



Article Complete Genome of Rose Myrtle, *Rhodomyrtus tomentosa*, and Its Population Genetics in Thai Peninsula

Matsapume Detcharoen ^{1,*}, Sara Bumrungsri¹ and Supayang Piyawan Voravuthikunchai^{2,3}

- ¹ Division of Biological Science, Faculty of Science, Prince of Songkla University, Hat Yai 90110, Thailand
- ² Center of Antimicrobial Biomaterial Innovation-Southeast Asia, Faculty of Science, Prince of Songkla University, Hat Yai 90110, Thailand
- ³ Natural Product Research Center of Excellence, Faculty of Science, Prince of Songkla University, Hat Yai 90110, Thailand
- * Correspondence: matsapume.d@psu.ac.th

Abstract: Several parts of rose myrtle, *Rhodomyrtus tomentosa*, exhibited profound antibacterial and anti-inflammatory activities, suggesting its potential in healthcare and cosmetics applications. During the past few years, the demand for biologically active compounds in the industrial sectors increased. Therefore, gathering comprehensive information on all aspects of this plant species is essential. Here, the genome sequencing using short and long reads was used to understand the genome biology of *R. tomentosa*. Inter-simple sequence repeats (ISSR) and simple sequence repeats (SSR) markers, and geometric morphometrics of the leaves of *R. tomentosa* collected across Thai Peninsula, were determined for population differentiation analysis. The genome size of *R. tomentosa* was 442 Mb, and the divergence time between *R. tomentosa* and *Rhodamnia argentea*, the white myrtle of eastern Australia, was around 15 million years. No population structure was observed between *R. tomentosa* on the eastern and western sides of the Thai Peninsula using the ISSR and SSR markers. However, significant differences in leaf size and shape of *R. tomentosa* were observed in all locations.

Keywords: Ceylon hill gooseberry; ISSR; microsatellite; Myrtaceae; Myrtales; *Rhodamnia argentea*; rose myrtle

1. Introduction

Rose myrtle, *Rhodomyrtus tomentosa* (Aiton) Hassk., is an evergreen flowering plant belonging to the order Myrtales of the family Myrtaceae. This plant species is native to Southeast Asia and is commonly found far and wide from South China to Indonesia [1]. In Thailand, *R. tomentosa* can be found on the east and west sides of the Malay Peninsula (Thai Peninsula) [2]. It thrives in poor acidic soils, especially near the coastlines and sunny environment [2–4]. Bees are the primary pollinators of *R. tomentosa* as this plant species cannot be pollinated by wind [5]. Several parts of this species are rich in sugar, vitamins, and minerals. Most parts of the plants have been used in traditional medicine. Leaves have been used to treat colic diarrhea [6], abscesses, hemorrhage [7], and unripe fruits for treating sepsis, colic, and diarrhea [6,8].

Numerous compounds of *R. tomentosa* have been investigated to search for alternative active compounds, given their detrimental effects against several symptoms. Rhodomyrtone, an acylphloroglucinol found in the leaves of *R. tomentosa*, exhibited diverse effects, such as antibacterial, anti-inflammatory, antioxidant, and antitumor activities [9–12]. It is used in cosmetics and therapeutics, such as acne creams [13]. Similarly, piceatannol is a primary phenolic compound found in this plant, possessing antioxidant, anti-cancer, and anti-inflammatory effects [14].

Myrtaceae family encompasses approximately 150 genera and 5000 species, with a genome size varying from 187 Mb in *Syzygium oleosum* to 962 Mb in *Eucalyptus rudis*. Despite the presence of several species in this family, genomic studies on Myrtaceae have



Citation: Detcharoen, M.; Bumrungsri, S.; Voravuthikunchai, S.P. Complete Genome of Rose Myrtle, *Rhodomyrtus tomentosa*, and Its Population Genetics in Thai Peninsula. *Plants* **2023**, *12*, 1582. https://doi.org/10.3390/ plants12081582

Academic Editor: Hisato Kunitake

Received: 13 February 2023 Revised: 4 April 2023 Accepted: 5 April 2023 Published: 7 April 2023



Copyright: © 2023 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/). mainly focused on *Eucalyptus* species [15]. In *R. tomentosa*, the chloroplast genome was approximately 156 kb in size [16], while the genome of *R. tomentosa* cultivar LFSTJN1 in China was around 470 Mb [17]. The cytological study reported that diploid cells contained 22 chromosomes [18]. Furthermore, inter-simple sequence repeats (ISSR) and simple sequence repeats (SSR) have been employed to examine the population genetics of *R. tomentosa* in Malaysia and China, respectively [19,20]. However, more information is needed on this species' population genetics and genetic diversity in other geographical regions [19].

Herein, we aimed to understand the genome biology, evolutionary history, and genetic diversity of *R. tomentosa*. Despite abundant knowledge of *R. tomentosa* organic compounds, evolution, and genome biology, its genetic diversity must be better known than other Myrtaceae members. The genome of *R. tomentosa* was sequenced and assembled through short and long reads, and the genetic diversity within the Thai Peninsula was detected using ISSR and SSR markers. We found that the genome of *R. tomentosa* was similar to those of other Myrtaceae members. However, no significant genetic differentiation was observed across *R. tomentosa* in the Thai Peninsula.

2. Results

2.1. Genome Assembly

The estimated genome size of *R. tomentosa* was between 442 and 485 Mb. The assembled genome of *R. tomentosa* was sequenced using short- $(45.1 \times \text{coverage})$ and long-read sequencing ($12.1 \times \text{coverage}$). After removing contaminant and chloroplast reads (Figure S1), the final length of the assembly was 442 Mb, with a GC content of 40.5%, 1431 contigs, and N50 of 868,979 bases (Table S1, BioProject accession PRJNA846150). The completeness of the assembly determined by BUSCO against the eudicots odb10 database of 2326 genes was 94.8% (2205 genes), of which 92.7% (2156 genes) were complete and single copies, 2.1% (49 genes) were complete and in duplicates, 1% (23 genes) was fragmented, and 4.2% (98 genes) were missing. Repeat analysis revealed that Gypsy/DIRS1 long terminal repeat retrotransposons were the most dominant (Table S2). *Rhodomyrtus tomentosa* contains more retroelements yet fewer DNA transposons and small RNAs than *R. argentea*.

Gene prediction identified 31,779 transcripts of 29,768 predicted genes. Five mean exons and four mean introns were observed per transcript, respectively. The mean gene length was 3157 bases. The longest coding sequence identified was 16,374 bases, with the closest match predicted to the midasin gene of *R. argentea* (XM_030670374.1) (Table S3). BUSCO completeness analysis of protein-coding sequences against the eudicots odb10 database reported completeness of 92.2%, of which 80.7% were complete and single copies, 11.5% were complete and in duplicate copies, 2.1% were partially complete, and 5.7% were missing. About 28,375 genes were annotated for gene annotation with most having an unknown function, cytochrome P450, protein kinases, and UDP-glycosyltransferase family (Table S4). Furthermore, KEGG pathway analysis indicated that most annotated genes were associated with secondary metabolites' metabolic pathways and biosynthesis (Table S5). The functional classification of the genes showed that most *R. tomentosa* genes belong to transcription, signal transduction mechanisms, post-translational modification, protein turnover, and chaperone functions (Figure 1).

For orthologous gene analysis, approximately 96% (704,329 genes) of genes present in all thirteen Myrtaceae species were assigned to 42,372 orthogroups. Of these orthogroups, 9858 existed in all species. In the case of *R. tomentosa*, 84% of its genes were assigned to 15,587 orthogroups. Among these, 833 orthogroups, representing 2308 genes, were specific to *R. tomentosa* (Table S6).

Based on 236 nuclear orthologous gene groups (Figure S2) or chloroplast genomes (Figure S3), the maximum likelihood phylogenetic trees had similar topologies. Myrtaceae species, along with *R. tomentosa*, formed a monophyletic clade. The clock-calibrated phylogenetic tree reported that the split between the closest taxa, *R. tomentosa* and *R. argentea*, occurred around 15.5 (7.3–25.2, 95% confidence interval) million years ago (Figure 2).



