diversity

IMPACT FACTOR 2.4

CITESCORE 3.1

Article

Identifying Molecular Markers for *Ficus erecta* Thunb. Based on Complete Plastome Sequences of Korean Figs (*Ficus* L., Moraceae)

Joonhyung Jung, Tae-Hee Kim, Seog Woo Kwon, Hyun Ji Park, In Suk Choi and Joo-Hwan Kim

MDPI

https://doi.org/10.3390/d16030129

diversity

MDPI

Article

Identifying Molecular Markers for *Ficus erecta* Thunb. Based on Complete Plastome Sequences of Korean Figs (*Ficus* L., Moraceae)

Joonhyung Jung [1], Tae-Hee Kim [2], Seog Woo Kwon [1], Hyun Ji Park [3], In Suk Choi [4] and Joo-Hwan Kim [1,*]

[1] Department of Life Sciences, Gachon University, Seongnam 13120, Republic of Korea; joonhyung0104@gmail.com (J.J.); jounchulbal@hanmail.net (S.W.K.)
[2] Division of Forest Biodiversity, Korea National Arboretum, Pocheon 11186, Republic of Korea; thkim3813@korea.kr
[3] Center for Consumer Health 1 Research, CHA Advanced Research Institute, Seongnam 13488, Republic of Korea; hyunjee@chamc.co.kr
[4] The 2nd Research Institute, CMG Pharmaceutical Co., Ltd., Seongnam 13488, Republic of Korea; inschoi@chamc.co.kr
* Correspondence: kimjh2009@gachon.ac.kr; Tel.: +82-317508827

Abstract: Plastome sequences are crucial in plant studies due to their role in examining genomic evolution, understanding phylogenetic relationships, and developing molecular markers. Despite the collection of information about Korean figs, their genomic data remain underexplored. We utilize next-generation sequencing and PCR techniques to investigate genomic data and to develop and validate molecular markers. In this study, we characterize the complete plastomes of Korean figs: *F. erecta*, *F. erecta* var. *sieboldii*, *F. sarmentosa* var. *nipponica*, and *F. sarmentosa* var. *thunbergii*, which range in length from 160,276 to 160,603 bp. These genomes comprise 78 plastid protein-coding genes, 30 tRNA, and four rRNA, with the exception of one pseudogene, *inf* A. We discovered that *F. erecta* and *F. erecta* var. *sieboldii* share identical plastome sequences. Phylogenomic analysis indicates the monophyly of *Ficus*, although the relationships among its subgenera remain unclear. We discovered that *Ficus* possesses 467 molecular diagnostic characters in its plastid protein-coding genes compared to other Moraceae groups, and *F. erecta* exhibits 33 molecular diagnostic characters. Single nucleotide polymorphisms in *ndh*D, *pet*A, and *rbc*L were effectively used to develop molecular markers for distinguishing *F. erecta* from other figs. Additionally, we provided a straightforward PCR protocol for utilizing these newly developed molecular markers.

Keywords: Korean figs; plastid genome (plastome); molecular markers

Citation: Jung, J.; Kim, T.-H.; Kwon, S.W.; Park, H.J.; Choi, I.S.; Kim, J.-H. Identifying Molecular Markers for *Ficus erecta* Thunb. Based on Complete Plastome Sequences of Korean Figs (*Ficus* L., Moraceae). *Diversity* **2024**, *16*, 129. https://doi.org/10.3390/d16030129

Academic Editor: Michael Wink

Received: 28 January 2024
Revised: 15 February 2024
Accepted: 17 February 2024
Published: 20 February 2024

1. Introduction

The genus *Ficus* L., commonly referred to as figs, represents the largest genus within Moraceae Gaudich., with over 800 species [1]. It is distributed across both tropical and temperate regions, encompassing various vegetative types including trees, creepers, stranglers, rheophytes, and lithophytes [2]. A distinct characteristic of figs is the syconium, which is an enclosed inflorescence containing either staminate (male) or pistillate (female) flowers. One notable species, *F. erecta* Thunb., is a deciduous shrub that typically grows 2–5 m in height and is commonly found on forest edges. It is native to East Asia and distinguishable from *F. erecta* var. *sieboldii* (Miq.) King, which has narrower lanceolate leaves, though the exact distinction between the two remains unresolved. Recent research has highlighted several beneficial properties of extracts from this species, including tyrosinase inhibition, thrombolytic action, antioxidant properties, cytotoxic effects, and neuroprotective capabilities [3–6]. In Korea, two more varieties, *F. sarmentosa* var. *nipponica* (Franch. & Sav.) Corner and *F. sarmentosa* var. *thunbergii* (Maxim.) Corner, are native, and another, *F. carica* L., is widely cultivated.

Chloroplasts are essential for the process of photosynthesis. Typically, the plastid genome (plastome) exhibits a unique four-part structure, comprised of two inverted repeat (IR) regions that connect the large single-copy (LSC) and the small single-copy (SSC) regions. Plastome sequences provide valuable data for determining phylogenetic relationships, developing molecular markers to identify specific species, and exploring biogeography, due to their highly conserved characteristics [7–10]. DNA molecular markers identify genetic differences among species or individuals, such as insertions, deletions, inversions, and mutations, by analyzing DNA sequence variations [11]. PCR-based molecular markers, including random amplified polymorphic DNA (RAPD) [12], amplified fragment length polymorphism (AFLP) [13], microsatellites or simple sequence repeats (SSRs) [14,15], and single nucleotide polymorphisms (SNPs), simplify genetic analysis by using primer pairs to specifically target and amplify DNA regions, thus eliminating the need for probe hybridization. Among these, SNPs are widely used for species identification due to the ease of direct DNA sequence comparison [16,17]. In recent years, there has been progress in species identification methods through molecular diagnostic characters (MDCs), such as species-specific SNPs [18]. These advancements complement traditional barcoding techniques, aiming to refine the criteria for successful species identification by evaluating MDCs across different taxonomic groups. With the current abundance of plastome sequences, SNP markers can now be developed in silico and employed to analyze individual differences.

In this study, we employed next-generation sequencing (NGS) to complete the plastome sequences of *F. erecta*, *F. erecta* var. *sieboldii*, *F. sarmentosa* var. *nipponica*, and *F. sarmentosa* var. *thunbergii*, all of which exhibit potential medicinal applications, bear edible fruits, and are native to Korea. Although the complete plastome sequences of these species are already available, our primary objective was to identify *F. erecta*-specific SNP sites based on individual genomic data for the development of molecular markers. We then conducted comparative genomic analyses to explore the genomic diversity among these species, focusing particularly on highly variable regions. Furthermore, we reconstructed the formerly ambiguous relationship between figs and related taxa based on plastid protein-coding genes. Concluding our research, we developed specific SNP markers for the putative medically informative species, *F. erecta*, based on the amplification refractory mutation system (ARMS) technique across three plastid protein-coding genes.

