

Article

Comparative Evaluation of Variant Calling Strategies for High-Density SNP Discovery in Polyploid Kiwifruit (*Actinidia* spp.)

Yumi Kim ^{1,2}, Mockhee Lee ³  and Daeil Kim ^{1,*} 

¹ Department of Horticulture, Chungbuk National University, Cheongju 28644, Republic of Korea; yumi6998@korea.kr

² Fruit Research Division, National Institute of Horticultural and Herbal Science, Rural Development Administration, Wanju 55365, Republic of Korea

³ Research Institute of Climate Change and Agriculture, National Institute of Horticultural and Herbal Science, Rural Development Administration, Jeju 63240, Republic of Korea; mockey92@korea.kr

* Correspondence: dkpomo@cbnu.ac.kr; Tel.: +82-43-261-2527

Highlights

What are the main findings?

- The methylation-sensitive *ApeKI*/*TfiI* restriction enzyme combination generated the highest proportion of DNA fragments within the target size range (200–500 bp) in the polyploid kiwifruit genome, outperforming methylation-insensitive combinations.
- GATK identified approximately 22-fold more SNPs (828,257 SNPs) than freebayes (38,703 SNPs) and bcftools (37,199 SNPs), with comparable transition/transversion (Ts/Tv) ratios across all three tools (1.4–1.6).
- Only the GATK-derived SNP dataset exceeded the minimum marker density threshold (~600,000 SNPs) required for high-density genomic coverage of the kiwifruit genome.

What are the implications of the main findings?

- The combination of methylation-sensitive restriction enzymes (*ApeKI*/*TfiI*) and GATK-based variant calling provided a practical approach for generating a high-density SNP dataset in mixed-ploidy kiwifruit germplasm. The observed differences represent differences in SNP discovery rather than comparative variant-calling accuracy.

Abstract

Single nucleotide polymorphisms (SNPs) are widely used for genetic diversity analysis, linkage mapping, genome-wide association studies (GWAS), and molecular marker development in crop plants. Genotyping-by-sequencing (GBS) enables cost-effective SNP discovery; however, achieving sufficient marker density in polyploid crops remains challenging because of complex genome structures, high sequence similarity among homologous chromosomes, and repetitive genomic regions. In this study, we optimized a GBS-based bioinformatics pipeline for polyploid kiwifruit (*Actinidia* spp.) by evaluating restriction enzyme combinations through in silico digestion analysis and comparing the SNP detection efficiency of three variant-calling tools, namely freebayes, bcftools, and Genome Analysis Tool Kit (GATK). The methylation-sensitive *ApeKI*/*TfiI* combination generated the highest proportion of DNA fragments within the target size range (200–500 bp) in the kiwifruit reference genome cv. Hongyang (*A. chinensis*). Using GATK, 828,257 SNPs were identified, approximately 22-fold higher than those detected using freebayes and bcftools, with a comparable transition/transversion (Ts/Tv) ratio. GATK also identified substantially higher absolute numbers of SNPs in genic regions, while the proportion of genic-region SNPs was



Academic Editor: Tommaso Giordani

Received: 21 June 2026

Revised: 20 July 2026

Accepted: 23 July 2026

Published: 25 July 2026

Copyright: © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

similar across all three tools. Notably, only the GATK-derived SNP dataset exceeded the estimated marker density discussed in this study for high-density genomic coverage of the kiwifruit genome. These results demonstrate that the combination of methylation-sensitive restriction enzymes and GATK-based variant calling generated a high-density SNP dataset for mixed-ploidy kiwifruit germplasm. Because independent validation of SNP accuracy was beyond the scope of this study, the observed differences should be interpreted as differences in SNP discovery rather than comparative variant-calling accuracy.

Keywords: autopolyploid; genotyping-by-sequencing; variant calling

1. Introduction

Genetic variants, including single nucleotide polymorphisms (SNPs), insertions/deletions (InDels), and structural variants, are widely distributed throughout plant genomes and are associated with important horticultural traits such as flesh color, fruit shape, flavor, and soluble solid content [1]. Among these variants, SNPs are particularly valuable for genetic diversity analysis, linkage map construction, quantitative trait locus (QTL) mapping, genome-wide association studies (GWAS), and molecular marker development [2,3]. The development of next-generation sequencing (NGS) technologies has enabled large-scale SNP discovery at relatively low cost; however, both the number and reliability of detected variants strongly depend on optimized bioinformatics pipelines [4,5].

Genotyping-by-sequencing (GBS) is a reduced-representation sequencing approach that detects SNPs adjacent to restriction enzyme sites in a cost-effective manner [6]. Because GBS sequences only restriction enzyme-digested genomic regions, genome coverage and marker density are strongly influenced by the choice of restriction enzymes [7]. To improve genome representation and sequencing efficiency, Poland et al. (2012) proposed a two-enzyme GBS approach, which has subsequently been applied in diverse plants [8]. However, the efficiency of restriction enzyme combinations could differ depending on genome complexity, repeat composition, methylation patterns, and ploidy level. Despite the availability of chromosome-level reference genomes and pangenome resources for *Actinidia*, GBS remains a cost-effective and widely adopted approach for large-scale germplasm characterization, genetic mapping, and molecular marker development because it enables simultaneous genotyping of numerous accessions with relatively low sequencing costs.

GBS data analysis generally involves quality control, read alignment, variant calling, annotation, and downstream filtering processes [9,10]. Among these steps, variant calling is a critical process for identifying genetic variants from sequence reads aligned to a reference genome [11]. Various variant-calling tools have been developed using different statistical models and variant detection strategies [12]. Freebayes utilizes a Bayesian haplotype-based model and has been widely used for polyploid species [13]. Bcftools performs variant detection based on likelihood models derived from read alignments [14,15]. Genome Analysis Tool Kit (GATK) employs local haplotype assembly and realignment approaches for variant detection [16]. Although these tools have been widely applied in short-read sequencing analyses, most variant-calling algorithms were originally developed for human diploid genomes [17], and their relative performance in terms of SNP detection efficiency could differ substantially depending on genome complexity and ploidy level.