Figure 1. Gene function of *Rhodomyrtus tomentosa* classified based on clusters of orthologous groups database.



Figure 2. Gene family evolution conducted on clock-calibrated Bayesian phylogenetic tree based on orthologous genes. Average expansion rate, number of genes expanded, and number of genes contracted are shown in orange, green, and red, respectively. Mean and 95% confidence interval of the estimated divergence time (million years ago) are shown above each node and blue horizontal bar, respectively.

The significance of gene family evolution across Myrtaceae species was based on the clock-calibrated phylogenetic tree. Overall, a negative rate of gene expansion/contraction (-0.05) was observed in *R. tomentosa* (Figure 2). A total of 352 gene families with significant expansion or contraction rates were detected in *R. tomentosa*. Numerous genes within these

families belong to protein kinases, leucine-rich repeats, and disease-resistance proteins (Table S7).

Comparing genome synteny between *R. tomentosa* and *R. argentea*, we found a strong synteny across the genomes of the two species (Figure 3). All pseudomolecules of *R. tomentosa* aligned well with the chromosomes of *R. argentea*. However, some synteny segments were on different chromosomes of *R. argentea*, particularly pseudomolecules 10 in *R. tomentosa*.



Figure 3. Genome synteny between *Rhodomyrtus tomentosa* pseudomolecules (Rt; colored) and *Rhodamnia argentea* chromosomes (Ra; gray). Contigs of *R. tomentosa* that are not able to scaffold into chromosomes are shown on top of the figure.

The terpene synthase genes of *R. tomentosa* and other Myrtaceae were examined. There were around sixty candidate terpene synthase genes in *R. tomentosa* (Table S4), and several contained all three domains (Table S8). The conserved motifs in these genes were organized similarly in *R. tomentosa*, *R. argentea*, and *E. grandis* (Figure S4). The phylogenetic analysis demonstrated that many *R. tomentosa* terpene synthase genes were similar to those of *Eucalyptus grandis*, while those of *Prunus dulcis* and *Punica granatum* formed different clades (Figure S5).

2.2. Population Genetics and Leaf Morphometrics of R. tomentosa on Thai Peninsula

Two markers, ISSR and SSR, were employed for population genetic analysis of *R. tomentosa* collected from twelve locations on the east and west sides of the Thai Peninsula (n = 80). For the ISSR marker, the mean expected heterozygosity per locus was 0.86. The ISSR primer, UBC825, had the highest expected heterozygosity and primer 56 had the most polymorphism information content (Table S9). Across the ten loci, we observed the highest expected heterozygosity in Phatthalung and Nakon-si-thammarat, and the lowest in Phuket (Table S10). The summary statistics of the SSR marker were similar to those of the ISSR marker. The expected heterozygosity per location ranged between 0.63 (Pattani) and 0.85 (Chumphon) (Table S10).

We used principal component analysis (PCA) using the Lynch distance to observe similarity among the samples by ISSR or SSR marker. PCA revealed that the samples from the east and west sides of the Peninsula did not form distinct clusters (Figure 4). However, compared to using the SSR marker, the PCA of the ISSR markers showed more grouping of samples within the location, for example, samples from Krabi, Satun, and Surat-thani (Figure 4A). In addition, admixture analysis of the ISSR and SSR markers revealed two distinct genetic clusters (K = 2), with no differentiation between the east and west sides of the Peninsula (Figure S6). Notably, Chumphon, Songkhla, and Krabi exhibited high similarity.



Figure 4. Principal component analysis based on ten ISSR loci (**A**) and thirteen SSR loci (**B**). Samples from the western and eastern side of Thai Peninsula are shown in triangles and circles, respectively.

The UPGMA dendrogram based on the genetic distance of the ISSR marker showed more similarity across the samples within locations than using the SSR marker (Figure 5). The ISSR dendrogram was categorized into three clades. The first consisted of Phang-nga and Narathiwas; Surat-thani samples were placed in the second clade, and samples from Krabi, Songkhla, and Chumphon comprised the third clade.

The SSR marker provided more information about G_{ST} than the ISSR markers for pairwise population differentiation. Population differentiation based on ISSR and SSR markers revealed the highest G_{ST} values between Phuket and Chumphon and Phuket and Pattani, respectively (Table S11).

An evaluation of population differentiation between the east and west sides of the Peninsula was conducted using an analysis of molecular variance. For the ISSR marker, there was no significant difference between the east and west sides (AMOVA: p = 0.49). The variation among samples within locations between the sides was 32.39%, and the differentiation between the two sides was 0.32. For the SSR marker, there was no difference between the two sides (AMOVA: p = 0.07), and the variation and differentiation among samples within locations between the two sides were 16.05% and 0.16, respectively.

Geometric morphometrics with four landmarks was used to determine whether there was a differentiation in leaf morphology among the locations. Leaf size and shape significantly varied across the locations (Procrustes ANOVA: size $F_{11,300} = 5.10$, p < 0.01 and shape $F_{44,1200} = 4.49$, p < 0.01). The longest Mahalanobis distance was observed between Phuket and Pattani, and the shortest was between Chumphon and Surat-thani (Table S12). We observed a high variation in the size and shape of leaves but no specific patterns based on the location or the east and west sides (Figure 6). In addition, Mantel tests were conducted to check the correlation between Nei's genetic distance and Mahalanobis distance of morphometrics data. The results indicated no significant correlation between genetic distances for ISSR (p = 0.42), SSR markers (p = 0.51), and morphometric distances.



Figure 5. UPGMA dendrograms based on Nei's genetic distance of *Rhodomyrtus tomentosa* samples collected from twelve locations in Thai Peninsula, using ISSR (**A**) and SSR markers (**B**). Bootstrap values higher than sixty are shown. Samples from the western and eastern side of the Thai Peninsula are shown in triangles and circles, respectively.



Figure 6. Canonical variate analysis of leaf geometric morphometrics across the twelve locations based on four landmarks. Populations in the western and eastern side of Thai Peninsula are shown in triangles and circles respectively.

3. Discussion

3.1. R. tomentosa Genome

Herein, we reported the genome assembly of *R. tomentosa*. In our study, the genome size of *R. tomentosa* (442 Mb) was similar to the cultivar LFSTJN-1 in China (470 Mb) [17]. It was found to be approximately half the size of its sister species, *Rhodomyrtus psidioides* (BioSample SAMN19602563), with a genome size of 931 Mb. The differences in genome size can be attributed to the polyphyletic nature of *Rhodomyrtus*. *Rhodomyrtus tomentosa* and *R. psidioides* were placed in different clades based on the nuclear internal transcribed spacer [21]. The genome size variation is common within Myrtaceae, for example, in *Eucalyptus* [22,23]. Nevertheless, the genome size of *R. tomentosa* is in the same range as those of other Myrtaceae members (i.e., between 187 (*Syzygium oleosum*) and 962 Mb (*Eucalyptus rudis*)).

For gene family evolution, many genes related to disease resistance (e.g., code for tobacco mosaic virus resistance protein N and disease-resistance nucleotide-binding site–leucine-rich repeat family) were significantly expanded or contracted, indicating the reduced susceptibility of the plant species against pests and diseases [24]. The proteins of these genes provide protection against plant pathogens, including fungi [25,26], insects [27], bacteria [28], and viruses [29], in numerous plants.

We compared the genome synteny of *R. tomentosa* with *R. argentea* as they belong to the Myrteae tribe and have the same chromosome numbers [18]. We found that *R. tomentosa* demonstrated high collinearity with *R. argentea*, supported by their recent divergence time (7.3–25.2 million years). *Rhodomyrtus* and *Rhodamnia* are closely related, share similar evolutionary histories, are distributed in the same area (Southeast Asia, Australia, and Pacific islands), and produce similar organic compounds [30,31]. *Rhodomyrtus tomentosa* has twice as many Gypsy elements as *R. argentea*. In addition, *Rhodomyrtus tomentosa* has more repetitive elements than most Myrtaceae species, including *R. argentea*, *Psidium guajava*, *Corymbia citriodora*, and *Metrosideros polymorpha* [32–34]. The high accumulation of Gypsy in the *R. tomentosa* genome suggested a specific activity of the long terminal repeats in this species.

