



OPEN ACCESS

EDITED BY

Patricio Hinrichsen,
Agricultural Research Institute, Chile

REVIEWED BY

Harun Bektaş,
Siirt University, Türkiye
Gerardo Tapia,
Agricultural Research Institute, Chile

*CORRESPONDENCE

Tao Li
✉ tianxing84@163.com

†These authors have contributed equally to this work

RECEIVED 08 June 2024

ACCEPTED 08 November 2024

PUBLISHED 03 December 2024

CITATION

Xu W, Gong C, Mai P, Li Z, Sun B and Li T (2024) Genetic diversity and population structure analysis of 418 tomato cultivars based on single nucleotide polymorphism markers. *Front. Plant Sci.* 15:1445734. doi: 10.3389/fpls.2024.1445734

COPYRIGHT

© 2024 Xu, Gong, Mai, Li, Sun and Li. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Genetic diversity and population structure analysis of 418 tomato cultivars based on single nucleotide polymorphism markers

Weijie Xu[†], Chao Gong[†], Peiting Mai, Zhenxing Li, Baojuan Sun and Tao Li*

Guangdong Key Laboratory for New Technology Research of Vegetables, Vegetable Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou, China

Introduction: Tomato (*Solanum lycopersicum*) is a highly valuable fruit crop. However, due to the lack of scientific and accurate variety identification methods and unified national standards, production management is scattered and non-standard, resulting in mixed varieties. This poses considerable difficulties for the cataloging and preservation of germplasm resources as well as the identification, promotion, and application of new tomato varieties.

Methods: To better understand the genetic diversity and population structure of representative tomato varieties, we collected 418 tomato varieties from the past 20 years and analyzed them using genome-wide single nucleotide polymorphism (SNP) markers. We initially assessed the population structure, genetic relationships, and genetic profiles of the 418 tomato germplasm resources utilizing simplified genome sequencing techniques. A total of 3,374,929 filtered SNPs were obtained and distributed across 12 chromosomes. Based on these SNP loci, the 418 tomatoes samples were divided into six subgroups.

Results: The population structure and genetic relationships among existing tomato germplasm resources were determined using principal component analysis, population structure analysis, and phylogenetic tree analysis. Rigorous selection criteria identified 15 additional high-quality DNA fingerprints from 50 validated SNP loci, effectively enabling the identification of the 418 tomato varieties, which were successfully converted into KASP (Kompetitive Allele Specific PCR) markers.

Discussion: This study represents the first comprehensive investigation assessing the diversity and population structure of a large collection of tomato varieties. Overall, it marks a considerable advancement in understanding the genetic makeup of tomato populations. The results broadened our understanding of

the diversity, phylogeny, and population structure of tomato germplasm resources. Furthermore, this study provides a scientific basis and reference data for future analysis of genetic diversity, species identification, property rights disputes, and molecular breeding in tomatoes.

KEYWORDS

genetic relationship, population structure, SNPs, DNA fingerprint, SNP markers

1 Introduction

Tomato (*Solanum lycopersicum*) is one of the most valuable vegetable crops worldwide. It is widely cultivated on all continents except Antarctica due to its good adaptability (Liu et al., 2023; Wang et al., 2023). However, modern tomato cultivars, similar to other crops, are facing problems caused by a narrowed genetic backgrounds. These issues are aggravated by climate change and frequent extreme weather events (Wang et al., 2023). Thus, it is necessary to collect and utilize more tomato germplasms to meet the future requirements of tomato breeding. Tomato landraces, as with most crops, are less inbred and have more genetic diversity than modern varieties. They have a long planting history and can adapt well to certain locations by harboring many biotic and abiotic stress-resistance genes. However, mistakes in naming landraces and duplications of genotypes have often occurred during the collection of landrace resources as landraces are typically maintained by local farmers over generations. Therefore, it is necessary to know the genetic backgrounds of these landraces for name corrections, duplicate removals, conservation, and tomato improvements. Although the genetic diversity of tomato landraces has previously been analyzed using simple sequence repeat (SSR) and SNP markers, numerous genotypes are still not well differentiated due to the lack of available markers or low polymorphism (Wang et al., 2023).

DNA molecular marker technology is extensively employed for the identification of plant varieties due to notable advantages such as independence from environmental conditions and species limitations, ease of operation, consistent and comprehensive locus identification, and considerable polymorphism (Yao et al., 2020; Liu, 2022). Owing to its unique advantages, molecular marker technology is widely applied in the research of forest plant species. In the field of plant biology, SSR and single nucleotide polymorphism (SNP) markers are utilized for the construction of fingerprint databases (Manivannan et al., 2021), the identification of genetic relationships, the development of genetic maps, and the isolation and cloning of diverse genes in various species including *Brassica oleracea* (Shen et al., 2021), *Triticum aestivum* (Li et al., 2018), *Camellia sinensis* (Karunaratna et al., 2021), and *Ipomoea batatas* (Wang et al., 2010). Previous research conducted genetic detection on 39 honeysuckle germplasm resources using eight pairs of high-quality SNP fluorescence markers (Liu, 2022). The development of SNP

markers exhibits a higher abundance and greater degree of polymorphism. In addition, SNP markers are also easy to detect and statistically analyze, and can be identified using high-throughput automatic detection methods, which can greatly reduce the investment of manpower and material resources (Pei et al., 2015; Hata et al., 2023). SNPs display high-density characteristics and demonstrate remarkable stability within the genome. This facilitates their seamless integration into automated genotyping methods applicable to the entire genome (Gong et al., 2016; Xu et al., 2023; Yamaguchi et al., 2023). Numerous other high-throughput, low-cost SNP genotyping platforms have also been developed and widely utilized in important crop species such as *Triticum aestivum*, *Zea mays*, *Glycine max*, and *Vigna unguiculata* (Allen et al., 2011; Hiremath et al., 2012; Wang et al., 2021a; Wang et al., 2021b).

The KASP (Competitive Allele Specific PCR) system, developed by KBiosciences, exhibits superior characteristics compared to alternative methods for SNP genotyping. It boasts remarkable precision, excellent versatility in terms of target sites, and cost-effectiveness. Most importantly, it has been demonstrated as suitable for accurately identifying SNP sites across a notable quantity of samples (Yang et al., 2018). SNP markers are important in the analysis of crop genetic diversity and the establishment of linkage maps. They are also capable of achieving comprehensive high-throughput detection typing throughout the entire process (Song et al., 2015). At present, KASP has been applied to many species such as *Triticum aestivum*, *Malvaceae*, and *Oryza sativa* (Trick et al., 2012; Byers et al., 2012; Yang et al., 2019).

Wild tomatoes have adapted to a variety of ecological environments in Asia and western South America, including the Galapagos Islands. They can be found at altitudes ranging from near sea level to 3,600 m and display a wide range of genetic and phenotypic diversity (Peralta et al., 2008; Tieman et al., 2017; Tieman et al., 2017). SNP molecular markers provide many advantages in tomato breeding, including improving extensive marker variation, standardizing variation data, clear marker position delineation, and mature technical methods (Val et al., 2020; Liu, 2022). We employed the markers recommended for constructing DNA fingerprint databases in the International Union for the Protection of New Plant Varieties Biological and Molecular Marker Technology Molecular Testing Guidelines and the General Principles of DNA Fingerprinting Methods for Plant Variety

Identification in China (NY/T 2594-2016) (Hayward et al., 2015; Fengge et al., 2018; Li et al., 2020). Therefore, it is crucial to effectively utilize the inherent diversity found in wild germplasm for the improvement of the genetic composition of tomatoes. To accomplish this objective, a thorough investigation into the extent of genetic diversity and population structure within tomato species is essential for genetic enhancement and preservation. However, previous research has raised concerns regarding the potential underestimation of tomato species diversity due to a lack of representativeness or a limited number of samples. There is an urgent need to investigate the molecular diversity present in tomatoes to enhance their productivity and quality. In our study, we analyzed the population composition, genetic affiliations, and genetic profiles of the 418 tomato germplasm resources from several regions worldwide. This was achieved using a simplified method of genome sequencing technology. High-quality DNA fingerprints were extracted from the validated 50 SNP loci and were confirmed to effectively identify the 418 tomato varieties using KASP platforms. These fingerprints have been gradually employed for the authentication of new cultivars, seed purity, and variety genuineness. We then analyzed the population structure and identified varieties within the tomato germplasm collection. The findings can significantly enhance future investigations in the fields of genetic diversity, species identification, property rights conflicts, and molecular breeding. This can be achieved by providing a scholarly foundation and benchmark data for these endeavors.

