Tropical Plant Biology (2025) 18:83
https://doi.org/10.1007/s12042-025-09450-z

RESEARCH

Novel InDel Marker Development in Pepper (*Capsicum annuum* L.) by ddRADSeq

Moin Qureshi[1] · Damla Ulusoy[2] · Serkan Kasapoğlu[2] · Hakan Fidan[3] · Engin Yol[1]

Received: 20 May 2025 / Accepted: 17 October 2025
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2025

Abstract

Recent breakthroughs in the genomics of *Capsicum annuum* have broadened molecular genetics research. Especially, marker technologies have contributed to studies of gene identification, genomic mapping, biodiversity, and gene tagging. Insertion-deletion (InDel) polymorphisms function as significant genetic markers for evaluating diversity, population structure, and marker-assisted breeding in crops. In this study, we identified and developed InDel markers in *Capsicum annuum* using double-digest restriction site-associated DNA sequencing (ddRADSeq) approach. A total of 82140 InDel regions were identified across 99 pepper lines, with chr 3 displaying the largest quantity of InDels (9587) and chr 5 the lowest (5765) indicating structural differences. The predominant InDels were single-nucleotide variations, with a declining number with increasing fragment size. According to these results, 88 InDel markers were developed with an indel size > 10 bp. These InDels were allocated over all 12 chromosomes of *Capsicum annuum*, with 75.6% located in intergenic regions and 24.4% were found within gene/transcript regions, including exons and coding sequences (CDS). Among developed InDel markers, 20 markers were chosen for validation with genetic diversity and population structure analysis. The Polymorphic Information Content (PIC) values varied from 0.048 to 0.372, underscoring their capacity to differentiate genetic diversity among lines. STRUCTURE analysis revealed three primary genetic clusters associated with certain fruit types supported by Principal Coordinate Analysis (PCoA) and phylogenetic clustering. The developed markers in this study therefore offer an extensive genetic database for the improvement of breeding and conservation studies.

Keywords Capsicum annuum · InDel markers · DdRADSeq · Marker-assisted selection · Pepper

Introduction

Pepper (*Capsicum* spp.) species are significant horticultural crops globally and are classified under the Solanaceae family. Pepper, originating in Mesoamerica (Kraft et al. 2014), have undergone extensive domestication over millennia. Native to South and Central America, they are commercially significant cash crops valued worldwide for their pungent (hot pepper) and non-pungent (sweet pepper) fruits, used as a spice and vegetable, respectively (Naresh et al. 2018; Duranova et al. 2022). Peppers are cultivated in several climatic locations, demonstrating their tolerance to tropical, subtropical, and temperate conditions (Gopalakrishnan 2007). Globally, their production has consistently increased over the past two decades, rising by 73% from 23.34 million tons in 2000 to 40.29 million tons in 2021 (Chillies and peppers, dry and green) (FAOSTAT 2021). Thirty-one species have been identified under the genus *Capsicum* (Moscone et al. 2007), of which five are domesticated: *C. annuum*, *C. chinense*, *C. frutescens*, *C. baccatum*, and *C. pubescens* (Pickersgill 1997). *C. annuum* is the most extensively grown species, renowned for its extensive phenotypic variation and adaptability to many agricultural settings (Zhang et al. 2016a, b, c). This diversity is prominently observed in the Mediterranean region of Turkiye, which is renowned for its unique pepper landraces, including Carliston, Sivri, and Kil varieties, each possessing distinct morphological and agronomic traits (Toker et al. 2025).

✉ Engin Yol
enginyol@akdeniz.edu.tr

[1] Department of Field Crops, Faculty of Agriculture, Akdeniz University, Antalya 07058, Türkiye

[2] Anamas Tarım Ltd. Şti., Antalya, Türkiye

[3] Department of Plant Protection, Faculty of Agriculture, Akdeniz University, Antalya, Türkiye

Published online: 08 November 2025

The rising need for superior and high-yield pepper cultivars has propelled the use of advanced methods and technologies in plant breeding and genetics. DNA marker technologies have contributed to this domain by allowing accurate genetic analysis and promoting the creation of enhanced cultivars. These technologies provide a dependable method for the quick and precise identification of plant species and their genetic variants (Guo et al. 2019). DNA-based molecular markers are extensively used in many applications, such as biodiversity research, gene tagging, genetic mapping, and marker-assisted selection (MAS) (Nadeem et al. 2017). Over the past three decades, DNA-based markers for *Capsicum* spp. have advanced significantly: early studies used RFLP and AFLP (Lefebvre et al. 1993; Tanksley et al. 1988); PCR-based markers like RAPD and SSR followed (Lanteri et al. 2003; Lee et al. 2004; Minamiyama et al. 2006); and recently, third-generation markers such as SNPs and InDels have become essential for pepper genetic research (Li et al. 2015; Tan et al. 2015). While next-generation sequencing has generated vast genomic resources and expanded molecular marker collections, validating these in silico markers and developing ones suitable for resource-limited infrastructure is essential. Recently, PCR-based InDels have received heightened interest owing to their co-dominant inheritance, repeatability, and user-friendliness (Jain et al. 2019). These attributes render InDels especially appropriate for genetic research and practical implementations in pepper breeding initiatives.

InDels are structural genomic differences resulting from the insertion or deletion of nucleotide sequences. These variants are plentiful and evenly dispersed throughout the genome, making them very valuable for genetic research (Li et al. 2015). Despite the SNP elevating the density of the pepper map to an unparalleled level (Qin et al. 2014; Kim et al. 2014), the overall quantity of PCR-based anchored markers remains constrained (Sugita et al. 2013). InDels are significant for their simplicity, codominance, and easy gel-based identification, with whole genome resequencing accelerating polymorphic InDel detection and enhancing uses in genetic mapping, population structure analysis, and molecular marker development. InDel markers have been developed in different crops such as rice (*Oryza sativa* L.) (Wu et al. 2013), barley (*Hordeum vulgare*) (Zhou et al. 2015), tomato (*Solanum lycopersicum* L.) (Yang et al. 2014), pepper (*Capsicum annuum*) (Li et al. 2015) and Brassica (*Brassica rapa*) (Liu et al. 2013) for possible use in population genetics, taxonomic diagnostics, genetic mapping, and association mapping. Similarly, the availability of pepper genome sequences (Qin et al. 2014; Kim et al. 2014) allows genome-wide InDel detection and development in *Capsicum* spp., utilized in genetic mapping (Li et al. 2015; Tan et al. 2015; Zhang et al. 2016a, b, c), germplasm classification, hybrid identification, and marker-assisted breeding (Li et al. 2015), while enhancing genetic variation understanding and supporting genotype improvement.

Double-digest restriction site-associated DNA sequencing (ddRADSeq) is a next-generation sequencing (NGS) genotyping technique that enables extensive detection of polymorphic markers across the genome. As a variation of restriction site-associated DNA sequencing (RADSeq), ddRADSeq employs two restriction enzymes to digest genomic DNA, generating a reduced-representation library that captures thousands of genetic markers without requiring prior genome sequence information (Peterson et al. 2012; Andrews et al. 2016). This approach is widely used in crop genetics to identify SNPs and InDels, evaluate genetic diversity, and research population structure at a lower cost than whole-genome sequencing (Lachenaud et al. 2018; Adenet et al. 2020; Osorio-Guarin et al. 2020). Various ddRADSeq approach can be used for different species by modifications in enzyme combinations and fragment size selection, making it a flexible and potent instrument for genome-wide marker identification and genotyping in both model and non-model organisms (Peterson et al. 2012).

ddRADSeq has been successfully utilized for InDel marker development in various crops, it has not yet been thoroughly applied to *Capsicum annuum*, underscoring the need for genome-wide detection of these polymorphisms in pepper. This study aims to bridge this gap by utilizing ddRADSeq for the first time in pepper to discover and develop stable and practical InDel markers based on sequencing data from 99 pepper parental lines. These developed markers should be used in breeding initiatives would facilitate the sustainable use of *Capsicum* spp. genetic resources, so assuring the enduring resilience and production of this essential crop.

Materials and methods

Plant material

The research used 99 pepper lines emphasizing the substantial genetic diversity of Mediterranean climate-adapted peppers (Supplement Table S1). The selected lines included Dolma (21 lines), Chilli (5 lines), Charleston (19 lines), Kapia (13 lines), Kil (11 lines), Sivri (24 lines), Jalapeno (2 lines), and Mazamort (4 lines) varieties, chosen for their differences in fruit morphology, flavor profiles (sweet or spicy), and customary culinary applications. Dolma peppers are large, blocky, and sweet, often used for stuffing, whereas Chilli peppers are small, pungent, and extensively utilized in hot sauces and spice mixture. Charleston peppers (*C. annuum var. longum*) possess elongated, curved fruits with mild to moderate heat, while Kapia peppers have thick, tapered

red fruits with a sweet taste, often roasted or consumed fresh. Kil peppers are characterized by their long, curly shape and variable pungency, frequently dried and ground into spice. Sivri peppers, extensively cultivated in Turkey, exhibit long, thin fruits with varying levels of heat, making them adaptable for multiple culinary uses. Jalapeno peppers, originating from Mexico, have thick-walled, moderately spicy fruits commonly consumed fresh or pickled. The rare Mazamort variety, with its triple-lobed fruit shape and mild flavor, represents a niche pepper type with breeding potential. Classification of these lines was performed according to the International Union for the Protection of New Varieties of Plants (UPOV) descriptors, emphasizing fruit morphology, dimensions, coloration, and pungency levels.

