Article

The Complete Plastid Genome of *Artocarpus camansi*: A High Degree of Conservation of the Plastome Structure in the Family Moraceae

Ueric José Borges de Souza ¹ , Luciana Cristina Vitorino ^{2,*} , Layara Alexandre Bessa ²
and Fabiano Guimarães Silva ²

¹ Graduate Program in the Biodiversity and Biotechnology of the Legal Amazon Region–BIONORTE, Federal University of Tocantins, UFT, Avenue NS-15, Quadra 109, Plano Diretor Norte, Palmas 77001-090, Tocantins, Brazil; uericjose@gmail.com

² Laboratory of Plant Mineral Nutrition, Instituto Federal Goiano campus Rio Verde, Highway Sul Goiana, Km 01, Rio Verde 75901-970, Goiás, Brazil; layara.bessa@ifgoiano.edu.br (L.A.B.); fabiano.silva@ifgoiano.edu.br (F.G.S.)

* Correspondence: luciana.vitorino@ifgoiano.edu.br; Tel.: +55-64-3620-5600

Received: 6 October 2020; Accepted: 4 November 2020; Published: 8 November 2020

Abstract: Understanding the plastid genome is extremely important for the interpretation of the genetic mechanisms associated with essential physiological and metabolic functions, the identification of possible marker regions for phylogenetic or phylogeographic analyses, and the elucidation of the modes through which natural selection operates in different regions of this genome. In the present study, we assembled the plastid genome of *Artocarpus camansi*, compared its repetitive structures with *Artocarpus heterophyllus*, and searched for evidence of synteny within the family Moraceae. We also constructed a phylogeny based on 56 chloroplast genes to assess the relationships among three families of the order Rosales, that is, the Moraceae, Rhamnaceae, and Cannabaceae. The plastid genome of *A. camansi* has 160,096 bp, and presents the typical circular quadripartite structure of the Angiosperms, comprising a large single copy (LSC) of 88,745 bp and a small single copy (SSC) of 19,883 bp, separated by a pair of inverted repeat (IR) regions each with a length of 25,734 bp. The total GC content was 36.0%, which is very similar to *Artocarpus heterophyllus* (36.1%) and other moraceous species. A total of 23,068 codons and 80 SSRs were identified in the *A. camansi* plastid genome, with the majority of the SSRs being mononucleotide (70.0%). A total of 50 repeat structures were observed in the *A. camansi* plastid genome, in contrast with 61 repeats in *A. heterophyllus*. A purifying selection signal was found in 70 of the 79 protein-coding genes, indicating that they have all been highly conserved throughout the evolutionary history of the genus. The comparative analysis of the structural characteristics of the chloroplast among different moraceous species found a high degree of similarity in the sequences, which indicates a highly conserved evolutionary model in these plastid genomes. The phylogenetic analysis also recovered a high degree of similarity between the chloroplast genes of *A. camansi* and *A. heterophyllus*, and reconfirmed the hypothesis of the intense conservation of the plastome in the family Moraceae.

Keywords: Artocarpeae; purifying selection; plastid genome; plastome; phylogenetic relationships

1. Introduction

The chloroplast, which has an independent circular genome, is an essential organelle in higher plants and plays a crucial role in the processes of photosynthesis and carbon fixation [1,2]. The plastid genome (cpDNA) of the angiosperms is highly conserved in the structure, order, and composition of its genes in comparison with the nuclear and mitochondrial genomes [3,4]. This, together with its maternal

inheritance, slow evolutionary rate, and its non-recombinant characteristics in most angiosperms, makes the plant plastid genome highly suitable for the investigation of phylogeographic patterns, both within and among populations, and for inferring evolutionary and phylogenetic relationships among taxa [1,5,6].

Typically, the plastome exhibits a quadripartite structure with two copies of an inverted repeat (IR) region separated by one large single-copy (LSC) and one small single-copy (SSC) region [7]. In general, the plastid genomes of land plants range in size from 120 kb to 160 kb [8], but can diverge considerably both within and among families. In the family Orobanchaceae, plastid genomes vary in size from 45,673 bp in *Conopholis americana* (L.) Wallr. [9] (NC_023131.1) to 190,233 bp in *Striga forbesii* Benth. [10] (MF780873.1). This variation in size is usually the result of the contraction and expansion of the inverted repeats (IRs), the independent loss of one IR region, or oscillations in the length of the intergenic spacers [8,9,11]. The plastid genomes quantified to date in the Moraceae range from 158,459 bp in *Morus mongolica* (Bureau) C.K.Schneid. [12] (NC_025772.2) to 162,594 bp in *Broussonetia luzonica* (Blanco) Bureau, 1873 (NC_047180.1; Unpublished).

Most angiosperm plastid genomes contain 70–90 protein coding genes that are involved in the photosynthesis process (such as photosystem I (PSI), photosystem II (PSII), ATP synthase and the cytochrome b6/f complex, the NADH dehydrogenase subunits, and the RuBisCo large subunit), transcription, and translation. The plastome also encodes approximately 30 transfer RNA (tRNA) genes and four ribosomal RNA (rRNA) genes [8,13,14]. The non-coding regions of the plastid genome of land plants vary considerably and include important regulatory sequences, while the introns are usually well conserved [1,15]. However, the loss of introns in protein-coding genes has been reported in *Bambusa oldhamii* [16], *Cicer arietinum* [17], *Dendrocalamus latiflorus* [16], *Hordeum vulgare* [18], and *Manihot esculenta* [19]. Genes with introns found in the plastid genome have a range of functions, including the coding of the Clp protease system (*clpP*), ATP synthase (*atpF*), RNA polymerase (*rpoC2*), and ribosomal proteins (*rps12*, *rps16*, and *rpl2*) [1,15].

The first complete plastid genomes, of *Nicotiana tabacum* [20] and *Marchantia polymorpha* [21], were sequenced in 1986. With the advent of next-generation sequencing technologies (NGS), the field of chloroplast genetics and genomics has expanded dramatically in recent years. Nowadays, investigators can use a range of bioinformatic tools to distinguish plastid reads from nuclear and mitochondrial reads, to assemble the plastid genome [22]. At the present time, approximately 4369 plant plastid genomes have been deposited as RefSeq in the NCBI Organelle Genome database (July 2020), although only 14 of these species belong to the mulberry family (Moraceae).

The Moraceae, a family of the rose order (Rosales), consists of approximately 39 genera and 1100 species distributed widely throughout tropical and temperate regions of the world [23–25]. In the most recent phylogenetic analysis of the family, Zerega and Gardner [23] recognized seven tribes (Artocarpeae, Castilleae, Ficeae, Dorstenieae, Maclureae, Moreae, and Parartocarpeae) based on the sequencing of 333 nuclear genes using target enrichment via hybridization (hybseq). *Artocarpus* J.R. Forster and G. Forster is the most diverse genus of the tribe Artocarpeae and the third largest moraceous genus, with approximately 70 species [24,25]. Several species of *Artocarpus* are important food sources for forest-dwelling animals, and a dozen species are important crops in the regions in which they occur, including the jackfruit (*Artocarpus heterophyllus* Lam.), cempedak (*Artocarpus integer* (Thunb.) Merr.), and terap (*Artocarpus odoratissimus* Blanco) [25].