The estimated divergence period of Myrtales in the late Cretaceous was in line with estimations of previous studies based on nuclear and plastid sequences, consolidating orthologous genes from publicly available genome data [35,36]. The divergence time of *R. tomentosa* and other Myrtaceae occurred in the late Miocene, between 7.3 and 25.2 million years ago. This result was in line with a previous study that estimated the divergence time of *Rhodomyrtus* around 14.64–24.64 million years ago using pollen fossil, nuclear, and chloroplast sequences [31].

Myrtaceae are well-known for their terpene diversity [37]. The number of terpene synthase genes in Myrtaceae varies between species, from 37 in *Melaleuca alternifolia* [38] to 113 in *E. grandis* [39]. In *R. tomentosa*, around 60 putative terpene synthase genes were identified and the structures of the conserved motifs of these genes were similar to those in other species analyzed. Heterogeneity in secondary metabolite gene numbers reflects evolutionary divergences among species driven by biotic and abiotic factors, including herbivory and temperature [37,40]. Phylogenetic analysis confirmed that the terpene synthase genes of *R. tomentosa* were phylogenetically related to other Myrtaceae than Lythraceae (*P. granatum*) and Rosaceae (*P. dulcis*). Several *R. tomentosa* terpene synthase genes were clustered with germacrene D synthase genes, whose product, germacrene D, is one of the common volatile compounds found in plants, including Myrtaceae [37] that can be used against insects [41].

Compared to the findings reported by a previous study of the *R. tomentosa* cultivar in China [17], our study reported similar results, including gene count, mean exon per gene, percentage of repeats in the genome, and predicted divergence time. These findings contributed to a more comprehensive understanding of the genomics of *R. tomentosa* and its evolutionary relationships within the Myrtaceae family.

3.2. Population Genetics and Geometric Morphometrics of R. tomentosa in Thai Peninsula

Comparative analysis of ISSR and SSR markers has been used to evaluate genetic diversity in numerous plant species, including the Myrtaceae family [42,43]. For this study, we used the ISSR and the SSR markers to study the population genetics of *R. tomentosa*. Both markers exhibited similar basic statistics per locus, including genetic diversity and heterozygosity. However, the dendrogram based on genetic distance and PCA revealed that the ISSR delivered better resolution for sample classification than the SSR markers. In addition, population analysis of *R. tomentosa* using the markers revealed high expected heterozygosity in the Thai Peninsula, suggesting high genetic diversity in the area, which was considerably greater than that observed in *R. tomentosa* populations in Malaysia [19] and China [20].

The PCA, genetic distance dendrogram, and admixture analysis indicated a lack of differentiation between the two sides of the Peninsula, despite the presence of mountain ranges as geographic barriers. This suggested that gene flow occurred between these two sides. Another population genetics study of *R. tomentosa* within Peninsular Malaysia did not find genetic differentiation between the east and west sides [19]. Unlike mangroves and other marine plants [44–46], mountain ranges did not act as geographic barriers for *R. tomentosa*. Nevertheless, PCA showed similarities among samples from the exact location, suggesting some degree of inbreeding. For instance, both markers showed similarities between samples from Krabi, possibly due to close proximity of the *R. tomentosa* collected in this area compared with those from other locations.

Based on population differentiation analysis, low G_{ST} values were observed between nearby locations; for example, collection sites at Songkhla and Phatthalung were around thirty kilometers apart with no geographic barriers. The highest G_{ST} values were observed between Phuket–Chumphon and Phuket–Pattani. Phuket was the only non-mainland location in this study, making it challenging for gene flow between Phuket and other locations. Additionally, *R. tomentosa* has a geitonogamy mating system and can only be outcrossed by insect pollinators (e.g., bees) but cannot be pollinated by wind [5]. The effects of geographic barriers on island populations have been documented in several plants, such as *Trifolium repens* [47] and *Lactuca watsoniana* [48].

Most morphometrics studies in Myrtaceae have focused on variations in complex species, such as *Myrcia* [49], *Eugenia* [50], and *Eucalyptus* [51]. Less is known about population variations. Herein, we observed high variation in leaves within locations. The leaves collected from Phuket and Pattani had the highest Mahalanobis distance, similar to the results from population differentiation analysis using the SSR marker. The study of the morphological variation of another Myrtaceae, *Myrceugenia fernandeziana*, could not find any differentiation across populations [52]. However, adding more morphological characteristics, such as flowers and peduncles, might increase the resolution of classification [51].

In conclusion, the assembled genome of *R. tomentosa* can aid in implementing the genetics and evolutionary studies of other Myrtaceae members. Herein, we assembled the 442 Mbp genome of *R. tomentosa* using short- and long-read sequencing. High synteny was observed between the genomes of *R. tomentosa* and *R. argentea*. The population genetic analysis of *R. tomentosa* collected from twelve distinct locations across the Thai Peninsula using ISSR and SSR markers demonstrated congruent results. There was neither strong genetic structure nor genetic differentiation between the east and west sides of the Peninsula, suggesting that mountain ranges did not act as a strong geographic barrier against this plant species.

4. Materials and Methods

4.1. Genome Assembly

The leaves of *R. tomentosa* were collected from Hat Yai, Thailand (7°00'24.6″ N, 100°30'14.8″ E), in October 2021. The DNA was extracted from a single young leaf using the cetyltrimethyl ammonium bromide (CTAB) method [53] and purified with Mag-Bind TotalPure NGS (Omega Bio-tek, Norcross, GA, USA) by the manufacturer's instructions.

The final product was sent to BGI Genomics (Hong Kong, China) for 2×150 bp short-read sequencing. The raw reads were checked for quality using Fastp version 0.23.0 [54] with a minimum length of 30 and a sliding window of four with a mean quality of 30. To estimate the genome size of *R. tomentosa*, we used Jellyfish version 2.3.0 [55] based on the filtered short reads with k-mer lengths of 19, 21, and 31, and the results were plotted using GenomeScope version 2.0 [56].

For long-read sequencing, DNA library was prepared using the NEBNext Companion Module for Oxford Nanopore Technologies ligation sequencing (New England BioLabs, Ipswich, MA, USA) and ligation sequencing kit (LSK-SQK109; Oxford Nanopore Technologies, Oxford, UK). The library was sequenced on an R9.4.1 flow cell on a MinION device.

Guppy version 6.0.1 (Oxford Nanopore Technologies, Oxford, UK) was used for basecalling with a high accuracy option and a minimum score of 9. The reads were assembled using Flye version 2.9 [57] with a minimum overlap set to 3000. The assembled contigs were polished with four rounds of Racon version 1.5.0 [58] and one round using Medaka version 1.5.0 (Oxford Nanopore Technologies). Following this, the assembled contigs were polished with short reads using four rounds of Pilon version 1.24 [59]. Purge_Dups version 1.2.5 [60] was used to check the haplotypic duplication of the assembly. BlobToolKit version 3.1.0 [61] was used to check for contaminant reads, which were manually removed. Quast version 5.0.2 [62] and Benchmarking Universal Single-Copy Orthologs (BUSCO) version 5.3.0 [63] with the eudicots odb10 database were used to evaluate the genome assembly.

For repeated gene prediction, the consensus sequence was soft-masked using Repeat-Modeler version 2.0.2 [64] and RepeatMasker version 4.1.2 [65] with the *E. grandis* repeat library from the Plant Repeat Database [66]. The masked consensus sequence was aligned with the publicly available *R. tomentosa* transcriptome data (Biosample SAMN06640542) using rnaSPAdes version 3.15.4 [67]. The masked sequence, aligned transcriptome, and protein sequences of other Myrtaceae downloaded from UniProt [68] were used as inputs in AUGUSTUS version 3.4.0 [69] and GeneMark_ES version 4.69 [70] incorporated in Braker2 version 2.1.6 pipeline [71]. BUSCO version 5.3.0 [63] with the eudicots odb10 database was used to evaluate the genome assembly and repeat gene prediction. Gene ontology and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways of the predicted genes were performed using eggNOG-mapper version 2.1.7 [72].

