3 Biotech (2023) 13:136
https://doi.org/10.1007/s13205-023-03563-6

ORIGINAL ARTICLE

Genome-wide characterization and development of SSR markers for genetic diversity analysis in northwestern Himalayas Walnut (*Juglans regia* L.)

H. Itoo[1] · Rafiq Ahmad Shah[1] · S. Qurat[2] · Afnan Jeelani[2] · Sheikh Khursheed[1] · Zahoor A. Bhat[1] · M. A. Mir[1] · G. H. Rather[1] · Sajad Majeed Zargar[3] · M. D. Shah[4] · Bilal A. Padder[4]

Received: 20 September 2022 / Accepted: 15 April 2023 / Published online: 25 April 2023

Abstract

In the present study, we designed and validated genome-wide polymorphic SSR markers (110 SSRs) by mining the walnut genome. A total of 198,924 SSR loci were identified. Among these, successful primers were designed for 162,594 (81.73%) SSR loci. Dinucleotides were the most predominant accounting for 88.40% (175,075) of total SSRs. The SSR frequency was 377.312 SSR/Mb and it showed a decreasing trend from dinucleotide to octanucleotide motifs. We identified 20 highly polymorphic SSR markers and used them to genotype 72 walnut accessions. Over all, we obtained 118 alleles that ranged from 2 to 12 with an average value of 5.9. The higher SSR PIC values indicate their robustness in discriminating walnut genotypes. Heat map, PCA, and population structure categorized 72 walnut genotypes into 2 distinct clusters. The genetic variation within population was higher than among population as inferred by analysis of molecular variance (AMOVA). For walnut improvement, it is necessary to have a large repository of SSRs with high discriminative power. The present study reports 150,000 SSRs, which is the largest SSR repository for this important nut crop. Scientific communities may use this repository for walnut improvement such as QTL mapping, genetic studies, linkage map construction, and marker-assisted selection.

Keywords *Juglans regia* · Genome-wide SSR markers · SSR validation · Walnut diversity · In silico PCR · Kashmir

Introduction

The walnut (*Juglans regia* L.) belongs to the Fagales order and family Juglandaceae (McGranahan and Leslie 2009; Hussain et al. 2021; Zaini et al. 2020). It is a widely cultivated nut crop in temperate regions of the world, including Indian state of Jammu and Kashmir (Shah et al. 2021, 2018). Walnuts are nutritionally dense and is considered as "bread of the future" (Jaćimović et al. 2020; Turdieva et al. 2012). Presently, China is a key walnut-producing country with share of 43.31% global production. The United States, Iran, Turkey, Mexico, and India each contribute 16.74%, 11.19%, 5.87%, 4.35%, and 0.88% of the global walnut output, respectively. Walnut is a native to Eurasia, growing from the Balkans to Southwest China (Aradhya et al. 2017; Feng et al. 2018; Khadivi-Khub et al. 2015; Pollegioni et al. 2004). All the species of genus *Juglans* are diploid with a karyotype of $2n = 32$ and have 16 linkage groups (Kefayati et al. 2017). There have been numerous previous attempts to create genetic linkage maps using RAPD, RFLP, and

✉ Bilal A. Padder
bapadder@rediffmail.com

1 Ambri Apple Research Centre, Pahnoo Shopian, Sheri-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, Srinagar, J&K 192303, India

2 Division of Fruit Science, Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, Faculty of Horticulture, Shalimar, Kashmir, Srinagar, J&K 190 025, India

3 Proteomics Laboratory, Division of Plant Biotechnology, Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, Faculty of Horticulture, Shalimar, Kashmir, Srinagar, J&K 190 025, India

4 Plant Virology and Molecular Plant Pathology Laboratory, Division of Plant Pathology, Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, Shalimar, Kashmir, 190 025 Srinagar, J&K, India

isozyme markers (Fjellstrom and Parfitt 1994; Woeste et al. 1996; Malvolti et al. 2001). However, there were not sufficient markers to cover all of the linkage groups, and some of the linked markers lacked sequence information. Recently, SNP and InDel markers were also used to construct genetic map of walnut (Zhu et al. 2015; Luo et al. 2015). The first SSR-based linkage map was constructed by Kefayati et al. (2017) with consensus map length of 1569.9 cM. Availability of walnut genome (Martínez-García et al. 2016) opened many frontier areas including fine mapping of economic traits (Bernard et al. 2019; Marrano et al. 2019a; Ji et al. 2021) and cracking of other Juglans species genomes (Stevens et al. 2018). The walnut genome version 1.0 (Martínez-García et al. 2016) was highly fragmented and was significantly improved in v1.5 genome assembly (Stevens et al. 2018). Recently, a high-quality chromosome-scale assembly (Chandler v2.0) helped to explain the complex biological processes in walnut (Marrano et al. 2020). This high-quality genome assembly was obtained by combining Oxford Nanopore long read sequencing with chromosome conformation capture (Hi-C) technology. A few genomic studies indicate that walnuts grown in South Asian countries particularly the Pakistani and Indian populations are ancestral (Aradhya et al. 2017; Bernard et al. 2020a; Gaisberger et al. 2020; Roor et al. 2017). However, new phylogenomic studies reveal the hybrid origin of *J. regia* (Zhang et al. 2019). In Jammu and Kashmir, the crop has not been exploited for any intensive breeding program; therefore, the natural population possess highest genetic diversity (Shah et al. 2022). To effectively harness the walnut latent potential, it is essential to accurately recognize high genetic diversity to breed new genitors and superior cultivars (Doğan et al. 2014). Phenotypic trait evaluation is a common approach to assess walnut diversity. However, such investigations are inefficient, expensive, and difficult to assess directly for complex polygenic traits (Nickravesh et al. 2023). These issues have been resolved by the development of DNA-based markers, which provide reliable results regardless of the external environment (Shah et al. 2020). Among the molecular markers, microsatellites (1–6 bp in length) are most reliable (Grover et al. 2007; Taheri et al. 2018), which are abundant and well distributed throughout the nuclear genome of eukaryotes (Kalia et al. 2011). Microsatellites are powerful and informative markers for assessing the genetic diversity, finding the relationships among different germplasm populations, linkage map construction, validate walnut scions, and source plants for reliable propagation and to investigate biotic or abiotic stresses (Ali Khan et al. 2016; Bernard et al. 2020a, b, 2019; Shah et al. 2018, 2020; Pollegioni et al. 2017; Doğan et al. 2014; Nickravesh et al. 2023). The genetic diversity in *J. regia* was first studied by Woeste et al. (2002) followed by other researchers (Bai et al. 2010; Chen et al. 2014; Dangl et al. 2005; Foroni et al. 2005, 2007; Hoban et al. 2008; Magige et al. 2022; Najafi et al. 2014; Robichaud et al. 2006, 2010; Ross-Davis et al. 2008; Topçu et al. 2015; Victory et al. 2006; Zhang et al. 2010, 2013). The first set of 13 SSR markers developed from *J. regia* was developed by Najafi et al. (2014). Second set of 94 SSR markers for walnut was developed by Topcu et al. (2015) and out of which only 19 SSRs markers were polymorphic. Topçu et al. (2015) developed another 276 SSR makers from enriched repeat region of genomic libraries. Among these, 185 SSR markers were polymorphic. In spite of the fact that molecular markers aid in deciphering *Juglans* species' population structure and differentiation (Victory et al. 2006; Foroni et al. 2005, 2007; Ross-Davis et al. 2008; Woeste et al. 2002), very few SSR have been developed so far. Although these SSRs have been routinely used to infer the walnut population structure (Bernard et al. 2020a, b; Wang et al. 2008; Ebrahimi et al. 2016), the number is less to construct dense linkage map, marker trait association studies, and QTL mapping. Recently walnut SNP chip, currently the largest chip available in crops, was developed by Marrano et al. (2019b) and is in vogue to map the complex traits (Marrano et al. 2019b; Arab et al. 2019, 2022; Bükücü et al. 2020; Sideli et al. 2020). However, it is difficult to access this chip by the scientific community from developing nations. Alternatively, SSRs being neutral can be used by the labs that do not have high-throughput genomics setup. The best and easiest way to develop large number of SSR markers is to use publicly available walnut genome (Martínez-García et al. 2016). With the aid of bioinformatics workflows, it is easy to mine huge number of genome-wide SSR markers. Many researchers exploited the genomic information to mine genome-wide SSR markers in different plant species in the past decade. For instance, genome-wide SSR markers were developed using bioinformatic approaches in pear (Liu et al. 2015), citrus (Hou et al. 2014), pomegranate (Patil et al. 2020b), spinach (Patil et al. 2020b), Lilium (Biswas et al. 2020), capsicum (Cheng et al. 2016), watermelon (Zhu et al. 2016), and Palmae (Manee et al. 2020).

To date there are only 1300 SSRs available for walnut (Foroni et al. 2005, 2007; Chen et al. 2014; Dangl et al. 2005; Hoban et al. 2008; Najafi et al. 2014; Robichaud et al. 2006; Ross-Davis et al. 2008; Topçu et al. 2015; Victory et al. 2006; Woeste et al. 2002; Zhang et al. 2010), hence we explored publicly available chandler genome to mine genome-wide SSR markers. We report a new set of 162,594 genome-wide SSR markers. Preliminary wet lab studies show that our SSRs are robust with high discriminatory power. Using these SSR markers, we found high diversity in walnut populations from northern India. Our SSR repository will help the scientific community actively working on walnut to saturate linkage map, phylogenetic analysis, and to map economically important traits. Further, this set will help to deduce the population structure of *Juglans* species as most of these SSR markers will show cross transferability.

Materials and methods

Genome-wide SSR mining

Walnut genome (Cv. Chandler) is publicly available at NCBI [*Juglans regia* (ID 17683)—Genome—NCBI (nih.gov)] and we downloaded it in a local server. We used GMATA v 2.0 tool (https://sourceforge.net/projects/gmata) to scan genome-wide SSRs markers as described previously (Wang and Wang 2016; Bhat et al. 2018). To design primers, standalone primer3 was used in batch mode with the following parameters: product size 140–400 bp; primer length 19–25 bp with optimal length 22 bp; primer annealing temperature with optimal Tm 60 °C; and primer must be at least 200 bp away from the microsatellite locus. To calculate amplicon size and number of alleles, we used standalone electronic PCR (e-PCR) module with default parameters. All text handlings were performed using in-house perl scripts.

Selection of plant material and DNA extraction

We collected young leaves from 72 walnut genotypes, that included 60 from Shopian (SW), 8 from Anantnag (AW), and 4 from Pulwama (PW). These populations were selected based on important growing districts of Kashmir, highly diverse agro-ecosystems, and high phenotypic plasticity. The sampling locations are geographically separated from each other (Fig. S1). The plants were selected based on the crucial morphological and pomological traits to include highly diverse genotypes for genotyping (Shah et al. 2021). The plant seedlings thrive in their natural habitat without the use of any management techniques. The genomic DNA was isolated using the CTAB technique (Doyle and Doyle 1987). RNase treatment was used to further purify the extract. On a 1% agarose gel, the DNA's purity was examined and DNA was quantified using a bio spectrometer (Eppendorf, Germany).

