TROPICAL LIFE SCIENCES RESEARCH

Delineating Population Structure and Genetic Erosion in Tropical Tomato Germplasm for Lowland Breeding Initiatives

Authors:

Siti Norhayati Ismail*, Nor Hazlina Mat Sa'at, Sharifah Munirah Syd Mohamad, Muhammad Fairuz Mohd Yusof and Shahril Ab Razak

Correspondence: yati@mardi.gov.my

Submitted; 8 October 2025; Accepted: 2 September 2026; Early view: 5 November 2026
To cite this article: Siti Norhayati Ismail*, Nor Hazlina Mat Sa'at, Sharifah Munirah Syd Mohamad, Muhammad Fairuz Mohd Yusof and Shahril Ab Razak (in press). Delineating population structure and genetic erosion in tropical tomato germplasm for lowland breeding initiatives. *Tropical Life Sciences Research* (Early view).

HIGHLIGHTS

- SSR and SNP profiling revealed moderate genetic variation paired with high inbreeding coefficients (*f*), signaling severe genetic erosion and restricted gene flow in MARDI tomato germplasm.

- Cluster analysis resolved the germplasm into two main groups (standard vs. cherry-type tomatoes), with population structure analysis identifying an optimal sub-structure of K = 6.

- WorldVeg accessions (CLN3125A, CLN3125Q, CLN3125L, CLN2026D, and CLN5915-206D4) were identified as key parental candidates for heat and disease tolerance to widen the genetic base for tropical lowland breeding.

© Penerbit Universiti Sains Malaysia 2026

EARLY VIEW

Delineating Population Structure and Genetic Erosion in Tropical Tomato Germplasm for Lowland Breeding Initiatives

[1]Siti Norhayati Ismail*, [2]Nor Hazlina Mat Sa'at, [1]Sharifah Munirah Syd Mohamad, [1]Muhammad Fairuz Mohd Yusof and [1]Shahril Ab Razak

[1]Agri-omics and Bioinformatics Programme, Biotechnology and Nanotechnology Research Centre, MARDI Headquarters, Persiaran MARDI-UPM, 43400 Serdang, Selangor, Malaysia
[2]Breeding Programme, Horticulture Research Centre, MARDI Headquarters, Persiaran MARDI-UPM, 43400 Serdang, Selangor, Malaysia

***Corresponding author:** yati@mardi.gov.my

Submitted; 8 October 2025; Accepted: 2 September 2026; Early view: 5 November 2026
To cite this article: Siti Norhayati Ismail*, Nor Hazlina Mat Sa'at, Sharifah Munirah Syd Mohamad, Muhammad Fairuz Mohd Yusof and Shahril Ab Razak (in press). Delineating population structure and genetic erosion in tropical tomato germplasm for lowland breeding initiatives. *Tropical Life Sciences Research* (Early view).

Running head: Delineating Population Structure in Tropical Tomato

Abstract. Understanding population structure and mitigating genetic erosion are critical for developing climate-resilient crop varieties tailored to tropical climates. Due to an intensive domestication history, cultivated tomato (*Solanum lycopersicum*) germplasm suffers from a narrowed genetic base that severely limits potential advancements in lowland breeding initiatives. This study delineates the population structure and assesses genetic erosion within a diverse collection of cultivated tomato accessions maintained by the Malaysian Agricultural Research and Development Institute (MARDI). The germplasm panel comprises breeding lines from MARDI and the World Vegetable Centre (WorldVeg), commercial varieties from local seed companies (Green World Genetics and Chia Hup Agriculture), regional accessions from Thailand, and landraces from Sabah and Sarawak. Genotypic profiling using simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers revealed moderate genetic variation coupled with high inbreeding coefficients (f). These metrics signify

severe genetic erosion and restricted gene flow, likely driven by historical bottlenecks and population stratification. Both SSR- and SNP-based cluster analyses resolved the germplasm into two primary clusters, clearly demarcating standard-type from cherry-type tomato accessions. Population structure analysis identified an optimal sub-population structure of K = 6, capturing the distinct genetic groups within the repository. Notably, WorldVeg-derived accessions (CLN3125A, CLN3125Q, CLN3125L, CLN2026D, and CLN5915-206D4) exhibited a highly distinct genetic background. Given their established heat and disease tolerance, these lines serve as vital parental candidates to counter genetic erosion and widen the genetic base in future tropical lowland tomato breeding programmes.

Keywords: Genetic Diversity, Cultivated Tomato, Simple Sequence Repeat (SSR), Single Nucleotide Polymorphism (SNP)

Abstrak. Pemahaman terhadap struktur populasi dan pengurangan hakisan genetik merupakan aspek penting dalam pembangunan varieti tanaman yang berdaya tahan terhadap perubahan iklim serta sesuai untuk persekitaran tropika. Sejarah domestikasi yang intensif telah menyebabkan germplasma kultivar tomato (*Solanum lycopersicum*) mengalami penyusutan asas genetik yang ketara, sekali gus mengehadkan potensi kemajuan program pembiakbakaan tomato tanah rendah. Kajian ini bertujuan untuk mencirikan struktur populasi dan menilai tahap hakisan genetik dalam koleksi germplasma tomato yang dipelihara oleh Institut Penyelidikan dan Kemajuan Pertanian Malaysia (MARDI). Panel germplasma yang dikaji terdiri daripada garis pembiakbakaan MARDI dan World Vegetable Centre (WorldVeg), varieti komersial daripada syarikat benih tempatan (Green World Genetics dan Chia Hup Agriculture), aksesi serantau dari Thailand serta varieti tradisional (*landrace*) dari Sabah dan Sarawak. Pencirian genotip menggunakan penanda ulangan jujukan ringkas (SSR) dan polimorfisme nukleotida tunggal (SNP) menunjukkan tahap kepelbagaian genetik yang sederhana dengan nilai pekali pembiakbakaan dalam (*f*) yang tinggi. Dapatan ini menunjukkan berlakunya hakisan genetik yang ketara serta aliran gen yang terhad, yang berkemungkinan berpunca daripada kesan penyempitan populasi dan penstrukturan populasi pada masa lalu. Analisis pengelompokan (*cluster*) berasaskan SSR dan SNP masing-masing mengelaskan germplasma kepada dua kelompok utama yang dapat membezakan aksesi tomato jenis standard dan tomato jenis ceri. Analisis struktur populasi mengenal pasti enam subpopulasi ($K = 6$) sebagai struktur populasi yang paling optimum, sekali gus mencerminkan kewujudan kumpulan genetik yang berbeza dalam koleksi tersebut. Aksesi yang berasal daripada WorldVeg, iaitu CLN3125A, CLN3125Q, CLN3125L, CLN2026D dan CLN5915-206D4, didapati mempunyai latar belakang genetik yang berbeza daripada aksesi-aksesi lain. Memandangkan aksesi ini mempunyai toleransi yang baik terhadap suhu tinggi dan penyakit,

ia berpotensi sebagai sumber induk yang penting untuk mengurangkan hakisan genetik serta memperluaskan asas genetik dalam program pembiakbakaan tomato tanah rendah tropika pada masa hadapan.

Kata kunci: Kepelbagaian Genetik, Tanaman Tomato, Ulangan Jujuan Ringkas (SSR), Polimorfisme Nukleotida Tunggal (SNP)

INTRODUCTION

Tomato (*Solanum lycopersicum* L.), is the most economically important edible nightshade in the Solanaceae family, which comprises more than 3,000 species. Cultivated tomato alongside its 12 wild relatives, belongs to the *Lycopersicon* clade (Bauchet and Causse 2012). Tomato can be consumed fresh in the form of salads or processed into purée and ketchup. In addition to serving as a rich source of essential micronutrients, particularly vitamins A and C (ascorbic acid), tomato contains high levels of carotenoids, principally lycopene, and diverse antioxidants that help mitigate the risk of cancer and cardiovascular diseases (Mazumder *et al.* 2021). It is also recognised as a model species for genetic research, particularly in examining fruit development, disease resistance, and metabolite accumulation (Rothan *et al.* 2019).

