Article

Joint Identification and Application of Microsatellite Markers in Genetic Diversity Study of Closely Related Species *Teucrium montanum*, *T. capitatum* and Their Natural Hybrid

Filip Varga [1,2] , Zlatko Liber [1,3,*] , Ante Turudić [1] , Jernej Jakše [4] , Lea Juzbašić [3] , Nina Jeran [2] , Martina Grdiša [1,2] , Miloš Zbiljić [5] and Zlatko Šatović [1,2]

1   Centre of Excellence for Biodiversity and Molecular Plant Breeding (CoE CroP-BioDiv), Svetošimunska Cesta 25, 10000 Zagreb, Croatia; fvarga@agr.hr (F.V.); aturudic@agr.hr (A.T.); mgrdisa@agr.hr (M.G.); zsatovic@agr.hr (Z.Š.)
2   Faculty of Agriculture, University of Zagreb, Svetošimunska Cesta 25, 10000 Zagreb, Croatia; njeran@agr.hr
3   Faculty of Science, University of Zagreb, Marulićev Trg 9a, 10000 Zagreb, Croatia
4   Biotechnical Faculty, University of Ljubljana, Jamnikarjeva 101, 1000 Ljubljana, Slovenia; jernej.jakse@bf.uni-lj.si
5   Department of Botany, Faculty of Pharmacy, University of Belgrade, Vojvode Stepe 450, 11060 Belgrade, Serbia; milos.zbiljic@pharmacy.bg.ac.rs
*   Correspondence: zlatko.liber@biol.pmf.hr

Abstract: *Teucrium montanum* L. and *T. capitatum* L. are two plant species with overlapping distribution in the Balkan Peninsula, especially in Croatia, where several occurrences of their putative hybrid species *T.* × *rohlenae* have been recorded. Next-generation sequencing of both species and de novo assembly was carried out resulting in 120 contigs for *T. montanum* and 1685 contigs for *T. capitatum* assembled. The *Dig-up primers* pipeline was used for SSR mining of both assemblies, applying different criteria that resulted in 112 SSR candidates for testing. A subset of 41 SSRs was selected and after two rounds of testing, twelve SSRs were developed and characterized. A total of 232 alleles were detected with 5 to 29 alleles per locus, based on the test sample. The genetic diversity analysis of three *Teucrium* taxa from a single location revealed a higher level of diversity in *T. montanum* than in *T. capitatum* with intermediate values for the hybrid species. The NeighborNet diagram and genetic structure analysis grouped the taxa into two separate clusters, the first of which consisted exclusively of *T. montanum*, while the other was composed of intermixed *T. capitatum* and *T.* × *rohlenae* individuals. The availability of SSR markers for two Teucrium taxa will allow in-depth analysis of genetic diversity and structure, as well as molecular identification of their putative hybrids in the future.

Keywords: *T. montanum*; *T. capitatum*; *T.* × *rohlenae*; genomic SSRs; Lamiaceae

1. Introduction

The genus *Teucrium* L., belonging to the family Lamiaceae, comprises a large (more than 250 species) and diverse group of herbaceous plants and shrubs distributed in different regions of the world [1]. Most *Teucrium* species are native to Europe and Central Asia, where they thrive in diverse habitats ranging from meadows and grasslands to rocky slopes and forests [2].

In Croatia, the genus *Teucrium* L. is well represented in the country's botanical diversity and in diverse landscapes, including coastal areas, mountainous regions and inland areas. It provides suitable conditions for 12 *Teucrium* species. Among these species, *T. montanum* L. and *T. capitatum* L., two Steno-Mediterranean taxa have the largest distribution area in Croatia. They can be found in karst habitats within the coastal and peri-coastal regions of the country [3], as can be seen in Figure 1. In certain areas where both species occur sympatrically, a naturally occurring species of hybridogenous origin, *T.* × *rohlenae* K. Malý,

has recently been recorded based on morphological analysis and biochemical profile [4]. *T. capitatum* and *T. polium* L. were until recently considered to be the same species [5,6], namely *T. capitatum* was treated as a synonym of *T. polium* subsp. *capitatum* (L.) Arcang.
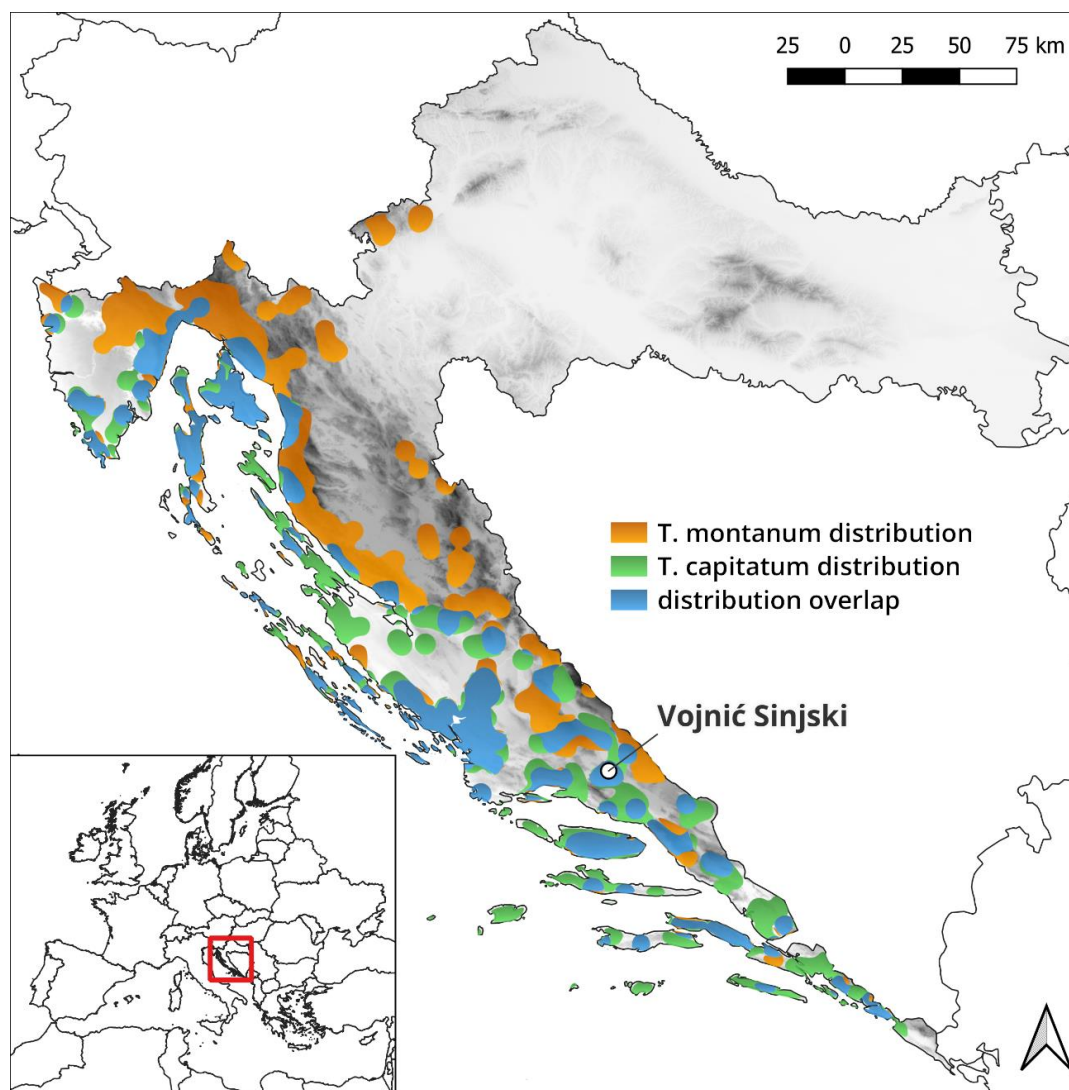
Figure 1. Distribution of *T. montanum* and *T. capitatum* in Croatia. Areas where the distribution of the two species overlap are marked in blue. Vojnić Sinjski, the sampled location where hybridization between two species was detected is marked with a white circle.

