

Article

Development of Microsatellite Markers for *Tanacetum cinerariifolium* (Trevis.) Sch. Bip., a Plant with a Large and Highly Repetitive Genome

Filip Varga ^{1,2}, Zlatko Liber ^{2,3,*}, Jernej Jakše ⁴, Ante Turudić ², Zlatko Šatović ^{1,2}, Ivan Radosavljević ^{2,3}, Nina Jeran ¹ and Martina Grdiša ^{1,2}

- ¹ Department of Seed Science and Technology, Faculty of Agriculture, University of Zagreb, Svetošimunska c. 25, 10000 Zagreb, Croatia; fvarga@agr.hr (F.V.); zsatovic@agr.hr (Z.Š.); njeran@agr.hr (N.J.); mgrdisa@agr.hr (M.G.)
- ² Centre of Excellence for Biodiversity and Molecular Plant Breeding (CoE CroP-BioDiv), Svetošimunska c. 25, 10000 Zagreb, Croatia; aturudic@agr.hr (A.T.); ivan.radosavljevic@biol.pmf.hr (I.R.)
- ³ Department of Biology, Faculty of Science, University of Zagreb, Marulićev trg 9a, 10000 Zagreb, Croatia
- ⁴ Department of Agronomy, Biotechnical Faculty, University of Ljubljana, Jamnikarjeva 101, 1000 Ljubljana, Slovenia; jernej.jakse@bf.uni-lj.si
- * Correspondence: zlatko.liber@biol.pmf.hr; Tel.: +385-1-4898-092

Abstract: Dalmatian pyrethrum (*Tanacetum cinerariifolium* (Trevis.) Sch. Bip.) is an outcrossing plant species ($2n = 18$) endemic to the eastern Adriatic coast and source of the natural insecticide pyrethrin. Due to the high repeatability and large genome (1C-value = 9.58 pg) our previous attempts to develop microsatellite markers using the traditional method were unsuccessful. Now we have used Illumina paired-end whole genome sequencing and developed a specific procedure to obtain useful microsatellite markers. A total of 796,130,142 high-quality reads (approx. $12.5\times$ coverage) were assembled into 6,909,675 contigs using two approaches (de novo assembly and joining of overlapped pair-end reads). A total of 31,380 contigs contained one or more microsatellite sequences, of which di-(59.7%) and trinucleotide (25.9%) repeats were the most abundant. Contigs containing microsatellites were filtered according to various criteria to achieve better yield of functional markers. After two rounds of testing, 17 microsatellite markers were developed and characterized in one natural population. Twelve loci were selected for preliminary genetic diversity analysis of three natural populations. Neighbor-joining tree, based on the proportion of shared alleles distances, grouped individuals into clusters according to population affiliation. The availability of codominant SSR markers will allow analysis of genetic diversity and structure of natural Dalmatian pyrethrum populations as well as identification of breeding lines and cultivars.

Keywords: *Tanacetum cinerariifolium*; Asteraceae; genomic SSRs



Citation: 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. <https://doi.org/10.3390/plants11131778>

Academic Editor: Oleg Sergeevich Alexandrov

Received: 6 June 2022

Accepted: 2 July 2022

Published: 5 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Balkans is a well-known European hotspot of plant biodiversity with a large number of aromatic and medicinal plant species distributed across the peninsula [1]. One such species is Dalmatian pyrethrum (*Tanacetum cinerariifolium* (Trevis.) Sch. Bip.; Asteraceae, Anthemideae) that inhabits dry grasslands with shallow, rocky soils [2] along the eastern Adriatic coast and is easily recognized during the flowering season by its flower heads, which consist of white petal-like ray florets at the edges and yellow disk florets in the centre of the flower head [3]. Pyrethrum plants produce pyrethrins (pyrethrin I and II, cinerin I and II, jasmolin I and II), which have a *knock-down* effect due to disruption of the insect's nervous system, followed by paralysis and death [4]. Pyrethrins are one of the most commercially used plant insecticides and the species has a long history of cultivation and domestic and agricultural use, with records of commercial cultivation in Dalmatia dating back to 1850s [5]. This region remained the world's leading producer of pyrethrum until

World War I. By then, production was introduced in other countries such as Japan [6] and Kenya [7], from where it spread further. Production in Croatia (part of Yugoslavia at the time) gradually declined and eventually ceased altogether, in part due to the discovery and widespread use of DDT [8]. Today, the largest producers of pyrethrum are concentrated in East Africa (Tanzania, Rwanda, and Kenya), Oceania (Papua New Guinea) [9] and Australia (Tasmania) [10].

The pyrethrin content and composition of Croatian natural Dalmatian pyrethrum populations have been largely investigated, especially in the last 15 years. Pyrethrin content reaches up to 1.35% of flower dry weight [11–13] which is lower than in commercial lines such as pyrethrum grown in Australia, where pyrethrin content can reach up to 2.5% of flower dry weight [14].

While pyrethrin content has been well studied in natural pyrethrum populations, far fewer studies have been conducted on the genetic diversity of *T. cinerariifolium* [15,16]. In addition, to date, no studies have explored the association between genetic and biochemical diversity in this species. One of the first studies attempted to clarify the relationship between ploidy level and certain morphological traits of Dalmatian pyrethrum in order to easily distinguish diploids from triploids based on morphology [17]. Recent research has focused more on identifying candidate genes involved in the biosynthesis of pyrethrin compounds, such as the chrysanthemyl diphosphate synthase gene [18–21] and genes involved in the in the 2-C-methyl-D-erythritol 4-phosphate pathway (MEP) pathway [22]. Another study that focused on the variation of tandem repeats in subtelomeres between individuals of this species showed high polymorphism of subtelomeres based on detailed FISH (fluorescence in situ hybridization) analysis [16]. A draft genome of the species was also constructed, revealing some genes potentially encoding enzymes specific for pyrethrin biosynthesis [23]. To date, only one study on the genetic diversity of natural populations of Dalmatian pyrethrum has been conducted [15]. In this comprehensive study (20 populations sampled along the Croatian Adriatic coast) based on Amplified Fragment Length Polymorphism (AFLP) markers, researchers discovered high gene diversity and high number of private alleles in the northern Adriatic populations, which gradually decreased towards the south, most likely due to historical overexploitation.