2 Materials and methods

2.1 DNA extraction and library construction

The tomato cultivar (*Solanum pimpinellifolium* (PIM), *Solanum lycopersicum* var. *cerasiforme* (CER), and *Solanum lycopersicum* (BIG)) was utilized to construct full-length DNA libraries extracted from leaves, as per previous research (Gramazio et al., 2020). Plants were grown in the field according to a randomized complete block design in Zhongluotan (23°23 N, 113°26 E), Guangzhou City, Guangdong Province, China. Each line consisted of three replicated blocks. After grinding the leaf samples, DNA was extracted using an RNeasy Kit (Qiagen) following the manufacturer's instructions. Following purification and DNA digestion with RNase-free DNase (Qiagen), DNA quality was assessed using a NanoDrop ND-2000 (A260/A280 1.9–2.1) and an Agilent 2100 Bioanalyzer (28S/18S 1.8–2.0). A total of 3 µg of DNA per sample was utilized for the library construction on the Illumina HiSeq 4000 platform as per the manufacturer's instructions and index codes. Library construction and paired-end sequencing were subsequently performed (Hanania et al., 2004; Trapnell et al., 2014).

2.2 Simplified genome sequencing and data analysis

Library construction and paired-end sequencing were performed by Beijing Biomarker Technologies. Subsequently, data

quality control was conducted on the original sequenced reads (Giovannoni, 2004; Dellaporta et al., 2016). The Burrows–Wheeler Aligner (BWA, v.0.7.17-R1188) was utilized to align the sequenced reads with the reference genome (Byers et al., 2012). The reference tomato genome was the 'Heinz 1706' genome (*S. lycopersicum*) (https://solgenomics.net/ftp/genomes/Solanum_lycopersicum/Heinz1706/assembly/build_4.00/), and the parameters were configured as -M -R. The insert size and coverage depth of each sample were assessed, and variations were identified by comparing the alignment positions of clean reads on the reference genome (Sato et al., 2012). The files generated from the comparison were then converted to bam format using Samtools (v.1.9; <https://samtools.sourceforge.net/>). Following this, Picard MarkDuplicates (v.2.21.2; <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM5970534>) was utilized to identify and remove duplicate tags, ensuring that only high-quality reads were retained for subsequent analysis.

2.3 Discovery and genotyping of SNP

To ensure the acquisition of high-quality SNPs, we employed a unique that included mapped paired-end reads for the SNP calling process. The genotype likelihood parameter of the genomic site for each individual was calculated using SOAPsnp (<https://sourceforge.net/projects/soapsnp/>) with default parameters (Gu et al., 2013). To validate the accuracy of our SNP calling results, we compared them with SNP data from a previous study involving 120 individuals and focusing on 10 candidate genes obtained through genome re-sequencing (Du et al., 2014). The SNP data in the previous study was determined using PCR-Sanger sequencing. The relative accuracy rate of our SNP calling exceeded 99%, demonstrating the high quality and reliability of our platform for SNP calling.

2.4 Genetic diversity and population structure analysis

Based on the high-quality SNPs obtained, we utilized VCFtools v.0.1.16 (<https://vcftools.sourceforge.net/>) to conduct principal component analysis (PCA) and other population divergence statistics (Petr et al., 2011; Abdelhafez et al., 2019) with default parameters. Subsequently, an evolutionary tree of the 418 tomato samples was constructed using the maximum likelihood (ML) method in FastTree (v.2.1.9). The genetic structure of the population was analyzed using PLINK1.9 (Sorkheh et al., 2007; Chang et al., 2015). PLINK1.9 and VCFtools64 v.0.1.16 were used to perform PCA and other population divergence statistics, including nucleotide diversity and genetic differentiation (FST).

2.5 Phylogenome analysis

The high-quality SNPs identified above underwent a second round of filtering to enhance the accuracy and efficiency of phylogenetic analysis as per previous research (Zhang et al., 2021).

Phylogenomic analysis was conducted by inferring orthologs and orthogroups using OrthoFinder (v.2.5.4) with the default settings and an activation of ‘-M msa’ (Emms and Kelly, 2018). The longest predicted protein of each individual gene was used as the representative input for the OrthoFinder analysis. TrimAl (v.1.4.12) was used to remove poorly aligned regions of protein multiple sequence alignments (Capella-Gutiérrez et al., 2009). RAXML (v.8.2.12) was employed to construct ML phylogenetic trees using the GAMMAJTT model (Stamatakis, 2014; Silvestro et al., 2012).

2.6 GO enrichment analysis

Candidate genes were analyzed using the AmiGO Term Enrichment tool (<http://amigo.geneontology.org/>) for GO analysis. The R (R Core Team) module CLUSTERPROFILER, available at Bioconductor (<http://bioconductor.org>), was utilized to identify enriched GO terms associated with candidate genes. We applied multiple testing corrections using the Benjamini and Hochberg FDR method (Green et al., 2007). GO terms with a corrected $P < 0.05$ were deemed significantly enriched.

2.7 Fingerprint construction and data analysis

The construction of DNA fingerprinting is achieved by utilizing high-quality SNPs (Xu et al., 2017; Obayes et al., 2024). To achieve simplicity, efficiency, and cost-effectiveness, we aimed to utilize the minimum number of markers possible, allowing us to identify the maximum number of varieties (Wang et al., 2009). Based on the size and frequency distribution of polymorphic information content (PIC) values, we meticulously selected 50 core markers of SNP sites with a high detection rate and considerable polymorphism. The obtained SNP loci were screened based on the following criteria: (1) they were evenly distributed across the different chromosomes; (2) no genotype data was missing; (3) the number of unmeasured materials at the site was < 20 and the minor allele frequency (MAF) was ≥ 0.34 ; (4) the PIC was > 0.35 ; (5) the Hardy–Weinberg equilibrium (HWE)-tested p -value was ≥ 0.01 ; and (6) there were no other mutations within 100 bp before and after the polymorphic site. These markers had the capability to distinguish all varieties and a DNA fingerprint was subsequently constructed. Furthermore, the 15 most efficient SNP combinations were further screened to identify tomato varieties with reduced costs and faster growth rates (Hiremath et al., 2012; Jones and Mackay, 2015).

2.8 Development and analysis of KASP markers

For each SNP site retained during the screening process, the surrounding sequences on both sides of the SNP were trimmed to a size of 100 bp. The KASP primers utilized in this study were developed and designed by LGC Genomics LLC (Beverly, MA,

USA). Here, we used Primer 5.0 (<https://primer-premier-5.software.informer.com/>) to develop PCR primers for the competitive allele-specific PCR-single nucleotide polymorphism (KASP-SNP) linked markers. PCR was performed on parental lines to denote the polymorphic efficacy of the designed primers. Based on this preliminary PCR analysis, the polymorphic markers were used to screen the bulks and backcross population. The sequences of primers are listed in [Supplementary Table S1](#).