2. Materials and Methods

2.1. Taxon Sampling, DNA Extraction, and Plastome Assembly

A total of 39 individuals of *F. erecta* (10 individuals), *F. erecta* var. *sieboldii* (4 individuals), *F. carica* (10 individuals), *F. sarmentosa* var. *nipponica* (7 individuals), and *F. sarmentosa* var. *thunbergii* (8 individuals) were collected in Korea (Table S1). After collecting the samples, we prepared voucher specimens for each and stored them in the Gachon University Herbarium (GCU), assigning distinct accession numbers. We extracted total genomic DNA (gDNA) of all samples using modified 2X CTAB method [19]. Fresh samples of Korean figs (*F. erecta*, *F. erecta* var. *sieboldii*, *F. sarmentosa* var. *nipponica*, and *F. sarmentosa* var. *thunbergii*) were used for next-generation sequencing (NGS) analysis using the Illumina Mi-seq platform (Illumina, Seoul, Republic of Korea; Table S2). We utilized the total raw reads for de novo assembly of plastome sequences using the GetOrganelle toolkit [20]. Then, we applied the map to reference tool to the plastome sequences to assess the coverage based on Geneious Prime 2022.0.2 program [21]. We annotated the gene content and sequence order with GeSeq [22]. All tRNAs underwent a secondary check using tRNAScan-SE in its default search mode [23]. In the end, OGDraw was employed to create visual representations of the full plastome sequences [24].

2.2. Phylogenetic Analysis

We downloaded 30 complete plastome sequences from the National Center for Biotechnology Information (NCBI), which included one from Moreae (*Streblus indicus* (Bureau) Corner; GenBank accession number MN065161), two from Dorstenieae (*Broussonetia kurzii*

(Hook.f.) Corner and *Malaisia scandens* (Lour.) Planch.; GenBank accession number MH118529 and MH189568, respectively), and one from Castilleae (*Antiaris toxicaria* (J.F.Gmel.) Lesch.; GenBank accession number MH606237) (Table S3). To infer the phylogenetic relationships of figs, we extracted 78 plastid protein-coding genes and aligned them with MUSCLE embedded in Geneious Prime 2022.0.2 program and performed Maximum Parsimony (MP), Maximum Likelihood (ML), and Bayesian Inference (BI). The MP analyses were conducted in PAUP* v4.0a with all characters treated as equally weighted and unordered [25]. Gaps were considered missing data. We executed searches of 1000 random taxon addition replicates using tree-bisection-reconnection (TBR) branch swapping, with MulTrees, allowing ten trees to be held at each step. Bootstrap analyses, labeled as parsimony bootstrap percentages (PBP), involved 1000 pseudoreplicates and were conducted to assess internal support using the same parameters. We used the IQ-TREE web server (http://iqtree.cibiv.univie.ac.at/; accessed on 23 October 2023) to make the ML searches [26]. Support value, denoted as mean bootstrap percentage (MBP), was calculated with 1000 replicates of ultrafast bootstrap. We used MEGA 11 to determine the optimal substitution model based on Bayesian Information Criterion (BIC) before conducting the Bayesian inference (BI) analysis (Table S4) [27]. Two simultaneous runs began from random trees and lasted for at least 1,000,000 generations with MrBayes v3.2.7 [28]. We sampled one tree every 1000 generations. Subsequently, 25% of the trees were discarded as burn-in samples. The remaining trees contributed to constructing a 50% majority-rule consensus tree, with the proportions of bifurcations in this consensus tree presented as posterior probability (PP) to estimate the robustness of half the BI tree. We then verified the effective sample size values (ESS) for model parameters, ensuring they were at least 200. The phylogenetic trees were edited using the FigTree v1.4.4 program [29].

2.3. Relative Synonymous Codon Usage (RSCU) Analysis

The DAMBE v7.3.11 was utilized to calculate the relative synonymous codon usage (RSCU) values for 78 protein-coding genes within plastomes [30].

2.4. Development of Molecular Markers for Ficus erecta

To identify specific SNP sites for *F. erecta*, analysis of MDCs was conducted using FastaChar v0.2.4 [18]. This examination utilized an alignment of 78 plastid protein-coding genes, comparing each species within the same genus. After confirming the SNPs specific to *F. erecta*, we conducted Sanger sequencing on individuals from different regions. As a result, three coding regions (*ndh*D, *pet*A, and *rbc*L) were selected to develop the molecular markers for *F. erecta*. Based on previous research, we designed an inner allele-specific primer to display the On/Off reaction of PCR products for *F. erecta* [31]. The PCR reactions, with a total volume of 25 µL for each primer pair, consist of a 2× premix of EmeraldAmp® GT PCR Master Mix and 30–50 ng of template DNA. The list of primer pairs, PCR mixture, and protocol is detailed in Tables 1 and S5. To verify the utility of these markers, we tested them on other species collected from different locations (Figure S1). The PCR products were examined using 1.5% agarose gel and the electrophoresis method.

Table 1. Primer sequences for ARMS PCR.

Primer	Sequence
Fic_ndhD_231F	5′-CTG GAG ATT GGG AAT AGA TGG A-3′
Fic_ndhD_899R	5′-TTA CGT TGA CCA GGA GAT GTT G-3′
Fic_ndhD_R	5′-CGA TTT GAC AGC AAA AGC GAG-3′
Fic_petA_130F	5′-TAT GAA AAT CCA CGA GAA GCG A-3′
Fic_petA_751R	5′-AAA CAA GAA GTT CTG GTC CTG G-3′
Fic_petA_F	5′-GAG GGG TTT GAA TTA GCC CTT A-3′
Fic_rbcL_5F	5′-CAC CAC AAA CAG AGA CTA AAG CA-3′
Fic_rbcL_565R	5′-CTG CTC TAC CGT AAT TCT TAG CG-3′
Fic_rbcL_F	5′-GAA TCT TCT ACT GGT ACA TGG AAA G-3′

3. Results

3.1. Plastome Features of Korean Figs

A total of 128,813 to 747,254 reads, accounting for 1.2–6.3% of the total 10,455,848 to 14,896,940 reads, were assembled (Table S2). The complete plastome map of the four figs exhibits a quadripartite structure, comprising a large single copy (LSC) region (88,397–88,641 bp), a small single copy (SSC) region (20,087–20,165 bp), and two inverted repeats (IRs) regions (25,889–25,899 bp) (Figure 1 and Table 2). There was a difference of only one base pair between the *rps*2-*rpo*C2 intergenic spacer (IGS) of *F. erecta* and *F. erecta* var. *sieboldii*. We determined that figs contain 129 genes, including one pseudogene (*inf* A), with 17 of these genes being duplicated in the IR regions. Of these genes, nine protein-coding genes (*atp*F, *ndh*A, *ndh*B, *pet*B, *pet*D, *rpl*2, *rpl*16, *rpo*C1, and *rps*16) and five tRNAs (*trn*K-UUU, *trn*G-UCC, *trn*L-UAA, *trn*I-GAU, and *trn*A-UGC) have one intron each. Meanwhile, three protein-coding genes (*clp*P1, *rps*12, and *pbf*1) possess two introns. The *rps*12 gene is trans-spliced, with its 5′ exon located in the LSC region and the 3′ exon and intron situated in the IR regions.