With a total production of approximately 70.1 million tonnes in 2023, China is the world's largest producer of tomatoes, followed by India (20.4 million tonnes) and Türkiye (13.3 million tonnes) (FAOSTAT, 2025). Tomato is a popular and nutritious vegetable crop in Malaysia. Given the total production of 188,842.24 tonnes from a planted area of 1,990.68 hectares, it is the highest-yielding crop produced in the country in 2023. Tomato is a temperate crop that grows best in cooler climates, with approximately 95% of national production concentrated in the limited highland areas of Pahang (110,227.31 tonnes) and Kelantan (70,433.63 tonnes) (DOA, 2025).

Cultivated tomato has a far more superior morphology than its wild ancestors in terms of size, colour, shape, fruit taste and quality owing to its evolution and domestication. Notwithstanding, the highly-diverse phenotypic variation in cultivated tomato does not reflect its genetic diversity. In line with Saulo *et al.* (2014), cultivated tomato genomes are more genetically-eroded than the broad genetic diversity observed in wild tomato species. The long history of domestication has led this vegetable crop to exhibit lower genetic diversity compared to other autogamous counterparts, thus significantly narrowing its genetic base. Such reduction in diversity is known as the domestication syndrome (Bauchet and Causse 2012 and Wang *et*

al. 2016), which is driven by prolonged inbreeding and selection, thus limiting tomato crop improvement.

Continued genetic erosion in modern germplasm could undermine long-term food and feed security (Lu *et al.* 2022; Long *et al.* 2026), following the increased susceptibility to abiotic (drought, salinity, and extreme temperature) and biotic (insect, fungi, bacteria and virus) threats (Zamani *et al.* 2025). Hence, breeders have been crossing elite varieties with wild relatives to enhance the genetic diversity of cultivated tomato (*Solanum lycopersicum*). A notable study, which involved crossing a yellow-fruited mutant line (e9292) with the wild species *S. pimpinellifolium*, revealed novel hybrid lines that exhibit improved tolerance to salinity and drought, as well as increased ascorbic acid and lycopene levels (Li *et al.* 2022). From a scholarly perspective, hybrids between *S. lycopersicum* and *S. chilense* (wild species) demonstrate higher tolerance to salt stress (Bigot *et al.* 2023).

Tomato breeding programmes are increasingly focusing on the development of cultivars with enhanced tolerance to biotic and abiotic stresses, alongside efforts to improve yield and fruit quality. The prevalence of these stresses could be driven by the accelerating impacts of climate change, which has significantly hindered global tomato production owing to greater variability and extremity in environmental conditions (Rezk *et al.* 2021). In highland areas, the expansion of agricultural activities has led to environmental issues pertaining to landslides and soil erosion. The steep, high-risk topography and limited availability of suitable land call for developing new tomato varieties or hybrids adapted for cultivation in the lowland regions of Malaysia (Nor *et al.* 2017). These challenges must be deeply understood to navigate the genetic diversity in tomato germplasm resources.

The primary objective of the present work was to delineate the population structure and assess the extent of genetic erosion within the MARDI tomato germplasm collection, thereby providing breeders with a strategic framework for selecting robust parental lines tailored for tropical lowland breeding initiatives. Both SNP and SSR markers were employed following their robustness, high abundance and polymorphism, enhanced reproducibility and reliability, broad genome-wide distribution, co-dominant inheritance, as well as suitability for high-throughput genotyping platforms. In line with past works, both SSR and SNP markers are suitable for assessing genetic diversity in tomato germplasm (Bhardwaj *et al.* 2016; Kaushal *et al.* 2017; Tranchida-Lombardo *et al.* 2019; Mata-Nicolás *et al.* 2020 and Al-Shammari *et al.* 2021).

MATERIALS AND METHODS

Plant Materials

This study evaluated 47 cultivated tomato accessions, which included 12 inbred lines developed by the Malaysian Agricultural Research and Development Institute (Table 1). Among these local accessions, MTI and MTII represent established lowland inbred varieties of the standard type. The MTI variety produces smaller fruits but exhibits superior tolerance to bacterial wilt disease compared to MTII (Nor *et al*. 2016). The newly developed lowland variety, MT3 (No. 32), which is also designated as MAHA18, was derived from a targeted hybridisation program involving MTI, MTII, and the Japanese cultivar Fumikura. Although MT3 was officially launched to the public during the Malaysia Agriculture, Horticulture and Agrotourism Show in 2018, evaluation of this advanced breeding line was initiated in 2017 (Nor *et al*. 2017). As a determinate variety, MT3 demonstrates high yield potential, elevated beta-carotene content, extended shelf life, and an attractive red fruit profile. Prior to its official commercial introduction, establishing a robust molecular identity for MT3 was essential. Conducting DNA fingerprinting at this stage served as a critical prerequisite to prove its genetic distinctness from existing commercial varieties and its parental lines, thereby ensuring variety authenticity and intellectual property protection.

Table 1: List of tomato accessions, their sources and types.

No.	Tomato Variety	Sources	Tomato type
1	MTI	MARDI	Standard
2	MTII	MARDI	Standard
3	MT3 (No.32/MAHA18)	MARDI	Standard
4	P7x32	MARDI	Standard
5	P20x32	MARDI	Standard
6	CherryTomato	MARDI	Standard
7	TC1	MARDI	Cherry
8	TC7	MARDI	Cherry
9	TC9	MARDI	Cherry
10	TC11	MARDI	Cherry
11	TC13	MARDI	Cherry
12	TC14	MARDI	Cherry
13	TC15	MARDI	Cherry
14	SK42	Sarawak, Malaysia	Standard
15	SK65	Sarawak, Malaysia	Standard
16	SK160	Sarawak, Malaysia	Standard
17	Tomato Sabah	Sabah, Malaysia	Standard
18	TO14	Green World Genetics (GWG)	Standard
19	GWG1	Green World Genetics (GWG)	Standard
20	GWG2	Green World Genetics (GWG)	Cherry
21	TC18	Green World Genetics (GWG)	Cherry
22	TO15	Chia Hup Agriculture	Standard
23	TO17	Chia Hup Agriculture	Standard
24	CLN1621F	World Vegetable Centre (WorldVeg)	Standard

25	CLN2366B	World Vegetable Centre (WorldVeg)	Standard
26	CLN3125L	World Vegetable Centre (WorldVeg)	Standard
27	CLN3125A	World Vegetable Centre (WorldVeg)	Standard
28	CLN3125Q	World Vegetable Centre (WorldVeg)	Standard
29	CLN3078A	World Vegetable Centre (WorldVeg)	Standard
30	CLN3078C	World Vegetable Centre (WorldVeg)	Standard
31	CLN2514B	World Vegetable Centre (WorldVeg)	Cherry
32	CLN2071B	World Vegetable Centre (WorldVeg)	Cherry
33	CLN2026D	World Vegetable Centre (WorldVeg)	Standard
34	CLN5915-206D4	World Vegetable Centre (WorldVeg)	Standard
35	TO16	Seed Company	Standard
36	Baccarat	Seed Company	Standard
37	Big Tomato	Seed Company	Standard
38	Black Tomato	Seed Company	Standard
39	Yellow Tomato	Seed Company	Standard
40	Yellow Stripe	Seed Company	Cherry
41	Red Stripe	Seed Company	Cherry
42	LS1	Seed Company	Standard
43	LS2	Seed Company	Standard
44	AGIAD	Thailand	Standard
45	AGIAD5	Thailand	Standard
46	AGIAD16	Thailand	Standard
47	ThaiF1Seed	Thailand	Standard

Seeds of all accessions were germinated and cultivated under controlled conditions at the Horticulture Research Centre, MARDI Headquarters, Serdang. Subsequent laboratory experiments, including molecular characterisation and fragment analysis, were executed in 2017 at the Centre for Marker Discovery and Validation laboratory, which is located within the Biotechnology and Nanotechnology Research Centre Building at the MARDI Headquarters.