Artocarpus camansi Blanco, known as the breadnut, is native to New Guinea and probably also the Moluccas, in Indonesia, and the Philippines [26,27]. This species is diploid and is cultivated widely in the tropics because of its large, edible seeds. The tree can grow to a height of 10–15 m and the trunk may reach 1 m or more in diameter [26]. The fruits and seeds are rich in nutrients, with appreciable amounts of proteins, carbohydrates, minerals, and unsaturated fatty acids. The fruit is normally eaten when immature, when it is sliced thinly and boiled as a vegetable in soups or stews [28]. The draft genome of *A. camansi* was reported recently. The genome was assembled in 388 Mbp and the N50 scaffold was 2574 bp [29]. These authors also provided 333 nuclear markers that are informative for

phylogenetic analyses, and have been sequenced successfully in a number of different genera using target enrichment [23,30].

The goals of this study were to assemble the complete plastid genome of *A. camansi* from whole genome sequence data, compare its repetitive structures with those of *A. heterophyllum*, and verified the plastome structure and synteny among the members of the family Moraceae. We also constructed a plastid phylogenomic tree to explore the relationships among three families (Moraceae, Rhamnaceae and Cannabaceae) of the order Rosales.

2. Materials and Methods

2.1. Sampling, Genome Assembly, and Annotation

Illumina paired-end sequencing data of *A. camansi* were obtained from the NCBI Sequence Read Archive (accession no. SRR2910988). The plant sampling, library preparation, and parameters used for high throughput sequencing are available in Gardner et al. [29]. The paired-end reads were assembled into a complete plastid genome using Fast-Plast pipeline v.1.2.8 [31] with the `-subsample` option defined as 45,000,000 and Rosales order as the *bowtie_index*. The assembly of the plastid genome was curated using the Bowtie2 software by aligning the sequence reads in the plastid [32]. The alignments were converted into binary BAM format, sorted, and indexed using the samtools platform [33]. The genome coverage was then estimated from the alignments in the BAM files and the `genomeCoverageBed` command in the BEDTools software [34].

The plastid genome was annotated using Geseq [35] and adjusted manually through comparisons with the annotations of *Artocarpus heterophyllum* (MK303549.1), *Ficus carica* (NC_035237.1), *Ficus racemosa* (NC_028185.1), *Morus indica* (NC_008359.1), *Morus mongolica* (NC_025772.2), and *Morus notabilis* (NC_027110.1) in the software Geneious 11.0.4 (Biomatters Ltd., Auckland, NZ). The tRNAscan-SE procedure was used to annotate the tRNA in organellar search mode with the default parameters [36]. The circular plastid genome map was drawn up using OrganellarGenomeDRAW (OGDRAW) [37]. The nucleotide composition and codon usage were analyzed in MEGA [38] on the Bioinformatics web server (https://www.bioinformatics.org/sms2/codon_usage.html).

2.2. Characterization of Repeat Sequences

The REPuter server was used to detect and locate forward, reverse, palindrome, and complementary repeat sequences with a minimum size of 30 bp, with a hamming distance of 3, and at least 90% identity [39]. Microsatellites, also known as SSRs (Simple Sequence Repeats), were detected using the MISA software v2.1 (available online at: <http://pgrc.ipk-gatersleben.de/misa/misa.html>) [40,41]. The SSR search was based on the following parameters: 10 repeat units for the mono-nucleotides, 5 repeat units for the di-nucleotides, 4 repeat units for the tri-nucleotides, and 3 repeat units for the tetra-, penta-, and hexa-nucleotides. For the comparative analysis among genera, the repeat analysis focused on the plastid genome of *A. camansi* and *A. heterophyllum*, the phylogenetically closest plastid genome available in the NCBI.

2.3. Non-Synonymous (K_a) and Synonymous (K_s) Substitution Rate Analysis and Nucleotide Diversity Analysis

To estimate the non-synonymous (K_a) and synonymous (K_s) substitution rates, the 79 protein coding genes of *A. camansi* and *A. heterophyllum* were aligned separately using the MAFFT tool [42] available in the software Geneious 11.0.4 (Biomatters Ltd., Auckland, NZ). The non-synonymous (K_a) and synonymous (K_s) substitutions and the K_a/K_s ratio were then estimated for each gene using the software DnaSP v6.12.03 [43].

To assess the nucleotide diversity (P_i), the complete plastid genome sequences of *A. camansi* and *A. heterophyllum* were first aligned using the MAFFT aligner tool available in the Geneious software. A sliding window analysis was then run to calculate the P_i values using the software DnaSP v6.12.03 [43] with a window length of 600 bp and a step size of 200 bp.

2.4. Comparative Plastid Genome Analysis

The complete plastid genome of *A. camansi* was compared with the plastid genomes available for six other moraceous species using the mVISTA program in the Shuffle-LAGAN mode [44]. In this analysis, the recently assembled and annotated *A. camansi* plastid genome was used as the reference. The border positions of the LSC, SSC, and IR regions (LSC/IRB/SSC/IRA) were plotted and compared between *A. camansi* and the six species using IRscope [45].

2.5. Phylogenetic Analyses

Fifty-six protein coding genes were recorded in 34 species from three plant families (Moraceae, Rhamnaceae, and Cannabaceae) and two species (*Glycine soja* and *Vigna unguiculata*) included as outgroups. All the genes were obtained from NCBI GenBank except those of *A. camansi* (see Supplementary Table S1 for species and accession numbers). The sequences of all the genes were aligned and concatenated, and used to obtain the priors provided by the evolutionary model. This model was selected using the Bayesian Information Criterion (BIC), implemented in the JMODELTEST 2 software [46]. The GTR+G model ($-\ln L = 187133.0707$, wBIC = 0.7612) was selected, with a gamma shape parameter equal to 0.2480.

The phylogenetic tree was assembled in MR BAYES v.3.2.7 [47], using Bayesian inference. The analysis was based on four independent runs of 10×10^6 generations, assigned to each chain, with the *a posteriori* probability distribution being determined every 500 generations. The first 2500 trees were discarded prior to the construction of the consensus tree, to ensure the convergence of the chains. The final tree was edited and visualized in FigTree v 1.4.4 [48].

3. Results and Discussion

3.1. Plastid Genome Assembly, Organization, and Features

A total of 36.9 Gbp of data with 182,485,953 Illumina paired-end raw reads was downloaded from the NCBI database and used to assemble the plastid genome of *A. camansi*. After filtering using the $-\text{subsample}$ option, a total of 21,841,505 paired-end raw reads (mean length of 99 bp) were retained and the plastid genome was assembled successfully using Fast-Plast. The mean genome coverage of the alignments was 10,733X, with a standard deviation of 1683 (median = 10,849; minimum = 2047; maximum = 17,649; Supplementary Figure S1). The sequence of the chloroplast genome was deposited to GenBank under the accession number MW149075.

The plastid genome of *A. camansi* is 160,096 bp in length and presents the circular quadripartite structure typical of the angiosperms, which comprises a large single copy (LSC) region of 88,745 bp and a small single copy (SSC) region of 19,883 bp, separated by a pair of inverted repeat (IR) regions, each with a length of 25,734 bp (Figure 1; Table 1). The plastid genome of *A. camansi* is similar in size to that of *A. heterophyllum* [49], and those of other moraceous species [12,50,51] (see also Table 1).

The total GC content (or guanine-cytosine content) was 36.0%, which is also very similar to that of *A. heterophyllum* (36.1%) and the plastomes of other moraceous species, whose overall GC content ranges from 35.6% to 36.4% (Table 1; see also Supplementary Table S1). The GC content of the IR regions (42.7%) was higher than that of the LSC (33.7%) and SSC (28.8%) regions (Supplementary Table S2). The high GC content of the IR region may be attributed to the presence of rRNA and tRNA genes with a relatively high GC content, which occupy the majority of this region [52,53].