Achieving sufficient marker density is particularly critical for high-density genomic applications such as linkage mapping, QTL analysis, and GWAS. A marker density approaching one SNP per kilobase has been suggested as an approximate guideline for some high-density genomic applications; however, the optimal density depends on linkage

disequilibrium, recombination rate, population structure, and the intended downstream analysis [18–20]. Given the approximately 650-Mb kiwifruit genome, this theoretical guideline would correspond to approximately 600,000 genome-wide SNPs [21]. However, accurate variant detection in polyploid plants remains challenging because of high sequence similarity among homologous chromosomes, complex allele dosage patterns, and ambiguous read mapping [19,22]. Kiwifruit (*Actinidia* spp.) is a highly heterozygous plant with diverse ploidy levels, making reliable and high-density SNP detection particularly difficult. In addition, the kiwifruit genome contains abundant repetitive elements and long terminal repeat retrotransposons (LTR-RTs), which further increase genome complexity and complicate read alignment and variant calling. Therefore, the selection of an appropriate variant-calling strategy that maximizes informative SNP recovery is critical for obtaining sufficient marker density in polyploid kiwifruit genomes.

Previous studies have evaluated variant-calling performance in several polyploid crops. Song et al. [23] compared freebayes, bcftools, and GATK in autopolyploid *Saccharum* species and reported that GATK showed superior SNP detection performance. Yao et al. [17] evaluated seven variant-calling tools and demonstrated that GATK and bcftools exhibited relatively high true-positive detection rates in large plant genomes. In kiwifruit, several studies have applied GBS-based SNP discovery using different variant-calling approaches, including GATK, freebayes, and GBS-SNP-CROP [24–26]. However, no previous study has comprehensively evaluated both restriction enzyme combinations and variant-calling strategies in terms of SNP detection efficiency and marker density for polyploid kiwifruit.

Therefore, this study aimed to evaluate a practical GBS-based bioinformatics pipeline for SNP discovery in mixed-ploidy *Actinidia* germplasm by evaluating restriction enzyme combinations through in silico digestion analysis and comparing the SNP detection performance of three widely used variant-calling tools (freebayes, bcftools, and GATK). Because the analyzed germplasm panel included multiple *Actinidia* species with diverse and, in some cases, unknown ploidy levels, the comparison focused on evaluating variant-calling performance under a practical analytical framework rather than allele dosage estimation using a predefined ploidy parameter. Given the minimum marker density requirements for high-density genomic applications in the large and complex kiwifruit genome, particular emphasis was placed on identifying a variant-calling strategy capable of generating sufficient informative SNPs for downstream breeding applications. Our findings provide practical guidance for SNP discovery in mixed-ploidy kiwifruit germplasm and offer useful insights for developing high-density molecular markers for genomics-assisted breeding.

2. Materials and Methods

2.1. Plant Materials and DNA Extraction

Young leaves of 55 kiwifruit accessions were collected from the Namhae branch, National Institute of Horticultural and Herbal Science (34°48′59.8″ N, 127°55′36.6″ E) (Table 1). DNA was extracted using DNeasy® Plant Mini Kit (Qiagen, Hilden, Germany). The extracted DNA quality and quantity were confirmed at 1.5% agarose gel electrophoresis using a DS-11 Spectrophotometer (DeNovix, Wilmington, DE, USA).

Table 1. List of 55 kiwifruit accessions used in this study.

Scientific Name	Accessions (Polyploidy)				No. of Accessions
<i>Actinidia deliciosa</i>	Chieftain (6×)	Elmwood	Garmrok (6×)	Matua (6×)	6
	Qinmei (6×)	Tomuri (6×)			
<i>A. chinensis</i>	Bliss Red	Bliss Yellow	DACT0101	Gold 9	15
	Golden Bold	Golden King	Hort16A (2×)	Jecy Gold (4×)	
	LCK Gold	New Gold	Redvita (4×)	SKK11	
	SKK13	Sunple (4×)	Sweet Gold (4×)		

Table 1. Cont.

Scientific Name	Accessions (Polyploidy)				No. of Accessions
<i>A. arguta</i>	Autumn Sense K5_2_3 K5_10_1	Chiak K5_2_7 K5_14_4	Hardy Red K5_2_13 Saehan	Ilse K5_2_18	11
<i>A. eriantha</i>	Bidan	Eriantha (2×)			2
<i>A. polygama</i>	S8				1
<i>A. macrosperma</i>	Bawoonty71	S7			2
<i>A. arguta</i> var. <i>purpurea</i>	S3				1
<i>A. deliciosa</i> × <i>A. arguta</i>	Choromi	Pohwa (6×)	Po-ok (6×)	SKK200 (6×)	4
<i>A. arguta</i> × <i>A. deliciosa</i>	Bangwoori (6×)	SKK202			2
(<i>A. arguta</i> × <i>A. deliciosa</i>) × <i>A. arguta</i>	Skinny Green (4×)				1
<i>A. chinensis</i> × <i>A. deliciosa</i>	Jecy Green	Mega Gold			2
<i>A. chinensis</i> × <i>A. arguta</i>	Greenmall (4×)				1
Unknown	CG-1 Pantam	CG-3-1	Haenam Gold	Jahyang	5
Total					55

2.2. In Silico Digestion Analysis

In silico digestion analysis was performed using the chromosome-level reference genome of *Actinidia chinensis* cv. Hongyang [21]. The 29 linkage groups were analyzed separately using the SimRAD R package, and the predicted fragment counts were subsequently combined across all linkage groups. Restriction enzymes were classified according to methylation sensitivity and evaluated as single-enzyme and two-enzyme combinations.

