forests

MDPI

Article

Complete Chloroplast Genome Sequence and Phylogenetic Inference of the Canary Islands Dragon Tree (*Dracaena draco* L.)

Konrad Celiński [1,*], Hanna Kijak [2] and Justyna Wiland-Szymańska [3]

1 Department of Genetics, Institute of Experimental Biology, Faculty of Biology, Adam Mickiewicz University in Poznań, Uniwersytetu Poznańskiego 6, 61-614 Poznań, Poland
2 Institute of Dendrology, Polish Academy of Sciences, Parkowa 5, 62-035 Kórnik, Poland; hkijak@man.poznan.pl
3 Department of Systematic and Environmental Botany, Institute of Environmental Biology, Faculty of Biology, Adam Mickiewicz University in Poznań, Uniwersytetu Poznańskiego 6, 61-614 Poznań, Poland; wiland@amu.edu.pl
* Correspondence: celinski@amu.edu.pl

Received: 20 January 2020; Accepted: 9 March 2020; Published: 12 March 2020

Abstract: *Dracaena draco*, which belongs to the genus *Dracaena*, is an endemic succulent of the Canary Islands. Although it is one of the most popular and widely grown ornamental plants in the world, little is known about its genomic variability. Next generation sequencing, especially in combination with advanced bioinformatics analysis, is a new standard in taxonomic and phylogenetic research. Therefore, in this study, the complete *D. draco* chloroplast genome (cp) was sequenced and analyzed in order to provide new genomic information and to elucidate phylogenetic relationships, particularly within the genus *Dracaena*. The *D. draco* chloroplast genome is 155,422 bp, total guanine-cytosine (GC) content is 37.6%, and it has a typical quadripartite plastid genome structure with four separate regions, including one large single copy region of 83,942 bp length and one small single copy region of 18,472 bp length, separated by two inverted repeat regions, each 26,504 bp in length. One hundred and thirty-two genes were identified, 86 of which are protein-coding genes, 38 are transfer RNAs, and eight are ribosomal RNAs. Seventy-seven simple sequence repeats were also detected. Comparative analysis of the sequence data of various members of Asparagales revealed mutational hotspots potentially useful for their genetic identification. Phylogenetic inference based on 16 complete chloroplast genomes of Asparagales strongly suggested that *Dracaena* species form one monophyletic group, and that close relationships exist between *D. draco*, *D. cochinchinensis* and *D. cambodiana*. This study provides new and valuable data for further taxonomic, evolutionary and phylogenetic studies within the *Dracaena* genus.

Keywords: Canary Islands dragon tree; *Dracaena*; Asparagales; taxonomy

1. Introduction

The genus *Dracaena* Vand. ex L. (Asparagaceae, Nolinoideae) consists of about 190 species [1–3]. The Dracenoid clade contains species belonging to genera, which now are incorporated into the *Dracaena* genus, but for a long time were treated as separate genera, such as *Sansevieria* Petagna with succulent leaves and mesophytic *Pleomele* Salisbury [3,4]. The morphology of the inflorescences of these two genera is very similar, but they are different from flowers of the species of dracaenas called the dragon trees [5]. This group contains 11 species, which are distinguished by succulent leaves and production of a red resin known as dragon blood [6]. These distinctive plants, which are tertiary relicts, are of great interest due to their cultural heritage to humanity, important ecological role as

umbrella species [6], and the fact that they do not form a strongly supported monophyletic clade within *Dracaena* sensu lato in molecular studies, despite similar morphology [1,4]. The first described species of the whole *Dracaena* genus and the most famous dragon tree is *Dracaena draco* L., which occurs in a subtropical climate in thermo-sclerophyllous zones in Morocco and Atlantic islands including the Canary Islands, Cabo Verde and Madeira. Due to its spectacular umbrella-like habit and excretion of a red resin, known as a dragon blood, it is called the Canary Islands dragon tree. Its natural distribution on the islands and the African continent was a subject of numerous studies [5,7–13]. Except the typical subspecies *D. draco* subsp. *draco* from the Canary Islands, two more were described: *D. draco* subsp. *ajgal* Benabid & Cuzin from Morocco [8] and *D. draco* subsp. *caboverdeana* Marrero Rodr. & R. Almeida from Cabo Verde [12]. Most previously conducted studies on *D. draco* focused on its habit, growth and morphology [5,9,12,14–19], leaf anatomy and function [20–22], formation and structure of vascular bundles [23–25] as well as root growth [26,27]. Many studies have concerned its flowers [28], pollen [29], seed propagation [13] or formation of the resin and its ecological significance [30,31]. *D. draco* being a plant icon of the Canary Islands, is widely cultivated as an ornamental plant and its resin has a great significance for human till today [31]. According to the International Union for Conservation of Nature, the current status of *D. draco* was defined as vulnerable [32].

In contrast to the works cited above, very few genetic studies have been performed on *D. draco*. Moreover, these studies were not very extensive and focused on a small number of DNA regions (mainly barcodes) [1,4,33,34]. So far, there are no more comprehensive genomic data on *D. draco*.

Next generation sequencing, especially in combination with advanced bioinformatic analysis, offers enormous opportunities and is becoming the new standard in biological research. Comparative analysis of complete chloroplast genomes (cp) not only allows advanced phylogenetic reconstruction [35,36], but also allows them to be used as barcode supercodes to distinguish and identify different taxa. This approach can be very useful especially in the study of closely related or recently divergent species [37].

The concept of using the entire chloroplast genome as a superbarcode or for reconstruction of phylogenesis is not new [38]. However, due to the decreasing costs of analysis, this approach is becoming a commonly used procedure and sets a new standard in taxonomic and phylogenetic studies, offering excellent genetic resolution. This approach also makes it possible to solve many complex taxonomic problems [39]. Furthermore, analysis of complete genomes is also recommended in studies of the genus *Dracaena* [40]. Recently, extensive genetic research has been conducted using next generation sequencing on six species of the genus *Dracaena*, mainly Asian [40]. To date, no similar genomic data have been published for *D. draco*. Therefore, the structure and organization of its chloroplast genome, as well as the level of genetic variation, remain unknown. In addition, phylogenetic relationships between different representatives of the genus *Dracaena* should include *D. draco*, and should also be clarified using data from high-performance techniques.

Therefore, our main aims in this study were: (1) sequencing, analyzing and characterizing the complete *D. draco* chloroplast genome using a next generation sequencing platform and bioinformatics tools; (2) conducting whole genome comparative analysis with published chloroplast genomes of other representatives of the genus *Dracaena*; (3) selecting mutational regions of chloroplast genome useful in identifying species in the order Asparagales; and (4) determining the position of *D. draco* within the genus based on the complete chloroplast genome sequence and phylogenetic inference.

The results obtained in this work provide a solid theoretical basis for future taxonomic and phylogenetic studies of the *Dracaena* genus. They also enable detailed analysis of *D. draco* cpDNA genetic resources and thus open the way to developing a program for the conservation of genetic resources of this important species.

