

Article

Optimized Pepper Target SNP-Seq Applied in Population Structure and Genetic Diversity Analysis of 496 Pepper (*Capsicum* spp.) Lines

Yihao Wang ^{1,†}, Xiaofen Zhang ^{2,†} , Jingjing Yang ², Bin Chen ^{1,3}, Jian Zhang ^{2,4}, Wenyue Li ⁵, Heshan Du ^{2,*} and Sansheng Geng ^{1,*}

¹ State Key Laboratory of Vegetable Biobreeding, Beijing Vegetable Research Center, Beijing Academy of Agriculture and Forestry Science, Beijing 100097, China; wangyihao@nercv.org (Y.W.); chenbin@nercv.org (B.C.)

² National Engineering Research Center for Vegetables, Beijing Vegetable Research Center, Beijing Academy of Agriculture and Forestry Science, Beijing 100097, China; zhangxiaofen@nercv.org (X.Z.); yangjingjing@nercv.org (J.Y.); zhangjian@nercv.org (J.Z.)

³ Beijing Key Laboratory of Vegetable Germplasms Improvement, Beijing 100097, China

⁴ Key Laboratory of Biology and Genetics Improvement of Horticultural Crops (North China), Beijing 100097, China

⁵ Henan OULAND Seed Industry Co., Ltd., Zhengzhou 450003, China; 13592613788@139.com

* Correspondence: duheshan@nercv.org (H.D.); gengsansheng@nercv.org (S.G.); Tel.: +86-(010)-51503701 (H.D.); +86-(010)-51503007 (S.G.)

† These authors contributed equally to this work.

Abstract: Peppers are a major vegetable crop worldwide. With the completion of additional genome assemblies, a multitude of single-nucleotide polymorphisms (SNPs) can be utilized for population structure and genetic diversity analysis. In this study, we used target SNP-sequencing as a new high-throughput sequencing technology, screening out 425 perfect SNPs for analyzing the genetic diversity and population structure among 496 pepper lines from five pepper species in China and abroad. The perfect SNP panel exhibited commendable discriminative ability, as indicated by the average values of polymorphism information content, observed heterozygosity, minor allele frequency, and genetic diversity, which were 0.346, 0.011, 0.371, and 0.449, respectively. Based on phylogenetic, population structure, and principal component analyses, 484 *C. annuum* lines were divided into four subpopulations according to the shape of fruit: blocky fruit, wide-horn fruit, narrow-horn fruit, and linear fruit. These subpopulations displayed clear clustering with minimal or no overlap. Moreover, *F* statistic (*F_{st}*) analysis revealed considerable distinctions among these subpopulations. Additionally, we established a set of 47 core SNPs that could effectively differentiate among all pepper lines. This core SNP set could precisely classify the *C. annuum* lines into four distinct fruit-shape groups. The blocky and narrow-horn fruit subpopulations displayed the lowest and highest genetic diversity, respectively. This study highlights the importance of fruit shape as a crucial trait in pepper breeding. Moreover, this work indicates the immense potential of optimized target SNP technology in the addition of foreground markers of important traits to improve molecular breeding efficiency, and demonstrates its broad application prospects in the genetic analysis and variety identification of peppers.

Keywords: pepper; target SNP-seq; fruit shape; genetic diversity; population structure



Citation: Wang, Y.; Zhang, X.; Yang, J.; Chen, B.; Zhang, J.; Li, W.; Du, H.; Geng, S. Optimized Pepper Target SNP-Seq Applied in Population Structure and Genetic Diversity Analysis of 496 Pepper (*Capsicum* spp.) Lines. *Genes* **2024**, *15*, 214. <https://doi.org/10.3390/genes15020214>

Academic Editor: Hailong Yu

Received: 4 January 2024

Revised: 26 January 2024

Accepted: 2 February 2024

Published: 7 February 2024



Copyright: © 2024 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/>).

1. Introduction

Pepper (*Capsicum* spp.) is an important vegetable crop that originated in South America and is cultivated worldwide [1]. According to the Food and Agriculture Organization, global pepper cultivation reached 2.05 million hectares in 2021, resulting in a total production of 36.29 million tons [2]. China is the largest pepper-producing country in the world,

with a cultivation area of 0.76 million hectares and a production of 16.75 million tons [2]. A total of 42 species of pepper have been reported to date, including five domesticated species: *Capsicum annuum* L., *Capsicum frutescens* L., *Capsicum chinense* Jacq., *Capsicum baccatum* L., and *Capsicum pubescens* Ruiz and Pav. [3]. Among them, *C. annuum* is the most widely cultivated domesticated species worldwide. Continuous selection and breeding efforts have contributed to the development of pepper varieties with diverse fruit shapes, colors, and flavors [4].

Over the past few decades, with the continuous expansion of the seed market, a growing number of commercial pepper varieties have emerged. The identification and certification of varieties have become increasingly important to protecting the interests of breeders and producers. Traditionally, variety identification was achieved via morphological characteristics determined in field investigations; however, this approach is time-consuming, labor-intensive, and susceptible to environmental interference [5]. The use of molecular markers and the establishment of fingerprints as an assisting method provide more convenience and efficiency for variety identification and DUS testing and are better suited to meet current high-throughput detection requirements [6]. With completion of the whole-genome sequences and assemblies of multiple pepper varieties, including Zunla-1, CM334, and Ca59 [1,7,8], many simple sequence repeat (SSR) and insertion–deletion (In-Del) markers have been developed and utilized to detect pepper genetic diversity and domestication traits, as well as for variety identification [9–13]. However, SSR and InDel markers, along with random amplified polymorphic DNA and restriction fragment length polymorphism markers, have limitations and cannot be widely used for identification on a large scale [14,15]. Alternatively, single-nucleotide polymorphisms (SNPs) are the most abundant and genetically stable variations in the genome, which offer several advantages as markers, including rapid and accurate high-throughput detection and easy integration of genotype information, making them ideal for analyzing genetic backgrounds [16,17].

SNP-based genotyping techniques such as cleaved amplified polymorphic sequences (CAPS) and derived CAPS have been extensively utilized in genetic research [18,19]. However, these methods have certain limitations owing to the inability of electrophoretic systems to meet the current demands of rapid and large-scale testing. With the continuous advancement of high-throughput sequencing technologies, SNP genotyping platforms have emerged, including competitive allele-specific polymerase chain reaction (KASP) and DNA microarrays [20,21]. Nevertheless, these methods are relatively expensive, and their usage is constrained by time-consuming design-related detection procedures, thereby limiting their suitability for genotyping a substantial number of SNPs and samples [22]. Currently, genotyping by target sequencing technology is widely employed as an SNP genotyping method that integrates multiplex PCR and next-generation sequencing technologies, which has also been successfully commercialized and applied to various crops [23–27].

In contrast to KASP, microarray, whole-genome resequencing, and other SNP genotyping methods, SNP sequencing (SNP-seq) is a more effective and cost-efficient method. SNP-seq offer flexibility in terms of the number of SNPs and samples used, while demonstrating high genotyping accuracy for highly polymorphic and conserved sequences within the genome [28]. In particular, target SNP-seq technology combines multiplex amplification with high-throughput sequencing. A higher sequencing depth (with an average depth of 1000×) ensures a higher SNP detection rate and accuracy in genotyping, thereby reducing the occurrence of false-positives and false-negatives. This technology was first applied to the genetic analysis and variety identification of 261 cucumber varieties [6]. Target SNP-seq technology has since been successfully employed to construct fingerprints in crops such as cucumber [28], melon [29], eggplant [30], watermelon [31], and pepper [32].