Validation of selected SSRs

A set of 110 SSR primers were selected from unique 136,582 SSR markers showing single allele in e-PCR and validated on 10 highly diverse samples that were chosen from geographically isolated places. For instance, From Shopian population, we selected four samples that were at least 200 km apart. Similarly, we selected walnut genotypes from other two districts. The criterion of selecting 110 SSR markers among a large SSR repository was based on the number of repeat motifs. The markers which failed to amplify or produced monomorphic fragments were discarded. From these, 35 markers were selected for validation to find out the highly polymorphic ones. Fifteen markers out of thirty-five markers although were polymorphic but produced low-resolution bands, thus were discontinued for fingerprinting. PCR amplification was carried out in 0.2 ml PCR tubes in a thermal cycler from Biometra T gradient (Gottingen, Germany) using 2 µl of genomic DNA (25 ng/µl), 1U of *Taq* polymerase (Thermo Scientific), 1.5 µl of 10 X Taq polymerase buffer, 1.5 mM $MgCl_2$, 200 µM of each dNTP, 0.4 µM of each primer, and 8.30 µl of deionized water in a final volume of 15 µl reaction. We used following temperature regimes; initial denaturation for 5 min at 94 °C, followed by 35 cycles of denaturation for 1 min at 94 °C, primer annealing for 30 s at 60 °C, primer extension for 30 s at 72 °C, and a final extension for 7 min at 72 °C. Amplified DNA fragments were resolved in 3% agarose gel. Product sizes of DNA fragments were determined using 100 bp DNA ladder (Thermo Scientific) as molecular size marker.

Data analysis

Genetic diversity and relationship analysis

Online marker efficiency iMEC program (https://irscope.shinyapps.io/iMEC/) was used to calculate multiple indices of marker efficiency such as number of alleles (Na), expected heterozygosity and discriminating power (Amiryousefi et al. 2018). DNA fragments of various molecular weight sizes generated by SSR markers were compared with the standard molecular weight marker and scored as discrete variables using 1 to indicate presence and 0 to indicate absence of a band. The heatmap was generated based on SSR data of 72 walnut genotypes constructed by Euclidean distance with Ward (unsquared distances) linkage method using Clust Vis Bio tools (https://bio.tools/clustvis) (Metsalu and Vilo 2015).

Genetic structure and admixture analysis

The population structure was analyzed using the Bayesian clustering algorithm implemented in STRUCTURE. The program STRUCTURE was run with *K* values from 1 to 12. A burn-in period of 50,000 iterations followed by 500,000 replications was used to estimate each value of *K*. No prior information was used to define the clusters. The number of populations was determined by maximizing Ln likelihood of data for different values of *K* (Evanno et al. 2005) and the optimal *K* depends on the peak of ΔK (Earl and VonHoldt 2012). Genotypes with affiliation probabilities of 60% or higher were designated as belonging to a specific group, while those with affiliation probabilities below 60% were classified as admixture. For the purpose of analyzing molecular variance, Arlequin software was employed (AMOVA). Based on the geographic location of the samples and the

findings of the population structure of the investigated genotypes, Arlequin was used to calculate the pair-wise genetic distances and the population differentiation coefficients within and among populations (Excoffier and Lischer 2010).

Results

Frequency of SSR's in walnut genome

A total of 198,924 SSR loci were identified in the 647 Mb walnut genome. Among these, successful primers were designed for 136,582 loci (Table S1). The frequency of SSRs per Mb within the genome was 428.71. Overall SSR motifs analysis shows that the frequency of SSRs falls with the increasing number of repeat motifs. Dinucleotides motifs were predominant and accounted for 88.40% (175,075) of total SSRs followed by trinucleotides (17,184) with a frequency of 8.3% while octanucleotides were least frequent (< 0.1%; Fig. 1a). Frequency of dinucleotide repeated motifs was 377.312 SSRs/Mb and the frequency of SSRs/Mb decreased with the increase in repeat motifs (Fig. 1b).

Motif type and motif repeats

We looked at the top 20 single and paired group motifs. In both solo and paired dinucleotide motifs, the dinucleotides came in first place, accounting for 88.40% and 88.6% motifs, respectively (Figs. 2, 3). In each class, we discovered that some motif types were more prevalent than others. For instance, the AT motif was significantly overrepresented in dinucleotide motifs (28%) (Fig. 2a). Additionally, an examination of various repeat counts revealed that dinucleotides (AT motif) had the highest frequency (114.33 SSRs/Mb). Among the trinucleotides, the AAT motif has the highest frequency (5.99 SSR/Mb), while tetranucleotides and pentanucleotides had less SSR repeats (Fig. 2b). The paired motifs AT/AT were more common and accounts for 28% alike that of single motif (AT) followed by TA/TA paired motifs (Fig. 3a). The highest number of SSRs/Mb was obtained in motifs AT/AT and TA/TA followed by other paired motif types (Fig. 3b). It was interesting to observe 81- and 62-time repetition of 2 trinucleotide SSR motifs (ATA and TAT). Another intriguing finding was that heptanucleotides had more repeats than tetranucleotides and pentanucleotides (Fig. 4).

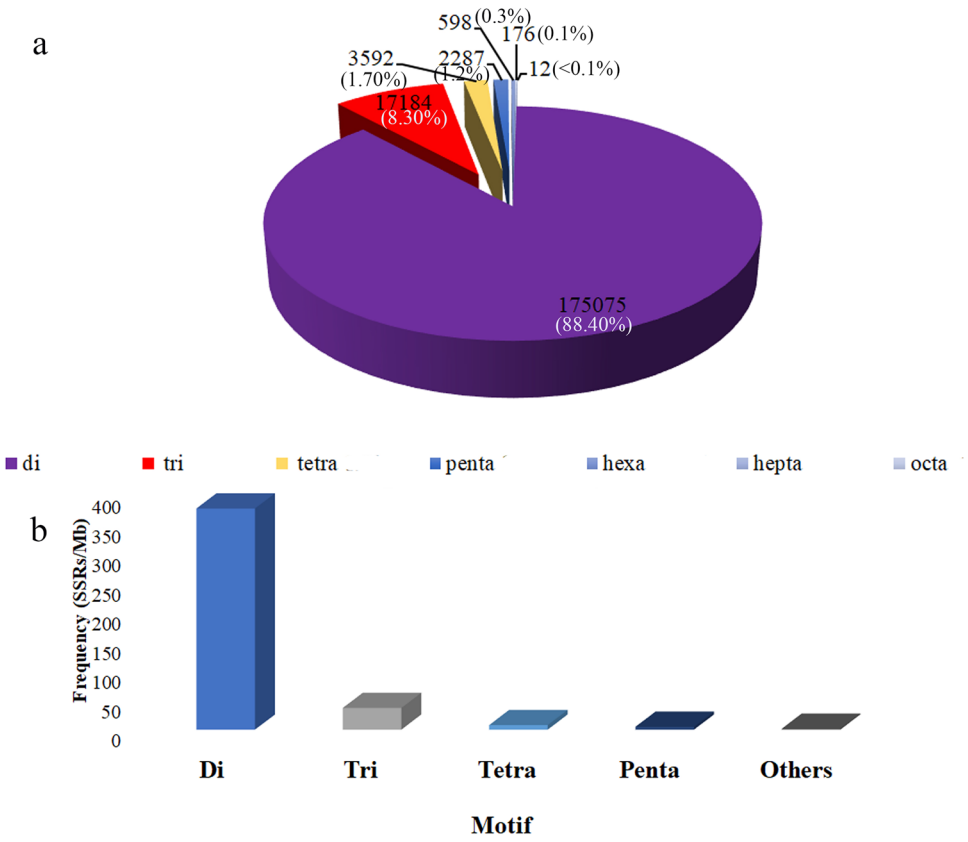

Fig. 1 Distribution of motifs and SSR density of walnut genome. Distribution of motif numbers and their percentage (%); the motifs from dinucleotides to octanucleotides are shown by colored boxes (**a**). Frequency distribution of SSR's/Mb from di-to-penta motifs. The horizontal axis depicts the motif type, whereas the vertical axis indicates the frequency of SSRs/Mb (**b**)

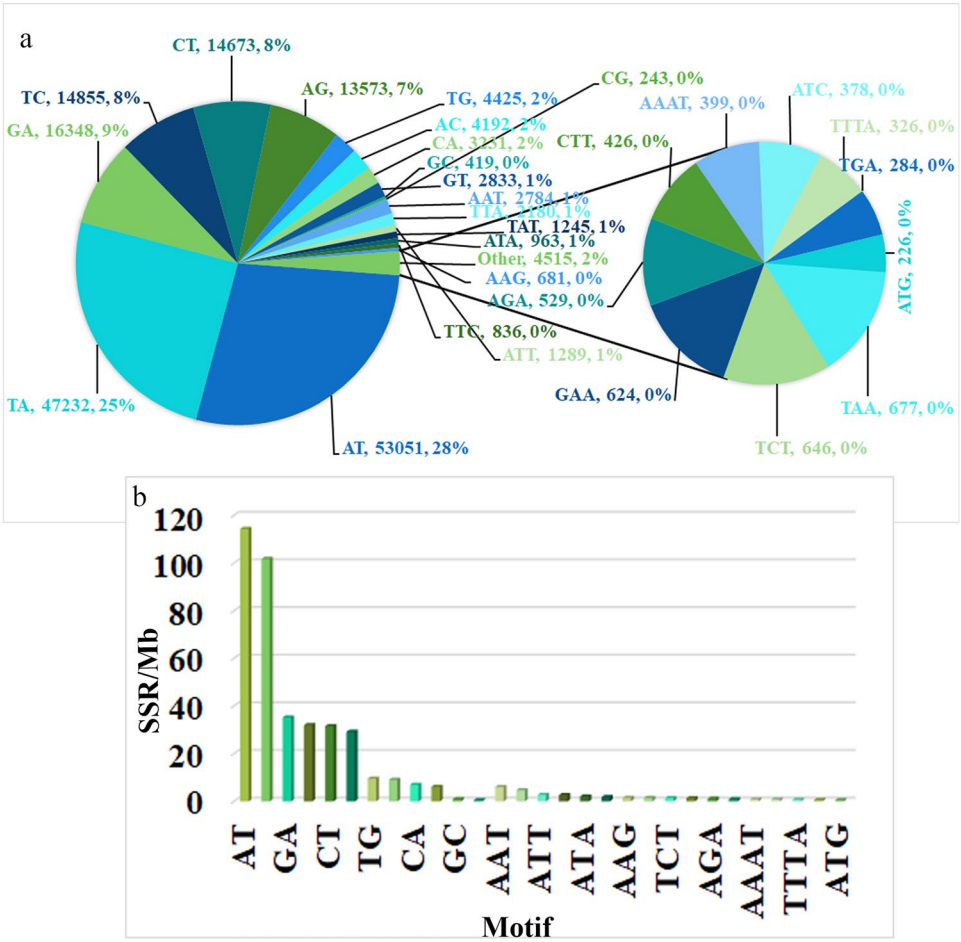

Fig. 2 Distribution of individual motifs and SSRs/Mb in Chandler walnut genome. Distribution of individual motif type, number, and percentage from dinucleotides to tetranucleotides, which are discriminated from each other by different colors (**a**). Frequency of individual SSR motifs from di to tetra (**b**). The horizontal axis depicts the motif type, whereas the vertical axis indicates the frequency of SSRs/Mb

In silico PCR

The unique SSR markers produced a single allele (84.35%) and the remaining markers produced greater than two alleles (15.65%; Table S2). The number of in silico alleles ranged from 1 to 131 and the average amplicons per mapped marker was 1.20. We also found 99.82% markers generating ≤ 10 in silico PCR products and 99.97% of markers generating ≤ 50 in silico products (Table S2).