2. Materials and Methods

2.1. Plant Material, DNA Extraction and Sequencing

Fresh and healthy leaves of *Dracaena draco* were collected from the Botanical Garden of the Adam Mickiewicz University in Poznań (Poland) (52° 25′N, 16° 53′E) from a specimen cultivated there since 1986 (collection number I_I005_001_0000_6986_0357) and stored at 4 °C until DNA extraction. Genomic DNA was isolated using the CTAB method [41]. The quality and integrity of isolated DNA were determined using agarose gel electrophoresis and measurement on a NanoDrop spectrophotometer (Thermo Fisher Scientific, Carlsbad, California, United States). Ion Torrent PGM libraries were made using the NEBNext Fast DNA Fragmentation & Library Prep Set for Ion Torrent (New England BioLabs), according to the manufacturer's instructions. Selection of the desired size of DNA fragments was performed by 2% agarose gel electrophoresis using the E-Gel Precast Agarose Electrophoresis System (Thermo Fisher Scientific). The library was sequenced on the Ion Torrent PGM platform located in the Laboratory of Molecular Biology, Faculty of Biology, Adam Mickiewicz University, Poznań (Poland).

2.2. Genome Assembly, Annotation and Identification of Simple Sequence Repeats

BBDuk Adapter/Quality Trimming V. 35.82 available in Geneious Prime 2020.0.49.1.7 [42] was used to filter low quality reads and trim low-quality ends and adapters. The filtered reads were de novo assembled into contigs using Geneious Assembler on default options with merging homopolymer variants. Contigs were mapped to the reference genome *Dracaena cambodiana* (NC_039776) using Geneious Mapper with minimum mapping quality: 30. Reads, which mapped to the reference genome, were used for the de novo assembly of the complete chloroplast genome of *D. draco*. The reference genome assembly database was created based on the available sequential data in the NCBI (National Center for Biotechnology Information) database for other *Dracaena* representatives, i.e., *D. angustifolia* (MN200193), *D. cambodiana* (MN200194), *D. cochinchinensis* (MN200195), *D. elliptica* (MN200196), *D. hokouensis* (MN200197) and *D. terniflora* (MN200198). Assembled genomes were initially annotated using CPGAVAS2, an integrated plastome sequence annotator [43], and GeSeq [44]. Location of large single copy region (LSC) and small single copy region (SSC) as well as calculation of GC content was done in Geneious Prime 2020.0.4 9.1.7 [42] by comparison with homologous sequences available to other *Dracaena* representatives. Transfer RNAs were also checked with tRNAscan-RE v2.0.3. [45] incorporated in GeSeq [44] using default settings. OrganellarGenomeDRAW (OGDRAW) version 1.3.1 [46] was used to draw a circular map chloroplast genome of *D. draco*. The complete *D. draco* chloroplast genome sequence has been deposited in GenBank under accession number MN990038.

Simple sequence repeats (SSRs) in the *D. draco* chloroplast genome were detected by MIcroSAtellite (MISA) [47], with parameters set at ≥10 for mononucleotides, 6≥ for dinucleotides and ≥5 for tri-, tetra-, penta- and hexanucleotides, respectively.

2.3. Genome Comparative Analysis and Identification of Divergent Hotspots

To study genome-wide evolutionary dynamics in *Dracaena* and to search evolutionary events such as gene loss, duplication, rearrangements and translocations multiple alignments were done using progressive MAUVE algorithm and default settings (Automatically calculate seed weight; Automatically calculate the minimum LCB score; Compute Locally Collinear Blocks; Full alignment; Gap Aligner: MUSCLE 3.6) implemented by MAUVE [48] plugin v1.1.1 available in Geneious Prime 2020.0.49.1.7 [42]. The complete chloroplast genome sequence of *D. draco* was employed as a reference and was compared with those previously published for other *Dracaena* representatives, i.e., *D. angustifolia* (MN200193), *D. cambodiana* (MN200194), *D. cochinchinensis* (MN200195), *D. elliptica* (MN200196), *D. hokouensis* (MN200197) and *D. terniflora* (MN200198) (Table 1) [40].

Table 1. GenBank information on complete chloroplast genomes of Asparagales species used in phylogenetic analyses in this study.

GenBank Accession	Species	Family
NC_035506	*Aloe vera*	Asphodelaceae
NC_034777	*Asparagus officinalis*	Asparagaceae
MH680946	*Convallaria keiskei*	Asparagaceae
MN200193	*Dracaena angustifolia*	Asparagaceae
MN200194	*Dracaena cambodiana*	Asparagaceae
MN200195	*Dracaena cochinchinensis*	Asparagaceae
MN990038	*Dracaena draco*	Asparagaceae
MN200196	*Dracaena elliptica*	Asparagaceae
MN200197	*Dracaena hokouensis*	Asparagaceae
MN200198	*Dracaena terniflora*	Asparagaceae
MH680945	*Liriope spicata*	Asparagaceae
KX790362	*Maianthemum bicolor*	Asparagaceae
KX931462	*Nolina atopocarpa*	Asparagaceae
KX822773	*Polygonatum stenophyllum*	Asparagaceae
MH356725	*Rohdea chinensis*	Asparagaceae
NC_032712	*Yucca filamentosa*	Asparagaceae

Identification of divergent hotspots was performed separately for representatives of the genus *Dracaena* and the Asparagales order based on seven and fifteen chloroplast genomes, respectively. The relevant chloroplast genomes were aligned using MAFFT 7.450 default option (Algorithm = Auto; Scoring matrix = 200PAM/k = 2; Gap openpenalty: 1.53 and Offset value: 0.123) [49], and then nucleotide diversity (Pi) were calculated through sliding window analysis using DnaSP version 6 [50]. The window length was set to 600 bp, with a step size 200 bp. The diversity thresholds for *Dracaena* 0.0152 and for *Asparagales* 0.0511 were calculated by sum of the average and double the standard deviation [51]. Regions with a level of nucleotide diversity higher than these thresholds were recommended as highly variable regions.

Evolutionary divergence between *Dracaena* species was estimated by calculating genetic distances using the p-distance method in MEGA X [52] using default settings (Substitution Type: Nucleotide; Substitution to Include: d: Transitions + Transversions; Rates among Sites: Uniform Rates; Gaps/Missing Data Treatment: Complete deletion) with 1000 bootstrap replicates.

2.4. Phylogenetic Inference

Phylogenetic inference was constructed by maximum likelihood (ML) and Bayesian inference (BI) analysis using 16 complete sequences of chloroplast genomes of various Asparagales representatives (including data obtained for *D. draco* in this study). The list of species included in the study along with GenBank accession numbers is given in Table 1. For a better elucidation of the tree topology, both closely related taxa, i.e., seven *Dracaena* species, and further related taxa, i.e., *Aloe vera* or *Yucca filamentosa*, were selected as the outgroups. Complete chloroplast genomes were aligned with MAFFT 7.450 using default settings (Algorithm = Auto; Scoring matrix = 200PAM/k = 2; Gap openpenalty: 1.53 and Offset value: 0.123) [49]. A General Time Reversible + Proportion Invariation + Gamma nucleotide substitution model (GTR + I + G) were selected according to Akaike's information criterion (AIC) [53] with MEGA X [52], as the best substitution model for the ML and BI analyses. The ML analyses were conducted in RaxML v8.2.11 [54], with 1000 rapid bootstrap replicates along with a search for the best-scoring ML tree in every run and parsimony random seed set to 10.