Microsatellite markers (single sequence repeats—SSRs) are more informative than AFLPs because they are codominant, locus-specific and more reproducible [24]. They are often used as molecular markers in population genetics due to the high level of polymorphism [25,26]. The presence of highly repetitive sequences and the large genome size have caused the development of SSR markers for this species to fail, as has been the case for species with similar genome features when a traditional approach consisting of restriction digestion, hybridization, library construction, cloning and Sanger sequencing has been used [27–29]. The advent of next generation sequencing (NGS) has enabled the efficient generation of a large amount of genome sequence data from which microsatellite primers can be identified and has significantly reduced the cost and time required for microsatellite marker development [30]. Compared to other plant species native to the Balkans that are agriculturally exploited (oregano, Dalmatian sage), genetic research on Dalmatian pyrethrum is severely underdeveloped. At the time of writing, only 418 nucleotide records for this species are available in NCBI (24 January 2022). Genomic characterization of this species using codominant molecular markers is urgently needed not only from the perspective of plant genetic resources conservation, but also for utilization in breeding programs aimed at reviving commercial production of Dalmatian pyrethrum in Croatia.

The objective of this research was to develop SSR markers by means of Next Generation Sequencing, which can then be used in the population genetics study of *T. cinerariifolium*.

2. Results and Discussion

2.1. Contigs Assembly and Genome-Wide Identification of SSR Markers

A total of 796,130,142 high-quality reads (120 Gbp in total) were obtained by next generation sequencing of *T. cinerariifolium* with a GC content of 35.89%. Low GC content

was previously observed in species with rather small or very large genomes such as *Elaeagnus angustifolia* L. (35.0%), *Trifolium striatum* L. (35.6%) [31] and *Helichrysum italicum* (Roth) G. Don (34.1%) [32]). The fact that GC base synthesis requires more biochemical cost than AT synthesis may be one of the main reasons why larger genomes tend to have lower GC content [33]. On the other hand, the prevalence of AT-rich motifs and low frequency of GC-rich motifs is the characteristic of dicotyledonous species compared to monocotyledonous species [34].

The first approach (de novo assembly of subset of the data) resulted in 923,207 contigs. The total length of the assembled contigs was 409.5 Mbp. The length of the longest contig was 82,765 bp and the N50 value of the assembly was 451 bp. The assembly had a mean GC content of 34.34%. The second approach (assembly of overlapping pair-end reads) yielded 5,986,468 sequences with a total length of 1311 Mbp (representing 13.7% of the estimated genome size) and a mean GC content of 34.41% (Table 1).

Table 1. Summary of de novo assembly and SSR mining.

Category	First Approach	Second Approach	Combined Results
Total number of contigs examined	923,207	5,986,468	6,909,675
Total size of examined contigs (bp)	409,503,869	1,311,396,942	1,720,900,811
Total number of identified SSRs	11,652	23,904	35,556
SSR containing contigs	9973	21,407	31,380
Contigs containing more than one SSR	1267	2005	3272
GC content (%)	34.34	34.41	35.89

Screening of the obtained contigs (6,909,675) for microsatellites without mononucleotide repeats revealed 35,556 microsatellite loci in 31,380 SSR-containing contigs. The majority of SSR-containing contigs (28,108—89.6%) contained only one microsatellite locus, while 3272 (10.4%) contained more than one microsatellite locus (Table 1). Dinucleotide repeats were the most common (nearly 60% of the remaining loci), followed by trinucleotide repeats (more than a quarter of the remaining loci). Tetra-, penta-, and hexanucleotides had similarly low frequencies (Figure 1). These results are consistent with other species from the Asteraceae family such as *Conyza canadensis* (L.) Cronquist. [35], *Cynara cardunculus* var. *scolymus* L. [36] and *H. italicum* [32]. The most common dinucleotide motifs were AT/AT, accounting for 29.6% of the total repeats, followed by AG/CT (15.1%) and AC/GT (15%). Among trinucleotides, AAT/ATT (10.4%) was the most common motif, followed by AAC/GTT (4.5%) and AGT/ATC (3.1%) (Figure 1).

2.2. Selection, Testing, and Characterization of SSR Markers

The 35,556 identified microsatellite loci were selected using described filtering criteria. Only di- and trinucleotide motifs were considered, as well as loci with GC content similar to average assembly GC content ($43 \pm 10\%$). Sequences with majority reads mapped with multi-mapping occurrences or sequences with average coverage not close to sequencing coverage ($12.5\times$) were also excluded, as were sequences with hits to repetitive elements occurrences, as identified by RepeatMasker. Filtering of the microsatellite loci according to the selection criteria resulted in 1006 sequences that were checked for the number of occurrences against the draft genome of *T. cinerariifolium* [23], which further narrowed the selection to 56 sequences that were used for primer design. These 56 primer pairs were tested on five samples from different populations. Even 39 of them were excluded from further analysis because they were not polymorphic between samples, more than two alleles were detected in the same sample, or PCR products were completely absent. The remaining 17 (30.4%) microsatellites that showed both good amplification and polymorphism were selected for further testing on 20 samples from one Dalmatian pyrethrum population (MAP02806). Similar approach using various multiple filtering criteria such as annealing temperature, GC content, product length etc. was previously successfully used to develop

SSR markers in other species [37,38]. Microsatellite loci abundance increases with genome size, but as species genome size increases, PCR amplification efficiency often decreases due to the dilution effect or multiple/non-specific priming sites. The proportion of available target DNA decreases in a template DNA volume, and the amount of non-specific primer binding increases [29]. This poses a major challenge in designing SSRs in *T. cinerariifolium* and other species with large genomes by the traditional laboratory method for screening genomic or enriched libraries. Next generation sequencing allows for rapid and cost-effective development of microsatellite markers in non-model species, although it should be noted that this method can still be costly and time-consuming in species with large genomes compared to other species [39]. The sequences of newly developed SSR markers were deposited into GenBank under accession numbers MW498263 to MW498279 (Table 2).

