheka” in relation to their locations. Again, this was based on the FAO and Orstom survey, based on preferentially spontaneous or sub-spontaneous coffee trees, 60 years ago, possibly before seed movements blurred this geographical structure. In our study, we do find a genetic structure in the “Core Ethiopia” cluster with only a slight east–west pattern.

Our study not only presents an opportunity to propose the first comprehensive scenario of the domestication and early movements of *C. arabica*; it also gives interesting new leads in the search for heterotic groups. Heterosis has been exploited in *C. arabica* through the creation of performing F1 hybrids in Ethiopia [56,57], in Central America [58–60] and Colombia [61]. In Ethiopia, F1 crosses were mainly from parents from the southwest and south regions [56,62] with a superiority of the hybrids over the best parent reaching 100% [56]. In Latin America, F1 crosses typically involve one Ethiopian parent from the FAO or Orstom survey and a traditional cultivated variety such as Caturra, Marsellesa or Castillo as the other parent [60]. Caturra is a variety deriving from the “Typica/Bourbon” genetic cluster, as most of the cultivated varieties worldwide do. Our results suggest that “New-Yemen” and “Harrar” could also be worthwhile heterotic groups in combination with Core Ethiopian parents. Intergroup crosses between “New-Yemen” or “Harrar” with “Typica/Bourbon” might be worth evaluating. Exploring new F1 hybrids in *C. arabica* with genetic clusters such as “Harrar” or “New-Yemen”, both thriving in marginal dry and hot coffee areas (Hararghe and Yemen) is a promising opportunity in the context of climate change in coffee growing areas around the world [63] for which resilient and climate smart varieties are needed [64]. This highlights the importance of the exchange of genetic materials in the framework of collaborative research with equitable benefit sharing.

5. Conclusions

To our knowledge, our work provides the first representation of genetic diversity of *C. arabica*, gathering more than 550 present day samples covering a wide geographical range of both Ethiopia and Yemen. The size and scale of the represented samples made it possible to confirm the present-day allelic richness of *C. arabica*, which was found to be in the same range as that found from the surveys in the 60's. Hence, while legitimate concerns have been raised on the impact of deforestation and climate change on the erosion of the genetic diversity of *C. arabica* in Ethiopia, our study did not detect such

genetic diversity loss for cultivated coffee. Our results confirm past findings of two main genetic clusters of *C. arabica* in its native habitat of Ethiopia, but with loose, if any, relation to a geographical pattern. Our results show for the first time that the Hararghe region in Ethiopia hosts a unique genetic cluster of *C. arabica* that we named “Harrar”. This “Harrar” genetic cluster is one of the “Domestication Pathway” clusters together with two other genetic clusters in Yemen: “Typica/Bourbon” and “New-Yemen”. All these clusters are deriving from the “Ethiopian Legacy” genetic cluster present in both Ethiopia and Yemen. The regions of Gedio and Sidamo in the south of Ethiopia are good candidates for the geographical origin of the “Ethiopian Legacy” cluster, leading to the domestication of *C. arabica* in Yemen and Hararghe. Only descendants of the “Ethiopian Legacy” and “Typica/Bourbon” genetic clusters exited Yemen to form the main varieties cultivated worldwide. This is in line with these two clusters being more represented in the central and southern parts of Yemen, close to the ports of Mokha and Aden, while “New-Yemen” is preponderant in the northern coffee region of Yemen. In addition to bringing knowledge on early movements of *C. arabica* during the first steps of domestication, our study also proposes new leads for the exploitation of hybrid vigor through crosses between genetically distant parents.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/agronomy12123203/s1>, Figure S1. Cluster analysis based on the five first components of the PCoA based on the Dice Index dissimilarity matrix of all the coffee samples of the study (up) and then for each main cluster separately (Bottom); Figure S2. Plots of ancestry coefficients of each pair of the 6 clusters after Admixture model. Each point represents a sample. Samples with intermediate values between 0 and 1, hence not on the vertical or horizontal axis, are admixed; Table S1. Repartition of studied coffee samples according to their geographical origin and genetic cluster detailing Ethiopian zones and Yemeni governorates; Table S2: Percentage of samples bearing most discriminating SSR alleles between identified *C. arabica* genetic clusters.