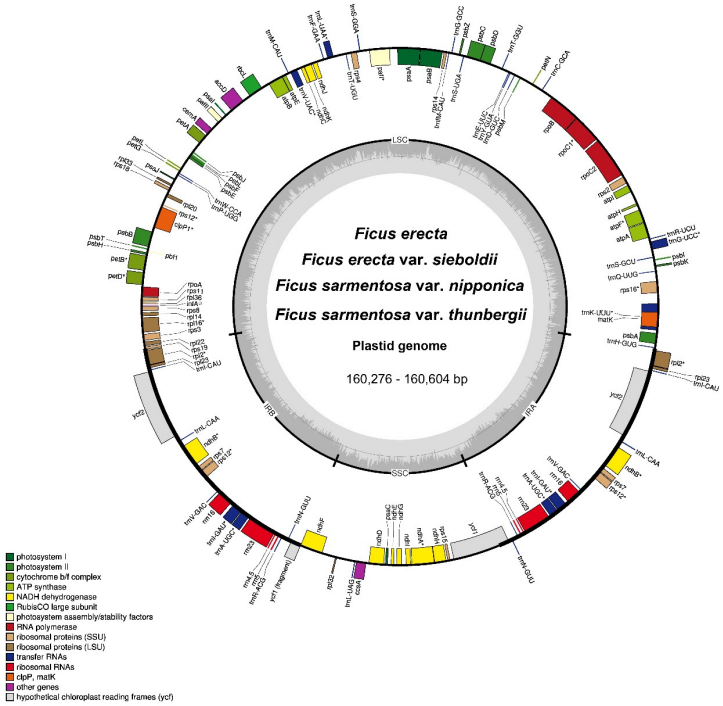

Figure 1. Complete plastome map of Korean figs and their gene contents. The colored boxes represent conserved plastid genes. Genes located inside the circle are transcribed clockwise, while those outside the circle are transcribed counterclockwise. The grey bar graphs in the inner circle indicate the GC content of the plastome.

Table 2. Features of the LSC, SSC, and IR of plastomes in this study.

Taxa	Subgenus	Length and G + C Content				GenBank Accession No.	Voucher
		LSC bp (G + C%)	SSC bp (G + C%)	IR bp (G + C%)	Total bp (G + C%)		
Ficus erecta Thunb.	*Ficus*	88,640 (33.5)	20,165 (28.9)	25,899 (42.6)	160,603 (35.9)	PP291718	JH220720001
Ficus erecta var. *sieboldii* (Miq.) King	*Ficus*	88,641 (33.5)	20,165 (28.9)	25,899 (42.6)	160,603 (35.9)	PP291717	JH220720003
Ficus sarmentosa var. *nipponica* (Franch. & Sav.) Corner	*Synoecia*	88,456 (33.6)	20,087 (29.1)	25,897 (42.6)	160,337 (36.0)	PP291719	JH190502001
Ficus sarmentosa var. *thunbergii* (Maxim.) Corner	*Synoecia*	88,397 (33.6)	20,101 (29.1)	25,889 (42.7)	160,276 (36.0)	PP291720	TH190616001

3.2. Phylogenetic Relationships with Related Taxa

We conducted MP, ML, and BI analyses and observed consistent topologies across the phylogenetic trees, which strongly supported the monophyly of figs (Figure 2). The sequence matrix encompassed 68,859 characters, with 64,761 (94.0%) being constant and 1580 (2.3%) being parsimony informative. We derived the most parsimonious tree with a tree length of 4656, consistency index (CI) of 0.919, and retention index (RI) of 0.905, as depicted in Figure 2. According to the IQ-TREE web server, the optimal model for ML was TVM + F + I + G4. The tribe Castilleae was identified as the sister tribe to Ficeae, which consists of the single genus, *Ficus*. Within Ficeae, two distinct clades emerged, indicating monophyletic relationships between the subgenus *Sycomorus* and *Urostigma*, as classified by Berg and Corner [32] (PBP = 82/MBP = 85/PP = 0.988). However, it was noted that the subgenera *Ficus* and *Synoecia* are paraphyletic (Figure 2). The phylogenetic trees further corroborate that *F. erecta* and *F. erecta* var. *sieboldii*, as well as *F. pumila* and *F. sarmentosa* var. *thunbergii*, are grouped within the same species.

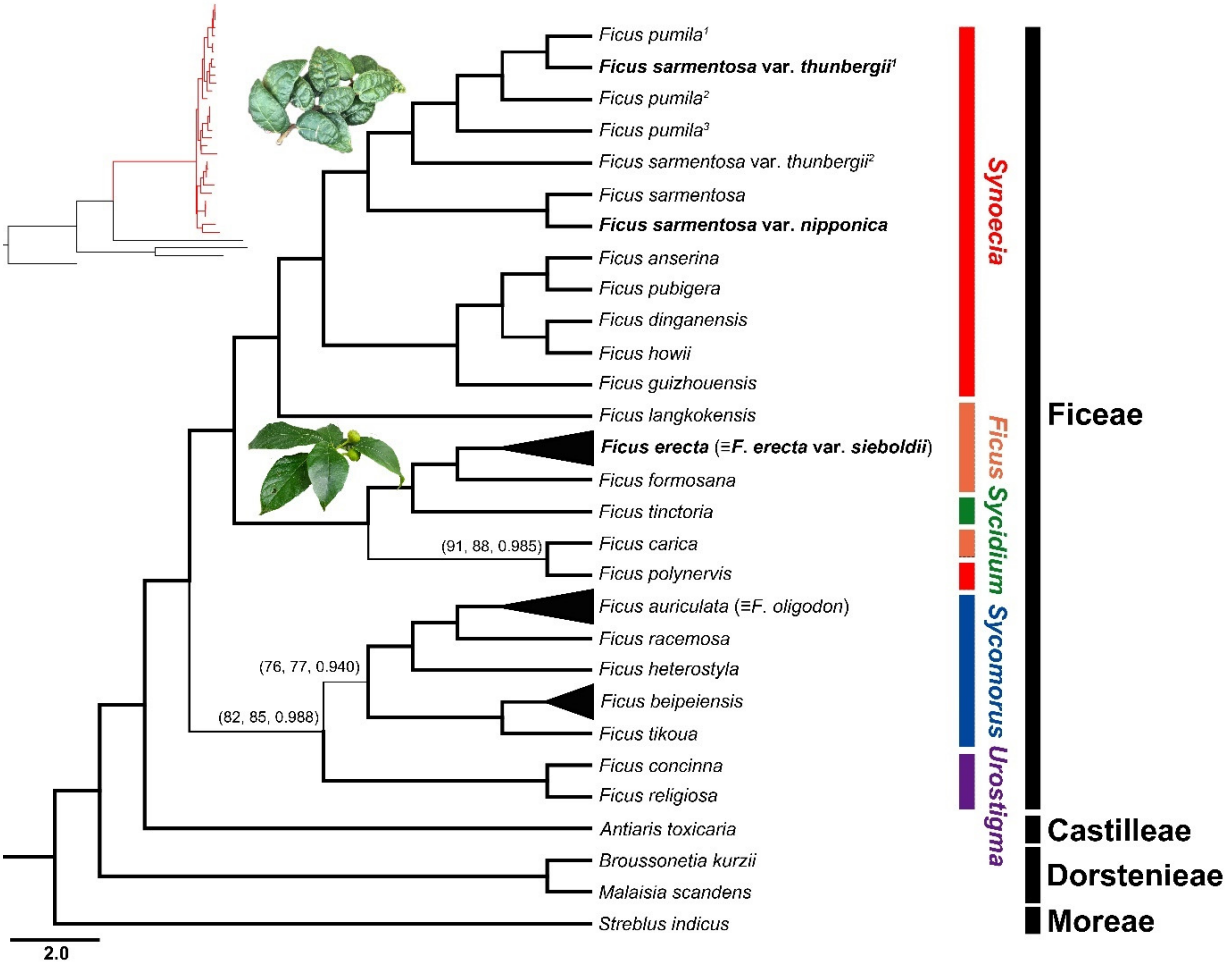

Figure 2. The Maximum Likelihood (ML) tree derived from 78 plastid protein-coding genes includes 34 taxa. Numbers indicate support values, represented as parsimony bootstrap percentages (PBP)/mean bootstrap percentages (MBP)/posterior probability (PP). Only support values with PBP $\leq$ 90%, MBP $\leq$ 95%, and PP $\leq$ 0.95 are displayed. The bold names indicate genomes obtained in this study, while the colored boxes represent the subgenera within the genus *Ficus* from Berg and Corner [32].