OrthoFinder version 2.5.4 [73] was used to identify the orthologous genes of *R. tomentosa* by running all-versus-all diamond BLAST against thirteen other annotated eurosid genomes, namely, *Arabidopsis thaliana* (GCF000001735.4), *Brassica napus* (GCF000686985.2), *Citrus clementina* (GCF000493195.1), *Carica papaya* (GCF000150535.2), *Durio zibethinus* (GCF002303985.1), *E. grandis* (GCF016545825.1), *Gossypium hirsutum* (GCF007990345.1), *P. dulcis* (GCF902201215.1), *P. granatum* (GCF007655135.1), *Pistacia vera* (GCF008641045.1), *Quercus suber* (GCF002906115.1), *R. argentea* (GCF900635035.1), and *Syzygium oleosum* (GCF900635055.1). A maximum likelihood phylogenetic tree based on orthologous genes (Supplementary data S1) was constructed using FastTree version 2.1.11 [74] with local support values based on the Shimodaira-Hasegawa test, and others were left as default.

To estimate the divergence time of *R. tomentosa*, we used the MCMCTree program in PAML version 4.9j [75] with a phylogenetic tree based on orthologous genes and the concatenated translated nucleotide alignments of orthologues as inputs. The estimated divergence time of 37.4–49.5 million years between *Arabidopsis* and *Brassica* [76], 70–74 million years between *Arabidopsis* and *Carica* [77], and 66.4–112.6 million years between *Eucalyptus* and *Punica* [78] were used as calibration points. We performed a Markov chain Monte Carlo (MCMC) run for one million generations and a burn-in of 100,000 generations. with the JC69 model. Two independent MCMC runs were performed.

CAFE version 4.2.1 was used to analyze gene family evolution of *R. tomentosa* [79]. The orthologous gene counts from OrthoFinder and the clock-calibrated tree was used as inputs. A gene family was counted as a significant expansion or contraction when its p-value was less than 0.01.

Synteny analysis between *R. tomentosa* and *R. argentea* was compared. *Rhodamnia argentea* was selected as it is the closest species to *R. tomentosa* with a publicly available genome, and both species have eleven chromosomes [18]. The *R. tomentosa* assembled contigs were scaffolded into eleven pseudomolecules using the RagTag version 2.1.0 [80]. These pseudomolecules were mapped to the *R. argentea* reference genome (GCF_020921035.1) in a one-to-one relationship using the nucmer function implemented in MUMmer version 3.1 [81]. The alignment was filtered to retain only segments with more than 80 alignment identities with a minimum of 3000 bases. The lengths of the generated pseudomolecules were similar to those of Li et al. [17]. The synteny plot was generated using Circos version 0.69-8 [82].

The phylogenetic relationships and conserved motifs of terpene synthase genes of *R. tomentosa* were investigated. The terpene synthase amino acid sequences from *R. argentea*, *E. grandis*, *P. granatum*, and *P. dulcis* were downloaded from GenBank. The sequences were aligned using ClustalW version 2 [83]. A phylogenetic tree was constructed using IQ-TREE version 2.2.0 [84] with the GTR20+F+R6 model, 1000 approximate likelihood ratio tests, and 10,000 ultrafast bootstraps. The domains of terpene synthase, RRx8W, RDR, and DDXXD, of *R. tomentosa* were analyzed. The conserved motifs of the proteins of *R. tomentosa*, *R. argentea*, and *E. grandis* were analyzed using the Multiple Em for Motif Elucidation (MEME) program version 5.5.0 [85] with options set to any number of repetitions and the number of motifs was set to 10.

We mapped the long-read sequences to the published *R. tomentosa* chloroplast genome (NC_043848.1) using NOVOPlasty version 4.3.1 [86] to retrieve the chloroplast genome. For phylogenetic analysis, the chloroplast genomes were aligned with other plant chloroplast genomes, *Angophora floribunda* (NC_022411.1), *A. thaliana* (NC_000932.1), *B. napus* (NC_016734.1), *C. clementina* (NC_059803.1), *Corymbia maculata* (NC_022408.1), *C. papaya* (NC_010323.1), *Campomanesia xanthocarpa* (KY392760.1), *D. zibethinus* (NC_036829.1), *Eucalyptus camaldulensis* (NC_022398.1), *E. grandis* (NC_014570.1), *G. hirsutum* (NC_007944.1), *Pistacia vera* (NC_034998.1), *P. guajava* (NC_033355.1), *P. dulcis* (NC_034696.1), *P. granatum* (NC_035240.1), and *Syzygium samarangense* (NC_060657.1) using MAFFT version 7.505 [87] (Supplementary data S2). A maximum likelihood tree was obtained using the IQ-TREE version 2.2.0 [84] with the UNREST+FO+R3 model, 1000 approximate likelihood ratio tests, and 10,000 ultrafast bootstraps.

4.2. Population Genetic Analysis

Leaves of *R. tomentosa* from twelve locations across the Thai Peninsula were collected in 2021 (Figure 7, Table 1). The leaves from three to eleven trees per location were collected as *R. tomentosa* was sparsely populated. The DNeasy Plant Mini Kit (QIAgen, Hilden, Germany) was used for DNA extraction. The DNA quality was analyzed with 1.5% agarose gel electrophoresis.



Figure 7. Twelve collection sites of *Rhodomyrtus tomentosa* across Thai Peninsula. Red circles and blue triangles denote the east and west sides of the Peninsula, respectively.

Location	Coordinate	Side	n Samples
Chumphon	9°53′49.8″ N 99°08′16.2″ E	East	11
Krabi	7°51′43.4″ N 99°10′08.4″ E	West	8
Narathiwas	6°17'30.0″ N 101°58'41.3″ E	East	7
Nakhon-si- thammarat	8°37′09.7″ N 99°57′07.2″ E	East	8
Phuket	8°02′09.1″ N 98°17′51.5″ E	West	4
Phang-nga	8°49′36.1″ N 98°20′33.0″ E	West	7
Phatthalung	7°09'46.3" N 100°04'58.4" E	East	6
Pattani	6°52′10.4″ N 101°27′58.8″ E	East	3
Songkhla	7°13′45.0″ N 100°22′21.5″ E	East	7
Surat-thani	9°25′49.2″ N 99°16′16.2″ E	East	7
Satun	6°56′46.2″ N 99°41′44.5″ E	West	4
Trang	7°14′35.9″ N 99°33′01.4″ E	West	8

Table 1. Collection sites, side of Thai Peninsula, and number of samples collected of *Rhodomyrtustomentosa* for population genetic analysis.

Population genetics was accessed using ISSR and SSR markers. Ten ISSR loci previously used by Hue et al. [19] to study the population genetics of *R. tomentosa* in Malaysia were selected. For the SSR marker, thirteen primers were designed by searching for sequences of di-, tri-, and tetranucleotide repeats in the *R. tomentosa* genome using WebSat [88] and Tandem Repeats Finder version 4.07b [89] (Table S13). All ISSR and SSR loci were amplified using AllTaq DNA polymerase (QIAgen, Hilden, Germany).

Fragment size scoring was performed using the R package Fragman version 1.09 [90] in R version 4.1.1 [91]. Simpson index, evenness index, expected heterozygosity, and analysis of molecular variance were calculated using the R package Poppr version 2.9.3 [92]. Mmod version 1.3.3 [93] was used to calculate Nei's coefficient of gene differentiation (G_{ST}) between the locations. The polymorphism information content of each ISSR and SSR locus was calculated using iMEC calculator [94]. Similarity among samples across populations was observed with PCA, calculated using Lynch distance using the R package adegenet version 2.1.6 [95]. The UPGMA dendrogram of the samples based on Nei's genetic distance was constructed using Structure version 2.9.3 [92] with 1000 bootstraps. The population structure was examined using Structure version 2.3.4 [96]. Five replicates were performed using a correlated allele frequency model for every 500,000 generations, with the number of genetic clusters (K) ranging from one to twelve. The first 50,000 burn-in iterations were discarded. The optimal number of populations was determined based on the highest ΔK value using Structure Harvester version 0.6.94 [97].