Figure 1. Gene map of the *A. camansi* plastid genome. The genes are color-coded according to their functional groups. The genes on the inside of the circle are transcribed in the clockwise direction, and those on the outside are transcribed counterclockwise. The inner circle shows the quadripartite structure of the chloroplast, that is, the small single copy (SSC), large single copy (LSC), and pair of inverted repeats (IRa and IRb). The darker and lighter gray shading in the inner circle correspond to guanine-cytosine (GC) and adenine-thymine (AT) content, respectively.

Table 1. Basic parameters of the plastid genome of the newly assembled *A. camansi* and other species of the family Moraceae.

Species	Size (bp)	LSC (bp)	SSC (bp)	IRs (bp)	GC Content (%)
<i>Artocarpus camansi</i>	160,096	88,745	19,883	25,734	36.0%
<i>Artocarpus heterophyllus</i>	160,389	89,077	19,896	25,708	36.1%
<i>Ficus carica</i>	160,602	88,661	20,137	25,902	35.9%
<i>Ficus racemosa</i>	159,473	88,110	20,007	25,678	35.9%
<i>Morus indica</i>	158,484	87,386	19,742	25,678	36.4%
<i>Morus mongolica</i>	158,459	87,367	19,736	25,678	36.3%
<i>Morus notabilis</i>	158,680	87,470	19,772	25,719	36.4%

The plastid genome of *A. camansi* was predicted to encode 113 different genes, with 79 protein-coding genes, 30 transfer RNA (tRNA), and four ribosomal RNA (rRNA) genes (Figure 1; Table 2). Eighteen genes were duplicated completely in the IR regions, including seven protein-coding genes (*ndhB*, *rpl2*, *rpl23*, *rps7*, *rps12*, *ycf2*, and *ycf15*), seven tRNAs (*trnI*-CAU, *trnL*-CAA, *trnV*-GAC, *trnI*-GAU, *trnA*-UGC, *trnR*-ACG, and *trnN*-GUU), all four rRNAs (*rrn4.5*, *rrn5*, *rrn16*, and *rrn23*), and part of the 5' end of *ycf1*. Of the remaining genes, the LSC region contains 22 tRNA and 61 protein-coding genes, while the SSC region consists of one tRNA gene and 10 protein-coding genes.

Table 2. List of the genes found in the plastid genome of *A. camansi*.

Category	Gene group	Genes
Self-replication	Large subunit of ribosomal proteins	<i>rpl2</i> ^{1,2} , <i>rpl14</i> , <i>rpl16</i> ¹ , <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> ² , <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>
	Small subunit of ribosomal proteins	<i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> ² , <i>rps8</i> , <i>rps11</i> , <i>rps12</i> ^{1,2} , <i>rps14</i> , <i>rps15</i> , <i>rps16</i> ¹ , <i>rps18</i> , <i>rps19</i>
	DNA-dependent RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> ¹ , <i>rpoC2</i>
	Ribosomal RNA genes	<i>rrn4.5</i> ² , <i>rrn5</i> ² , <i>rrn16</i> ² , <i>rrn23</i> ²
Transfer RNA genes		<i>trnA</i> -UGC ^{1,2} , <i>trnC</i> -GCA, <i>trnD</i> -GUC, <i>trnE</i> -UUC, <i>trnF</i> -GAA, <i>trnG</i> -CAU, <i>trnG</i> -UCC ¹ , <i>trnG</i> -UCC, <i>trnH</i> -GUG, <i>trnI</i> -CAU ² , <i>trnI</i> -GAU ^{1,2} , <i>trnK</i> -UUU ¹ , <i>trnL</i> -CAA ² , <i>trnL</i> -UAA ¹ , <i>trnL</i> -UAG, <i>trnM</i> -CAU, <i>trnN</i> -GUU ² , <i>trnP</i> -UGG, <i>trnQ</i> -UUG, <i>trnR</i> -ACG ² , <i>trnR</i> -UCU, <i>trnS</i> -GCU, <i>trnS</i> -UGA, <i>trnS</i> -GGA, <i>trnT</i> -UGU, <i>trnT</i> -GGU, <i>trnV</i> -UAC ¹ , <i>trnV</i> -GAC ² , <i>trnW</i> -CCA, <i>trnY</i> -GUA
	Photosynthesis	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i>
	Photosystem I	<i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbH</i> , <i>psbI</i> , <i>psbJ</i> , <i>psbK</i> , <i>psbL</i> , <i>psbM</i> , <i>psbN</i> , <i>psbT</i> , <i>psbZ</i>
	Photosystem II	<i>ndhA</i> ¹ , <i>ndhB</i> ^{1,2} , <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhJ</i> , <i>ndhK</i>
NADH dehydrogenase	NADH dehydrogenase	
	Cytochrome b/f complex	<i>petA</i> , <i>petB</i> ¹ , <i>petD</i> ¹ , <i>petG</i> , <i>petL</i> , <i>petN</i>
	ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> ¹ , <i>atpH</i> , <i>atpI</i>
	RubisCo large subunit	<i>rbcL</i>
Other genes	Maturase K	<i>matK</i>
	Envelope membrane protein	<i>cemA</i>
	Subunit of acetyl-CoA carboxylase	<i>accD</i>
	C-type cytochrome synthesis gene	<i>ccsA</i>
	Protease	<i>clpP</i> ¹
	Conserved hypothetical chloroplast open reading frames	<i>ycf1</i> , <i>ycf2</i> ² , <i>ycf3</i> ¹ , <i>ycf4</i> , <i>ycf15</i> ²

¹—Gene with introns; ²—Gene duplicated completely in the inverted repeat.

Nine of the protein-coding genes annotated here (*atpF*, *ndhA*, *ndhB*, *petB*, *petD*, *rpl2*, *rpl16*, *rpoC1*, and *rps16*) and six tRNAs (*trnK*-UUU, *trnG*-UCC, *trnL*-UAA, *trnV*-UAC, *trnI*-GAU, and *trnA*-UGC) each contained one intron, while three genes (*rps12*, *clpP* and *ycf3*) each contained two introns (Table 2; see also Supplementary Table S3). The *rps12* gene presents signatures of trans-splicing, with the 5' end located in the LSC region, whereas the duplicated 3' end found in the IRs has an identical sequence, in an opposite transcriptional direction. The *rps12* gene is thought to be trans-splicing and highly conserved in most angiosperms, although this gene appears to be even more conserved in the pteridophytes. However, this gene lacks the second intron in some species of the three basal fern lineages, the Psilotales, Ophioglossales, and Equisetales [54–56]. The *trnK*-UUU has the largest intron, which encompasses the *matK* gene, with 2556 bp, whereas the intron of *trnL*-UAA is the smallest (501 bp). In general, the gene content, order, and organization of the *A. camansi* plastid genome is highly similar to that of the closely-related *A. heterophyllum*, reported previously [49].