Validation and marker efficiency of microsatellite markers

When fingerprinted, majority of SSR primers (65) produce monomorphic band. Therefore, 72 walnut genotypes were fingerprinted using a set of 20 highly polymorphic SSR markers which produced 118 alleles. The primers generated alleles with values ranging from 2 to 12 with an average of 5.2 alleles per primer. The primer WSSR008 yielded the most alleles (12), followed by WSSR001 and WSSR026 with ten alleles each (Fig. 5; Table 1). The primer WSSR3 amplified the minimum of two bands. We observed the amplicon size 110–500 bp that matched to the e-PCR band size (161–393 bp) of SSRs. The polymorphic information content (PIC) of 75% of the markers was ≥ 0.5 and 25% of the markers produced PIC value less than 0.5 with overall values ranging from 0.391 to 0.605 and an average value of 0.184 (Table 1). The expected heterozygosity index (*H*) ranged between 0.081 and 0.625 with a mean value of 0.514. The discrimination power had a mean value of 0.474 and a range of 0.081–0.590 (Table 1).

Genetic relationship and admixture analysis

To determine which genotypes are similar and which individuals differ from one another, it is vital to analyze molecular data matrices using methods like heatmaps and principal component analysis (PCoA). The heatmap created from the SSR molecular data set using the Ward's linkage clustering approach and Euclidean distance indicated two unique groups (Fig. 6). The genotypes are clustered as shown by the PCoA ellipses (Fig. 7), with PC1 and PC2 accounting for 22.2% and 10.1%, respectively, of the molecular variation. The results of the PCoA matrix showed that the walnut accessions were divided into two primary clusters (Fig. 7). According to the PCoA results, the accessions from

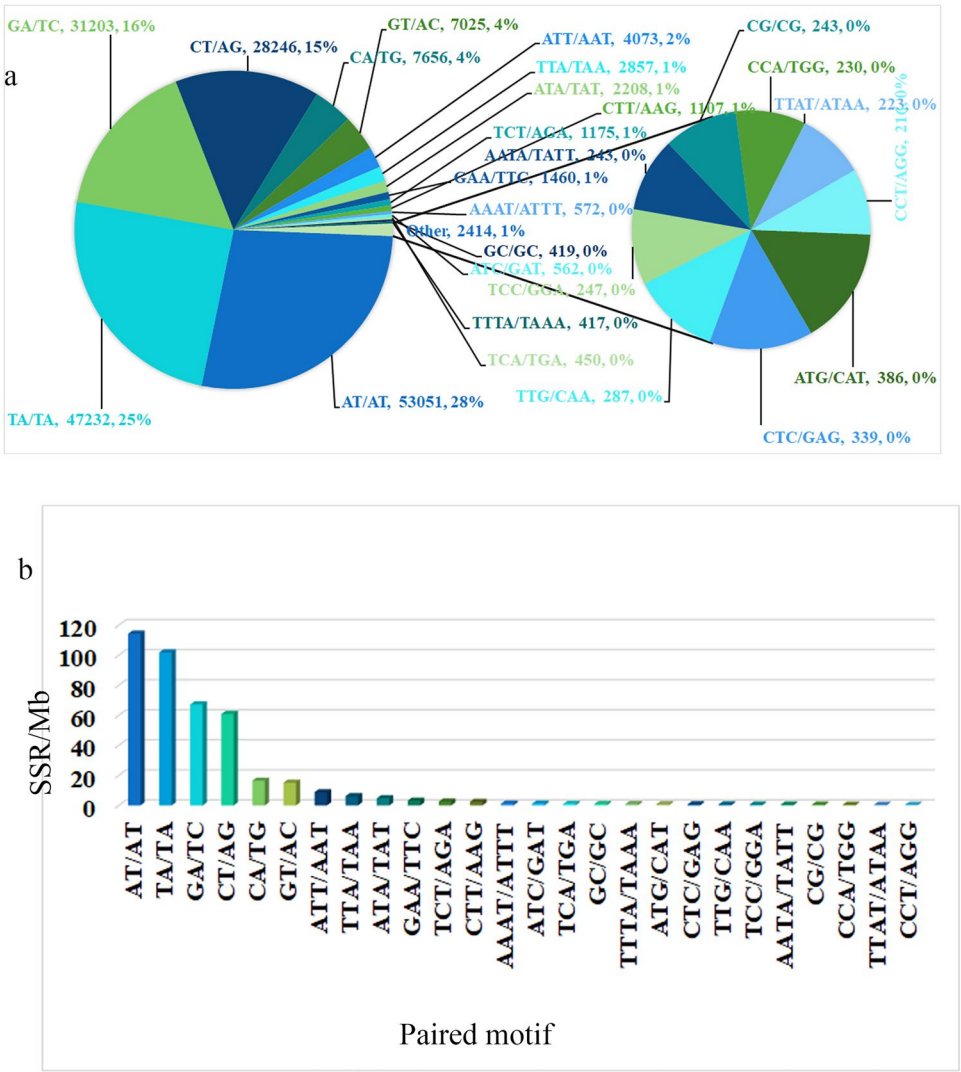

Fig. 3 Distribution of motif type, quantity, and percentage of paired nucleotides (di to tetra), that can be distinguished from one another by their respective colors (**a**). Pair-wise frequency distribution for di-, tri- and tetra-SSR motifs. The vertical axis shows the frequency of SSR's/Mb and the horizontal axis displays the paired motif type (**b**)

Anantnag and Pulwama form a single group, and are clustered within the Shopian population that is encircled by a red ellipse, except a single genotype at the circumference's edge. Many Shopian accessions were present in the other cluster. The clustering pattern of PCoA and the heatmap are in agreement.

We used a model-based approach to study the genetic structure of walnut. To identify the true number of populations, two distinguished methods, non-parametric (Wilcoxon test) and delta *K* method, were applied. The non-parametric method could not give the exact number of populations. Therefore, delta *K* method was applied (Fig. 8). According to the distribution of delta *K* values, there was only one peak (Fig. 8a) at $K=2$ indicating two distinct populations. Among 72 genotypes, 28 genotypes were placed in subpopulation I and 43 were placed in subpopulation II (Fig. 8b). The single genotype SW-46 showed admixtures. Furthermore, the analysis revealed that the overall proportion of membership of the samples in each of the two clusters was 39.43% in cluster I and 59.72% in cluster II excluding admixture member. Statistical analysis revealed that the percentage of genotypes having ≥ 90 membership was 87.5%, 11.11% exhibited membership coefficient $\geq 60\%$, and 1.39% of the genotypes exhibited membership coefficient percentage of $\leq 5\%$. The membership coefficient in the bar plot revealed that accessions SW-05 and SW-25 have gene flow from the cluster II (green) and accessions SW-01, SW-29, SW-37, and SW-38 received genetic material from the cluster I (red). Similarly, allele frequency among two sub-populations (net nucleotide distance) was 0.0669 and average distance (expected heterozygosity or gene diversity) between individuals in same cluster was found almost similar in cluster I (0.2303) and cluster II (0.2312). Mean value of fixation indices (F_{ST}) measures the genetic differentiation among the populations. It is one of the most important and frequently used parameters in explaining the population structure. The F_{ST} measured by the STRUCTURE program revealed greater F_{ST} in subpopulation I (0.3134) than in subpopulation II (0.2389).