The methylation-sensitive enzymes tested were *ApeKI*, *Acil*, *TfiI*, *HhaI*, *HinfI*, and *Sau96I*, whereas the methylation-insensitive enzymes included *AvrII*, *BsrGI*, *HindIII*, *EcoRI*, *MfeI*, *SphI*, and *MseI*. The two-enzyme combinations evaluated included methylation-sensitive/methylation-sensitive combinations (*ApeKI/Acil*, *ApeKI/HhaI*, *ApeKI/TfiI*, *ApeKI/HinfI*, and *ApeKI/Sau96I*), a methylation-sensitive/methylation-insensitive combination (*ApeKI/MseI*), and a methylation-insensitive/methylation-insensitive combination (*MfeI/MseI*).

Restriction fragments were predicted using the `insilico.digest()` function, and fragment-size distributions were summarized using the `size.select()` function in 100-bp intervals. The proportion of fragments within the target range of 200–500 bp was calculated for each enzyme combination. The *ApeKI/TfiI* combination was selected for GBS library construction because it generated the highest proportion of predicted fragments within the target size range among the combinations tested.

2.3. GBS Library Preparation and Sequencing

Genomic DNA from 55 kiwifruit accessions was submitted to a commercial sequencing service provider (Seeders Co., Ltd., Daejeon, Republic of Korea) for GBS library construction. Libraries were prepared using a double-digestion protocol with *ApeKI* and *TfiI* restriction enzymes. Following restriction digestion, barcode adapters and common adapters were ligated to the digested DNA fragments, after which the samples were pooled and amplified by multiplex PCR according to the provider's standard GBS library preparation workflow. Library quality was evaluated prior to sequencing, and quality-control analysis confirmed that the library fragments were predominantly distributed between approximately 250 and 350 bp. Finally, the libraries were sequenced on the Illumina HiSeq X platform (Illumina, Inc., San Diego, CA, USA) using 151-bp paired-end sequencing.

2.4. GBS Data Processing and SNP Calling

The GBS data were processed from raw sequencing reads (Figure 1). The raw reads were demultiplexed using Sabre (<https://github.com/najoshi/sabre>, accessed on 22 July 2026). Adapter sequences and low-quality bases were removed using Cutadapt [27]. The reads were aligned to the reference genome of cv. Hongyang (*A. chinensis*) [21] using BWA-MEM [28]. The resulting SAM files were converted to BAM format, sorted, and indexed using SAMtools prior to variant calling.

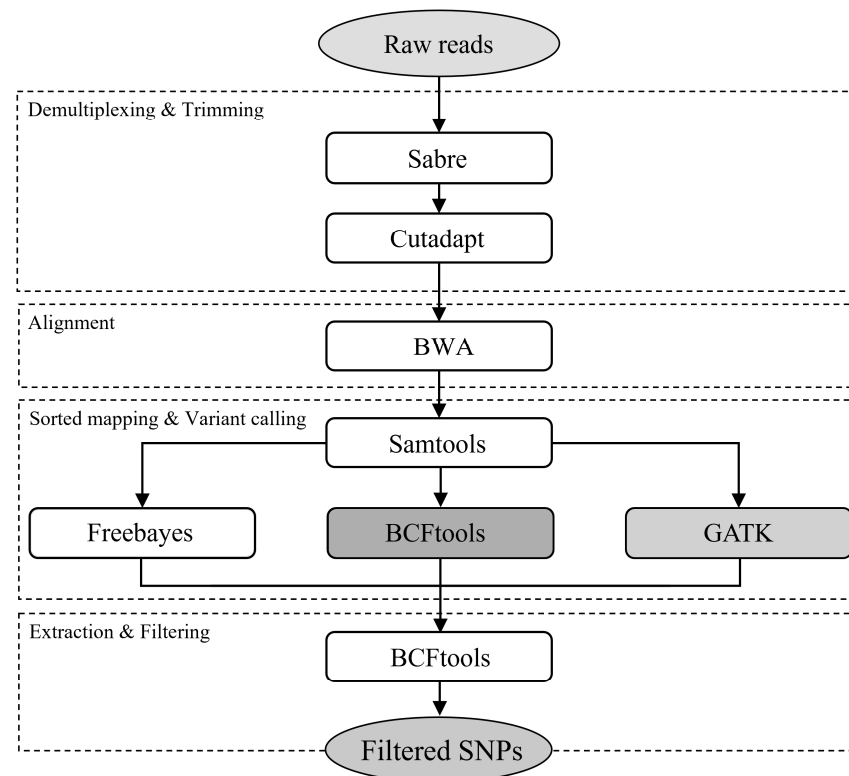


Figure 1. The flowchart of GBS data processing. Variant calling was conducted in the direction of the arrow.

To compare variant-calling performance, SNP discovery was independently performed using three widely used variant callers: FreeBayes, bcftools, and GATK. For GATK, the HaplotypeCaller, CombineGVCFs, and GenotypeGVCFs workflow was used to generate multi-sample variant calls. To ensure a fair comparison among the three variant callers, identical post-calling filtering criteria were applied to all datasets using bcftools. SNPs were retained according to the following criteria: (1) variant quality (QUAL) > 10, (2) read depth (DP) > 5, (3) minor allele frequency (MAF) > 5%, and (4) missing value < 30%.

To evaluate SNP quality, the transition/transversion (Ts/Tv) ratio was calculated. Although the Ts/Tv ratio is a commonly used indicator of SNP quality, it is important to note that GBS-based reduced-representation sequencing approaches may inherently yield lower Ts/Tv ratios than whole-genome sequencing because of partial genome coverage and potential biases introduced by restriction enzyme digestion patterns.

Because the analyzed germplasm panel consisted of multiple *Actinidia* species with different ploidy levels, including diploid, tetraploid, hexaploid, and several accessions with unknown ploidy status, a uniform ploidy parameter was not specified during variant calling, as a single ploidy value could not appropriately represent all accessions included in this study. Instead, all three variant callers were applied under an identical analytical

framework to enable a consistent comparison of SNP discovery performance across the mixed-ploidy *Actinidia* germplasm panel.