Selection for Simple Sequence Repeat (SSR) markers

Thirty-two potential SSR markers with relatively high heterozygosity and an elevated number of alleles were selected for screening (He *et al*. 2003; Benor *et al*. 2008; Zhou *et al*. 2015). Only seven polymorphic markers were selected for post-PCR optimisation and fragment analysis. The remaining markers were excluded due to non-specific binding, monomorphic properties, or poor amplification during PCR optimisation in 2% pre-stained agarose gel electrophoresis and fragment analysis via a DNA Analyser (ABI3730XL, Applied Biosystems, USA). Table 2 details the characteristics of these seven markers.

Table 2: Detail information on seven polymorphic SSR markers.

Marker Name	Marker ID	Repeat Motive	Na[a]	H_e[b]	H_o[c]	PIC[d]	MAF[e]	f[f]
AI773078	T01	(AAT)14	3	0.48	0.1	0.41	0.65	0.77
AW037347	T03	(AAT)12	4	0.64	0.06	0.57	0.46	0.91
TMS9	T08	(GATA)26	6	0.78	0.25	0.74	0.31	0.69
SSR50	T12	(TC)6, (CCTTC)2	2	0.4	0.02	0.32	0.73	0.95
LEag001	T15	(AG)11	2	0.48	0.12	0.36	0.61	0.76
LEat002	T16	(AT)9	2	0.49	0.07	0.37	0.57	0.88
LEat016	T17	(AT)9	4	0.61	0.38	0.56	0.56	0.40
Mean			3.29	0.55	0.14	0.48	0.56	0.75

Legend: Na[a] = number of alleles; H_e[b] = expected heterozygosity H_o[c] = observed heterozygosity and PIC[d] = polymorphic information content; MAF[e] = major allele frequency and f[f] = inbreeding coefficient.

Selection for Single Nucleotide Polymorphism (SNP) Markers

The complete genome sequence of *Solanum lycopersicum* was retrieved from the National Centre for Biotechnology Information (NCBI) and employed as the reference genome for downstream analyses. Raw sequencing datasets for *S. arcanum* (ERR410227), *S. pimpinellifolium*, and *S. pennellii* (ERR538191) were downloaded from the European Nucleotide Archive (ENA). Read pre-processing was performed using Platanus trim (v1.0.7) with a quality threshold of q = 20, during which reads shorter than 50 bp and adapter contaminations (including PhiX control sequences) were removed. The resulting high-quality reads were aligned to the *S. lycopersicum* reference genome using the Burrows-Wheeler Aligner (BWA-MEM, v0.7.17). Furthermore, SAM format files were converted to BAM format using Picard tools, which were also utilised to identify and mark duplicate reads.

The SNP calling was executed using the HaplotypeCaller module from the Genome Analysis Toolkit (GATK). The resulting three VCF files were merged using the CombineVariants function, with the shared variants across all files extracted via the SelectVariants function to isolate intersecting SNPs. To ensure the selection of high-quality variants, several SNP filtration steps were applied, which involved retaining SNPs with a Phred quality score (Q) greater than 30 and a minimum depth of coverage (DP) greater than 10. SNPs located within 100 bp of one another were excluded to minimize potential linkage disequilibrium or data redundancy. Lastly, 100 bp flanking sequences upstream and downstream of the selected SNPs were extracted from the reference genome using BEDTools.

DNA Extraction

Xin and Chen's (2012) modified high-throughput protocol was adapted in this study to extract genomic DNA. A 5 mm tungsten bead was placed into each well of a 96-well block plate (Corning Incorporated, USA), followed by approximately 0.5 mg of fresh young tomato leaf tissue (two weeks old). The samples were frozen at −80°C overnight and, subsequently, ground into a fine powder using a TissueLyser (Qiagen, Netherlands) at 30 Hz for 1 minute. Notably, 600 µL of DNA extraction buffer (100 mM Tris-HCl, pH 8.0; 200 mM EDTA; 2% CTAB; 1.2 M NaCl; 0.1% β-mercaptoethanol) was added to each well using a multi-channel, multi-dispense pipette (Thermo Scientific™, USA) immediately after homogenisation. To ensure thorough mixing, the contents were gently inverted and incubated at 60°C for 1 hour in a water bath (Memmert, Germany). Samples were cooled to room temperature for 15 minutes post-incubation before adding 350 µL of chloroform:isoamyl alcohol (24:1, v/v). The plate was centrifuged at 5500 rpm for 15 minutes (Beckman Coulter, USA), with the aqueous phase carefully transferred to a fresh 96-well block plate. Furthermore, the DNA was precipitated by adding an equal volume of cold isopropanol, followed by a wash with 200 µL of 70% ethanol. The DNA pellets were air-dried for approximately 2 hours and resuspended in 50 µL of TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0).

DNA Qualification and Quantification

The extracted genomic DNA quality was evaluated via 0.8% agarose gel electrophoresis, pre-stained with ethidium bromide. Essentially, DNA bands were visualised using the Quantum ST4 3000 imaging system under UV light (Thermo Scientific, USA). Only samples displaying sharp, intact bands without smearing were selected for downstream analysis. Quantification was performed with Epoch® microplate reader (Agilent Technologies, USA). Subsequently, the DNA concentrations were normalised to 20 ng/µL with the standard dilution formula ($C_1V_1 = C_2V_2$).

PCR Amplification and Fragment Analysis

High-throughput polymerase chain reaction (PCR) amplifications were executed using semi-automated workflow. The master mix was initially prepared in 1.5 mL microcentrifuge tubes and dispensed into a 96-well PCR plate (Axygen™, USA) using an electronic multi-channel, multi-dispense pipette (Thermo Scientific™, USA). Each well received 9 µL of the master mix, formulated using the following initial stock concentrations: 1× PCR buffer (Invitrogen™, USA),

50 mM $MgCl_2$ (Invitrogen™, USA), 2 mM dNTPs (Invitrogen™, USA), 10 µM of an unlabelled forward primer with an M13 tail (Schuelke, 2000), 10 µM of an unlabelled reverse primer, 5 µM of a fluorescently dye-labelled M13 tail (labelled with either FAM, VIC, PET, or NED), and 5 U/µL of Taq DNA Polymerase (Invitrogen™, USA). Subsequently, 1 µL of genomic DNA template (30 ng/µL) was transferred into each well using a Biomek automated liquid handling system (Beckman Coulter™, USA), yielding a final total reaction volume of 10 µL per sample. The PCR amplifications were performed on a thermal cycler (Applied Biosystems, USA) under the following conditions: an initial denaturation at 95°C for 5 minutes; 34 cycles of denaturation at 95°C for 30 seconds, primer-specific annealing temperatures (see Table 1) for 30 seconds, and extension at 72°C for 1 minute, followed by a final extension at 72°C for 10 minutes. Amplified PCR products were subjected to fragment analysis using the ABI 3730XL DNA Analyzer (Applied Biosystems, USA). Fragments ranging from 35 to 500 base pairs (bp) were size-called using the GeneScan™ 500 LIZ size standard (Applied Biosystems, USA).

Out of the initial 47 tomato accessions compiled for this study, 40 accessions were successfully genotyped using the designated SSR primers. Six accessions, namely 'P20x32', 'CherryTomato', 'TC15', 'CLN2071B', 'TO16', ‘LS1’ and 'Yellow Tomato', were completely excluded from downstream analysis due to a high ratio of missing allelic data (>10%). This strict filtering was implemented to ensure maximum data integrity during cluster analysis.

SNP Genotyping

SNP genotyping was performed using the MassARRAY® system (Agena Bioscience, USA). The procedure was carried out according to the iPLEX Pro and Gold Reagents user guide. Out of the initial 47 tomato accessions compiled for this study, 46 accessions were successfully genotyped and retained for downstream genetic analysis. A single accession, 'Red Stripe', was completely excluded from the final dataset due to a low sample genotype call rate (<90%). This strict exclusion criterion was implemented to preserve overall data integrity and prevent the introduction of artifactual parameters during population structure and cluster analysis.