In our previous study, we used target SNP-seq technology to analyze the genetic diversity of 271 commercial pepper varieties based on 92 perfect SNPs [32]. However, there were a limited number of SNPs for this analysis, and the collected pepper varieties were restricted to commercial hybrids. Therefore, in the present study, we optimized target SNP-seq technology by selecting 425 perfect SNP loci and adding several important trait

loci, significantly enriching the density and quality of markers. In addition, we collected 496 pepper lines from five pepper species in China and abroad. Genetic diversity and population structure analyses of these pepper lines based on SNP markers provide valuable insights into artificial selection in pepper breeding. The selected core-SNPs set, consisting of 47 SNP markers, can also be used for pepper variety identification and intellectual property protection. Adding important foreground markers in the SNP panel can improve the efficiency of pepper molecular breeding.

2. Materials and Methods

2.1. Plant Material and DNA Isolation

A total of 496 pepper lines were included in this study (Table S1): 69 were obtained from the U.S. National Plant Germplasm System, 115 were provided by Henan Ouland Seed Industry Co., Ltd. (Zhengzhou, China), 222 were sourced from the pepper genetic breeding group of Beijing Vegetable Research Center (BVRC), and 90 were introduced from the Vegetables Germplasm System of Beijing Academy of Agriculture and Forestry Sciences. Additional information regarding these pepper lines is provided in Table S1. DNA was extracted from the fresh leaves of seedlings of each inbred line using the modified CTAB method [33], and DNA quality was determined by 1.5% agarose gel electrophoresis.

2.2. Resequencing and Genome-Wide SNP Discovery

Genome-wide SNP discovery was performed using resequencing data from 31 pepper inbred lines that were previously sequenced by the Pepper Genetic Breeding Group of BVRC (publicly accessible at <http://bigd.big.ac.cn/gsa> accessed on 8 August 2023, accession number CRA001576) [32], along with reported sequencing data from four cultivars: Dempsey, Zunla-1, Perennial, and Chiltepin. All high-quality reads were aligned to the reference genome Zunla-1 v2.0 [1] using the Burrows–Wheeler alignment (BWA) tool with default parameters, and the mapped reads were filtered out by PCR duplication. Subsequently, perfect SNPs were selected from the whole genome of pepper using Genome Analysis Toolkit (GATK, v2.4-7g5e89f01) [34], with the following stringent filtering criteria: (1) minor allele frequency (MAF) > 0.4; (2) missing rate < 0.2; (3) heterozygosity < 0.2; (4) no sequence variations in the flanking region of 100 bp; and (5) two alleles per SNP locus.

2.3. SNP Genotype Analysis by Target SNP-Seq

Library construction for target SNP-seq was performed according to the target SSR-seq protocol reported previously [6], which involved two rounds of PCR. At first, a multiplex PCR panel was used with primers targeting the 150-bp flanking regions of the perfect SNPs. In the second round, a unique barcode was introduced to the amplified products of each sample to enable their differentiation. The purified PCR products from each sample were then combined to prepare the target SNP-seq library, which was subsequently subjected to sequencing on an Illumina HiSeq X Ten platform (Molbreeding Biotechnology Company, Shijiazhuang, China).

The Illumina bcl2fastq pipeline (Illumina, San Diego, CA, USA) was employed to demultiplex the raw data obtained from target SNP-seq. This pipeline uses a sample-unique barcode to determine the precise genotype of each inbred line. Next, the clean data were filtered using Trimmomatic, and the reads of each inbred line were aligned to the reference genome of pepper, Zunla-1 v2.0, using BWA with default parameters. SNP genotypes were subsequently named using GATK. Based on the high-throughput sequencing results, alleles with the highest and second-highest read counts were defined as the major and minor alleles, respectively. A locus was identified as homozygous if the read frequency of the major allele exceeded 0.7. If both the major and minor allele read frequencies were >0.35, the locus was classified as heterozygous.

2.4. Genetic Structure Analysis of Pepper Lines

Three methods, namely, principal component analysis (PCA), population structure, and phylogenetic tree construction, were used to analyze the genetic relationships among pepper lines. STRUCTURE v2.3 [35], which is based on a Bayesian model approach, was employed to analyze the population structure following standard procedures. Twenty independent runs were conducted with the cluster number (K) ranging from 1 to 15, a burn-in period of 100,000 iterations, and 100,000 Monte Carlo Markov Chain simulations. The K value was determined using Evanno's delta K method [36].

PCA was performed using the FactoMineR package in R [37]. An unrooted phylogenetic tree was constructed using the neighbor-joining method with the Ape and Poppr packages in R [38] and was further enhanced using iTOL for improved visualization [39].

2.5. Population Diversity Analysis of Pepper Lines

Perl script was used to calculate the MAF, genetic diversity (GD), observed heterozygosity (Ho), and polymorphism information content (PIC) of subpopulations with different fruit shapes [6]. We performed AMOVA within and between groups [40], as well as pairwise Fst analysis [41] using the poppr and Hierfstat R packages, respectively, to determine the genetic differences between populations and subpopulations.

2.6. Selection of the Core-SNPs Set for Inbred Line Discrimination

A Perl script was used to analyze the genetic diversity at each SNP site to screen the minimum number of SNPs required to distinguish the maximum number of pepper lines [6]. The selected SNPs were recognized as the core-SNP set with high discrimination. After obtaining the core-SNP set, saturation curves were generated by conducting pairwise comparisons of the genotypes for all pepper lines.

2.7. Selection of Core Pepper Lines

Within each subpopulation, a pairwise comparison matrix was constructed by counting the number of differential SNP genotypes between each pair of pepper lines and considering missing genotypes as null values [6,42]. A lower number of differential SNP genotypes for a given pepper line compared with the others indicated a closer relationship. The top 10% of pepper lines with a closer relationship than the others in each subpopulation were considered the core pepper lines for this analysis.

3. Results

3.1. Genome-Wide Perfect SNPs Identified

We successfully mapped our previously sequenced resequencing data of 31 inbred pepper lines [32] to the reference genome Zunla-1 v2.0 [1]. We also identified raw polymorphic SNPs across the entire genome sequences of these 31 inbred pepper lines and four cultivars with available full-genome sequences: Dempsey, Zunla-1, Perennial, and Chiltepin [7]. Considering the large number of highly repetitive sequences in the pepper genome, we refined the selection of candidate SNPs for target SNP genotyping.

A total of 425 perfect SNPs were selected for target SNP genotyping and were evenly distributed across the 12 pepper chromosomes, with an average of 35 SNPs on each chromosome. The average physical distance between SNP markers was 5.92 Mb (Figure 1a; Tables S2 and S3).

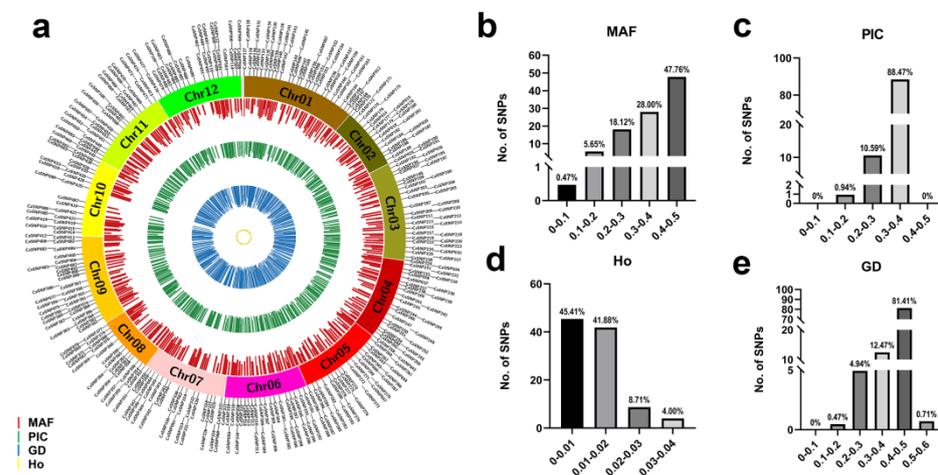


Figure 1. Genetic characteristics of 425 perfect single-nucleotide polymorphisms (SNPs) in 496 pepper lines. (a) Distribution of 425 SNP loci on the 12 pepper chromosomes. (b) Minor allele frequency (MAF). (c) Polymorphic information content (PIC). (d) Observed heterozygosity (Ho). (e) Genetic diversity (GD).