3 Results

3.1 Genome sequencing and alignment to the tomato reference genome

First, we constructed high-quality DNA fingerprints of 418 tomato samples from around the world from the past 20 years using high-quality SNPs ([Supplementary Figure S1](#)). The acquisition of this high-quality, clean data provides a robust foundation for our work ([Supplementary Table S1](#)). The original raw Illumina reads obtained through GBS (genotyping by sequencing) underwent quality control and data filtering to produce high-quality clean data, which serves as the foundation of this study. To obtain a high-quality genome for the 418 tomatoes, the sequencing of the tomato samples on the Illumina NovaSeq 6000 PE150 yielded a total of 196,612,427 PE reads (55.44 Gb) of clean data ([Figures 1A, B](#) and [Supplementary Table S1](#)). This is consistent with the estimated genome size obtained from the Illumina short read sequencing data. The average coverage depth of the loci reached 492.87, and the obtained clean reads were mapped to the ‘Heinz 1706’ reference genome with a mapping efficiency of 93.25%. The genotype deletion rate was determined as 4.71% ([Supplementary Table S1](#)). This assembly also demonstrated high quality compared to other organisms with sequenced reference genomes in previous studies (<https://www.biorxiv.org/content/10.1101/2021.05.04.441887v1>).

3.2 Identification and annotation of high-quality SNPs in the 418 tomato samples

After completing the sequencing, the Genome Analysis Toolkit (GATK; <https://gatk.broadinstitute.org/hc/en-us>) was used to identify a notable number of SNP variants in the 418 tomato samples. The identified SNPs were filtered, resulting in a total of 13,143,268 SNPs generated during the sequencing process ([Figure 2](#)). Furthermore, a total of 3,374,929 filtered SNPs were obtained and distributed across 12 chromosomes in the tomato genome ([Figure 2B](#)). Notably, the distribution of SNPs on each chromosome was depicted based on the number and density of SNPs. There were approximately 400,000 SNPs on chromosome 4 and 600,000 SNPs on chromosome 10 ([Figures 2A, B](#)). Chromosomes 3–9 exhibited a similar number of SNPs, with approximately 300,000 SNPs on each chromosome.

The analysis of these SNP-predicted mutation types revealed that, out of the six possible single base mutations, C/T and A/G transitions were the most prevalent, accounting for 26.43% and

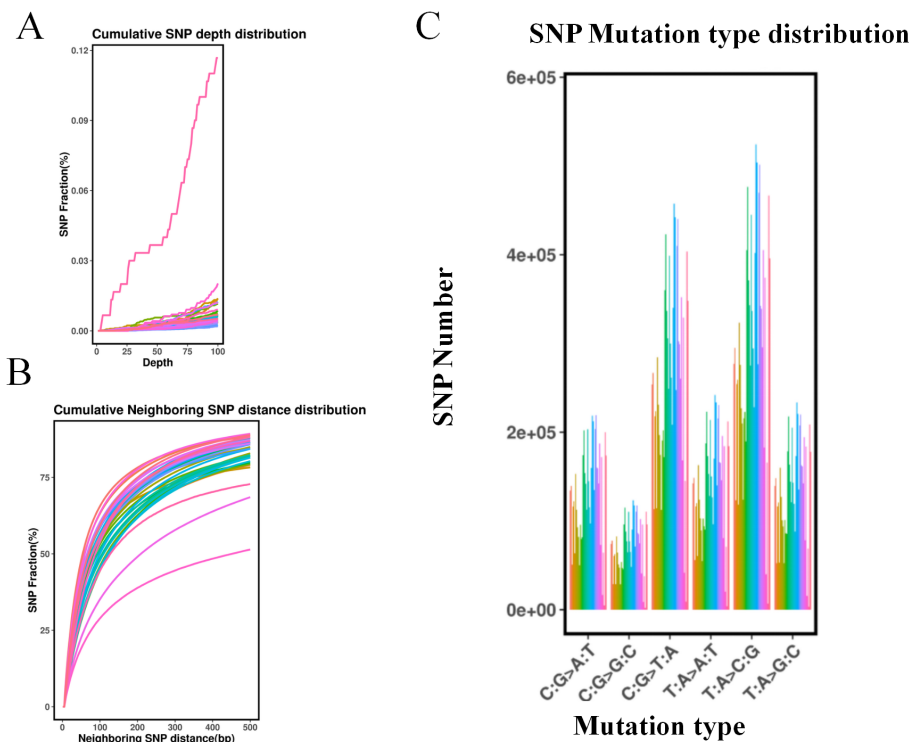


FIGURE 1 Single nucleotide polymorphism (SNP) type and distribution identification of 418 samples. **(A)** Cumulative SNP depth distribution. The horizontal axis represents the depth, and the vertical axis represents the fraction of SNPs. **(B)** Cumulative neighboring SNP distance distribution. The horizontal axis represents the neighboring SNP distance(bp) and the vertical axis represents the fraction of SNPs. **(C)** SNP mutation type distribution on each chromosome. The horizontal axis represents the mutation type, and the vertical axis represents the SNP number.

28.51% of the total, respectively. Further analysis of the distribution of the SNPs in the genome revealed that 70.29% were located in intergenic regions, 6.85% were located in introns, 8.59% were located in the 1-kb region downstream of the transcription termination site, 0.05% were simultaneously located in the 1-kb

region upstream of one gene and in the 1-kb region downstream of another gene, and 0.03 and 0.02% were located in the 5' and 3' UTRs (Supplementary Figure S2). Figure 1C presents the mutation type distribution of the SNPs in the genomic, where T:A replaces C: G, which is the most common type of variation. A total of 94,457

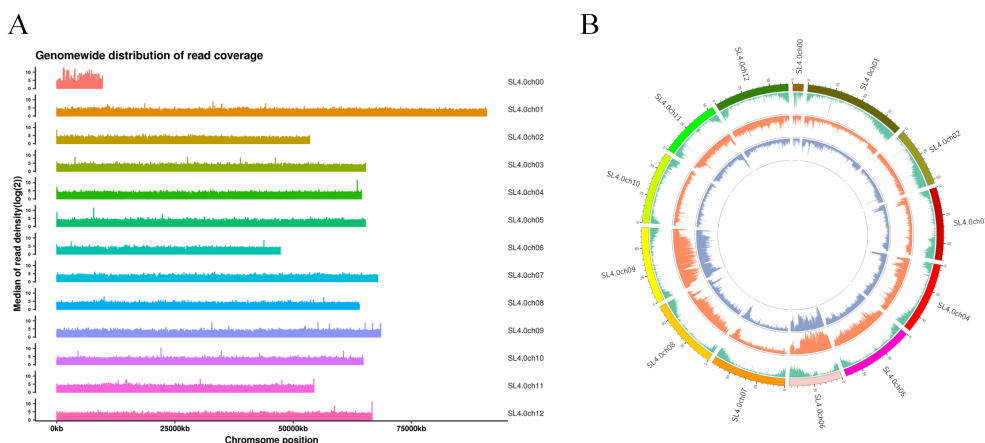


FIGURE 2 Single nucleotide polymorphism (SNP) type and distribution identification of 418 samples. **(A)** Number of SNPs on each chromosome. The horizontal axis represents the chromosome number, and the vertical axis represents the number of SNPs. **(B)** SNP density distribution on each chromosome. The horizontal axis represents the chromosome length, and the vertical axis represents the chromosome number. Different colors represent the number of SNPs in different regions.