DNA extraction and genotyping by ddRADSeq

The seeds from the pepper lines were germinated in a controlled greenhouse. Leaf tissues were collected from plants at the four-leaf stage for genomic DNA extraction. The CTAB procedure (Doyle and Doyle 1990) was used for DNA extraction, subsequently followed by quality evaluation on 1% agarose gels. DNA concentrations were calibrated to 100 ng/µL using lambda DNA as a reference standard. The double digest restriction-site associated DNA sequencing (ddRADSeq) procedure (Peterson et al. 2012) was performed with minor modifications in library preparation, genomic DNA from all samples was cleaved using the restriction enzymes *Vsp*I and *Msp*I. The resultant fragments were isolated by electrophoresis on a 2% agarose gel, and those within the size range of 400–500 bp were chosen for library construction. Illumina 150 bp paired-end sequencing was conducted on the prepared libraries.

Bioinformatic analysis

The pipeline for analysis on genomic data:

1. Demultiplexing: Barcoded reads were demultiplexed with Je-Demultiplex software (Galaxy Version 1.0.1) (Girardot et al. 2016).
2. Quality Control: The quality of the sequence was assessed, and low-quality reads were eliminated using fastp (Galaxy Version 1.0.1) (Chen et al. 2018). High-quality readings were preserved for further processing.
3. Mapping: Clean reads were aligned to the *Capsicum annuum* reference genome (GCF_002878395.1_UCD10Xv1.1_genomic) using Bowtie2 (Langmead and Salzberg 2012) obtained from NCBI genbank.
4. Variant Calling: Variants were identified using FreeBayes (Galaxy Version 1.3.10) (Garrison and Marth 2012) with genotype-specific individual alignment data in BAM format by applying the following parameters: Require at least this coverage to process a site: 8x.
5. VCF filter: SNPs were omitted at this stage using VCFfilter (Galaxy Version 1.0.0) to concentrate only on insertion and deletion polymorphisms.
6. Data Integration: Variant call format (VCF) files containing InDel data were merged via VCFcombine (Galaxy Version 1.0.0).
7. Selection: The consolidated variation file was processed in Microsoft Excel to eliminate duplicate regions and organize the InDels according to their sizes.

Primer design and marker validation

The Integrated Genome Browser (IGB) V10.1.0 (Freese et al. 2016) was used to verify InDel regions which have size > 10 bp. Primer pairs for these locations were constructed using PrimerPlus3 (Untergasser et al. 2007) with the following criteria: primer length of 18–27 bp, GC concentration of 30–70%, melting temperature (Tm) ranging from 49 to 60 °C, and product sizes of 200–600 bp. Primer specificity was confirmed via IGB to remove any duplication. Marker IDs (names) were assigned according to a defined format: Pepper [P] – [Deletion (D)/Insertion (I)] – [Chromosome Number] – [Chromosomal Region]. This nomenclature guarantees uniformity and offers comprehensive details on the kind of variant (deletion or insertion), its chromosomal position, and particular genomic area, enabling accurate identification and interpretation of the markers.

For markers validation, EcoTech PCR Kit was used to conduct polymerase chain reaction (PCR) following the manufacturer's guidelines. The reaction mixtures were formulated to a final volume of 20 µL, including 2 µL of genomic DNA (20 ng/µL), 8 µL of 2x EcoTech Master Mix, 1 µL of each primer (10 µM), and 8 µL of nuclease-free water. The PCR cycling parameters included an initial denaturation at 95 °C for 2.3 min, followed by 35 cycles including denaturation at 95 °C for 45 s, annealing at 50 °C for 20 s, and extension at 60 °C for 50 s. Products were analyzed on 3% agarose gels under UV illumination.

Genetic diversity

Genetic diversity analysis was conducted with GenAlEx V6.5 (Peakall and Smouse 2006). The essential metrics like observed heterozygosity (Ho), anticipated heterozygosity (He), number of allele (Na), and genetic differentiation

(FST) was calculated for pepper population. Polymorphism Information Content (PIC) values were also computed for selected InDel marker using PowerMarker software (Liu and Muse 2005). Principal Coordinate Analysis (PCoA) was performed using GenAlEx to explore the genetic relationships among the pepper lines. For phylogenetic analysis, a genetic distance matrix was generated using PowerMarker software (Liu and Muse 2005) and a phylogenetic tree was constructed with MEGA 12 software (Kumar et al. 2024).

Population structure was laid out with the use of STRUCTURE V2.3.4 (Pritchard et al. 2010) employing the admixture model. The investigation included a burn-in phase of 10,000 iterations and 100,000 MCMC iterations for K values between 2 and 10. The ideal quantity of clusters (K) was established using Structure Selector (Li and Liu 2018). Genotypes were categorized into clusters according to Q-values, and a bar plot depicting the population structure was created using STRUCTUREPLOT V2.0 (Ramasamy et al. 2014).

Results

A total of 334.5 million reads were obtained using 150 bp paired-end sequencing on Illumina HiSeq technology. After quality filtering, 318.76 million high-quality reads were retained (Fig. S1). The average GC content was measured at 39.70% and 37.66% before and after filtering, respectively. The pepper line, ANS121 (Kapia type) exhibited the highest read as 5.21 M, whilst ANS30 (Charleston type) had the lowest read as 0.76 M. Considering GC content, ANS65 representing the Kil type exhibited the highest percentage at 38.72%, whereas ANS118 from the Sivri type displayed the lowest at 36.90%.

The analysis of InDel variations across the genome revealed significant variability in their distribution across chromosomes. A total of 82140 InDel regions were discovered, with varying contributions from deletions and insertions on each chromosome (Table 1). The distribution of InDels across the *Capsicum annuum* genome was analyzed using a 1 Mb window size across 12 chromosomes (Fig. 1). Chr 3 had the highest concentration of InDel, while chr 5 showed the lowest InDel density. Additionally, Chr 3 had the largest quantity of InDel variations, amounting to 9587, including 8158 deletions and 1429 insertions, with a frequency of 35.43 InDels/Mb. Conversely, Chr 5 had the lowest InDel count, with 5765 variations, including 4939 deletions and 826 insertions, along with the lowest frequency of 26.07 InDels/Mb (Table 1)..

The study of InDel variants in the pepper lines indicates a pronounced prevalence of smaller InDels across the dataset. A persistent trend was also seen in which deletions exceeded insertions across all chromosomes. A total of 69419 deletions and 12721 insertions were discovered, with deletions comprising about 85% of the total changes, thus being substantially more prevalent (Table 2). Whereas single-base pair (1 bp) deletions were the most common, accounting for 87.61% of all deletions, whereas 2 bp deletions comprised 6.86%. Agarose-gel-use-suitable deletions, namely those between 10 and 20 bp and over 20 bp, were infrequent, accounting for about 0.89% and 0.19% of the overall deletions, respectively. Likewise, insertions were mostly 1bp in size, accounting for 75.52% of all insertions, with 2 bp insertions comprising 11.37%. Agarose-gel-use-suitable insertions (10–20 bp and >20 bp) were infrequent, comprising 2.65% and 0.20% of the total insertions, respectively. The longest deletion, measuring 29 bp, was identified at five positions: three on Chr 6, one on Chr 10, and one on Chr 12. The largest insertion, spanning 23 bp, was identified at ten different positions across multiple chromosomes. Table 3 represents the InDel found greater than 10 bp across pepper genome. Where chr 3 exhibited the largest number of InDels (150), followed by Chr 11 (115), whilst Chr 7 had the lowest count (70). Chr 3 had the highest number of deletions at 94, whilst Chr 7 had the lowest at 53. Chr 3 had the highest number of insertions (56), whilst Chr 7 showed the lowest value (17).

A total of 88 InDel regions with a length of >10 bp were selected in our study, ensuring their resolution on agarose gel (Supplement Table S2). We designed markers for all these InDels evenly distributed all chromosomes, covering diverse genomic regions. The developed markers represented both insertion and deletion variants, with size changes ranging from 200 bp to 600 bp. PCR amplification of selected markers generated clear and distinct amplicons, demonstrating their reliability for genotyping applications (Fig. S2). The InDels with a length of

Table 1 The distribution of insertion-deletion (InDel) variants across genome

Chromosome	Numbers of Insertion-deletion	Numbers of Deletion	Numbers of Insertion	Frequency (InDels/Mb)
Chromosome 1	8627	7435	1192	33.75
Chromosome 2	6572	5589	983	42.58
Chromosome 3	9587	8158	1429	35.43
Chromosome 4	6806	5818	988	29.40
Chromosome 5	5765	4939	826	26.07
Chromosome 6	6971	5814	1157	30.54
Chromosome 7	6136	5251	885	27.01
Chromosome 8	5865	5092	773	33.75
Chromosome 9	5867	4862	1005	26.78
Chromosome 10	6562	5417	1145	29.60
Chromosome 11	6924	5570	1354	29.35
Chromosome 12	6458	5474	984	27.72
Total	82140	69419	12721	