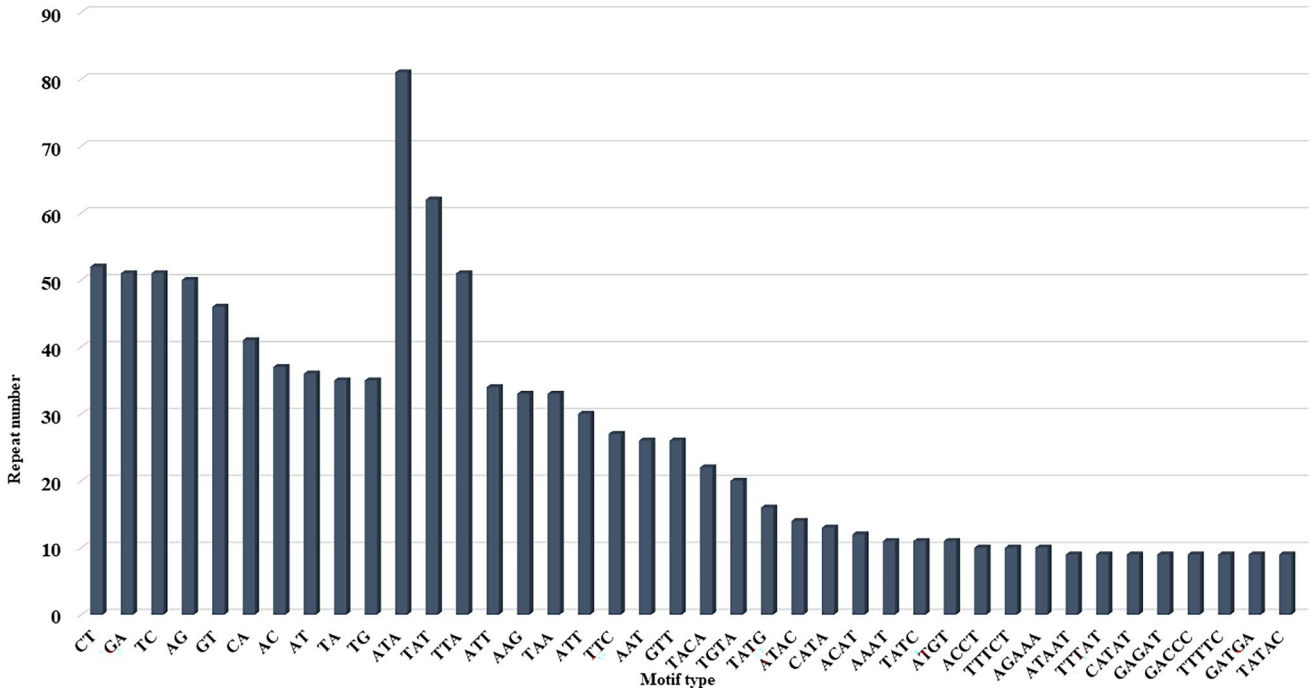

Fig. 4 Distribution of top ten motifs (di to penta) with their repeat numbers

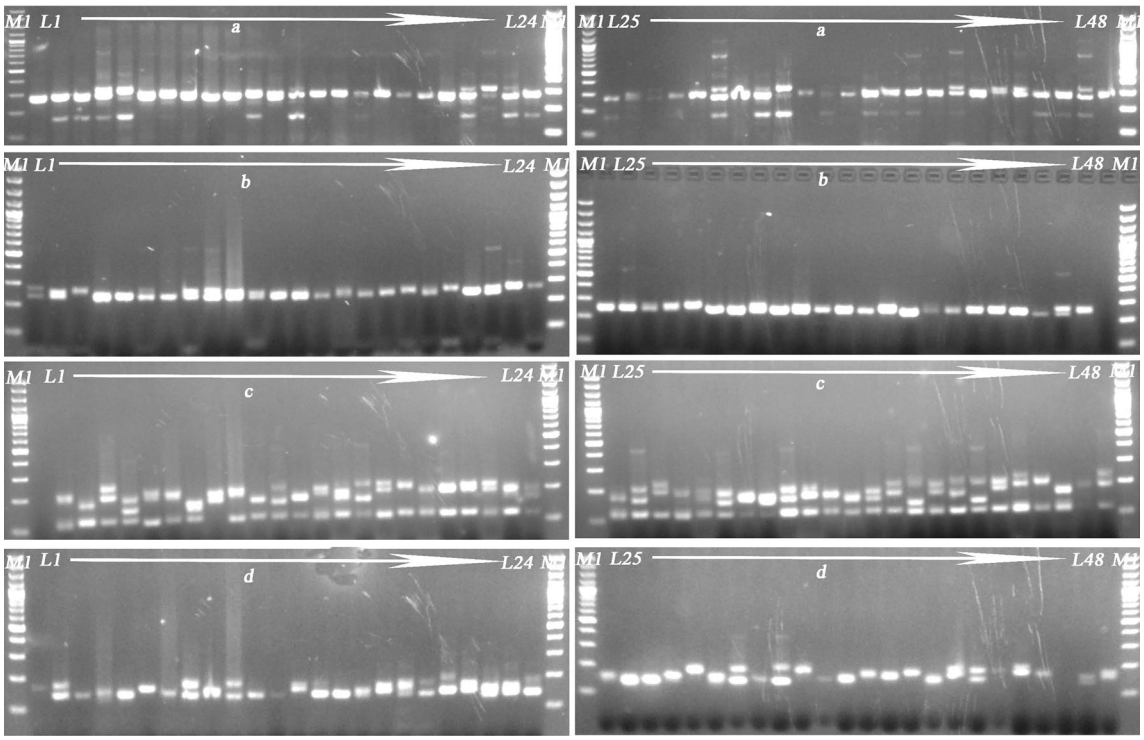

Fig. 5 Electrophoretic monograph of four SSR markers. Lane M1 is 100 bp DNA marker. Lane L1–L48 are walnut genotypes and *a* = WSSR1; *b* = WSSR016; *c* = WSSR018; *d* = WSSR002

Table 1 List of highly polymorphic 20 SSR markers and their summary statistics

S. no.	Primer locus	Chromosome number	Left primer sequence 5′–3′	Right primer sequence 5′–3′	No. of alleles (Na)	Polymorphic information content (PIC)	Expected Heterozygosity index (*H*)	Discriminating power (*D*)	Expected allele size (bp)	Observed allele size range (bp)
1	WSSR1	14	TGCACGAACTCGCTTGTATC	CGGGATTGACAAGAGTAGGG	3	0.530	0.596	0.475	316	280–320
2	WSSR 2	2	TATTTCCAACCCACCGATGT	AGCTCAAAGGCGTACGAAAC	4	0.540	0.589	0.547	197	150–200
3	WSSR 3	9	GAAATCGGAGAGGGAGGTTG	GACCCAAGATTTAAGGATTGGAC	2	0.605	0.081	0.081	292	280–290
4	WSSR 6	1	TTTGTGATGCAGGAGACAGG	CGTTGAGATCGGACATGTTG	3	0.545	0.554	0.539	311	300–320
5	WSSR 8	16	TGTGCGACGGTATGAGAGAG	GGCAGCCACCATCAATATCT	4	0.551	0.506	0.460	316	220–300
6	WSSR 9	2	TCGAGTGGGCTAGCTTTGAT	CCCTTAAAATAAGACCTGAAGGTGA	4	0.585	0.290	0.291	288	210–320
7	WSSR001	2	TTGTGCACCTTTGAGTTTCG	TGGAGGATGCTCTTCATGTG	10	0.520	0.560	0.549	299	210–300
8	WSSR 004	2	AAGCTTCCCAAAACAAGCAA	GAAAGAGAGGGTCATGCACAG	3	0.533	0.625	0.457	320	180–320
9	WSSR 007	11	TTAGCAAGATGCAGGCTAACA	ATCAGCAACATTCGCATCAG	6	0.523	0.606	0.580	351	250–350
10	WSSR 008	12	ATCAGCCGGTTCACAAATTC	TTCCACCGAATCCATCTCTC	12	0.541	0.497	0.488	294	200–300
11	WSSR015	11	GGGCCCCTTCAATGACTACT	CGAAGGCAATGGAAATTTGA	8	0.497	0.587	0.581	161	110–220
12	WSSR016	8	GTAGCTAGAGCCCACGATGC	CACCCATCAACATCGTTCAG	6	0.553	0.513	0.477	296	250–300
13	WSSR018	11	GAGGAGCCATAGGTGGTTGA	GTGGGTCTTTTCCCCTTAGC	6	0.552	0.472	0.445	206	120–250
14	WSSR020	5	CCGATTGATCACGATTGGTA	AGAGCGCATTCTCTCAAAGC	8	0.391	0.475	0.481	191	190–230
15	WSSR023	15	CCATTTCGGTTGGGTTACTG	TGCATCCAAACATGGGTTTA	5	0.502	0.589	0.573	385	390–420
16	WSSR024	1	CGAGTTCGGTCTCTTTCCTG	ACCTTAGGACCCACCCTGAC	8	0.475	0.534	0.490	295	250–300
17	WSSR026	7	TGCATGATGTTTGTGCTTCT	CATGAGCTTTGTACTTCGGCTA	10	0.416	0.467	0.461	264	200–280
18	WSSR034	8	TGTGTAACATGATTTGCTCGAC	TAGACCCTGGTGGCTCATTC	8	0.440	0.554	0.550	266	200–290
19	WSSR035	5	CGATGGGGTGGATTTTAGTG	AGAAGCCCTTACCCAGTCGT	4	0.535	0.602	0.590	355	300–500

Table 1 (continued)

S. no.	Primer locus	Chromosome number	Left primer sequence 5′–3′	Right primer sequence 5′–3′	No. of alleles (Na)	Polymorphic information content (PIC)	Expected Heterozygosity index (*H*)	Discriminating power (*D*)	Expected allele size (bp)	Observed allele size range (bp)
20	WSSR036	8	CTTCCCACCCACCAAGACTA	TTGTTAGCTCGCCTCTTTGC	4	0.524	0.589	0.571	393	300–400
	Maximum				12	0.605	0.625	0.590		
	Minimum				2	0.391	0.081	0.081		
	Average				5.9	0.518	0.514	0.4843		

The AMOVA based on geographical origin of samples revealed significant molecular variation within populations (92.04%) than among populations (7.96%). Whereas, analysis based on population structure ($K=2$) showed 87.38% molecular variance within the population and 12.62% among the populations (Table 2). The F_{ST} among the populations was 0.06 to 0.12 (0.05–0.25), indicating moderate level of genetic differentiation.

Discussion

In the present study, walnut genome downloaded from NCBI (National Centre of Biotechnology Information) was mined to develop large number of microsatellite markers. Genome-wide SSR markers have been successfully developed in various plant species including jujube (Xiao et al. 2015), apple (Zhang et al. 2012), citrus (Biswas et al. 2014; Duhan et al. 2020; Liu et al. 2013), pomegranate (Patil et al. 2020b, 2021), *Bunium persicum* (Bansal et al. 2022), pear (Xue et al. 2018), watermelon (Zhu et al. 2016), and bottle gourd (Bonthala et al. 2022). In the current investigation, we thoroughly detailed 162,594 genome-wide microsatellite markers for this significant crop. To the best of our knowledge, this is the first study on *J. regia* that presents enormous number of genome-wide microsatellite markers. Because of its larger genomic size (647 Mb), the number of SSRs in walnut is comparatively large than other crops. For instance, only 28,342 and 39,523 SSRs were mined from foxtail and watermelon genomes, because of their smaller genomic sizes (Zhu et al. 2016; Pandey et al. 2013). In comparison, the density of SSRs within the genome was 428.71 SSRs/Mb. However, it is surprising that SSR densities among the various woody plants did not differ considerably (Liu et al. 2018a). According to other studies, genome size and SSR density are negatively correlated (Cavagnaro et al. 2010; Liu et al. 2013; Morgante et al. 2002). It may be due to variation in search parameters used to mine SSRs from the genomes (Zhu et al. 2016) or, the different sequencing and assembly methods (Xu et al. 2013). This SSR set after validation will help the scientific community for developing saturated linkage map and mapping of useful traits in walnut that were impossible with limited number of available SSR markers. In addition, a large set of SSR markers will make it easier to map QTLs precisely, identify and exploit genes that control critical traits, conduct genome-wide association studies, enable selective breeding through genomic selection, and infer population structure. Microsatellite markers play a major role in genetic improvement of cereals and grasses but are yet to be explored in horticultural crops. For instance, SSRs shed light on gene regulation and genome organization, genetic diversity (Zhao et al. 2014; Göl et al. 2017), crop domestication (Zhao et al. 2014), variety and scion source