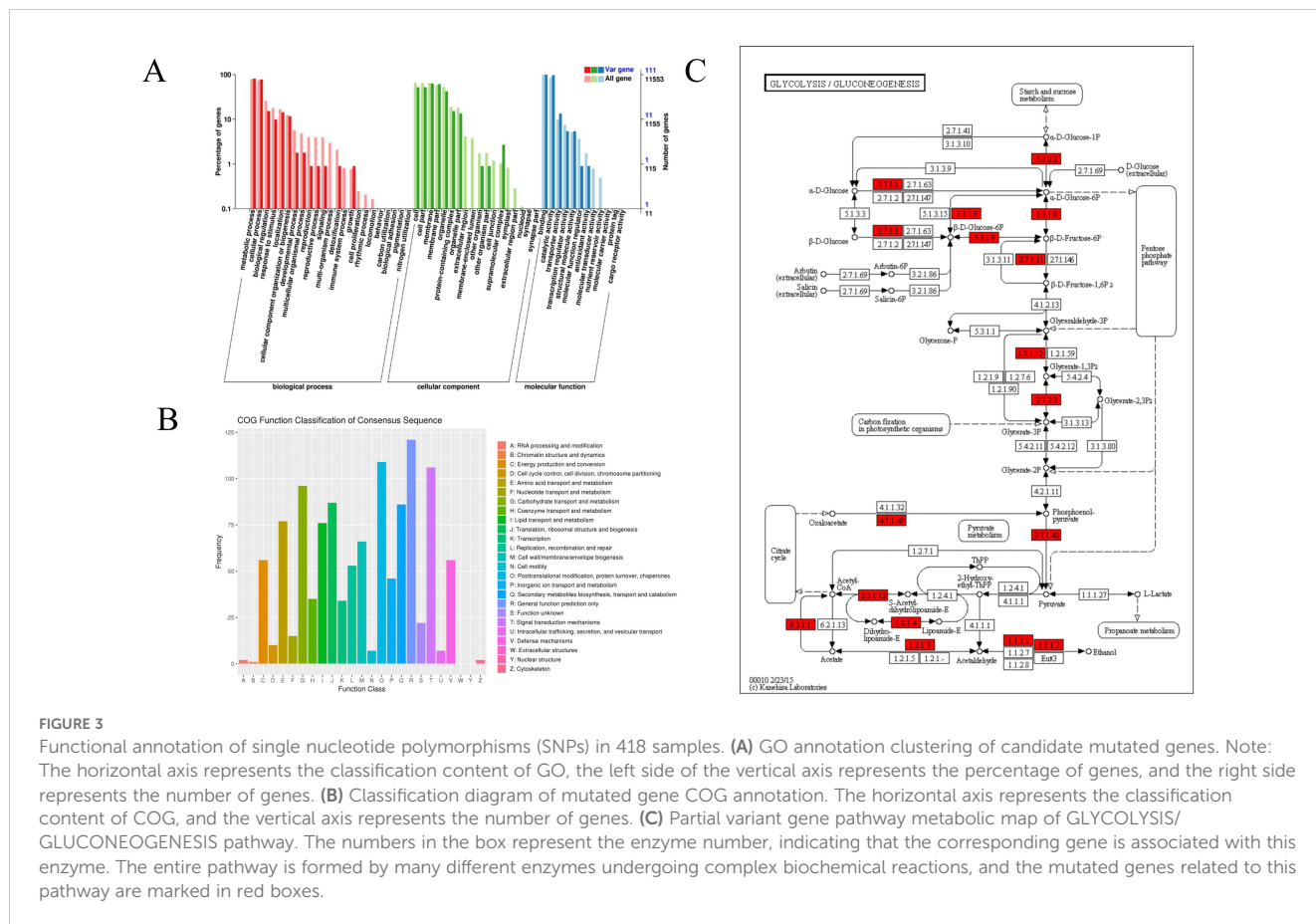
non-synonymous single nucleotide variants (SNVs) were predicted to cause changes in coding amino acids, while 62,439 synonymous SNVs were predicted not to alter amino acid sequences. The ratio of non-synonymous SNVs to synonymous SNVs was 1.51.

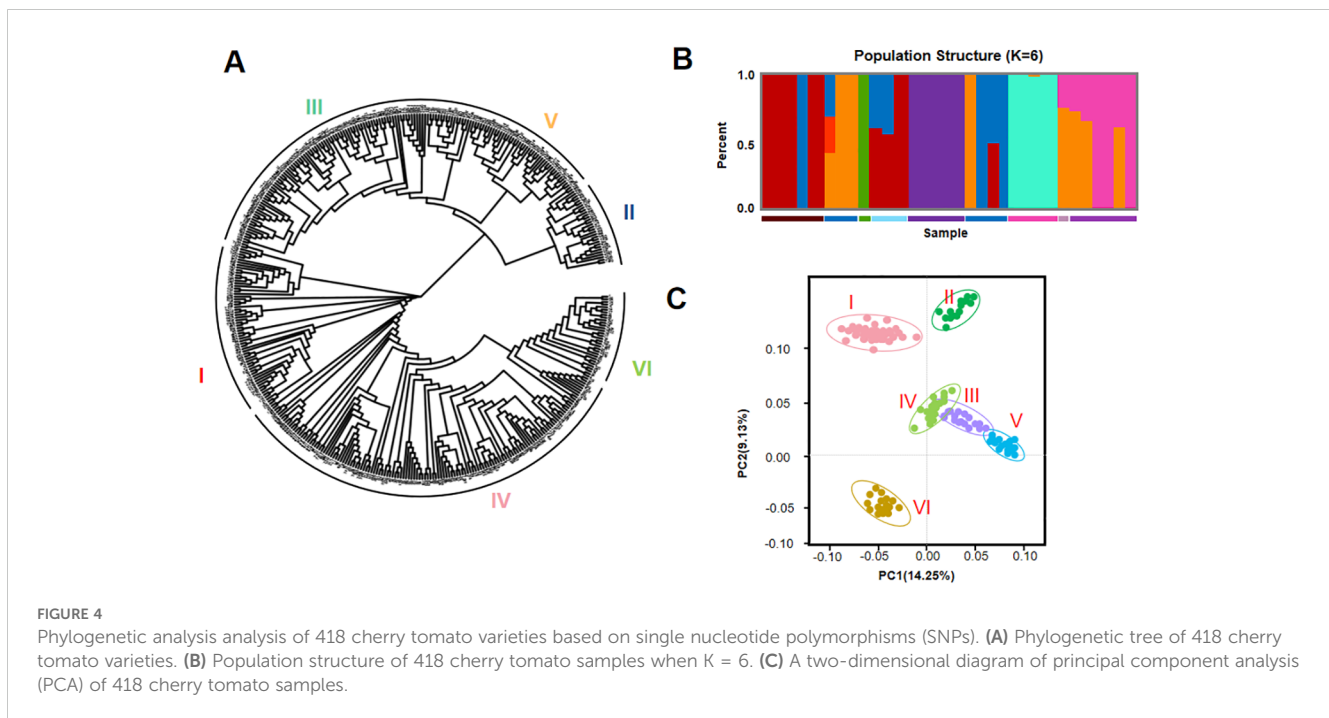
By employing a range of strategies, such as *de novo* prediction, homology-based methods, and transcriptome alignment, we identified and annotated genes. A total of 6,001 protein-coding genes were predicted in SNP, with 98.8% of these predicted genes exhibiting homologues in at least one of the six functional protein databases. Furthermore, a total of 5,910 genes were distributed across different chromosomes, representing approximately 90.14% of the total predicted genes. Among them, 28.4% of SNP sites caused changes in amino acids, including 1,678 candidate genes found in our study. To further understand the potential impact of mutation types between different varieties on the phenotype, the GO annotation of 1,678 candidate genes for several enriched biological pathways, including “RNA processing and modification,” “RNA processing and modification,” “basal transcription factor,” “starch and sucrose regulatory metabolism,” “ubiquitin-mediated proteolysis,” “terpenoid backbone biosynthesis,” “pentose and glucuronate inter conversions,” “glycolipid metabolism,” “flavonoid biosynthesis,” “fatty acid biosynthesis,” “elongation and degradation,” and “ascorbate and aldarate metabolism” (Figure 3). This suggests that a potential mechanism to overcome deleterious mutations occurred in important genes related to basic biological functions. For example, the *SRC2* gene encodes an activator of a calcium-

dependent pathway that mediates reactive oxygen species production in response to cold stress, indicating the adaptation of various samples to different selections, climates, and environments.