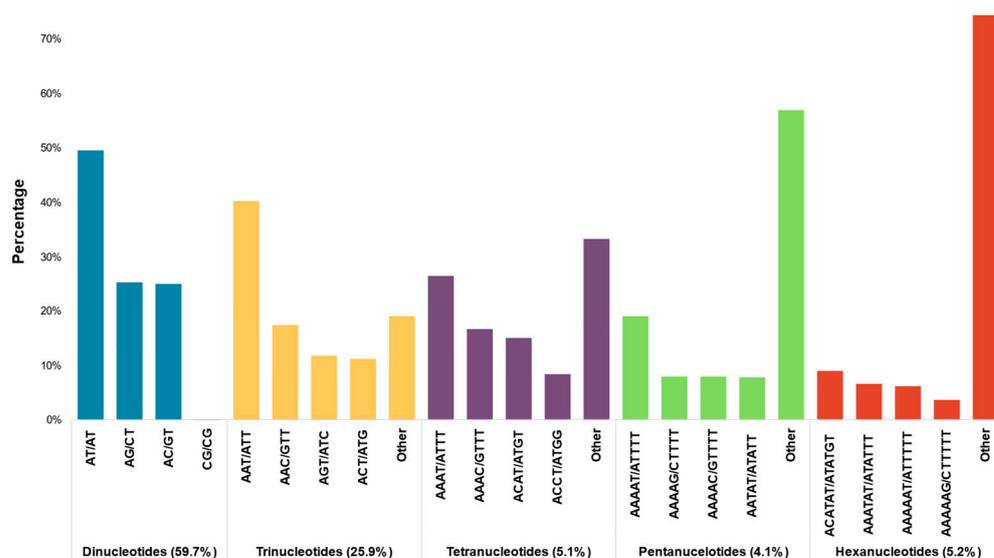


Figure 1. Frequency of SSR repeat motifs in *T. cinerariifolium* assembly.

The analysis performed in MicroChecker revealed no evidence of scoring errors due to stuttering or large allele dropout. The presence of null alleles was detected at four loci (TcUniZg007, TcUniZg009, TcUniZg020 and TcUniZg037). A total of 94 alleles were detected ranging from 3 (TcUniZg020, TcUniZg023 and TcUniZg032) to 13 (TcUniZg006) with an average of 5.53 alleles per locus. Values of expected heterozygosity ranged from 0.197 (TcUniZg020) to 0.864 (TcUniZg008) with an average value of 0.545. Observed heterozygosity ranged from 0.083 (TcUniZg020) to 0.708 (TcUniZg008) with an average value of 0.483 (Table 3). Similar values of observed heterozygosity were previously recorded in *H. italicum*, another outcrossing perennial species characterized by extensive gene flow [32]. Only TcUniZg037 showed significant deviation from Hardy-Winberg equilibrium ($p < 0.05$). The probability of identity varied between loci from 0.043 (TcUniZg008) to 0.668 (TcUniZg020) with combined non-exclusion probability of identity for all loci of 1.94×10^{-11} . The polymorphic information content varied from 0.178 (TcUniZg020) to 0.824 (TcUniZg008) with an average value of 0.492. Of the 17 markers tested, eight were classified as moderately polymorphic ($PIC > 0.44$), while two microsatellite markers were classified as highly polymorphic ($PIC > 0.70$) [40]. A low PIC was observed at the microsatellite loci TcUniZg004, TcUniZg013, TcUniZg020 and TcUniZg037 ($PIC < 0.29$), where only a small number of alleles were present ($N_a < 5$), suggesting that these loci are not suitable for use in genetic diversity analysis of the species (Table 3).

Table 2. List of 17 newly developed SSR markers for *T. cinerariifolium*.

Locus Name	GenBank Acc. No.	Primer Sequence (5' to 3')	Repeat Motif	Size (bp)
TcUniZg004	MW498263	F*: TGATCTTTTAAATTTGGAAGTAA R: GAAAGCTTCCTCTACCTCCTTG	(GT) ₁₀	248–256
TcUniZg005	MW498264	F: CCAGATCATTACTTAAAATGGAACA R: TACAACACTGGTGGCGTCAT	(AC) ₈	217–223
TcUniZg006	MW498265	F: CGACGGTTGGTGTGTGTATC R: CCATACGTGTCTCTTTCTCTTTGA	(GA) ₁₀	224–280
TcUniZg007	MW498266	F: GCTTCACATGGTTCGTCTCTG R: GCTTCACATGGTTCGTCTCTG	(CA) ₈	191–203
TcUniZg008	MW498267	F: TGCGATGATGATGATTGAGAG R: ATGGCAGAACATTCAACACAA	(GT) ₁₁	116–146
TcUniZg009	MW498268	F: TCTCCTTCTCCTCCTGCAA R: GGATGTTTGTGTGTTCAATTGG	(CA) ₁₁	106–118
TcUniZg010	MW498269	F: CATACTCCGCCCTTGATTA R: CCAAGACCCACTTTTTGGTG	(TG) ₈	180–194
TcUniZg012	MW498270	F: TCATCATCAACAAAATATCCAAGAA R: CCACCGACCACCTCATAATC	(CA) ₁₀	244–254
TcUniZg013	MW498271	F: ACATAACGTCGGAGGCATCA R: TGAGTTGGGTGCGTTACAAA	(TA) ₈	216–222
TcUniZg014	MW498272	F: AGCATAGACTGACTGTTCTTCA R: CCATATTCATCACAGCCTACGA	(TG) ₁₂	216–230
TcUniZg017	MW498273	F: AAGGCTGCGCTTCTTAACAG R: TAGCCATGCCTGGTACTTC	(TA) ₁₀	258–274
TcUniZg019	MW498274	F: AATGTGTGACTAATGGTCCTCAGA R: TGTTACTTAATTATAACATGCGGCCTA	(TA) ₈	116–124
TcUniZg020	MW498275	F: ACCACCAATACAAATACACCTTC R: GCAGAGGCTCGAGCTAGGAC	(CA) ₇	113–117
TcUniZg023	MW498276	F: CACAAATCCTTCACCTGTCAAA R: GCCAGTGGCAGAAGAGAAGT	(AC) ₉	240–250
TcUniZg032	MW498277	F: GAAATCAAGTGC GGATACGA R: TTTCCATATTGTGTTTTGGGTTC	(CAT) ₈	106–115
TcUniZg037	MW498278	F: GGACGGGATTACAGAAGGTG R: TCGACCTCATTATGCTGCTG	(CAA) ₇	249–258
TcUniZg038	MW498279	F: GGAGCCAAATACTAGCCTTCAA R: CGTTAGTCATCCGTGAGCAA	(TTG) ₆	151–163

* Each forward primer had M13 tail (TGTA AAAACGACGGCCAGT) at the 5' end.