The codon usage was analyzed using 79 protein-coding gene sequences, which have a strong (63%) AT bias (Supplementary Table S2). A total of 23,068 codons were identified in the *A. camansi* plastid genome (Table 3), with 64 codon types being identified, which encode all 20 amino acids. The most abundant amino acid was leucine, with 2523 codons, that is, approximately 10.94% of the total

number of codons, whereas, excluding stop codons, cysteine was the least abundant one (270 codons, ~1.17% of the total number). A prevalence of leucine and reduced quantities of cysteine were also observed by Somaratne et al. [57] in their evaluation of the plastid genome of species of the bush clover genus *Lespedeza*. The most abundant codon was ATT, which encodes isoleucine. Only one codon was identified for the tryptophan (TGG) and methionine (ATG) amino acids (Table 3).

Table 3. Codon usage in the plastid genome of *A. camansi*.

Amino Acid	Codon	Number	Fraction	Amino Acid	Codon	Number	Fraction
Alanine	GCG	142	0.12	Lysine	AAG	316	0.25
	GCA	319	0.27		AAA	935	0.75
	GCT	535	0.45	Methionine	ATG	557	1.00
	GCC	181	0.15		TTT	898	0.68
	AGG	150	0.11	Phenylalanine	TTC	413	0.32
Arginine	AGA	422	0.32		CCG	154	0.16
	CGG	91	0.07		CCA	281	0.30
	CGA	300	0.23	Proline	CCT	348	0.37
	CGT	282	0.21		CCC	160	0.17
	CGC	86	0.06		AGT	356	0.21
Asparagine	AAT	870	0.76	Serine	AGC	108	0.06
	AAC	277	0.24		TCG	147	0.09
Aspartic Acid	GAT	698	0.80		TCA	356	0.21
	GAC	178	0.20		TCT	498	0.29
Cysteine	TGT	205	0.76		TCC	262	0.15
	TGC	65	0.24		ACG	139	0.12
Glutamine	CAG	185	0.23	Threonine	ACA	349	0.30
	CAA	635	0.77		ACT	464	0.40
Glutamic Acid	GAG	288	0.24		ACC	200	0.17
	GAA	893	0.76	Tryptophan	TGG	400	1.00
	GGG	231	0.15	Tyrosine	TAT	699	0.80
Glycine	GGA	619	0.41		TAC	173	0.20
	GGT	501	0.33		GTG	177	0.14
	GGC	161	0.11	Valine	GTA	521	0.40
Histidine	CAT	412	0.76		GTT	448	0.35
	CAC	132	0.24		GTC	148	0.11
	ATA	673	0.33	Stop codon	TGA	34	0.24
Isoleucine	ATT	998	0.49		TAG	43	0.31
	ATC	370	0.18		TAA	62	0.45
	TTG	518	0.21				
Leucine	TTA	805	0.32				
	CTG	172	0.07				
	CTA	337	0.13				
	CTT	531	0.21				
	CTC	160	0.06				

The GC content of the codon positions is the principal factor determining biased codon usage in many organisms. In the *A. camansi*, the mean GC content of the first codon position was 44.7%, while it was 37.1% for the second position, and 29.4%, for the third position (Supplementary Table S2). This indicates a strong bias toward A/T in the third codon position, which is consistent with the previous studies of plastid genomes [57,58]. The presence of translation-preferred codons demonstrates the importance of evolutionary processes in the plastid genome that result from both natural selection and mutation preferences [58]. This variation in the codon bias is highly similar to that found in other moraceous species [12,50], consistent with the fact that they are highly conserved.

3.2. Repetitive Sequences

Microsatellites, or SSRs, are short tandem repeat DNA sequences of one to six base pairs, which are distributed throughout the plastid genome. A total of 80 SSRs were identified in the *A. camansi* plastid genome using MISA, of which 56 are mononucleotide (70.0%), 10 dinucleotide (12.5%), 4 trinucleotide (5.0%), 9 tetranucleotide (11.3%), and 1 pentanucleotide (1.3%), with a length of at least 10 bp and 3 to

14 repeats. A total of 74 SSRs of at least 10 bp were detected in *A. heterophyllum*, with between 3 and 18 repeats. In this species, 52 of the SSRs were mononucleotide (70.3%), 10 were dinucleotide (13.5%), 3 were trinucleotide (4.1%), and nine were tetranucleotide (12.2%). No hexanucleotide SSRs were detected in either of the two species, and only one pentanucleotide repeat was identified, in *A. camansi* (Figure 2A). These findings are consistent with previous studies, which recorded low frequencies of tri-, tetra-, penta-, and hexa-nucleotide repeat motifs in plastid genomes, while most motifs were mono- and di-nucleotide repeats [52,57,59,60].

The A/T mononucleotide repeats were the most abundant SSRs identified in the *A. camansi* and *A. heterophyllum* plastid genomes, accounting for 67.5% (54) and 66.2% (49) of the SSRs, respectively (Figure 2B). As expected, since the higher plant plastid genomes are generally AT-rich, most of the di-, tri-, and tetra-nucleotides were AT motifs, showing a strong AT bias in the SSRs of the plastid genomes of these two *Artocarpus* species, which is consistent with the pattern observed in *Morus mongolica*, which has an even higher AT content (99.7%) of SSRs in its genome [12]. The SSRs were distributed in the LSC, SSC, and IR regions of *A. camansi*, but were more abundant in the LSC than in the SSC and IR regions (Figure 2C). Overall, 52 of the 80 SSRs identified in the *A. camansi* plastid genome were located in intergenic regions (65.0%), while 15 were identified in the introns (18.8%), and 13 in coding regions (16.3%), including *atpF*, *rpoB*, *atpB*, *ndhF*, *rps15*, and *ycf1*. The *ycf1* gene contained five SSRs, and the *ndhF* gene contained four (Figure 2D). In *A. heterophyllum*, in comparison, 57 SSRs were located in intergenic regions (77.03%), 11 in introns (14.86%), and 6 in coding regions (8.11%), including *rpoC2*, *rpoB*, *atpB*, and *ndhF* (Figure 2D). These differences are probably due to the fact that the *ycf1* gene of *A. heterophyllum* does not encompass the IR-SSC boundary, as observed in *A. camansi* (see below). The *ycf1* gene of *A. camansi* encodes a protein of 1867 amino acids, and the portion of the *ycf1* located in the IR region is short and conserved, while that located in the SSC region is highly variable in most land plants [61,62]. This region has also been shown to be more variable than *matK* in many taxa and would thus be suitable for molecular systematics at low taxonomic levels, and for DNA barcoding [61–63].

The repetitive structure of the plastid genome may promote rearrangement and increase the genetic diversity of plant populations [64,65]. A total of 50 tandem repeat structures were identified in the *A. camansi* plastid genome (Forward = 21; Palindromic = 29; Figure 3A), whereas 61 were identified in the plastid genome of *A. heterophyllum* (Forward = 32; Palindromic = 27; Reverse = 2; Figure 3B). No complementary repeats were found in both plastid genomes whereas reverse repeats were detected only in *A. heterophyllum* (Figure 3B). The repeats were also larger in *A. heterophyllum* (30–124 bp) in comparison with *A. camansi* (30–69 bp). The LSC region contained more repeats than the IR and SSC regions in both species (Figure 3C). Moreover, a large number of tandem repeat structures were found in intergenic regions and introns, while only seven repeats were identified in the coding regions of *A. heterophyllum*, and nine in that of *A. camansi*, in the *ycf2*, *psaB*, *trnG-UCC*, and *trnS-UGA* sequences of both species (Figure 3D). The largest number of repeats were located in the *ycf2* gene (four repeats in both *A. camansi* and *A. heterophyllum*). This can be attributed to the length of the *ycf2* sequence (6888 bp in *A. camansi* and 6846 bp in *A. heterophyllum*). This is consistent with the pattern observed in *Carya illinoensis* [52], *Citrus sinensis* [66], *Nasturtium officinale* [67], and *Toxicodendron vernicifluum* [65].