3.3 Genetic relationships and population structure analysis

The phylogenetic tree showed that the 418 tomato varieties clustered into six groups, denoted as clades I, II, III, IV, V, and VI (Figures 4A, B). This classification is consistent with the variations in the five morphological traits of the row-type and adherence of the hulls. Each clade contained 39–128 members. Interestingly, three clades (I, III, and VI) were present in all continents, with high supporting values. In clade II, only 39 varieties were observed, suggesting this clade was lost during transition, selection, and breeding. Notably, we observed an extensive diversity of clade IV with most varieties. PCA was also utilized on the sequences of the high-quality SNPs identified in the 418 tomato accessions. The two-dimensional graph in Figure 4 illustrates the values of each sample in the first (PC1) and second (PC2) principal components. The values in parentheses on the axis labels indicate the percentage of the total variation explained by PC1 and PC2, which were 14.25% and 9.13%, respectively. The geographical origins of the varieties and the PCA results jointly reveal that each of the four groups contained tomato accessions originating from different geographic regions (Figure 3).





This study demonstrates that there is no considerable correlation between the PCA results and the geographical origin of the germplasm. In addition, we employed Plink and Frappe to analyze the population structure of the 418 tomato accessions. The determination of the number of clusters is typically based on the cross-validation error rate, with the optimal number identified as that with the lowest cross-validation error rate. As illustrated in Figure 3, the cross-validation error rate is minimized when $K = 6$, indicating that the 418 tomato accessions can be classified into six clusters. This is consistent with the phylogenetic analysis results. In addition, the test materials were not fully categorized based on their geographical origins, which aligns with the classification derived from the PCA. The neighbor-joining method in Tree Best was utilized to construct a phylogenetic tree for the 418 tomato accessions. The first three axes of the PCA further confirmed this population structure but showed more divergence between subgroups I and VI (Figure 4C). The remaining accessions did not cluster solely based on geographical origin. The other clades contained tomato accessions from different geographical regions, and those originating from the same region were distributed across all five clades. However, note that while these tomato accessions from the same source are complex and diverse in terms of genetic backgrounds, their clustering within the clade indicates interrelatedness and independence among them.

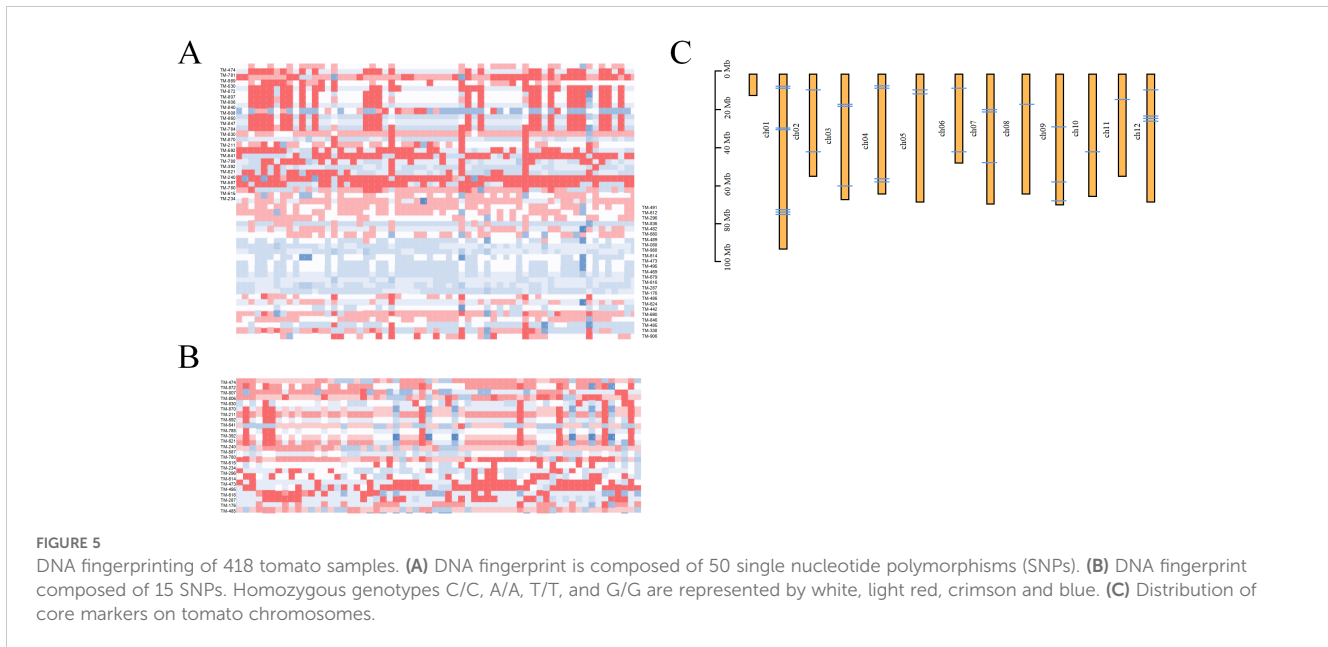
3.4 Design of KASP primers and selection of core SNP sites

SNP sites containing other mutations within 100 bp before and after the SNP were excluded. This resulted in a total of 737 remaining sites in the 418 tomato varieties. The flanking sequences of these sites were then extracted from the corresponding scaffold and aligned with the tomato reference genome at the chromosome level. Only the sites

located on the chromosomes were retained, leading to a final selection of 71 SNP sites for the subsequent experiment. The 715 SNP sites demonstrated considerable polymorphism and a high degree of variety discrimination. The MAF and PIC values for these sites ranged from 0.346 to 0.500, and 0.350 to 0.375, with average values of 0.45 and 0.37, respectively. Notably, the percentage of PIC values falling within the range of 0.370–0.375 was found to be at 70.6%, indicating a strong level of polymorphism for these markers. The observed heterozygosity for the 715 variable SNP sites ranged from 0 to 0.64, with an average of 0.10. The genetic diversity within the germplasm collection was assessed and found to range from 0.45 to 0.50, with an average of 0.49. The combined MAF, PIC, and observed heterozygosity values were utilized for further screening of the SNP sites distributed across the 12 chromosomes. These sites did not have missing genotypic data, were predominantly homozygous variant sites, and were detected in as many individuals as possible. This screening yielded a total of 163 SNP sites. Subsequently, KASP primers were designed for these identified SNP loci, with successful conversion into KASP markers achieved for all 73 primer sets. The KASP markers were then typed and verified using GBS-genotyped tomato accessions ($n = 418$). Based on the typing results, we identified 46 KASP markers that demonstrate high accuracy (Supplementary Figure S3). These markers were selected as candidate markers for screening verification populations. Among them, those with superior performance were designated as core markers and utilized as candidate core primers for further analysis.