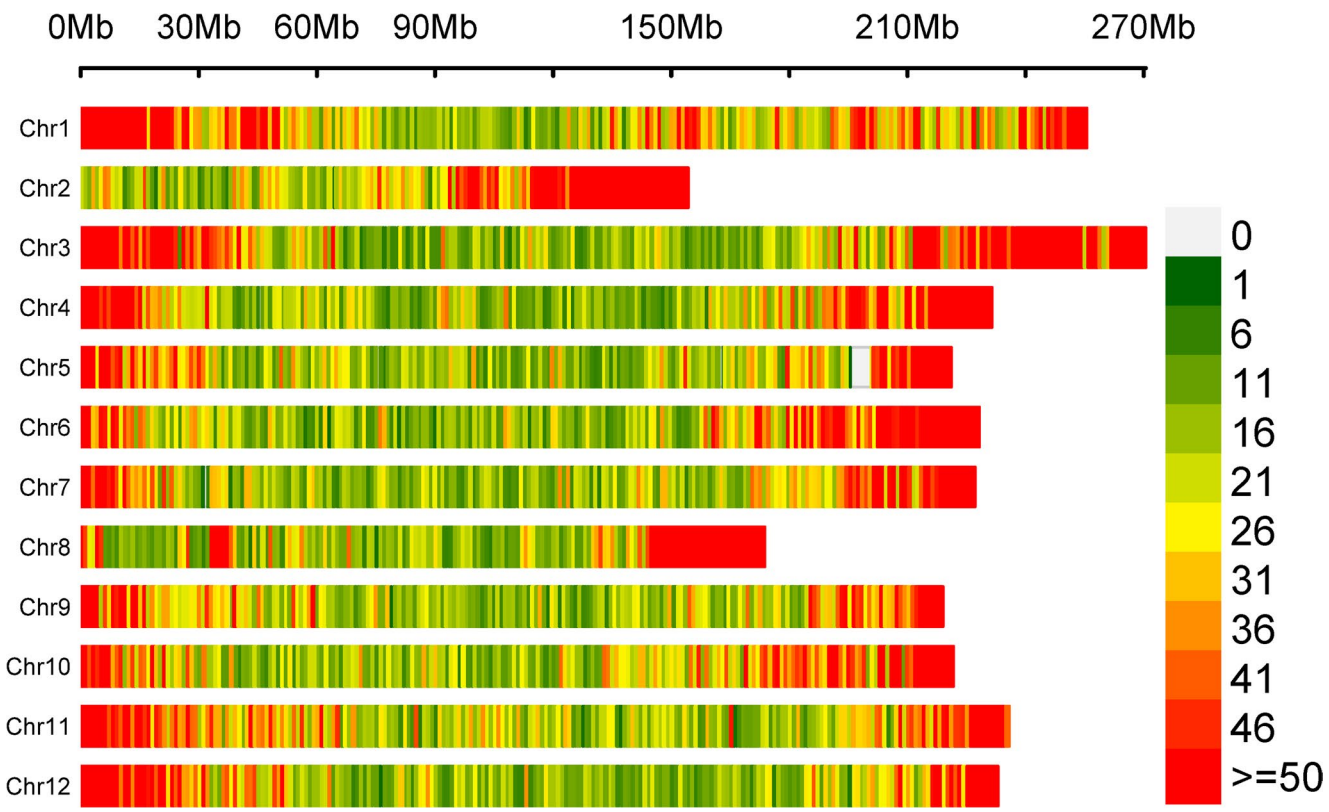


Fig. 1 Distribution of InDels across the *Capsicum annuum* genome using a 1 Mb window size

Table 2 The distribution of insertion-deletion (InDel) variants according to indel size

InDel Size (bp)	Number of Deletion	Frequency (%)	Number of Insertion	Frequency (%)
1	60,821	87.61	9607	75.52
2	4762	6.86	1447	11.37
3	1782	2.57	468	3.68
4	465	0.67	264	2.08
5	272	0.39	166	1.3
6	216	0.31	129	1.01
7	118	0.17	123	0.97
8	105	0.15	74	0.58
9	129	0.19	80	0.63
10–20	615	0.89	337	2.65
>20	134	0.19	26	0.2
Total	69419	100	12721	100

Table 3 Distribution of insertions and deletions of size≥10 bp across genome

Chromosome	Insertion-deletion	Deletion	Insertion
Chr 1	83	59	24
Chr 2	86	59	27
Chr 3	150	94	56
Chr 4	73	47	26
Chr 5	85	54	31
Chr 6	92	59	33
Chr 7	70	53	17
Chr 8	75	55	20
Chr 9	109	73	36
Chr 10	96	58	38
Chr 11	115	83	32
Chr 12	78	55	23
Total	1112	749	363

>10 bp were predominantly located in intergenic regions, accounting for 75.6% of the total identified variants (Table S2). Notably, 20 InDels (24.4%) were found within gene/transcript regions, including exons and coding sequences (CDS). Furthermore, we identified a total of 20 gene-associated InDels across the pepper genome, with notable variants in exons, coding sequences (cds), and gene/transcript regions (Supplement Table S2). Among the 20 genic variants, deletions predominated (17/20) over insertions (3/20).

A recurrent deletion was identified at LOC107856485 on Chromosome 1 (152480720). Chromosome 2 harbored three deletions impacting LOC107858732 (position: 96360218), LOC107858837 (position: 102826450), and LOC107859229 (position: 122975253). On Chromosome 3, both deletions and insertions were found, including deletions in LOC107864880 (position: 245593472), LOC107865844 (position: 257894855), and LOC107865981 (position: 246402008), as well as insertions in LOC107863119 (position: 230024211) and LOC107865811 (position:

267223816). A solitary deletion was found on Chromosome 6 at LOC107873958 (position: 204673151). On Chromosome 9, deletions occurred in LOC107840785 (position: 218725798) and in LOC107842307 at two distinct positions (2167026 and 190695762). Chromosome 10 exhibited deletions in LOC107843714 (position: 212791487) and LOC107843864 (position: 218234605). Finally, Chromosome 11 exhibited the highest variation, with deletions in LOC107847218 (position: 8624050), LOC107847748 (position: 213807012), LOC107872574 (position: 234862974), and LOC124888540 (position: 189705623), alongside an insertion in LOC107847832 (position: 233449175).

Among the developed 88 developed markers, 20 InDel markers were selected for validation with the population and diversity analysis (Table 4). The selected InDel markers (200–500 bp) were distributed across multiple chromosomes and mapped to four distinct pepper varieties—Dolma, Charleston, Kapya, and Sivri (Table 4). Five markers (AS-08, AS-10, AS-13, AS-14, and AS-25) were discovered in Dolma, exhibiting a combination of insertions and deletions with size variations between 18 bp and 19 bp. Charleston type exhibited five markers (AS-28, AS-32, AS-33, AS-35, and AS-46), characterized by predominant deletions and size variations between 12 bp and 29 bp. In Kapya, five markers (AS-54, AS-55, AS-56, AS-60, and AS-63) were identified, including both insertions and deletions, with size variations ranging from −15 bp to 23 bp. Sivri had five markers (AS-87, AS-92, AS-94, AS-97, and AS-99), equally apportioned between insertions and deletions, exhibiting size variations ranging from −11 bp to 29 bp.

Genetic diversity analysis was conducted with the use of selected InDel markers across 20 pepper lines (Table 5). The observed number of alleles (Na) varied from 1.25 to 2.0, with most markers displaying either one or two alleles. The effective number of alleles (Ne) presented a nuanced perspective on allele frequencies, ranging from 1.055 (P-I-AS97-873) to 1.829 (P-D-AS60-787). Shannon's Information Index (I), indicating the genetic variety of the markers, varied from 0.081 (P-I-AS97-873) to 0.635 (P-D-AS60-787), implying moderate to high diversity among the loci. Observed heterozygosity (Ho) values exhibited significant variation, ranging from 0.050 (P-I-AS97-873) to 0.600 (P-D-AS54-556), indicating disparities in the percentage of heterozygous individuals for each marker. Expected heterozygosity (He), indicating the likelihood of two randomly chosen alleles differing, varied from 0.045 (P-I-AS97-873) to 0.445 (P-D-AS60-787). The fixation index (F) values indicated departures from Hardy-Weinberg equilibrium. Negative F values, seen in several markers (e.g., P-I-AS08-994 and P-D-AS32-079). In contrast, positive F values, shown by P-D-AS87-120 (0.344), indicate a deficiency of heterozygotes. Polymorphic Information Content (PIC) values varied from 0.048 (P-I-AS97-873) to 0.372 (P-D-AS54-556 and P-D-AS10-026), reflecting the differential informativeness of the markers. Values of PIC over 0.3 are deemed highly informative, rendering markers such as P-D-AS10-026, P-D-AS54-556, and P-D-AS60-787 very significant for genetic research and breeding initiatives.

The analysis of genetic diversity was also conducted on the basis of pepper commercial types—Dolma, Charleston, Kapya, and Sivri—revealed notable variations in studied parameters (Table 6). The observed number of alleles (Na) ranged from 1.650 in Dolma to 1.700 in Charleston, Kapya, and Sivri. The effective number of alleles (Ne) ranged from 1.315 in Sivri to 1.489 in Charleston. Shannon's Information Index (I) was 0.326 in Sivri and 0.396 in Charleston, with Dolma and Kapya showing values of 0.336 and 0.356, respectively. Observed heterozygosity (Ho) ranged from 0.210 in Dolma to 0.310 in Charleston. Expected heterozygosity (He) varied from 0.207 in Sivri to 0.271 in Charleston, with unbiased expected heterozygosity (uHe) ranging from 0.230 in Sivri to 0.301 in Charleston.

The STRUCTURE analysis was conducted with 20 selected InDel markers across 20 pepper lines belonging to four different commercial types with K = 3, reveals three genetic clusters among the markers, shown by red, green, and blue components in the bar plot shown in Fig. 2. Each bar represents the chance of a certain pepper line belonging to one of these groups. Some pepper type, like Sivri, have distinct genetic differentiation with a prominent blue genetic component, but others, like Charleston, Kapia, and Dolma, reveal varied degrees of mixing. Furthermore, Charleston has stronger genetic mixing than other kinds, indicating increased genetic diversity. The findings correspond with the dendrogram analysis, emphasizing genetic divergence and mixing among the pepper varieties. The Principal Coordinates Analysis (PCoA) plot reveals three distinct genetic clusters among the pepper lines (Fig. 3). Cluster 1 (Dolma) is tightly grouped in the lower-right quadrant, indicating high genetic similarity. Cluster 2 (Charleston and Kapia) forms a central grouping with some overlap, suggesting shared ancestry. Cluster 3 (Sivri) in the lower-left quadrant shows substantial genetic diversity with a wider spread of lines. These findings align with the dendrogram and STRUCTURE analysis, confirming three main genetic groups and highlighting genetic overlap in Charleston and Kapia and variability within Sivri. The UPGMA dendrogram illustrates genetic relationships among 20 pepper lines across four types: Dolma, Charleston, Kapia, and Sivri (Fig. 4). Dolma forms a distinct clade, indicating high genetic uniformity. Charleston shows close relationships but with some divergence. Kapia and Sivri display broader diversity and overlapping clusters. The tree scale (0.009) reflects low overall genetic divergence, indicating close evolutionary relationships.