Data Analysis

The genotype data derived from fragment analysis were scored and assessed through the GeneMapper Software v4.0 (Applied Biosystems, USA), as well as converted into multiple formats using CONVERT to facilitate compatibility with various analysis programmes (Glaubitz, 2004). Each locus was screened for the presence of null alleles, scoring errors, and

large allele dropouts using Micro-Checker v2.2 (Van Oosterhout *et al.* 2004). PowerMarker v3.5 was employed to calculate the number of alleles per locus (NA), observed and expected heterozygosity values, and polymorphism information content (PIC) (Liu and Muse, 2005). Based on Nei's genetic distance (Nei, 1983), a UPGMA dendrogram was constructed via PowerMarker and visualised with MEGA11 (Tamura *et al.* 2016). The Bayesian analysis of population structure was conducted using STRUCTURE version 2.3.4 (Pritchard *et al*. 2000). Meanwhile, the optimal number of clusters (K) was determined based on a plot of the mean likelihood for each K value, L(K) (Pritchard *et al*. 2000; Evanno *et al*. 2005). The Evanno method was applied using StructureSelector (https://lmme.qdio.ac.cn/StructureSelector/index.html) to calculate ΔK and the mean of Ln P(K).

RESULTS

Genetic Diversity Assessment Using SSR Markers

Out of the 47 baseline tomato accessions, a final cohort of 40 successfully genotyped accessions was subjected to downstream genetic analysis. Capillary electrophoresis of the remaining cohorts resolved a total of 23 alleles across the evaluated loci. The number of alleles ranged from 2 to 6, with an average of 3.29 alleles per locus (see Table 2). Of these markers, T1, T12, T15, and T16 exhibited moderate PIC values ranging from 0.32 to 0.48. Markers T3 and T17 exhibited slightly high PIC values of 0.56 and 0.57, respectively, while T8 recorded the highest PIC value at 0.74. The overall mean for PIC value (0.48) demonstrated a moderate level of polymorphism. Ranging from 0.31 (T8) to 0.72 (T12), the major allele frequency (MAF) values indicated moderate to high levels of polymorphism. The observed heterozygosity values ranged from 0.02 (T12) to 0.38 (T17), with an average of 0.14, while the expected heterozygosity values, ranged from 0.40 (T12) to 0.78 (T8), with an average of 0.55. Derived from deviations between observed and expected heterozygosity, the inbreeding coefficient (f) ranged from 0.40 (T17) to 0.95(T12) with a mean of 0.75. The Unweighted Pair Group Method with Arithmetic Mean (UPGMA) was employed to construct a dendrogram (see Figure 1), in which two major clusters corresponded to standard-type tomato (blue), with average fruit diameters more than 30 mm and cherry-type tomato accessions (red), typically characterised by fruit diameters of less than 29 mm (FAO 2003). Three sub-clusters were observed within the cherry-type tomato cluster: Sub-cluster 1 entailed lowland-type accessions derived from WorldVeg (CLN2514B) and GWG (GWG2 and TC18); sub-cluster 2 encompassed MARDI's highland-type accessions (TC1, TC9, TC13, TC14, TC11, and TC7); and a distinct sub-cluster 3 was formed by the Red-Striped (Figure 2) and Yellow-Striped

accessions, possibly influenced by their unique stripe pattern colouration. Excluding MTI and MT3, most of the accessions were clearly distinguishable from one another. The lowest genetic distances (data not shown) were recorded between T13 and T14 (0.0026), followed by T11 and T17 (0.0074). Based on these findings, the respective pairs may share a recent common ancestry or possess closely-related genetic background.

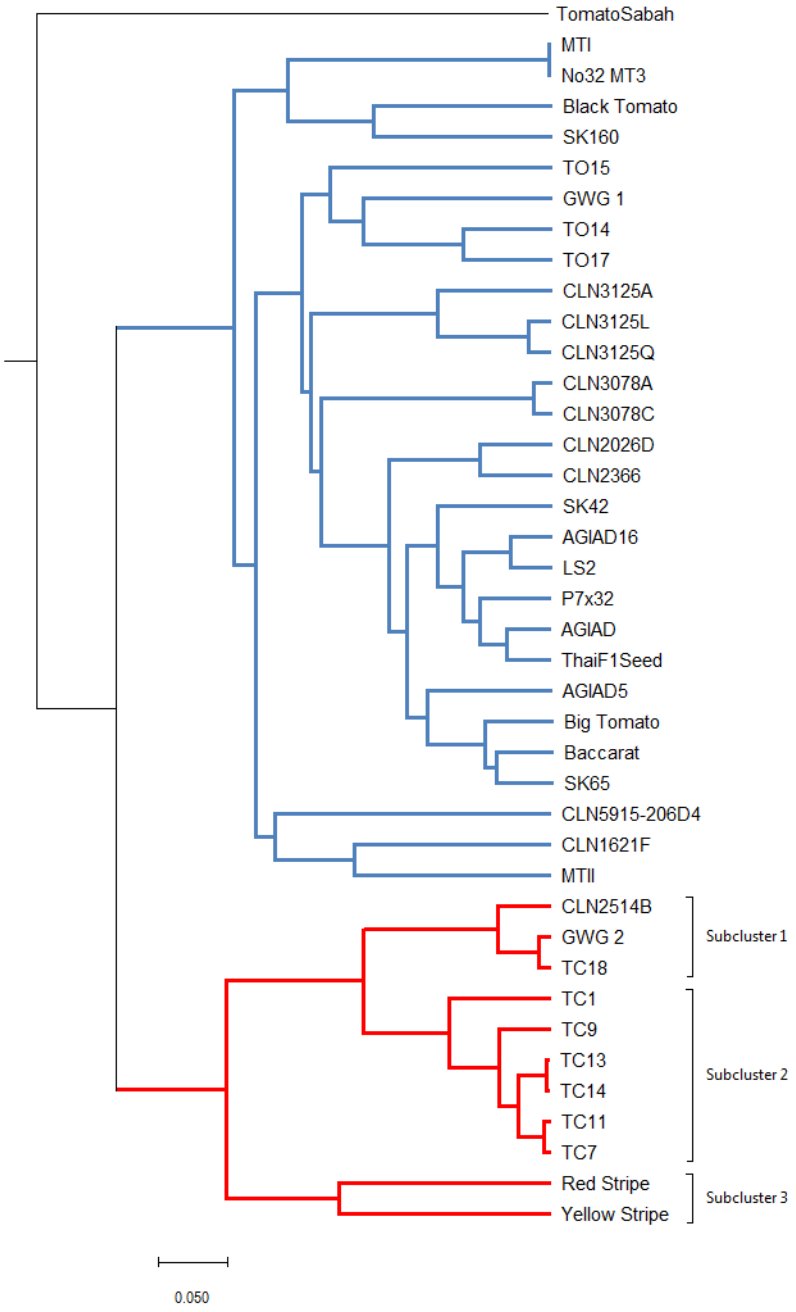


Figure 1: Dendrogram (UPGMA) illustrating the genetic relationships among 40 tomato accessions based on seven SSR markers. Two major clusters are evident: one comprising

standard-type tomatoes (blue) and the other consisting of cherry-type tomatoes (red). Within the cherry-type cluster, three distinct subclusters are also observed.



Figure 2: Red Stripe cherry tomato accession exhibiting a distinct stripe pattern coloration and oval-shaped fruits.