3.2. Genotyping of Pepper Lines Using Target SNP-seq

A total of 496 pepper lines were genotyped with the 425 perfect SNP loci using target SNP-seq. Additionally, fingerprints for each of the pepper lines were established using these 425 perfect SNPs (Table S4). Overall, 47.76% of these perfect SNPs exhibited an MAF between 0.4 and 0.5, suggesting a better discriminating ability (Figure 1b). The PIC values of the perfect SNPs ranged from 0.100 to 0.384, with a mean of 0.346. Furthermore, 81.41% of the SNPs showed a PIC greater than 0.3, while only four SNPs showed the PIC less than 0.2 (Figure 1c). The Ho values for each SNP ranged from 0 to 0.038, with a mean of 0.011, indicating that the 496 pepper lines had high homozygosity (Figure 1d). The average GD of the 425 SNPs was 0.449, ranging from 0.106 to 0.504 (Figure 1e). Notably, three SNPs exhibited an inbreeding coefficient of 1, indicating the absence of heterozygous genotypes in all pepper lines (Table S4). Overall, these results revealed that the 425 perfect SNPs of pepper possess high discrimination power and are suitable for genetic diversity analysis.

3.3. Genetic Structure of Pepper Lines

The genetic structure of the 496 pepper lines was analyzed based on the 425 perfect SNPs selected above. Most of these pepper lines were bred in China, with only a small proportion imported from abroad. A phylogenetic neighbor-joining tree was constructed, and the genetic distances were well-arranged according to interspecific relationships. Most of the pepper lines belonged to *C. annuum*, with three lines each belonging to *C. baccatum* and *C. frutescens*, one line belonging to *C. chacoense*, and five lines belonging to *C. chinense* (Figure 2; Table S1).

To further analyze the genetic structure of the 484 *C. annuum* lines, the most likely number of clusters (K-value) was determined to be two, suggesting the presence of two main populations (Figure 3a), which were designated POP1, comprising 314 pepper lines (64.9%), and POP2, comprising 170 pepper lines (35.1%). Increasing the K-value to three revealed that POP1 could be further divided into two subpopulations based on fruit shape: POP1A, which predominantly consisted of blocky fruit pepper lines, nine wide-horn fruit lines, and one narrow-horn fruit line; and POP1B, comprising mostly wide-horn fruit lines and a small percentage of narrow-horn fruit lines. When the K-value was set to four, POP2 was further divided into two subpopulations: POP2A, dominated by narrow-horn fruit lines with a few linear fruit lines and one wide-horn fruit line (similar to POP1); and POP2B, comprising only 11 linear fruit lines. Overall, the four subpopulations (POP1A, POP1B, POP2A, and POP2B) exhibited a clear structure with minimal mixing, primarily driven by different fruit shapes (Figure 3b; Table S5).

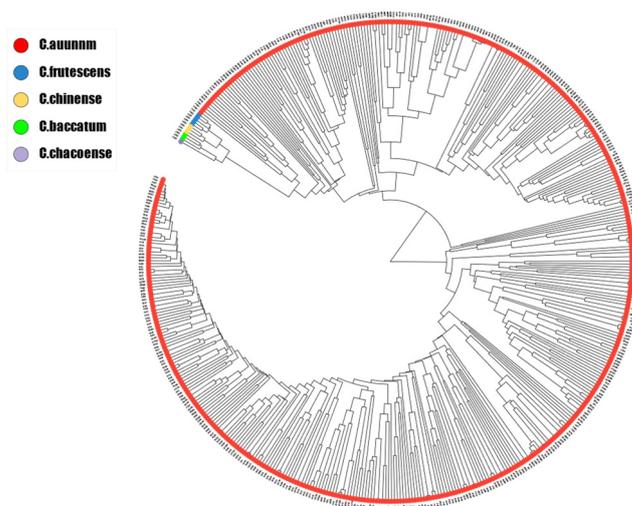


Figure 2. Neighbor-joining tree of 496 pepper lines based on 425 perfect single-nucleotide polymorphisms.

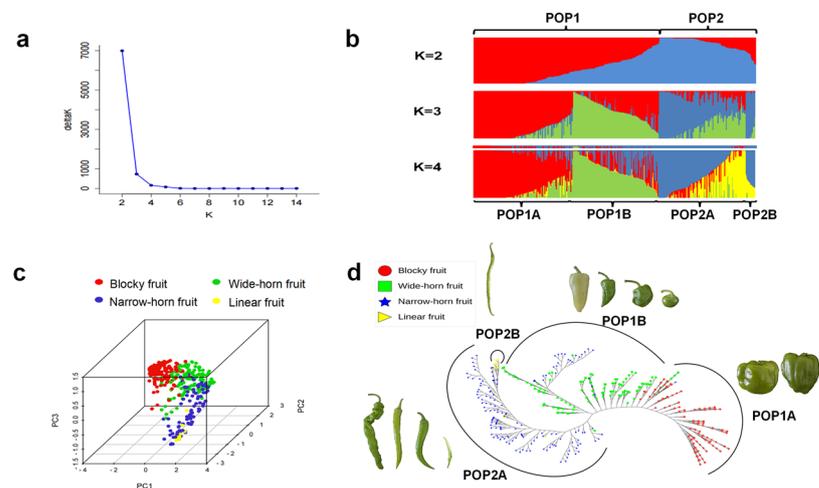


Figure 3. Population structure analysis of 484 *C. annuum* lines. (a) Delta K value derived from target SNP-seq genotype data using STRUCTURE. (b) All *C. annuum* lines were divided into two main populations (POP1 and POP2) when $K = 2$, and four subpopulations (POP1A, POP1B, POP2A, and POP2B) when $K = 4$. (c) Principal component analysis (PCA) of 484 *C. annuum* lines. (d) Unrooted neighbor-joining tree of the four subpopulations.

PCA was conducted to investigate the clusters of the 484 *C. annuum* lines using the 425 perfect SNPs. The PCA results were consistent with those of the analyses described above, demonstrating that the four populations clustered separately according to different fruit shapes. The blocky and wide-horn fruit lines exhibited more concentrated clustering, whereas the narrow-horn and linear fruit lines exhibited greater dispersal along the PCA plot. Interestingly, the linear and blocky fruit lines did not overlap in the cluster, indicating higher diversity and more pronounced genetic differentiation. Notably, the narrow-horn and linear fruit populations appeared to be more closely related despite their distinguishable fruit shapes (Figure 3c).

The unrooted neighbor-joining tree of the *C. annuum* lines constructed based on pairwise genetic distances provided additional support for the PCA- and model-based population structure analyses, revealing obvious differences among the four populations in terms of fruit shape, although some mixtures were observed between the wide- and narrow-horn fruit lines (Figure 3d). The representative lines corresponding to the fruit shapes of the subpopulations are depicted in Figure 3d. Collectively, the division of *C. annuum* lines into four genetic populations associated with different fruit shapes was strongly supported

by three independent analyses. This suggests that fruit shape has played a significant role in the selection and breeding of these *C. annuum* lines.