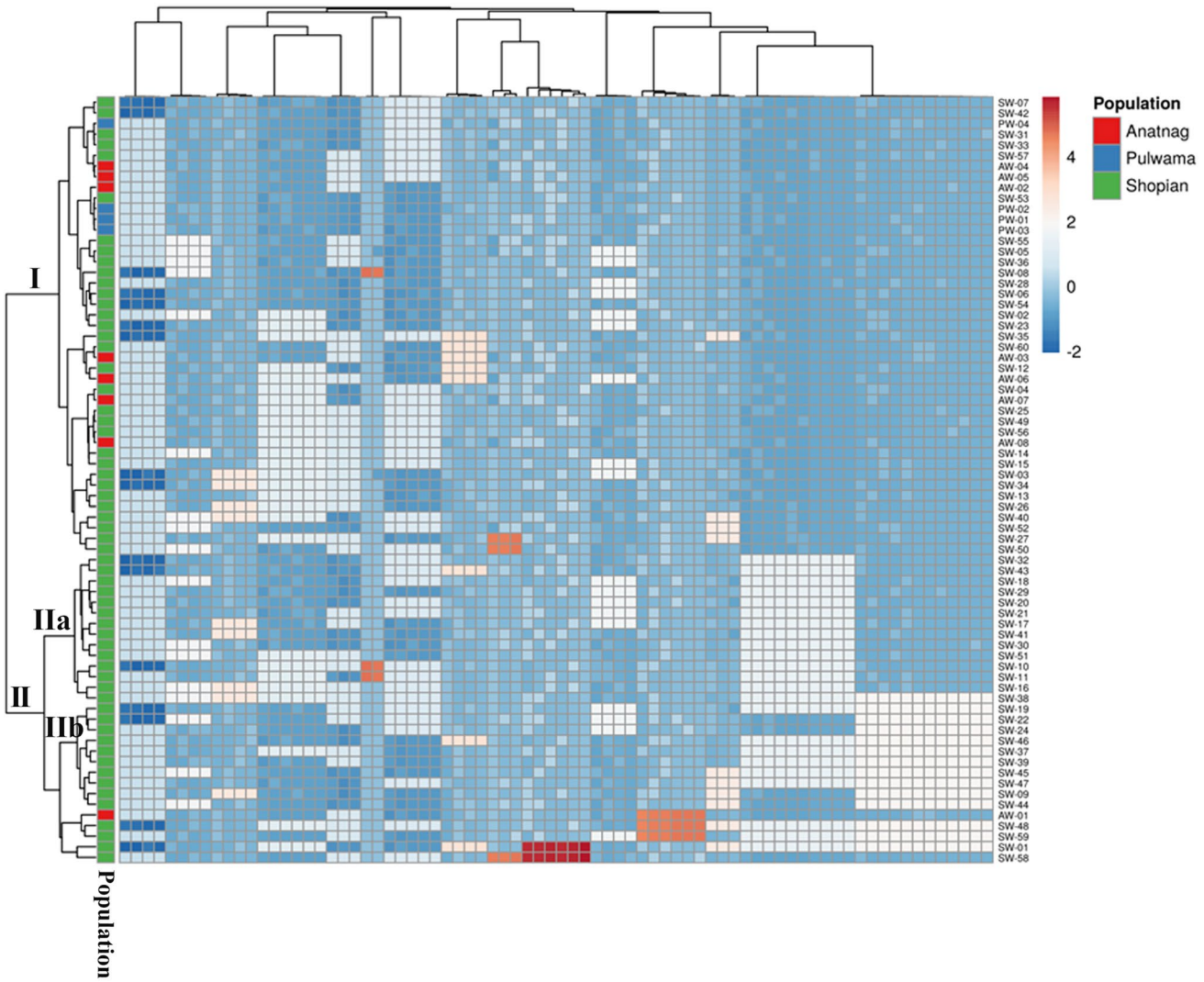

Fig. 6 Heatmap categorize 72 walnut genotypes into 3 populations. The blue and light square plots of the heatmap indicate the presence (1) and absence of the loci (0) of the particular sample. The red, blue, and green represent the three populations

validation (Arab et al. 2022; Nickravesh et al. 2023), comparative mapping (Zhu et al. 2016; Wu et al. 2017), genetic map construction (Bali et al. 2015; Tan et al. 2013), and breeding studies (Dossa et al. 2017).

Out of 192,924 SSR loci identified, successful primers were designed for 162,594 (84.27%) loci. In the present investigation, the options of 200 bp flanking SSR region must be responsible for not designing SSR marker for 15.73% loci. Most of these SSR loci were present either in the beginning or end of the scaffold. The failure to develop successful primer pairs for each detected SSR locus in plants genomes is consistent with earlier observations (Pandey et al. 2013; Sonah et al. 2011; Parida et al. 2009). The SSR primers designed were subjected to electronic PCR module (e-PCR) to check the amplification efficiency. It is difficult to validate each primer pair through a thermocycler but e-PCR module is very useful for rapid screening and effective identification of informative markers (Patil et al. 2020a, 2021; Duhan et al. 2020). Hence, each microsatellite created in the present study was confirmed using the e-PCR module with default settings. When subjected to in silico PCR, the majority of SSRs produced a single allele; however, few SSR primers produced multiple bands. To validate the microsatellites generated from plant genomes, many researchers have used in silico PCR amplification modules (Biswas et al. 2020; Shi et al. 2014; Wang et al. 2015). Out of the designed primers, 110 microsatellite markers with different motifs and longest repeats were selected for validation purpose because longer repeats in the genome have higher mutation rates, which can result in a high frequency of polymorphism (Bhat et al. 2018; Cavagnaro et al. 2010; Wren et al. 2000).

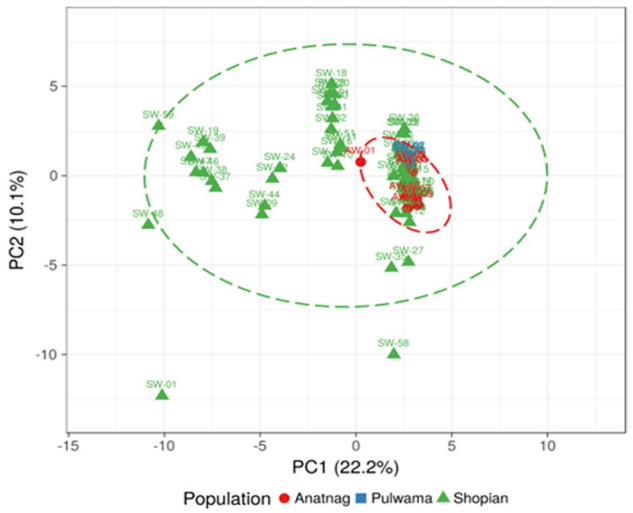

Fig. 7 PCA biplot categorizes the genotypes into single cluster (encircled by green) with admixture from Pulwama and Anantnag (encircled by red). The Anantnag population is shown by red circle, Pulwama population by blue square, and Shopian population by green triangle

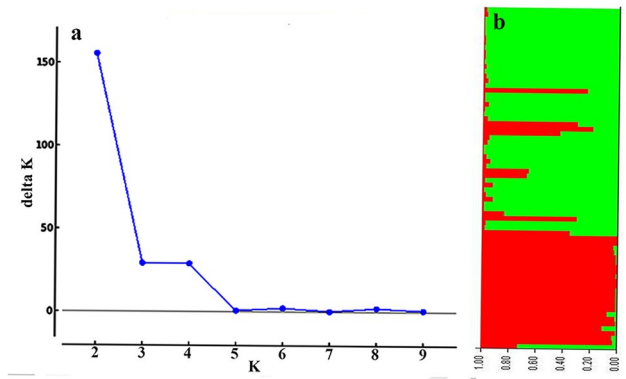

Fig. 8 Structure stratification indicates 2 populations of 72 walnut genotypes (**a**). The red and green represent the members of the two groups or clusters inferred by STRUCTURE harvester (**b**)

The frequency of microsatellites is negatively correlated with the number of nucleotides among the different nucleotide types. Frequency analyses of different nucleotide repeats in walnut revealed that dinucleotide repeats are most abundant SSRs, accounting for 88.4% of total SSRs while hepta-nucleotide repeats were least abundant, representing only 0.1% of total microsatellites. These results are in agreement with numerous studies examining various crop species (Liu et al. 2013; Najafi et al. 2014; Tangphatsornruang et al. 2009; Topçu et al. 2015; Xu et al. 2013; Zhang et al. 2007; Zhu et al. 2012). Microsatellite abundances considerably reduced with the increase in number of motif repeats. The dinucleotide repeats experienced the slowest rate of change while other longer repeats experienced a higher rate of change. The results were inconsistent with those of other studies, as *Cucumis sativa*, *Medicago truncatula*, *Populus trichocarpa*, and *Vitis vinifera* had the highest tetranucleotide repeats, while *Glycine max*, *Arabidopsis thaliana*, *Oryza sativa*, *Setaria italica*, and *Sorghum bicolor* had the highest trinucleotide repeats (Cavagnaro et al. 2010). This is most likely a result of the various SSR identification criteria being used. Dinucleotides and trinucleotides were found to have SSRs with a greater repetition count, whereas tetranucleotides, pentanucleotides, and hexanucleotides had less repeats of the SSR motif. Several plant species showed similar tendencies as well such as citrus (Liu et al. 2013) and watermelon (Zhu et al. 2016).

There were apparent differences in the frequency of the motifs. The AT/AT motif was the most prevalent dinucleotide repeat in the walnut genome. Likewise, to this, the trinucleotide and tetranucleotide repeats of the motifs ATA/TAT and AAAT/ATTT were the most common, indicating that they are the most frequent motifs throughout the entire walnut genome. Since AT motifs are unlikely to undergo mutations. For instance, AG/CT is the most abundant motif in rice (Zhang et al. 2007) and citrus (Liu et al. 2013). However, AT/TA motif is abundant in maize (Xu et al. 2013),

Table 2 Analysis of molecular variance of 72 walnut genotypes partitioned into populations based on their geographic location and structure differentiation

Populations	Source of variation	*df*	Sum of squares	Estimated variability	Percentage of variation (%)	*p* value
Geographic origin of samples	Among populations	2	7.930	0.17322[a]	7.96	< 0.05
	Within populations	69	140.140	2.00199[b]	92.04	< 0.05
	Total	71	148.069	2.17521		
	F_{ST}	0.07963				
Population structure	Among populations	1	10.036	0.28892[a]	12.62	< 0.05
	Within populations	70	138.033	2.00048[b]	87.38	< 0.05
	Total	71	148.069	2.28940		
	F_{ST}	0.126				

[a,b] indicate significant difference among and within population, respectively

cucumber (Cavagnaro et al. 2010), pomegranate (Patil et al. 2021), pepper (Zhong et al. 2021), and watermelon (Zhu et al. 2012). Such studies indicate overrepresentation of different motifs in different plant species.

Molecular diversity analysis of *J. regia* genotypes based on 20 microsatellite markers revealed a high level of polymorphism in different genotypes of walnut indicating a suitability of these markers for studying genetic diversity. Microsatellite markers are suitable for studying the walnut genetic diversity (Ahmed et al. 2012; Bai et al. 2010; Dangl et al. 2005; Foroni et al. 2005; Gunn et al. 2010; Shah et al. 2020; Victory et al. 2006; Woeste et al. 2002; Karimi et al. 2010). All primers showed high rate of amplification success. In the present study, some of the primers were unable to amplify all genotypes indicating that these genotypes are distant to the Chandler. Walnut being diploid so the SSRs produced a maximum of two bands per locus and the results are in accordance with earlier reports (Ahmed et al. 2012; Najafi et al. 2014; Mahmoodi et al. 2019). However, some primers (Walnut primer-7 and Walnut primer-11) produced multiple bands suggesting their multi-loci nature.

The substantial impact on the utilization of the SSR markers depends on the SSR markers, the accuracy of the genotypic data acquisition, and the planting material (Liu et al. 2018b, 2017). We were able to find 20 highly polymorphic SSR markers which amplified distinct and consistent bands across 72 walnut genotypes. The size of the amplified products was at par with the expected size value of each locus. This shows the primer binding site of primers was highly conserved. Surprisingly the few SSR markers produced low PIC value < 0.5 and majority of the markers produced PIC value > 0.5. The low PIC value may likely be due to location of these markers in the coding regions of the genotypes. The SSRs found in coding regions are less prone to mutation than non-coding genomic SSRs (Kalia et al. 2011). The average PIC value of our SSR markers was comparatively lesser than reported by Guney et al. (2021). The variations in PIC value may be due to sampling technique, number of SSR markers, the size and type of SSR motifs repeats, and the location of the SSR motifs in the genome (Orhan et al. 2020). The PIC value of the majority of the newly developed SSR markers is > 0.5 demonstrating their suitability for phylogenetic and diversity studies as well as construction of linkage maps (Biswas et al. 2014). The present study reports 5.2 alleles per primer and is significantly lower than 23.8 alleles per primer reported by Victory et al. (2006). It is interesting to note that compared to agarose, metaphor gel electrophoresis polyacrylamide gel electrophoresis and the automated capillary DNA fragment analyzer significantly contribute to higher polymorphism (Ebrahimi et al. 2011; Dangl et al. 2005; Patil et al. 2020a). We anticipated that our polymorphic SSR markers can reveal higher number of alleles if assayed through automated capillary systems or polyacrylamide gel electrophoresis. The variation in number of alleles amplified may also be due to highly diverse nature of the samples and number of SSRs tested.