BI analyses were conducted using MrBayes v 3.2.6 [55,56]. The Markov Chain Monte Carlo (MCMC) algorithm was run for 100,000 generations and the trees were sampled every 100 generations. The remaining analysis parameters were set as follows: heated chains: 4 and heated chain temperature: 0.2; random seed: 23,364; unconstrained branch lengths: GammaDir (1;0,1;1;1) with shape parameter exponential: 10. The first 25% of the trees were discarded as a burn-in and remaining trees were used

to generate the consensus tree, including clade posterior probability (PP). *Aloe vera* was used as an outgroup. Convergence was determined by examining the average standard deviation of the split frequencies (<0.01).

3. Results and Discussion

3.1. Genome Features

The *Dracaena draco* chloroplast genome (Figure 1) has a typical quadripartite structure that is characteristic and observed in many flowering plants [35,36,57]. It contains a pair of inverted repeat regions (IRa and IRb) that comprise 26,504 bp each. The two IR regions divide the genome into a large single copy (LSC) region and a small single copy (SSC) region.

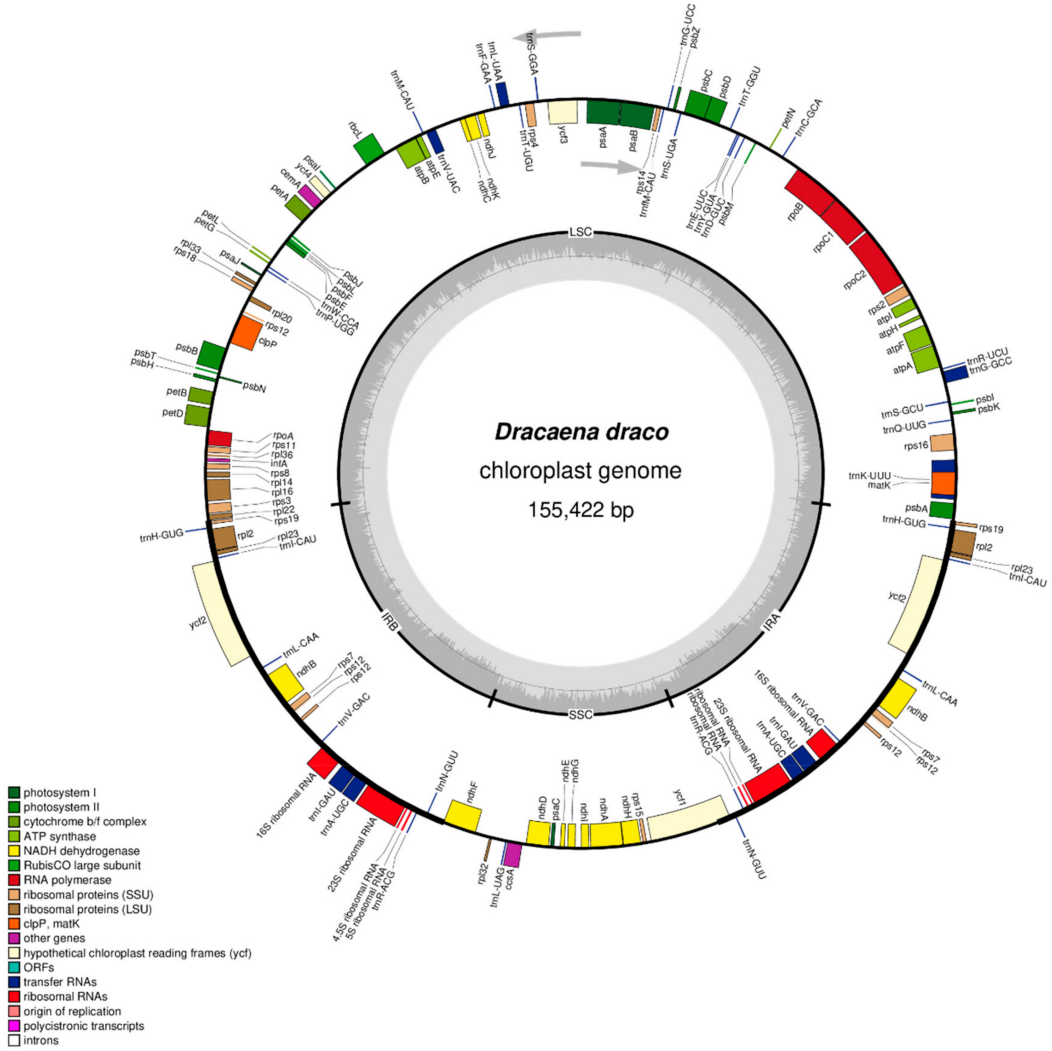

Figure 1. Map of the chloroplast genome of *D. draco*. The genes inside the circle are transcribed clockwise, and those outside are transcribed counterclockwise. Genes of different functions are color coded. The darker gray in the inner circle shows the GC content, while the lighter gray shows the AT content. IRA, IRB, inverted repeats; LSC, large single copy region; SSC, small single copy region.

The LSC region is 83,942 bp, whereas the SSC region is 18,472 bp. The complete chloroplast genome of *D. draco* is 155,422 bp in length (GenBank acc. MN990038). The total percentage of GC content is 37.6% and ranges from 31.2% in the SSC region, through 35.6% in the LSC region to 42.9% in the IR regions. Slightly higher GC content in IR regions is likely due to duplicate ribosomal RNA

genes in these regions [58]. A total of 132 genes were annotated in the *D. draco* chloroplast genome (113 unique genes excluding duplicate ones), including 86 protein-coding genes, 38 transfer RNA genes, and eight ribosomal RNA genes (Figure 1, Table 2).

Table 2. List of genes present in the chloroplast genome of *D. draco* sequenced in this study.

No.	Classification of Genes	Name of Genes	Number
1	Photosystem I	*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*	5
2	Photosystem II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*	15
3	Cytochrome b/f complex	*petA*, *petB**, *petD**, *petG*, *petL*, *petN*,	6
4	ATP synthase	*atpA*, *atpB*, *atpE*, *atpF**, *atpH*, *atpI*,	6
5	NADH dehydrogenase	*ndhA**, *ndhB**(x2), *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*	12
6	RubisCO large subunit	*rbcL*	1
7	RNA polymerase	*rpoA*, *rpoB*, *rpoC1**, *rpoC2*	4
8	Ribosomal proteins – small units (SSU)	*rps2*, *rps3*, *rps4*, *rps7*(x2), *rps8*, *rps11*, *rps12**(x2), *rps14*, *rps15*, *rps16**, *rps18*, *rps19*(x2),	15
9	Ribosomal proteins – large units (LSU)	*rpl2**(x2), *rpl14*, *rpl16**, *rpl20*, *rpl22*, *rpl23*(x2), *rpl32*, *rpl33*, *rpl36*,	11
10	Other genes/Miscellaneous	*accD*, *ccsA*, *cemA*, *clpP***, *infA*, *matK*,	6
11	Protein of unknown function	*ycf1*, *ycf2*(x2), *ycf3***, *ycf4*	5
12	Transfer RNAs	*trnA-UGC*(x2), *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnfM-CAU*, *trnG-GCC*, *trnG-UCC*, *trnH-GUG*(x2), *trnI-CAU*(x2), *trnI-GAU**(x2), *trnK-UUU**, *trnL-CAA*(x2), *trnL-UAA**, *trnL-UAG*, *trnM-CAU*, *trnN-GUU*(x2), *trnP-UGG*, *trnQ-UUG*, *trnR-ACG*(x2), *trnR-UCU*, *trnS-GCU*, *trnS-GGA*, *trnS-UGA*, *trnT-GGU*, *trnT-UGU*, *trnV-GAC*(x2), *trnV-UAC**, *trnW-CCA*, *trnY-GUA*	38
13	Ribosomal RNAs	*rrn4.5*(x2), *rrn5*(x2), *rrn16*(x2), *rrn23*(x2)	8
	Total		132

*Gene contains one intron. **Gene contains two introns. (x2) indicates that the number of the repeat unit is 2.