2.3. Preliminary Genetic Diversity Study of Dalmatian Pyrethrum Populations

Based on the descriptive parameters, a subset of 12 developed microsatellite markers (TcUniZg005, TcUniZg006, TcUniZg008, TcUniZg010, TcUniZg012, TcUniZg013, TcUniZg014, TcUniZg017, TcUniZg019, TcUniZg023, TcUniZg032, and TcUniZg038) was used in the preliminary genetic diversity study of three Dalmatian pyrethrum natural populations (Table 4). Four microsatellite markers found to have null alleles (TcUniZg007, TcUniZg009, TcUniZg020, and TcUniZg037) and one marker with a very low PIC value (TcUniZg004) were not used in this study.

Table 3. Characterization of developed microsatellite markers of *T. cinerariifolium* based on molecular analysis of 20 samples from Hvar population.

Locus	N_a	H_O	H_E	F_{IS}	Sign	P_{null}	PIC	PI
TcUniZg004	4	0.167	0.235	0.29	ns	-	0.219	0.605
TcUniZg005	4	0.304	0.449	0.322	ns	-	0.384	0.37
TcUniZg006	13	0.542	0.633	0.145	ns	-	0.605	0.158
TcUniZg007	5	0.333	0.505	0.339	ns	0.091	0.432	0.318
TcUniZg008	11	0.708	0.864	0.18	ns	-	0.824	0.043
TcUniZg009	6	0.583	0.774	0.246	ns	0.111	0.716	0.098
TcUniZg010	6	0.667	0.757	0.12	ns	-	0.697	0.11
TcUniZg012	5	0.391	0.542	0.277	ns	-	0.485	0.266
TcUniZg013	4	0.375	0.323	-0.16	ns	-	0.288	0.495
TcUniZg014	6	0.542	0.531	-0.021	ns	-	0.488	0.263
TcUniZg017	7	0.542	0.704	0.23	ns	-	0.639	0.146
TcUniZg019	5	0.583	0.72	0.19	ns	-	0.654	0.137
TcUniZg020	3	0.083	0.197	0.576	ns	0.111	0.178	0.668
TcUniZg023	3	0.458	0.433	-0.059	ns	-	0.35	0.405
TcUniZg032	3	0.5	0.592	0.155	ns	-	0.506	0.25
TcUniZg037	4	0.125	0.299	0.582	*	0.146	0.266	0.528
TcUniZg038	5	0.542	0.702	0.228	ns	-	0.634	0.15

N_a —number of alleles; H_O —observed heterozygosity; H_E —expected heterozygosity; F_{IS} —inbreeding coefficient; Sign—significant deviations from Hardy-Weinberg equilibrium after sequential Bonferroni corrections: ns—non-significant; *— $p < 0.05$; P_{null} —null allele frequency; PIC—polymorphic information content; PI—probability of identity.

Table 4. Sampling sites of Dalmatian pyrethrum for preliminary genetic diversity study.

Accession Number ^a	Location	Elevation (m)	Latitude ^b	Longitude ^b
MAP02814	Mali Lošinj	39	44.574	14.420
MAP02807	Čiovo	211	43.498	16.303
MAP02809	Biokovo—Kotiški Stanovi	1350	43.314	17.062

^a Accession number from The Collection of Medicinal and Aromatic Plants, as available at the CPGRD (<https://cpgrd.hapih.hr> (accessed on 14 February 2022)); ^b Latitude and longitude are expressed in decimal degrees.

The highest D_{PSA} (0.875) was recorded between several pairs of individuals from Mali Lošinj and Mt. Biokovo populations, and lowest between two individuals from Čiovo population. The average D_{PSA} between 30 individuals was 0.557. Based on the genetic distance matrix of 30 Dalmatian pyrethrum samples, a Neighbor-joining tree was constructed (Figure 2), showing grouping of sampled individuals into their populations of origin. Only the separation of Mt Biokovo clade is supported with a bootstrap value larger than 50%. Similar results were obtained in previous study employing AFLP markers, where three populations from Mt Biokovo differentiated from the rest of the Croatian populations, possibly due to geographic and genetic isolation [15]. These preliminary results on the genetic diversity of Dalmatian pyrethrum are encouraging for a future full-scale study across the entire distribution range of the species with a larger sample size.

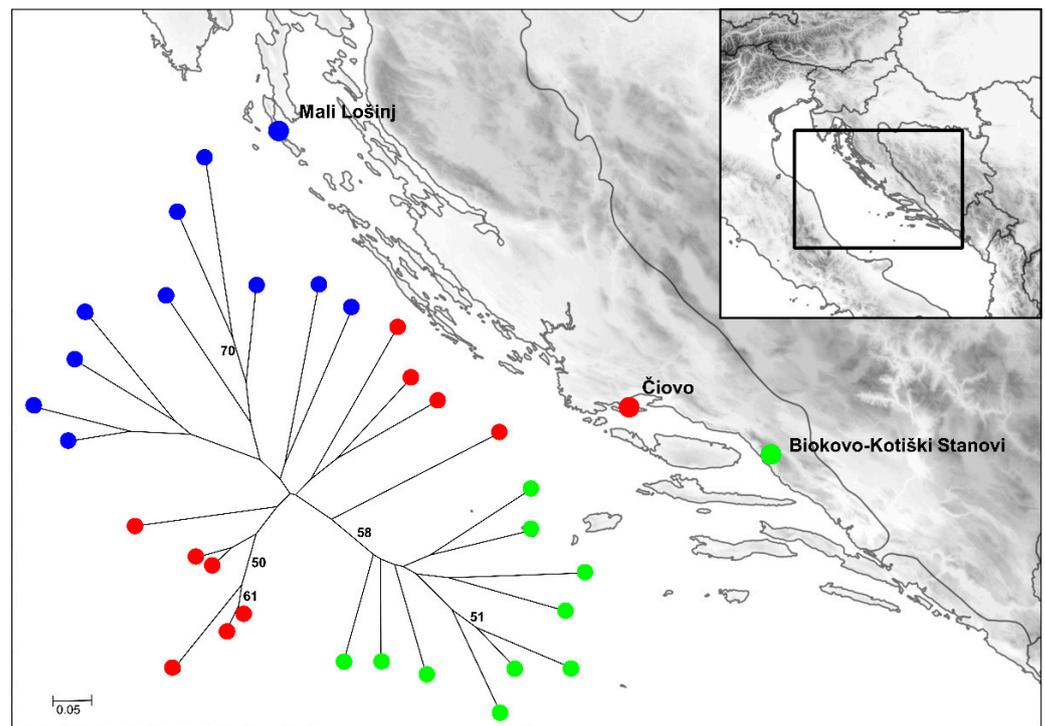


Figure 2. NJ tree based on D_{PSA} between 30 Dalmatian pyrethrum individuals from three distinct populations. Bootstrap values greater than 50% are indicated on the tree.