4.3. Geometric Morphometrics

The leaves of *R. tomentosa* from the twelve locations used in the population genetics analysis were used in geometric morphometrics. A total of 11–40 leaves per location

were photographed. As the leaf venation patterns of *R. tomentosa* were challenging to observe, four landmarks were selected (Figure S7). A TPS file was created using tpsUtil64

observe, four landmarks were selected (Figure S7). A TPS file was created using tpsUtil64 version 1.81 (http://sbmorphometrics.org/soft-utility.html, 9 February 2023). Each leaf was digitized twice using the tpsDig2 version 2.31 (http://sbmorphometrics.org/soft-dataacq.html, 9 February 2023), and the mean landmarks of the two photographs were used for further analyses to reduce digitization errors. Procrustes analysis of variance of leaves among populations, regression, and canonical variate analysis was performed in MorphoJ version 1.07a [98]. Nei's genetic distance of the ISSR and SSR markers and Mahalanobis distance of the geometric morphometrics data were used to test the correlation between genetic and morphometrics data. The correlation analysis was performed using the Mantel test function in vegan version 2.6-4 [99].

Supplementary Materials: The following supporting information can be downloaded at: https:// www.mdpi.com/article/10.3390/plants12081582/s1, Figure S1: Blobplot of Rhodomyrtus tomentosa final assembly; Figure S2: Maximum likelihood phylogenetic tree based on 236 orthologous gene groups; Figure S3: Maximum likelihood tree based on chloroplast genome; Figure S4: Conserved motif analysis of terpene synthase genes in Rhodomyrtus tomentosa, Rhodamnia argentea, Eucalyptus grandis, Punica granatum, and Prunus dulcis; Figure S5: Phylogenetic tree of terpene synthase genes in Rhodomyrtus tomentosa, Rhodamnia argentea, Eucalyptus grandis, Punica granatum, and Prunus dulcis; Figure S6: Admixture analysis bar plots of genetic structure inferred by Bayesian clustering at K = 2 and 3 using ISSR and SSR data; Figure S7: Four geometric morphometrics landmarks used to digitized Rhodomyrtus tomentosa leaves; Table S1: Genome assembly statistics; Table S2: Repeats statistics of *Rhodomyrtus tomentosa* compared with *Rhodamnia argentea* genome; Table S3. Predicted genes statistics; Table S4: Gene ontology annotation of Rhodomyrtus tomentosa generated using eggNOG; Table S5. KEGG pathways identified from Rhodomyrtus tomentosa genes; Table S6: Statistics of orthologous gene analysis per species; Table S7: Orthogroups with significant expansion or contraction rate of Rhodomyrtus tomentosa; Table S8. Terpene synthase domains (RRx8W, RxR, and DDxxD) of Rhodomyrtus tomentosa; Table S9. Summary statistics per locus of ISSR and SSR markers; Table S10. Genotypic diversity per location calculated using ISSR and SSR markers; Table S11. Pairwise G_{ST} between *Rhodomyrtus tomentosa* locations based on ISSR and SSR markers; Table S12. Mahalanobis distances among Rhodomyrtus tomentosa populations; Table S13. Simple sequence repeats (SSR) loci developed from genomic data of Rhodomyrtus tomentosa.

Author Contributions: M.D., S.B. and S.P.V. designed the study. S.B. collected the samples. M.D. analyzed the data. M.D. and S.P.V. interpreted the results. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by National Science, Research and Innovation Fund (NSRF) and Prince of Songkla University (Grant No. SCI6505182M).

Data Availability Statement: All supplementary data, tables, and figures are available on OSF (DOI 10.17605/OSF.IO/ZHJDN, https://osf.io/zhjdn/ (9 February 2023)). Sequences of *R. tomentosa* are available in GenBank, BioProject accession number PRJNA846150.

Acknowledgments: We thank Sirithorn Kongseng for helping in the microsatellite analysis, and Hattaya Jaroensap for plant collection.

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

References

- 1. Wu, T. (Ed.) *Check List of Hong Kong Plants. Agriculture, Fisheries and Conservation Department Bulletin 1 (Revised Edn);* Hong Kong Herbarium and the South China Institute of Botany: Hong Kong, China, 2002.
- Winotai, A.; Wright, T.; Goolsby, J.A. Herbivores in Thailand on *Rhodomyrtus tomentosa* (Myrtaceae), an Invasive Weed in Florida. *Fla. Entomol.* 2005, 88, 104–105. [CrossRef]
- 3. Ren, H.; Peng, S.; Sun, G.; Yu, Z. The Ecological Comparison of *Psychotria rubra* and *Rhodomyrtus tomentosa* in South China. *Chin. J. Plant Ecol.* **1997**, *21*, 386–392.
- 4. Langeland, K.A.; Craddock Burks, K. Identification and Biology of Non-Native Plants in Florida's Natural Areas; University of Florida: Gainesville, FL, USA, 1998.