Structure Analysis and Genetic Diversity Assessment Using SNP Markers

While ninety-seven high-quality SNPs were initially selected for genotyping using the MassARRAY® System (Agena Bioscience, USA), only 85 SNPs passed the assay design criteria based on analysis using Typer 4.0 software. Of these, 67 SNPs were retained for downstream analysis. Monomorphic alleles, low call rates (< 90%), or a major allele frequency exceeding 0.95 led to the exclusion of 19 SNPs (Gunundu *et al.* 2025). This study genotyped 46 tomato accessions using 67 SNP markers. The PIC values ranged from 0.03 to 0.38 with an average of 0.25 (data not shown). Of the total markers, 34 SNPs (50.75%) were classified as moderately informative (PIC between 0.25 and 0.38) following Gunundu *et al.'s* (2025) threshold. With PIC values below 0.20, 18 SNPs (27.27%) were categorised as less informative. The observed heterozygosity (Ho) across loci ranged from 0.004 to 0.764, with a mean of 0.183, while the expected heterozygosity (He) ranged from 0.032 to 0.500, with a mean of 0.289. The major allele frequency varied between 0.50 and 0.98, averaging at 0.75. The inbreeding coefficient (f) ranged from 0.12 to 0.98, with a mean value of 0.67. The optimal number of genetic clusters (K) was determined via STRUCTURE analysis, which involved using a burn-in period of 10,000 and 100,000 MCMC (Markov Chain Monte Carlo) repetitions. There were 10 independent runs for each K value ranging from 1 to 10. To determine the most likely K based on the rate of change in log probability values (calculated as delta K [ΔK]), the Evanno method was implemented using the StructureSelector function (Evanno *et al*. 2005). The highest ΔK value was observed at K = 6 (ΔK = 127.84), as shown in Table 3, indicating

the presence of six major genetic clusters. Figure 3, which depicts the scatter plot of ΔK values across K = 1 to 10, clearly shows the primary peak at K = 6 and the secondary peak at K = 2. In line with Figure 3, the bar plot for K = 6 (Figure 4) illustrates the inferred population structure, in which each colour represents one of the key genetic clusters. Two major clusters were revealed in the SNP-based dendrogram (see Figure 5), which was constructed with 67 SNP markers. Cluster 1 (blue) predominantly involved standard-type tomato accessions, excluding Yellow Tomato. Although cluster 2 primarily constituted cherry-type tomatoes, several standard-type accessions (including Big Tomato, CLN2366B, Black Tomato, and Yellow Tomato) fell under this cluster. Meanwhile, CLN3125A, CLN3125Q, CLN3125L, CLN2026D, CLN5915-206D4, and Tomato Sabah accessions suggest a divergent genetic background given their genetically-distinct group outside the two major clusters.

Table 3: ΔK values and log-likelihood estimates for K = 1 to 10 identifying the optimal number of genetic clusters generated by StructureSelector.

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	10	-33965	0.13499	NA	NA	NA
2	10	-29277	66.8952	4687.78	1591.33	23.7884
3	10	-26181	247.163	3096.45	1310.66	5.30281
4	10	-24395	463.394	1785.79	397.56	0.85793
5	10	-23007	1072.61	1388.23	391.21	0.36473
6	10	-21228	21.7563	1779.44	2781.4	127.843
7	10	-22230	6913.79	-1002	3058.55	0.44238
8	10	-20173	2643.19	2056.59	1834.82	0.69417
9	10	-19951	2313.98	221.77	466.59	0.20164
10	10	-20196	5914.32	-244.82	NA	NA

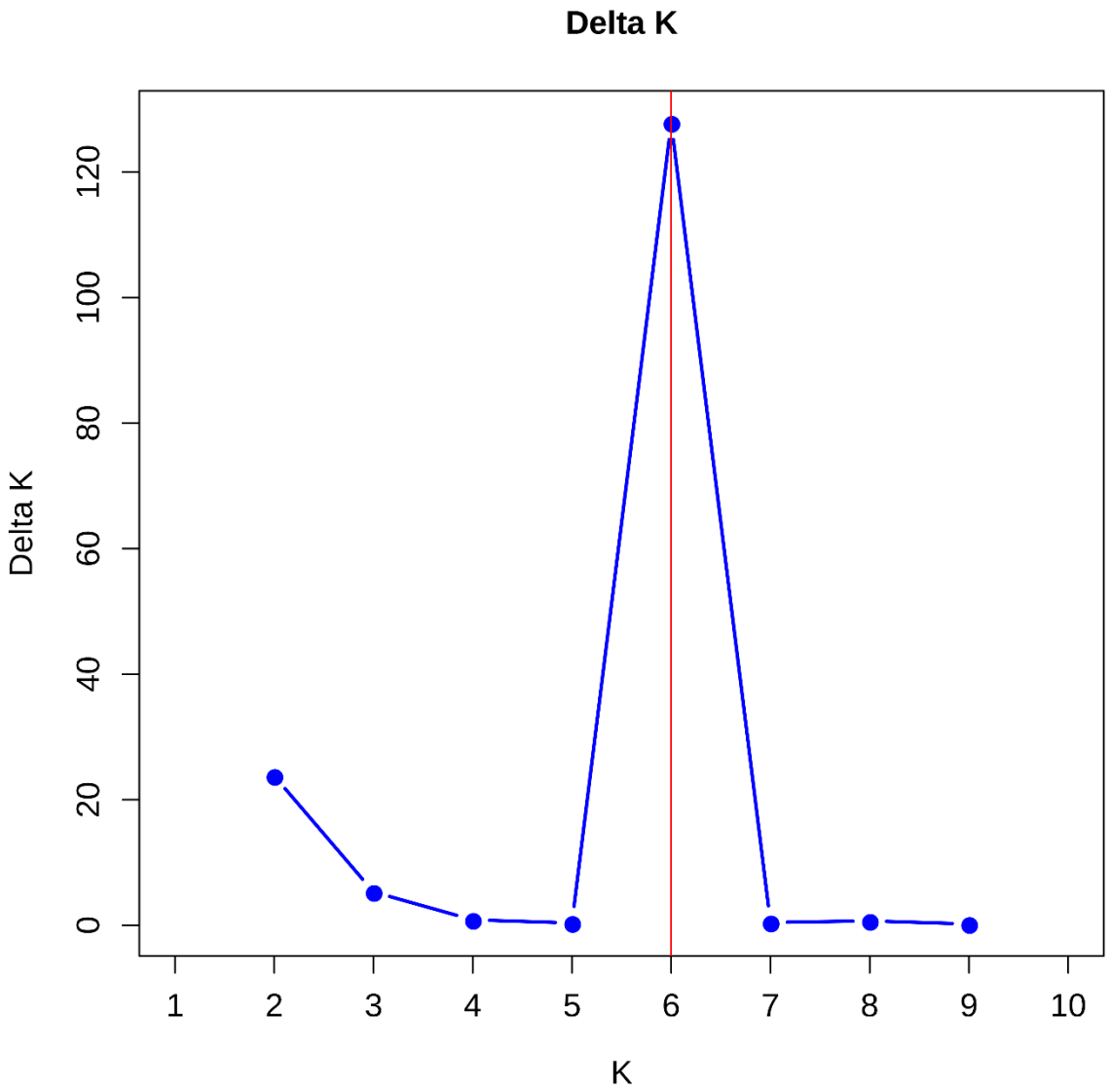


Figure 3: Scatter plot of ΔK values from the Evanno method showing the optimal K = 6 and secondary peak at K = 2.

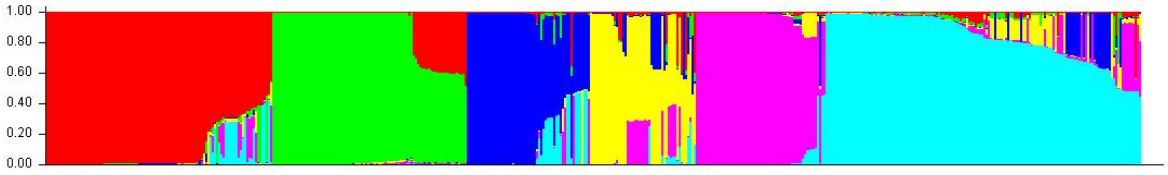


Figure 4: Population structure of tomato accessions at K = 6 based on STRUCTURE analysis.