3.4. Identification of a Core-SNP Set

The fraction of informative SNPs that can differentiate the genotypes of all materials are typically selected as the core-SNP set, which is commonly used in variety and kinship identification. We found that 47 SNPs could accurately differentiate genotypes among the 484 *C. annuum* lines with a 100% accuracy rate (Figure 4a; Table S6). These 47 SNPs comprised the core-SNP set, which guaranteed that at least two distinct SNPs were present in 83% of the lines (Figure 4b). In addition, the neighbor-joining tree constructed using the 47 core SNPs clearly showed separation of the four distinct clusters (Figure 4c). Consequently, these 47 core SNPs were highly consistent with the analysis using the entire set of 425 perfect SNPs, demonstrating their suitability to effectively represent the genetic diversity of the 484 *C. annuum* lines, serving as reliable markers for similar varieties' identification.

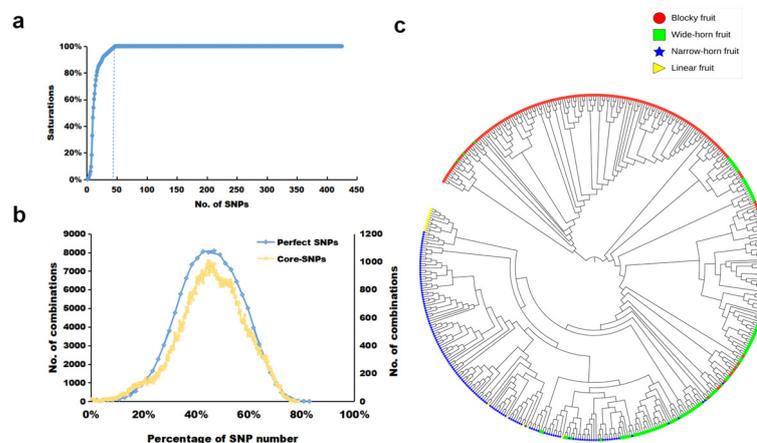


Figure 4. Identification ability of 47 core single-nucleotide polymorphisms (SNPs) in 496 pepper lines. (a) Discriminating saturation curve of 47 core SNPs. (b) Comparison of 47 core SNPs with 425 perfect SNPs. The horizontal axis and vertical axis denote the percentage of SNPs contributing to different genotype combinations and the number of different genotype combinations between 2 and 496 pepper lines, respectively. The left and right vertical axes correspond to the 47 core SNPs and 425 perfect SNPs, respectively. (c) Neighbor-joining tree of the 496 pepper lines based on the 47 core SNPs.

3.5. Genetic Variation in Pepper Populations

The analysis of molecular variance (AMOVA) for the population structure of the 484 *C. annuum* lines showed that population differences contributed to a minimum variation of 1.53%, whereas 9.32% of the total variation was attributed to differences between subpopulations (Table 1). The largest variation (86.6%) was observed within pepper lines in the subpopulations, whereas the contribution of variation within pepper lines accounted for only 2.55% of the total variation (Table 1). Concurrently, pairwise F statistic (F_{st}) estimates between populations and subpopulations (Table 2) revealed a high level of differentiation between POP1 and POP2 ($F_{st} = 0.3027$). The lowest pairwise F_{st} value (0.1672) among the four subpopulations was observed between POP2B (consisting of linear fruit lines) and POP2A (mainly composed of narrow-horn fruit lines), indicating relatively low genetic differentiation between these subpopulations. In contrast, the highest pairwise F_{st} value of 0.6229 was observed between POP1A (mostly comprising blocky fruit lines) and POP2B, indicating a high level of genetic differentiation.

Table 1. Analysis of molecular variance results between and within populations and subpopulations.

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Variation Component	Percentage of Variation
Between populations	1	3612.23	3612.23	1.52	1.53
Between subpopulations within populations	2	3886.47	1943.24	9.26	9.32
Between samples within subpopulations	480	85,948.64	174.69	86.08	86.6
Within samples	484	1257.78	2.54	2.54	2.55
Total	967	94,705.12	95.57	99.4	100

Table 2. Pairwise F statistic (Fst) estimates among subpopulations.

(Sub) Populations	POP1	POP1B	POP2A	POP2B
POP2	0.3027			
POP1A		0.1988	0.4363	0.6229
POP1B			0.2215	0.3982
POP2A				0.1672

3.6. Genetic Diversity and Core Pepper Lines of Subpopulations

The results of the polymorphism analysis of 425 SNP markers across the four subpopulations showed that POP1A and POP1B had 420 and 425 polymorphic SNP markers, respectively, whereas all SNP markers in POP2A displayed polymorphisms (Table 3). In contrast, POP2B had the lowest number of polymorphic SNP markers, with a total of 291, which is potentially attributable to the smaller number of pepper lines in this subpopulation (Table 3). These findings underscore the distinctions among the four subpopulations. Furthermore, we compared the genetic diversity of these subpopulations characterized by fruit shape using average genetic parameters and constructed a genetic similarity matrix by calculating the number of different SNP genotypes between every pair of DNA samples (Figure 5; Table S7). POP1A had the lowest PIC (0.223), MAF (0.189), Ho (0.007), and GD (0.270) values, along with a high inbreeding coefficient, indicating lower genetic diversity in the blocky fruit lines (Table 3). By contrast, POP2A, primarily comprising narrow-horn fruit lines, displayed higher genetic diversity with elevated PIC (0.301), MAF (0.285), Ho (0.021), and GD (0.379) values, along with a lower inbreeding coefficient (Table 3). Among the 484 *C. annuum* lines, the imported inbred line “21_G28” (PI 640889) exhibited the highest genetic similarity with other pepper lines, whereas the Chinese inbred line “old_147” showed the lowest genetic similarity with other pepper lines (Table S7). For each subpopulation, the top 10% of pepper lines with the fewest differing SNP genotypes were selected as core pepper lines, resulting in a total of 49 lines (Table S8).

Table 3. Genetic parameters in four subpopulations.

Subpopulation	Number of Pepper Lines	Number of Polymorphic Markers	PIC	MAF	Ho	GD	Inbreeding Coefficient
POP1A	183	420	0.223	0.189	0.006	0.27	0.972
POP1B	170	424	0.312	0.319	0.007	0.4	0.981
POP2A	120	425	0.301	0.285	0.021	0.379	0.936
POP2B	11	291	0.23	0.197	0.01	0.279	0.956

Abbreviations: PIC, polymorphism information content; MAF, minimum allele frequency; Ho, observed heterozygosity; GD, genetic diversity.

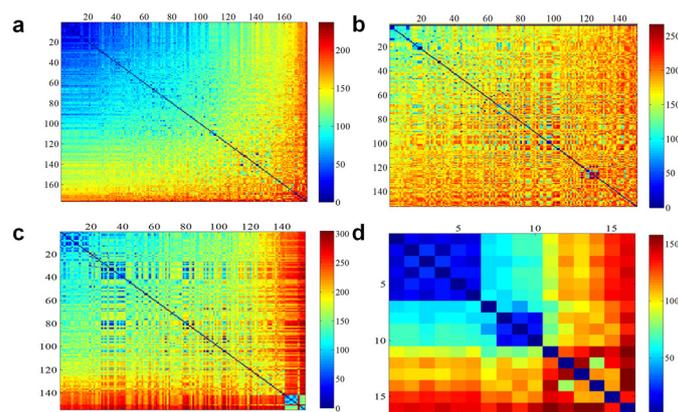


Figure 5. Pairwise comparison matrix heat maps of different single-nucleotide polymorphism (SNP) genotypes in subpopulations (a) POP1A, (b) POP1B, (c) POP2A, and (d) POP2B. The range from blue to red indicates an increase in the number of different SNP genotypes.