The consumption of *T. montanum* and/or *T. capitatum* for the treatment of different medical conditions such as gastrointestinal and liver ailments, inflammation of the respiratory tract, and strengthening of the immune system has been well documented in the Balkan region [1,7–10]. Most of the research on *T. montanum* and *T. capitatum* has focused on investigating their phytochemical composition and potential pharmacological applications. Analysis of the essential oil extracted from the aerial parts of *T. montanum* and *T. capitatum* revealed a similar chemical profile [11], with sesquiterpenes such as β-caryophyllene and β-pinene being the dominant compounds [4]. The essential oil of *T. capitatum* has also been investigated for its antinociceptive potential by performing the twitch test in mice with positive results [12]. The antimicrobial activity of *T. capitatum* has also been thoroughly investigated, especially against *Klebsiella pneumoniae* [13] and *Salmonella typhimurium* [14]. The essential oils of *T. montanum* and *T. capitatum* have also been studied for their antiphytoviral activity against cucumber mosaic virus (CMV) in quinoa, which is due to the high content

of β-caryophyllene (52% in *T. capitatum*) and germacrene D (17.2%), β-pinene (12.3%) and limonene (4.6%) in the case of *T. montanum* [11]. In the majority of the research papers cited here that investigated the biochemical composition of the essential oil *T. capitatum* and its pharmacological potential, the species was referred to as *T. polium* or *T. polium* subsp. *capitatum*. From the authors' statements on the origin of the plant material (Balkan Peninsula, Iran) and the known distribution areas of *T. polium* and *T. capitatum* [15], we can conclude that they indeed refer to *T. capitatum*.

Compared to research on the molecular diversity and phylogenetics of *T. montanum* and *T. capitatum*, there is a larger number of studies on biochemical diversity. In other *Teucrium* species, there is a considerable amount of research on population genetics using different marker systems, such as the combination of chloroplast (*rpL32-trnL* and *trnL-trnF*) and nuclear regions (ITS) in *T. flavum* L. [16] and *T. scorodonia* L. [17], AFLP markers in *T. arduini* L. [18] or ISSR markers in *T. stocksianum* Boiss. [19]. There are only a handful of molecular studies on *T. capitatum*, which usually focus on the systematics of the *Teucrium* section *Polium* [20], an entire *Teucrium* genus [6] as a starting point for conservation [21]. A single molecular study investigated the genetic differences between *T. polium* and *T. capitatum* (referred to in the paper as *T. polium* subsp. *Capitatum*) [22]. A review of the literature revealed no studies addressing the molecular diversity of *T. montanum*.

Currently, no microsatellite markers for *T. montanum* or *T. capitatum* have been developed. Microsatellite markers (widely known as simple sequence repeats or SSRs) consist of a large number of copies of a nucleotide motif [23]. As a marker system, they are codominant and highly informative, making them ideal for population genetics, conservation genetics and phylogenetic studies [24,25]. Microsatellite markers are usually species-specific, i.e., they must be developed de novo for each species, which limits interspecies transfer in plants [26]. Another advantage of microsatellite markers lies in their potential to identify hybrid individuals in plant species with high levels of hybridization [27], especially when combined with increasingly new open-source statistical tools for the analysis of hybrid species [28,29].

The traditional approach to microsatellite marker development is both time-consuming and expensive. With the emergence of NGS sequencing, [30] this process has been significantly shortened and has become much more affordable. Additionally, SSRs developed in-silico exhibit a high level of polymorphism, especially in species of hybridogenous origin making them ideal for the identification of hybrid individuals on a molecular level [31]. Wide-range availability of NGS coupled with various pipelines for the rapid and effective development of highly polymorphic microsatellite markers [31,32] has led in recent years to the identification and research of hybrids in a variety of plant species such as silverleaf sunflower (*Helianthus argophyllus* Torr. & A.Gray) [33], bur-reed (*Sparganium* L.) [34] and cherries (*Prunus* × *yedoensis* Matsum.) [35] to name a few.

The aim of this research was: (a) to develop SSR markers for *T. montanum* and *T. capitatum* using the combination of NGS and a recently developed pipeline for SSR identification in related species, (b) to characterize the developed SSRs, and (c) to analyze their suitability for genetic diversity studies and the identification of putative hybrids.

2. Materials and Methods

2.1. Plant Material and DNA Isolation

All samples were collected in June 2021 in the central Adriatic hinterland, near Vojnić Sinjski where both *T. montanum* and *T. capitatum* as well as their hybrid *T.* × rohlenae are naturally distributed. In total, 34 individuals of *T. montanum*, 30 individuals of *T. capitatum* and 18 individuals of *T.* × *rohlenae* were sampled.

High molecular weight genomic DNA for NGS sequencing was extracted from fresh leaf tissue (500 mg) of one individual of each parental species using the CTAB isolation procedure [36] with some modifications. Immediately before extraction, 1% 2-mercaptoethanol, 1% polyvinylpyrrolidone (PVP) and 1% Rnase A (100 mg/mL) were added to the genomic lysis buffer. The isopropanol precipitation was replaced by precipitation with 2.5 volumes

of cold ethanol. The precipitated DNA was not centrifuged but fished out with the curved tip of a micropipette and transferred twice to 2 mL of 70% ethanol for washing. The isolated DNA was checked for concentration and purity using the NanoPhotometer P300 (Implen, Munich, Germany) and the Qubit™ fluorometer (Invitrogen, Carlsbad, CA, USA). The extracted DNA samples are stored at the Laboratory of Genetic Diversity, Phylogeny and Molecular Systematics of Plants, Faculty of Science, University of Zagreb, under accession numbers ZAGR 51322 (*T. montanum*) and ZAGR 51323 (*T. capitatum*).