Author Contributions: Conceptualization, C.M., F.S., D.D. and A.D.B.; methodology, C.M., F.S. and T.B.; validation, C.M., F.S., A.D.B., D.D. and T.B.; formal analysis, C.M.; investigation, C.M. and T.B.; resources, F.S., A.D.B. and D.D.; data curation, C.M.; writing—original draft preparation, C.M.; writing—review and editing, C.M., F.S., T.B. and A.D.B.; visualization, C.M. and T.B.; supervision, A.D.B. and F.S.; project administration, F.S., D.D. and A.D.B.; funding acquisition, F.S., D.D. and A.D.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Qima Coffee for the analysis of Yemeni Sample. Funding for the analysis of Ethiopian samples was from the USAID Ethiopia Feed the Future Value Chain Activity.

Data Availability Statement: The datasets generated during the current study are available from the corresponding author on reasonable request.

Acknowledgments: Authors would like to thank Ethiopian and Yemeni coffee farmers for maintaining coffee genetic resources for generations.

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

References

1. Lashermes, P.; Combes, M.C.; Robert, J.; Trouslot, P.; D’Hont, A.; Anthony, F.; Charrier, A. Molecular characterisation and origin of the *Coffea arabica* L. genome. *Mol. Gen. Genet.* **1999**, *261*, 259–266. <https://doi.org/10.1007/s004380050965>.
2. Bawin, Y.; Ruttink, T.; Staelens, A.; Haegeman, A.; Stoffelen, P.; Mwanga Mwanga, J.C.I.; Roldan-Ruiz, I.; Honnay, O.; Janssens, S.B. Phylogenomic analysis clarifies the evolutionary origin of *Coffea arabica*. *J. Syst. Evol.* **2020**, *59*, 953–963. <https://doi.org/10.1111/jse.12694>.
3. Scalabrin, S.; Toniutti, L.; Di Gaspero, G.; Scaglione, D.; Magris, G.; Vidotto, M.; Pinosio, S.; Cattonaro, F.; Magni, F.; Jurman, I.; et al. A single polyploidization event at the origin of the tetraploid genome of *Coffea arabica* is responsible for the extremely low genetic variation in wild and cultivated germplasm. *Sci. Rep.* **2020**, *10*, 4642. <https://doi.org/10.1038/s41598-020-61216-7>.
4. IFC. *Performance Standard 6: Biodiversity Conservation and Sustainable Management of Living Natural Resources*; International Finance Corporation: Washington, DC, USA, 2012. Available online: <https://www.ifc.org/ps6> (accessed on 30 October 2022).
5. Meyer, F.G. Notes on wild *Coffea arabica* from Southwestern Ethiopia, with some historical considerations. *Econ. Bot.* **1965**, *19*, 136–151. <https://doi.org/10.1007/BF02862825>.