3.3. RSCU

A total of 78 genes coding for proteins were analyzed across three varieties of figs to assess their usage of relatively synonymous codons (RSCU), omitting the stop codons UAA, UAG, and UGA from the study (Figure 3). The codon count varied slightly among the species, with *F. erecta* having 23,472 codons and *F. sarmentosa* var. *thunbergii* slightly higher at 23,488 (Table S6). Among the amino acids, leucine (L) emerged as the most prevalent, making up 10.75% of the total, while cysteine (C) was the least, at just 1.10%.

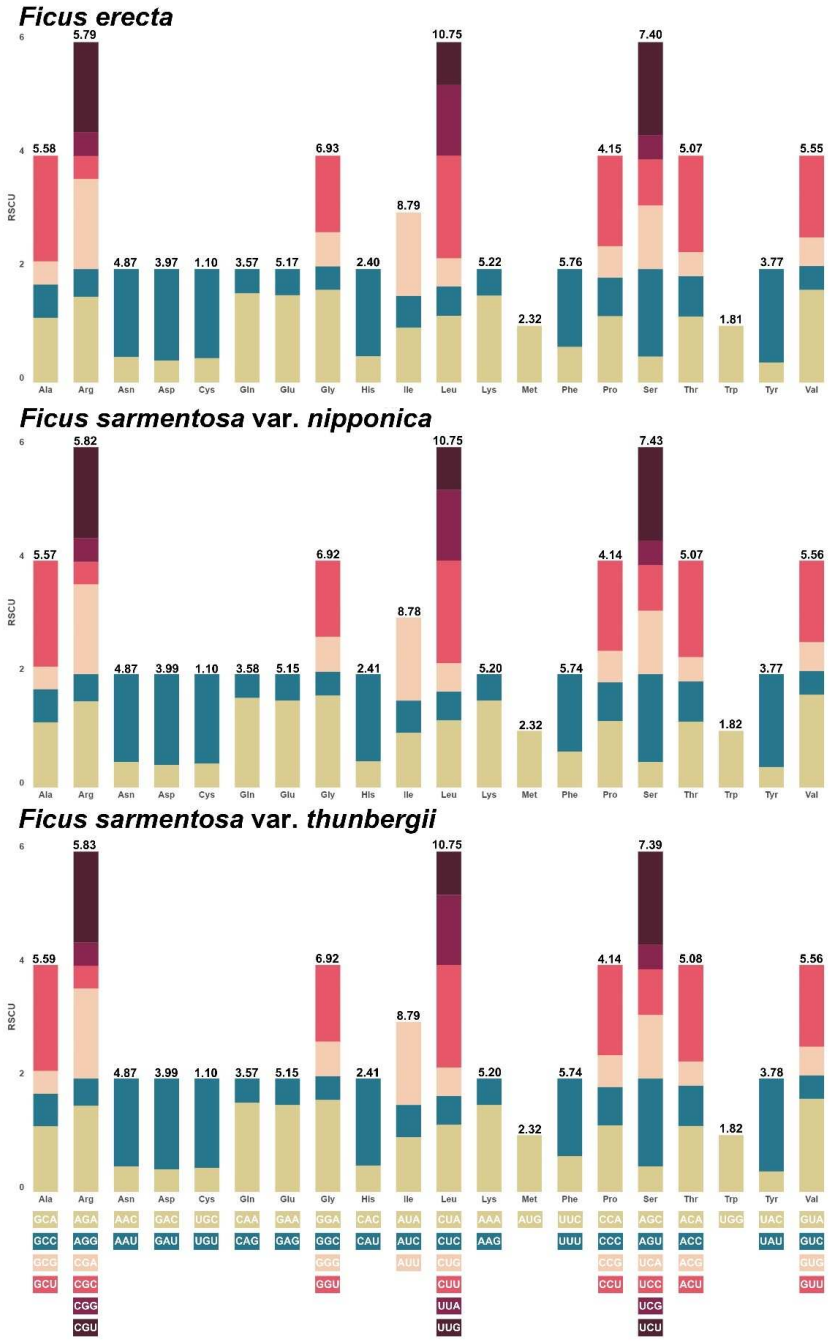

Figure 3. Relative synonymous codon usage (RSCU) analysis of 20 amino acids in protein-coding genes of the complete plastomes of Korean figs. The values at the top of each stack indicate the frequency of use of each amino acid.

3.4. MDCs and SNP Markers for Ficus erecta

Based on our alignment data of 78 plastid protein-coding genes, we identified 467 sites as MDCs for the genus *Ficus* when compared with other groups in the Moraceae, such as *Antiaris*, *Broussonetia*, *Malaisia*, and *Streblus*. Additionally, we identified MDCs for each species by comparing them with other figs, varied from 0 to 175 (Figure 4). Four species, *F. concinna*, *F. polynervis*, *F. racemosa*, and *F. religiosa*, have more than 80 MDCs; however, two species, *F. oligodon* and *F. tikoua*, exhibit 0 MDCs indicating the absence of specific SNP sites in the plastid protein-coding genes (Figure 4). A total of 33 sites were identified as MDCs for *F. erecta* (including *F. erecta* var. *sieboldii*). Among these, 24 were deletions in the *rps*11 gene, which codes for the eight amino acids 'RKNRHMSS' in other figs. In the remaining nine MDCs, six (located in the *atp*I, *ndh*D, *pet*A, *rbc*L, *rpo*C2, and *ycf*1 genes) induced alterations in translation, leading to changes in the protein structure depending on whether the substitutions were conservative or non-conservative. The specific primer pairs designed for *F. erecta* (which includes *F. erecta* var. *sieboldii*) based on *ndh*D, *pet*A, and *rbc*L yielded bands of 445, 412, and 388 bp, respectively (Figure 5). For *F. carica*, *F. sarmentosa* var. *nipponica*, and *F. sarmentosa* var. *thunbergii*, no bands were observed, with the exception of control bands measuring 669, 622, and 561 bp from *ndh*D, *pet*A, and *rbc*L, respectively. These primer pairs proved effective for other species examined from various locations in Korea (Figure S1).