2.5. Functional Characteristics of SNPs

For analysis of SNP locations in the genome, the general feature format (GFF3) file of the kiwifruit reference genome cv. Hongyang was downloaded from NCBI. Based on the positional information of the reference genome, SNP positions were annotated using R packages and the GFF3 annotation file of the kiwifruit reference genome. Intergenic and genic regions, including untranslated regions (UTRs), introns, and coding DNA sequences, were classified.

SNP annotation was performed in R studio using the Bioconductor packages GenomicRanges and rtracklayer. The GFF3 annotation file was imported using rtracklayer, and overlaps between SNP coordinates and annotated genomic features were identified using GenomicRanges. The corresponding R script has been provided as Supplementary File S1.

3. Results

3.1. Selection of Restriction Enzyme Combinations

Expected DNA fragment distributions were predicted through in silico digestion analysis (Figure 2a). Both methylation-sensitive and methylation-insensitive restriction enzyme combinations generated DNA fragments within the preferred size range (200–500 bp) in the kiwifruit reference genome cv. Hongyang. However, methylation-sensitive restriction enzyme combinations produced higher proportions of suitable DNA fragments than methylation-insensitive combinations. Among the tested two-enzyme combinations, *ApeKI/TfiI* generated the highest proportion of DNA fragments within the preferred size range (200–500 bp) (Figure 2b). Therefore, GBS libraries were constructed using the *ApeKI/TfiI* combination.

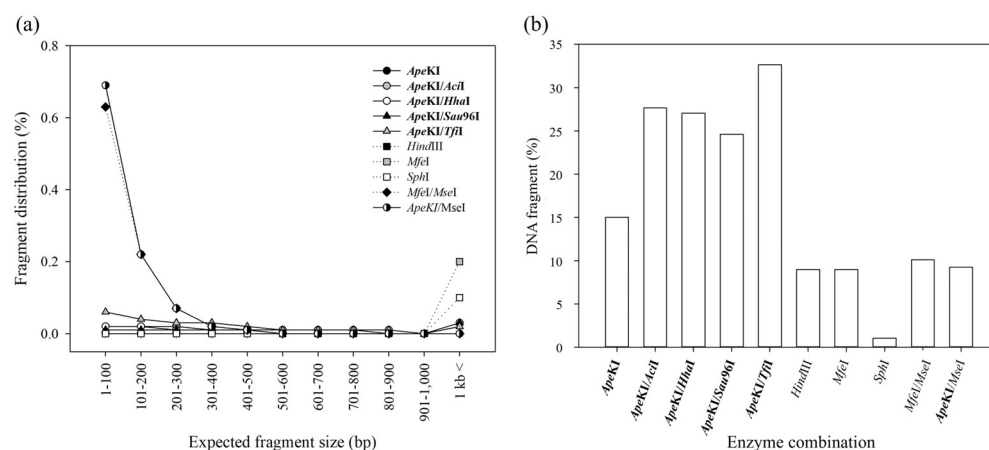


Figure 2. Percentage of expected fragment size (bp) (a) and DNA fragments (200–500 bp) (b) in kiwifruit reference genome cv. Hongyang. Methylation-sensitive restriction enzyme combinations were indicated in bold. Methylation-insensitive restriction enzymes were represented by dotted lines.

3.2. Genotyping-by-Sequencing

A total of 540,308,634 reads were generated in 55 kiwifruit accessions. An average of 9,823,793 reads were detected per individual after GBS. Among generated reads, 528,538,871 reads were mapped to the kiwifruit reference genome with a mapping rate of 97.84%. The average sequencing depth of mapped regions was 16.28 $\times$. The reference genome was covered by 45% (Supplementary Table S1).

3.3. SNP Calling

A total of 38,703 and 37,199 SNPs were identified using freebayes and bcftools, respectively. Using GATK, 828,257 SNPs were identified, which was approximately 22-fold higher than the numbers detected using freebayes (38,703 SNPs) and bcftools (37,199 SNPs). A total of 13,063 SNPs were commonly identified among the three variant callers. The numbers of unique SNPs identified for each variant caller were 8147 (21.05%), 7418 (19.94%), and 790,783 (95.47%) using freebayes, bcftools, and GATK, respectively (Figure 3a). In freebayes, the number of SNPs varied from 733 (LG 4) to 1866 (LG 3). The number of SNPs ranged from 691 (LG 4) to 1817 (LG 19) in bcftools. Using GATK, the number of SNPs ranged from 18,292 (LG 4) to 40,502 (LG 3) (Figure 3b). Notably, the lowest number of SNPs was consistently detected on LG 4 across all three variant-calling tools, suggesting that this pattern reflects a genomic characteristic of LG 4 rather than a tool-specific bias.