6. Koehler, J. *Where the Wild Coffee Grows: The Untold Story of Coffee from the Cloud Forests of Ethiopia to Your Cup*; Bloomsbury Publishing: New York, NY, USA, 2017.
7. Montagnon, C.; Bouharmont, P. Multivariate analysis of phenotypic diversity of *Coffea arabica*. *Genet. Resour. Crop Evol.* **1996**, *43*, 221–227. <https://doi.org/10.1007/BF00123274>.
8. Lashermes, P.; Trouslot, P.; Anthony, F.; Combes, M.C.; Charrier, A. Genetic diversity for RAPD markers between cultivated and wild accessions of *Coffea arabica*. *Euphytica* **1996**, *87*, 59–64. <https://doi.org/10.1007/BF00022965>.
9. Davis, A.P.; Gole, T.W.; Baena, S.; Moat, J. The impact of climate change on indigenous arabica coffee (*Coffea arabica*): Predicting future trends and identifying priorities. *PLoS ONE* **2012**, *7*, e47981. <https://doi.org/10.1371/journal.pone.0047981>.
10. Mairal, M.; Sanmartín, I.; Herrero, A.; Pokorny, L.; Vargas, P.; Aldasoro, J.J.; Alarcón, M. Geographic barriers and Pleistocene climate change shaped patterns of genetic variation in the Eastern Afrotropical biodiversity hotspot. *Sci. Rep.* **2017**, *7*, 45749. <https://doi.org/10.1038/srep45749>.
11. Thomas, A.S. The wild Arabica coffee on the Boma Plateau. *Angle-Egypt. Sudan. Emp. J. Exp. Agric.* **1942**, *10*, 207–212.
12. Krishnan, S.; Pruvot-Woehl, S.; Davis, A.P.; Schilling, T.; Moat, J.; Solano, W.; Al Hakimi, A.; Montagnon, C. Validating South Sudan as a Center of Origin for *Coffea arabica*: Implications for Conservation and Coffee Crop Improvement. *Front. Sustain. Food Syst.* **2021**, *5*, 445. <https://doi.org/10.3389/fsufs.2021.761611>.
13. Anthony, F.; Berthaud, J.; Guillaumet, J.L.; Lourd, M. Collecting wild *Coffea* species in Kenya and Tanzania. *Plant Genet. Resour. Newsl.* **1987**, *69*, 23–29.
14. De La Roque, J. *Voyage de l'Arabie Heureuse, par l'Océan Oriental, et le Détroit de la Mer Rouge: Fait par les François Pour la Première Fois, Dans les Années 1708, 1709 et 1710*; André Cailleau: Paris, France, 1716. Available online: <https://play.google.com/books/reader?id=3fsOAAAAQAAJ&hl=fr&num=10&printsec=frontcover&pg=GBS.PP7> (accessed on 30 October 2022).
15. Ukers, M.A. *All about Coffee*; The Tea and Coffee Trade Journal: New York, NY, USA, 1922.
16. Chevalier, A. *Les Caféiers du Globe. I. Généralités sur les Caféiers*. *Encyclopédie Biologique*; Paul Lechevalier: Paris, France, 1929.
17. Cramer, P.J.S. *A Review of Literature of Coffee Research in Indonesia (from about 1602 to 1945)*; IICA: Turrialba, Costa Rica, 1957.
18. Haarer, A.E. *Modern Coffee Production*. Ebenezer Baylis and Son; The Trinity Press: London, UK, 1958.
19. Tuchscherer, M. Commerce et production du café en mer Rouge au XVI^e siècle. In *Le Commerce du Café Avant L'ère des Plantations Coloniales*; Tuchscherer, M., Ed.; Institut Français d'Archéologie Orientale, Cahier des Annales Ismologiques: Cairo, Egypt, 2001; Volume 20, pp. 69–90.
20. Berthaud, J. L'origine et la distribution des caféiers dans le monde. In *Le Commerce du Café Avant L'ère des Plantations Coloniales*; Tuchscherer, M., Ed.; Institut Français d'Archéologie Orientale, Cahier des annales ismologiques: Cairo, Egypt, 2001; Volume 20, pp. 361–370.