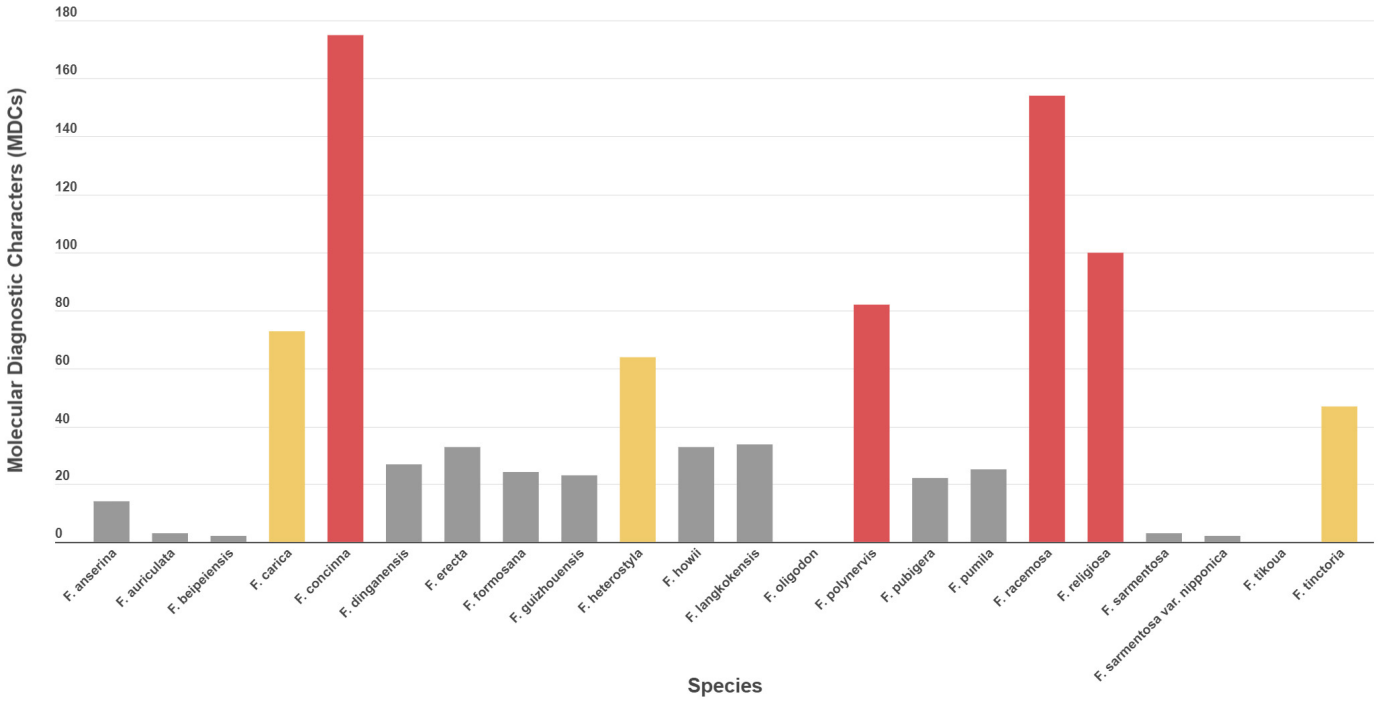

Figure 4. Quantitative analysis of MDCs in 78 plastid protein-coding genes across 30 individuals of figs. The red bar represents more than 80 MDCs; the yellow bar shows 40 to 80 MDCs; and the grey bar indicates fewer than 40 MDCs.

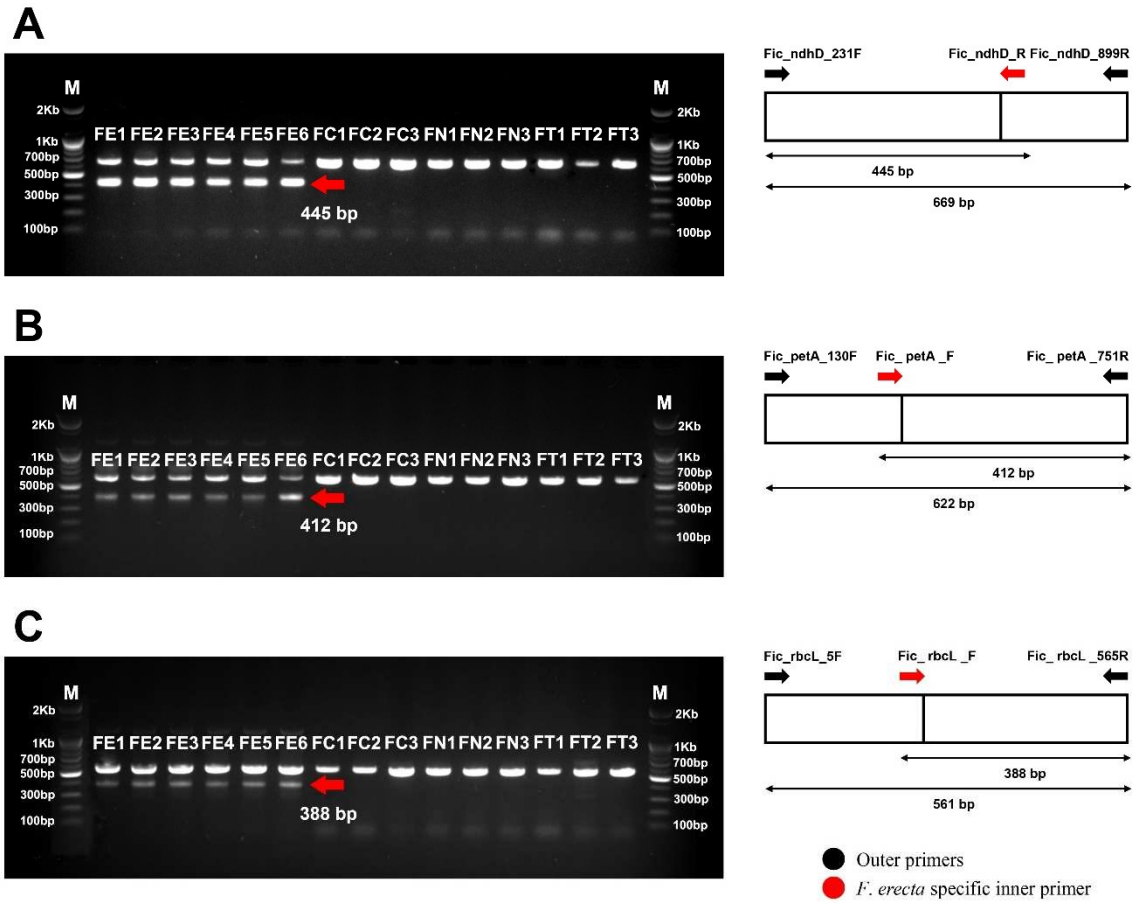

Figure 5. Multiplex PCR results based on specific primer pairs for *F. erecta* and the design of primer pairs: (**A**) Combination of Fic_ndhD_231F, Fic_ndhD_899R, and Fic_ndhD_R; (**B**) combination of Fic_petA_130F, Fic_petA_751R, and Fic_petA_F; (**C**) combination of Fic_rbcL_5F, Fic_rbcL_565R, and Fic_rbcL_F; FE1-6: *F. erecta*; FC1-3: *F. carica*; FN1-3: *F. sarmentosa* var. *nipponica*; FT1-3: *F. sarmentosa* var. *thunbergii*.