Unrevealing the degree of genetic diversity is necessary for accelerating the walnut genetic improvement. To achieve this, molecular marker technologies, such as SSRs, have become a promising method for identifying genetic variation in a set of genotypes. In this context, the heatmap, PCoA, and structure analysis methods were effectively used to measure the genetic relationships and population differentiation (Ebrahimi et al. 2016; Shah et al. 2020; Pollegioni et al. 2011, 2015). According to Roor et al. (2017), the Himalayan range of Jammu and Kashmir is the native range of the *J. regia.* The fragmentation and geographic isolation of the walnut populations in this area occurs due to genes flow barrier and other natural factors (Pollegioni et al. 2015). This led to population differentiation in natural range of walnut. However, there are other factors such as human activities, which can contribute to the genetic structure of the autochthonous population (Gunn et al. 2010). Therefore, the population genetic structure revealed by our genetic data needs to be integrated with historical and linguistic sources to find whether this is the product of natural factors or anthropogenic dispersal or human cultural interactions. We observed higher molecular variance within the walnut populations, which may be attributed to the predominant cross-pollination of walnut (Victory et al. 2006; Pollegioni et al. 2014) and the higher gene flow. The low molecular variance among populations is related to long separation, avoidance of long-distance pollination, and fragmented character of populations, which causes pollinations within near relatives only. These results are in accordance with other earlier studies (Magige et al. 2022; Wang et al. 2022; Zhang et al. 2022). Therefore, when selecting the populations of *J. regia* with high genetic diversity, the individuals should be selected from within the population for genetic improvement of the walnut.

Conclusion

Walnut is an economically important nut crop with high diversity. The long juvenile period is a bottleneck for its genetic improvement. For walnut speed breeding, it is imperative to identify the markers tightly linked to the economic traits. Rapid progress has been made in the development of genomic tools over the past few years, such as the release of the genome sequence, which created new prospects for the development of numerous genetic markers like SSRs. To explore this opportunity, we identified 198,924 SSR loci and successfully designed primers for 162,594 SSR loci. As 100 out of 110 SSRs amplified the various walnut genotypes, the e-PCR module demonstrated that each SSR created in

the current study will generate an amplicon across all of the walnut genotypes. The majority of our SSRs had PIC values above 0.5, which shows their robustness for predicting genetic diversity and population structure. To the best of our knowledge, this is the first study of scanning SSRs from the walnut genome, and we present a microsatellite repository for the walnut scientific community. These SSRs will be helpful for walnut improvement such as development of saturated genetic linkage map, genetic structure, QTL mapping, and marker-assisted selection.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13205-023-03563-6.

Acknowledgements First and corresponding authors thank Department of Biotechnology, Ministry of Science and Technology India for providing funding (Grant No: BT/PR31931/AGIII/103/1128/2019). We also thank Halima E. Awale, Michigam State University, East Lansing USA for helping us in language editing.

Author contributions HI and BAP conceived and designed the experiments. RAS, SQ, and AJ performed the experiments. BAP and RAS analyzed the data. HI, BAP, MDS, and RAS contributed reagents/materials/analysis tools. BAP and RAS wrote the paper. HI, BAP, RAS. SK, SMZ, MAM, and ZAB are advisee members to AJ and SQ.

Data availability The data is publicly available at NCBI [Juglans regia (ID 17683)—Genome—NCBI (nih.gov)].

Declarations

Conflict of interest The authors declare that they have no conflict of interest in the publication.

References

Ahmed N, Mir JI, Mir RR, Rather NA, Rashid R, Wani S, Shafi W, Mir H, Sheikh MA (2012) SSR and RAPD analysis of genetic diversity in walnut (Juglans regia L) genotypes from Jammu and Kashmir, India. Physiol Mol Biol Plants 18(2):149–160

Ali Khan M, Shahid Ul I, Mohammad F (2016) Extraction of natural dye from walnut bark and its dyeing properties on wool yarn. J Nat Fibers 13(4):458–469

Amiryousefi A, Hyvönen J, Poczai P (2018) iMEC: online marker efficiency calculator. Appl Plant Sci 6(6):e01159

Arab MM, Marrano A, Abdollahi-Arpanahi R, Leslie CA, Askari H, Neale DB, Vahdati K (2019) Genome-wide patterns of population structure and association mapping of nut-related traits in Persian walnut populations from Iran using the Axiom J. regia 700K SNP array. Sci Rep 9(1):1–14

Arab MM, Brown PJ, Abdollahi-Arpanahi R, Sohrabi SS, Askari H, Aliniaeifard S, Mokhtassi-Bidgoli A, Mesgaran MB, Leslie CA, Marrano A (2022) Genome-wide association analysis and pathway enrichment provide insights into the genetic basis of photosynthetic responses to drought stress in Persian walnut. Hortic Res 9:uhac124. https://doi.org/10.1093/hr/uhac124

Aradhya M, Velasco D, Ibrahimov Z, Toktoraliev B, Maghradze D, Musayev M, Bobokashvili Z, Preece JE (2017) Genetic and ecological insights into glacial refugia of walnut (*Juglans regia* L.). PLoS ONE 12(10):e0185974

Bai WN, Liao WJ, Zhang DY (2010) Nuclear and chloroplast DNA phylogeography reveal two refuge areas with asymmetrical gene flow in a temperate walnut tree from East Asia. New Phytol 188(3):892–901

Bali S, Mamgain A, Raina SN, Yadava SK, Bhat V, Das S, Pradhan AK, Goel S (2015) Construction of a genetic linkage map and mapping of drought tolerance trait in Indian beveragial tea. Mol Breed 35(5):1–20

Bansal S, Kumar A, Lone AA, Khan MH, Malhotra EV, Singh R (2022) Development of novel genome-wide simple sequence repeats (SSR) markers in *Bunium persicum*. Ind Crops Prod 178:114625

Bernard A, Marrano A, Donkpegan A, Brown PJ, Leslie CA, Neale DB, Lheureux F, Dirlewanger E (2019) Association and Linkage Mapping to Unravel Genetic Architecture of Phenology-Related Traits and Lateral Bearing in Persian Walnut (*Juglans regia* L.).

Bernard A, Barreneche T, Donkpegan A, Lheureux F, Dirlewanger E (2020a) Comparison of structure analyses and core collections for the management of walnut genetic resources. Tree Genet Genom 16(5):1–14

Bernard A, Marrano A, Donkpegan A, Brown PJ, Leslie CA, Neale DB, Lheureux F, Dirlewanger E (2020b) Association and linkage mapping to unravel genetic architecture of phenological traits and lateral bearing in Persian walnut (*Juglans regia* L.). BMC Genom 21(1):1–25

Bhat NN, Padder BA, Shah MD, Dar MS, Nabi A, Bano A, Rasool RS (2018) Microsatellite mining in the genus *Colletotrichum*. Gene Rep 13:84–93

Biswas MK, Xu Q, Mayer C, Deng X (2014) Genome wide characterization of short tandem repeat markers in sweet orange (*Citrus sinensis*). PLoS ONE 9(8):e104182

Biswas MK, Bagchi M, Nath UK, Biswas D, Natarajan S, Jesse DMI, Park J-I, Nou I-S (2020) Transcriptome wide SSR discovery cross-taxa transferability and development of marker database for studying genetic diversity population structure of Lilium species. Sci Rep 10(1):1–13

Bonthala B, Abdin MZ, Arya L, Pandey CD, Sharma V, Yadav P, Verma M (2022) Genome-wide SSR markers in bottle gourd: development, characterization, utilization in assessment of genetic diversity of National Genebank of India and synteny with other related cucurbits. J Appl Genet 63(2):237–263

Bükücü ŞB, Sütyemez M, Kefayati S, Paizila A, Jighly A, Kafkas S (2020) Major QTL with pleiotropic effects controlling time of leaf budburst and flowering-related traits in walnut (*Juglans regia* L.). Sci Rep 10(1):1–10

Cavagnaro PF, Senalik DA, Yang L, Simon PW, Harkins TT, Kodira CD, Huang S, Weng Y (2010) Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). BMC Genom 11(1):1–18

Chen L, Ma Q, Chen Y, Wang B, Pei D (2014) Identification of major walnut cultivars grown in China based on nut phenotypes and SSR markers. Sci Hortic 168:240–248

Cheng J, Zhao Z, Li B, Qin C, Wu Z, Trejo-Saavedra DL, Luo X, Cui J, Rivera-Bustamante RF, Li S (2016) A comprehensive characterization of simple sequence repeats in pepper genomes provides valuable resources for marker development in Capsicum. Sci Rep 6(1):1–12

Dangl GS, Woeste K, Aradhya MK, Koehmstedt A, Simon C, Potter D, Leslie CA, McGranahan G (2005) Characterization of 14 microsatellite markers for genetic analysis and cultivar identification of walnut. J Am Soc Hort Sci 130(3):348–354

Doğan Y, Kafkas S, Sütyemez M, Akça Y, Türemiş N (2014) Assessment and characterization of genetic relationships of walnut (*Juglans regia* L.) genotypes by three types of molecular markers. Sci Hortic 168:81–87