4. Discussion

High-throughput genotyping technologies are increasingly used in genetic research and breeding. In this study, based on our previous resequencing data of 35 inbred lines [32], we optimized target SNP-seq technology by identifying and expanding 425 perfect SNP markers and conducted genetic diversity analysis on 496 pepper lines. The perfect SNPs effectively distinguished the pepper lines, with the PIC values of 81.41% of the lines exceeding 0.3, and the MAF values of 47.76% of the lines exceeding 0.4. Compared to the previous study [32], we found richer variation among pepper materials, and the SNPs are more informative, making them widely applicable for pepper variety identification. It is worth noting that the H_o values analyzed in this study were lower than in previous the genetic diversity-related studies of pepper [9,11]. This indicated that the examples in our pepper lines with high homozygosity are suitable for resequencing for genome-wide association studies of important traits in the future, and the genetic polymorphism of domestic breeding resources were abundant. Furthermore, we selected a set of 47 core SNPs that could successfully distinguish between all pepper lines.

Over a prolonged period of artificial selection, five commonly cultivated species gradually emerged, with *C. annuum* being the most widely grown species at present. There has been substantial genetic research on pepper species, which has mainly focused on exploring their genetic relationships and clustering using molecular markers, whole-genome sequencing, and genotyping by sequencing [17,43–53]. Most of these findings suggest that *C. frutescens* is most closely related to *C. annuum*, followed by *C. chinense* [45,48,50,53]. In addition, some studies demonstrated that *C. chinense* is closely related to *C. annuum* [16,51,52]. Another study suggested that *C. frutescens* and *C. chinense* are more closely related to *C. baccatum* than to *C. annuum*, which was hypothesized to result from the geographical locations of the respective origin centers for different cultivated species [49]. The origin center of *C. baccatum* is in Bolivia, which is in close proximity to the Amazon basin, the origin site for *C. frutescens* and *C. chinense* [54,55]. Therefore, this discrepancy among studies may be attributed to the impact of artificial selection during the domestication of diverse pepper varieties in various regions and their adaptation to a wide range of global agricultural climates. Intriguingly, all of the aforementioned studies indicate that *C. frutescens* and *C. chinense* are the closest relatives, potentially because they share a common origin. The present genetic analysis of 496 pepper lines representing five pepper species revealed that *C. frutescens* was most closely related to *C. annuum*, with *C. baccatum* as the most distant species, consistent with the findings of Nicola et al. [48].

We used 484 pepper lines, including excellent pepper lines from our research team, breeding companies in China, and imported lines, to analyze the genetic diversity of *C. annuum* using phylogenetic analysis, population structure analysis, and PCA. All three methods divided the pepper lines into distinct populations according to fruit shape. In

population structure analysis, 484 *C. annuum* lines were separated into two main populations, POP1 and POP2, which were each subsequently separated into two subpopulations, POP1A/POP1B and POP2A/POP2B, respectively. These four subpopulations are closely related to fruit shape, corresponding to blocky, wide-horn, narrow-horn, and linear fruits, respectively. PCA and phylogenetic analysis showed that the four subpopulations divided by fruit shape were categorized distinctly with little or no overlap, indicating that fruit shape is an important selection criterion in pepper breeding. Although our previous study [32] also identified subpopulations based on four different fruit shapes, they were divided into five subgroups, with long-horn fruit lines split into two subpopulations. Additionally, from the phylogenetic and principal component analyses of our previous study, blocky fruit lines seemed to be more closely related to long-horn fruit lines. Similarly, in a recent study [56], population structure and phylogenetic analyses of 186 pepper varieties resulted in four different fruit shape groups, with TJ and NJ seemingly representing the blocky fruit and narrow-horn fruit subpopulations of our analysis, respectively, being the most closely related; whereas in our study, the lines of blocky fruit were more strongly associated with wide-horn fruit lines, which seemed more reasonable considering the regularity of fruit shape variation. These differences may be due to the different varieties and criteria used for fruit shape classification among studies.

In terms of genetic diversity parameters in different subpopulations, POP1A, which mainly consisted of blocky fruit, had the lowest MAF, PIC, Ho, and GD values and the highest inbreeding coefficient, indicating that this subpopulation is characterized by the lowest genetic diversity and genetic background, with limited genetic materials used in breeding processes. In contrast, narrow-horn fruits showed higher diversity, with the highest Ho value and the lowest inbreeding coefficient, consistent with previous findings [32,57,58]. Different from previous studies, the PIC, Ho, and MAF of POP1B, which mainly consisted of wide-horn fruit, were higher than those of narrow-horn fruits, which also showed high diversity. Therefore, fruit shape, as one of major traits in artificial selection breeding, significantly influences the genetic structure of peppers. This study provides valuable insights for future research on the population structure variation of peppers based on fruit shape.

With the continuous development of genome sequencing and resequencing technologies, as well as the assembly of large numbers of plant genomes, marker-assisted selection has become increasingly important in modern breeding practice. In our study, in addition to using perfect SNPs as background markers in target SNP-seq, we also added foreground selection markers which use specific primers to screen target genes, enabling the selection of individual plants with disease resistance and excellent traits at the seedling stage. Based on the individual disease-resistant loci added in a previous study [32,57,58], we successfully added specific primers to the flanking region of other high-quality trait-related functional sites (Table S9), such as genic male sterile genes *msc-1* and *msc-2* [59,60], capsaicin synthesis genes *Pun1* and *Pun3* [61,62], male fertility restoration gene *CaRf032* [63], functional site of fruit shape gene *Fs3.1* [51], fasciculate inflorescence gene *Cafa* [64], and fruit color (yellow/red) gene *CaCCS* [65], which were added into the perfect SNPs panel as foreground markers, combining with the background markers, and can be simultaneously detected in many pepper germplasms using target SNP-seq (Figure 6). This technology, with its advantages of low cost (in terms of both time and money) and high efficiency and flexibility, integrating foreground and background markers, can improve molecular breeding efficiency in the future and is suitable for a wide range of other applications, including genetic research and variety identification (Figure 6).

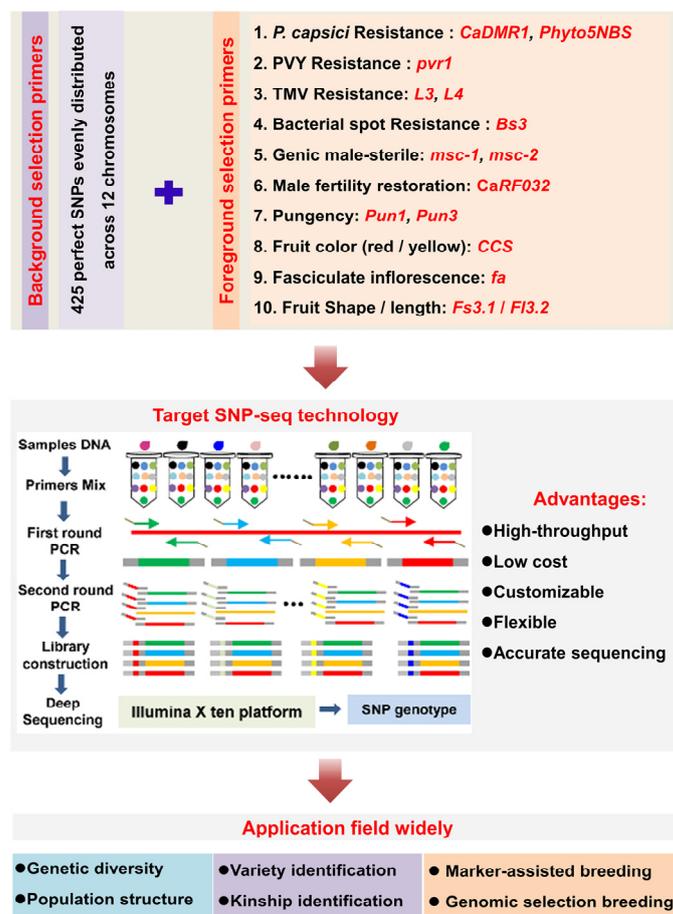


Figure 6. The flow chart of screening pepper materials by foreground and background markers based on optimized target SNP-seq and the application scope of this technology. References to functional sites of foreground markers are shown in Table S9.