4. Discussion

4.1. Complete Plastomes of Figs and Phylogenetic Relationships

In this study, we completed plastome sequences of Korean figs, noted for their highly conserved characteristics. Aligning with recent research findings, we annotated the *ycf*3 and *ycf*4 genes, which are crucial for assembling the photosystem I complex, as *paf*I and *paf*II, respectively. Furthermore, we annotated the *psb*N gene, responsible for encoding the photosystem biogenesis factor 1, as *pbf*1. It was observed that the *inf*A gene, which encodes translation initiation factor 1, underwent pseudogenization, a common occurrence in many members of Moraceae [33–36]. Utilizing 78 plastid protein-coding genes, we achieved consistent topologies across MP, ML, and BI trees, supported by high bootstrap values and posterior probability (Figure 2). This phylogenetic tree was similar to those in previous molecular studies that deviated from the traditional classification of figs based on morphological characteristics [33,36–43]. In the first plastid phylogenomic study, only subgenus *Synoecia* was shown to be monophyletic, while the study also highlighted discordant relationships based on comparisons of plastid and nuclear data [40]. Based on involucral bracts, inflorescence morphology, and pollination syndromes with robust sampling, the Involucrata was newly introduced, containing *Ficus* and Castilleae [41]. The classification of figs was also revised into 19 clades to better recognize their evolutionary history [41]. However, this revised classification, despite contributing to ecological and in-depth morphological studies of figs, reveals non-monophyly in the present study when the proposed clades are substituted (not shown). Although the morphological classification

of figs is still ambiguous, it has been suggested that genera of pollinating fig wasps show a high degree of co-evolution [33,39,44]. To better understand the relationships among figs, we need further genomic data, encompassing not only plastidial but also nuclear genes, coupled with an improved phylogeny of the wasps.

Given their identical sequences, *F. erecta* var. *sieboldii* is nested within *F. erecta*. Traditionally distinguished by its more slender, lanceolate leaves and serrated young sprouts, it has recently been classified as a synonym for *F. erecta* by Plants of the World Online (POWO), based on a previous study [45]. Our plastome sequences and phylogenetic analyses support the recognition of *F. erecta* and its variety as the same species. Additionally, among the individuals of *F. pumila* and *F. sarmentosa* var. *thunbergii*, we identified 58 SNPs out of 68,399 in the protein-coding sequences, grouping them together. The sequencing of our newly collected *F. sarmentosa* var. *thunbergii*, embedded within *F. pumila* as indicated by a previous study, calls for a taxonomic revision [36]. Moreover, interspecific hybrids have been reported for both species, yet the evolutionary significance of these hybrids is still not clearly understood [46].

4.2. RSCU

Examining codon usage within plastid protein-coding genes can reveal valuable information about mutation trends, selection influences, and genetic drift at the species level. Generally, most amino acids are encoded by two to six synonymous codons, except for methionine (M) and tryptophan (W). Codons that have RSCU values above one are indicative of higher usage frequency, whereas those with values below one are less frequently used [47]. Our study found that 29 codons showed elevated RSCU values (greater than one), predominantly ending in A or U. In contrast, codons with lower RSCU values (less than one) often ended in G or C (Table S6). It was particularly noteworthy that the codon AUU, encoding for isoleucine (Ile), was the most commonly used, aligning with findings in other figs [36].

4.3. Implications of MDCs Data and Molecular Markers for Ficus erecta

In recent years, DNA barcoding has become a common tool for herb identification, enhancing safety in the herbal medicine industry. Beyond the internal transcribed spacer (ITS), partial plastid, and nuclear genes, genomic research has rapidly advanced super-barcoding, making species identification more accessible. The molecular characters provide distinct MDCs found in DNA sequences, which can be utilized for describing species. Although our study included only a limited selection of Moraceae members to identify specific sites for *Ficus*, the number of MDCs in plastid protein-coding genes for *Ficus* will offer fundamental super-barcoding information. Based on analyses of MDCs, we identified 24 deletions in the *rps*11 gene of *F. erecta*, accounting for 5.75% of its sequence, which encodes ribosomal protein S11. To assess the presence of an open reading frame (ORF) with a conserved domain, we utilized the NCBI Conserved Domains Database (CDD) [48]. Our analysis confirmed that the gene retains an intact functional domain, indicating no issues with its functionality.

A variety of DNA-based markers, such as SNPs, Inter Simple Sequence Repeats (ISSRs), RAPD, and Sequence Characterization of Amplified Regions (SCARs), has been developed for verifying the authenticity of medicinal plants, thus ensuring their safety and effectiveness [49–51]. Among these, SNPs have been particularly valuable for population genetics and phylogenetic studies, especially in angiosperms, as evidenced by numerous studies [52,53]. Additionally, a range of molecular markers, specifically designed for different species, has been developed from SNPs in plastome sequences [7,54,55]. In our research, we successfully created molecular markers based on MDC data that effectively differentiate *F. erecta* from its relatives *F. carica*, *F. sarmentosa* var. *nipponica*, and *F. sarmentosa* var. *thunbergii* (Figures 5 and S1). These markers are based on SNPs that are altered in translation, leading to changes in protein structure. Consequently, these markers are expected to significantly contribute to precise species distinction. Furthermore, these

markers have demonstrated that variations in nucleotide sequences can provide a rapid method for the molecular identification of *F. erecta*. Our focus on *F. erecta* is due to its medicinal effects, as discussed in recent studies [5,6]. Despite the limited and geographically confined number of samples, we designed primers that incorporate the published plastome sequences of Japanese *F. erecta* (GenBank accession No. MT093220) and relative figs. This approach ensures that these markers can be applied on a global scale and increase accuracy. While our study primarily utilized the SNP in the three protein-coding regions, exploring the complete plastome sequences of figs could enable the discovery of additional SNPs, aiding in the creation of more comprehensive molecular markers for *F. erecta* and related species.

5. Conclusions

This study provides vital insights into the plastomes of Korean figs. We conducted a comparison of the plastome sequences between *F. erecta* and *F. erecta* var. *sieboldii*, confirming them as the same species. Additionally, our research outlines the phylogenetic relationships among Korean figs and identifies the non-monophyly within the subgenera. Our genomic research provides foundational super-barcoding information for *Ficus*, a genus of significant economic and medicinal value. The methods and protocols developed in this study for generating molecular markers based on SNPs for *F. erecta* are straightforward and have the potential to serve as a standard approach for similar research across other Moraceae species.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/d16030129/s1, Figure S1: The PCR results of specific primer pairs for *Ficus erecta*; Table S1: List of sampling taxa in this study; Table S2: List of sampling taxa and plastome assembly information; Table S3: List of species used for phylogenomic analyses; Table S4: Maximum Likelihood fits of 24 different nucleotide substitution models; Table S5: The composition of PCR mixture and the PCR protocol for testing molecular markers; Table S6: Relative synonymous codon usage values of Korean figs.

Author Contributions: Conceptualization, J.-H.K.; Methodology, J.J., T.-H.K. and H.J.P.; Validation, H.J.P. and I.S.C.; Formal analysis, T.-H.K. and S.W.K.; Investigation, J.J.; Resources, J.J., T.-H.K. and S.W.K.; Data curation, J.J., T.-H.K. and S.W.K.; Writing—original draft, J.J. and T.-H.K.; Writing—review and editing, S.W.K., H.J.P., I.S.C. and J.-H.K.; Visualization, J.J. and T.-H.K.; Supervision, T.-H.K.; Project administration, H.J.P., I.S.C. and J.-H.K.; Funding acquisition, I.S.C. and J.-H.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry (IPET) through Technology Commercialization Support Program, funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA) (122046-3).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The genome sequence data that support the findings of this study are openly available in the GenBank of NCBI at https://www.ncbi.nlm.nih.gov (Registered Date on 5 Feburary 2024) under accession nos. PP291717 (*F. erecta* var. *sieboldii*), PP291718 (*F. erecta*), PP291719 (*F. sarmentosa* var. *nipponica*), and PP291720 (*F. sarmentosa* var. *thunbergii*).