3. Materials and Methods

3.1. Plant Material

Leaves for the NGS were sampled from Dalmatian pyrethrum plant, Accession No. MAP02806; the Collection of Medicinal and Aromatic Plants, Zagreb, Croatia as available at the Croatian Plant Genetic Resources Database—<https://cpgrd.hapih.hr> (accessed on 14 February 2022) grown in the Botanical Garden of the Faculty of Science, University of Zagreb. For initial testing of newly developed SSR markers, five plant samples, each originating from distinct population (MAP02769, MAP02797, MAP02799, MAP02813 and MAP02809 from the Collection of Medicinal and Aromatic Plants) were collected in 2018 from a field experiment laid out in randomized complete block design at the experimental field station of the Institute for Adriatic Crops and Karst Reclamation in Kaštel Stari, Croatia. For additional testing and subsequent characterization of SSRs, 20 samples from one population (MAP02806) were collected from the same field experiment.

To evaluate the effectiveness of the developed SSR markers in the genetic diversity study of Dalmatian pyrethrum, 10 samples of *T. cinerariifolium* were collected at each of the three geographically distinct sites in Croatia (Table 4). The collection of samples from natural habitats in Croatia, including Biokovo Nature Park, was approved by the authority of the Ministry of Environmental Protection and Energy of the Republic of Croatia (UP/I-612-07/17-48/47, 517-07-1-1-1-17-6; 21 April 2017).

3.2. DNA Isolation

DNA for NGS was isolated from 100 mg of fresh plant tissue using the OmniPrep™ for Plant Kit (G-Biosciences, St. Louis, MO, USA). The manufacturer's instructions were followed except that 1% 2-mercaptoethanol and 1% polyvinylpyrrolidone (PVP) were added to the genomic lysis buffer. The concentration and purity of the isolated DNA were measured using the NanoPhotometer P300 spectrophotometer (Implen, Munich, Germany) and Qubit™ Fluorometer (Invitrogen, Carlsbad, CA, USA). The extracted DNA is stored at the Laboratory of genetic diversity, phylogeny and molecular systematics of plants of the Faculty of Science, University of Zagreb under the accession number ZAGR 47776.

Total genomic DNA of samples for both testing and characterization of the microsatellite loci, and genetic diversity study was isolated from 25 mg of silica gel-dried plant leaf tissue using the GenElute™ Plant Genomic DNA Miniprep Kit (Sigma-Aldrich®, Steinheim, Germany). Prior to isolation leaf tissue was ground to a fine powder using the TissueLyser II (Qiagen®, Hilden, Germany). The concentration and purity of the isolated DNA were measured using the NanoPhotometer P300 spectrophotometer (Implen, Munich, Germany).

3.3. Next Generation Sequencing, DNA Assembly and SSR Identification

The sequencing library was prepared by random fragmentation of the DNA sample, followed by 5' and 3' adapter ligation using the TruSeq DNA PCR Free kit (Illumina®, San Diego, CA, USA). Fragmentation was verified using the High Sensitivity DNA kit on the Agilent 2100 Bioanalyzer (Agilent Technologies®, Santa Clara, CA, USA). The library was sequenced using the Illumina HiSeq X Ten sequencer at Macrogen Europe®. The FastQC tool was used to check the quality of the raw sequences and the possible presence of the sequence adapters [41]. Raw sequencing reads were deposited in Sequence Read Archive (SRA) of NCBI under accession number SRX13877838.

Two approaches were used in the development of microsatellite markers for *T. cinerariifolium*. The first approach was based on *de novo* assembly of 20% of the total sequencing data randomly subsampled using the reformat.sh tool, part of the BBDMap/BBTools package [42]. The assembly was performed on this set of data 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.

In the second approach, evidence of contiguity from paired end reads were used. The complete dataset was searched for pairs with overlapping evidences using the bbmerge tool from the BBDMap/BBTools package, which were then deduplicated using the dedupe.sh tool from the BBDMap/BBTools package [42]. The obtained contigs from both approaches were re-mapped with the full set of sequencing data using the Map Reads to Contigs tool from CLC Genomics Server to obtain coverage information. Screening for possible repeat associated annotations of contigs was performed using RepeatMasker ver. 4.1.0 against combined Dfam 3.1 and RepBase (20170127) element databases [43].

Sequences obtained from both approaches were combined. To find positions of microsatellite repeats in the contigs, the MISA tool script [44] was used with search parameters for motif length of 2–6 nucleotides and with defined minimum microsatellite repeat lengths (2–8, 3–6, 4–5, 5–4, 6–4). The sequences were further analysed by BLASTn algorithm against the locally formatted database of draft genome of *T. cinerariifolium* [23] for a set of occurrences using the locally installed BLAST 2.10.1+ executable [45].

Sequences were filtered for further characterization of SSR markers and testing in natural populations of Dalmatian pyrethrum using the following selection criteria: (1) loci with compound microsatellite repeats were removed; (2) sequences with any repetitive element occurrences were removed; (3) reads mapped with multi-mapping occurrences were removed; (4) sequences with an average coverage of mapped reads close to the expected sequencing coverage depth and a length of more than 130 nucleotides were considered; (5) sequences with one or two occurrences in the draft genome of *T. cinerariifolium* were considered; (6) dinucleotide and trinucleotide motifs with longer repeats were preferred; and (7) sequences with guanine-cytosine (GC) content closer to the GC-content of the genome ($43 \pm 10\%$) were preferred. Sequences that met the listed requirements were used for primer design using Primer3 [46,47] with the parameter 'one primer for each seq' to obtain a single primer pair for each locus.