- 5. Wei, M.S.; Chen, Z.H.; Ren, H.; Yin, Z.Y. Reproductive Ecology of *Rhodomyrtus tomentosa* (Myrtaceae). *Nord. J. Bot.* 2009, 27, 154–160. [CrossRef]
- 6. Ong, H.C.; Nordiana, M. Malay Ethno-Medico Botany in Machang, Kelantan, Malaysia. Fitoterapia 1999, 70, 502–513. [CrossRef]
- 7. Wei, F. Manufacture of Oral Liquid Contaning Traditional Chinese Medicine Extract for Treating Gynecopathy; Guangxi Huahong Pharmaceutical Co., Ltd.: Liuzhou, China, 2006.
- Zhao, Z.; Wu, L.; Xie, J.; Feng, Y.; Tian, J.; He, X.; Li, B.; Wang, L.; Wang, X.; Zhang, Y.; et al. *Rhodomyrtus tomentosa* (Aiton.): A Review of Phytochemistry, Pharmacology and Industrial Applications Research Progress. *Food Chem.* 2020, 309, 125715. [CrossRef]
- Limsuwan, S.; Trip, E.N.; Kouwen, T.R.H.M.; Piersma, S.; Hiranrat, A.; Mahabusarakam, W.; Voravuthikunchai, S.P.; van Dijl, J.M.; Kayser, O. Rhodomyrtone: A New Candidate as Natural Antibacterial Drug from *Rhodomyrtus tomentosa*. *Phytomedicine* 2009, 16, 645–651. [CrossRef]
- 10. Leejae, S.; Hasap, L.; Voravuthikunchai, S.P. Inhibition of Staphyloxanthin Biosynthesis in *Staphylococcus aureus* by Rhodomyrtone, a Novel Antibiotic Candidate. *J. Med. Microbiol.* **2013**, *62*, 421–428. [CrossRef]
- 11. Chorachoo, J.; Lambert, S.; Furnholm, T.; Roberts, L.; Reingold, L.; Auepemkiate, S.; Voravuthikunchai, S.P.; Johnston, A. The Small Molecule Rhodomyrtone Suppresses TNF-∞ and IL-17A-Induced Keratinocyte Inflammatory Responses: A Potential New Therapeutic for Psoriasis. *PLoS ONE* **2018**, *13*, e0205340. [CrossRef]
- Saising, J.; Nguyen, M.T.; Härtner, T.; Ebner, P.; Al Mamun Bhuyan, A.; Berscheid, A.; Muehlenkamp, M.; Schäkermann, S.; Kumari, N.; Maier, M.E.; et al. Rhodomyrtone (Rom) Is a Membrane-Active Compound. *Biochim. Biophys. Acta Biomembr.* 2018, 1860, 1114–1124. [CrossRef]
- 13. Wunnoo, S.; Bilhman, S.; Amnuaikit, T.; Ontong, J.C.; Singh, S.; Auepemkiate, S.; Voravuthikunchai, S.P. Rhodomyrtone as a New Natural Antibiotic Isolated from *Rhodomyrtus tomentosa* Leaf Extract: A Clinical Application in the Management of Acne Vulgaris. *Antibiotics* **2021**, *10*, 108. [CrossRef]
- 14. Lai, T.N.H.; Herent, M.F.; Quetin-Leclercq, J.; Nguyen, T.B.T.; Rogez, H.; Larondelle, Y.; André, C.M. Piceatannol, a Potent Bioactive Stilbene, as Major Phenolic Component in *Rhodomyrtus tomentosa*. *Food Chem.* **2013**, *138*, 1421–1430. [CrossRef] [PubMed]
- 15. Grattapaglia, D.; Vaillancourt, R.E.; Shepherd, M.; Thumma, B.R.; Foley, W.; Külheim, C.; Potts, B.M.; Myburg, A.A. Progress in Myrtaceae Genetics and Genomics: *Eucalyptus* as the Pivotal Genus. *Tree Genet. Genomes* **2012**, *8*, 463–508. [CrossRef]
- 16. Huang, Y.; Yang, Z.; Huang, S.; An, W.; Li, J.; Zheng, X. Comprehensive Analysis of *Rhodomyrtus tomentosa* Chloroplast Genome. *Plants* **2019**, *8*, 89. [CrossRef] [PubMed]
- 17. Li, F.; Xu, S.; Xiao, Z.; Wang, J.; Mei, Y.; Hu, H.; Li, J.; Liu, J.; Hou, Z.; Zhao, J.; et al. Gap-Free Genome Assembly and Comparative Analysis Reveal the Evolution and Anthocyanin Accumulation Mechanism of *Rhodomyrtus tomentosa*. *Hortic. Res.* **2023**, *10*, uhad005. [CrossRef] [PubMed]
- 18. Vijayakumar, N.; Subramanian, D. Cytotaxonomical Studies in South Indian Myrtaceae. Cytologia 1985, 50, 513–520. [CrossRef]
- 19. Hue, T.S.; Abdullah, T.L.; Abdullah, N.A.P.; Sinniah, U.R. Genetic Variation in *Rhodomyrtus tomentosa* (Kemunting) Populations from Malaysia as Revealed by Inter-Simple Sequence Repeat Markers. *Genet. Mol. Res.* **2015**, *14*, 16827–16839. [CrossRef]
- 20. Sun, L.; Li, J.; Sun, K.; Wang, H.; Yang, K.; Chen, Q.; Lin, M. Development and Characterization of EST-SSR Markers in *Rhodomyrtus tomentosa* Based on Transcriptome. *Genet. Resour. Crop Evol.* **2022**. [CrossRef]
- Snow, N.; McFadden, J.; Evans, T.M.; Salywon, A.M.; Wojciechowski, M.F.; Wilson, P.G. Morphological and Molecular Evidence of Polyphyly in *Rhodomyrtus* (Myrtaceae: Myrteae). *Syst. Bot.* 2011, *36*, 390–404. [CrossRef]
- 22. Carvalho, G.M.A.; Carvalho, C.R.; Soares, F.A.F. Flow Cytometry and Cytogenetic Tools in Eucalypts: Genome Size Variation × Karyotype Stability. *Tree Genet. Genomes* 2017, *13*, 106. [CrossRef]
- 23. Wang, W.; Das, A.; Kainer, D.; Schalamun, M.; Morales-Suarez, A.; Schwessinger, B.; Lanfear, R. The Draft Nuclear Genome Assembly of *Eucalyptus pauciflora*: A Pipeline for Comparing de novo Assemblies. *Gigascience* **2020**, *9*, giz160. [CrossRef]
- 24. Lai, T.N.H.; André, C.; Rogez, H.; Mignolet, E.; Nguyen, T.B.T.; Larondelle, Y. Nutritional Composition and Antioxidant Properties of the Sim Fruit (*Rhodomyrtus tomentosa*). *Food Chem.* **2015**, *168*, 410–416. [CrossRef] [PubMed]
- 25. Narusaka, M.; Kubo, Y.; Hatakeyama, K.; Imamura, J.; Ezura, H.; Nanasato, Y.; Tabei, Y.; Takano, Y.; Shirasu, K.; Narusaka, Y. Interfamily Transfer of Dual NB-LRR Genes Confers Resistance to Multiple Pathogens. *PLoS ONE* **2013**, *8*, e55954. [CrossRef]
- Dang, P.M.; Lamb, M.C.; Chen, C.Y. Association of Differentially Expressed R-Gene Candidates with Leaf Spot Resistance in Peanut (*Arachis hypogaea* L.). *Mol. Biol. Rep.* 2021, 48, 323–334. [CrossRef] [PubMed]
- 27. Rossi, M.; Goggin, F.L.; Milligan, S.B.; Kaloshian, I.; Ullman, D.E.; Williamson, V.M. The Nematode Resistance Gene Mi of Tomato Confers Resistance against the Potato Aphid. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 9750–9754. [CrossRef] [PubMed]
- Kubaa, R.A.; Giampetruzzi, A.; Altamura, G.; Saponari, M.; Saldarelli, P. Infections of the *Xylella fastidiosa* Subsp. *pauca* Strain "De Donno" in Alfalfa (*Medicago sativa*) Elicits an Overactive Immune Response. *Plants* 2019, *8*, 335. [CrossRef] [PubMed]
- 29. Zhu, C.; Li, X.; Zheng, J. Transcriptome Profiling Using Illumina- and SMRT-Based RNA-Seq of Hot Pepper for in-Depth Understanding of Genes Involved in CMV Infection. *Gene* **2018**, *666*, 123–133. [CrossRef]
- Brophy, J.J.; Goldsack, R.J.; Forster, P.I. The Leaf Essential Oils of the Australian Species of *Rhodamnia* (Myrtaceae). *Flavour Fragr. J.* 1997, 12, 345–354. [CrossRef]
- Vasconcelos, T.N.C.; Proença, C.E.B.; Ahmad, B.; Aguilar, D.S.; Aguilar, R.; Amorim, B.S.; Campbell, K.; Costa, I.R.; De-Carvalho, P.S.; Faria, J.E.Q.; et al. Myrteae Phylogeny, Calibration, Biogeography and Diversification Patterns: Increased Understanding in the Most Species Rich Tribe of Myrtaceae. *Mol. Phylogenetics Evol.* 2017, 109, 113–137. [CrossRef]