3.5 Fingerprint construction and verification of varieties

Based on the simplified sequencing results, SNP loci with a PIC value greater than 0.30—which were uniformly distributed on



different chromosomes—were selected as the core loci for constructing the fingerprint of the tomato varieties in our study (Figure 5). Finally, we screened 50 high-quality SNP loci and constructed a fingerprint, successfully distinguishing the 418 tomato varieties. Among the 50 core SNPs, the PIC values for site markers 8, 17, 23, and 32 were maximized at 0.38, while the PIC values for Marker 4 were minimized at 0.30. The average PIC value was calculated as 0.34, indicating moderate polymorphism. Paired comparisons were conducted on the 418 tomato samples using this developed core SNP locus combination. Statistical analysis reveals considerable differences in the number of different loci among the 418 tomato samples, and different loci were observed in each pair of samples, demonstrating the high reliability of this locus combination. This indicates a substantial diversity in the genetic markers present within the different samples, highlighting the complexity and richness of the genetic information being studied. This study aimed to efficiently and cost-effectively differentiate tomato varieties. We focused on identifying the minimum number of SNPs required to distinguish 418 tomato varieties from the 50 core loci obtained. As a result, we obtained the simplest SNP combination, which is evenly distributed on 12 chromosomes (Figure 5). When utilizing a combination of 15 SNPs for sample differentiation, we determined at least one SNP variation between the two samples. This allows for effective discrimination among the 418 tomato varieties. Note that certain tomato samples could only be differentiated based on a single SNP. For example, “C20F31” and “C20F02” could only be distinguished by Marker 7, while “003” and “065” (SPL004) could only be distinguished by Marker 6. These results also indicate that the development of fingerprint maps is of great significance in the complex and chaotic genetic relationship background of tomato germplasm.

4 Discussion

4.1 Analysis of SNP-based genetic relationships among the 418 tomato germplasm reveals considerable polymorphism

Much horticultural research has been conducted on tomatoes due to their ease of cultivation, large number of plants, and wide planting area (Dargiri et al., 2021). In addition, a comprehensive overview of the variations in medicinal history, sources, characteristics, chemical composition, and other aspects of tomatoes has been provided (Tiemann et al., 2017). However, although a wide variety of tomato cultivars have been extensively grown in previous studies, there is a lack of standardized and decentralized management. Further investigation is necessary to elucidate the genetic relationships and population structure of existing tomato varieties.

In this study, 418 tomato germplasm resources from around the world were collected and grouped into six subgroups using simplified sequencing, PCA, population structure analysis, and evolutionary tree construction. The results show similar outcomes. Subgroup IV comprises tomato cultivars that demonstrate high yields, robust resistance to pests and diseases, and desirable traits currently valued in the market. These findings emphasize the complex nature of wild tomato species lineages, in contrast to the relatively simpler lineages observed in systematically bred tomatoes (Liu, 2022). In the population structure analysis, all individuals displayed ancestral lines from the four subgroups. However, the genetic relationship between “C20F02” and “C20F07” was found to be direct and derived from a single pedigree. This discovery effectively explains their placement

within subgroup IV on the evolutionary tree. Furthermore, four additional hybrids were included in this study. All hybrids inherited genetic traits from their parents, which aligns largely with expectations. Currently, our research group is conducting a more comprehensive analysis of hybrid characteristics.

Many genes previously reported to be involved in climate adaptation were identified here with variants to be associated with environmental variables, a significant functional enrichment could be detected in our study, the GO annotation of 1678 candidate genes about several enriched biological pathways were significantly shown in 'basal transcription factor', 'starch and sucrose regulatory metabolism', 'ubiquitin-mediated proteolysis', 'terpenoid backbone biosynthesis', 'pentose and glucuronate inter conversions', 'fatty acid biosynthesis', 'elongation and degradation', as well as 'ascorbate and aldarate metabolism'. The variants of these genes also related to precipitation showed similar geographic distribution in frequencies. For example, this transcription factor with a lateral organ boundary (LOB) domain plays an essential role in root development in response to flooding and drought stresses. The SRC2 gene encodes an activator of a calcium-dependent pathway that mediates reactive oxygen species production in response to cold stress.

4.2 Efficient and convenient SNP fingerprints for the identification of the 418 tomato germplasm resources

Unlike the majority of previous studies that concentrate on tomato morphology and quality assessments (Zhang et al., 2011; Beckles et al., 2012; Li et al., 2022), the current study aimed to analyze the population structure of tomatoes, explore the differences among tomato varieties, and utilize SNP analysis to distinguish between different tomato varieties. This study not only provides a clear explanation of the population relationships among the existing 418 tomato germplasm resources but also offers further insights into the genetic relationships between different tomato varieties. These findings serve as valuable references and a foundation for the management of tomato germplasm resources, holding considerable importance for the future breeding of new tomato varieties. This research contributes to a comprehensive understanding of the genetic diversity within tomato germplasm resources and provides essential knowledge for their effective utilization in breeding programs. The implications of these findings are far-reaching, with the potential to considerably impact the development of improved tomato varieties with desirable traits (Muir et al., 2014). In our study, we utilized SNP molecular marker technology and DNA fingerprinting to obtain 196,612,427 SNPs from the 418 tomato samples. Among these, 50 core SNPs demonstrated strong identification ability for tomato germplasm. Furthermore, specific DNA fingerprinting analysis was conducted and developed. The fingerprint map, which was created using 50 core SNP markers, effectively distinguishes between different species and varieties of tomatoes. This study also identified 15 SNPs as the most compact combination of sites and the DNA fingerprints constructed based on the loci of these 15 SNP loci were utilized for the detection of various tomato samples.

At least one distinct locus was identified in each variety of tomato, allowing for the rapid, accurate, and cost-effective differentiation of different species and varieties (Elbasyoni et al., 2018). Given the highly complex genetic background of tomato germplasm, this study conducted further research on varieties that exhibited several different loci in the fingerprints of the 15 SNP loci. After further analysis, it was found that these varieties showed high repeatability in DNA fingerprints and similarity in origin, species, and characteristics. Thus, when the DNA fingerprint difference between two varieties is ≤ 1 locus, the varieties share similarities in species, origin, parentage, and flower color. This indicates that the two varieties are likely to be either identical or highly similar.

4.3 Reliability and stability of the 15 tomato SNP core fingerprints

To evaluate the reliability of the 15 SNP loci DNA fingerprints, four samples were randomly selected from the initial 418 germplasm resources. PCR and Sanger sequencing were used for the analysis. The results of the SNP verification were consistent with previous findings, demonstrating reliability and stability with previous research (Liu, 2022). A remarkably high genetic stability was observed among these samples, with a consistency rate of 97.33%. This level of stability is consistent with the typical characteristics exhibited by tomatoes. However, it is important to note that the genotypes of samples "C20F31," "C20F19," "C20F15," and "022," underwent a change from A/A to A/G in Marker 10. This result may be attributed to the presence of a second allele of the SNP. In summary, the validation results of the 15 SNP loci and the genetic stability of tomato were as expected. The analysis confirmed the authenticity and reliability of the DNA fingerprint composed of these loci, and also verified the high genetic stability of tomato. This provides valuable data and theoretical support for tomato germplasm identification and SNP fingerprint methods.

This research provides new insights into the genetic diversity of tomato varieties, domestically and internationally, over the past 20 years. It also reveals important information about the genetic background of domestic tomato varieties. The findings of this study have major implications for basic research in tomato breeding, the protection of new plant varieties (PVPs), and genetic diversity.

Data availability statement

Raw data are available at the BIGD Genome Sequence Archive (<https://bigd.big.ac.cn>) under accession number CRA007031.

Author contributions

XW: Writing – original draft, Writing – review & editing, CG: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft. MP: Conceptualization, Data curation,

Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft. LZ: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft. BS: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft. TL: Writing – original draft, Writing – review & editing, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The present study was financially supported by National Key Research and Development Program of China (2023YFD1201504), Key-Area Research and Development Program of Guangdong Province (2020B020220001, 2020B0202090002), Natural Science Foundation of China (32372072), the Department of agriculture and rural areas of Guangdong province of China (2022-NJS-00-005, 2023-NJS-00-003, 2022-NPY-00-025, 2024-NPY-00-021 2023KJ110, 2023KJ106) and

the Special fund for scientific innovation strategy-construction of high leveled Academy of Agriculture Science (R2020PY-JX003, R2023PY-JX008).