21. Montagnon, C.; Mahyoub, A.; Solano, W.; Sheibani, F. Unveiling a unique genetic diversity of cultivated *Coffea arabica* L. in its main domestication center: Yemen. *Genet. Resour. Crop Evol.* **2021**, *68*, 2411–2422. <https://doi.org/10.1007/s10722-021-01139-y>.
22. Anthony, F.; Bertrand, B.; Quiros, O.; Wilches, A.; Lashermes, P.; Berthaud, J.; Charrier, A. Genetic diversity of wild coffee (*Coffea arabica* L.) using molecular markers. *Euphytica* **2001**, *118*, 53–65. <https://doi.org/10.1023/A:1004013815166>.
23. Silvestrini, M.; Junqueira, M.G.; Favarin, A.C.; Guerreiro-Filho, O.; Maluf, M.P.; Silvarolla, M.B.; Colombo, C.A. Genetic diversity and structure of Ethiopian, Yemen and Brazilian *Coffea arabica* L. accessions using microsatellites markers. *Genet. Resour. Crop Evol.* **2007**, *54*, 1367–1379. <https://doi.org/10.1007/s10722-006-9122-4>.
24. Aga, E.; Bekele, E.; Bryngelsson, T. Inter-simple sequence repeat (ISSR) variation in forest coffee trees (*Coffea arabica* L.) populations from Ethiopia. *Genetica* **2005**, *124*, 213–221. <https://doi.org/10.1007/s10709-005-1484-6>.
25. Tesfaye, K.; Oljira, T.; Govers, K.; Belkele, E.; Borsh, T. Genetic diversity and population structure of wild *Coffea arabica* populations in Ethiopia using molecular markers. In *Coffee Diversity and Knowledge*; Adugna, G., Bellachew, B., Taye, E., Kufa, T., Eds; Ethiopian Institute of Agricultural Research: Addis Ababa, Ethiopia, 2008; pp. 35–44.
26. Dida, G.; Bantte, K.; Disasa, T. Molecular characterization of Arabica Coffee (*Coffea arabica* L.) germplasms and their contribution to biodiversity in Ethiopia. *Plant Biotechnol. Rep.* **2021**, *15*, 791–804. <https://doi.org/10.1007/s11816-021-00721-1>.
27. Anthony, F.; Combes, M.C.; Astorga, C.; Bertrand, B.; Graziosi, G.; Lashermes, P. The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. *Theor. Appl. Genet.* **2002**, *104*, 894–900. <https://doi.org/10.1007/s00122-001-0798-8>.
28. FAO. *FAO Coffee Mission to Ethiopia: 1964–1965*; FAO: Rome, Italy, 1968.
29. Charrier, A. *Etude de la Structure et de la Variabilité Génétique des Caféiers: Résultats des Etudes et des Expérimentations Réalisées au Cameroun, en Côte d'Ivoire et à Madagascar sur l'espèce Coffea arabica L. Collectée en Ethiopie par une Mission Orstom en 1966*; Bulletin IFCC n° 14; FRA: Paris, France, 1978.
30. Bonneuil, C.; Goffaux, R.; Bonnin, I.; Montalent, P.; Hamon, C.; Balfourier, F.; Goldringer, I. A new integrative indicator to assess crop genetic diversity. *Ecol. Indic.* **2012**, *23*, 280–289. <https://doi.org/10.1016/j.ecolind.2012.04.002>.
31. Gnapi, D.E.; Pokou, D.N.D.; Legnate, H.; Dapeng, Z.; Montagnon, C.; Bertrand, B.; N'guetta, A.S.P. Is the genetic integrity of wild *Coffea canephora* from Ivory Coast threatened by hybridization with introduced coffee trees from Central Africa? *Euphytica* **2022**, *218*, 62. <https://doi.org/10.1007/s10681-022-03004-0>.
32. Pernès, J. *Gestion des Ressources Génétiques des Plantes: Tome 2-Manuel*; Lavoisier: Paris, France, 1984.
33. Combes, M.C.; Andrzejewski, S.; Anthony, F.; Bertrand, B.; Rovelli, P.; Graziosi, G.; Lashermes, P. Characterization of microsatellite loci in *Coffea arabica* and related coffee species. *Mol. Ecol.* **2002**, *9*, 1178–1180. <https://doi.org/10.1046/j.1365-294x.2000.00954-5.x>.