Our results obtained in this study are fully consistent with those previously published for other *Dracaena* representatives, i.e., *D. angustifolia* (MN200193), *D. cambodiana* (MN200194), *D. cochinchinensis* (MN200195), *D. elliptica* (MN200196), *D. hokouensis* (MN200197) and *D. terniflora* (MN200198) [40], in terms of size, length of LSC, SSC and IR regions, number of predicted genes or GC content (Table 3).

Table 3. Basic features of seven *Dracaena* chloroplast genomes.

Species	*D. draco*	*D. cochinchinensis*	*D. cambodiana*	*D. angustifolia*	*D. terniflora*	*D. hokouensis*	*D. elliptica*
Total length (bp)	155,422	155,459	155,291	155,332	155,347	155,340	155,055
LSC length (bp)	83,942	83,907	83,752	83,807	83,794	83,796	83,621
SSC length (bp)	18,472	18,492	18,489	18,465	18,493	18,494	18,456
IR length (bp)	53,008	53,050	53,050	53,060	53,060	53,050	52,978
Overall GC content (%)	37.6	37.5	37.5	37.5	37.5	37.5	37.5
Total gene number	132	130	130	130	130	130	130
Total SSR number	77	69	69	67	64	70	71
GenBank accession	MN990038	MN200195	MN200194	MN200193	MN200198	MN200197	MN200196
Reference	This study	[40]	[40]	[40]	[40]	[40]	[40]

Some studies indicate that evolution phylogenetic relationships [40,59] can be analyzed based on GC content. However, in the case of *Dracaena* species, the GC content is at a similar level (from 37.5% to 37.6%), so it is difficult to draw far-reaching conclusions about this type of phylogenetic nature only on this basis.

Simple sequence repeats (SSRs or microsatellites) are very often used in population, ecological and conservation genetics as effective molecular markers mainly due to the high level of genetic polymorphism detected by them and wide distribution throughout the whole chloroplast genome [60–63]. Initially, isolation and characterization of microsatellite markers and the development of primers amplifying these regions was quite a laborious and expensive task. Therefore, an attempt was made to design universal primers that could be used for many species on a cross-amplification basis [64,65]. Chloroplast genome sequences are generally highly conserved; hence there is a good chance that such primers amplifying repeating regions may also be used in the analysis of several other related species. Some studies support the high effectiveness of such procedures (especially in the case of chloroplast SSRs), whereas others do not (in the case of nuclear SSRs) [66,67].

In this study, a total of seventy-seven SSRs of at least 10 bp length were detected in the *D. draco* chloroplast genome. The number of SSRs detected in this study is similar to previously published results obtained for six other *Dracaena* species, in which the number of SSRs ranged from 64 for *D. terniflora* to 71 for *D. elliptica* (Table 3). Most of identified SSRs (Table 4) had a mononucleotide motif (89.61%). There were definitely fewer repeats with a di- and trinucleotide motifs (7.79% and 2.60%, respectively).

Table 4. Summary of the number of individual types of SSRs identified in cp genome of *D. draco.* (including complementary sequences).

Repeats	5	6	7	8	9	10	11	12	13	14	15	16	17	Total
A/T	-	-	-	-	-	29	16	5	12	2	1	1	2	68
C/G	-	-	-	-	-	-	-	-	1	-	-	-	-	1
AT/AT	-	1	2	3	-	-	-	-	-	-	-	-	-	6
AAT/AAT	2	-	-	-	-	-	-	-	-	-	-	-	-	2

Our results are also confirmed by the observations from other studies in which SSRs in chloroplast genomes have a motif composed mainly of short polyadenine (polyA) or polythymine (polyT) repeats and much less often contain guanine (G) or cytosine (C) tandem repeats [59]. The SSRs identified in this study can be used in the further genetic studies on *D. draco* in order to characterize genetic resources and geographical patterns of diversity.

3.2. Genome Comparative Analysis and Identification of Divergent Hotspots

The *D. draco* chloroplast genome was aligned with the chloroplast genomes of six other other *Dracaena* representatives, i.e., *D. angustifolia* (MN200193), *D. cambodiana* (MN200194), *D. cochinchinensis* (MN200195), *D. elliptica* (MN200196), *D. hokouensis* (MN200197) and *D. terniflora* (MN200198) to compare the organisation of these genomes (Figure 2). Figure 2 shows only one locally collinear block (LCB) between all analyzed chloroplast genomes, which suggest a high level of similarity in genome organisation between *Dracaena* species.

In general, whole-genome alignment of the chloroplast sequences revealed no rearrangement or inversion events among *Dracaena* chloroplast genomes and confirm the close evolutionary relationships between *Dracaena* species. However, a common break was observed in the regions (~87,000–107,000; ~130.00–150.00), characterized by the high variation in gene sequence among the aligned chloroplast genomes. Our findings are consistent with previous studies [40], which also noted high sequence homology (with a few exceptions) among *Dracaena* chloroplast genomes.

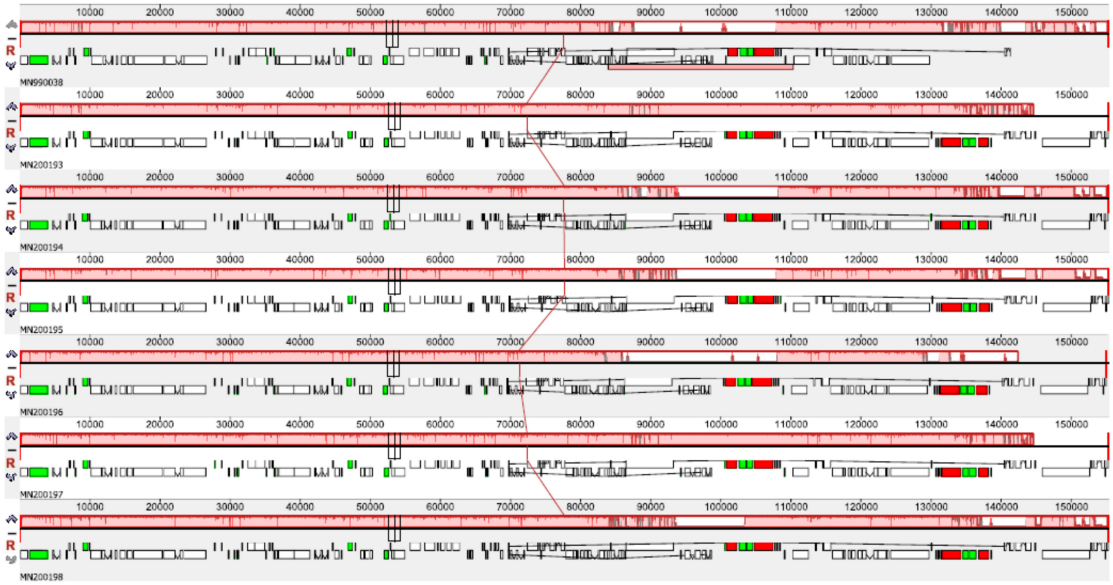

Figure 2. MAUVE alignment of seven *Dracaena* chloroplast genomes. The *D. draco* genome is shown at the top as the reference genome. Within each of the alignments, local collinear blocks are represented by blocks of the same colour connected by lines.