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2024.1445734/full#supplementary-material>

References

- Abdelhafez, O. H., Othman, E. M., Fahim, J. R., Desoukey, S. Y., and Abdelmohsen, U. R. (2019). Metabolomics analysis and biological investigation of three malvaceae plants. *Phytochem. Anal. PCA* 31, 204–214. doi: 10.1002/pca.2883
- Allen, A. M., Barker, G. L. A., and Berry, S. T. (2011). Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (*Triticum aestivum* L.). *Plant Biotechnol. J.* 9, 1086–1099. doi: 10.1111/j.1467-7652.2011.00628.x
- Beckles, D. M., Hong, N., Stamova, L., and Luengwilai, K. (2012). Biochemical factors contributing to tomato fruit sugar content: a review. *Fruits* 1. doi: 10.1051/FRUITS/2011066
- Byers, R. L., Harker, D. B., Yourstone, S. M., Maughan, P. J., and Udall, J. A. (2012). Development and mapping of SNP assays in allotetraploid cotton. *Theor. Appl. Genet.* 124, 1201–1214. doi: 10.1007/s00122-011-1780-8
- Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348
- Chang, C. C., Chow, C. C., Cam, T. L., Shashaank, V., Purcell, S. M., and Lee, J. J. (2015). Second-generation plink: rising to the challenge of larger and richer datasets. *GigaScience* 4, 1 (2015-02-25), 7. doi: 10.1186/s13742-015-0047-8
- Dargiri, S. A., Samsampour, D., Seyahooei, M. A., and Bagheri, A. (2021). The effect of bacterial endophyte (*exigobacterium aurantiacum*) isolated from *salsola imbricata* on growth characteristics of tomato seedlings under salinity stress. *J. Hortic. Sci.* 35, 153–167. doi: 10.22067/jhs.2021.61963.0
- Dellaporta, S., Fragoso, C., Heffelfinger, C., and Morena, M. (2016). Flexible and scalable genotyping-by-sequencing strategies for population studies. *BMC Genomics* 15, 979. doi: 10.1186/1471-2164-15-979
- Du, Q., Xu, B., Gong, C., Yang, X., Wei, P., Tian, J., et al. (2014). Variation in growth, leaf, and wood property traits of chinese white poplar (*Populus tomentosa*), a major industrial tree species in northern china. *Can. J. Forest. Res.* 44, 326–339. doi: 10.1139/cjfr-2013-0416
- Elbasyoni, I. S., Lorenz, A. J., Guttieri, M., Frels, K., Baenziger, P. S., Poland, J., et al. (2018). A comparison between genotyping-by-sequencing and arraybased scoring of SNPs for genomic prediction accuracy in winter wheat. *Plant Sci.* 270, 123–130. doi: 10.1016/j.plantsci.2018.02.019
- Emms, D. M., and Kelly, S. (2018). Orthofinder2: fast and accurate phylogenomic orthology analysis from gene sequences. *Cold Spring Harbor Lab.* (1). doi: 10.1101/466201
- Fengge, W., Hongli, T., Hongmei, Y., Han, Z., Yongxue, H., Meng, K., et al. (2018). Principle and strategy of dna fingerprint identification of plant variety. *Mol. Plant Breed.* doi: 10.5376/mpb.2019.10.0011
- Giovanoni, J. J. (2004). Genetic regulation of fruit development and ripening. *Plant Cell* 16, S170–S180. doi: 10.1105/tpc.019158
- Gong, D. P., Huang, L., Xu, X. H., Wang, C. Y., Ren, M., Wang, C. K., et al. (2016). Construction of a high-density SNP genetic map in flue-cured tobacco based on SLAF-seq. *Mol. Breed.* 36, 100. doi: 10.1007/s11032-016-0514-7
- Gramazio, P., Pereira-Dias, L., Vilanova, S., Prohens, J., Soler, S., Esteras, J., et al. (2020). Morphoagronomic characterization and wholegenome resequencing of eight highly diverse wild and weedy *s.pimpinellifolium* and *s.lycopersicum* var. *cerasiforme* accessions used for the first interspecific tomato magic population. *Hortic. Res.* 7, 16. doi: 10.1007/978-3-319-99208-2_9
- Green, G. H., and Diggle, P. J. (2007). On the operational characteristics of the benjamini and hochberg false discovery rate procedure. *Genet. Mol. Biol.* 6, 27.
- Gu, S., Fang, L., and Xu, X. (2013). Using soapaligner for short reads alignment. *Curr. Protoc. Bioinf.* 11, 1–17. doi: 10.1002/0471250953.bi1111s44
- Hanania, U., Velcheva, M., Sahar, N., and Perl, A. (2004). An improved method for isolating high-quality DNA from *Vitis vinifera* nuclei. *Plant Mol. Biol. Rep.* 22, 173–177. doi: 10.1007/BF02772724
- Hata, H., Taniguchi, R., Yamashita, N., Hashiguchi, Y., Nakajima, J., and Takeyama, T. (2023). Genotyping of two congeneric bitterling fish species by nuclear snp markers and the detection of hybridization in a sympatric region. *Ecol. Res.* doi: 10.1111/1440-1703.12387
- Hayward, A. C., Tollenaere, R., Dalton, M. J., and Batley, J. (2015). Molecular marker applications in plants. *Methods Mol. Biol.* 21245, 13–27. doi: 10.1007/978-1-4939-1966-6_2