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

References

1. Clement, W.L.; Weiblen, G.D. Morphological Evolution in the Mulberry Family (Moraceae). *Syst. Bot.* **2009**, *34*, 530–552. [CrossRef]
2. Harrison, R.D.; Shanahan, M. Seventy-seven ways to be a fig: Overview of a diverse plant assemblage. In *Pollination Ecology and the Rain Forest*; Springer: New York, NY, USA, 2005; pp. 111–127.
3. Park, S.H.; Oh, T.-H.; Kim, S.S.; Kim, J.E.; Lee, S.J.; Lee, N.H. Constituents with tyrosinase inhibitory activities from branches of *Ficus erecta* var. *sieboldii King. J. Enzym. Inhib. Med. Chem.* **2012**, *27*, 390–394. [CrossRef]
4. Al Faysal, A.; Mian, Y.; Rahman, M.M.; Akter, M.; Rahman, M.; Tuhin, T.H.; Bilkiss, M. In vitro thrombolytic activity, antioxidant and cytotoxic properties of fruit extracts of *Ficus erecta* (Thunb.). *J. Med. Plants Res.* **2018**, *12*, 50–54.

5. Sohn, E.; Kim, Y.J.; Kim, J.H.; Jeong, S.J. *Ficus erecta* Thunb Leaves Alleviate Memory Loss Induced by Scopolamine in Mice via Regulation of Oxidative Stress and Cholinergic System. *Mol. Neurobiol.* **2021**, *58*, 3665–3676. [CrossRef]
6. Sohn, E.; Kim, Y.J.; Kim, J.-H.; Jeong, S.-J. *Ficus erecta* Thunb. Leaves Ameliorate Cognitive Deficit and Neuronal Damage in a Mouse Model of Amyloid-β-Induced Alzheimer's Disease. *Front. Pharmacol.* **2021**, *12*, 607403. [CrossRef] [PubMed]
7. Do, H.D.K.; Jung, J.; Hyun, J.; Yoon, S.J.; Lim, C.; Park, K.; Kim, J.H. The newly developed single nucleotide polymorphism (SNP) markers for a potentially medicinal plant, *Crepidiastrum denticulatum* (Asteraceae), inferred from complete chloroplast genome data. *Mol. Biol. Rep.* **2019**, *46*, 3287–3297. [CrossRef] [PubMed]
8. Dobrogojski, J.; Adamiec, M.; Luciński, R. The chloroplast genome: A review. *Acta Physiol. Plant.* **2020**, *42*, 98. [CrossRef]
9. Jung, J.; Kim, C.; Kim, J.H. Insights into phylogenetic relationships and genome evolution of subfamily Commelinoideae (Commelinaceae Mirb.) inferred from complete chloroplast genomes. *BMC Genom.* **2021**, *22*, 231. [CrossRef] [PubMed]
10. Kim, T.-H.; Kim, J.-H. Molecular Phylogeny and Historical Biogeography of *Goodyera* R. Br.(Orchidaceae): A Case of the Vicariance between East Asia and North America. *Front. Plant Sci.* **2022**, *13*, 1255. [CrossRef]
11. Yang, H.-B.; Kang, W.-H.; Nahm, S.-H.; Kang, B.-C. Methods for developing molecular markers. In *Current Technologies in Plant Molecular Breeding: A Guide Book of Plant Molecular Breeding for Researchers*; Springer: Dordrecht, The Netherlands, 2015; pp. 15–50.
12. Williams, J.G.; Kubelik, A.R.; Livak, K.J.; Rafalski, J.A.; Tingey, S.V. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* **1990**, *18*, 6531–6535. [CrossRef]
13. Vos, P.; Hogers, R.; Bleeker, M.; Reijans, M.; Lee, T.v.d.; Hornes, M.; Friters, A.; Pot, J.; Paleman, J.; Kuiper, M. AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Res.* **1995**, *23*, 4407–4414. [CrossRef]
14. Litt, M.; Luty, J.A. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* **1989**, *44*, 397. [PubMed]
15. Jacob, H.J.; Lindpaintner, K.; Lincoln, S.E.; Kusumi, K.; Bunker, R.K.; Mao, Y.-P.; Ganten, D.; Dzau, V.J.; Lander, E.S. Genetic mapping of a gene causing hypertension in the stroke-prone spontaneously hypertensive rat. *Cell* **1991**, *67*, 213–224. [CrossRef] [PubMed]
16. Dwiningsih, Y.; Rahmaningsih, M.; Alkahtani, J. Development of single nucleotide polymorphism (SNP) markers in tropical crops. *Adv. Sustain. Sci. Eng. Technol.* **2020**, *2*, 343558. [CrossRef]
17. Yoo, E.; Haile, M.; Ko, H.-C.; Choi, Y.-M.; Cho, G.-T.; Woo, H.-J.; Wang, X.; Sung, P.; Lee, J.; Lee, J. Development of SNP markers for *Cucurbita* species discrimination. *Sci. Hortic.* **2023**, *318*, 112089. [CrossRef]
18. Merckelbach, L.M.; Borges, L.M. Make every species count: Fastachar software for rapid determination of molecular diagnostic characters to describe species. *Mol. Ecol. Resour.* **2020**, *20*, 1761–1768. [CrossRef]
19. Doyle, J.J.; Doyle, J.L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **1987**, *19*, 11–15.
20. Jin, J.-J.; Yu, W.-B.; Yang, J.-B.; Song, Y.; DePamphilis, C.W.; Yi, T.-S.; Li, D.-Z. GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* **2020**, *21*, 241. [CrossRef]
21. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [CrossRef]
22. Tillich, M.; Lehwark, P.; Pellizzer, T.; Ulbricht-Jones, E.S.; Fischer, A.; Bock, R.; Greiner, S. GeSeq–versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **2017**, *45*, W6–W11. [CrossRef]
23. Chan, P.P.; Lowe, T.M. tRNAscan-SE: Searching for tRNA genes in genomic sequences. In *Gene Prediction: Methods and Protocols*; Humana: New York, NY, USA, 2019; pp. 1–14.
24. Greiner, S.; Lehwark, P.; Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.3. 1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* **2019**, *47*, W59–W64. [CrossRef]
25. Swofford, D. PAUP* 4.0 b. 4a. In *Phylogenetic Analysis Using Parsimony*(and Other Methods)*; Sinauer: Sunderland, MA, USA, 2000.
26. Trifinopoulos, J.; Nguyen, L.-T.; von Haeseler, A.; Minh, B.Q. W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **2016**, *44*, W232–W235. [CrossRef] [PubMed]
27. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027. [CrossRef] [PubMed]
28. Ronquist, F.; Teslenko, M.; Van Der Mark, P.; Ayres, D.L.; Darling, A.; Höhna, S.; Larget, B.; Liu, L.; Suchard, M.A.; Huelsenbeck, J.P. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **2012**, *61*, 539–542. [CrossRef] [PubMed]
29. Rambaut, A. *FigTree, a Graphical Viewer of Phylogenetic Trees (Version 1.4.4)*; Institute of evolutionary biology, University of Edinburgh: Edinburgh, UK, 2018.
30. Xia, X.; Xie, Z. DAMBE: Software package for data analysis in molecular biology and evolution. *J. Hered.* **2001**, *92*, 371–373. [CrossRef]
31. You, F.M.; Huo, N.; Gu, Y.Q.; Luo, M.-C.; Ma, Y.; Hane, D.; Lazo, G.R.; Dvorak, J.; Anderson, O.D. BatchPrimer3: A high throughput web application for PCR and sequencing primer design. *BMC Bioinform.* **2008**, *9*, 253. [CrossRef]
32. Berg, C.C.; Corner, E.J.H. Moraceae: Ficeae. *Flora Malesiana-Ser. 1 Spermatophyta* **2005**, *17*, 1–702.
33. Zhang, Z.-R.; Yang, X.; Li, W.-Y.; Peng, Y.-Q.; Gao, J. Comparative chloroplast genome analysis of *Ficus* (Moraceae): Insight into adaptive evolution and mutational hotspot regions. *Front. Plant Sci.* **2022**, *13*, 965335. [CrossRef]