Dossa K, Yu J, Liao B, Cisse N, Zhang X (2017) Development of highly informative genome-wide single sequence repeat markers

for breeding applications in sesame and construction of a web resource: SisatBase. Front Plant Sci 8:1470
Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bullet 19:11–15
Duhan N, Meshram M, Loaiza CD, Kaundal R (2020) citSATdb: genome-wide simple sequence repeat (SSR) marker database of Citrus species for germplasm characterization and crop improvement. Genes 11(12):1486
Earl DA, VonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4(2):359–361
Ebrahimi A, Fatahi R, Zamani Z (2011) Analysis of genetic diversity among some Persian walnut genotypes (*Juglans regia* L.) using morphological traits and SSRs markers. Sci Hortic 130(1):146–151
Ebrahimi A, Zarei A, Lawson S, Woeste KE, Smulders MJM (2016) Genetic diversity and genetic structure of Persian walnut (*Juglans regia*) accessions from 14 European, African, and Asian countries using SSR markers. Tree Genet Genom 12(6):1–12
Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14(8):2611–2620
Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10(3):564–567
Feng X, Zhou H, Zulfiqar S, Luo X, Hu Y, Feng L, Malvolti ME, Woeste K, Zhao P (2018) The phytogeographic history of common walnut in China. Front Plant Sci 9:1399
Fjellstrom RG, Parfitt DE (1994) RFLP inheritance and linkage in walnut. Theor Appl Genet 89:665–670
Foroni I, Rao R, Woeste K, Gallitelli M (2005) Characterisation of *Juglans regia* L. with SSR markers and evaluation of genetic relationships among cultivars and the 'Sorrento' landrace. J Hortic Sci Biotechnol 80(1):49–53
Foroni I, Woeste K, Monti LM, Rao R (2007) Identification of 'Sorrento' walnut using simple sequence repeats (SSRs). Genet Resour Crop Evol 54(5):1081–1094
Gaisberger H, Legay S, Andre C, Loo J, Azimov R, Aaliev S, Bobokalonov F, Mukhsimov N, Kettle C, Vinceti B (2020) Diversity under threat: connecting genetic diversity and threat mapping to set conservation priorities for *Juglans regia* L. populations in Central Asia. Front Ecol Evol 8:171
Göl Ş, Göktay M, Allmer J, Doğanlar S, Frary A (2017) Newly developed SSR markers reveal genetic diversity and geographical clustering in spinach (*Spinacia oleracea*). Mol Genet Genom 292(4):847–855
Grover A, Aishwarya V, Sharma PC (2007) Biased distribution of microsatellite motifs in the rice genome. Mol Genet Genom 277(5):469–480
Guney M, Kafkas S, Keles H, Zarifikhosroshahi M, Gundesli MA, Ercisli S, Necas T, Bujdoso G (2021) Genetic diversity among some walnut (*Juglans regia* L.) genotypes by SSR markers. Sustainability 13(12):6830
Gunn BF, Aradhya M, Salick JM, Miller AJ, Yongping Y, Lin L, Xian H (2010) Genetic variation in walnuts (*Juglans regia* and *J. sigillata*; Juglandaceae): species distinctions, human impacts, and the conservation of agrobiodiversity in Yunnan, China. Am J Bot 97(4):660–671
Hoban S, Anderson R, McCleary TIM, Schlarbaum S, Romero-Severson J (2008) Thirteen nuclear microsatellite loci for butternut (*Juglans cinerea* L.). Mol Ecol Resour 8(3):643–646
Hou X-J, Liu S-R, Khan MRG, Hu C-G, Zhang J-Z (2014) Genome-wide identification, classification, expression profiling, and SSR marker development of the MADS-box gene family in Citrus. Plant Mol Biol Rep 32(1):28–41
Hussain SZ, Naseer B, Qadri T, Fatima T, Bhat TA (2021) Walnut (*Juglans regia*)-morphology, taxonomy, composition and health benefits. In: Hussain SZ (ed) Fruits grown in highland regions of the Himalayas: nutritional and health benefits, pp 269–281. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-75502-7_21
Jaćimović V, Adakalić M, Ercisli S, Božović D, Bujdoso G (2020) Fruit quality properties of walnut (*Juglans regia* L.) genetic resources in Montenegro. Sustainability 12(23):9963
Ji F, Ma Q, Zhang W, Liu J, Feng Y, Zhao P, Song X, Chen J, Zhang J, Wei X (2021) A genome variation map provides insights into the genetics of walnut adaptation and agronomic traits. Genome Biol 22(1):1–22
Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK (2011) Microsatellite markers: an overview of the recent progress in plants. Euphytica 177(3):309–334
Karimi R, Ershadi A, Vahdati K, Woeste K (2010) Molecular characterization of Persian walnut populations in Iran with microsatellite markers. HortScience 45(9):1403–1406
Kefayati S, Ikhsan A, Sütyemez M, Paizila A, Topçu H, Bükücü Ş, Kafkas S (2017) A genetic linkage map for walnut based on SSR markers. In: VIII international symposium on Walnut, Cashew and Pecan 1318, pp 39–44
Khadivi-Khub A, Ebrahimi A, Sheibani F, Esmaeili A (2015) Phenological and pomological characterization of Persian walnut to select promising trees. Euphytica 205(2):557–567
Liu S-R, Li W-Y, Long D, Hu C-G, Zhang J-Z (2013) Development and characterization of genomic and expressed SSRs in citrus by genome-wide analysis. PLoS ONE 8(10):e75149
Liu Q, Song Y, Liu L, Zhang M, Sun J, Zhang S, Wu J (2015) Genetic diversity and population structure of pear (Pyrus spp.) collections revealed by a set of core genome-wide SSR markers. Tree Genet Genom 11(6):1–22
Liu S, Liu H, Wu A, Hou Y, An Y, Wei C (2017) Construction of fingerprinting for tea plant (*Camellia sinensis*) accessions using new genomic SSR markers. Mol Breed 37(8):1–14
Liu S, An Y, Li F, Li S, Liu L, Zhou Q, Zhao S, Wei C (2018a) Genome-wide identification of simple sequence repeats and development of polymorphic SSR markers for genetic studies in tea plant (*Camellia sinensis*). Mol Breed 38:1–13
Liu S, An Y, Li F, Li S, Liu L, Zhou Q, Zhao S, Wei C (2018b) Genome-wide identification of simple sequence repeats and development of polymorphic SSR markers for genetic studies in tea plant (*Camellia sinensis*). Mol Breed 38(5):1–13
Luo M-C, You FM, Li P, Wang J-R, Zhu T, Dandekar AM, Leslie CA, Aradhya M, McGuire PE, Dvorak J (2015) Synteny analysis in Rosids with a walnut physical map reveals slow genome evolution in long-lived woody perennials. BMC Genom 16:1–17
Magige EA, Fan P-Z, Wambulwa MC, Milne R, Wu Z-Y, Luo Y-H, Khan R, Wu H-Y, Qi H-L, Zhu G-F (2022) Genetic diversity and structure of Persian walnut (*Juglans regia* L.) in Pakistan: implications for conservation. Plants 11(13):1652
Mahmoodi R, Dadpour MR, Hassani D, Zeinalabedini M, Vendramin E, Micali S, Nahandi FZ (2019) Development of a core collection in Iranian walnut (*Juglans regia* L.) germplasm using the phenotypic diversity. Sci Hortic 249:439–448
Malvolti ME, Fornari B, Maccaglia E, Cannata F (2001) Genetic linkage mapping in an intraspecific cross of walnut (*Juglans regia* L.) using molecular markers. Acta Hortic 544:179–186
Manee MM, Al-Shomrani BM, Al-Fageeh MB (2020) Genome-wide characterization of simple sequence repeats in Palmae genomes. Genes Genom 42(5):597–608
Marrano A, Martínez-García PJ, Bianco L, Sideli GM, Di Pierro EA, Leslie CA, Stevens KA, Crepeau MW, Troggio M, Langley CH (2019a) A new genomic tool for walnut (*Juglans regia* L.):

development and validation of the high-density Axiom™ J. regia 700K SNP genotyping array. Plant Biotechnol J 17(6):1027–1036
Marrano A, Sideli GM, Leslie CA, Cheng H, Neale DB (2019b) Deciphering of the genetic control of phenology, yield, and pellicle color in Persian walnut (*Juglans regia* L.). Front Plant Sci 10:1140
Marrano A, Britton M, Zaini PA, Zimin AV, Workman RE, Puiu D, Bianco L, Pierro EAD, Allen BJ, Chakraborty S (2020) High-quality chromosome-scale assembly of the walnut (*Juglans regia* L.) reference genome. Gigascience 9(5):giaa050
Martínez-García PJ, Crepeau MW, Puiu D, Gonzalez-Ibeas D, Whalen J, Stevens KA, Paul R, Butterfield TS, Britton MT, Reagan RL (2016) The walnut (*Juglans regia*) genome sequence reveals diversity in genes coding for the biosynthesis of non-structural polyphenols. Plant J 87(5):507–532
McGranahan G, Leslie C (2009) Breeding walnuts (*Juglans regia*). In: Breeding plantation tree crops: temperate species. Springer, pp 249–273
Metsalu T, Vilo J (2015) ClustVis: a web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. Nucleic Acids Res 43(W1):W566–W570
Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet 30(2):194–200
Najafi F, Mardi M, Fakheri B, Pirseyedi S, Mehdinejad N, Farsi M (2014) Isolation and characterization of novel microsatellite markers in walnut (*Juglans regia* L.). Am J Plant Sci 5(3):409–415. https://doi.org/10.4236/ajps.2014.53054
Nickravesh MH, Vahdati K, Amini F, Di Pierro EA, Amiri R, Woeste K, Arab MM (2023) Reliable propagation of Persian walnut varieties using SSR marker-based true-to-type validation. HortScience 58(1):64–66
Orhan E, Eyduran SP, Poljuha D, Akin M, Weber T, Ercisli S (2020) Genetic diversity detection of seed-propagated walnut (L.) germplasm from Eastern Anatolia using SSR markers. Folia Hortic 32(1):37–46
Pandey G, Misra G, Kumari K, Gupta S, Parida SK, Chattopadhyay D, Prasad M (2013) Genome-wide development and use of microsatellite markers for large-scale genotyping applications in foxtail millet [*Setaria italica* (L.)]. DNA Res 20(2):197–207
Parida SK, Kalia SK, Kaul S, Dalal V, Hemaprabha G, Selvi A, Pandit A, Singh A, Gaikwad K, Sharma TR (2009) Informative genomic microsatellite markers for efficient genotyping applications in sugarcane. Theor Appl Genet 118(2):327–338
Patil PG, Singh NV, Parashuram S, Bohra A, Mundewadikar DM, Sangnure VR, Babu KD, Sharma J (2020a) Genome wide identification, characterization and validation of novel miRNA-based SSR markers in pomegranate (*Punica granatum* L.). Physiol Mol Biol Plants 26(4):683–696
Patil PG, Singh NV, Parashuram S, Bohra A, Sowjanya R, Gaikwad N, Mundewadikar DM, Sangnure VR, Jamma SM, Injal AS (2020b) Genome-wide characterization and development of simple sequence repeat markers for genetic studies in pomegranate (*Punica granatum* L.). Trees 34(4):987–998
Patil PG, Singh NV, Bohra A, Raghavendra KP, Mane R, Mundewadikar DM, Babu KD, Sharma J (2021) Comprehensive characterization and validation of chromosome-specific highly polymorphic SSR markers from Pomegranate (*Punica granatum* L.) cv. Tunisia Genome. Front Plant Sci 12:337
Pollegioni P, Major A, Bartoli S, Ducci F, Proietti R, Malvolti ME (2004) Application of microsatellite and dominant molecular markers for the discrimination of species and interspecific hybrids in genus *Juglans*. In: V International walnut symposium, vol 705, pp 191–197
Pollegioni P, Woeste K, Olimpieri I, Marandola D, Cannata F, Emilia Malvolti M (2011) Long-term human impacts on genetic structure of Italian walnut inferred by SSR markers. Tree Genet Genom 7(4):707–723
Pollegioni P, Woeste KE, Chiocchini F, Olimpieri I, Tortolano V, Clark J, Hemery GE, Mapelli S, Malvolti ME (2014) Landscape genetics of Persian walnut (*Juglans regia* L.) across its Asian range. Tree Genet Genom 10(4):1027–1043
Pollegioni P, Woeste KE, Chiocchini F, Del Lungo S, Olimpieri I, Tortolano V, Clark J, Hemery GE, Mapelli S, Malvolti ME (2015) Ancient humans influenced the current spatial genetic structure of common walnut populations in Asia. PLoS ONE 10(9):e0135980
Pollegioni P, Woeste K, Chiocchini F, Del Lungo S, Ciolfi M, Olimpieri I, Tortolano V, Clark J, Hemery GE, Mapelli S (2017) Rethinking the history of common walnut (*Juglans regia* L.) in Europe: its origins and human interactions. PLoS ONE 12(3):e0172541
Robichaud RL, Glaubitz JC, Rhodes OE, Woeste K (2006) A robust set of black walnut microsatellites for parentage and clonal identification. New for 32(2):179–196
Robichaud RL, Glaubitz JC, Rhodes OE, Woeste K (2010) Genetic consequences of harvest in a mature second-growth stand of black walnut (*Juglans nigra* L.). Ann for Sci 67(7):702
Roor W, Konrad H, Mamadjanov D, Geburek T (2017) Population differentiation in common walnut (*Juglans regia* L.) across major parts of its native range—insights from molecular and morphometric data. J Hered 108(4):391–404
Ross-Davis A, Huang Z, McKenna J, Ostry M, Woeste K (2008) Morphological and molecular methods to identify butternut (*Juglans cinerea*) and butternut hybrids: relevance to butternut conservation. Tree Physiol 28(7):1127–1133
Shah UN, Mir JI, Ahmed N, Fazili KM (2018) Assessment of germplasm diversity and genetic relationships among walnut (*Juglans regia* L.) genotypes through microsatellite markers. J Saudi Soc Agric Sci 17(4):339–350
Shah RA, Baksi P, Jasrotia A, Bhat DJI, Gupta R, Bakshi M (2020) Genetic diversity of walnut (*Juglans regia* L.) seedlings through SSR markers in north-western Himalayan region of Jammu. Bangladesh J Bot 49(4):1003–1012
Shah RA, Bakshi P, Sharma N, Jasrotia A, Itoo H, Gupta R, Singh A (2021) Diversity assessment and selection of superior Persian walnut (*Juglans regia* L.) trees of seedling origin from North-Western Himalayan region. Resour Environ Sustain 3:100015
Shah RA, Bakshi P, Jasrotia A, Itoo H, Gupta R (2022) Bio-Chemical composition of some Walnut (*Juglans regia* L.) genotypes of North-Western Himalayan Region. Bangladesh J Bot 51(1):93–101
Shi J, Huang S, Zhan J, Yu J, Wang X, Hua WEI, Liu S, Liu G, Wang H (2014) Genome-wide microsatellite characterization and marker development in the sequenced Brassica crop species. DNA Res 21(1):53–68
Sideli GM, Marrano A, Montanari S, Leslie CA, Allen BJ, Cheng H, Brown PJ, Neale DB (2020) Quantitative phenotyping of shell suture strength in walnut (*Juglans regia* L.) enhances precision for detection of QTL and genome-wide association mapping. PLoS ONE 15(4):e0231144
Sonah H, Deshmukh RK, Sharma A, Singh VP, Gupta DK, Gacche RN, Rana JC, Singh NK, Sharma TR (2011) Genome-wide distribution and organization of microsatellites in plants: an insight into marker development in Brachypodium. PLoS ONE 6(6):e21298
Stevens KA, Woeste K, Chakraborty S, Crepeau MW, Leslie CA, Martínez-García PJ, Puiu D, Romero-Severson J, Coggeshall M, Dandekar AM (2018) Genomic variation among and within six Juglans species. G3 Genes Genomes Genet 8(7):2153–2165
Taheri S, Lee Abdullah T, Yusop MR, Hanafi MM, Sahebi M, Azizi P, Shamshiri RR (2018) Mining and development of novel SSR markers using next generation sequencing (NGS) data in plants. Molecules 23(2):399