Total genomic DNA of *T. montanum* and *T. capitatum* used for testing and characterization of the developed microsatellites, as well as from both parental species and the hybrids used in the genetic diversity study, was isolated from 25 mg of silica gel-dried leaf tissue. Prior to DNA isolation, the leaf tissue was ground to a fine powder using the TissueLyser II (Qiagen, Hilden, Germany). The DNeasy® Plant Mini Kit (Qiagen, Hilden, Germany) was used for DNA isolation, and the concentration and purity of the isolated DNA were measured with the NanoPhotometer P300 (Implen, Munich, Germany).

2.2. Next-Generation Sequencing, DNA Assembly and SSR Mining

The PacBio HiFi library was prepared using the SMRTbell™ Template Prep Kit 1.0 (Pacific Biosciences of California, Inc, San Diego, CA, USA). The library was sequenced using the PacBio 8M SMRT cell and the PacBio Sequel II System (Pacific Biosciences of California, Inc, San Diego, CA, USA) in the DNA Link Sequencing lab (DNA Link Inc., Seoul, Republic of Korea). The FastQC tool was used to check the quality of the raw sequences and the possible presence of the sequence adapters [37]. De novo assembly was performed using the de novo Assembly algorithm of the CLC Genomics Server ver.20.0.2 (Qiagen Bioinformatics, Aarhus, Denmark) with default parameters and high similarity fraction (0.95) and mapping option ON.

For SSR mining, the recently developed *Dig-up Primers* pipeline was utilized [31] in Python version 3 with the Biopython module [38]. The assemblies of *T. montanum* (Assembly A) and *T. capitatum* (Assembly B) were used as input in FASTA format. The pipeline consists of six steps and uses several external software: MISA [39] (https://mybiosoftware.com/misa-microsatellite-identification-tool.html), RepeatMasker 4.1.6 [40], Primer3 4.1.0 [41] and BLAST 2.15.0 [42]. In the first step, SSRs were identified using MISA. SSR motifs with the length of two and three nucleotides were searched for, whereby a minimum repeat length was defined (2: 10; 3: 7). In the second step, the SSR regions were analyzed and SSRs were filtered out based on several criteria: simplicity (composite SSRs were excluded), proximity to other SSRs, and proximity to the end of the contigs (a default value of 200 bp was used for the last two criteria). In the third step, primers were designed with Primer3 using the default settings for primer properties [31]. The fourth step included a low complexity and proximity to coding regions check with RepeatMasker. The SSR markers from both assemblies were amplified in silico with BLAST in the fifth step. In the final step, the markers were selected for testing if they were amplified once in both assemblies, the difference in the length of the amplified regions was due to polymorphism in the SSR length, and the amplified regions contained SSRs with the same motif as that for which the primers were designed.

2.3. Testing and Characterization of Developed SSR Markers

For initial testing, a subset of markers was selected that showed polymorphisms in the number of repeat motifs between two parental species. The selected SSR loci were first tested on five genotypes of each parental species to determine amplification success and polymorphisms. Amplification was performed in the ProflexTM PCR System (Thermo Fisher Scientific Inc., Waltham, MA, USA) using a two-step PCR protocol with an initial touchdown cycle: 94 °C for 5 min, followed by five cycles of 45 s at 94 °C, 30 s at 60 °C, 10 lowered by 1 °C in each cycle, 1 min 30 s at 72 °C, followed by 25 cycles of 45 s at 94 °C, 30 s at 55 °C, 1 min 30 s at 72 °C, and ending with an 8-min extension step at 72 °C. PCR products were run on an ABI 12 3730XL analyzer using a commercial fragment analysis

service (Macrogen Inc., Amsterdam, The Netherlands). The results were analyzed using GeneMapper 5.0 software (Thermo Fisher Scientific Inc).

Loci that were successfully amplified and determined to be polymorphic in initial testing were further tested in the same way, but with 28 *T. montanum*, 24 *T. capitatum*, and 18 hybrid individuals (according to their morphology).

For each developed SSR, the number of alleles (N_a), polymorphic information content (*PIC*) and probability of identity (PI) were determined using Cervus v. 3.0 [43]. All loci were checked for scoring errors, allelic dropout and the presence of null alleles using Micro-Checker v. 2.2.3 [44] and null allele frequencies were estimated in FreeNA [45] using the expectation-maximization algorithm.

2.4. Teucrium Diversity Analysis

The genetic diversity of *T. montanum*, *T. capitatum* and hybrid species (*T. × rohlenae*) was assessed by calculating the observed heterozygosity (H_O), expected heterozygosity (H_E) and inbreeding coefficient (F_{IS}) in Genepop v. 4.7 [46]. Deviations from Hardy–Weinberg equilibrium were determined in Genepop and significance levels were adjusted using the Bonferroni correction for multiple testing in SAS v. 9.4 [47]. The allelic richness (N_{ar}) of each taxon was estimated using HP-Rare v. 1.0 [48].

Pairwise genetic distances based on the proportion of shared alleles (D_{psa} [49]) between all individuals were calculated using MICROSAT [50] and used to construct a NeighborNet diagram using SplitsTree 4 [51]. Population differentiation was determined using pairwise F_{ST} estimates in FSTAT v. 2.9.3.2 [52], with *p*-values calculated after 1000 random permutations.

The genetic structure of the three *Teucrium* species was assessed using STRUCTURE v2.3.4 [53] with two approaches. The first approach included the standard analysis with a number of clusters of K = 1–6 to assess the overall structure. In the second analysis, the individuals belonging to the parental species (*T. montanum* and *T. capitatum*) were assigned a priori to their predefined clusters (POPFLAG = 1), while the unflagged individuals (POPFLAG = 0) were the hybrids (*T. × rohlenae*). Allele frequency estimation was thus based solely on the parental species (option PFROMPOPFLAGONLY) and the number of clusters was set to K = 2. The calculations were performed on the Isabella computer cluster at the University of Zagreb (Croatia), University Computing Center (SRCE). Thirty runs per K were performed, with each run including a burn-in period of 200,000 steps followed by 1,000,000 MCMC replicates using the Admixture model with correlated allele frequencies. In the first approach, the optimal number of clusters was determined by calculating ΔK [54] using the web-based software StructureSelector [55], which also allows clustering and merging of the resulting runs according to the CLUMPAK method described in Kopelman et al. [56]. An individual was assigned to a cluster if an arbitrary value of 75% of its genome was estimated to belong to that genetic cluster [57], while the individuals with a membership probability Q < 75% for all clusters were considered 'admixed'.

3. Results and Discussion

3.1. DNA Assembly and SSR Mining

Next-generation sequencing of *T. montanum* yielded a total of 1,564,267 high-quality reads, which were assembled into 120 contigs. The total length of the contigs was 387.6 Mbp and the length of the longest contig was 14,167,793 bp. The assembly had a mean GC content of 37.19% and an N50 value of 5,151,942 bp. A similar GC content was previously observed in *T. montanum* and other Teucrium species [58]. Next-generation sequencing of *T. capitatum* yielded a total of 1,962,544 high-quality reads assembled into 1685 contigs. The total length of the contigs was 816.4 Mbp and the length of the longest contig was 7,799,184. The assembly had an N50 value of 1,030,709 bp and a mean GC content of 32.88%, which is significantly lower than in other *Teucrium* taxa [58].