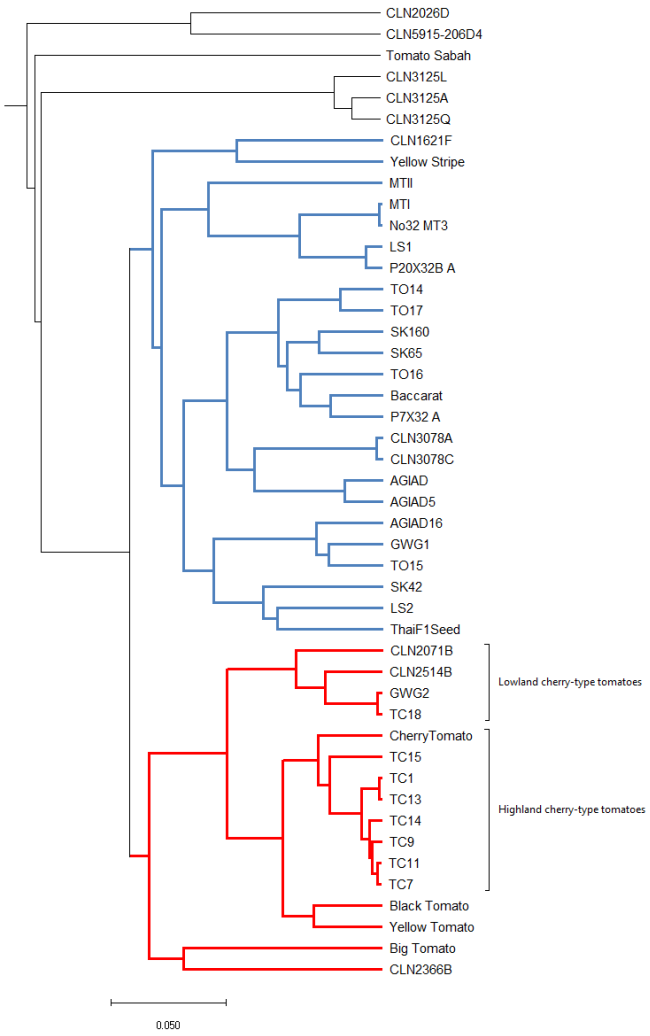


Figure 5: Dendrogram (UPGMA), illustrating the genetic relationships among 46 tomato accessions based on 67 SNP markers. Two major clusters are observed: the first predominantly comprises standard-type tomatoes (blue), while the second consists primarily of cherry-type tomatoes (red). Within the second cluster, further subclusters corresponding to lowland-type and highland-type cherry tomatoes can also be distinguished.

DISCUSSIONS

Genetic Diversity Assessment Using SSR Markers

Polymorphic information content (PIC) values are crucial for assessing genetic marker informativeness. A PIC value exceeding 0.5 is deemed highly informative, values between 0.25 and 0.5 are moderately informative, and values under 0.25 indicate low informativeness (Singh *et al*. 2022). The overall mean PIC value of 0.48 observed in this study indicates a moderate level of polymorphism. Similarly, Benor *et al*. (2008) reported an average PIC value

of 0.31 when screening 39 inbred tomato lines with 35 SSR markers. Zhou *et al*. (2015) evaluated 50 cultivated and wild tomato accessions using 15 genomic and 13 EST-SSR markers, obtaining mean PIC values of 0.49 and 0.45, respectively. Based on these comparable ranges, the seven polymorphic SSR markers deployed here exhibit robust informativeness, validating the baseline detection of moderate genetic diversity across the evaluated germplasm.

The major allele frequency (MAF) represents the frequency of the most common allele at a given locus. Contrary to the polymorphism information content (PIC), higher MAF values reflect lower polymorphism, as they indicate the dominance of a single allele. For instance, the SSR marker T12 exhibited the highest MAF (0.72) and consequently the lowest PIC (0.32), whereas marker T8 displayed the lowest MAF (0.31) and highest PIC (0.74), indicating a superior level of polymorphism (Table 2).

The observed heterozygosity values across all evaluated SSR loci notably deviated from expected heterozygosity, implying a distinct departure from Hardy–Weinberg equilibrium (HWE). This imbalance and the high level of inbreeding (*F*) indicated in the study dataset coincide with the self-pollinating nature of cultivated tomato, which naturally accumulates homozygosity over generations due to intensive selection during domestication. While SSR markers exhibit high polymorphic potential, the elevated *F* values observe indicate strong genetic fixation across most accessions. This marked homozygosity could be attributed to population substructure (the Wahlund effect), historical founder events, or genetic drift typical of small, isolated breeding lines, all of which restrict overall gene flow. Conversely, the relatively lower *F* value observed in 'TO17' may reflect higher residual heterozygosity or the presence of distinct allelic variants at those specific loci.

The commercial varieties 'MTI' and 'MT3' could not be differentiated based on the UPGMA dendrogram. Historically, 'MT3' was developed using 'MTI', 'MTII', and the Japanese cultivar 'Fumikura' as parental lines. The MTI and MTII were selected for their tolerance to bacterial wilt disease, a key limitation in lowland tomato cultivation. Conversely, Fumikura was chosen for its superior fruit quality, attractive red colour, extended shelf life, and to introduce genetic diversity into the local germplasm (Nor *et al.* 2017). Because 'MTI' served as a primary direct parent for 'MT3', the high degree of genetic similarity between them is expected. The genetic distance between MTI and MTII (0.30) (data not shown) indicates a slightly moderate variation level. Based on the current outcomes, additional or more informative molecular markers may be required to achieve finer resolution in distinguishing closely-related accessions.

Positioned as an outlier in the dendrogram (Figure 1), 'Tomato Sabah' (a local cultivar) indicates substantial genetic divergence from the other accessions. Benor *et al.* (2008) similarly observed that 'Usa-16', a tomato cultivar originating from the USA, was positioned as an outlier in a constructed UPGMA dendrogram. Factors such as geographical isolation, genetic drift, localized adaptation, or introgression via natural or intentional crosses with wild relatives or landraces may influence this divergence, leading to the presence of unique alleles. Notably, 'Tomato Sabah' (standard-type) exhibited an absolute Nei's (1983) genetic distance of 1.00 from accessions 'GWG2', 'TC', 'TC7', 'TC11', 'TC13', 'TC14', and 'TC18' (cherry-types) (data not shown). Furthermore, 'GWG2' displayed complete divergence (Nei's genetic distance = 1.00) from both 'MTI' and MT3, while 'MTI' and 'MT3' were entirely divergent from 'TC18'. Phenotypically, 'TC18' (Figure 6) and 'GWG2' (Figure 7) are cherry tomato types characterized by oval-shaped fruits, whereas 'MTI' (Figure 8) and 'MT3/No.32' (Figure 9) are standard tomato types with round-shaped fruits.



Figure 6: TC 18, cherry-type tomato accession exhibiting oval-shaped fruits.



Figure 7: GWG 2, cherry-type tomato accession exhibiting oval-shaped fruits.



Figure 8: MT1, standard-type tomato accession exhibiting round-shaped fruits.



Figure 9: No. 32 (MT3), standard-type tomato accession exhibiting round-shaped fruits.

While complete dissimilarity at all loci is highly improbable among closely related cultivated tomato accessions, these absolute divergence values must be interpreted with caution. This phenomenon is largely attributed to the highly polymorphic nature of the simple sequence repeat (SSR) markers coupled with the limited panel of seven loci utilized in this specific analysis. Because seven polymorphic markers provide a constrained genome-wide resolution for a diverse pool of 40 accessions, these artifactual values do not accurately reflect the true, comprehensive genomic relationships across the germplasm. Instead, this SSR-based profiling serves exclusively as a preliminary baseline, necessitating denser genome-wide marker platforms to validate absolute ancestral boundaries. Reflecting this unrefined genetic background, anecdotal reports from local growers (unpublished data) indicate that 'Tomato Sabah' remains highly susceptible to diseases, produces non-uniform fruits, and exhibits dull, unattractive fruit colouration. Nevertheless, the presence of such genetically distinct accessions contributes to the overall genetic diversity and enhances the expected heterozygosity within the population. In addition, 'MT3' (a MARDI-released variety) exhibits considerable genetic distances from 'CLN1621F' (0.66), 'CLN3078' (0.61), and 'CLN3125Q'

(0.71), suggesting their potential as excellent parental candidates for the future improvement of 'MT3'.