Table 4 The primer sequences of the 20 indel markers used for genetic diversity analysis

Type	ID	Chromosome	Physical Position	InDel Type	InDel Size (bp)	Marker Name	Forward Primer (5′ to 3′)	Reverse Primer (5′ to 3′)	Product Length (bp)
Dolma	AS08	Chr 8	155,090,994	insertion	19	PIAS08994	TGTTTCTTAGGTAAGCAGTGGGG	TTACACAAAATGAAGCCTGTCATCA	301
Dolma	AS10	Chr 9	2,167,026	deletion	18	PDAS10026	CCTCGAGCATTTCTTCCCCT	GTCAGAGAGTTTGTTTTCACCACA	301
Dolma	AS13	Chr 10	150,479,568	insertion	18	PIAS13568	AGGCATTCTATTGATTTGGAACGC	TCAACTCTCTTTCAGAAGATTTCCCT	301
Dolma	AS14	Chr 4	223,189,071	deletion	14	PDAS14071	ATGCCTGCATTAGATCCATTGC	AATCGTGGTGTCCGGACAAG	406
Dolma	AS25	Chr 8	155,090,994	insertion	19	PIAS25994	GTAAGCAGTGGGGAATGGGT	ACAAGACTATCACACCAAATCGA	318
Charleston	AS28	Chr 6	191,918,103	deletion	29	PDAS28103	ACACGACAATTAATTTGGGATGAA	GGGGCAACATTTTGGATATTTTACG	339
Charleston	AS32	Chr 6	226,327,079	deletion	28	PDAS32079	TGAGGAAACAGCAATAGAATGC	GGGTTGATCTTTTACCTGTCAGG	403
Charleston	AS33	Chr 7	206,601,350	deletion	24	PDAS33350	TCTCGTTTCAGTTTCCGGCT	AGCAGAGTTCTATAGATTTGTATAGGA	311
Charleston	AS35	Chr 11	234,862,976	deletion	12	PDAS35976	AACACCGCCGCCATATACAA	ACTAAAATTACACCTGCAgctattt	362
Charleston	AS46	Chr 2	138,051,702	insertion	13	PIAS46702	TGATTTGGTGGAAAGACGTCA	GCTATGCCTCTTCATTCAGTCC	410
Kapya	AS54	Chr 2	80,328,556	deletion	20	PDAS54556	GGTGTTTATGTGGAATCATCATGTAGA	AGACAAGGCTTTAAGGGTGGATAG	414
Kapya	AS55	Chr 3	0264986911	insertion	15	PIAS55911	GGCATGGTGTTAGGCCGT	ACTTTTCCGCATACAATCTCTTTTGA	341
Kapya	AS56	Chr 3	0264986911	insertion	15	PIAS56911	TTTGGCCAAAGATAAGATTAATCCA	TGCACTATATATCCGCGGATTATTAA	307
Kapya	AS60	Chr 12	221,456,787	deletion	23	PDAS60787	TCCAATGATCCTTAAATCAATGCTT	ACCTTGCTTGCATCAGTCAGA	439
Kapya	AS63	Chr 3	4,760,403	deletion	14	PDAS63403	TCATTTCGTTTGAATGGGTGTGT	TCATTTGAGTTGTTTTCTCTTGCG	248
Sivri	AS87	Chr 10	218,234,605	deletion	14	PDAS87120	CCGGCATGTGATGAGGAGAT	CCAACTCGGATTAATTTCGCCA	356
Sivri	AS92	Chr 1	2,350,468	insertion	11	PIAS92468	GCAAGCTAGCTTCAACACCA	TCTGAAGACCCTTTTGACTTGCT	427
Sivri	AS94	Chr 5	6,819,300	deletion	18	PDAS94300	TAAATGTGTTTGTAAAATCTGCATCCT	ATTCGACCAGTGCTAGTTGC	301
Sivri	AS97	Chr 2	139,621,873	insertion	15	PIAS97873	TCACGAGTTTTGTTCTAGGAGCA	AAGTGGGGGACGACCTATCT	233
Sivri	AS99	Chr 6	197,759,819	deletion	29	PDAS99819	ACGATGTCTTACTACATTTTTGCA	ATTTTGTCGTCTCTCCGGGG	334

Table 5 Genetic diversity parameters of selected indel markers

Markers	Na*	Ne*	I*	Ho*	He*	uHe*	F*	PIC*
P-I-AS08-994	1.500	1.173	0.206	0.150	0.125	0.139	−0.181	0.288
P-D-AS10-026	2.000	1.697	0.570	0.400	0.390	0.433	−0.052	0.372
P-I-AS13-568	1.750	1.290	0.331	0.250	0.205	0.228	−0.204	0.195
P-D-AS14-071	2.000	1.634	0.543	0.350	0.365	0.406	−0.007	0.342
P-I-AS25-994	1.500	1.173	0.206	0.150	0.125	0.139	−0.181	0.288
P-D-AS28-103	2.000	1.710	0.599	0.300	0.410	0.456	0.241	0.332
P-D-AS32-079	1.750	1.529	0.446	0.350	0.305	0.339	−0.171	0.288
P-D-AS33-350	1.500	1.173	0.206	0.150	0.125	0.139	−0.181	0.129
P-D-AS35-976	1.750	1.579	0.462	0.400	0.320	0.356	−0.250	0.305
P-I-AS46-702	1.750	1.227	0.288	0.200	0.170	0.189	−0.157	0.164
P-D-AS54-556	2.000	1.697	0.570	0.600	0.390	0.433	−0.468	0.372
P-I-AS55-911	1.250	1.118	0.125	0.100	0.080	0.089	−0.250	0.090
P-I-AS56-911	1.250	1.118	0.125	0.100	0.080	0.089	−0.250	0.090
P-D-AS60-787	2.000	1.829	0.635	0.550	0.445	0.494	−0.238	0.369
P-D-AS63-403	1.500	1.110	0.163	0.100	0.090	0.100	−0.111	0.090
P-D-AS87-120	2.000	1.477	0.461	0.150	0.295	0.328	0.344	0.359
P-I-AS92-468	1.750	1.354	0.359	0.300	0.230	0.256	−0.263	0.365
P-D-AS94-300	2.000	1.647	0.571	0.350	0.385	0.428	0.122	0.319
P-I-AS97-873	1.250	1.055	0.081	0.050	0.045	0.050	−0.111	0.048
P-D-AS99-819	1.250	1.118	0.125	0.100	0.080	0.089	−0.250	0.090
Mean	1.688	1.385	0.354	0.255	0.233	0.259	−0.131	0.245

* Number of alleles (Na). Number of effective alleles (Ne). Shannon diversity index (I). Expected heterozygosity (He). Unexpected heterozygosity (uHe). Polymorphic information content (PIC)

Table 6 Genetic diversity parameters of four *Capsicum* commercial types with the use of 20 selected indel markers

Types	Na*	Ne*	I*	Ho*	He*	uHe*	F*
Dolma	1.650	1.361	0.336	0.210	0.221	0.246	0.052
Charleston	1.700	1.489	0.396	0.310	0.271	0.301	−0.145
Kapia	1.700	1.376	0.356	0.250	0.233	0.259	−0.102
Sivri	1.700	1.315	0.326	0.250	0.207	0.230	−0.196

* Number of alleles (Na). Number of effective alleles (Ne). Shannon diversity index (I). Expected heterozygosity (He). Unexpected heterozygosity (uHe). Polymorphic information content (PIC)

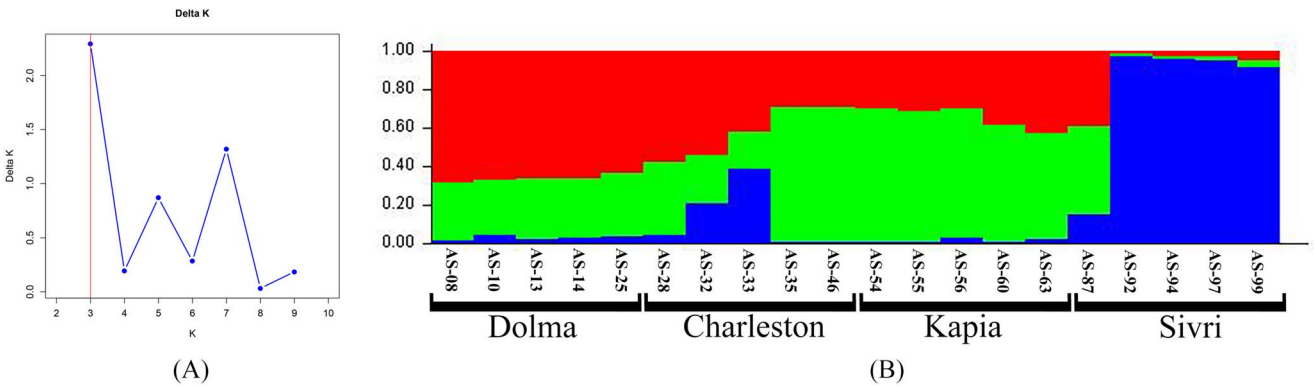


Fig. 2 (**A**) Delta K plot showing the highest peak at K = 3, indicating the presence of three optimal genetic clusters in the studied pepper lines (**B**) STRUCTURE plot (K = 3) depicting genetic clusters and admixture among 20 pepper lines