3.4. Testing and Characterization of Developed SSR Markers

The selected microsatellite primers were first tested on five samples from different populations (MAP02769, MAP02797, MAP02799, MAP02813 and MAP02809 from the Collection of Medicinal and Aromatic Plants). Polymorphic microsatellite markers with high amplification rate were further tested on 20 DNA samples from one population

(MAP02806) using a tailed primer protocol [48]. The 20 μ L of reaction mix contained 0.06 μ M of tailed forward SSR primer, 0.25 μ M of reverse SSR primer, 0.25 μ M of FAM. NED, VIC or PET 5' labeled M13 primer (5'-TGTAACACGACGGCCAGT-3), 1 \times PCR buffer with 1.5 mM MgCl₂, 0.2 μ M of each dNTP, 0.5 U TaqTM HS DNA polymerase (Takara[®] Bio Inc., Shiga, Japan), and 5 ng of template DNA. A PCR protocol with an initial touchdown cycle (94 °C for 5 min; 5 cycles of 45 s at 94 °C, 30 s at 60 °C, which was lowered by 1 °C in each cycle, and 90 s at 72 °C; 25 cycles of 45 s at 94 °C, 30 s at 55 °C, and 90 s at 72 °C; and 8-min extension step at 72 °C) was used [49]. Fluorescently labeled PCR products were detected on an ABI 3730XL (Applied Biosystems[®], Foster City, CA, USA) by the service Fragment Analysis (Macrogen Europe[®]). Allele sizes of PCR products were estimated using GeneMapper 4.0 software (Applied Biosystems[®], Foster City, CA USA).

For each microsatellite locus, the average number of alleles per locus (N_a), observed (H_E) and expected (H_O) heterozygosity, the probability of deviations from Hardy-Weinberg equilibrium, and the inbreeding coefficient (F_{IS}) were calculated using GENEPOP v. 4.4 [50]. Sequential Bonferroni corrections [51] were applied when multiple statistical tests were performed in SAS. For each locus, the frequency of null alleles (F_{null}) was calculated, and loci were analysed for scoring errors and allelic dropout using MicroChecker v. 2.2.3 [52]. Polymorphic information content (PI_C) and the probability of identity (PI) were calculated using Cervus v. 3.0.7 [53].

3.5. Preliminary Study of Dalmatian Pyrethrum Genetic Diversity

To evaluate effectiveness of developed SSR markers in differentiation of Dalmatian pyrethrum populations, the same tailed primer protocol as described in previous section was utilized. The distance based on the proportion of shared alleles (D_{PSA}) [54] between all pairs of individuals from three populations based on 12 microsatellite loci was calculated using MICROSAT [55]. Cluster analysis using the neighbor-joining (NJ) method with 1000 bootstraps was performed with the programs NEIGHBOR and CONSENSE (PHYLIP package) to construct a dendrogram [56].

4. Conclusions

With ever increasing urbanization in the Croatian coastal region Dalmatian pyrethrum populations are faced with the risk of habitat loss and subsequent decline in species richness. Newly developed SSR markers will be a powerful tool in the assessment of genetic diversity and structure of the species. In the future, these markers will serve as genetic background control in genome-wide association studies (GWAS) based on DArTseq-derived SNP markers in the hope of identifying quantitative trait nucleotides (QTNs) associated with variation in pyrethrin content and composition.