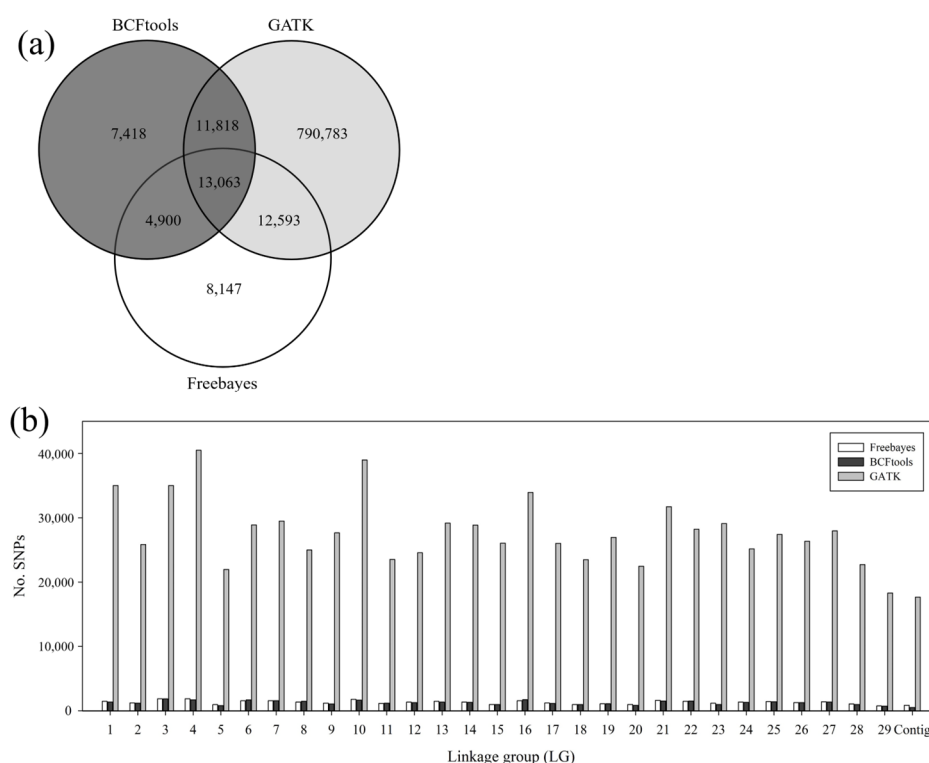


Figure 3. Number of single nucleotide polymorphisms (SNPs) detected by three variant callers, namely freebayes, bcftools, and Genome Analysis Tool Kit (GATK). (a) Venn diagram showing the detected SNPs by freebayes (white), bcftools (dark gray), and GATK (gray); (b) SNP distribution across linkage groups by freebayes, bcftools, and GATK represented as white, dark gray, and gray colors, respectively.

The Ts/Tv ratio was analyzed to evaluate SNP quality. Ts/Tv ratios were 1.4, 1.5, and 1.6 in freebayes, bcftools, and GATK, respectively (Supplementary Table S2).

3.4. Classification of SNPs Based on Positions in the Kiwifruit Genome

Among 38,703 SNPs detected by freebayes, 11,797 (30.49%) and 29,906 (69.51%) SNPs were located in intergenic and genic regions (untranslated region, intron, and coding DNA sequences), respectively. In bcftools, 10,864 (29.21%) SNPs were distributed in the intergenic region, and 26,335 (70.79%) SNPs were detected in the genic region. Using GATK, 248,223 (29.97%) SNPs were identified in the intergenic region, and 580,034 (70.03%) SNPs were detected in the genic region. The proportions of genic-region SNPs were 69.51%, 70.79%, and 70.03% in freebayes, bcftools, and GATK, respectively (Figure 4), indicating

that the distribution of SNPs between genic and intergenic regions was largely consistent across all three tools. In absolute terms, the number of genic-region SNPs identified using GATK (580,034 SNPs) was approximately 21-fold higher than those detected by freebayes (29,906 SNPs) and bcftools (26,335 SNPs).

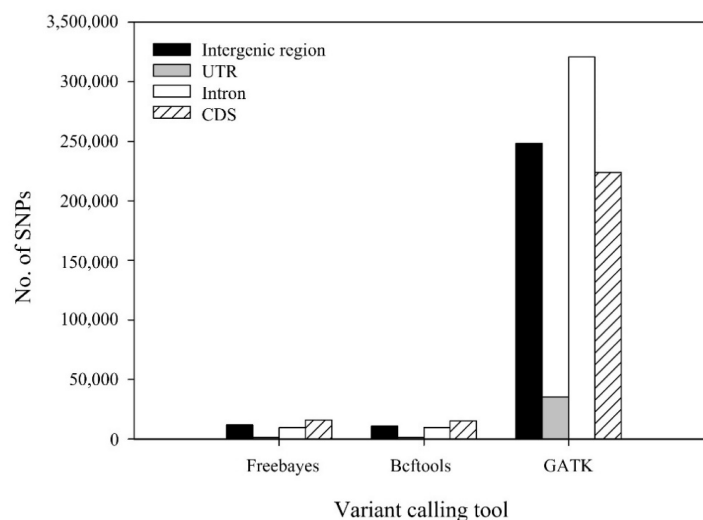


Figure 4. The number of SNPs based on the position in the kiwifruit genome from three variant-calling tools. Intergenic region, DNA sequences located between genes; UTR, untranslated region; CDS, coding DNA sequence. The intergenic region, UTR, and intron were indicated as black, gray, and white, respectively. The CDS regions were represented by diagonal stripes.

4. Discussion

The development of next-generation sequencing (NGS) technologies has enabled large-scale and cost-effective SNP discovery in crop plants. However, the efficiency and reliability of SNP detection are strongly influenced by bioinformatics pipelines and genome complexity [17,23,29]. Compared with diploid genomes, polyploid plant genomes present additional challenges for variant detection because of high sequence similarity among homologous chromosomes, ambiguous read mapping, and complex allele dosage patterns [19]. Furthermore, achieving sufficient marker density for high-density genomic applications requires not only accurate but also efficient SNP discovery strategies in complex polyploid genomes. Therefore, optimization of both restriction enzyme selection and variant-calling strategy is particularly important for complex polyploid crops such as kiwifruit.

The *Actinidia* genome is approximately 43.42% composed of repetitive elements. Among these, long terminal repeat retrotransposons (LTR-RTs), particularly Gypsy and Copia elements, dominate, accounting for 23.38% of the genome. The recent expansion of intact LTR-RTs, with most insertions occurring within the last one million years, underscores the dynamic evolution of the kiwifruit genome [21]. This high repeat content not only increases genome complexity but also creates challenges for restriction enzyme-based sequencing approaches, as repetitive regions may disproportionately consume sequencing capacity without contributing informative variants.