The p-distance values calculated as an estimator of evolutionary divergence (Table 5) differ between *Dracaena* species from 0.00147 in pair of *D. cochinchinensis* and *D. cambodiana* to 0.00548 in pair of *D. cochinchinensis* and *D. hokouensis*, with an average of 0.00420 for all seven *Dracaena* species.

Table 5. Estimates of evolutionary divergence between Dracaena species. The number of base differences per site from between sequences are shown. Standard error estimate(s) are shown above the diagonal and were obtained by a bootstrap procedure (1000 replicates). This analysis involved 7 nucleotide sequences. All positions containing gaps and missing data were eliminated (complete deletion option). There were a total of 154,129 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [52].

	D. angustifolia	*D. terniflora*	*D. hokouensis*	*D. elliptica*	*D. cambodiana*	*D. cochinchinensis*	*D. draco*
D. angustifolia		0.00011	0.00012	0.00016	0.00018	0.00018	0.00019
D. terniflora	0.00176		0.00012	0.00016	0.00018	0.00018	0.00018
D. hokouensis	0.00207	0.00214		0.00016	0.00018	0.00019	0.00018
D. elliptica	0.00396	0.00389	0.00405		0.00018	0.00019	0.00018
D. cambodiana	0.00507	0.00500	0.00533	0.00503		0.00010	0.00014
D. cochinchinensis	0.00517	0.00520	0.00548	0.00534	0.00147		0.00014
D. draco	0.00518	0.00524	0.00537	0.00520	0.00299	0.00317	

DnaSP was used to carry out two sliding window analyses in order to identify mutational regions. One analysis concerned *Dracaena* species (Figure 3A), while the other concerned representatives of Asparagales (Figure 3B). The results in Figure 3A clearly show that for *Dracaena* species there were six divergent hotspots with a high Pi value (>0.0152), i.e., *psbI-atpA*, *petA-psbJ*, *clpP*, *rps12-ndhF*, *ndhF-ccsA* and *ycf1-rps12*. For Asparagales, a total of ten unique mutational regions with a high Pi value (>0.0511) were detected, i.e., *psbK-rps16*, *rpoB-petN*, *psbM-psbD*, *ndhK-atpE*, *petA-psbJ*, *ycf2-ndhB*, *rps12-ndhF*, *ndhI-ndhA*, *ycf1* and *ndhB-ycf2* (Figure 3B). The average value of nucleotide diversity (Pi) was 0.00420 and 0.02133 for *Dracaena* and Asparagales, respectively. In summary, the Pi value was over five times higher in Asparagales than in *Dracaena*. This result is in line with expectations because the analysis included more distant species representing 10 genera and two families, not just seven closely related species from the genus *Dracaena*. However, a previous study on Asian *Dracaena* species also indicated a low level of genetic polymorphism [40].

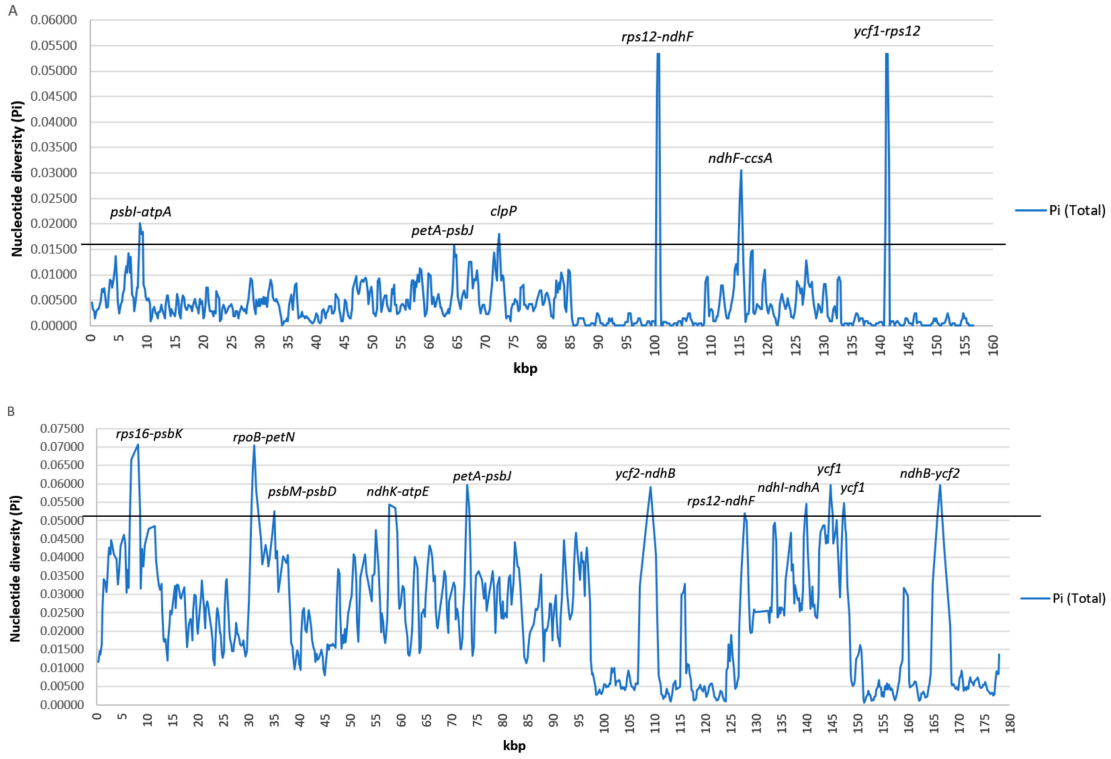

Figure 3. Sliding window analysis of the whole chloroplast genomes. Window length: 600 bp; step size: 200 bp. *X*-axis: position of the midpoint of a window. *Y*-axis: nucleotide diversity of each window. (**A**) Pi between seven *Dracaena* species. (**B**) Pi among all 16 representatives of Asparagales. The black horizontal line on the graph sets the threshold separately for *Dracaena* (0.0152) and separately for Asparagales (0.0511).

The chloroplast DNA regions selected in this study can be used as specific barcodes dedicated for further taxonomic studies of Asparagales members. A species-specific barcode is defined as a fragment of DNA sequence that has a sufficiently high mutation rate to enable species identification within a given taxonomic group [38]. The *ycf1* region seems particularly interesting in this respect for Asparagales. Several studies show that this region has a high discriminatory power and much more potential than universal barcodes, including for species other than *Dracaena* [37,68,69].