Author Contributions: Conceptualization, Z.Š., J.J. and Z.L.; methodology, J.J. and A.T.; validation, Z.L.; formal analysis, Z.Š. and F.V.; investigation, N.J. and I.R.; writing—original draft preparation, F.V., A.T. and J.J.; writing—review and editing, F.V., Z.Š., M.G., J.J. and Z.L.; visualization, F.V.; supervision, Z.Š. and M.G.; project administration, M.G. and Z.Š. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Croatian Science Foundation, under the project 'Genetic background of Dalmatian pyrethrum (*Tanacetum cinerariifolium*/Trevir./Sch. Bip.) insecticidal potential'-(PyrDiv) grant number IP-06-2016-9034 and by Slovenian Research Agency (ARRS) grant number P4-0077. The APC was funded by project KK.01.1.1.01.0005, Biodiversity and Molecular Plant Breeding, at the Centre of Excellence for Biodiversity and Molecular Plant Breeding (CoE CroP-BioDiv), Zagreb, Croatia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Eastwood, W.J. East Mediterranean Vegetation and Climate Change. In *Balkan Biodiversity—Pattern and Process in the European Hotspot*; Griffiths, H.I., Kryštufek, B., Reed, J.M., Eds.; Springer: Dordrecht, The Netherlands, 2004; pp. 25–48.
2. Grdiša, M.; Carović-Stanko, K.; Kolak, I.; Šatović, Z. Morphological and Biochemical Diversity of Dalmatian Pyrethrum (*Tanacetum cinerariifolium* (Trevir.) Sch. Bip.). *Agric. Conspec. Sci.* **2009**, *74*, 73–80.
3. Bhat, B.K. Breeding Methodologies Applicable to Pyrethrum. In *Pyrethrum Flowers: Production, Chemistry, Toxicology, and Uses*; Casida, J.E., Quistad, G.B., Eds.; Oxford University Press: New York, NY, USA, 1995; pp. 67–94.
4. Davies, T.G.E.; Field, L.M.; Usherwood, P.N.R.; Williamson, M.S. DDT, Pyrethrins, Pyrethroids and Insect Sodium Channels. *IUBMB Life* **2007**, *59*, 151–162. [[CrossRef](#)]
5. Bakarić, P. Buhač—Prirodni Insekticid. *Gospod. List* **2005**, *17*, 41–45. (In Croatian)
6. Hitmi, A.; Coudret, A.; Barthomeuf, C. The Production of Pyrethrins by Plant Cell and Tissue Cultures of *Chrysanthemum Cinerariaefolium* and *Tagetes* Species. *Crit. Rev. Biochem. Mol. Biol.* **2000**, *35*, 317–337. [[CrossRef](#)] [[PubMed](#)]
7. Wandahwa, P.; Van Ranst, E.; Van Damme, P. Pyrethrum (*Chrysanthemum cinerariaefolium* Vis.) Cultivation in West Kenya: Origin, Ecological Conditions and Management. *Ind. Crops Prod.* **1996**, *5*, 307–322. [[CrossRef](#)]
8. Ožanić, S. *Poljoprivreda Dalmacije u Prošlosti (in Croatian)*; Agronomic Society NRH: Split, Croatia, 1955.
9. FAO. *Food and Agriculture Organization of the United Nations. Production: Crops*; FAO: Rome, Italy, 2020.
10. BRA. Botanical Resources Australia. Available online: <https://www.botanicalresources.com/> (accessed on 14 March 2022).
11. Ambrožič Dolinšek, J.; Kovač, M.; Žel, J.; Camloh, M. Pyrethrum (*Tanacetum cinerariifolium*) from the Northern Adriatic as a Potential Source of Natural Insecticide. *Ann. Ser. Hist. Nat.* **2007**, *17*, 39–46.
12. Grdiša, M.; Babić, S.; Periša, M.; Carović-Stanko, K.; Kolak, I.; Liber, Z.; Jug-Dujaković, M.; Satovic, Z. Chemical Diversity of the Natural Populations of Dalmatian Pyrethrum (*Tanacetum cinerariifolium* /Trevir./Sch.Bip.) in Croatia. *Chem. Biodivers* **2013**, *10*, 460–472. [[CrossRef](#)]
13. Varga, F.; Jeran, N.; Šatović, Z.; Biošić, M.; Grdiša, M. High Diversity of Natural Dalmatian Pyrethrum Based on Pyrethrin Composition at Intra- and Interpopulation Level. *Phytochemistry* **2021**, *192*, 112934. [[CrossRef](#)]
14. Morris, S.E.; Davies, N.W.; Brown, P.H.; Groom, T. Effect of Drying Conditions on Pyrethrins Content. *Ind. Crops Prod.* **2006**, *23*, 9–14. [[CrossRef](#)]
15. Grdiša, M.; Liber, Z.; Radosavljević, I.; Carović-Stanko, K.; Kolak, I.; Satovic, Z. Genetic Diversity and Structure of Dalmatian Pyrethrum (*Tanacetum Cinerariifolium* Trevir./Sch./Bip., Asteraceae) within the Balkan Refugium. *PLoS ONE* **2014**, *9*, e105265. [[CrossRef](#)]
16. Mlinarec, J.; Skuhala, A.; Jurković, A.; Malenica, N.; McCann, J.; Weiss-Schneeweiss, H.; Bohanec, B.; Besendorfer, V. The Repetitive DNA Composition in the Natural Pesticide Producer *Tanacetum Cinerariifolium*: Interindividual Variation of Subtelomeric Tandem Repeats. *Front. Plant Sci.* **2019**, *10*, 613. [[CrossRef](#)] [[PubMed](#)]
17. Ottaro, W.G.W. The Relationship between the Ploidy Level and Certain Morphological Characteristics of *Chrysanthemum cinerariaefolium* Vis. *Pyrethrum Post* **1977**, *14*, 10–14.
18. Liu, P.L.; Wan, J.N.; Guo, Y.P.; Ge, S.; Rao, G.Y. Adaptive Evolution of the Chrysanthemyl Diphosphate Synthase Gene Involved in Irregular Monoterpene Metabolism. *BMC Evol. Biol.* **2012**, *12*, 214. [[CrossRef](#)] [[PubMed](#)]
19. Rivera, S.B.; Swedlund, B.D.; King, G.J.; Bell, R.N.; Hussey, C.E.; Shattuck-Eidens, D.M.; Wrobel, W.M.; Peiser, G.D.; Poulter, C.D. Chrysanthemyl Diphosphate Synthase: Isolation of the Gene and Characterization of the Recombinant Non-Head-to-Tail Monoterpene Synthase from *Chrysanthemum cinerariaefolium*. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 4373–4378. [[CrossRef](#)] [[PubMed](#)]
20. Sultana, S.; Hu, H.; Gao, L.; Mao, J.; Luo, J.; Jongsma, M.A.; Wang, C. Molecular Cloning and Characterization of the Trichome Specific Chrysanthemyl Diphosphate/Chrysanthemol Synthase Promoter from *Tanacetum cinerariifolium*. *Sci. Hortic.* **2015**, *185*, 193–199. [[CrossRef](#)]
21. Tang, L.; Li, J.; Khalil, R.; Yang, Y.; Fan, J.; Liu, M.; Li, Z. Cloning and Functional Analysis of CDS_CC12: A *Tanacetum Cinerariaefolium* Chrysanthemyl Diphosphate Synthase Gene. *Plant Growth Regul.* **2012**, *67*, 161–169. [[CrossRef](#)]
22. Khan, S.; Upadhyay, S.; Khan, F.; Tandon, S.; Shukla, R.K.; Ghosh, S.; Gupta, V.; Banerjee, S.; Ur Rahman, L. Comparative Transcriptome Analysis Reveals Candidate Genes for the Biosynthesis of Natural Insecticide in *Tanacetum cinerariifolium*. *BMC Genom.* **2017**, *18*, 54. [[CrossRef](#)]
23. Yamashiro, T.; Shiraiishi, A.; Satake, H.; Nakayama, K. Draft Genome of *Tanacetum cinerariifolium*, the Natural Source of Mosquito Coil. *Sci. Rep.* **2019**, *9*, 18249. [[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*, 312–328. [[CrossRef](#)]
25. Vaz Patto, M.C.; Satovic, Z.; Pêgo, S.; Fevereiro, P. Assessing the Genetic Diversity of Portuguese Maize Germplasm Using Microsatellite Markers. *Euphytica* **2004**, *137*, 63–72. [[CrossRef](#)]
26. Belaj, A.; Muñoz-Diez, C.; Baldoni, L.; Porceddu, A.; Barranco, D.; Satovic, Z. Genetic Diversity and Population Structure of Wild Olives from the North-Western Mediterranean Assessed by SSR Markers. *Ann. Bot.* **2007**, *100*, 449–458. [[CrossRef](#)]
27. Kostia, S.; Varvio, S.L.; Vakkari, P.; Pulkkinen, P. Microsatellite Sequences in a Conifer, *Pinus sylvestris*. *Genome* **1995**, *38*, 1244–1248. [[CrossRef](#)]