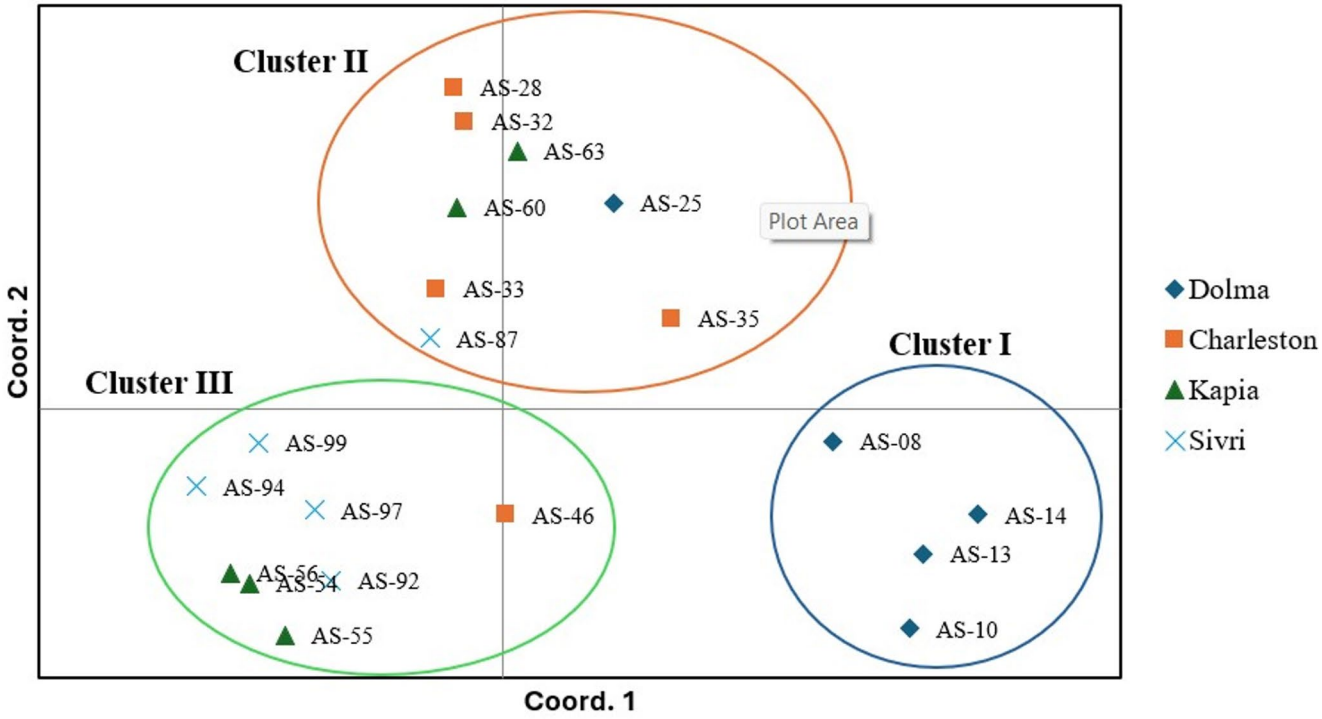


Fig. 3 Principal Coordinates Analysis (PCoA) plot showing genetic diversity and clustering of 20 pepper lines

Discussion

The advancement of next-generation sequencing technologies, like ddRADSeq, has transformed genetic marker discovery by facilitating the detection of genome-wide variants in high density (Pandey et al. 2016). By enabling the identification of DNA polymorphisms, NGS has emerged as an essential instrument for the progression of plant research (Albert and Chang 2014). The ddRADSeq technique, using two restriction enzymes to reduce genomic complexity which reduces the library preparation cost, making it a powerful and economical variant diagnostic (DaCosta and Sorenson 2014). This method was used for sequencing 99 pepper lines, resulting in the discovery of InDel locations across the pepper genome.

In this study, ddRADSeq approach was used to identify InDels in 99 pepper lines acclimated to Mediterranean climatic conditions specially in Turkiye. This methodology produced around 318 million raw sequence reads overall. The read counts for individual lines varied from 0.76 to 5.21 million, yielding a robust dataset for genetic research. The variation in sequencing reads among the 99 pepper lines can be attributed to genetic and structural factors and also technical issues. Moreover, the abundance of repetitive elements, which constitute approximately 81% of the pepper genome, further complicates read alignment and can result in ambiguous data in bioinformatic process (Kim et al. 2014; Lee et al. 2022). This repetitive structure also heightens genomic instability and the probability of deletion events during DNA replication and repair mechanisms (Qin et al. 2014). This evident tendency towards deletions indicates possible selection causes that promote genome compactness, perhaps improving metabolic efficiency or genomic stability (Peterson and Arick 2019). Such deletions may result in loss-of-function mutations, modifying gene expression and inducing phenotypic changes crucial for adaptation (Olsen and Wendel 2013).

The InDel markers demonstrated considerable heterogeneity in distribution, size, and abundance, underscoring the structural diversity of the genome and its relevance for breeding and genetic research. The distribution of InDels across chromosomes reveals significant variability, with Chr 3 displaying the highest number of InDels, whilst Chr 5 recorded the lowest number. Research indicates that deletions are very common in certain chromosomal locations of *Capsicum annuum*, often associated with transposable elements and other genomic alterations (Rudenko and Korotkov 2023). These disparities may indicate localized genomic evolution or differing selection pressures that affect gene function, environmental adaptability, and disease resistance (Cuc et al. 2008). The identified inclination towards deletions underscores their essential function in influencing the structural diversity of the pepper genome. Deletions facilitate evolutionary adaptations by modifying gene expression

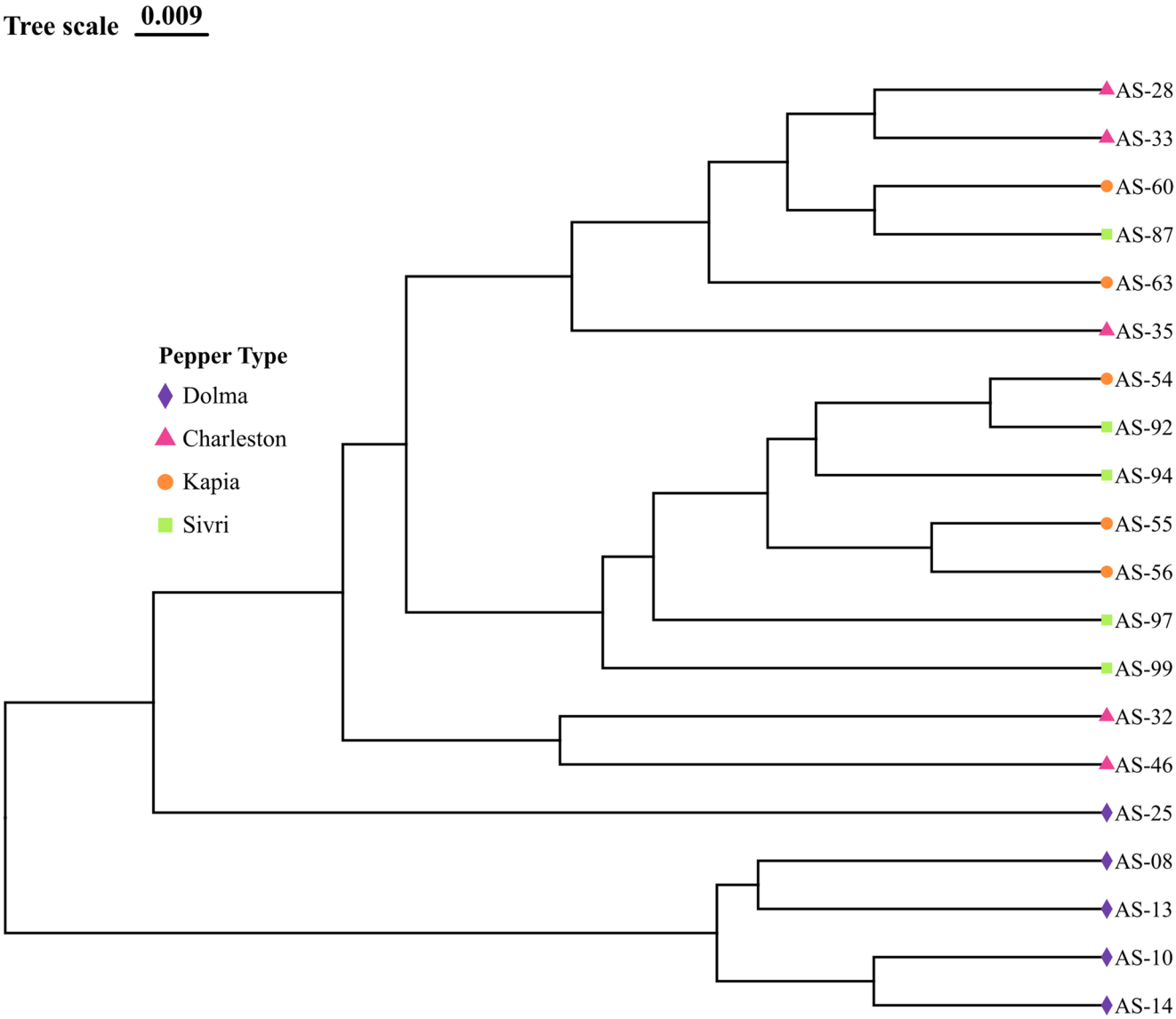


Fig. 4 UPGMA dendrogram showing genetic relationships among 20 pepper lines grouped by type (Dolma, Charleston, Kapia, and Sivri)

and promoting phenotypic variety, which are crucial for domestication and diversification (Gaut et al. 2018). Extensive genomic research has shown that the accumulation of deletions plays a role in the evolutionary history of pepper species, illuminating their genetic diversity and adaptability (Stewart et al. 2005). Furthermore, Petrov (2002) proposed that thermodynamic asymmetry could explain the deletion bias in longer DNA segments. Unlike long insertions, deletions do not require the dissociation of a large portion of already replicated DNA. This makes deletions more thermodynamically favorable than insertions. The variation in InDel distribution across chromosomes underscores the complex nature of genomic architecture and its impact on adaptation and functional divergence.