Structure Analysis and Genetic Diversity Assessment Using SNP Markers

The majority of SNP loci exhibited marked discrepancies between observed and expected heterozygosity, coinciding with patterns observed in the SSR marker data. This finding indicates deviations from HWE. Unlike the multi-allelic nature of SSRs, SNP markers are inherently bi-allelic and possess a lower theoretical maximum for PIC (capped at 0.5). This bi-allelic structure also contributes to higher MAF observed in SNP datasets. Characterised by moderate PIC values, significant deviations from HWE, and elevated inbreeding coefficients (f), the SNP marker results closely paralleled those obtained from SSR genotyping. These collective trends reinforce the hypothesis of limited gene flow in the studied population, possibly due to population sub-structure, historical bottlenecks, or founder effects. The optimal number of genetic clusters (K) was determined as 6, consistent with Xu *et al.*'s (2024) analysis of 418 tomato cultivars' genetic diversity using SNP markers. A secondary peak at K = 2 (ΔK = 23.79) suggests a broader genetic split, possibly reflecting a hierarchical population structure within the dataset.

Based on the SNP-based dendrogram, most of the WorldVeg-derived standard-type tomato varieties (excluding CLN3078A, CLN3078C, and CLN1621F) formed a genetically distinct cluster relative to the other accessions. This divergence may be attributed to their shared breeding origin, where selection efforts have historically focused on developing tolerance to heat stress and resistance to Tomato Yellow Leaf Curl Virus (TYLCV), traits that confer an advantage in tropical environments. Classified as an oval-shaped fruit, CLN3125L exhibits genetic distinction from round-shaped CLN3125A and CLN3125Q. The distinct genetic placement of Tomato Sabah reinforces the presence of unique allelic variation based on the SSR-based dendrogram, distinguishing it from the rest of the collection. Despite being categorised as a cherry-type tomato, Yellow Stripe appeared genetically isolated from the cherry-type cluster that might be due to introgression from genetically-distant lineages. The genetic structure of tomato is influenced by fruit morphology, breeding history, as well as selection for abiotic and biotic stress resilience. The observed genetic diversity offers valuable genetic resources for future breeding programmes, especially those focused on elevating stress tolerance and broadening the genetic base of cultivated tomatoes.

Implications for Parental Selection and Heterosis Optimization

The complementary use of both SSR and SNP marker systems provides a robust framework for informing parental selection in tomato breeding programs. SSR markers, being highly polymorphic and multiallelic, excel at identifying specific allelic variations and resolving close pedigree relationships at targeted loci. Conversely, high-density SNPs offer uniform, genome-wide coverage that provides a precise estimate of overall genetic distance across the entire germplasm. By integrating the specific allelic insights from SSRs with the comprehensive genomic background data from SNPs, breeders can strategically select parental pairs that maximize genetic distance. This dual-marker approach ensures the combination of distinct target alleles while optimizing heterosis and minimizing the risk of inbreeding depression in subsequent generations.

CONCLUSIONS

The genetic diversity assessment of MARDI's tomato germplasm via SSR and SNP markers revealed a moderate level of genetic variation among the accessions. Thus, introgression with genetically-diverse cultivated varieties or wild tomato relatives is highly recommended to widen the collection's genetic base. Such efforts potentially broaden allelic diversity, facilitate the development of improved cultivars, and strengthen the germplasm's potential for future breeding initiatives. The genetic analyses based on SSR and SNP markers aligned with tomato-type morphology, clearly distinguishing between standard and cherry types. Intriguingly, SSR markers further classified the cherry group into three sub-clusters: lowland-type, highland-type, and striped-fruit types. The WorldVeg-derived accessions, particularly CLN3125A, CLN3125Q, CLN3125L, CLN2026D, and CLN5915-206D4, emerged as potential parental candidates for breeding programmes following their tolerance to heat and diseases, as well as their distinct genetic profiles compared to the other accessions. Furthermore, CLN1621F, CLN3078, and CLN3125Q appear to be potential parental candidates for the future improvement of MT3, a MARDI-released tomato variety. The genetic distance data generated in this study provide a valuable preliminary reference for tomato breeders in identifying optimal parental combinations, given that greater genetic divergence generally enhances the likelihood of heterotic expression.

ACKNOWLEDGMENTS

This research was funded by the National Key Economic Area (NKEA) Entry Point Project 14 (EPP 14) grant. The authors wish to thank the project team members from the Biotechnology

and Nanotechnology Research Centre and the Horticulture Research Centre for their direct and indirect contributions to this study.

AUTHORS' CONTRIBUTIONS

Siti Norhayati: Conceptualization, experimental design, methodology, investigation (SNP and SSR genotyping), data analysis, writing – original draft and writing – review & editing
Nor Haslina: Methodology writing – review & editing
Muhammad Fairuz: Investigation (bioinformatics) and writing – review & editing
Shahril: Experimental design and writing – review & editing

REFERENCES

Al-Shammari, A M A, Hamdi, G J, Al-Mahdawi M A S and Mohammed N K. (2021). Genetic diversity analysis and DNA fingerprinting of tomato breeding lines using SSR markers. *Journal of Agricultural Science* 32(1): 1–7. https://doi.org/ 10.15159/jas.21.13

Bauchet G and Causse M. (2012). Genetic diversity in tomato (Solanum lycopersicum) and its wild relatives. *Genetic diversity in plants* 8: 134–162. https://hal.inrae.fr/hal-02805788/file/48918_20120322115308414_1.pdf

Benor, Solomon, Mengyu Zhang, Zhoufei Wang, and Hongsheng Zhang. Assessment of genetic variation in tomato (Solanum lycopersicum L.) inbred lines using SSR molecular markers. *Journal of Genetics and Genomics* 35, no. 6 (2008): 373-379. https://doi.org/10.1016/S1673-8527(08)60054-5

Bhardwaj A, Dhar Y V, Asif M H and Bag S K. (2016). In silico identification of SNP diversity in cultivated and wild tomato species: insight from molecular simulations. *Scientific Reports* 6(1): 38715. https://doi.org/10.1038/srep38715

Bigot S, Leclef C, Rosales C, Martínez J P, Lutts S and Quinet M. (2023). Comparison of the salt resistance of Solanum lycopersicum x Solanum chilense hybrids and their parents. *Frontiers in Horticulture 2: 1130702*. https://doi.org/10.3389/fhort.2023.1130702

Department of Agriculture (DOA). Statistik tanaman sayur-sayuran dan tanaman kontan 2023. http://www.doa.gov.my/resources/aktiviti_sumber/sumber_awam/penerbitan/statistik_tanaman/statistik_tanaman_sayur_tanaman_kontan_2023.pdf (accessed on 28 July 2025).

Department of Agriculture (DOA). *Vegetables and Cash Crops Statistic*. http://www.doa.gov.my/index/resources/aktiviti_sumber/sumber_awam/maklumat_pe

rtanian/perangkaan_tanaman/perangkaan_sayur_tnmn_ladang_2018.pdf (accessed on 23 November 2018).

Evanno G S, Regnaut S and Goudet J. (2005). Detecting the number of clusters ofindividuals using the software structure: A simulation study. *Molecular Ecology 14:2611–2620.* https://doi.org/10.1111/j.1365-294X.2005.02553.x

FAO (2003). *Proposed draft standard for tomatoes. Report of eleventh session of the codex committee on fresh fruits and vegetables.* https://www.fao.org/4/j0695e/j0695e06.htm? (assessed on 22 July 2025).