As a result of *in silico* digestion, methylation-sensitive RE combinations produced the highest proportions of suitable DNA fragments (Figure 2). Among methylation-sensitive RE combinations, *ApeKI/TfiI* generated the highest proportion of DNA fragments within the target size range. This restriction enzyme combination also generated the highest proportion of suitable DNA fragments in Asian pear [30]. The kiwifruit genome contains abundant repetitive elements, including LTRs such as Copia and Gypsy retrotransposons, which are heavily methylated as part of epigenetic regulation [21,31,32]. Because methylation-

sensitive restriction enzymes preferentially avoid highly methylated repetitive regions, *ApeKI/TfiI* could improve genome representation in low-copy and gene-proximal genomic regions of kiwifruit. Therefore, these results suggest that *ApeKI/TfiI* is a suitable restriction enzyme combination for efficient genotyping of complex polyploid kiwifruit genomes. However, because only the *ApeKI/TfiI* combination was experimentally evaluated in this study, additional comparisons using other candidate restriction enzyme combinations will be necessary to further validate its relative performance under experimental conditions.

To evaluate variant-calling efficiency in kiwifruit, SNPs were extracted using three variant-calling tools. Among the three tools, GATK identified approximately 22-fold more candidate SNPs (828,257 SNPs) than freebayes (38,703 SNPs) and bcftools (37,199 SNPs) (Figure 3). The SNP counts and Ts/Tv ratios were similar between freebayes and bcftools, with Ts/Tv ratios of 1.4 and 1.5, respectively (Supplementary Table S2). Freebayes has been reported to exhibit reduced sensitivity in detecting SNPs within highly divergent regions of the genome [13,33]. Similarly, bcftools has shown low recall in SNP detection, indicating a reduced ability to identify true positives accurately in highly diverse populations [17]. GATK HaplotypeCaller performs local haplotype assembly and realignment during variant detection, enabling de novo assembly of haplotype variants and the identification of SNPs and InDels [16]. This local haplotype assembly strategy may contribute to the larger number of candidate SNPs identified by GATK, particularly in complex genomic regions containing repetitive sequence elements and LTRs.

The Ts/Tv ratios observed across all three tools were comparable, ranging from 1.4 to 1.6, and were within the range reported in previous plant variant discovery studies [34]. However, the Ts/Tv ratio alone cannot be used as the sole indicator of variant accuracy. Therefore, although GATK identified substantially more candidate SNPs, the authenticity of these additional candidate SNPs should be confirmed through independent experimental validation in future studies.

The detected SNPs across all three variant-calling tools were predominantly located in genic regions, with genic SNPs approximately two-fold more abundant than intergenic SNPs (Figure 4). As discussed above, this distribution pattern was consistent across all three tools, with genic SNP proportions of 69.51%, 70.79%, and 70.03% for freebayes, bcftools, and GATK, respectively, indicating that this pattern primarily reflects the characteristics of the *ApeKI/TfiI* restriction enzyme combination rather than differences among variant-calling algorithms. SNPs located in genic regions are often more directly associated with phenotypic traits and are widely applied in marker-assisted selection, facilitating the identification of plants with desirable traits at early developmental stages [35,36]. While the proportions of genic SNPs were similar across the three variant callers, the absolute number of genic-region SNPs identified by GATK (580,034 SNPs) was approximately 21-fold higher than those identified by freebayes (29,906 SNPs) and bcftools (26,335 SNPs). This increased number of genic-region candidate SNPs may provide a larger pool of markers for downstream applications, including linkage mapping, QTL analysis, and molecular breeding in kiwifruit. However, the biological validity of these additional candidate SNPs should be confirmed through independent experimental validation before their application in breeding programs.

High marker density is particularly important for genomic applications such as linkage mapping, QTL analysis, genome-wide association studies (GWAS), marker development, and genomics-assisted breeding, especially in complex polyploid genomes [18–20,37]. Given that the kiwifruit genome is approximately 650 Mb [21], a marker density approaching one SNP per kilobase would theoretically correspond to approximately 600,000 genome-wide SNPs, potentially providing sufficient marker density for high-resolution genomic analyses in this species.

The SNP datasets generated by freebayes (38,703 SNPs) and bcftools (37,199 SNPs) were substantially lower than this estimated marker density under the current GBS conditions. In contrast, GATK identified 828,257 candidate SNPs, exceeding the estimated marker density proposed for genome-wide coverage in kiwifruit. These results suggest that, under the analytical framework adopted in this study, GATK recovered a substantially larger number of candidate SNPs that may provide sufficient marker density for high-density genomic applications, including linkage mapping, QTL analysis, and association studies in polyploid kiwifruit. However, because these additional candidate SNPs were not independently validated, their authenticity and practical utility should be confirmed through experimental validation before application in downstream breeding studies.

The consistently low SNP count observed on LG4 across all three variant-calling tools may be partially attributable to the relatively small physical size of this chromosome. According to the *A. chinensis* v3.0 genome assembly [21], LG4 (13.86 Mb) is substantially smaller than the average chromosome size (~20.18 Mb) and is the smallest among the 29 linkage groups, which could result in fewer restriction enzyme cleavage sites and, consequently, fewer detectable SNPs. However, chromosome size alone is unlikely to fully explain the reduced SNP density observed on LG4. Other genomic features, including repetitive sequence composition, GC content, local recombination rate, and sequencing coverage, may also have contributed to this pattern and warrant further investigation.