The distribution of InDels across all 12 chromosomes of pepper exhibited considerable diversity in frequency, aligning with earlier research results. Guo et al. (2019) performed a genome-wide investigation and detected 1,664,770 InDels between PBC688 and CM334, spanning all 12 chromosomes. The variety in InDel distribution highlights the diverse nature of genetic polymorphisms in pepper genomes. Further research has highlighted that InDels are not uniformly distributed, with certain chromosomes exhibiting higher frequencies than others (Fig. 1), reflecting the complex genomic architecture of *Capsicum annuum*. These results highlight the extensive yet unequal distribution of InDels throughout pepper chromosomes, offering significant insights into the genetic diversity and structural differences within pepper.

In our study, Chr 3 displayed the maximum number of InDels, while Chr 5 recorded the lowest total (Table 1). This variability may be affected by chromosome-specific characteristics like size, gene density, and repeat content. The elevated frequency of InDels on Chr 3 may be ascribed to its greater size or the existence of recombination hotspots, whereas the lowered count on Chr 5 indicated a higher ratio of conserved areas with decreased recombination rates (Li et al. 2015). Similar patterns have been observed in previous studies. For instance, Moreno-Contreras et al. (2024) identify variations (SNPs and InDels) between the genomes of two types of peppers and QO (*Capsicum annuum* from Querétaro, Mexico) and CHMX_Ch1 (*Capsicum annuum* from Chihuahua, Mexico) exhibited the highest number of InDels on Chr 3, with 661 and 779 InDels, respectively, while the lowest numbers were detected on Chr 5 in QO (285 InDels) and Chr 11 in CHMX_Ch1 (389 InDels). The size distribution of InDels indicated an abundance of single-nucleotide InDels, with frequency decreasing as base-pair size increased (Table 2). This trend corresponds with findings in other crops, such as maize (Batley et al. 2003), Brassica rapa (Liu et al. 2013), tea (Liu et al. 2019), peanut (Qureshi et al. 2024), sorghum (Guden et al. 2024) and chickpea (Jain et al. 2019). Mechanistically, single-nucleotide insertions and deletions (InDels) are more prevalent owing to their frequent occurrence during DNA replication and their comparatively easier repair methods (Chen and Zhang 2015). Minor modifications may have substantial biological effects, such as interrupting the reading frame of a gene, resulting in frameshift mutations that change the amino acid sequence and possibly impact protein structure and function (Garcia-Diaz and Kunkel 2006). These mutations may lead to phenotypic changes crucial for adaptation and survival. The infrequent incidence of longer InDels (>20 bp) reinforces the idea that these events are uncommon and often face heightened selection pressures because of their capacity to impair gene function (Ramakrishna et al. 2018).

Although less prevalent, larger InDels perform essential functions by initiating significant structural alterations in the genome (Gregory 2004). These alterations may affect regulatory components, disrupt functional domains, or generate new alleles, so influencing the plant's adaptation to environmental conditions or its resistance to infections. The results in *Capsicum annuum* provide a significant resource for comprehending genetic variation and investigating possible uses in breeding programs designed to enhance crop resilience and production.

InDel markers have shown effective as essential tools for studying genetic diversity and for marker-assisted selection in breeding initiatives. A notable benefit of InDels compared to SNPs is their variable length, which offers enhanced flexibility in marker generation and applications (Liu et al. 2023). Additionally, PCR-based InDels offer advantages including co-dominance, cost-effectiveness, and considerable polymorphism, facilitating their straightforward identification via gel electrophoresis (Jain et al. 2019). With this the identified InDels greater than 10 bp selected for additional studies to ensure resilient amplification and repeatability. A set of 88 markers was developed from the identified InDels, focusing on locations with high polymorphism and the right size ranges for efficient PCR amplification. They should be used for breeding efforts such as selection and identification of lines in pepper.

The 88 InDel markers were developed across various genomic regions of *Capsicum annuum*. These markers were distributed over both intergenic and genic regions, including transcripts, exons, and coding sequences. Most InDels (75.6%) occurred in intergenic regions, aligning with prior research indicating that intergenic regions are more susceptible to structural changes owing to less selection constraints. The 20 discovered InDels inside gene/transcript regions, including exons and CDS, are particularly noteworthy owing to their potential to impair gene function and influence phenotypic variance. Our research indicated that most InDel events transpired inside non-coding regions (Supplement Table S2), consistent with the results of Guo et al. (2019). InDels inside functionally significant gene regions, particularly in coding sequence (CDS) locations, may impact gene function by causing frameshifts and altering protein structure (Zhang et al. 2016a, b, c).

The analysis of 20 selected InDel markers demonstrated notable genetic diversity across the studied pepper types, highlighting their genetic distinctiveness and potential for breeding applications (Table 5). The observed number of alleles (Na) varied from 1.250 to 2.000, indicating restricted allelic diversity across the majority of markers, a common characteristic of domesticated crops resulting from selection forces that diminish genetic variety. The effective number of alleles (Ne), varying from 1.055 to 1.829, offered a more detailed perspective, with markers like P-D-AS60-787 demonstrating a balanced allele distribution, highlighting their effectiveness in differentiating closely related genotypes. Markers such as P-D-AS54-556, exhibiting elevated Ho values, indicated persistent recombination or diminished selection pressures, whereas markers with lowered Ho values signified loci subjected to intensified selection or the consequences of inbreeding (Gao and Gao 2016). An analysis of genetic characteristics indicated that the blocky-fruited population had the lowest heterozygosity (Ho), while the horn-fruited group had the greatest Ho. These findings correspond with earlier research indicating less genetic diversity in non-pungent blocky-fruited varieties relative to pungent varieties (Hill et al. 2013; Du et

al. 2019; Solomon et al. 2019; Tam et al. 2009). The low genetic diversity in blocky-fruited types is probably a result of inbreeding within a confined gene pool.

The examination of genetic diversity among several pepper subpopulations yielded significant insights into their genetic architecture. The data in the table indicate that the blocky fruit type (Dolma) displayed the lowest values for several genetic diversity parameters, including number of alleles (Na) at 1.650, effective alleles (Ne) at 1.361, Shannon diversity index (I) at 0.336, observed heterozygosity (Ho) at 0.210, and expected heterozygosity (He) at 0.221. This group had the greatest inbreeding coefficient (0.052). The results indicate that Dolma peppers have less genetic diversity, probably due to rigorous selection practices emphasizing blocky fruit characteristics during cultivation. Conversely, Charleston peppers had the greatest recorded heterozygosity (Ho) at 0.310, coupled with a negative inbreeding coefficient of −0.145, indicating a greater genetic diversity. Kapia and Sivri peppers exhibited considerable variety, shown by intermediate Na values (1.700), I value (0.356 and 0.326, respectively), and negative inbreeding coefficients (−0.102 and − 0.196, respectively). Negative inbreeding coefficients in these subpopulations indicate a propensity for outcrossing, perhaps facilitating the preservation of genetic variety.

These discoveries correspond with the findings documented in other research (Du et al. 2019; Solomon et al. 2019), which indicated that narrow-horn fruits had more genetic diversity. Nonetheless, disparities were seen in the wide-horn fruit variety, where PIC, Ho, and MAF values were surprisingly elevated in comparison to narrow-horn fruits. This underscores the impact of fruit form as a crucial determinant in the genetic architecture of pepper populations. The genetic diversity of blocky fruit varieties has been considerably reduced owing to their heightened selection for breeding driven by commercial demand. Conversely, horn-shaped peppers (both narrow and broad) exhibit more genetic variety, presumably due to their lesser use in breeding schemes. The clustering shown in the phylogenetic tree also indicates that the population structure is intricately related to cultivar shape types. The distinction among the primary cultivar types—Dolma, Charleston, Kapia, and Sivri—within the evolutionary tree illustrates the impact of selected pressures stemming from varied market needs and consumption patterns. This aligns with other research that links genetic diversity to fruit characteristics, consumption preferences, and regional agroclimatic adaptation (González-Pérez et al. 2014; Zhang et al. 2016a, b, c; Du et al. 2019; Pereira-Dias et al. 2019).

Selective breeding for certain features, including fruit size, shape, and taste, has facilitated the domestication and diversification of *Capsicum annuum* cultivars. The phylogenetic tree corroborates the findings from the PCA and model-based population structure studies, demonstrating a strong separation among the four fruit morphologies, although considerable mixing across groups. This discovery corresponds with the recognized fruit form categorization system and highlights the critical importance of breeding efforts that emphasize fruit morphology. These programs have significantly facilitated the differentiation of genetic structures among pepper cultivars.

Conclusion

This development of InDel markers by ddRADSeq have yielded significant insights into the genetic diversity, population structure, and breeding potential of pepper cultivars. The 88 developed InDel markers aimed at distinct genomic areas exhibited exceptional specificity, enabling the accurate identification of hybrids, evaluation of genetic diversity, and marker-assisted selection. The results underscore the impact of fruit form selection on the genetic structure of pepper varieties, with consequences for breeding initiatives focused on preserving and improving *Capsicum* genetic resources.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12042-025-09450-z.