Food and Agriculture Organization of the United Nations (FAOSTAT). *Crops and livestock products*. https://www.fao.org/faostat/ (accessed on 14 April 2025).

Glaubitz J C. (2004). CONVERT: user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Molecular Ecology Notes 4(2): 309-310.* https://doi.org/10.1111/j.1471-8286.2004.00597.x

Gunundu R, Shimelis H, Tesfamariam SA. (2025). Genetic diversity and population structure analyses of tropical maize inbred lines using Single Nucleotide Polymorphism markers. *PloS one. 24;20(1): e0315463.* https://doi.org/10.1371/journal.pone.0315463

He, C., V. Poysa, and K. Yu. Development and characterization of simple sequence repeat (SSR) markers and their use in determining relationships among Lycopersicon esculentum cultivars. *Theoretical and Applied Genetics 106, no. 2 (2003): 363-373*. https://doi.org/10.1007/s00122-002-1076-0

Kaushal A, Singh A, and Jeena A S. (2017). Genetic diversity in tomato (Solanum lycopersicum L.) genotypes revealed by simple sequence repeats (SSR) markers. *Journal of Applied and Natural Science 9(2): 966*. http://jans.ansfoundation.org/previous-issues/volume-9-year-2017-issue-2#TOC-Genetic-diversity-in-tomato-Solanum-lycopersicum-L.-genotypes-revealed-by-simple-sequence-repeats-SSR-markers

Li W, Li Y, Liang Y, Ni L, Huang H, Wei Y, Wang M, Zhang L and Zhao L. (2022). Generating novel tomato germplasm using the ancestral wild relative of Solanum pimpinellifolium. *Horticulturae 9(1): 34.* https://doi.org/10.3390/horticulturae9010034

Liu K and Muse S V. (2005). PowerMarker: An Integrated Analysis Environment for Genetic Marker Data. *Bioinformatics 21(9): 2128–2129.* https://doi.org/10.1093/bioinformatics/bti282

Long Y., Paengkoum S., Lu S., Niu X., Thongpea S., Taethaisong N., Han Y. and Paengkoum P. (2026). Degradation kinetics and protein intestinal digestibility of tomato pomace in goats. *Frontiers in Veterinary Science* 12: 1724783. https://doi.org/10.3389/fvets.2025.1724783

Lu S., Chen S., Li H., Paengkoum S., Taethaisong N., Meethip W. and Paengkoum P. (2022). Sustainable valorization of tomato pomace (*Lycopersicon esculentum*) in animal nutrition: a review. *Animals* 12(23): 3294. https://doi.org/10.3390/ani12233294

Mata-Nicolás E, Montero-Pau J, Gimeno-Paez E, Garcia-Carpintero V, Ziarsolo P, Menda N and Mueller L A. (2020). Exploiting the diversity of tomato: the development of a phenotypically and genetically detailed germplasm collection. *Horticulture Research 7.* https://doi.org/10.1038/s41438-020-0291-7

Mazumder M N N, Misran A, Ding P, Wahab P E M and Mohamad A. (2021). Preharvest foliar spray of calcium chloride on growth, yield, quality, and shelf-life extension of different lowland tomato varieties in Malaysia. *Horticulturae 7(11): 466.* https://doi.org/10.3390/horticulturae7110466

Nor H, Rozlaily Z, Sharizan A, Nur S G, Rahayu A, Nuradliza B, Farahuda S, Zaulia O, Sebrina S and Selmiah M. (2017). *National technology transfer conference (Confertech).* MT3: Varieti tomato baharu tanah rendah. Berjaya Waterfront Hotel, Johor, 14–16 November 2017.

Nor H, Sharizan A and Rozlaily Z. (2016). *Buletin Teknologi MARDI. Evaluation and Selection of Selected Tomato Lines for Lowland Cultivation Bil.10: 17 – 21*

Pritchard J K, Stephens M and Donnelly P. (2000). Inference of population structure using multilocus genotype data. *Genetics 155(2): 945–959*. https://doi.org/10.1093/genetics/155.2.945

Rezk A, Abhary M, and Akhkha A. (2021). Tomato (Solanum lycopersicum L.) Breeding Strategies for Biotic and Abiotic Stresses. In: Al-Khayri J M, Jain S M, Johnson D V. (eds) Advances in Plant Breeding Strategies: *Vegetable Crops. Springer, Cham.* https://doi.org/10.1007/978-3-319-22521-0

Rothan C, Diouf I, and Causse M. (2019). Trait discovery and editing in tomato. *The Plant Journal 97(1): 73–90.* https://doi.org/10.1111/tpj.14152

Saulo A, Schijlen E, de Jong H, de Ridder D, Smit S, Finkers R, and Visser R. (2014). Exploring genetic variation in the tomato (Solanum section Lycopersicon) clade by whole-genome sequencing. *The Plant Journal 80(1): 136–148.* https://doi.org/10.1111/tpj.12616

Schuelke M. (2000). An Economic Method for the Fluorescent Labeling of PCR Fragments. *Nature Biotechnology, 18(2), 233–234.* https://doi.org/10.1038/72708

Singh K H, Singh L, Parmar N, Kumar S, Nanjundan J, Singh G and Thakur A K. (2022). Molecular characterization and genetic diversity analysis in Indian mustard (Brassica juncea L. Czern & Coss.) varieties using SSR markers. *PLOS ONE 17(8): e0272914.* https://doi.org/10.1371/journal.pone.0272914

Tamura K, Stecher G, and Kumar S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution 38(7): 3022–3027.* https://doi.org/10.1093/molbev/msab120

Tranchida-Lombardo V, Mercati F, Avino M, Punzo P, Fiore M C, Poma I, Patanè, C. (2019). Genetic diversity in a collection of Italian long storage tomato landraces as revealed by SNP markers array. *Plant Biosystems 153(2): 288–297.* https://doi.org/10.1080/11263504.2018.1478900

Van Oosterhout C, Hutchinson W F, Wills D P M and Shipley P. (2004). MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes 4(3): 535–538.* https://doi.org/10.1080/11263504.2018.1478900

Wang T, Zou Q D, Qi S Y, Wang X F, Wu Y Y, Liu N, Zhang Y M, Zhang Z J and Li H T. (2016). Analysis of genetic diversity and population structure in a tomato (Solanum lycopersicum L.) germplasm collection based on single nucleotide polymorphism markers. *Genetics and Molecular Research 15(3): 1–12.*

Xin Z and Chen J. (2012). A high throughput DNA extraction method with high yield and quality. *Plant Methods 8: 26.* https://doi.org/10.1186/1746-4811-8-26

Xu W, Gon 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. *Frontiers in Plant Science 15: p.1445734.* https://doi.org/10.3389/fpls.2024.1445734

Yeh, F. C., R.-C. Yang, T. B. J. Boyle, Z.-H. Ye, and J. X. Mao. (1997). POPGENE: The User-Friendly Shareware for Population Genetic Analysis. *Edmonton: Molecular Biology and Biotechnology Centre, University of Alberta*.

Zamani K, Mohsenpour M, Moslemnejad K, Motamed E, and Shariatpanahi M E. (2025). Genetic Engineering for Enhancing Resistance to Biotic and Abiotic Stresses in Crop Plants in Omics Era. In: Sharma K. (ed) Omics and Genome Editing. *Springer, Cham*. https://doi.org/10.1007/978-3-031-81639-0_7

Zhou R, Wu Z, Cao X and Jiang F L. (2015). Genetic diversity of cultivated and wild tomatoes revealed by morphological traits and SSR markers. *Genetics and Molecular Research 14(4): 13868–13879.* http://dx.doi.org/10.4238/2015.October.29.7

Zhou, R., Z. Wu, F. L. Jiang, and M. Liang. Comparison of gSSR and EST-SSR markers for analyzing genetic variability among tomato cultivars (Solanum lycopersicum L.). *Genet Mol Res* 14, no. 4 (2015): 13184-13194. https://doi.org/10.4238/2015.October.26.14