Figure 2. Distribution of the microsatellites identified in the *A. camansi* plastid genome. (A) The number of different microsatellite types detected; (B) The number of different microsatellite motifs detected; (C) The number of microsatellites in the different regions of the plastid genome; (D) The number of microsatellites in intergenic and coding regions, and the introns.

Figure 3. Distribution of the repeat structures in the *Artocarpus* plastid genome. (A) The number of repeat structures in the *A. camansi* plastid genome; (B) The number of repeat structures in the *A. heterophyllus* plastid genome; (C) The number of repeat structures in the different regions of the plastid genomes of the two species; (D) The number of structures in intergenic and coding regions, and the introns.

3.3. The *Ka/Ks* Ratio and Nucleotide Diversity

In coding regions, substitutions can be either synonymous (*Ks*) or non-synonymous (*Ka*). Synonymous substitutions, or silent mutations, do not alter the amino acid composition of the encoded protein, whereas non-synonymous substitutions do modify this composition. The *rps19* gene, associated with the small subunit of ribosomal proteins, had the highest non-synonymous rate (0.0334), while the *psbT* gene, which is related to the core complex of photosystem II (PSII), had the highest synonymous rate of 0.0864 (Supplementary Table S4).

The ratio of the non-synonymous to the synonymous rate (Ka/Ks) was determined for the 79 protein-coding genes common to the plastid genomes of *A. camansi* and *A. heterophyllum* (Figure 4 and Supplementary Table S4). The Ka/Ks ratios of the two *Artocarpus* species ranged from 0.000 to 2.299 (mean = 0.317). The lowest Ka/Ks ratios were observed in genes that encode subunits of photosystem I (mean = 0.010) and photosystem II (0.038), the large subunit of the RuBisCO (0.033), and the subunits of the ATP synthase (0.094).

four of these loci are located in the intergenic spacer of the LSC region, whereas the latter is located in the SSC region. The intergenic region *rpl32-trnL* that has the highest nucleotide diversity ($P_i = 0.04$) in the present study has also been found to be highly variable in the *Machilus*, *Morus*, and *Solanum* plastid genomes, and in those of the spermatophyte species [61,62,70,73,74]. Given their previous designation as hotspots, two intergenic spacers appear to be universal in the family Moraceae. The *trnT-trnL* were found to be highly variable between two species of the genus *Morus*, as was *psbE-petL* in all five *Morus* plastomes [12,70]. The *trnH-psbA* spacer of *Artocarpus* was also highly variable ($P_i = 0.03667$). This region has been widely used as a plant DNA barcode and, when combined with ITS2, it was significantly more efficient in the identification of taxa than *matK+rbcL* in 18 families and 21 genera, indicating that it may be an optimal marker for species identification [75]. Overall, then, all these regions may be especially valuable for further phylogenetic analysis of the genus, and have good potential for use as barcode markers.

Figure 5. Results of the sliding window analysis of the complete plastid genomes of *A. camansi* and *A. heterophyllus*.

3.4. Comparative Plastid Genome Structure

The comparative analysis of structural characteristics of the chloroplast among seven moraceous species revealed a high level of sequence similarity, indicating a highly conserved evolutionary model for these plastid genomes (Figure 6). The analyses also demonstrated clearly that the IR regions are more conserved than the SSC and LSC regions, which may be due to the copy correction of the IR regions by gene conversion [76]. The most variable loci in the seven chloroplast genomes compared here were found in the *matK*, *rpoC2*, *rps19*, *ndhF*, and *ycf1* genes. These genes were also found to be highly divergent in other plastid genomes [61,67,77,78], and may thus constitute potentially useful markers for phylogenetic analyses at different taxonomic levels. In the case of the noncoding regions, a high level of variation was observed in the intergenic spacers, including *trnH-GUG-psbA*, *matK-rps16*, *rps16-psbK*, *trnG-UCC-trnR-UCU*, *trnT-UAA-trnL-GAA*, *trnL-UAA-trnF-GAA*, *trnF-GAA-ndhJ*, *accD-psaI*, *psbE-petL*, and *rpl32-trnL-UAG*. Some of these noncoding regions have the highest levels of observed nucleotide diversity.

Figure 6. The alignment of the plastid genomes of seven species of the family Moraceae, with *A. camansi* as a reference. The vertical scale indicates the percentage of identity, which ranges from 50% to 100%. The coding regions are shown in purple, and the non-coding regions in red. The gray arrows above the alignment indicate the orientation of the genes.

The contraction and expansion of the IR region and the SSC boundaries can be considered to be the primary mechanism of variation in the length of the plastid genomes of higher plants [79]. The length of the IR regions was highly similar in the seven moraceous species analyzed here however, ranging from 25,678 bp in *F. racemosa*, *M. indica*, and *M. mongolica* to 25,902 in *F. carica* (Figure 7). In most species, the *rps19* gene is located to the left of the LSC-IRb boundary (JLB), and *rpl2* is located to the right. The exception was *F. carica*, in which the LSC-IRb boundary was located within the *rps19* sequence, and had a length of 108 bp, located in the IRb (Figure 7). The IRb-SSC boundary (JSB) is located within the *ndhF* gene, so that from 13 bp (in *F. carica*) to 42 bp (in *A. camansi*) of its coding sequence is in the IRb region. In *A. heterophyllus* and *F. carica* the *ycf1* is also located within the IRb/SSC region, with a length of 4 bp and 84 bp, respectively (Figure 7). The SSC-IRa boundaries (JSA) were located within the *ycf1* pseudogene, with the fragment located in the IRa region ranging from 983 bp in *A. camansi* to 1079 bp in *F. racemosa* (Figure 7). In *A. heterophyllus*, the *ycf1* pseudogene embedded in the SSC-IRa was unannotated. The IRa-LSC (JLA) boundary is located upstream from the *rpl2* and downstream from the *trnH*. The *trnH* gene was unannotated in *F. racemosa*. The present study is the first to compare the IR boundaries in species of the family Moraceae, in order to better evaluate the evolution of the plastome. The analysis demonstrated clearly that the IR and the size genome are highly conserved in the study species.

Figure 7. Comparison of the borders of the LSC (light blue), SSC (light green), and IR (orange) regions among the seven moraceous plastid genomes analyzed in the present study.

3.5. Phylogenetic Analyses

The phylogeny recovered from the sequencing of the 56 protein coding genes in 33 species of the families Moraceae, Rhamnaceae, and Cannabaceae (Figure 8) confirms the hypothesis of a high level of plastome conservation in the Moraceae. The families were arranged in three well-defined clusters, and based on the analysis, the Moraceae is the sister group to the Cannabaceae, and the Rhamnaceae is, in turn, sister to that clade. This arrangement also reinforces the genetic similarity of the plastid genomes of *A. camansi* and *A. heterophyllus*.

Figure 8. Phylogeny of 33 species of the families Moraceae, Rhamnaceae, and Cannabaceae, based on the sequences of 56 chloroplast genes. Numbers represent the Bayesian posterior probability given to each node. *Glycine soja* and *Vigna unguiculata* are the outgroups.