34. Pruvot-Woehl, S.; Krishnan, S.; Solano, W.; Schilling, T.; Toniutti, L.; Bertrand, B.; Montagnon, C. Authentication of *Coffea arabica* Varieties through DNA Fingerprinting and its Significance for the Coffee Sector. *J. AOAC Int.* **2020**, *103*, 325–334. <https://doi.org/10.1093/jaoacint/qs003>.
35. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **2000**, *155*, 945–959. <https://doi.org/10.1534/genetics.116.195164>.
36. Pritchard, J.K.; Wen, X.; Falush, D. 2010 Documentation for Structure Software: Version 2.3. Available online: https://web.stanford.edu/group/pritchardlab/structure_software/release_versions/v2.3.4/structure_doc.pdf (accessed on 9 December 2022).
37. Perrier, X.; Jacquemoud-Collet, J.P. DARwin Software. 2006. Available online: <http://darwin.cirad.fr/darwin> (accessed on 30 October 2022).
38. Addinsoft. *XLSTAT Statistical and Data Analysis Solution*; Addinsoft: Paris, France, 2022. Available online: <https://www.xlstat.com/fr> (accessed on 30 October 2022).
39. Ward, J.H. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **1963**, *58*, 238–244.
40. Patterson, N.; Price, A.L.; Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2006**, *2*, e190. <https://doi.org/10.1371/journal.pgen.0020190>.
41. Montagnon, C.; Rossi, V.; Guercio, C.; Sheibani, F. Vernacular Names and Genetics of Cultivated Coffee (*Coffea arabica*) in Yemen. *Agriculture* **2022**, *22*, 1970. <https://doi.org/10.3390/agriculture12081970>.
42. Murray, D.G. *Tableau Your Data!: Fast and Easy Visual Analysis with Tableau Software*; John Wiley & Sons: Hoboken, NJ, USA, 2013.
43. Arwa, A. Yemeni Coffee Trade. Master's Thesis, University of Sanaa, Sana'a, Yemen, 2004.
44. Morris, J. *Coffee: A Global History*; Reaktion Books: London, UK, 2019.
45. Schaefer, C.G.H. Coffee unobserved: Consumption and Commoditization of Coffee in Ethiopia before the Eighteenth century. In *Le Commerce du Café Avant L'ère des Plantations Coloniales*; Tuchscherer, M., Ed.; Institut Français d'Archéologie Orientale, Cahier des annales ismologiques: Cairo, Egypt, 2001; Volume 20, pp. 23–33.
46. Sylvain, P.G. Ethiopian coffee—Its significance to world coffee problems. *Econ. Bot.* **1958**, *12*, 111–139.
47. Ahmed, W.M. *History of Harar and the Hararis. Refined Version*; Harari People Regional State Culture, Heritage and Tourism Bureau: Harar, Ethiopia, 2015. Available online: https://everythingharar.com/files/History_of_Harar_and_Harari-HNL.pdf (accessed on 30 October 2022).
48. Niebuhr, C. *Beschreibung von Arabien. Auseigenen Beobachtungen und im Landeselbstgesammelten Nachrichten*; Nicolaus Müller: Kopenhagen, 1772. Available online: <https://doi.org/10.11588/diglit.5544> (accessed on 30 October 2022).
49. Benti, T. Progress in Arabica Coffee breeding in Ethiopia: Achievements, challenges and prospects. *Int. J. Sci. Basic Appl. Res.* **2017**, *33*, 15–25.
50. Hill, T.; Bekele, G. *A Reference Guide to Ethiopian Coffee Varieties*; Counter Culture Coffee: Durham, NC, USA, 2018.
51. Benti, T.; Gebre, E.; Tesfaye, K.; Berecha, G.; Lashermes, P.; Kyallo, M.; Kouadio Yao, N. Genetic diversity among commercial arabica coffee (*Coffea arabica* L.) varieties in Ethiopia using simple sequence repeat markers. *J. Crop Improv.* **2021**, *35*, 147–168. <https://doi.org/10.1080/15427528.2020.1803169>.
52. Labouisse, J.P.; Bellachew, B.; Kotecha, S.; Bertrand, B. Current status of coffee (*Coffea arabica* L.) genetic resources in Ethiopia: Implications for conservation. *Genet. Resour. Crop Evol.* **2008**, *55*, 1079–1093. <https://doi.org/10.1007/s10722-008-9361-7>.