34. Ravi, V.; Khurana, J.P.; Tyagi, A.K.; Khurana, P. The chloroplast genome of mulberry: Complete nucleotide sequence, gene organization and comparative analysis. *Tree Genet. Genomes* **2006**, *3*, 49–59. [CrossRef]
35. Yang, J.; Chu, Q.; Meng, G.; Kong, W. The complete chloroplast genome sequences of three *Broussonetia* species and comparative analysis within the Moraceae. *PeerJ* **2022**, *10*, e14293. [CrossRef] [PubMed]
36. Zhang, Z.; Zhang, D.-S.; Zou, L.; Yao, C.-Y. Comparison of chloroplast genomes and phylogenomics in the *Ficus sarmentosa* complex (Moraceae). *PLoS ONE* **2022**, *17*, e0279849. [CrossRef]
37. Azuma, H.; Harrison, R.D.; Nakamura, K.; Su, Z.-H. Molecular phylogenies of figs and fig-pollinating wasps in the Ryukyu and Bonin (Ogasawara) islands, Japan. *Genes Genet. Syst.* **2010**, *85*, 177–192. [CrossRef]
38. Li, H.-Q.; Wang, S.; Chen, J.-Y.; Gui, P. Molecular phylogeny of *Ficus* section *Ficus* in China based on four DNA regions. *J. Syst. Evol.* **2012**, *50*, 422–432. [CrossRef]
39. Cruaud, A.; Ronsted, N.; Chantarasuwan, B.; Chou, L.S.; Clement, W.L.; Couloux, A.; Cousins, B.; Genson, G.; Harrison, R.D.; Hanson, P.E.; et al. An extreme case of plant-insect codiversification: Figs and fig-pollinating wasps. *Syst. Biol.* **2012**, *61*, 1029–1047. [CrossRef]
40. Bruun-Lund, S.; Clement, W.L.; Kjellberg, F.; Ronsted, N. First plastid phylogenomic study reveals potential cyto-nuclear discordance in the evolutionary history of *Ficus* L. (Moraceae). *Mol. Phylogenet. Evol.* **2017**, *109*, 93–104. [CrossRef]
41. Clement, W.L.; Bruun-Lund, S.; Cohen, A.; Kjellberg, F.; Weiblen, G.D.; Rønsted, N. Evolution and classification of figs (*Ficus*, Moraceae) and their close relatives (Castilleae) united by involucral bracts. *Bot. J. Linn. Soc.* **2020**, *193*, 316–339. [CrossRef]
42. Wang, G.; Zhang, X.; Herre, E.A.; McKey, D.; Machado, C.A.; Yu, W.B.; Cannon, C.H.; Arnold, M.L.; Pereira, R.A.S.; Ming, R.; et al. Genomic evidence of prevalent hybridization throughout the evolutionary history of the fig-wasp pollination mutualism. *Nat. Commun.* **2021**, *12*, 718. [CrossRef] [PubMed]
43. Huang, Y.; Li, J.; Yang, Z.; An, W.; Xie, C.; Liu, S.; Zheng, X. Comprehensive analysis of complete chloroplast genome and phylogenetic aspects of ten *Ficus* species. *BMC Plant Biol.* **2022**, *22*, 253. [CrossRef] [PubMed]
44. Machado, C.A.; Robbins, N.; Gilbert, M.T.P.; Herre, E.A. Critical review of host specificity and its coevolutionary implications in the fig/fig-wasp mutualism. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6558–6565. [CrossRef]
45. Govaerts, R.; Nic Lughadha, E.; Black, N.; Turner, R.; Paton, A. The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. *Sci. Data* **2021**, *8*, 215. [CrossRef] [PubMed]
46. Tsai, L.; Hayakawa, H.; Fukuda, T.; Yokoyama, J. A breakdown of obligate mutualism on a small island: An interspecific hybridization between closely related fig species (*Ficus pumila* and *Ficus thunbergii*) in Western Japan. *Am. J. Plant Sci.* **2015**, *6*, 126. [CrossRef]
47. Sharp, P.M.; Tuohy, T.M.; Mosurski, K.R. Codon usage in yeast: Cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res.* **1986**, *14*, 5125–5143. [CrossRef] [PubMed]
48. Wang, J.; Chitsaz, F.; Derbyshire, M.K.; Gonzales, N.R.; Gwadz, M.; Lu, S.; Marchler, G.H.; Song, J.S.; Thanki, N.; Yamashita, R.A. The conserved domain database in 2023. *Nucleic Acids Res.* **2023**, *51*, D384–D388. [CrossRef] [PubMed]
49. Bi, D.; Chen, D.; Khayatnezhad, M.; Hashjin, Z.S.; Li, Z.; Ma, Y. Molecular identification and genetic diversity in *Hypericum* L.: A high value medicinal plant using RAPD markers markers. *Genetika* **2021**, *53*, 393–405. [CrossRef]
50. Zheng, K.; Cai, Y.; Chen, W.; Gao, Y.; Jin, J.; Wang, H.; Feng, S.; Lu, J. Development, identification, and application of a germplasm specific SCAR Marker for *Dendrobium officinale* Kimura et Migo. *Front. Plant Sci.* **2021**, *12*, 669458. [CrossRef] [PubMed]
51. Kumar, A.; Mishra, P.; Singh, S.C.; Sundaresan, V. Efficiency of ISSR and RAPD markers in genetic divergence analysis and conservation management of *Justicia adhatoda* L., a medicinal plant. *Plant Syst. Evol.* **2014**, *300*, 1409–1420. [CrossRef]
52. Rendón-Anaya, M.; Ibarra-Laclette, E.; Méndez-Bravo, A.; Lan, T.; Zheng, C.; Carretero-Paulet, L.; Perez-Torres, C.A.; Chacón-López, A.; Hernandez-Guzmán, G.; Chang, T.-H. The avocado genome informs deep angiosperm phylogeny, highlights introgressive hybridization, and reveals pathogen-influenced gene space adaptation. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 17081–17089. [CrossRef]
53. Slimp, M.; Williams, L.D.; Hale, H.; Johnson, M.G. On the potential of Angiosperms353 for population genomic studies. *Appl. Plant Sci.* **2021**, *9*, e11419. [CrossRef]
54. Giang, V.N.L.; Waminal, N.E.; Park, H.-S.; Kim, N.-H.; Jang, W.; Lee, J.; Yang, T.-J. Comprehensive comparative analysis of chloroplast genomes from seven *Panax* species and development of an authentication system based on species-unique single nucleotide polymorphism markers. *J. Ginseng Res.* **2020**, *44*, 135–144.
55. Hyun, J.; Do, H.D.K.; Jung, J.; Kim, J.-H. Development of molecular markers for invasive alien plants in Korea: A case study of a noxious weed, *Cenchrus longispinus* (Hack.) Fernald, based on next generation sequencing data. *PeerJ* **2019**, *7*, e7965. [CrossRef]