5. Conclusions

In this study, we analyzed the genetic diversity of 496 pepper lines from five species with target SNP-seq. The analysis of 484 *C. annuum* lines showed that their population structure was significantly influenced by fruit shape. A core set containing 47 SNPs could effectively distinguish all pepper lines, and 49 core *C. annuum* lines were identified. The core-SNP panel with important foreground markers could be an effective tool for variety identification and marker-assisted breeding.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/genes15020214/s1>. Table S1: The species of each pepper line; Table S2: The average number of SNPs and physical distance between SNP markers on each chromosome; Table S3: Perfect SNP markers used in this study; Table S4: Genotyping of 496 pepper lines by 425 perfect SNPs; Table S5: The subpopulation and fruit shape of each pepper line; Table S6: The selected core SNPs in this study; Table S7: The number of different SNP genotypes between every pair of pepper lines in POP1A; Table S8: Core pepper lines of each subpopulations; Table S9: Foreground primers were used for the detection of the resistance loci and other trait-related loci through Target SNP-seq.

Author Contributions: S.G. designed the research. B.C., X.Z. and W.L. provided the pepper materials. J.Y., H.D., J.Z. and Y.W. analysed the data. Y.W. and X.Z. wrote the manuscript. H.D. revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by National Key Research and Development program of China, 2022YFD1200501; National Natural Science Foundation of China, 32072565; Construction Program of Science and Technology Innovation Capability of Beijing Academy of Agriculture and Forestry Sciences, KJCX20210421; Beijing Joint Research Program for Germplasm Innovation and New Variety Breeding, G20220628003-01; Beijing Innovation consortium of Agriculture Research System, BAIC01-2023-01; Development Program of Beijing Vegetable Research Center, KJCX20230126 and KYCX202301; and Henan Province Agricultural Variety Joint Research, 2022010501.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the Supplementary Material, further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The Author Wenyue Li was employed by the Henan OULAND Seed Industry Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Qin, C.; Yu, C.S.; Shen, Y.O.; Fang, X.D.; Chen, L.; Min, J.M.; Cheng, J.W.; Zhao, S.C.; Xu, M.; Luo, Y.; et al. Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 5135–5140. [CrossRef] [PubMed]
2. Food and Agriculture Organization of the United Nations, FAOSTAT. Available online: <https://www.fao.org/faostat/en/#data/QCL/visualize> (accessed on 8 August 2023).
3. Barboza, G.; García, C.; Scaldaferrero, M.; Bohs, L. *Capsicum* An amazing new (Solanaceae) species from the Andean-Amazonian Piedmont. *PhytoKeys* **2020**, *167*, 13–29. [CrossRef] [PubMed]
4. Taitano, N.E.K. Exploring Genetic Diversity Underlying Fruit Diversity in Tomato and Pepper. Ph.D. Thesis, University of Georgia, Athens, GA, USA, 2020.
5. Tommasini, L.; Batley, J.; Arnold, G.; Cooke, R.; Donini, P.; Lee, D.; Law, J.; Lowe, C.; Moule, C.; Trick, M.; et al. The development of multiplex simple sequence repeat (SSR) markers to complement distinctness, uniformity and stability testing of rape (*Brassica napus* L.) varieties. *Theor. Appl. Genet.* **2003**, *106*, 1091–1101. [CrossRef] [PubMed]
6. Yang, J.; Zhang, J.; Han, R.; Zhang, F.; Mao, A.; Luo, J.; Dong, B.; Liu, H.; Tang, H.; Zhang, J.; et al. Target SSR-Seq: A Novel SSR Genotyping Technology Associate with Perfect SSRs in Genetic Analysis of Cucumber Varieties. *Front. Plant Sci.* **2019**, *10*, 531. [CrossRef] [PubMed]
7. Kim, S.; Park, M.; Yeom, S.; Kim, Y.; Lee, J.; Lee, H.; Seo, E.; Choi, J.; Cheong, K.; Kim, K.; et al. Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nat. Genet.* **2014**, *46*, 270–278. [CrossRef] [PubMed]
8. Liao, Y.; Wang, J.; Zhu, Z.; Liu, Y.; Chen, J.; Zhou, Y.; Liu, F.; Lei, J.; Gaut, B.; Cao, B.; et al. The 3D architecture of the pepper genome and its relationship to function and evolution. *Nat. Commun.* **2022**, *13*, 3479. [CrossRef]
9. Zhang, X.M.; Zhang, Z.H.; Gu, X.Z.; Mao, S.L.; Li, X.X.; Chadoeuf, J.; Palloix, A.; Wang, L.H.; Zhang, B.X. Genetic diversity of pepper (*Capsicum* spp.) germplasm resources in China reflects selection for cultivar types and spatial distribution. *J. Integr. Agric.* **2016**, *15*, 1991–2001. [CrossRef]
10. Wu, B.D.; Fan, R.; Hu, L.S.; Wu, H.S.; Hao, C.Y. Genetic diversity in the germplasm of black pepper determined by EST-SSR markers. *Genet. Mol. Res.* **2016**, *15*, 9. [CrossRef]
11. Gu, X.Z.; Cao, Y.C.; Zhang, Z.H.; Zhang, B.X.; Zhao, H.; Zhang, X.M.; Wang, H.P.; Li, X.X.; Wang, L.H. Genetic diversity and population structure analysis of *Capsicum* germplasm accessions. *J. Integr. Agric.* **2019**, *18*, 1312–1320. [CrossRef]
12. Jo, Y.D.; Park, J.; Kim, J.; Song, W.; Hur, C.G.; Lee, Y.H.; Kang, B.C. Complete sequencing and comparative analyses of the pepper (*Capsicum annuum* L.) plastome revealed high frequency of tandem repeats and large insertion/deletions on pepper plastome. *Plant Cell Rep.* **2011**, *30*, 217–229. [CrossRef]
13. Esposito, S.; Cigliano, R.A.; Cardi, T.; Tripodi, P. Whole-genome resequencing reveals genomic footprints of Italian sweet and hot pepper heirlooms giving insight into genes underlying key agronomic and qualitative traits. *BMC Genom. Data* **2022**, *23*, 16. [CrossRef]
14. Yasuda, K.; Yamaguchi, H. Genetic diversity of vegetable water pepper (*Persicaria hydropiper* (L.) Spach) as revealed by RAPD markers. *Breed. Sci.* **2005**, *55*, 7–14. [CrossRef]
15. Lefebvre, V.; Palloix, A.; Rives, M. Nuclear RFLP between pepper cultivars (*Capsicum annuum* L.). *Euphytica* **1993**, *71*, 189–199. [CrossRef]
16. Lee, H.; Ro, N.; Jeong, H.; Kwon, J.; Jo, J.; Ha, Y.; Jung, A.; Han, J.; Venkatesh, J.; Kang, B. Genetic diversity and population structure analysis to construct a core collection from a large *Capsicum* germplasm. *BMC Genet.* **2016**, *17*, 142. [CrossRef] [PubMed]