The assembly of *T. montanum* was used as Assembly A in the *Dig-up Primers* pipeline, as it is more suitable for initial SSR mining due to its lower number of contigs (the first four

steps of the pipeline). The *T. capitatum* assembly was used as Assembly B in the last two steps of the pipeline (Figure 2). We identified a total of 11,820 SSRs in Assembly A, most of which were dinucleotides (79%), which is consistent with other species in the Lamiaceae family such as *Ocimum basilicum* L. [59] and *Tectona grandis* L. [60]. After the second step of the pipeline (analysis of SSR regions), the number of SSRs was reduced to 10,601 based on the selection criteria. The number of eligible SSRs was further reduced to 5275 in the third step based on the selected primer design parameters. In the fourth step, regions with low complexity or close to the coding regions were excluded, leaving 2631 SSRs that were used in the fifth step (in silico PCR on both assemblies). In *T. montanum* 2095 SSRs were amplified, while in *T. capitatum* 219 SSRs were amplified. The number of SSRs that were amplified only once in both assemblies and were polymorphic was 112. Complex filtering criteria have already been used to successfully develop SSRs for several species with larger and more complex genomes than that of *Teucrium* taxa [61,62].
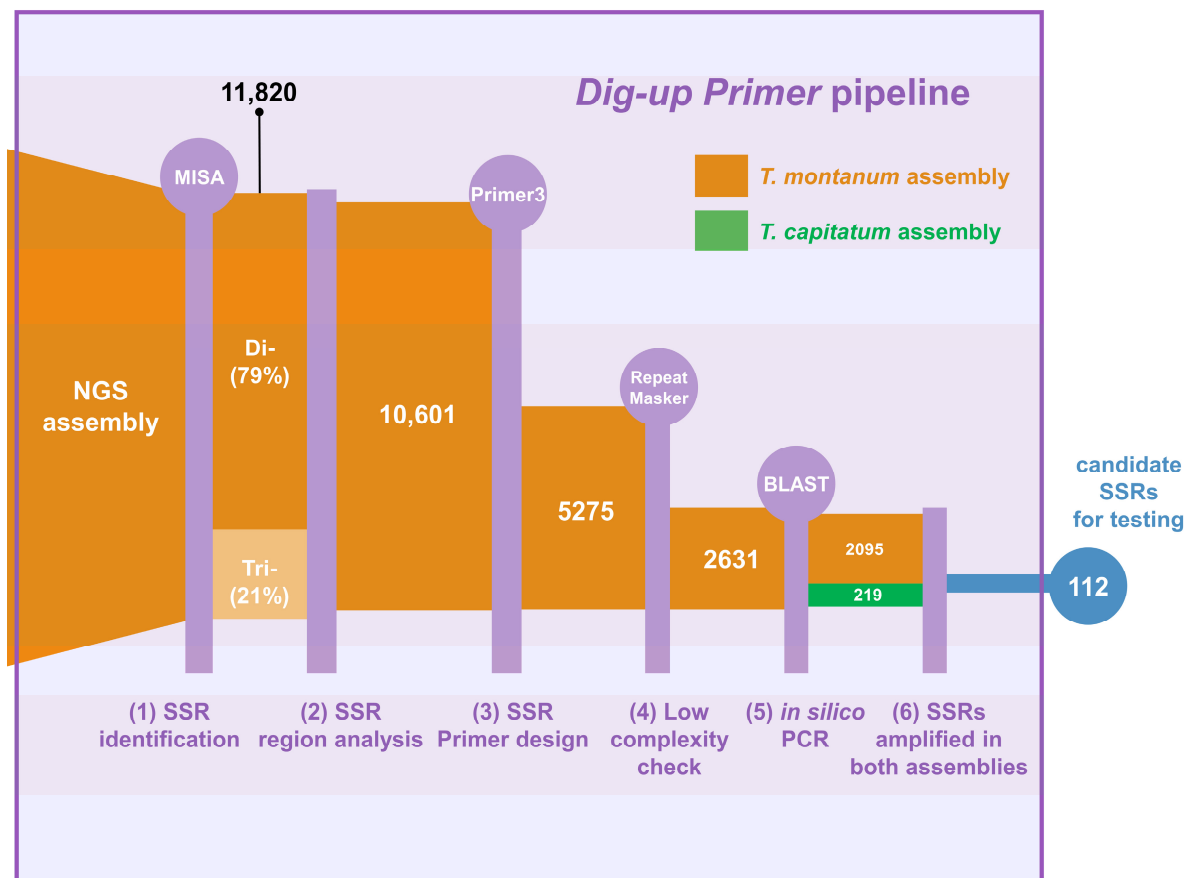
Figure 2. Identification of SSR candidate markers for *T. montanum* and *T. capitatum* using the *Dig-up Primer* pipeline.

3.2. Testing and Characterisation of Developed Markers

Of the 112 SSRs, 41 were selected on the basis of polymorphisms in the number of repeat motifs between two parental species for initial testing on five samples of each species. Eighteen loci were polymorphic, while 23 of them were either monomorphic or did not amplify at all. Further tests on individuals of *T. montanum* (28) and *T. capitatum* (24) and their hybrids (18) revealed twelve amplified loci in all individuals examined. Six of the original loci occurred in the form of so-called stuttering peaks, which were very difficult to analyze, with more than two amplified fragments in the same individual, or no amplification occurred in more than 25% of the individuals examined. The sequences of the newly developed SSRs were deposited in GenBank under accession numbers PP001804–PP001829 (Table 1).

Table 1. List of twelve newly developed SSRs for *T. montanum*/*T. capitatum* and their characterization based on the Vojnić Sinjski population.