- Hiremath, P. J., Kumar, A., and Penmetsa, R. V. (2012). Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. *Plant Biotechnol. J.* 10, 716–732. doi: 10.1111/j.1467-7652.2012.00710.x
- Jones, H., and Mackay, I. (2015). Implications of using genomic prediction within a high-density SNP dataset to predict DUS traits in barley. *TAG. Theor. Appl. Genet. Theoretische und angewandte Genetik* 128, 2461–2470. doi: 10.1007/s00122-015-2601-2
- Karunaratna, K. H. T., Mewan, K. M., Weerasena, O.V.D.S.J., Perera, S. A. C. N., and Edirisinghe, E. N. U. (2021). A functional molecular marker for detecting blister blight disease resistance in tea (*Camellia sinensis* L.). *Plant Cell Rep.* 40, 351–359. doi: 10.1007/s00299-020-02637-6
- Li, C. X., Xu, W. G., Guo, R., Zhang, J. Z., Qi, X. L., Hu, L., et al. (2018). Molecular marker assisted breeding and genome composition analysis of zhengmai 7698, an elite winter wheat cultivar. *Rep.* 8, 322. doi: 10.1038/s41598-017-18726-8
- Li, J. J., Yu, X. J., Shan, Q. R., Shi, Z. B., Li, J. H., Zhao, X. T., et al. (2022). Ntegrated volatile metabolomic and transcriptomic analysis provides insights into the regulation of floral scents between two contrasting varieties of *Lonicera japonica*. *Front. Plant Sci.* doi: 10.3389/fpls.2022.989036
- Li, M. D., Ao, D. M., Chen, J., Liu, S., Xu, C. Q., Xu, R., et al. (2020). Research progress on identification and quality evaluation of *Lonicera japonica* thubn. *China Med. Herald* 17, 29–32. doi: 10.4268/cjcm20161309
- Liu, G., Hu, Q., Zhang, J., Li, E., Yang, X., et al. (2023). Genome-wide identification of the plant u-box (PUB) gene family and their global expression analysis in tomato (*Solanum lycopersicum*). *Vegetable Res.* 3, 16. doi: 10.48130/VR-2023-0016
- Liu, Y. J. (2022). The development of agricultural products promotes the revitalization of county economy—a case study of honeysuckle industry in fengqiu county, henan province. *Guangdong Sericulture* 56, 130–132.
- Manivannan, A., Choi, S., Jun, T. H., Yang, E. Y., and Ahn, Y. K. (2021). Genotyping by sequencing-based discovery of snp markers and construction of linkage map from F5 population of pepper with contrasting powdery mildew resistance trait. *BioMed. Res. Int.* 2021, 1–8. doi: 10.1155/2021/6673010
- Muir, C. D., Hangarter, R. P., Moyle, L. C., and Davis, P. A. (2014). Morphological and anatomical determinants of mesophyll conductance in wild relatives of tomato (*Solanum* sect. *Lycopersicon*, sect. *Lycopersicoideae*; solanaceae). *Plant Cell Environ.* 37 (6), 1415–1426. doi: 10.1111/pce.12245
- Obayes, S. K., Enayah, S. H., Al-Oubaidy, S. A., and Ghaleb, R. A. (2024). Detection of new single nucleotide polymorphisms in plasma cells carcinoma (myeloma) by polymerase chain reaction—single strand conformation polymorphism and sequencing technical methods. *Med. J. Babylon* 21, 52–56. doi: 10.4103/MJBL.MJBL_41_23
- Pei, L., Guo, M., Wang, C., Liu, X., and Quan, Z. (2015). An overview of SNP interactions in genome-wide association studies. *Brief. Funct. Genom.* 14, 143–155. doi: 10.1093/bfpg/elu036
- Peralta, I. E., Spooner, D. M., and Knapp, S. (2008). Taxonomy of wild tomatoes and their relatives (*Solanum* sect. *Lycopersicoideae*, sect. *Juglandifolia*, sect. *Lycopersicon*; Solanaceae). *Syst. Bot. Monogr.* 84, 1–186. doi: 10.1007/s001220050031
- Petr, D., Adam, A., Goncalo, A., Albers, C. A., Eric, B., Depristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330
- Sato, S., Tabata, S., Hirakawa, H., Asamizu, E., and Gianese, G. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485(7400). doi: 10.1038/nature11119
- Shen, Y., Wang, J., Shaw, R. K., Yu, H., Sheng, X., Zhao, Z., et al. (2021). 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.* 12. doi: 10.3389/fpls.2021.655254
- Silvestro, D., and Michalak, I. (2012). raxmlGUI: a graphical front-end for RAxML. *Org. Divers. Evol.* 12, 335–337. doi: 10.1007/s13127-011-0056-0
- Song, X., Ge, T., Li, Y., and Hou, X. (2015). Genome-wide identification of SSR and SNP markers from the non-heading chinese cabbage for comparative genomic analyses. *BMC Genomics* 16, 328. doi: 10.1186/s12864-015-1534-0
- Sorkheh, K., Shiran, B., Gradziel, T. M., Epperson, B. K., Martinez-Gomez, P., and Asadi, E. (2007). Amplified fragment length polymorphism as a tool for molecular characterization of almond germplasm: genetic diversity among cultivated genotypes and related wild species of almond, and its relationships with agronomic traits. *Euphytica* 156, 327–344. doi: 10.1007/s10681-007-9382-x
- Stamatidakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Tieman, D., Zhu, G., Resende, MFR, Lin, T., and Klee, H. (2017). A chemical genetic roadmap to improved tomato flavor. *Science* 355, 391–394. doi: 10.1126/science.aal1556
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., and Kelley, D. R. (2014). Erratum: differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. *Nat. Protoc.* 7 (3), 562–578. doi: 10.1038/nprot.2012.016
- Trick, M., Adamski, N. M., Mugford, S. G., Jiang, C. C., Febrer, M., and Uauy, C. (2012). Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. *BMC Plant Biol.* 12, 1. doi: 10.1186/1471-2229-12-14
- Val, A. D. B. D., Breves, S., Cancado, G., and Pasqual, M. (2020). Use of molecular markers SSR and scar for identification of olive accessions. *Biosci. J.* doi: 10.1007/s00122-002-1079-x
- Wang, S., Qiang, Q., Xiang, L., Fernie, A. R., and Yang, J. (2023). Targeted approaches to improve tomato fruit taste [J]. *Hortic. Res.* 10 (1), 57–68. doi: 10.48130/VR-2023-0016
- Wang, M. L., Zhu, C., Barkley, N. A., Chen, Z., Erpelding, J. E., Murray, S. C., et al. (2009). Genetic diversity and population structure analysis of accessions in the US historic sweet sorghum collection. *Theor. Appl. Genet.* 120, 13–23. doi: 10.1007/s00122-009-1155-6
- Wang, Z., Fang, B., Chen, J., Zhang, X., Luo, Z., Huang, L., et al. (2010). *De novo* assembly and characterization of root transcriptome using illumina paired-end sequencing and development of cSSR markers in sweetpotato (*Ipomoea batatas*). *BMC Genomics* 11, 726. doi: 10.1186/1471-2164-11-726
- Wang, Y., Lv, H., Xiang, X., Yang, A., Feng, Q., Dai, P., et al. (2021a). Construction of a SNP fingerprinting database and population genetic analysis of cigar tobacco germplasm resources in China. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2023.1080691
- Wang, Y., Lv, H., Xiang, X., Yang, A., Feng, Q., Dai, P., et al. (2021b). Construction of a SNP fingerprinting database and population genetic analysis of cigar tobacco germplasm resources in China. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.618133
- Xu, C., Ren, Y., Jian, Y., Guo, Z., Zhang, Y., Xie, C., et al. (2017). Development of a maize 55 k SNP array with improved genome coverage for molecular breeding. *Mol. Breed.* 37, 1–12. doi: 10.1007/s11032-017-0622-z
- Xu, W., Wang, Y., Xie, J., Tan, S. X., Wang, H., Zhao, Y., et al. (2023). Growth-regulating factor 15-mediated gene regulatory network enhances salt tolerance in poplar. *Plant Physiol.* 00, 1–18.
- Yamaguchi, A., Mukai, Y., Sakuma, T., Suganuma, Y., Furugen, A., Narumi, K., et al. (2023). Molecular characteristic analysis of single-nucleotide polymorphisms in slc16a9/hmct9. *Life Sci.* 334. doi: 10.1016/j.lfs.2023.122205
- Yang, G. L., Chen, S. P., Chen, L. K., Sun, K., Huang, C. H., Zhou, D. H., et al. (2019). Development of a core SNP arrays based on the KASP method for molecular breeding of rice. *Rice* 12, 21. doi: 10.1186/s12284-019-0272-3
- Yang, S., Yuan, Y., Wei, X., Wang, Z., and Zhang, X. (2018). Optimization and establishment of KASP reaction system for Chinese cabbage (*Brassica rapa* L. SSP. *Pekinensis*). *Acta Horticulturae Sinica.* 14 (12), 2442–2452. doi: 10.16420/j.issn.0513-353X.2018-0274
- Yao, J., Shen, Z., Zhang, Y., Wu, X., and Chen, S. (2020). *Populus euphratica* WRKY1 binds the promoter of PeHA1 to enhance gene expression and salt tolerance. *J. Exp. Bot.* 71 (4), 1527–1539. doi: 10.1093/jxb/erz493
- Zhang, C. W., et al. (2011). Analysis and assessment on nutritional quality of different tomato varieties. *China Vegetables.* doi: 10.1080/00288233.2011.613403
- Zhang, Y., Song, M. F., Li, H. T., Sun, H. F., and Zhang, Z. L. (2021). DNA barcoding identification of original plants of a rare medicinal material *resina draconis* and related dracaena species. *Chin. J. Traditional Chin. Med.* 46, 2173–2181. doi: 10.19540/j.cnki.cjcm.20210124.104