17. Lozada, D.N.; Bhatta, M.; Coon, D.; Bosland, P.W. Single nucleotide polymorphisms reveal genetic diversity in New Mexican chile peppers (*Capsicum* spp.). *BMC Genom.* **2021**, *22*, 12. [[CrossRef](#)] [[PubMed](#)]
18. Semagn, K.; Bjornstad, A.; Ndjiondjop, M.N. An overview of molecular marker methods for plants. *AFR. J. Biotechnol.* **2006**, *5*, 2540–2568.
19. Yanagisawa, T.; Kiribuchi-Otobe, C.; Hirano, H.; Suzuki, Y.; Fujita, M. Detection of single nucleotide polymorphism (SNP) controlling the waxy character in wheat by using a derived cleaved amplified polymorphic sequence (dCAPS) marker. *Theor. Appl. Genet.* **2003**, *107*, 84–88. [[CrossRef](#)] [[PubMed](#)]
20. Semagn, K.; Babu, R.; Hearne, S.; Olsen, M. Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): Overview of the technology and its application in crop improvement. *Mol. Breed.* **2014**, *33*, 1–14. [[CrossRef](#)]
21. Meaburn, E.; Butcher, L.; Schalkwyk, L.; Plomin, R. Genotyping pooled DNA using 100K SNP microarrays: A step towards genome-wide association scans. *Nucleic Acids Res.* **2006**, *34*, e27. [[CrossRef](#)] [[PubMed](#)]
22. Reuter, J.; Spacek, D.; Snyder, M. High-throughput sequencing technologies. *Mol. Cell* **2015**, *58*, 586–597. [[CrossRef](#)]
23. Ma, J.; Cao, Y.Y.; Wang, Y.Z.; Ding, Y. Development of the maize 5.5K loci panel for genomic prediction through genotyping by target sequencing. *Front. Plant Sci.* **2022**, *13*, 13. [[CrossRef](#)] [[PubMed](#)]
24. Yang, Q.; Zhang, J.A.; Shi, X.L.; Chen, L.; Qin, J.; Zhang, M.C.; Yang, C.Y.; Song, Q.J.; Yan, L. Development of SNP marker panels for genotyping by target sequencing (GBTS) and its application in soybean. *Mol. Breed.* **2023**, *43*, 11. [[CrossRef](#)]
25. Guo, Z.F.; Wang, H.W.; Tao, J.J.; Ren, Y.H.; Xu, C.; Wu, K.S.; Zou, C.; Zhang, J.N.; Xu, Y.B. Development of multiple SNP marker panels affordable to breeders through genotyping by target sequencing (GBTS) in maize. *Mol. Breed.* **2019**, *39*, 12. [[CrossRef](#)]
26. Shen, Y.S.; Wang, J.S.; Shaw, R.K.; Yu, H.F.; Sheng, X.G.; Zhao, Z.Q.; Li, S.J.; Gu, H.H. Development of GBTS and KASP Panels for Genetic Diversity, Population Structure, and Fingerprinting of a Large Collection of Broccoli (*Brassica oleracea* L. var. *italica*) in China. *Front. Plant Sci.* **2021**, *12*, 17. [[CrossRef](#)]
27. Guo, Z.F.; Yang, Q.; Huang, F.F.; Zheng, H.J.; Sang, Z.Q.; Xu, Y.F.; Zhang, C.; Wu, K.S.; Tao, J.J.; Prasanna, B.M.; et al. Development of high-resolution multiple-SNP arrays for genetic analyses and molecular breeding through genotyping by target sequencing and liquid chip. *Plant Commun.* **2021**, *2*, 15. [[CrossRef](#)]
28. Zhang, J.; Yang, J.J.; Zhang, L.K.; Luo, J.; Zhao, H.; Zhang, J.A.; Wen, C.L. A new SNP genotyping technology Target SNP-seq and its application in genetic analysis of cucumber varieties. *Sci. Rep.* **2020**, *10*, 11. [[CrossRef](#)]
29. Zhang, J.; Yang, J.; Lv, Y.; Zhang, X.; Xia, C.; Zhao, H.; Wen, C. Genetic diversity analysis and variety identification using SSR and SNP markers in melon. *BMC Plant Biol.* **2023**, *23*, 39. [[CrossRef](#)]
30. Liu, W.L.; Qian, Z.W.; Zhang, J.; Yang, J.J.; Wu, M.S.; Barchi, L.; Zhao, H.Y.; Sun, H.H.; Cui, Y.L.; Wen, C.L. Impact of fruit shape selection on genetic structure and diversity uncovered from genome-wide perfect SNPs genotyping in eggplant. *Mol. Breed.* **2019**, *39*, 13. [[CrossRef](#)]
31. Yang, J.J.; Zhang, J.; Du, H.S.; Zhao, H.; Mao, A.J.; Zhang, X.F.; Jiang, L.; Zhang, H.Y.; Wen, C.L.; Xu, Y. Genetic relationship and pedigree of Chinese watermelon varieties based on diversity of perfect SNPs. *Hortic. Plant J.* **2022**, *8*, 489–498. [[CrossRef](#)]
32. Du, H.; Yang, J.; Chen, B.; Zhang, X.; Zhang, J.; Yang, K.; Geng, S.; Wen, C. Target sequencing reveals genetic diversity, population structure, core-SNP markers, and fruit shape-associated loci in pepper varieties. *BMC Plant Biol.* **2019**, *19*, 578. [[CrossRef](#)]
33. Fulton, T.M.; Tanksley, C.S.D. Microprep protocol for extraction of DNA from tomato and other herbaceous plants. *Plant Mol. Biol. Rep.* **1995**, *13*, 207–209. [[CrossRef](#)]
34. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [[CrossRef](#)] [[PubMed](#)]
35. Pritchard, J.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **2000**, *155*, 945–959. [[CrossRef](#)] [[PubMed](#)]
36. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **2005**, *14*, 2611–2620. [[CrossRef](#)] [[PubMed](#)]
37. Husson, F.; Josse, J.; Pagès, J. Principal component methods—Hierarchical clustering—Partitional clustering: Why would we need to choose for visualizing data? Unpublished work. 2010.
38. Kamvar, Z.N.; Tabima, J.F.; Grünwald, N.J. *Poppr*: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2014**, *2*, e281. [[CrossRef](#)] [[PubMed](#)]
39. Letunic, I.; Bork, P. Interactive Tree of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [[CrossRef](#)] [[PubMed](#)]
40. Excoffier, L.; Smouse, P.; Quattro, J. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **1992**, *131*, 479–491. [[CrossRef](#)] [[PubMed](#)]
41. de Meeûs, T.; Goudet, J. A step-by-step tutorial to use HierFstat to analyse populations hierarchically structured at multiple levels. *Infect. Genet. Evol.* **2007**, *7*, 731–735. [[CrossRef](#)]
42. Jones, H.; Norris, C.; Smith, D.; Cockram, J.; Lee, D.; O’Sullivan, D.M.; Mackay, I. Evaluation of the use of high-density SNP genotyping to implement UPOV Model 2 for DUS testing in barley. *Theor. Appl. Genet.* **2013**, *126*, 901–911. [[CrossRef](#)]
43. Moreira, A.F.P.; Ruas, P.M.; Ruas, C.d.F.; Baba, V.Y.; Giordani, W.; Arruda, I.M.; Rodrigues, R.; Goncalves, L.S.A. Genetic diversity, population structure and genetic parameters of fruit traits in *Capsicum chinense*. *Sci. Hortic.* **2018**, *236*, 1–9. [[CrossRef](#)]