Previous studies have confirmed the phylogenetic proximity of *Artocarpus* and *Morus*. In a phylogeny based on the *ndhF* chloroplast gene, Datwyler and Weiblen [80] recovered a cluster that included *A. camansi*, *A. heterophyllus*, *A. altilis*, *A. vriesiana*, *M. nigra*, and *M. alba*. Zerega et al. [25] analyzed nuclear ITS sequences and chloroplast sequences from the *trnLF* region, and also confirmed the genetic proximity of *Artocarpus* to *M. alba*, with these species diverging from *Humulus lupulus* and *Cannabis sativa*. The complete sequencing of the *A. camansi* genome is consistent with the inference that *Artocarpus* separated from *Morus* through an event of the total duplication of the genome in the tribe Artocarpeae [29].

4. Conclusions

The present study describes the plastid genome of *A. camansi*, and confirms a highly conserved structure of the plastome in the family Moraceae. In particular, the study corroborates the hypothesis of the intense conservation of plastid genomes during the evolutionary of these plants, supporting the understanding of the genomic features of the chloroplast genes that are common to the different moraceous species. The chloroplast microsatellites, and the most diverse coding and noncoding regions identified in the present study may also provide valuable molecular markers for further research into the evolution of the Moraceae.

Supplementary Materials: Supplementary Materials can be found at <http://www.mdpi.com/1999-4907/11/11/1179/s1>. Supplementary Figure S1. Genome sequencing coverage distribution of all reads that aligned in the in *A. camansi* plastid genome. Supplementary Table S1. Accession number and sampled plastid genomes obtained from GenBank and used to recover phylogeny for Moraceae, Rhamnaceae, and Cannabaceae. Supplementary Table S2. Base composition of the *Artocarpus camansi* plastid genome. Supplementary Table S3. Genes with intron in the *Artocarpus camansi* plastid genome, including the exon and intron length. Supplementary Table S4. The Ka, Ks and Ka/Ks ratio of *A. camansi* and *A. heterophyllus* plastid genome for individual genes and region.

Author Contributions: Conceptualization, U.J.B.d.S., L.C.V. and L.A.B.; Formal analyses, U.J.B.d.S. and L.C.V.; Supervision, L.C.V.; Writing—original draft preparation, U.J.B.d.S.; Writing—review and editing, L.C.V. and L.A.B.; Project administration, F.G.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors are grateful to the Fundação de Amparo à Pesquisa do Estado de Goiás (Goiás State Research Foundation, FAPEG) and the Goiano Federal Institute Rio Verde Campus (IFGoiano), for the infrastructure and the assistants involved in this study.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that influenced the work reported in this paper.

References

1. Daniell, H.; Lin, C.-S.; Yu, M.; Chang, W.-J. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biol.* **2016**, *17*, 134. [CrossRef] [PubMed]
2. Leister, D. Chloroplast research in the genomic age. *Trends Genet.* **2003**, *19*, 47–56. [CrossRef]
3. Dobrogojski, J.; Adamiec, M.; Luciński, R. The chloroplast genome: A review. *Acta Physiol. Plant.* **2020**, *42*, 98. [CrossRef]
4. Shaw, J.; Lickey, E.B.; Schilling, E.E.; Small, R.L. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. *Am. J. Bot.* **2007**, *94*, 275–288. [CrossRef]
5. Provan, J.; Powell, W.; Hollingsworth, P.M. Chloroplast microsatellites: New tools for studies in plant ecology and evolution. *Trends Ecol. Evol.* **2001**, *16*, 142–147. [CrossRef]
6. Ravi, V.; Khurana, J.P.; Tyagi, A.K.; Khurana, P. An update on chloroplast genomes. *Plant Syst. Evol.* **2008**, *271*, 101–122. [CrossRef]
7. Jansen, R.K.; Ruhlmann, T.A. Plastid genomes of seed plants. In *Genomics of Chloroplasts and Mitochondria*; Bock, R., Knoop, V., Eds.; Springer: Dordrecht, The Netherlands, 2012; Volume 35, pp. 103–126.
8. Palmer, J.D. Comparative organization of chloroplast genomes. *Annu. Rev. Genet.* **1985**, *19*, 325–354. [CrossRef]

9. Wicke, S.; Müller, K.F.; de Pamphilis, C.W.; Quandt, D.; Wickett, N.J.; Zhang, Y.; Renner, S.S.; Schneeweiss, G.M. Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. *Plant Cell* **2013**, *25*, 3711–3725. [\[CrossRef\]](#)
10. Frailey, D.C.; Chaluvadi, S.R.; Vaughn, J.N.; Coatney, C.G.; Bennetzen, J.L. Gene loss and genome rearrangement in the plastids of five Hemiparasites in the family Orobanchaceae. *BMC Plant Biol.* **2018**, *18*, 30. [\[CrossRef\]](#)
11. Raubeson, L.A.; Jansen, R.K. Chloroplast genomes of plants. In *Plant Diversity and Evolution: Genotypic Variation in Higher Plants*; Henry, R.J., Ed.; CABI Publishing: Wallingford, UK, 2005; pp. 45–68.
12. Kong, W.; Yang, J. The complete chloroplast genome sequence of *Morus mongolica* and a comparative analysis within the Fabidae clade. *Curr. Genet.* **2016**, *62*, 165–172. [\[CrossRef\]](#)
13. Bock, R. Structure, function, and inheritance of plastid genomes. In *Cell and Molecular Biology of Plastids*; Bock, R., Ed.; Topics in Current Genetics; Springer: Heidelberg, Germany, 2007; pp. 29–63.
14. Wicke, S.; Schneeweiss, G.M.; Depamphilis, C.W.; Müller, K.F.; Quandt, D. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Mol. Biol.* **2011**, *76*, 273–297. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Jansen, R.K.; Cai, Z.; Raubeson, L.A.; Daniell, H.; dePamphilis, C.W.; Leebens-Mack, J.; Müller, K.F.; Guisinger-Bellian, M.; Haberle, R.C.; Hansen, A.K. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19369–19374. [\[CrossRef\]](#)
16. Wu, F.-H.; Kan, D.-P.; Lee, S.-B.; Daniell, H.; Lee, Y.-W.; Lin, C.-C.; Lin, N.-S.; Lin, C.-S. Complete nucleotide sequence of *Dendrocalamus latiflorus* and *Bambusa oldhamii* chloroplast genomes. *Tree Physiol.* **2009**, *29*, 847–856. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Jansen, R.K.; Wojciechowski, M.F.; Sanniyasi, E.; Lee, S.-B.; Daniell, H. Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of *rps12* and *clpP* intron losses among legumes (Leguminosae). *Mol. Phylogenet. Evol.* **2008**, *48*, 1204–1217. [\[CrossRef\]](#)
18. Saski, C.; Lee, S.-B.; Fjellheim, S.; Guda, C.; Jansen, R.K.; Luo, H.; Tomkins, J.; Rognli, O.A.; Daniell, H.; Clarke, J.L. Complete chloroplast genome sequences of *Hordeum vulgare*, *Sorghum bicolor* and *Agrostis stolonifera*, and comparative analyses with other grass genomes. *Theor. Appl. Genet.* **2007**, *115*, 571–590. [\[CrossRef\]](#)
19. Daniell, H.; Wurdack, K.J.; Kanagaraj, A.; Lee, S.-B.; Saski, C.; Jansen, R.K. The complete nucleotide sequence of the cassava (*Manihot esculenta*) chloroplast genome and the evolution of *atpF* in Malpighiales: RNA editing and multiple losses of a group II intron. *Theor. Appl. Genet.* **2008**, *116*, 723. [\[CrossRef\]](#)
20. Shinozaki, K.; Ohme, M.; Tanaka, M.; Wakasugi, T.; Hayashida, N.; Matsubayashi, T.; Zaita, N.; Chunwongse, J.; Obokata, J.; Yamaguchi-Shinozaki, K. The complete nucleotide sequence of the tobacco chloroplast genome: Its gene organization and expression. *EMBO J.* **1986**, *5*, 2043–2049. [\[CrossRef\]](#)
21. Ohyama, K.; Fukuzawa, H.; Kohchi, T.; Shirai, H.; Sano, T.; Sano, S.; Umesono, K.; Shiki, Y.; Takeuchi, M.; Chang, Z. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature* **1986**, *322*, 572–574. [\[CrossRef\]](#)
22. Twyford, A.D.; Ness, R.W. Strategies for complete plastid genome sequencing. *Mol. Ecol. Resour.* **2017**, *17*, 858–868. [\[CrossRef\]](#)
23. Zerega, N.J.C.; Gardner, E.M. Delimitation of the new tribe Parartocarpeae (Moraceae) is supported by a 333-gene phylogeny and resolves tribal level Moraceae taxonomy. *Phytotaxa* **2019**, *388*, 253–265. [\[CrossRef\]](#)
24. Williams, E.W.; Gardner, E.M.; Harris III, R.; Chaveerach, A.; Pereira, J.T.; Zerega, N.J.C. Out of Borneo: Biogeography, phylogeny and divergence date estimates of *Artocarpus* (Moraceae). *Ann. Bot.* **2017**, *119*, 611–627. [\[CrossRef\]](#)
25. Zerega, N.J.C.; Supardi, N.; Motley, T.J. Phylogeny and recircumscription of Artocarpeae (Moraceae) with a focus on *Artocarpus*. *Syst. Bot.* **2010**, *35*, 766–782. [\[CrossRef\]](#)
26. Ragone, D. *Artocarpus camansi* (breadnut) ver 2.1. In *Species Profiles for Pacific Island Agroforestry*; Elevitch, C.R., Ed.; Permanent Agriculture Resources (PAR): Holualoa, Hawaii, 2006; pp. 1–11.
27. Jarrett, F.M. Studies in *Artocarpus* and allied genera, III. A revision of *Artocarpus* subgenus *Artocarpus*. *J. Arnold Arbor.* **1959**, *40*, 113–155.
28. Adeleke, R.O.; Abiodun, O.A. Nutritional composition of breadnut seeds (*Artocarpus camansi*). *Afr. J. Agric. Res.* **2010**, *5*, 1273–1276.