Locus	Primer Sequences (5′–3′)	Repeat Motif	*T. montanum*		*T. capitatum*		Size Range (bp)	N_a	PIC	PI
			No. of Repeats	GeneBank Acc. No.	No. of Repeats	GeneBank Acc. No.				
TmUZ05/TpcUZ05	F: GGTAGGTCTTAATCTTGGGCACA R: TCGACAAACGGCTTCCTCTT	(TC)	14	PP001804	9	PP001805	196–274	25	0.925	0.009
TmUZ08/TpcUZ08	F: GCCTAGTTAAAGTGCTTCCAGC R: GGAAGACTGGCGGTAGTACG	(GA)	16	PP001806	31	PP001807	188–260	29	0.940	0.006
TmUZ09/TpcUZ09	F: ACTCAAGTCGTGTTTGGATCCA R: CCTCCTTCGATCGGTGCTTT	(CT)	15	PP001808	8	PP001809	140–178	18	0.857	0.029
TmUZ11/TpcUZ11	F: TTCCTTGCTTAGCCTGGAGC R: ACCTTGATAAAATGGAGCAGCC	(AG)	25	PP001810	33	PP001811	142–192	24	0.936	0.007
TmUZ14/TpcUZ14	F: GGGATTCCATGTGATTTGCCC R: CCATGAGAAGCACAAGGCCT	(TG)	19	PP001812	22	PP001813	80–146	26	0.922	0.010
TmUZ20/TpcUZ20	F: TGTGGAACGGTATGAAGCCT R: GTCAAGGTGGTGGGGTTGAT	(CT)	13	PP001814	7	PP001815	127–161	13	0.624	0.150
TmUZ26/TpcUZ26	F: GGATGTGCAATGTTGTGTCACT R: TGCAACCAACAGATGTGCTTG	(TC)	20	PP001816	12	PP001817	232–274	18	0.912	0.013
TmUZ31/TpcUZ31	F: GAGGAGAAGAGCATCACCCG R: CTCAATTCCTGAGGACGGCT	(AT)	15	PP001820	17	PP001821	232–286	23	0.941	0.006
TmUZ32/TpcUZ32	F: GCGGCTTTTCCTCCCTACAA R: AGTGACGGATGCCCATTGG	(TC)	13	PP001822	9	PP001823	243–291	22	0.906	0.013
TmUZ35/TpcUZ35	F: TCGGGCCAGGATAAGGTGTA R: TCGATGGTCGAGACTCAGGT	(AG)	18	PP001824	22	PP001825	170–230	23	0.929	0.009
TmUZ43/TpcUZ43	F: GACCACTAAACCAGAAGGGCA R: CAAGCCTCTCTCCAACCGAG	(TAA)	7	PP001826	4	PP001827	207–228	5	0.616	0.162
TmUZ44/TpcUZ44	F: CTTCCCCTCACACAAGTCCC R: GCCTCCTCGCTTTACTTTGC	(GAA)	8	PP001828	4	PP001829	233–248	6	0.552	0.209

N_a—number of alleles; *PIC*—polymorphic information content; *PI*—probability of identity.

A total of 232 alleles ranging from 5 (TmUZ43/TpcUZ43) to 29 (TmUZ08/TpcUZ08) were detected with an average of 19.33 alleles per locus. The probability of identity ranged from 0.006 (TmUZ08/TpcUZ08 and TmUZ31/TpcUZ31) to 0.209 (TmUZ44/TpcUZ44) with a combined probability of identity for all loci of 1.04×10^{-17}. On the basis of polymorphic information content (*PIC*), nine of the twelve SSRs were classified as highly informative (*PIC* > 0.70) and the other three as moderately polymorphic with a *PIC* higher than 0.44 [63].

3.3. Teucrium Genetic Diversity in Vojnić Sinjski

The genetic parameters used to analyze three *Teucrium* taxa at the Vojnić Sinjski locality showed similar patterns. In *T. montanum*, the highest values were obtained for allelic richness (13.17), observed (0.807) and expected heterozygosity (0.873). In contrast, *T. capitatum* showed the lowest values of allelic richness (10.27), observed (0.656) and expected heterozygosity (0.713). The values of the genetic parameters in the putative hybrid species *T.* × *rohlenae* were intermediate (Table 2). A significant deviation from Hardy–Weinberg equilibrium was observed in all three taxa. Null alleles were found in three of 36 loci × taxa combinations, one in *T. montanum* (estimated null allele frequency of 0.15 for TmUZ31/TpcUZ31) and two in the putative hybrid species (estimated null frequencies of 0.079 and 0.061 for TmUZ32/TpcUZ32 and TmUZ35/TpcUZ35, respectively). Examination of the shared alleles revealed that of 232 alleles, 36.21% were shared between all three taxa (Figure 3A). A further 13.79% of alleles were shared between *T. montanum* and the putative hybrid species, while 20.26% of alleles were found only in *T. montanum*.

Table 2. Genetic diversity of three *Teucrium* taxa from Vojnić Sinjski, Croatia, based on twelve SSR markers.

| Taxon	n	N_a	N_{ar}	N_{pr}	H_O	H_E	F_{IS}	p (F_{IS})
T. montanum	28	15.417	13.166	47	0.807	0.873	0.076	***
T. × *rohlenae*	18	12.417	12.417	18	0.759	0.805	0.056	**
T. capitatum	24	11.250	10.269	14	0.656	0.713	0.080	***

n—sample size; N_a—average number of alleles; N_{ar}—allelic richness; N_{pr}—number of private alleles; H_O—observed heterozygosity; H_E—expected heterozygosity; F_{IS}—inbreeding coefficient (significance levels: **—significant at $p < 0.01$; ***—significant at $p < 0.001$).

The lowest pairwise differentiation between taxa was found for *T. capitatum* and the putative hybrid species (F_{ST} = 0.008) while the highest differentiation was found between the parent taxa (F_{ST} = 0.153). The NeighborNet diagram based on the proportion of shared alleles (D_{psa}) between all 70 individuals clearly separated *T. montanum* from *T. capitatum*. The hybrid species was mostly admixed with *T. capitatum*, with no evidence of grouping (Figure 3B).

The optimal number of genetic clusters was first determined using standard STRUCTURE analysis, with no prior taxa information considered. The highest ΔK was observed at K = 2 (37.03), while the second best (at K = 3; ΔK = 8.21), as expected, did not group the hybrid individuals as members of the putative third genetic cluster. At K = 2, all but one individual of *T. montanum* were assigned to cluster A, with the proportion of membership being higher than Q > 0.75. Similarly, all but two individuals of *T. capitatum* were assigned to cluster B. The hybrid individuals (*T.* × *rohleane*) were clearly admixed, as their proportion of membership was lower than Q < 0.75 in both clusters, with the exception of one of 17 individuals.