Acknowledgements This study was supported by The Scientific and Technological Research Council of Türkiye (TUBITAK) with the grant number of 5220044. We appreciate the Scientific Research Projects Coordination Unit of Akdeniz University for continuous support.

Author Contributions EY designed the research study. DU and SK collected and provided pepper lines. MQ, DU and EY performed DNA extraction, library preparation, PCR and sequencing analysis. EY and HF performed data analysis. DU and MQ validated markers. MQ and EY wrote original draft of the manuscript. EY and HF review and edit the manuscript. All authors approved the final version of this manuscript.

Funding This study was supported by The Scientific and Technological Research Council of Türkiye (TUBITAK) with the grant number of 5220044.

Data Availability All the data produced in this manuscript are either presented in the main manuscript or as supplementary files. The sequence data used in this manuscript are available from Sequence Read Archives of the National Center for Biotechnology Information with the lines number of PRJNA1129584.

Declarations

Competing Interests The authors declare no competing interests.

References

Adenet S, Regina F, Rogers D, Bharath S, Argout X, Rochefort K, Cilas C (2020) Study of the genetic diversity of cocoa populations (*Theobroma cacao* L.) of Martinique (FWI) and potential

for processing and the cocoa industry. Genet Resour Crop Evol 67:1969–1979. https://doi.org/10.1007/s10722-020-00953-0

Albert VA, Chang TH (2014) Evolution of a hot genome. Proc Natl Acad Sci U S A 111:5069–5070. https://doi.org/10.1073/pnas.1402597111

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet 17:81–92. https://doi.org/10.1038/nrg.2015.28

Batley J, Barker G, O'Sullivan H, Edwards KJ, Edwards D (2003) Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence Tag data. Plant Physiol 132:84–91. https://doi.org/10.1104/pp.102.019422

Chen W, Zhang L (2015) The pattern of DNA cleavage intensity around indels. Sci Rep 5:8333. https://doi.org/10.1038/srep08333

Chen S, Zhou Y, Chen Y, Gu J (2018) Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Cuc LM, Mace ES, Crouch JH, Quang VD, Long TD, Varshney RK (2008) Isolation and characterization of novel microsatellite markers and their application for diversity assessment in cultivated groundnut (*Arachis hypogaea*). BMC Plant Biol 8:1–11. https://doi.org/10.1186/1471-2229-8-55

DaCosta JM, Sorenson MD (2014) Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol. PLoS ONE 9:e106713. https://doi.org/10.1371/journal.pone.0106713

Doyle JJ, Doyle JL (1990) A rapid total DNA preparation procedure for fresh plant tissue. Focus 12:13–15

Du H, Yang J, Chen B, Zhang X, Zhang J, Yang K, Geng S, Wen C (2019) Target sequencing reveals genetic diversity, population structure, core-SNP markers, and fruit shape-associated loci in pepper varieties. BMC Plant Biol 19:1–16. https://doi.org/10.1186/s12870-019-2122-2

Duranova H, Valkova V, Gabriny L (2022) Chili peppers (*Capsicum* spp.): the spice not only for cuisine purposes: an update on current knowledge. Phytochem Rev 21:1379–1413. https://doi.org/10.1007/s11101-021-09789-7

FAOSTAT (2021) About Trade: Crops and livestock products. Accessed 19 Apr 2022

Freese NH, Norris DC, Loraine AE (2016) Integrated genome browser: visual analytics platform for genomics. Bioinformatics 32:2089–2095. https://doi.org/10.1093/bioinformatics/btw069

Gao Z, Gao CW (2016) Lowered diversity and increased inbreeding depression within peripheral populations of wild rice *Oryza rufipogon*. PLoS ONE 11:e0150468. https://doi.org/10.1371/journal.pone.0150468

Garcia-Diaz M, Kunkel TA (2006) Mechanism of a genetic glissando*: structural biology of indel mutations. Trends Biochem Sci 31:206–214. https://doi.org/10.1016/j.tibs.2006.02.003

Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. ArXiv Preprint. https://doi.org/10.48550/arXiv.1207.3907. arXiv:1207.3907

Gaut BS, Seymour DK, Liu Q, Zhou Y (2018) Demography and its effects on genomic variation in crop domestication. Nat Plants 4:512–520. https://doi.org/10.1038/s41477-018-0210-1

Girardot C, Scholtalbers J, Sauer S, Su SY, Furlong EE (2016) Je, a versatile suite to handle multiplexed NGS libraries with unique molecular identifiers. BMC Bioinformatics 17:1–6. https://doi.org/10.1186/s12859-016-1284-2

González-Pérez S, Garcés-Claver A, Mallor C, Saenz de Miera LE, Fayos O, Pomar F, Merino F, Silvar C (2014) New insights into *Capsicum* spp. Relatedness and the diversification process of *Capsicum annuum* in Spain. PLoS ONE 9:e116276. https://doi.org/10.1371/journal.pone.0116276

Gopalakrishnan TR (2007) Solanaceous fruit vegetables. Vegetable Crops 63–102

Guden B, Kiemde O, Yol E, Uzun B (2024) Development and utilization of genome-wide indel markers in sorghum (*Sorghum bicolor* L. Moench). Euphytica 220:75. https://doi.org/10.1007/s10681-024-03335-0

Guo G, Zhang G, Pan B, Diao W, Liu J, Ge W, Gao C, Zhang Y, Jiang C, Wang S (2019) Development and application of indel markers for *Capsicum* spp. based on whole-genome re-sequencing. Sci Rep 9:3691. https://doi.org/10.1038/s41598-019-40244-y

Gregory TR (2004) Insertion-deletion biases and the evolution of genome size. Gene 324:15–34. https://doi.org/10.1016/j.gene.2003.09.030

Hill TA, Ashrafi H, Reyes-Chin-Wo S, Yao J, Stoffel K, Truco MJ, Kozik A, Michelmore RW, Van Deynze A (2013) Characterization of *Capsicum annuum* genetic diversity and population structure based on parallel polymorphism discovery with a 30K unigene pepper genechip. PLoS ONE 8:e56200. https://doi.org/10.1371/journal.pone.0056200

Jain A, Roorkiwal M, Kale S, Garg V, Yadala R, Varshney RK (2019) Indel markers: an extended marker resource for molecular breeding in chickpea. PLoS ONE 14:e0213999. https://doi.org/10.1371/journal.pone.0213999

Kim S, Park M, Yeom SI, Kim YM, Lee JM, Lee HA, Seo E, Choi J, Cheong K, Kim KT, Jung K (2014) Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. Nat Genet 46:270–278. https://doi.org/10.1038/ng.2877

Kraft KH, Brown CH, Nabhan GP, Luedeling E, Luna Ruiz JDJ, Coppens d'Eeckenbrugge G, Robert JH, Gepts P (2014) Multiple lines of evidence for the origin of domesticated Chili pepper, *Capsicum annuum*, in Mexico. Proc Natl Acad Sci U S A 111:6165–6170. https://doi.org/10.1073/pnas.1308933111

Kumar S, Stecher G, Suleski M, Sanderford M, Sharma S, Tamura K (2024) MEGA12: molecular evolutionary genetic analysis version 12 for adaptive and green computing. Mol Biol Evol 41:msae263. https://doi.org/10.1093/molbev/msae263

Lachenaud P, Clément D, Argout X, Scalabrin S, Doare F (2018) The Guiana cacao genetic group (*Theobroma Cacao* L.): a new core collection in French Guiana. Bot Lett 165:248–254. https://doi.org/10.1080/23818107.2018.1465466

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923

Lanteri S, Acquadro A, Quagliotti L, Portis E (2003) RAPD and AFLP assessment of genetic variation in a landrace of pepper (*Capsicum annuum* L.) grown in north-west Italy. Genet Resour Crop Evol 50:723–735. https://doi.org/10.1023/A:1025075118200

Lee JM, Nahm SH, Kim YM, Kim B (2004) Characterization and molecular genetic mapping of microsatellite loci in pepper. Theor Appl Genet 108:619–627. https://doi.org/10.1007/s00122-003-1467-x

Lee JH, Venkatesh J, Jo J, Jang S, Kim GW, Kim JM, Han K, Ro N, Lee HY, Kwon JK, Kim YM (2022) High-quality chromosome-scale genomes facilitate effective identification of large structural variations in hot and sweet peppers. Hortic Res 9:uhac210. https://doi.org/10.1093/hr/uhac210

Lefebvre V, Palloix A, Rives M (1993) Nuclear RFLP between pepper cultivars (*Capsicum annuum* L). Euphytica 71:189–199. https://doi.org/10.1007/BF00040408

Li YL, Liu JX (2018) Structureselector: a web-based software to select and visualize the optimal number of clusters using multiple methods. Mol Ecol Resour 18:176–177. https://doi.org/10.1111/1755-0998.12719

Li W, Cheng J, Wu Z, Qin C, Tan S, Tang X, Cui J, Zhang L, Hu K (2015) An InDel-based linkage map of hot pepper (*Capsicum annuum*). Mol Breed 35:1–10. https://doi.org/10.1007/s11032-015-0219-3

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129. https://doi.org/10.1093/bioinformatics/bti282

Liu B, Wang Y, Zhai W, Deng J, Wang H, Cui Y, Cheng F, Wang X, Wu J (2013) Development of indel markers for *Brassica rapa* based on whole-genome re-sequencing. Theor Appl Genet 126:231–239. https://doi.org/10.1007/s00122-012-1976-6