3.3. Phylogenetic Inference

Chloroplast genome sequences have been successfully used to study the phylogeny of many different plant groups [35,36,40,57]. In this study, we were particularly interested in the position of *D. draco* within the genus *Dracaena*. Therefore, phylogenetic trees were constructed using ML and Bayesian algorithms and nucleotide sequences of chloroplast genomes of sixteen Asparagales representatives. As shown in Figure 4, both obtained ML and Bayesian phylogenetic trees clearly indicated that all species representing the *Dracaena* genus formed a separate cluster within Asparagales. However, this cluster consists of two groups, confirmed by a very high probability. One on them includes *D. draco* obtained in this study and, originating from one node, a sister clade composed of *D. cochinchinensis* and *D. cambodiana*. The second comprises *D. hokouensis*, *D. elliptica* and the sister clade of *D. terniflora* and *D. angustifolia*. The second major group also consists of two smaller clades with representatives of six other genera: *Convallaria*, *Liriope*, *Maianthemum*, *Nolina*, *Polygonum* and *Rohdea*. The first sister clade, which includes *Convallaria keiskei* and *Rohdea chinensis*, is characteristic of both trees. The other species form different sister groups depending on the method. In the Bayesian tree *Liriope spicata* groups in a sister clade with *Nolina atopocarpa*, while in the ML tree *L. spicata* is sister to *Maianthemum bicolor* and *N. atopocarpa* is sister to *Polygonatum enophyllum*. It is worth noting that

the probability of a correct tree solution is higher in the Bayesian method than in ML. However, to discuss the phylogenetic relationship of these genera, a larger number of representatives is needed. Two species of the Asparagaceae family, i.e., *Asparagus officinalis* (genus *Asparagus*) and *Yucca filamentosa* (genus *Yucca*), were outside the two main distinguished groups both in ML and Bayesian trees.

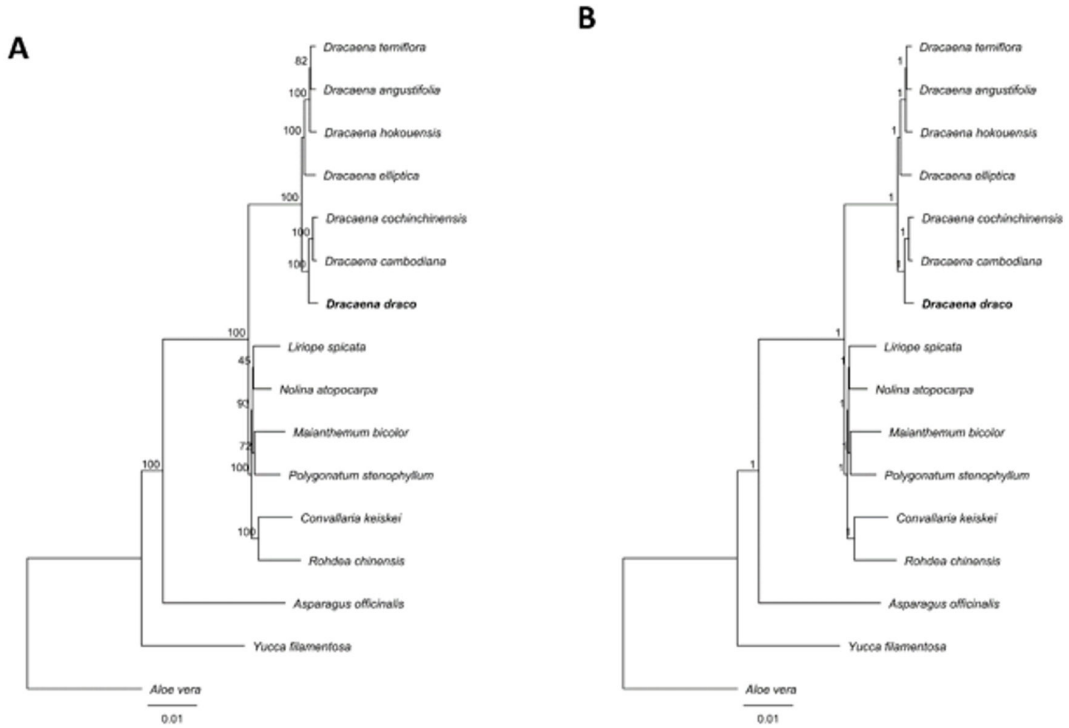

Figure 4. Phylogenetic relationships among the 16 Asparagales species based on complete chloroplast genomes inferred from: Maximum Likelihood (**A**) and Bayesian inference (**B**). The numbers near each node are bootstrap support values obtained by ML (**A**) or show the posterior probabilities according to BI (**B**).

Our results are consistent with previous phylogenetic studies of the *Dracaena* genus, both those based on complete chloroplast genomes but not including *D. draco* [40] and those based on selected cpDNA regions in which *D. draco* was included [34]. Our analyses increase the knowledge of *Dracaena's* phylogeny and provide valuable genetic information for future research into the evolutionary history of this group.

4. Conclusions

In this study, we reported and analyzed the complete chloroplast genome of *D. draco*. The structure of the chloroplast genome, its organization, length as well as the order and number of genes are similar to those recently published for six Asian *Dracaena* species. Due to the high level of polymorphism and its length, the *ycf1* region appears to be very useful for identifying taxa within Asparagales, though not necessarily of the genus *Dracaena*. Conducted phylogenetic analyses revealed that *D. draco* is much closer to *D. cochinchinensis* and *D. cambodiana* than to *D. terniflora*, *D. angustifolia*, *D. hokouensis* and *D. elliptica*.