- 32. Feng, C.; Feng, C.; Lin, X.; Liu, S.; Li, Y.; Kang, M. A Chromosome-Level Genome Assembly Provides Insights into Ascorbic Acid Accumulation and Fruit Softening in Guava (*Psidium guajava*). *Plant Biotechnol. J.* **2021**, *19*, 717–730. [CrossRef]
- 33. Healey, A.L.; Shepherd, M.; King, G.J.; Butler, J.B.; Freeman, J.S.; Lee, D.J.; Potts, B.M.; Silva-Junior, O.B.; Baten, A.; Jenkins, J.; et al. Pests, Diseases, and Aridity Have Shaped the Genome of *Corymbia citriodora*. *Commun. Biol.* **2021**, *4*, 537. [CrossRef]
- Izuno, A.; Wicker, T.; Hatakeyama, M.; Copetti, D.; Shimizu, K.K. Updated Genome Assembly and Annotation for *Metrosideros polymorpha*, an Emerging Model Tree Species of Ecological Divergence. *G3 Genes Genomes Genet.* 2019, *9*, 3513. [CrossRef] [PubMed]
- Thornhill, A.H.; Ho, S.Y.W.; Külheim, C.; Crisp, M.D. Interpreting the Modern Distribution of Myrtaceae Using a Dated Molecular Phylogeny. *Mol. Phylogenetics Evol.* 2015, 93, 29–43. [CrossRef] [PubMed]
- 36. Balbinott, N.; Rodrigues, N.F.; Guzman, F.L.; Turchetto-Zolet, A.C.; Margis, R. Perspectives in Myrtaceae Evolution from Plastomes and Nuclear Phylogenies. *Genet. Mol. Biol.* **2022**, *45*, e20210191. [CrossRef]
- 37. Padovan, A.; Keszei, A.; Külheim, C.; Foley, W.J. The Evolution of Foliar Terpene Diversity in Myrtaceae. *Phytochem. Rev.* 2014, 13, 695–716. [CrossRef]
- Calvert, J.; Baten, A.; Butler, J.; Barkla, B.; Shepherd, M. Terpene Synthase Genes in *Melaleuca alternifolia*: Comparative Analysis of Lineage-Specific Subfamily Variation within Myrtaceae. *Plant Syst. Evol.* 2018, 304, 111–121. [CrossRef]
- Külheim, C.; Padovan, A.; Hefer, C.; Krause, S.T.; Köllner, T.G.; Myburg, A.A.; Degenhardt, J.; Foley, W.J. The *Eucalyptus* Terpene Synthase Gene Family. *BMC Genom.* 2015, 16, 450. [CrossRef]
- Hanada, K.; Zou, C.; Lehti-Shiu, M.D.; Shinozaki, K.; Shiu, S.H. Importance of Lineage-Specific Expansion of Plant Tandem Duplicates in the Adaptive Response to Environmental Stimuli. *Plant Physiol.* 2008, 148, 993–1003. [CrossRef]
- 41. Ravi Kiran, S.; Sita Devi, P. Evaluation of Mosquitocidal Activity of Essential Oil and Sesquiterpenes from Leaves of *Chloroxylon swietenia* DC. *Parasitol. Res.* 2007, 101, 413–418. [CrossRef]
- Rai, M.K.; Phulwaria, M.; Harish; Gupta, A.K.; Shekhawat, N.S.; Jaiswal, U. Genetic Homogeneity of Guava Plants Derived from Somatic Embryogenesis Using SSR and ISSR Markers. *Plant Cell Tissue Organ Cult.* 2012, 111, 259–264. [CrossRef]
- Quezada, M.; Pastina, M.M.; Ravest, G.; Silva, P.; Vignale, B.; Cabrera, D.; Hinrichsen, P.; Garcia, A.A.F.; Pritsch, C. A First Genetic Map of *Acca sellowiana* Based on ISSR, AFLP and SSR Markers. *Sci. Hortic.* 2014, 169, 138–146. [CrossRef]
- 44. Liao, P.C.; Chiang, Y.C.; Huang, S.; Wang, J.C. Gene Flow of *Ceriops tagal* (Rhizophoraceae) across the Kra Isthmus in the Thai Malay Peninsula. *Bot. Stud.* **2009**, *50*, 193–204.
- 45. Nguyen, V.X.; Detcharoen, M.; Tuntiprapas, P.; Soe-Htun, U.; Sidik, J.B.; Harah, M.Z.; Prathep, A.; Papenbrock, J. Genetic Species Identification and Population Structure of *Halophila* (Hydrocharitaceae) from the Western Pacific to the Eastern Indian Ocean. *BMC Evol. Biol.* **2014**, *14*, 92. [CrossRef]
- Wee, A.K.S.; Noreen, A.M.E.; Ono, J.; Takayama, K.; Kumar, P.P.; Tan, H.T.W.; Saleh, M.N.; Kajita, T.; Webb, E.L. Genetic Structures across a Biogeographical Barrier Reflect Dispersal Potential of Four Southeast Asian Mangrove Plant Species. *J. Biogeogr.* 2020, 47, 1258–1271. [CrossRef]
- Hargreaves, S.; Maxted, N.; Hirano, R.; Abberton, M.; Skøt, L.; Ford-Lloyd, B.V. Islands as Refugia of *Trifolium repens* Genetic Diversity. *Conserv. Genet.* 2010, 11, 1317–1326. [CrossRef]
- 48. Dias, E.F.; Moura, M.; Schaefer, H.; Silva, L. Geographical Distance and Barriers Explain Population Genetic Patterns in an Endangered Island Perennial. *AoB Plants* **2016**, *8*, plw072. [CrossRef]
- 49. Burton, G.P.; Campbell, K.C.S.E.; Lucas, E.J. Morphometric Analysis as a Tool to Resolve a Taxonomic Complex in *Myrcia* Sect. Calyptranthes (Myrtaceae, Myrteae). *Syst. Bot.* **2022**, *46*, 1016–1025. [CrossRef]
- 50. Dalastra, C.; Sausen, T.L.; Capellesso, E.S.; Fornel, R. Variations in Leaf Size and Leaf Shape in Four Species of *Eugenia* (Myrtaceae) Using Geometric Morphometrics Approach. *Pesqui. Botânica* **2021**, *75*, 143–154.
- 51. Jones, R.C.; Steane, D.A.; Potts, B.M.; Vaillancourt, R.E. Microsatellite and Morphological Analysis of *Eucalyptus globulus* Populations. *Can. J. For. Res.* 2002, 32, 59–66. [CrossRef]
- Jensen, R.J.; Schwoyer, M.; Crawford, D.J.; Stuessy, T.F.; Anderson, G.J.; Baeza, C.M.; Mario Silva, O.; Ruiz, E. Patterns of Morphological and Genetic Variation among Populations of *Myrceugenia fernandeziana* (Myrtaceae) on Masatierra Island: Implications for Conservation. *Syst. Bot.* 2002, 27, 534–547.
- 53. Doyle, J.J.; Doyle, J.L. A Rapid DNA Isolation Procedure from Small Quantities of Fresh Leaf Material. *Phytochem Bull* **1987**, 19, 11–15.
- Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor. *Bioinformatics* 2018, 34, i884–i890. [CrossRef] [PubMed]
- Marçais, G.; Kingsford, C. A Fast, Lock-Free Approach for Efficient Parallel Counting of Occurrences of k-mers. *Bioinformatics* 2011, 27, 764–770. [CrossRef] [PubMed]
- 56. Ranallo-Benavidez, T.R.; Jaron, K.S.; Schatz, M.C. GenomeScope 2.0 and Smudgeplot for Reference-Free Profiling of Polyploid Genomes. *Nat. Commun.* **2020**, *11*, 1432. [CrossRef]
- 57. Kolmogorov, M.; Yuan, J.; Lin, Y.; Pevzner, P.A. Assembly of Long, Error-Prone Reads Using Repeat Graphs. *Nat. Biotechnol.* 2019, 37, 540–546. [CrossRef]
- Vaser, R.; Sović, I.; Nagarajan, N.; Šikić, M. Fast and Accurate de novo Genome Assembly from Long Uncorrected Reads. *Genome Res.* 2017, 27, 737–746. [CrossRef]

- Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K.; et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS ONE* 2014, 9, e112963. [CrossRef] [PubMed]
- Guan, D.; Guan, D.; McCarthy, S.A.; Wood, J.; Howe, K.; Wang, Y.; Durbin, R.; Durbin, R. Identifying and Removing Haplotypic Duplication in Primary Genome Assemblies. *Bioinformatics* 2020, *36*, 2896–2898. [CrossRef]
- 61. Laetsch, D.R.; Blaxter, M.L. BlobTools: Interrogation of Genome Assemblies. *F1000Res* 2017, *6*, 1287. [CrossRef]
- 62. Mikheenko, A.; Prjibelski, A.; Saveliev, V.; Antipov, D.; Gurevich, A. Versatile Genome Assembly Evaluation with QUAST-LG. *Bioinformatics* **2018**, *34*, i142–i150. [CrossRef]
- 63. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [CrossRef]
- 64. Flynn, J.M.; Hubley, R.; Goubert, C.; Rosen, J.; Clark, A.G.; Feschotte, C.; Smit, A.F. RepeatModeler2 for Automated Genomic Discovery of Transposable Element Families. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 9451–9457. [CrossRef]
- Smit, A.; Hubley, R.; Green, P. RepeatMasker Open-4.0 2015. Available online: https://www.repeatmasker.org/ (accessed on 9 February 2023).
- Luo, X.; Chen, S.; Zhang, Y. PlantRep: A Database of Plant Repetitive Elements. *Plant Cell Rep.* 2022, 41, 1163–1166. [CrossRef]
 [PubMed]
- 67. Bushmanova, E.; Antipov, D.; Lapidus, A.; Prjibelski, A.D. RnaSPAdes: A *de novo* Transcriptome Assembler and Its Application to RNA-Seq Data. *Gigascience* **2019**, *8*, giz100. [CrossRef] [PubMed]
- 68. Bateman, A. UniProt: A Worldwide Hub of Protein Knowledge. Nucleic Acids Res. 2019, 47, D506–D515. [CrossRef]
- 69. Stanke, M.; Keller, O.; Gunduz, I.; Hayes, A.; Waack, S.; Morgenstern, B. AUGUSTUS: Ab initio Prediction of Alternative Transcripts. *Nucleic Acids Res.* 2006, *34*, W435–W439. [CrossRef]
- Lomsadze, A.; Ter-Hovhannisyan, V.; Chernoff, Y.O.; Borodovsky, M. Gene Identification in Novel Eukaryotic Genomes by Self-Training Algorithm. *Nucleic Acids Res.* 2005, 33, 6494–6506. [CrossRef]
- 71. Brůna, T.; Hoff, K.J.; Lomsadze, A.; Stanke, M.; Borodovsky, M. BRAKER2: Automatic Eukaryotic Genome Annotation with GeneMark-EP+ and AUGUSTUS Supported by a Protein Database. *NAR Genom. Bioinform.* **2021**, *3*, lqaa108. [CrossRef]
- Cantalapiedra, C.P.; Hernandez-Plaza, A.; Letunic, I.; Bork, P.; Huerta-Cepas, J. EggNOG-Mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol. Biol. Evol.* 2021, 38, 5825–5829. [CrossRef]
- 73. Emms, D.M.; Kelly, S. OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics. *Genome Biol.* 2019, 20, 238. [CrossRef]
- Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2—Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* 2010, 5, e9490. [CrossRef]
- 75. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol. Biol. Evol. 2007, 24, 1586–1591. [CrossRef]
- Beilstein, M.A.; Nagalingum, N.S.; Clements, M.D.; Manchester, S.R.; Mathews, S. Dated Molecular Phylogenies Indicate a Miocene Origin for *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA 2010, 107, 18724–18728. [CrossRef]
- Wikström, N.; Savolainen, V.; Chase, M.W. Evolution of the Angiosperms: Calibrating the Family Tree. *Proc. R. Soc. B Biol. Sci.* 2001, 268, 2211–2220. [CrossRef] [PubMed]
- 78. Qin, G.; Xu, C.; Ming, R.; Tang, H.; Guyot, R.; Kramer, E.M.; Hu, Y.; Yi, X.; Qi, Y.; Xu, X.; et al. The Pomegranate (*Punica granatum* L.) Genome and the Genomics of Punicalagin Biosynthesis. *Plant J.* **2017**, *91*, 1108–1128. [CrossRef] [PubMed]
- 79. Han, M.V.; Thomas, G.W.C.; Lugo-Martinez, J.; Hahn, M.W. Estimating Gene Gain and Loss Rates in the Presence of Error in Genome Assembly and Annotation Using CAFE 3. *Mol. Biol. Evol.* **2013**, *30*, 1987–1997. [CrossRef] [PubMed]
- 80. Alonge, M.; Soyk, S.; Ramakrishnan, S.; Wang, X.; Goodwin, S.; Sedlazeck, F.J.; Lippman, Z.B.; Schatz, M.C. RaGOO: Fast and Accurate Reference-Guided Scaffolding of Draft Genomes. *Genome Biol.* **2019**, *20*, 224. [CrossRef]
- Kurtz, S.; Phillippy, A.; Delcher, A.L.; Smoot, M.; Shumway, M.; Antonescu, C.; Salzberg, S.L. Versatile and Open Software for Comparing Large Genomes. *Genome Biol.* 2004, 5, R12. [CrossRef]
- Krzywinski, M.; Schein, J.; Birol, I.; Connors, J.; Gascoyne, R.; Horsman, D.; Jones, S.J.; Marra, M.A. Circos: An Information Aesthetic for Comparative Genomics. *Genome Res.* 2009, 19, 1639–1645. [CrossRef]
- 83. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; Mcgettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; et al. Clustal W and Clustal X Version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [CrossRef]
- Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; Von Haeseler, A.; Lanfear, R.; Teeling, E. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* 2020, 37, 1530–1534. [CrossRef]
- Bailey, T.L.; Johnson, J.; Grant, C.E.; Noble, W.S. The MEME Suite. *Nucleic Acids Res.* 2015, 43, W39–W49. [CrossRef]
 Dierckxsens, N.; Mardulyn, P.; Smits, G. NOVOPlasty: De novo Assembly of Organelle Genomes from Whole Genome Data.
- Nucleic Acids Res. 2017, 45, gkw955. [CrossRef]
 87. Katoh, K.; Misawa, K.; Kuma, K.I.; Miyata, T. MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform. Nucleic Acids Res. 2002, 30, 3059–3066. [CrossRef]
- 88. Martins, W.S.; Lucas, D.C.S.; Neves, K.F.S.; Bertioli, D.J. WebSat—A Web Software for Microsatellite Marker Development. *Bioinformation* **2009**, *3*, 282–283. [CrossRef] [PubMed]
- Benson, G. Tandem Repeats Finder: A Program to Analyze DNA Sequences. Nucleic Acids Res. 1999, 27, 573–580. [CrossRef] [PubMed]

- 90. Covarrubias-Pazaran, G.; Diaz-Garcia, L.; Schlautman, B.; Salazar, W.; Zalapa, J. Fragman: An R Package for Fragment Analysis. BMC Genet. 2016, 17, 62. [CrossRef] [PubMed]
- 91. R Core Team R. A Language and Environment for Statistical Computing; R Core Team R: Vienna, Austria, 2020.
- 92. Kamvar, Z.N.; Tabima, J.F.; Grünwald, N.J. Poppr: An R Package for Genetic Analysis of Populations with Clonal, Partially Clonal, and/or Sexual Reproduction. *PeerJ* 2014, 2014, e281. [CrossRef]
- 93. Winter, D.J. MMOD: An R Library for the Calculation of Population Differentiation Statistics. *Mol. Ecol. Resour.* 2012, 12, 1158–1160. [CrossRef]
- 94. Amiryousefi, A.; Hyvönen, J.; Poczai, P. IMEC: Online Marker Efficiency Calculator. Appl. Plant Sci. 2018, 6, e01159. [CrossRef]
- 95. Jombart, T.; Ahmed, I. Adegenet 1.3-1: New Tools for the Analysis of Genome-Wide SNP Data. *Bioinformatics* 2011, 27, 3070–3071. [CrossRef]
- 96. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 2000, 155, 945–959. [CrossRef] [PubMed]
- Earl, D.A.; Vonholdt, B.M. STRUCTURE HARVESTER: A Website and Program for Visualizing STRUCTURE Output and Implementing the Evanno Method. *Conserv. Genet. Resour.* 2012, *4*, 359–361. [CrossRef]
- Klingenberg, C.P. MorphoJ: An Integrated Software Package for Geometric Morphometrics. *Mol. Ecol. Resour.* 2011, 11, 353–357. [CrossRef] [PubMed]
- Oksanen, J.; Blanchet, F.G.; Friendly, M.; Kindt, R.; Legendre, P.; McGlinn, D.; Minchin, P.R.; O'Hara, R.B.; Simpson, G.L.; Solymos, P. Vegan: Community Ecology Package. R Package Version 2.5-6. 2019. Available online: https://cran.r-project.org/package= vegan (accessed on 9 February 2023).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.