The second approach, using predefined clusters for individuals belonging to the parental species, produced a similar result. In addition, the gradual transition of hybrids from one species to another was shown, suggesting a long-term introgression between the parental species involving the formation of individuals belonging to different degrees of hybridization, including not only F1 or F2 hybrids but also recurrent backcrosses to the parental genotypes.
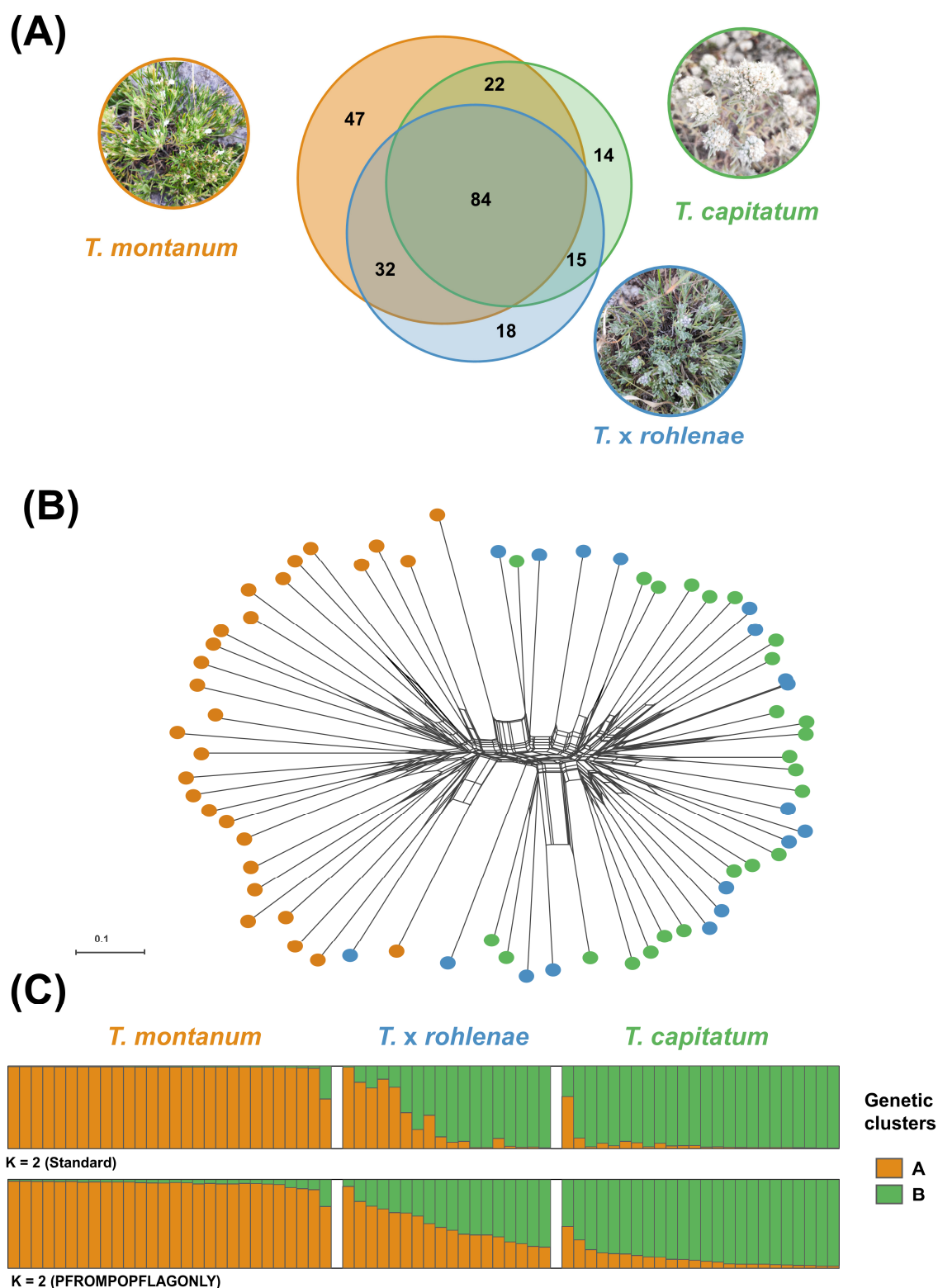
Figure 3. Genetic diversity and structure of three *Teucrium* taxa (70 individuals) from Vojnić Sinjski, Croatia based on 12 newly developed SSR markers. (**A**) Venn diagram visualizing the proportion of private and shared alleles between the three taxa. (**B**) NeighborNet diagram based on the proportion of shared alleles distance (D_{psa}). (**C**) Genetic structure based on the standard STRUCTURE analysis (at K = 2) and using the PFROMPOPFLAGONLY option. Each individual is represented by a column and the probability of membership to each of the two genetic clusters is shown on a scale of 0–100%.

4. Conclusions

Our results confirm that the recently developed *Dig-up Primers* pipeline can be used to mine for SSRs in closely related taxa and selected markers were successfully amplified in individuals used for testing.

The newly developed SSRs for *T. montanum* and *T. capitatum* can be successfully used to analyze the genetic diversity of both species and to assess the extent of their hybridization process in situ. Genetic diversity and genetic structure analysis suggested that *T. × rohleane* is a species of hybridogenous origin, which is in agreement with previous research on the three *Teucrium* taxa conducted on the morphological and biochemical levels.

The preliminary results, based on a single location where all three taxa occur, are very encouraging for future studies on hybridization between *T. montanum* and *T. capitatum* at other locations throughout the distribution range of both species and will allow the elucidation of the complex taxonomy and phylogeography of these closely related species.

References

1. Dinić, J.; Novaković, M.; Pešić, M. Potential for cancer treatment: Natural products from the Balkans. In *Biodiversity and Biomedicine: Our Future*; Ozturk, M., Egamberdieva, D., Pešić, M., Eds.; Academic Press: Cambridge, MA, USA, 2020; pp. 137–159, ISBN 9780128195413.
2. Navarro, T.; El Oualidi, J. Synopsis of *Teucrium* L. (Labiatae) in the Mediterranean region and surrounding areas. *Flora Medit.* **2000**, *10*, 349–363.
3. Nikolić, T. (Ed.) Flora Croatica Database. Available online: http://hirc.botanic.hr/fcd (accessed on 14 February 2024).
4. Zbiljić, M.; Lakušić, B.; Marčetić, M.; Bogdanović, S.; Lakušić, D. Morphological and chemical evidence of *Teucrium × rohlenae* K. Malý (Lamiaceae), a new hybrid in Croatia. *Acta Bot. Croat.* **2021**, *80*, 48–55. [CrossRef]
5. Navarro, T. *Teucrium* L. In *Flora Iberica 12*; Morales, R., Quintanar, A., Cabezas, F., Pujadas, A.J., Cirujano, S., Eds.; Real Jardín Botánico, CSIC: Madrid, Spain, 2010; pp. 30–166.
6. Salmaki, Y.; Kattari, S.; Heubl, G.; Bräuchle, C. Phylogeny of non-monophyletic *Teucrium* (Lamiaceae: Ajugoideae): Implications for character evolution and taxonomy. *Taxon* **2016**, *65*, 805–822. [CrossRef]
7. Redzić, S.S. The ecological aspect of ethnobotany and ethnopharmacology of population in Bosnia and Herzegovina. *Coll. Antropol.* **2007**, *31*, 869–890.
8. Šarić-Kundalić, B.; Dobeš, C.; Klatte-Asselmeyer, V.; Saukel, J. Ethnobotanical study on medicinal use of wild and cultivated plants in middle, south and west Bosnia and Herzegovina. *J. Ethnopharmacol.* **2010**, *131*, 33–55. [CrossRef]
9. Zlatković, B.K.; Bogosavljević, S.S.; Radivojević, A.R.; Pavlović, M.A. Traditional use of the native medicinal plant resource of Mt. Rtanj (Eastern Serbia): Ethnobotanical evaluation and comparison. *J. Ethnopharmacol.* **2014**, *151*, 704–713. [CrossRef]
10. Varga, F.; Šolić, I.; Jug Dujaković, M.; Łuczaj, Ł.; Grdiša, M. The first contribution to the ethnobotany of inland Dalmatia: Medicinal and wild food plants of the Knin area, Croatia. *Acta Soc. Bot. Pol.* **2019**, *88*, 1–20. [CrossRef]
11. Bezić, N.; Vuko, E.; Dunkić, V.; Ruščić, M.; Blažević, I.; Burčul, F. Antiphytoviral Activity of Sesquiterpene-Rich Essential Oils from Four Croatian *Teucrium* Species. *Molecules* **2011**, *16*, 8119. [CrossRef]