29. Gardner, E.M.; Johnson, M.G.; Ragone, D.; Wickett, N.J.; Zerega, N.J.C. Low-coverage, whole-genome sequencing of *Artocarpus camansi* (Moraceae) for phylogenetic marker development and gene discovery. *Appl. Plant Sci.* **2016**, *4*, 1600017. [\[CrossRef\]](#)
30. Johnson, M.G.; Gardner, E.M.; Liu, Y.; Medina, R.; Goffinet, B.; Shaw, A.J.; Zerega, N.J.C.; Wickett, N.J. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Appl. Plant Sci.* **2016**, *4*, 1600016. [\[CrossRef\]](#)
31. McKain, M.R.; Wilson, M. Fast-Plast: Rapid de Novo Assembly and Finishing for Whole Chloroplast Genomes. Available online: <https://github.com/mrmckain/Fast-Plast2017> (accessed on 11 July 2020).
32. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357. [\[CrossRef\]](#)
33. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [\[CrossRef\]](#)
34. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842. [\[CrossRef\]](#)
35. Tillich, M.; Lehwork, 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. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Lowe, T.M.; Chan, P.P. tRNAscan-SE On-line: Integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* **2016**, *44*, W54–W57. [\[CrossRef\]](#)
37. Lohse, M.; Drechsel, O.; Bock, R. OrganellarGenomeDRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* **2007**, *52*, 267–274. [\[CrossRef\]](#)
38. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [\[CrossRef\]](#)
39. Kurtz, S.; Choudhuri, J.V.; Ohlebusch, E.; Schleiermacher, C.; Stoye, J.; Giegerich, R. REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **2001**, *29*, 4633–4642. [\[CrossRef\]](#)
40. Beier, S.; Thiel, T.; Münch, T.; Scholz, U.; Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Thiel, T.; Michalek, W.; Varshney, R.; Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* **2003**, *106*, 411–422. [\[CrossRef\]](#)
42. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [\[CrossRef\]](#)
43. Rozas, J.; Ferrer-Mata, A.; Sánchez-DelBarrio, J.C.; Guirao-Rico, S.; Librado, P.; Ramos-Onsins, S.E.; Sánchez-Gracia, A. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **2017**, *34*, 3299–3302. [\[CrossRef\]](#)
44. Frazer, K.A.; Pachter, L.; Poliakov, A.; Rubin, E.M.; Dubchak, I. VISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* **2004**, *32*, W273–W279. [\[CrossRef\]](#)
45. Amiryousefi, A.; Hyvönen, J.; Pocai, P. IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* **2018**, *34*, 3030–3031. [\[CrossRef\]](#)
46. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. jModelTest 2: More models, new heuristics and parallel computing. *Nat. Methods* **2012**, *9*, 772. [\[CrossRef\]](#)
47. 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. [\[CrossRef\]](#)
48. Rambaut, A. *FigTree Version 1.4.3 Software*; Institute of Evolutionary Biology, University of Edinburgh: Edinburgh, Scotland, UK, 2016.
49. Liu, J.; Niu, Y.-F.; Ni, S.-B.; Liu, Z.-Y.; Zheng, C.; Shi, C. The complete chloroplast genome of *Artocarpus heterophyllus* (Moraceae). *Mitochondrial DNA Part B* **2018**, *3*, 13–14. [\[CrossRef\]](#)
50. 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. [\[CrossRef\]](#)
51. Bruun-Lund, S.; Clement, W.L.; Kjellberg, F.; Rønsted, 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. [\[CrossRef\]](#)