53. Davis, A.P.; Chadburn, H.; Moat, J.; O'Sullivan, R.; Hargreaves, S.; Lughadha, E.N. High extinction risk for wild coffee species and implications for coffee sector sustainability. *Sci. Adv.* **2019**, *5*, eaav3473. <https://doi.org/10.1126/sciadv.aav3473>.
54. FAO. Food and Agriculture Data. 2022. Available online: <https://www.fao.org/faostat/en/#home> (accessed on 30 October 2022).
55. Berecha, G.; Aerts, R.; Vandepitte, K.; Van Glabeke, S.; Muys, B.; Roldán-Ruiz, I.; Honnay, O. Effects of forest management on mating patterns, pollen flow and intergenerational transfer of genetic diversity in wild Arabica coffee (*Coffea arabica* L.) from Afromontane rainforests. *Biol. J. Linn. Soc.* **2014**, *112*, 76–88. <https://doi.org/10.1111/bij.12274>.
56. Bellachew, B.; Atero, B.; Tefera, F.; Ayano, A.; Benti, T. Genetic diversity and heterosis in arabica coffee. In *Coffee Diversity and Knowledge*; Adugna, G., Bellachew, B., Taye, E., Kufa, T., Eds; Ethiopian Institute of Agricultural Research: Addis Ababa, Ethiopia, 2008; pp. 50–57.
57. Alemayehu, D. Review on genetic diversity of coffee (*Coffea arabica* L.) in Ethiopia. *Int. J. For. Hortic.* **2017**, *3*, 18–27. <https://doi.org/10.20431/2454-9487.0302003>.
58. Bertrand, B.; Alpizar, E.; Lara, L.; Santacrea, R.; Hidalgo, M.; Quijano, J.M.; Montagnon, C.; Georget, F.; Etienne, H. Performance of *Coffea arabica* F1 hybrids in agroforestry and full-sun cropping systems in comparison with American pure line cultivars. *Euphytica* **2011**, *181*, 147–158. <https://doi.org/10.1007/s10681-011-0372-7>.
59. Georget, F.; Marie, L.; Alpizar, E.; Courtel, P.; Bordeaux, M.; Hidalgo, J.M.; Marraccini, P.; Breitler, J.C.; Déchamp, E.; Ponçon, C.; et al. Starmaya: The first Arabica F1 coffee hybrid produced using genetic male sterility. *Front. Plant Sci.* **2019**, *10*, 1344. <https://doi.org/10.3389/fpls.2019.01344>.
60. Marie, L.; Abdallah, C.; Campa, C.; Courtel, P.; Bordeaux, M.; Navarini, L.; Lonzarich, V.; Bosselmann, A.S.; Turreira-Garcia, N.; Alpizar, E.; et al. G × E interactions on yield and quality in *Coffea arabica*: New F1 hybrids outperform American cultivars. *Euphytica* **2020**, *216*, 78. <https://doi.org/10.1007/s10681-020-02608-8>.
61. Bertrand, B.; Villegas Hincapié, A.M.; Marie, L.; Breitler, J.C. Breeding for the main agricultural farming of arabica coffee. *Front. Sustain. Food Syst.* **2021**, *5*, 709901. <https://doi.org/10.3389/fsufs.2021.709901>.

-
62. Gebreselassie, H.; Atinafu, G.; Degefa, M.; Ayano, A. Arabica Coffee (*Coffea arabica* L.) Hybrid Genotypes Evaluation for Growth Characteristics and Yield Performance under Southern Ethiopian Growing Condition. *Acad. Res. J. Agric. Sci. Res.* **2018**, *6*, 89–96. <https://doi.org/10.14662/ARJASRD2017.071>.
 63. Bunn, C.; Läderach, P.; Jimenez, J.G.P.; Montagnon, C.; Schilling, T. Multiclass classification of agro-ecological zones for Arabica coffee: An improved understanding of the impacts of climate change. *PLoS ONE* **2015**, *10*, e0140490. <https://doi.org/10.1371/journal.pone.0140490>.
 64. Breitler, J.C.; Etienne, H.; Lérant, S.; Marie, L.; Bertrand, B. Description of an Arabica coffee ideotype for agroforestry cropping systems: A guideline for breeding more resilient new varieties. *Plants* **2022**, *11*, 2133. <https://doi.org/10.3390/plants11162133>.