12. Abdollahi, M.; Karimpour, H.; Monsef-Esfehani, H.R. Antinociceptive effects of *Teucrium polium* L total extract and essential oil in mouse writhing test. *Pharmacol. Res.* **2003**, *48*, 31–35. [CrossRef]
13. Raei, F.; Ashoori, N.; Eftekhar, F.; Yousefzadi, M. Chemical composition and antibacterial activity of *Teucrium polium* essential oil against urinary isolates of *Klebsiella pneumoniae*. *J. Essent. Oil Res.* **2014**, *26*, 65–69. [CrossRef]
14. Mahmoudi, R.; Zare, P.; Nosratpour, S. Application of *Teucrium polium* Essential Oil and *Lactobacillus casei* in Yoghurt. *J. Essent. Oil Bear. Plants* **2015**, *18*, 477–481. [CrossRef]
15. POWO. Plants of the World Online (POWO). Available online: https://powo.science.kew.org/ (accessed on 10 February 2024).
16. Djabou, N.; Battesti, M.J.; Allali, H.; Desjobert, J.M.; Varesi, L.; Costa, J.; Muselli, A. Chemical and genetic differentiation of Corsican subspecies of *Teucrium flavum* L. *Phytochemistry* **2011**, *72*, 1390–1399. [CrossRef]
17. Djabou, N.; Allali, H.; Battesti, M.J.; Tabti, B.; Costa, J.; Muselli, A.; Varesi, L. Chemical and genetic differentiation of two Mediterranean subspecies of *Teucrium scorodonia* L. *Phytochemistry* **2012**, *74*, 123–132. [CrossRef]
18. Kremer, D.; Bolarić, S.; Ballian, D.; Bogunić, F.; Stešević, D.; Karlović, K.; Kosalec, I.; Vokurka, A.; Rodríguez, J.V.; Randić, M.; et al. Morphological, genetic and phytochemical variation of the endemic *Teucrium arduini* L. (Lamiaceae). *Phytochemistry* **2015**, *116*, 111–119. [CrossRef]
19. Kamali, M.; Samsampour, D.; Bagheri, A.; Mehrafarin, A.; Homaei, A. Association analysis and evaluation of genetic diversity of *Teucrium stocksianum* Boiss. populations using ISSR markers. *Genet. Resour. Crop Evol.* **2023**, *70*, 691–709. [CrossRef]
20. El Oualidi, J.; Verneau, O.; Puech, S.; Dubuisson, J.Y. Utility of rDNA ITS sequences in the systematics of *Teucrium* section *Polium* (*Lamiaceae*). *Plant Syst. Evol.* **1999**, *215*, 49–70. [CrossRef]
21. Marzouk, R.I.; El-Badan, G.E. Research article molecular characterization of *Teucrium* L. (lamiaceae) as a prerequisite for its conservation. *J. Biol. Sci.* **2018**, *11*, 16–23.
22. Djabou, N.; Muselli, A.; Allali, H.; Dib, M.E.A.; Tabti, B.; Varesi, L.; Costa, J. Chemical and genetic diversity of two Mediterranean subspecies of *Teucrium polium* L. *Phytochemistry* **2012**, *83*, 51–62. [CrossRef]
23. Zane, L.; Bargelloni, L.; Patarnello, T. Strategies for microsatellite isolation: A review. *Mol. Ecol.* **2002**, *11*, 1–16. [CrossRef]
24. Vieira, M.L.C.; Santini, L.; Diniz, A.L.; Munhoz, C.d.F. Microsatellite markers: What they mean and why they are so useful. *Genet. Mol. Biol.* **2016**, *39*, 312. [CrossRef]
25. Kalia, R.K.; Rai, M.K.; Kalia, S.; Singh, R.; Dhawan, A.K. Microsatellite markers: An overview of the recent progress in plants. *Euphytica* **2010**, *177*, 309–334. [CrossRef]
26. Barbará, T.; Palma-Silva, C.; Paggi, G.M.; Bered, F.; Fay, M.F.; Lexer, C. Cross-species transfer of nuclear microsatellite markers: Potential and limitations. *Mol. Ecol.* **2007**, *16*, 3759–3767. [CrossRef]
27. Kong, S.; Kubatko, L.S. Comparative Performance of Popular Methods for Hybrid Detection using Genomic Data. *Syst. Biol.* **2021**, *70*, 891–907. [CrossRef]
28. Goulet, B.E.; Roda, F.; Hopkins, R. Hybridization in Plants: Old Ideas, New Techniques. *Plant Physiol.* **2017**, *173*, 65–78. [CrossRef]
29. Neophytou, C. Bayesian clustering analyses for genetic assignment and study of hybridization in oaks: Effects of asymmetric phylogenies and asymmetric sampling schemes. *Tree Genet. Genomes* **2014**, *10*, 273–285. [CrossRef]
30. Zalapa, J.E.; Cuevas, H.; Zhu, H.; Steffan, S.; Senalik, D.; Zeldin, E.; McCown, B.; Harbut, R.; Simon, P. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *Am. J. Bot.* **2012**, *99*, 193–208. [CrossRef]
31. Turudić, A.; Liber, Z.; Grdiša, M.; Jakše, J.; Varga, F.; Poljak, I.; Šatović, Z. *Dig-up Primers*: A Pipeline for Identification of Polymorphic Microsatellites Loci within Assemblies of Related Species. *Int. J. Mol. Sci.* **2024**, *25*, 3169. [CrossRef]
32. Wang, H.; Gao, S.; Liu, Y.; Wang, P.; Zhang, Z.; Chen, D. A pipeline for effectively developing highly polymorphic simple sequence repeats markers based on multi-sample genomic data. *Ecol. Evol.* **2022**, *12*, e8705. [CrossRef]
33. Makarenko, M.S.; Azarin, K.V.; Gavrilova, V.A. Mitogenomic Research of Silverleaf Sunflower (*Helianthus argophyllus*) and Its Interspecific Hybrids. *Curr. Issues Mol. Biol.* **2023**, *45*, 4841–4849. [CrossRef]
34. Belyakov, E.A.; Александрович, Б.Е.; Machs, E.M.; Модрисович, М.Э.; Mikhailova, Y.V.; Владимировна, М.Ю.; Rodionov, A.V.; Викеньтьевич, Р.А. The study of hybridization processes within genus *Sparganium* L. Subgenus Xanthosparganium holmb. Based on data of next generation sequencing (NGS). *Ecol. Genet.* **2019**, *17*, 27–35. [CrossRef]
35. Baek, S.; Choi, K.; Kim, G.B.; Yu, H.J.; Cho, A.; Jang, H.; Kim, C.; Kim, H.J.; Chang, K.S.; Kim, J.H.; et al. Draft genome sequence of wild *Prunus yedoensis* reveals massive inter-specific hybridization between sympatric flowering cherries. *Genome Biol.* **2018**, *19*, 127. [CrossRef]
36. Doyle, J.J.; Doyle, J.L. Isolation of plant DNA from fresh tissue. *Focus* **1990**, *12*, 13–15.
37. Andrews, S. FastQC—A Quality Control Tool for High Throughput Sequence Data. *Babraham Bioinforma.* 2010. Available online: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on 10 December 2023).
38. Cock, P.J.A.; Antao, T.; Chang, J.T.; Chapman, B.A.; Cox, C.J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; et al. Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **2009**, *25*, 1422–1423. [CrossRef]
39. Thiel, T.; Michalek, W.; Varshney, R.K.; Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* **2003**, *106*, 411–422. [CrossRef]
40. Smit, A.; Hubley, R.; Green, P. RepeatMasker Open-4.0.6 2013–2015. 2018. Available online: http://www.repeatmasker.org (accessed on 11 November 2023).