52. Mo, Z.; Lou, W.; Chen, Y.; Jia, X.; Zhai, M.; Guo, Z.; Xuan, J. The chloroplast genome of *Carya illinoensis*: Genome structure, adaptive evolution, and phylogenetic analysis. *Forests* **2020**, *11*, 207. [\[CrossRef\]](#)
53. He, Y.; Xiao, H.; Deng, C.; Xiong, L.; Yang, J.; Peng, C. The complete chloroplast genome sequences of the medicinal plant *Pogostemon cablin*. *Int. J. Mol. Sci.* **2016**, *17*, 820. [\[CrossRef\]](#)
54. Kuo, L.; Qi, X.; Ma, H.; Li, F. Order-level fern plastome phylogenomics: New insights from Hymenophyllales. *Am. J. Bot.* **2018**, *105*, 1545–1555. [\[CrossRef\]](#)
55. Liu, S.; Wang, Z.; Wang, H.; Su, Y.; Wang, T. Patterns and rates of plastid rps 12 gene evolution inferred in a phylogenetic context using plastomic data of ferns. *Sci. Rep.* **2020**, *10*, 1–12.
56. Lu, J.; Zhang, N.; Du, X.; Wen, J.; Li, D. Chloroplast phylogenomics resolves key relationships in ferns. *J. Syst. Evol.* **2015**, *53*, 448–457. [\[CrossRef\]](#)
57. Somaratne, Y.; Guan, D.-L.; Wang, W.-Q.; Zhao, L.; Xu, S.-Q. The complete chloroplast genomes of two *Lespedeza* species: Insights into codon usage bias, RNA editing sites, and phylogenetic relationships in Desmodieae (Fabaceae: Papilionoideae). *Plants* **2020**, *9*, 51. [\[CrossRef\]](#)
58. Yang, Y.; Zhu, J.; Feng, L.; Zhou, T.; Bai, G.; Yang, J.; Zhao, G. Plastid genome comparative and phylogenetic analyses of the key genera in Fagaceae: Highlighting the effect of codon composition bias in phylogenetic inference. *Front. Plant Sci.* **2018**, *9*, 82. [\[CrossRef\]](#) [\[PubMed\]](#)
59. Dong, W.-L.; Wang, R.-N.; Zhang, N.-Y.; Fan, W.-B.; Fang, M.-F.; Li, Z.-H. Molecular evolution of chloroplast genomes of orchid species: Insights into phylogenetic relationship and adaptive evolution. *Int. J. Mol. Sci.* **2018**, *19*, 716. [\[CrossRef\]](#)
60. George, B.; Bhatt, B.S.; Awasthi, M.; George, B.; Singh, A.K. Comparative analysis of microsatellites in chloroplast genomes of lower and higher plants. *Curr. Genet.* **2015**, *61*, 665–677. [\[CrossRef\]](#)
61. Dong, W.; Xu, C.; Li, C.; Sun, J.; Zuo, Y.; Shi, S.; Cheng, T.; Guo, J.; Zhou, S. *ycf1*, the most promising plastid DNA barcode of land plants. *Sci. Rep.* **2015**, *5*, 8348. [\[CrossRef\]](#)
62. Dong, W.; Liu, J.; Yu, J.; Wang, L.; Zhou, S. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS ONE* **2012**, *7*, e35071. [\[CrossRef\]](#) [\[PubMed\]](#)
63. Neubig, K.M.; Whitten, W.M.; Carlsward, B.S.; Blanco, M.A.; Endara, L.; Williams, N.H.; Moore, M. Phylogenetic utility of *ycf1* in orchids: A plastid gene more variable than *matK*. *Plant Syst. Evol.* **2009**, *277*, 75–84. [\[CrossRef\]](#)
64. Qian, J.; Song, J.; Gao, H.; Zhu, Y.; Xu, J.; Pang, X.; Yao, H.; Sun, C.; Li, C.; Liu, J. The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PLoS ONE* **2013**, *8*, e57607. [\[CrossRef\]](#) [\[PubMed\]](#)
65. Wang, L.; He, N.; Li, Y.; Fang, Y.; Zhang, F. Complete chloroplast genome sequence of Chinese lacquer tree (*Toxicodendron vernicifluum*, Anacardiaceae) and its phylogenetic significance. *Biomed Res. Int.* **2020**, *1*, 1–13. [\[CrossRef\]](#)
66. Bausher, M.G.; Singh, N.D.; Lee, S.-B.; Jansen, R.K.; Daniell, H. The complete chloroplast genome sequence of *Citrus sinensis* (L.) Osbeck var 'Ridge Pineapple': Organization and phylogenetic relationships to other angiosperms. *BMC Plant Biol.* **2006**, *6*, 21. [\[CrossRef\]](#)
67. Yan, C.; Du, J.; Gao, L.; Li, Y.; Hou, X. The complete chloroplast genome sequence of watercress (*Nasturtium officinale* R. Br.): Genome organization, adaptive evolution and phylogenetic relationships in Cardamineae. *Gene* **2019**, *699*, 24–36. [\[CrossRef\]](#)
68. Yang, Z.; Nielsen, R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* **2000**, *17*, 32–43. [\[CrossRef\]](#) [\[PubMed\]](#)
69. Nei, M.; Kumar, S. *Molecular Evolution and Phylogenetics*; Oxford University Press: New York, NY, USA, 2000.
70. Kong, W.Q.; Yang, J.H. The complete chloroplast genome sequence of *Morus cathayana* and *Morus multicaulis*, and comparative analysis within genus *Morus* L. *PeerJ* **2017**, *5*, e3037. [\[CrossRef\]](#) [\[PubMed\]](#)
71. Li, H.; Xie, D.-F.; Chen, J.-P.; Zhou, S.-D.; He, X.-J. Chloroplast genomic comparison of two sister species *Allium macranthum* and *A. fasciculatum* provides valuable insights into adaptive evolution. *Genes Genom.* **2020**, *42*, 507–517. [\[CrossRef\]](#)
72. Kong, H.; Liu, W.; Yao, G.; Gong, W. A comparison of chloroplast genome sequences in *Aconitum* (Ranunculaceae): A traditional herbal medicinal genus. *PeerJ* **2017**, *5*, e4018. [\[CrossRef\]](#)
73. Sarkinen, T.; George, M. Predicting plastid marker variation: Can complete plastid genomes from closely related species help? *PLoS ONE* **2013**, *8*, e82266. [\[CrossRef\]](#)
74. Song, Y.; Dong, W.; Liu, B.; Xu, C.; Yao, X.; Gao, J.; Corlett, R.T. Comparative analysis of complete chloroplast genome sequences of two tropical trees *Machilus yunnanensis* and *Machilus balansae* in the family Lauraceae. *Front. Plant Sci.* **2015**, *6*, 662. [\[CrossRef\]](#)

75. Pang, X.; Liu, C.; Shi, L.; Liu, R.; Liang, D.; Li, H.; Cherny, S.S.; Chen, S. Utility of the *trnH-psbA* intergenic spacer region and its combinations as plant DNA barcodes: A meta-analysis. *PLoS ONE* **2012**, *7*, e48833. [[CrossRef](#)]
76. Khakhlova, O.; Bock, R. Elimination of deleterious mutations in plastid genomes by gene conversion. *Plant J.* **2006**, *46*, 85–94. [[CrossRef](#)]
77. Li, B.; Lin, F.; Huang, P.; Guo, W.; Zheng, Y. Complete chloroplast genome sequence of *Decaisnea insignis*: Genome organization, genomic resources and comparative analysis. *Sci. Rep.* **2017**, *7*, 1–10. [[CrossRef](#)] [[PubMed](#)]
78. Ivanova, Z.; Sablok, G.; Daskalova, E.; Zahmanova, G.; Apostolova, E.; Yahubyan, G.; Baev, V. Chloroplast genome analysis of resurrection tertiary relict *Haberlea rhodopensis* highlights genes important for desiccation stress response. *Front. Plant Sci.* **2017**, *8*, 204. [[CrossRef](#)]
79. Niu, Y.-T.; Jabbour, F.; Barrett, R.L.; Ye, J.-F.; Zhang, Z.-Z.; Lu, K.-Q.; Lu, L.-M.; Chen, Z.-D. Combining complete chloroplast genome sequences with target loci data and morphology to resolve species limits in *Triplostegia* (Caprifoliaceae). *Mol. Phylogenet. Evol.* **2018**, *129*, 15–26. [[CrossRef](#)] [[PubMed](#)]
80. Datwyler, S.L.; Weiblen, G.D. On the origin of the fig: Phylogenetic relationships of Moraceae from *ndhF* sequences. *Am. J. Bot.* **2004**, *91*, 767–777. [[CrossRef](#)]

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

© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).