Because the analyzed germplasm included diverse and partially unknown ploidy levels, ploidy-specific variant-calling parameters were not applied in this study. This approach enabled a consistent comparison among variant-calling tools across diverse *Actinidia* germplasm; however, it also represents an important limitation of the present study. In particular, the default diploid assumption implemented in GATK HaplotypeCaller may influence genotype calling and allele dosage estimation in tetraploid and hexaploid accessions and may have contributed to differences in SNP detection patterns relative to freebayes and bcftools. Future studies incorporating ploidy-aware variant-calling strategies, such as those implemented in specialized polyploid genotyping tools, are expected to improve genotype accuracy and allele dosage estimation in polyploid kiwifruit. In addition, because the germplasm panel included multiple *Actinidia* species with unbalanced representation among species and ploidy groups, potential species-specific differences in mapping performance, sequencing depth, and missing data patterns were not evaluated in the present study. Future studies using more balanced germplasm panels will facilitate a more comprehensive assessment of these factors. The *in silico* digestion analysis was performed using the chromosome-level reference genome of *A. chinensis* cv. Hongyang. Although this reference genome provides a valuable framework for evaluating restriction enzyme combinations, the predicted restriction-site distribution may not fully represent the genomic diversity of other *Actinidia* species included in the present germplasm panel. Future studies incorporating multiple reference genomes or pangenome resources would enable a more comprehensive evaluation of restriction enzyme performance across diverse *Actinidia* germplasm. Overall, the present study suggests that the combination of *Ape*KI/*Tfi*I and GATK could provide a practical framework for SNP discovery in mixed-ploidy *Actinidia* germplasm. Although additional validation and ploidy-aware analyses are still required, the workflow presented here provides useful guidance for future genomic studies and genomics-assisted breeding in kiwifruit.

5. Conclusions

The methylation-sensitive *Ape*KI/*Tfi*I combination and GATK-based variant calling could provide a practical framework for high-density SNP discovery in polyploid kiwifruit. GATK identified substantially more candidate SNPs (828,257 SNPs) than freebayes and

bcftools under the analytical conditions used in this study. Because independent validation of SNP accuracy was beyond the scope of this study, these findings should be interpreted as differences in SNP discovery rather than comparative variant-calling accuracy. The resulting high-density SNP dataset provides a valuable resource for linkage mapping, QTL analysis, and molecular breeding in *Actinidia* species.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/horticulturae12080922/s1>, Table S1: Summary of GBS data, including the number of reads, mapping rate, depth of mapped region and reference genome coverage with 55 kiwifruit accessions; Table S2: Ratio of transitions/transversions in variants from three variant callers; File S1: R script for SNP annotation using GenomicRanges and rtracklayer.

Author Contributions: Conceptualization, D.K.; Methodology, Y.K.; Formal analysis, Y.K.; Investigation, Y.K.; Resources, M.L.; Data curation, M.L.; Writing—original draft preparation, Y.K.; Writing—review and editing, D.K.; Supervision, D.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was carried out with the support of the “Cooperative Research Program for Agriculture Science and Technology Development (Project No. RS-2020-RD009281)” Rural Development Administration, Republic of Korea.

Data Availability Statement: The data supporting the conclusions of this article are openly available in the NCBI Sequence Read Archive (SRA), accession number PRJNA1470575, available at <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1470575> (accessed on 22 July 2026).

Acknowledgments: During the preparation of this manuscript, the author(s) used Claude (Anthropic) for the purposes of English grammar editing. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

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

Abbreviations

The following abbreviations are used in this manuscript:

CDS	Coding DNA sequence
GATK	Genome Analysis Tool Kit
GBS	Genotyping-by-sequencing
GFF3	General feature format version 3
GWAS	Genome-wide association study
InDel	Insertion/Deletion
LTR-RT	Long terminal repeat retrotransposon
NGS	Next-generation sequencing
QTL	Quantitative trait locus
RE	Restriction enzyme
SNP	Single nucleotide polymorphism
Ts/Tv	Transition/transversion
UTR	Untranslated region

References

1. Saxena, R.K.; Edwards, D.; Varshney, R.K. Structural variations in plant genomes. *Brief. Funct. Genom.* **2014**, *13*, 296–307. [CrossRef]
2. Ganai, M.W.; Altmann, T.; Röder, M.S. SNP identification in crop plants. *Curr. Opin. Plant Biol.* **2009**, *12*, 211–217. [CrossRef] [PubMed]
3. Gao, Q.; Yue, G.; Li, W.; Wang, J.; Xu, J.; Yin, Y. Recent progress using high-throughput sequencing technologies in plant molecular breeding. *J. Integr. Plant Biol.* **2012**, *54*, 215–227. [CrossRef] [PubMed]
4. Metzker, M.L. Sequencing technologies—The next generation. *Nat. Rev. Genet.* **2010**, *11*, 31–46. [PubMed]