41. Untergasser, A.; Cutcutache, I.; Koressaar, T.; Ye, J.; Faircloth, B.C.; Remm, M.; Rozen, S.G. Primer3-new capabilities and interfaces. *Nucleic Acids Res.* **2012**, *40*, e115. [CrossRef]

42. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef]

43. Kalinowski, S.T.; Taper, M.L.; Marshall, T.C. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* **2007**, *16*, 1099–1106. [CrossRef]

44. Van Oosterhout, C.; Hutchinson, W.F.; Wills, D.P.M.; Shipley, P. MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* **2004**, *4*, 535–538. [CrossRef]

45. Chapuis, M.P.; Estoup, A. Microsatellite null alleles and estimation of population differentiation. *Mol. Biol. Evol.* **2007**, *24*, 621–631. [CrossRef]

46. Rousset, F. genepop'007: A complete re-implementation of the genepop software for Windows and Linux. *Mol. Ecol. Resour.* **2008**, *8*, 103–106. [CrossRef]

47. SAS Institute Inc. *Base SAS 9.4 Procedures Guide: Statistical Procedures*; SAS Institute: Cary, NC, USA, 2011; ISBN 978-1-60764-944-1.

48. Kalinowski, S.T. HP-RARE 1.0: A computer program for performing rarefaction on measures of allelic richness. *Mol. Ecol. Notes* **2005**, *5*, 187–189. [CrossRef]

49. Bowcock, A.M.; Ruiz-Linares, A.; Tomfohrde, J.; Minch, E.; Kidd, J.R.; Cavalli-Sforza, L.L. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **1994**, *368*, 455–457. [CrossRef] [PubMed]

50. Minch, E.; Ruiz-Linares, A.; Goldstein, D.; Feldman, M.; Cavalli-Sforza, L.L. *Microsat (Version 1.5b): A Computer Program for Calculating Various Statistics on Microsatellite Allele Data*; Stanford University: Stanford, CA, USA, 1997.

51. Huson, D.H.; Bryant, D. Application of Phylogenetic Networks in Evolutionary Studies. *Mol. Biol. Evol.* **2006**, *23*, 254–267. [CrossRef] [PubMed]

52. Goudet, J. FSTAT, Version 2.9.4, A Program to Estimate and Test Gene Diversities and Fixation Indices; 2002. Available online: https://www2.unil.ch/popgen/softwares/fstat.htm (accessed on 11 November 2023).

53. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **2000**, *155*, 945–959. [CrossRef] [PubMed]

54. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **2005**, *14*, 2611–2620. [CrossRef] [PubMed]

55. Li, Y.L.; Liu, J.X. StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. *Mol. Ecol. Resour.* **2018**, *18*, 176–177. [CrossRef] [PubMed]

56. Kopelman, N.M.; Mayzel, J.; Jakobsson, M.; Rosenberg, N.A.; Mayrose, I. Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* **2015**, *15*, 1179–1191. [CrossRef] [PubMed]

57. Matsuoka, Y.; Vigouroux, Y.; Goodman, M.M.; Sanchez, J.G.; Buckler, E.; Doebley, J. A single domestication for maize shown by multilocus microsatellite genotyping. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 6080–6084. [CrossRef] [PubMed]

58. Šmarda, P.; Knápek, O.; Březinová, A.; Horová, L.; Grulich, V.; Danihelka, J.; Veselý, P.; Šmerda, J.; Rotreklová, O.; Bureš, P. Genome sizes and genomic guanine+cytosine (GC) contents of the Czech vascular flora with new estimates for 1700 species. *Preslia* **2019**, *91*, 117–142. [CrossRef]

59. Rastogi, S.; Meena, S.; Bhattacharya, A.; Ghosh, S.; Shukla, R.K.; Sangwan, N.S.; Lal, R.K.; Gupta, M.M.; Lavania, U.C.; Gupta, V.; et al. De novo sequencing and comparative analysis of holy and sweet basil transcriptomes. *BMC Genom.* **2014**, *15*, 588. [CrossRef]

60. Yasodha, R.; Vasudeva, R.; Balakrishnan, S.; Sakthi, A.R.; Abel, N.; Binai, N.; Rajashekar, B.; Bachpai, V.K.W.; Pillai, C.; Dev, S.A. Draft genome of a high value tropical timber tree, Teak (*Tectona grandis* L. f): Insights into SSR diversity, phylogeny and conservation. *DNA Res.* **2018**, *25*, 409–419. [CrossRef]

61. Varga, F.; Liber, Z.; Jakše, J.; Turudić, A.; Šatović, Z.; Radosavljević, I.; Jeran, N.; Grdiša, M. Development of Microsatellite Markers for *Tanacetum cinerariifolium* (Trevis.) Sch. Bip., a Plant with a Large and Highly Repetitive Genome. *Plants* **2022**, *11*, 1778. [CrossRef]

62. Ebrahimi, A.; Mathur, S.; Lawson, S.S.; LaBonte, N.R.; Lorch, A.; Coggeshall, M.V.; Woeste, K.E. Microsatellite Borders and Micro-sequence Conservation in Juglans. *Sci. Rep.* **2019**, *9*, 3748. [CrossRef]

63. Hildebrand, C.E.; Torney, D.C.; Wagner, R.P. Informativeness of polymorphic DNA markers. *Los Alamos Sci.* **1992**, *20*, 100–102. [CrossRef]