Author Contributions: K.C., H.K., and J.W.-S. conceived of and designed the research framework; K.C. and H.K. performed most of the experiments and data analysis; K.C. and J.W.-S. wrote the original draft manuscript as well as reviewing and editing the final manuscript; H.K. assembled and annotated the genome; J.W.-S. collected the sample. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Faculty of Biology of the Adam Mickiewicz University, Poznań, Poland. The APC was funded by the Faculty of Biology of the Adam Mickiewicz University, Poznań, Poland.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Lu, P.-L.; Morden, C.W. Phylogenetic Relationships among Dracaenoid Genera (Asparagaceae:Nolinoideae) Inferred from Chloroplast DNA Loci. *Syst. Bot.* **2014**, *39*, 90–104. [CrossRef]
2. Angiosperm Phylogeny Group. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* **2016**, *181*, 1–20. [CrossRef]
3. Govaerts, R.; Zonneveld, B.J.M.; Zona, S.A.; World Checklist of Asparagaceae. Facilitated by the Royal Botanic Gardens, Kew. Available online: http://apps.kew.org/wcsp/ (accessed on 22 December 2019).
4. Takawira-Nyenya, R.; Mucina, L.; Cardinal-Mcteague, W.; Thiele, K. *Sansevieria* (Asparagaceae, 1101 Nolinoideae) is a herbaceous clade within *Dracaena*: Inference from non-coding plastid and nuclear DNA 1102 sequence data. *Phytotaxa* **2018**, *376*, 254–276. [CrossRef]
5. Marrero, A.; Almeida, S.R.; Martín-González, M. A new species of the wild Dragon Tree, *Dracaena* (Dracaenaceae) from Gran Canaria and its taxonomic and biogeographic Implications. *Bot. J. Linn. Soc.* **1998**, *128*, 291–314.
6. Maděra, P.; Forrest, A.; Hanáček, P.; Vahalík, P.; Gebauer, R.; Plichta, R.; Jupa, R.; Rensburg, J.J.V.; Morris, M.; Nadezhdina, N.; et al. What We Know and What We Do Not Know About Dragon Trees? *Forests* **2020**, *11*, 236. [CrossRef]
7. Turland, N.J. Agavaceae. In *Flora of Madeira*; Press, J.R., Short, M.J., Eds.; Natural History Museum (HMSO): London, UK, 1995; pp. 391–392.
8. Benabid, A.; Cuzin, F. Populations de dragonnier (*Dracaena draco* L. subsp. aigal Benabid et Cuzin) au Maroc: Valeurs taxinomique, biogéographique et phytosociologique. *C. R. Acad. Sci. Paris Sci. Vie* **1997**, *320*, 267–277. [CrossRef]
9. Marrero, A. *Dracaena tamaranae*, el género dracaena y otros afines: Análisis morfológico para un aproximación filogenética. *Museo Canario* **2000**, *55*, 301–334.
10. Almeida Pérez, R.S. Sobre la presencia de *Dracaena draco* (L.) L. En gran Canaria (Islas Canarias): Aportación corológica, estado actual y significación biogeográfica. *Bot. Macarónesica* **2003**, *24*, 17–38.
11. Almeida Pérez, R.S. Dracaena draco (L.). In *Atlas y Libro Rojo de la Flora Vascular Amenazada de España*, 2nd ed.; Bañares, A., Blanca, G., Güemes, J., Moreno, J.C., Ortiz, S., Eds.; Publicaciones de O.A.P.N.: Madrid, Spain, 2004; pp. 680–681.
12. Marrero, A.; Almeida, S.R. A new subspecies, *Dracaena draco* (L.) L. subsp. caboverdeana Marrero Rodr. & R. Almeida (Dracaenaceae) from Cape Verde Island. *Int. J. Geobot. Res.* **2012**, *2*, 35–40.
13. González-Castro, A.; Pérez-Pérez, D.; Romero, J.; Nogales, M. Unraveling the Seed Dispersal System of an Insular "Ghost" Dragon Tree (*Dracaena draco*) in the Wild. *Front. Ecol. Evol.* **2019**, *7*, 39. [CrossRef]
14. Pütter, A. Altersbestimmung an Drachenbäumen von Tenerife. Sitzungsberichte der Heidelberger Akademie der Wissenschäften. *Math.-Nat. Klasse* **1925**, *12*, 12–18.
15. Byström, K. *Dracaena draco* L. in the Cape Verde Islands. *Acta Horti-Gotobg.* **1960**, *23*, 179–214.
16. Symon, D.E. The growth of *Dracaena draco*—dragon's blood tree. *J. Arnold Arbor.* **1974**, *55*, 51–58.
17. Mägdefrau, K. Das Alter der Drachenbäume auf Tenerife. *Flora* **1975**, *164*, 347–357. [CrossRef]
18. Beyhl, F.E. Two different growth forms of *Dracaena draco* L. (Monocotyledones: Liliales: Agavaceae). *Boletin Museu Municipal Funchal* **1995**, *4*, 91–95.
19. Krawczyszyn, J.; Krawczyszyn, T. Photomorphogenesis in *Dracaena draco*. *Trees Struct. Funct.* **2016**, *30*, 647–664. [CrossRef]
20. Wiland-Szymańska, J.; Klimko, M. Differentiation of leaf anatomy of the genera *Dracaena* L and *Sansevieria* Thunb. (Dracaenaceae). In Proceedings of the XVII International Botanical Congress. 100 years after the II IBC in Vienna 1905, Vienna, Austria, 12–16 July 2005; p. 328.

21. Nadezhdina, N.; Plichta, R.; Nadezhdin, V.; Gebauer, R.; Jupa, R.; Habrová, H.; Maděra, P. A comparative structural and functional study of leaf traits and sap flow in *Dracaena cinnabari* and *Dracaena draco* seedlings. *Funct. Plant Biol.* **2015**, *42*, 1092–1105. [CrossRef]
22. Klimko, M.; Nowińska, R.; Wilkin, P.; Wiland-Szymańska, J. Comparative leaf micromorphology and anatomy of the dragon tree group of *Dracaena* (Asparagaceae) and their taxonomic implications. *Plant Syst. Evol.* **2018**. [CrossRef]
23. Jura-Morawiec, J. Formation of amphivasal vascular bundles in *Dracaena draco* stem in relation to rate of cambial activity. *Trees Struct. Funct.* **2015**, *29*, 1493–1499. [CrossRef]
24. Jura-Morawiec, J.; Wiland-Szymańska, J. A novel insight into the structure of amphivasal secondary bundles on the example of *Dracaena draco* L. stem. *Trees Struct. Funct.* **2014**, *28*, 871–877. [CrossRef]
25. Jura-Morawiec, J. Atypical origin, structure and arrangement of secondary tracheary elements in the stem of the monocotyledonous dragon tree, *Dracaena draco*. *Planta* **2017**, *245*, 93–99. [CrossRef] [PubMed]
26. Jura-Morawiec, J. Rhythmic growth and age estimation of aerial roots in *Dracaena draco* (Asparagaceae). *Trees* **2019**, *33*, 1513–1518. [CrossRef]
27. Krawczyszyn, J.; Krawczyszyn, T. Massive aerial roots growth and form of *Dracaena draco*. *Trees Struct. Funct.* **2014**, *28*, 757–768. [CrossRef]
28. Brown, N.E. Notes on the genera *Cordyline*, *Dracaena*, *Pleomele*, *Sansevieria*, and *Taetsia*. *Bull. Misc. Inf.* **1914**, *8*, 273–279. [CrossRef]
29. Klimko, M.; Nowińska, R.; Jura-Morawiec, J.; Wiland-Szymańska, J.; Wilkin, P. Pollen morphology of selected species of the genera *Chrysodracon* and *Dracaena* (Asparagaceae, subfamily Nolinoideae) and its systematic implications. *Plant Syst. Evol.* **2018**. [CrossRef]
30. Jura-Morawiec, J.; Tulik, M. Morpho-anatomical basis of dragon's blood secret in *Dracaena draco* stem. *Flora* **2015**, *213*, 1–5. [CrossRef]
31. Jura-Morawiec, J.; Tulik, M. Dragon's blood secretion and its ecological significance. *Chemoecology* **2016**, *26*, 101–105. [CrossRef]
32. Walter, K.S.; Gillett, H.J. (Eds.) Dracaena draco. In *1997 IUCN Red List of Threatened Plants*; IUCN: Gland, Switzerland, 1998.
33. Lu, P.L.; Morden, C. Phylogenetics of the plant genera *Dracaena* and *Pleomele* (Aparagaceae). *Bot. Orient. J. Plant Sci.* **2010**, *7*, 64–72. [CrossRef]
34. Edwards, C.E.; Bassüner, B.; Birkinshaw, C.; Camara, C.; Lehavana, A.; Lowry, P.P.; Miller, J.S.; Wyatt, A.; Jackson, P.W. A botanical mystery solved by phylogenetic analysis of botanical garden collections: The rediscovery of the presumed-extinct *Dracaena umbraculifera*. *Oryx* **2018**, *52*, 427–436. [CrossRef]
35. Li, D.M.; Zhao, C.Y.; Liu, X.F. Complete chloroplast genome sequences of *Kaempferia galanga* and *Kaempferia elegans*: Molecular structures and comparative analysis. *Molecules* **2019**, *24*, 474. [CrossRef]
36. Vu, H.T.; Tran, N.; Nguyen, T.D.; Vu, Q.L.; Bui, M.H.; Le, M.T.; Le, L. Complete chloroplast genome of *Paphiopedilum delenatii* and phylogenetic relationships among Orchidaceae. *Plants* **2020**, *9*, E61. [CrossRef] [PubMed]
37. Celiński, K.; Kijak, H.; Wojnicka-Półtorak, A.; Buczkowska-Chmielewska, K.; Sokołowska, J.; Chudzińska, E. Effectiveness of the DNA barcoding approach for closely related conifers discrimination: A case study of the *Pinus mugo* complex. *Comptes Rendus Biol.* **2017**, *340*, 339–348. [CrossRef] [PubMed]
38. Li, X.; Yang, Y.; Henry, R.J.; Rossetto, M.; Wang, Y.; Chen, S. Plant DNA barcoding: From gene to genome. *Biol. Rev. Camb. Philos. Soc.* **2015**, *90*, 157–166. [CrossRef] [PubMed]
39. Parks, M.; Cronn, R.; Liston, A. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol.* **2009**, *7*, 84. [CrossRef] [PubMed]
40. Zhang, Z.; Zhang, Y.; Song, M.; Guan, Y.; Ma, X. Species identification of *Dracaena* using the complete chloroplast genome as a super-barcode. *Front. Pharmacol.* **2019**. [CrossRef]
41. Doyle, J.J.; Doyle, J.L. Isolation of plant DNA from fresh tissue. *Focus* **1990**, *12*, 13–15.
42. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [CrossRef]
43. Shi, L.; Chen, H.; Jiang, M.; Wang, L.; Wu, X.; Huang, L.; Liu, C. CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res.* **2019**, *47*, W65–W73. [CrossRef]