Tan L-Q, Wang L-Y, Wei K, Zhang C-C, Wu L-Y, Qi G-N, Cheng H, Zhang Q, Cui Q-M, Liang J-B (2013) Floral transcriptome sequencing for SSR marker development and linkage map construction in the tea plant (*Camellia sinensis*). PLoS ONE 8(11):e81611

Tangphatsornruang S, Somta P, Uthaipaisanwong P, Chanprasert J, Sangsrakru D, Seehalak W, Sommanas W, Tragoonrung S, Srinives P (2009) Characterization of microsatellites and gene contents from genome shotgun sequences of mungbean (*Vigna radiata* (L.) Wilczek). BMC Plant Biol 9(1):1–12

Topcu H, Coban N, Woeste K, Sutyemez M, Kafkas S (2015) Developing new microsatellite markers in walnut (*Juglans regia* L.) from Juglans nigra genomic GA enriched library. Ekin J Crop Breed Genet 1(2):93–99

Topçu H, Ikhsan AS, Sütyemez M, Çoban N, Güney M, Kafkas S (2015) Development of 185 polymorphic simple sequence repeat (SSR) markers from walnut (*Juglans regia* L.). Sci Hortic 194:160–167

Turdieva MK, Kayimov AK, Baymetov KI, Mustafina FU, Butkov EA (2012) Conservation and sustainable use of biodiversity of fruit crops and wild fruit species. In: Proceedings of International scientific and practical conference, 23–26 August 2011, Tashkent, Uzbekistan

Victory ER, Glaubitz JC, Rhodes OE Jr, Woeste KE (2006) Genetic homogeneity in *Juglans nigra* (Juglandaceae) at nuclear microsatellites. Am J Bot 93(1):118–126

Wang X, Wang L (2016) GMATA: an integrated software package for genome-scale SSR mining, marker development and viewing. Front Plant Sci 7:1350

Wang H, Pei D, Gu R-s, Wang B-q (2008) Genetic diversity and structure of walnut populations in central and southwestern China revealed by microsatellite markers. J Am Soc Hortic Sci 133(2):197–203

Wang Q, Fang L, Chen J, Hu Y, Si Z, Wang S, Chang L, Guo W, Zhang T (2015) Genome-wide mining, characterization and development of microsatellite markers in Gossypium species. Sci Rep 5(1):1–10

Wang Z, Zhang H, Tong B, Han B, Liu D, Zhang P, Hu D (2022) The study on genetic diversity and genetic structure of *Juglans mandshurica* in Shandong Province based on EST-SSR

Woeste K, McGranahan GH, Bernatzky R (1996) Randomly amplified polymorphic DNA loci from a walnut backcross [(*Juglans hindsii* × *J. regia*) × *J. regia*]. J Am Soc Hortic Sci 121(3):358–361

Woeste K, Burns R, Rhodes O, Michler C (2002) Thirty polymorphic nuclear microsatellite loci from black walnut. J Hered 93(1):58–60

Wren JD, Forgacs E, Fondon Iii JW, Pertsemlidis A, Cheng SY, Gallardo T, Williams RS, Shohet RV, Minna JD, Garner HR (2000) Repeat polymorphisms within gene regions: phenotypic and evolutionary implications. Am J Hum Genet 67(2):345–356

Wu J, Cheng F, Cai C, Zhong Y, Jie X (2017) Association mapping for floral traits in cultivated *Paeonia rockii* based on SSR markers. Mol Genet Genom 292(1):187–200

Xiao J, Zhao J, Liu M, Liu P, Dai L, Zhao Z (2015) Genome-wide characterization of simple sequence repeat (SSR) loci in Chinese jujube and jujube SSR primer transferability. PLoS ONE 10(5):e0127812

Xu JIE, Liu L, Xu Y, Chen C, Rong T, Ali F, Zhou S, Wu F, Liu Y, Wang J (2013) Development and characterization of simple sequence repeat markers providing genome-wide coverage and high resolution in maize. DNA Res 20(5):497–509

Xue H, Zhang P, Shi T, Yang J, Wang L, Wang S, Su Y, Zhang H, Qiao Y, Li X (2018) Genome-wide characterization of simple sequence repeats in *Pyrus bretschneideri* and their application in an analysis of genetic diversity in pear. BMC Genom 19(1):1–13

Zaini PA, Feinberg NG, Grilo FS, Saxe HJ, Salemi MR, Phinney BS, Crisosto CH, Dandekar AM (2020) Comparative proteomic analysis of walnut (*Juglans regia* L.) pellicle tissues reveals the regulation of nut quality attributes. Life 10(12):314

Zhang Z, Deng Y, Tan J, Hu S, Yu J, Xue Q (2007) A genome-wide microsatellite polymorphism database for the indica and japonica rice. DNA Res 14(1):37–45

Zhang R, Zhu A, Wang X, Yu J, Zhang H, Gao J, Cheng Y, Deng X (2010) Development of Juglans regia SSR markers by data mining of the EST database. Plant Mol Biol Rep 28(4):646–653

Zhang Q, Ma B, Li H, Chang Y, Han Y, Li J, Wei G, Zhao S, Khan MA, Zhou Y (2012) Identification, characterization, and utilization of genome-wide simple sequence repeats to identify a QTL for acidity in apple. BMC Genom 13(1):1–12

Zhang ZY, Han JW, Jin Q, Wang Y, Pang XM, Li YY (2013) Development and characterization of new microsatellites for walnut (*Juglans regia*). Genet Mol Res 12(4):4723–4734

Zhang B-W, Xu L-L, Li N, Yan P-C, Jiang X-H, Woeste KE, Lin K, Renner SS, Zhang D-Y, Bai W-N (2019) Phylogenomics reveals an ancient hybrid origin of the Persian walnut. Mol Biol Evol 36(11):2451–2461

Zhang Q, Ree RH, Salamin N, Xing Y, Silvestro D (2022) Fossil-informed models reveal a boreotropical origin and divergent evolutionary trajectories in the walnut family (Juglandaceae). Syst Biol 71(1):242–258

Zhao D-w, Yang J-b, Yang S-x, Kato K, Luo J-p (2014) Genetic diversity and domestication origin of tea plant *Camellia taliensis* (Theaceae) as revealed by microsatellite markers. BMC Plant Biol 14(1):1–12

Zhong Y, Cheng Y, Ruan M, Ye Q, Wang R, Yao Z, Zhou G, Liu J, Yu J, Wan H (2021) High-throughput SSR marker development and the analysis of genetic diversity in *Capsicum frutescens*. Horticulturae 7(7):187

Zhu H, Senalik D, McCown BH, Zeldin EL, Speers J, Hyman J, Bassil N, Hummer K, Simon PW, Zalapa JE (2012) Mining and validation of pyrosequenced simple sequence repeats (SSRs) from American cranberry (*Vaccinium macrocarpon* Ait.). Theor Appl Genet 124(1):87–96

Zhu Y, Yin Y, Yang K, Li J, Sang Y, Huang L, Fan S (2015) Construction of a high-density genetic map using specific length amplified fragment markers and identification of a quantitative trait locus for anthracnose resistance in walnut (*Juglans regia* L.). BMC Genom 16(1):1–13

Zhu H, Song P, Koo D-H, Guo L, Li Y, Sun S, Weng Y, Yang L (2016) Genome wide characterization of simple sequence repeats in watermelon genome and their application in comparative mapping and genetic diversity analysis. BMC Genom 17(1):1–17