28. Pfeiffer, A.; Olivieri, A.M.; Morgante, M. Identification and Characterization of Microsatellites in Norway Spruce (*Picea Abies* K.). *Genome* **1997**, *40*, 411–419. [[CrossRef](#)] [[PubMed](#)]
29. Garner, T.W.J. Genome Size and Microsatellites: The Effect of Nuclear Size on Amplification Potential. *Genome* **2002**, *45*, 212–215. [[CrossRef](#)] [[PubMed](#)]
30. Purwoko, D.; Cartealy, I.C.; Tajuddin, T.; Dinarti, D.; Sudarsono, S. SSR Identification and Marker Development for Sago Palm Based on NGS Genome Data. *Breed. Sci.* **2019**, *69*, 1–10. [[CrossRef](#)]
31. 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](#)]
32. Baruca Arbeiter, A.; Hladnik, M.; Jakše, J.; Bandelj, D. First Set of Microsatellite Markers for Immortelle (*Helichrysum italicum* (Roth) G. Don): A Step towards the Selection of the Most Promising Genotypes for Cultivation. *Ind. Crops Prod.* **2021**, *162*, 113298. [[CrossRef](#)]
33. Singh, R.; Ming, R.; Yu, Q. Comparative Analysis of GC Content Variations in Plant Genomes. *Trop. Plant Biol.* **2016**, *9*, 136–149. [[CrossRef](#)]
34. Cserhati, M. Motif Content Comparison between Monocot and Dicot Species. *Genom. Data* **2015**, *3*, 128. [[CrossRef](#)]
35. Peng, Y.; Lai, Z.; Lane, T.; Nageswara-Rao, M.; Okada, M.; Jasieniuk, M.; O’geen, H.; Kim, R.W.; Douglas Sammons, R.; Rieseberg, L.H.; et al. De Novo Genome Assembly of the Economically Important Weed Horseweed Using Integrated Data from Multiple Sequencing Platforms. *Plant Physiol.* **2014**, *166*, 1241–1254. [[CrossRef](#)]
36. Scaglione, D.; Reyes-Chin-Wo, S.; Acquadro, A.; Froenicke, L.; Portis, E.; Beitel, C.; Tirone, M.; Mauro, R.; Lo Monaco, A.; Mauromicale, G.; et al. The Genome Sequence of the Outbreeding Globe Artichoke Constructed de Novo Incorporating a Phase-Aware Low-Pass Sequencing Strategy of F1 Progeny. *Sci. Rep.* **2016**, *6*, 19427. [[CrossRef](#)] [[PubMed](#)]
37. 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](#)] [[PubMed](#)]
38. Staton, M.; Best, T.; Khodwekar, S.; Owusu, S.; Xu, T.; Xu, Y.; Jennings, T.; Cronn, R.; Arumuganathan, A.K.; Coggeshall, M.; et al. Preliminary Genomic Characterization of Ten Hardwood Tree Species from Multiplexed Low Coverage Whole Genome Sequencing. *PLoS ONE* **2015**, *10*, e0145031. [[CrossRef](#)] [[PubMed](#)]
39. Schoebel, C.N.; Brodbeck, S.; Buehler, D.; Cornejo, C.; Gajurel, J.; Hartikainen, H.; Keller, D.; Leys, M.; Říčanová, Š.; Segelbacher, G.; et al. Lessons Learned from Microsatellite Development for Nonmodel Organisms Using 454 Pyrosequencing. *J. Evol. Biol.* **2013**, *26*, 600–611. [[CrossRef](#)] [[PubMed](#)]
40. Hildebrand, C.E.; Torney, D.C.; Wagner, R.P. Informativeness of Polymorphic DNA Markers. *Los Alamos Sci.* **1992**, *20*, 100–102.
41. Andrews, S. *FastQC—A Quality Control Tool for High Throughput Sequence Data*; Babraham Institute: Cambridge, UK, 2010.
42. Bushnell, B.; Rood, J.; Singer, E. BBMerge—Accurate Paired Shotgun Read Merging via Overlap. *PLoS ONE* **2017**, *12*, e0185056. [[CrossRef](#)]
43. Smit, A.; Hubley, R.; Grenn, P. *RepeatMasker Open-4.0*; Institute for System Biology: Seattle, WA, USA, 2013.
44. 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](#)]
45. Zhang, Z.; Schwartz, S.; Wagner, L.; Miller, W. A Greedy Algorithm for Aligning DNA Sequences. *J. Comput. Biol.* **2000**, *7*, 203–214. [[CrossRef](#)]
46. Koressaar, T.; Remm, M. Enhancements and Modifications of Primer Design Program Primer3. *Bioinformatics* **2007**, *23*, 1289–1291. [[CrossRef](#)]
47. 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](#)]
48. Schuelke, M. An Economic Method for the Fluorescent Labeling of PCR Fragments. *Nat. Biotechnol.* **2000**, *18*, 233–234. [[CrossRef](#)] [[PubMed](#)]
49. Radosavljević, I.; Jakse, J.; Javornik, B.; Satovic, Z.; Liber, Z. New Microsatellite Markers for *Salvia Officinalis* (Lamiaceae) and Cross-Amplification in Closely Related Species. *Am. J. Bot.* **2011**, *98*, e316–e318. [[CrossRef](#)] [[PubMed](#)]
50. Rousset, F. Genepop’007: A Complete Re-Implementation of the Genepop Software for Windows and Linux. *Mol. Ecol. Resour.* **2008**, *8*, 103–106. [[CrossRef](#)]
51. Holm, S. A Simple Sequentially Rejective Multiple Test Procedure. *Scand. J. Stat.* **1979**, *6*, 65–70.
52. van Oosterhout, C.; Hutchinson, W.F.; Wills, D.P.M.; Shipely, P. Micro-Checker: Software for Identifying and Correcting Genotyping Errors in Microsatellite Data. *Mol. Ecol. Notes* **2004**, *4*, 535–538. [[CrossRef](#)]
53. 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](#)]
54. 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](#)]
55. Minch, E.; Ruiz-Linares, A.; Goldstein, D.; Feldman, M.; Cavalli-Sforza, L.L. *MICROSAT: A Computer Program for Calculating Various Statistics on Microsatellite Allele Data*; Stanford University: Stanford, CA, USA, 1997.
56. Felsenstein, J. *PHYLIP (Phylogenetic Inference Package, Version 3.6b)*; Department of Genetics, University of Washington: Seattle, WA, USA, 1993.