Liu S, An Y, Tong W, Qin X, Samarina L, Guo R, Xia X, Wei C (2019) Characterization of genome-wide genetic variations between two varieties of tea plant (*Camellia sinensis*) and development of indel markers for genetic research. BMC Genomics 20:1–16. https://doi.org/10.1186/s12864-019-6347-0

Liu Z, Zhao Y, Zhang Y, Xu L, Zhou L, Yang W, Zhao H, Zhao J, Wang F (2023) Development of omni indel and supporting database for maize. Front Plant Sci 14:1216505. https://doi.org/10.3389/fpls.2023.1216505

Minamiyama Y, Tsuro M, Hirai M (2006) An SSR-based linkage map of *Capsicum annuum*. Mol Breed 18:157–169. https://doi.org/10.1007/s11032-006-9024-3

Moreno-Contreras VI, Delgado-Gardea MCE, Ramos-Hernández JA, Mendez-Tenorio A, Varela-Rodríguez H, Sánchez-Ramírez B, Muñoz-Ramírez ZY, Infante-Ramírez R (2024) Genome-wide identification and characterization of SNPs and indels of *Capsicum annuum* var. *Glabriusculum* from Mexico based on whole genome sequencing. Plants 13:3248. https://doi.org/10.3390/plants13223248

Moscone EA, Scaldaferro MA, Grabiele M, Cecchini NM, Sánchez García Y, Jarret R, Daviña JR, Ducasse DA, Barboza GE, Ehrendorfer F (2007) The evolution of Chili peppers (Capsicum-Solanaceae): a cytogenetic perspective. Acta Hortic 745:137

Nadeem MA, Nawaz MA, Shahid MQ, Doğan Y, Comertpay G, Yıldız M, Hatipoğlu R, Ahmad F, Alsaleh A, Labhane N, Ozkan H, Chung G, Baloch FS (2017) DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. Biotechnol Biotechnol Equip 32(2):261–285. https://doi.org/10.1080/13102818.2017.1400401

Naresh P, Lin SW, Lin CY, Wang YW, Schafleitner R, Kilian A, Kumar S (2018) Molecular markers associated to two non-allelic genic male sterility genes in peppers (*Capsicum annuum* L). Front Plant Sci 9:1343. https://doi.org/10.3389/fpls.2018.01343

Olsen KM, Wendel JF (2013) Crop plants as models for understanding plant adaptation and diversification. Front Plant Sci 4:290. https://doi.org/10.3389/fpls.2013.00290

Osorio-Guarín JA, Berdugo-Cely JA, Coronado-Silva RA, Baez E, Jaimes Y, Yockteng R (2020) Genome-wide association study reveals novel candidate genes associated with productivity and disease resistance to *Moniliophthora* spp. in cacao (*Theobroma cacao* L). G3 Genes|Genomes|Genetics 10:1713–1725. https://doi.org/10.1534/g3.120.401153

Pandey MK, Roorkiwal M, Singh VK, Ramalingam A, Kudapa H, Thudi M, Chitikineni A, Rathore A, Varshney RK (2016) Emerging genomic tools for legume breeding: current status and future prospects. Front Plant Sci 7:455. https://doi.org/10.3389/fpls.2016.00455

Peakall ROD, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295. https://doi.org/10.1111/j.1471-8286.2005.01155.x

Pereira-Dias L, Vilanova S, Fita A, Prohens J, Rodríguez-Burruezo A (2019) Genetic diversity, population structure, and relationships in a collection of pepper (*Capsicum* spp.) landraces from the Spanish centre of diversity revealed by genotyping-by-sequencing (GBS). Hortic Res 6:54. https://doi.org/10.1038/s41438-019-0132-8

Peterson DG, Arick M (2019) Sequencing plant genomes. Prog Bot 80:109–193. https://doi.org/10.1007/124_2018_18

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest radseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS ONE 7:e37135. https://doi.org/10.1371/journal.pone.0037135

Petrov DA (2002) Mutational equilibrium model of genome size evolution. Theor Popul Biol 61:531–544. https://doi.org/10.1006/tpbi.2002.1595

Pickersgill B (1997) Genetic resources and breeding of *Capsicum* spp. Euphytica 96:129–133

Pritchard JK, Wen X, Falush D (2010) Documentation for structure software: version 2.3, vol 1. Univ Chicago, Chicago, IL, p 37

Qin C, Yu C, Shen Y, Fang X, Chen L, Min J, Cheng J, Zhao S, Xu M, Luo Y, Yang Y (2014) Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. Proc Natl Acad Sci U S A 111:5135–5140. https://doi.org/10.1073/pnas.1400975111

Qureshi M, Guden B, Tosun HŞ, Kordrostami M, Gangurde SS, Uzun B, Yol E (2024) Selective indel marker identification across the peanut (*Arachis hypogaea* L.) genome using DdRADSeq. Turk J Bot 48:441–453. https://doi.org/10.55730/1300-008X.2826

Ramakrishna G, Kaur P, Nigam D, Chaduvula PK, Yadav S, Talukdar A, Singh NK, Gaikwad K (2018) Genome-wide identification and characterization of indels and SNPs in *Glycine max* and *Glycine Soja* for contrasting seed permeability traits. BMC Plant Biol 18:1–15. https://doi.org/10.1186/s12870-018-1321-2

Ramasamy RK, Ramasamy S, Bindroo BB, Naik VG (2014) STRUCTURE PLOT: a program for drawing elegant STRUCTURE bar plots in a user-friendly interface. SpringerPlus 3:1–3. https://doi.org/10.1186/2193-1801-3-431

Rudenko V, Korotkov E (2023) Detection of tandem repeats in the *Capsicum annuum* genome. DNA Res 30:dsad007. https://doi.org/10.1093/dnares/dsad007

Solomon AM, Han K, Lee JH, Lee HY, Jang S, Kang BC (2019) Genetic diversity and population structure of Ethiopian *Capsicum* germplasms. PLoS ONE 14:e0216886. https://doi.org/10.1371/journal.pone.0216886

Stewart C, Kang BC, Liu K, Mazourek M, Moore SL, Yoo EY, Kim BD, Paran I, Jahn MM (2005) The *Pun1* gene for pungency in pepper encodes a putative acyltransferase. Plant J 42:675–688. https://doi.org/10.1111/j.1365-313X.2005.02410.x

Sugita T, Semi Y, Sawada H, Utoyama Y, Hosomi Y, Yoshimoto E, Maehata Y, Fukuoka H, Nagata R, Ohyama A (2013) Development of simple sequence repeat markers and construction of a high-density linkage map of *Capsicum annuum*. Mol Breed 31:909–920. https://doi.org/10.1007/s11032-013-9844-x

Tam SM, Lefebvre V, Palloix A, Sage-Palloix AM, Mhiri C, Grandbastien MA (2009) LTR-retrotransposons Tnt1 and T135 markers reveal genetic diversity and evolutionary relationships of domesticated peppers. Theor Appl Genet 119:973–989. https://doi.org/10.1007/s00122-009-1102-6

Tan S, Cheng JW, Zhang L, Qin C, Nong DG, Li WP, Tang X, Wu ZM, Hu KL (2015) Construction of an interspecific genetic map based on indel and SSR for mapping the QTLs affecting the initiation of flower primordia in pepper (*Capsicum* spp). PLoS ONE 10:e0119389. https://doi.org/10.1371/journal.pone.0119389

Tanksley SD, Bernatzky R, Lapitan NL, Prince JP (1988) Conservation of gene repertoire but not gene order in pepper and tomato. Proc Natl Acad Sci U S A 85:6419–6423. https://doi.org/10.1073/pnas.85.17.6419

Toker TP, Ulusoy D, Doğan B, Kasapoğlu S, Fidan H, Reddy UK, Kordrostami M, Yol E et al (2025) Genomic insights into mediterranean pepper diversity using DdRADSeq. PLoS ONE 20(2):e0318105. https://doi.org/10.1371/journal.pone.0318105

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA (2007) Primer3plus, an enhanced web interface to Primer3.

Nucleic Acids Res 35:W71–W74. https://doi.org/10.1093/nar/gkm306

Wu DH, Wu HP, Wang CS, Tseng HY, Hwu KK (2013) Genome-wide indel marker system for application in rice breeding and mapping studies. Euphytica 192:131–143. https://doi.org/10.1007/s10681-013-0925-z

Yang J, Wang Y, Shen H, Yang W (2014) *In Silico* identification and experimental validation of insertion–deletion polymorphisms in the tomato genome. DNA Res 21:429–438. https://doi.org/10.1093/dnares/dsu008

Zhang JZ, Liu SR, Hu CG (2016a) Identifying the genome-wide genetic variation between precocious trifoliate orange and its wild type and developing new markers for genetics research. DNA Res 23(4):403–414. https://doi.org/10.1093/dnares/dsw017

Zhang XF, Sun HH, Xu Y, Chen B, Yu SC, Geng SS, Wang Q (2016b) Development of a large number of SSR and indel markers and construction of a high-density genetic map based on a RIL population of pepper (*Capsicum annuum* L). Mol Breed 36:1–10. https://doi.org/10.1007/s11032-016-0517-4

Zhang XM, Zhang ZH, Gu XZ, Mao SL, Li XX, Chadœuf J, Palloix A, Wang LH, Zhang BX (2016c) Genetic diversity of pepper (*Capsicum* spp.) germplasm resources in China reflects selection for cultivar types and spatial distribution. J Integr Agric 15:1991–2001. https://doi.org/10.1016/S2095-3119(16)61364-3

Zhou G, Zhang Q, Tan C, Zhang XQ, Li C (2015) Development of genome-wide indel markers and their integration with SSR, DArT, and SNP markers in a single barley map. BMC Genomics 16:1–8. https://doi.org/10.1186/s12864-015-2027-x

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.