5. Peterson, G.W.; Dong, Y.; Horbach, C.; Fu, Y.B. Genotyping-by-sequencing for plant genetic diversity analysis: A lab guide for SNP genotyping. *Diversity* **2014**, *6*, 665–680. [\[CrossRef\]](#)
6. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Chung, Y.S.; Jun, T.; Kim, C. Digestion efficiency differences of restriction enzymes frequently used for genotyping-by-sequencing technology. *Korean J. Agric. Sci.* **2017**, *44*, 318–323.
8. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Nielsen, R.; Korneliussen, T.; Albrechtsen, A.; Li, Y.; Wang, J. SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. *PLoS ONE* **2012**, *7*, e37558. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Roy, S.; Coldren, C.; Karunamurthy, A.; Kip, N.S.; Klee, E.W.; Lincoln, S.E.; Leon, A.; Pullambhala, M.; Temple-Smolkin, R.L. Standards and guidelines for validating next-generation sequencing bioinformatics pipelines: A joint recommendation of the Association for Molecular Pathology and the College of American Pathologists. *J. Mol. Diagn.* **2018**, *20*, 4–27. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Xu, C. A review of somatic single nucleotide variant calling algorithms for next-generation sequencing data. *Comput. Struct. Biotechnol. J.* **2018**, *16*, 15–24. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Zverinova, S.; Guryev, V. Variant calling: Considerations, practices, and developments. *Hum. Mutat.* **2022**, *43*, 976–985. [\[PubMed\]](#)
13. Garrison, E.; Marth, G. Haplotype-based variant detection from short-read sequencing. *arXiv* **2012**, arXiv:1207.3907.
14. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **2011**, *27*, 2987–2993. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Lefouili, M.; Nam, K. The evaluation of Bcftools mpileup and GATK HaplotypeCaller for variant calling in non-human species. *Sci. Rep.* **2022**, *12*, 11331. [\[CrossRef\]](#) [\[PubMed\]](#)
16. DePristo, M.A.; Banks, E.; Poplin, R.; Garimella, K.V.; Maguire, J.R.; Hartl, C.; Philippakis, A.A.; del Angel, G.; Rivas, M.A.; Hanna, M.; et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **2011**, *43*, 491–498. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Yao, Z.; You, F.M.; N'Diaye, A.; Knox, R.E.; McCartney, C.; Hiebert, C.W.; Xu, W. Evaluation of variant calling tools for large plant genome re-sequencing. *BMC Bioinform.* **2020**, *21*, 1–16. [\[CrossRef\]](#)
18. Bertoli, D.J.; Ozias-Akins, P.; Chu, Y.; Dantas, K.M.; Santos, S.P.; Gouvea, E.; Guimarães, P.M.; Leal-Bertoli, S.C.M.; Knapp, S.J.; Moretzsohn, M.C. The use of SNP markers for linkage mapping in diploid and tetraploid peanuts. *G3-Genes Genomes Genet.* **2014**, *4*, 89–96. [\[CrossRef\]](#)
19. Clevenger, J.P.; Ozias-Akins, P. SWEEP: A tool for filtering high-quality SNPs in polyploid crops. *G3-Genes Genomes Genet.* **2015**, *5*, 1797–1803. [\[CrossRef\]](#)
20. Makhoul, M.; Rambla, C.; Voss-Fels, K.P.; Hickey, L.T.; Snowdon, R.J.; Obermeier, C. Overcoming polyploidy pitfalls: A user guide for effective SNP conversion into KASP markers in wheat. *Theor. Appl. Genet.* **2020**, *133*, 2413–2430. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Wu, H.; Ma, T.; Kang, M.; Ai, F.; Zhang, J.; Dong, G.; Liu, J. A high-quality *Actinidia chinensis* (kiwifruit) genome. *Hortic. Res.* **2019**, *6*, 117. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Cooke, D.P.; Wedge, D.C.; Lunter, G. Benchmarking small-variant genotyping in polyploids. *Genome Res.* **2022**, *32*, 403–408. [\[PubMed\]](#)
23. Song, J.; Yang, X.; Resende, M.F., Jr.; Neves, L.G.; Todd, J.; Zhang, J.; Comstock, J.C.; Wang, J. Natural allelic variations in highly polyploid *Saccharum* complex. *Front. Plant Sci.* **2016**, *7*, 804. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Huang, Y.F.; Poland, J.A.; Wight, C.P.; Jackson, E.W.; Tinker, N.A. Using genotyping-by-sequencing (GBS) for genomic discovery in cultivated oat. *PLoS ONE* **2014**, *9*, e102448. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Melo, A.T.; Guthrie, R.S.; Hale, I. GBS-based deconvolution of the surviving North American collection of cold-hardy kiwifruit (*Actinidia* spp.) germplasm. *PLoS ONE* **2017**, *12*, e0170580. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Oh, S.; Lee, M.; Kim, K.; Han, H.; Won, K.; Kwack, Y.B.; Kim, D. Genetic diversity of kiwifruit (*Actinidia* spp.), including Korean native *A. arguta*, using single nucleotide polymorphisms derived from genotyping-by-sequencing. *Hortic. Environ. Biotechnol.* **2019**, *60*, 105–114. [\[CrossRef\]](#)
27. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **2011**, *17*, 10–12. [\[CrossRef\]](#)
28. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Popowski, E.; Thomson, S.J.; Knäbel, M.; Tahir, J.; Crowhurst, R.N.; Davy, M.; Chagné, D. Construction of a high-density genetic map for hexaploid kiwifruit (*Actinidia chinensis* var. *deliciosa*) using genotyping by sequencing. *G3-Genes Genomes Genet.* **2021**, *11*, jkab142. [\[CrossRef\]](#)
30. Hwang, K.; Oh, S.; Kim, K.; Han, H.; Oh, Y.; Lim, H.; Kim, Y.K.; Kim, D. Genotyping-by-sequencing approaches using optimized two-enzyme combinations in Asian pears (*Pyrus* spp.). *Mol. Breed.* **2019**, *39*, 1–9. [\[CrossRef\]](#)

31. Choi, J.Y.; Purugganan, M.D. Evolutionary epigenomics of retrotransposon-mediated methylation spreading in rice. *Mol. Biol. Evol.* **2018**, *35*, 365–382. [[PubMed](#)]
32. Ou, S.; Jiang, N. LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* **2018**, *176*, 1410–1422. [[PubMed](#)]
33. Tian, S.; Yan, H.; Neuhauser, C.; Slager, S.L. An analytical workflow for accurate variant discovery in highly divergent regions. *BMC Genom.* **2016**, *17*, 1–15. [[CrossRef](#)]
34. Uitdewilligen, J.G.A.M.; Wolters, A.M.A.; D'hoop, B.B.; Borm, T.J.A.; Visser, R.G.F.; van Eck, H.J. A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. *PLoS ONE* **2013**, *8*, e62355. [[CrossRef](#)] [[PubMed](#)]
35. Korte, A.; Farlow, A. The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods* **2013**, *9*, 29. [[CrossRef](#)] [[PubMed](#)]
36. DeFaveri, J.; Viitaniemi, H.; Leder, E.; Merilä, J. Characterizing genic and nongenic molecular markers: Comparison of microsatellites and SNPs. *Mol. Ecol. Resour.* **2013**, *13*, 377–392. [[CrossRef](#)] [[PubMed](#)]
37. Bourke, P.M.; Voorrips, R.E.; Visser, R.G.F.; Maliepaard, C. Tools for genetic studies in experimental populations of polyploids. *Front. Plant Sci.* **2018**, *9*, 513. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.