44. 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]
45. Chan, P.P.; Lowe, T.M. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. *Methods Mol. Biol.* **2019**, *1962*, 1–14.
46. 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] [PubMed]
47. 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]
48. Darling, A.E.; Mau, B.; Perna, N.T. Progressive Mauve: Multiple Genome Alignment with Gene Gain, Loss, and Rearrangement. *PLoS ONE* **2010**, *5*, e11147. [CrossRef]
49. 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] [PubMed]
50. 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 Datasets. *Mol. Biol. Evol.* **2017**, *34*, 3299–3302. [CrossRef] [PubMed]
51. Bi, Y.; Zhang, M.-F.; Xue, J.; Dong, R.; Du, Y.-P.; Zhang, X.-H. Chloroplast genomic resources for phylogeny and DNA barcoding: A case study on Fritillaria. *Sci. Rep.* **2018**, *8*, 1184. [CrossRef]
52. 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]
53. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **1974**, *19*, 716–723. [CrossRef]
54. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [CrossRef]
55. Huelsenbeck, J.P.; Ronquist, F. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* **2001**, *17*, 754–755. [CrossRef]
56. Ronquist, F.; Huelsenbeck, J.P. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **2003**, *19*, 1572–1574. [CrossRef] [PubMed]
57. Ge, J.; Cai, L.; Bi, G.Q.; Chen, G.; Sun, W. Characterization of the complete chloroplast genomes of *Buddleja colvilei* and *B. sessilifolia*: Implications for the taxonomy of *Buddleja* L. *Molecules* **2018**, *23*, E1248. [CrossRef] [PubMed]
58. Li, Y.; Sylvester, S.P.; Li, M.; Zhang, C.; Li, X.; Duan, Y.; Wang, X. The Complete Plastid Genome of *Magnolia zenii* and Genetic Comparison to *Magnoliaceae* species. *Molecules* **2019**, *24*, 261. [CrossRef] [PubMed]
59. Nikbakht, H.; Xia, X.; Hickey, D.A. The evolution of genomic GC content undergoes a rapid reversal within the genus *Plasmodium*. *Genome* **2014**, *57*, 507–511. [CrossRef] [PubMed]
60. Ellegren, H. Microsatellites: Simple sequences with complex evolution. *Nat. Rev. Genet.* **2004**, *5*, 435–445. [CrossRef] [PubMed]
61. Oliveira, E.J.; Pádua, J.G.; Zucchi, M.I.; Vencovsky, R.; Vieira, M.L.C. Origin, evolution and genome distribution of microsatellites. *Genet. Mol. Biol.* **2006**, *29*, 294–307. [CrossRef]
62. Gómez, A.; González-Martínez, S.C.; Collada, C.; Climent, J.; Gil, L. Complex population genetic structure in the endemic Canary Island pine revealed using chloroplast microsatellite markers. *Theor. Appl. Genet.* **2003**, *107*, 123–131. [CrossRef]
63. Urbaniak, L.; Wojnicka-Półtorak, A.; Celiński, K.; Lesiczka, P.; Pawlaczyk, E.; Aučina, A. Genetic resources of relict populations of *Pinus sylvestris* (L.) in Western Carpathians assessed by chloroplast microsatellites pine revealed using chloroplast microsatellite. *Biologia* **2019**, *74*, 1077–1086. [CrossRef]
64. Vendramin, G.G.; Lelli, L.; Rossi, P.; Morgante, M. A set of primers for the amplification of 20 chloroplast microsatellites in Pinaceae. *Mol. Ecol.* **1996**, *5*, 595–598. [CrossRef]
65. Echt, C.S.; Vendramin, G.G.; Nelson, C.D.; Marquardt, P. Microsatellite DNA as shared genetic markers among conifer species. *Can. J. For. Res.* **1999**, *29*, 365–371. [CrossRef]
66. Celiński, K.; Pawlaczyk, E.M.; Wojnicka-Półtorak, A.; Chudzińska, E.; Prus-Głowacki, W. Cross-species amplification and characterization of microsatellite loci in *Pinus mugo* Turra. *Biologia* **2013**, *68*, 621–626. [CrossRef]

67. González-Martínez, S.C.; Robledo-Arnuncio, J.J.; Collada, C.; Díaz, A.; Williams, C.G.; Alía, R.; Cervera, M.T. Cross-amplification and sequence variation of microsatellite loci in Eurasian hard pines. *Theor. Appl. Genet.* **2004**, *109*, 103–111. [CrossRef] [PubMed]
68. 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] [PubMed]
69. Olsson, S.; Grivet, D.; Cid-Vian, J. Species-diagnostic markers in the genus Pinus: Evaluation of the chloroplast regions *matK* and *ycf1*. *For. Syst.* **2018**, *27*, e016. [CrossRef]