44. Colonna, V.; D'Agostino, N.; Garrison, E.; Albrechtsen, A.; Tripodi, P. Genomic diversity and novel genome-wide association with fruit morphology in *Capsicum*, from 746k polymorphic sites. *Sci. Rep.* **2019**, *9*, 10067. [[CrossRef](#)] [[PubMed](#)]
45. González-Pérez, S.; Garcés-Claver, A.; Mallor, C.; Sáenz de Miera, L.; Fayos, O.; Pomar, F.; Merino, F.; Silvar, C. New insights into *Capsicum* spp. relatedness and the diversification process of *Capsicum annuum* in Spain. *PLoS ONE* **2014**, *9*, e116276. [[CrossRef](#)] [[PubMed](#)]
46. Pérez-Martínez, A.; Eguiarte, L.; Mercer, K.; Martínez-Ainsworth, N.; McHale, L.; van der Knaap, E.; Jardón-Barbolla, L. Genetic diversity, gene flow, and differentiation among wild, semiwild, and landrace chile pepper (*Capsicum annuum*) populations in Oaxaca, Mexico. *Am. J. Bot.* **2022**, *109*, 1157–1176. [[CrossRef](#)] [[PubMed](#)]
47. Feng, S.; Liu, Z.; Hu, Y.; Tian, J.; Yang, T.; Wei, A. Genomic analysis reveals the genetic diversity, population structure, evolutionary history and relationships of Chinese pepper. *Hortic. Res.* **2020**, *7*, 158. [[CrossRef](#)] [[PubMed](#)]
48. Nicola, M.; Cantet, M.; Lefebvre, V.; Sage-Palloix, A.M.; Palloix, A. Genotyping a large collection of pepper (*Capsicum* spp.) with SSR loci brings new evidence for the wild origin of cultivated *C. annuum* and the structuring of genetic diversity by human selection of cultivar types. *Genet. Resour. Crop Evol.* **2013**, *60*, 2375–2390. [[CrossRef](#)]
49. Pereira-Dias, L.; Vilanova, S.; Fita, A.; Prohens, J.; Rodríguez-Burruezo, A. Genetic diversity, population structure, and relationships in a collection of pepper (*Capsicum* spp.) landraces from the Spanish centre of diversity revealed by genotyping-by-sequencing (GBS). *Hortic. Res.* **2019**, *6*, 54. [[CrossRef](#)]
50. Ibiza, V.P.; Blanca, J.; Cañizares, J.; Nuez, F. Taxonomy and genetic diversity of domesticated *Capsicum* species in the Andean region. *Genet. Resour. Crop Evol.* **2012**, *59*, 1077–1088. [[CrossRef](#)]
51. Cao, Y.; Zhang, K.; Yu, H.; Chen, S.; Xu, D.; Zhao, H.; Zhang, Z.; Yang, Y.; Gu, X.; Liu, X.; et al. Pepper variome reveals the history and key loci associated with fruit domestication and diversification. *Mol. Plant* **2022**, *15*, 1744–1758. [[CrossRef](#)]
52. Costa, M.; do Rêgo, M.; da Silva, A.; do Rêgo, E.; Barroso, P. Characterization and genetic diversity of pepper (*Capsicum* spp.) parents and interspecific hybrids. *Genet. Mol. Res. GMR* **2016**, *15*, gmr.15027652. [[CrossRef](#)]
53. Wahyuni, Y.; Ballester, A.; Tikunov, Y.; de Vos, R.; Pelgrom, K.; Maharijaya, A.; Sudarmonowati, E.; Bino, R.; Bovy, A. Metabolomics and molecular marker analysis to explore pepper (*Capsicum* spp.) biodiversity. *Metabolomics* **2013**, *9*, 130–144. [[CrossRef](#)]
54. Mcleod, M.J.; Guttman, S.I.; Eshbaugh, W.H.; Rayle, R.E. An Electrophoretic Study of Evolution in *Capsicum* (Solanaceae). *Evolution* **1983**, *37*, 562–574. [[CrossRef](#)]
55. Moscone, E.A.; Scaldaferrero, M.A.; Grabielle, M.; Cecchini, N.M.; Sánchez García, Y.; Jarret, R.; Daviña, J.R.; Ducasse, D.A.; Barboza, G.E.; Ehrendorfer, F. The evolution of chili peppers (*Capsicum*-Solanaceae): A cytogenetic perspective. *Acta Hortic.* **2007**, *745*, 137–170. [[CrossRef](#)]
56. Xu, W.; Huang, H.; Li, X.; Yang, M.; Chi, S.; Pan, Y.; Li, N.; Paterson, A.; Chai, Y.; Lu, K. CaHMA1 promotes Cd accumulation in pepper fruit. *J. Hazard. Mater.* **2023**, *460*, 132480. [[CrossRef](#)]
57. Tam, S.; Lefebvre, V.; Palloix, A.; Sage-Palloix, A.; Mhiri, C.; Grandbastien, M. LTR-retrotransposons Tnt1 and T135 markers reveal genetic diversity and evolutionary relationships of domesticated peppers. *Theor. Appl. Genet.* **2009**, *119*, 973–989. [[CrossRef](#)]
58. Solomon, A.; Han, K.; Lee, J.; Lee, H.; Jang, S.; Kang, B. Genetic diversity and population structure of Ethiopian *Capsicum* germplasms. *PLoS ONE* **2019**, *14*, e0216886. [[CrossRef](#)] [[PubMed](#)]
59. Cheng, Q.; Wang, P.; Liu, J.; Wu, L.; Zhang, Z.; Li, T.; Gao, W.; Yang, W.; Sun, L.; Shen, H. Identification of candidate genes underlying genic male-sterile *msc-1* locus via genome resequencing in *Capsicum annuum* L. *Theor. Appl. Genet.* **2018**, *131*, 1861–1872. [[CrossRef](#)] [[PubMed](#)]
60. Cheng, Q.; Li, T.; Ai, Y.; Lu, Q.; Wang, Y.; Wu, L.; Liu, J.; Sun, L.; Shen, H. Phenotypic, genetic, and molecular function of *msc-2*, a genic male sterile mutant in pepper (*Capsicum annuum* L.). *Theor. Appl. Genet.* **2020**, *133*, 843–855. [[CrossRef](#)] [[PubMed](#)]
61. Stewart, C.; Kang, B.; Liu, K.; Mazourek, M.; Moore, S.; Yoo, E.; Kim, B.; Paran, I.; Jahn, M. The Pun1 gene for pungency in pepper encodes a putative acyltransferase. *Plant J.* **2005**, *42*, 675–688. [[CrossRef](#)] [[PubMed](#)]
62. Han, K.; Jang, S.; Lee, J.; Lee, D.; Kwon, J.; Kang, B. A MYB transcription factor is a candidate to control pungency in *Capsicum annuum*. *Theor. Appl. Genet.* **2019**, *132*, 1235–1246. [[CrossRef](#)] [[PubMed](#)]
63. Zhang, Z.; Zhu, Y.; Cao, Y.; Yu, H.; Bai, R.; Zhao, H.; Zhang, B.; Wang, L. Fine mapping of the male fertility restoration gene *CaRf032* in *Capsicum annuum* L. *Theor. Appl. Genet.* **2020**, *133*, 1177–1187. [[CrossRef](#)] [[PubMed](#)]
64. Lv, J.; Liu, Y.; Liu, Z.; Wang, J.; Zou, X. Mapping and identifying candidate genes involved in the novel fasciculate inflorescence in pepper (*Capsicum annuum* L.). *Mol. Breed.* **2019**, *39*, 148. [[CrossRef](#)]
65. Popovskiy, S.; Paran, I. Molecular genetics of the γ locus in pepper: Its relation to capsanthin-capsorubin synthase and to fruit color. *Theor. Appl. Genet.* **2000**, *101*, 86–89. [[CrossRef](#)]

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.