

Article

Comparative Plastomics of Plantains (*Plantago*, Plantaginaceae) as a Tool for the Development of Species-Specific DNA Barcodes

Furrukh Mehmood^{1,2}, Mingai Li^{1,2}, Alessio Bertolli³ , Filippo Prosser³ and Claudio Varotto^{1,2,*} 

¹ Ecogenomics Unit, Research and Innovation Centre, Fondazione Edmund Mach, 38098 San Michele all'Adige, Italy

² National Biodiversity Future Center (NBFC), 90133 Palermo, Italy

³ Fondazione Museo Civico di Rovereto, 38068 Rovereto, Italy

* Correspondence: claudio.varotto@fmach.it

Abstract: *Plantago* (plantains, Plantaginaceae) is a cosmopolitan genus including over 250 species used as functional foods, forage, and traditional medicine. Among them, *Plantago lanceolata* is commonly used as an ingredient of herbal products, but the close similarity to other *Plantago* species can cause misidentifications with potentially serious consequences for product safety/quality. To test the possibility of developing species-specific barcoding markers, we de novo assembled plastome sequences of individuals of *Plantago argentea*, *Plantago atrata*, *P. lanceolata*, and *Plantago maritima*. These genomes were characterized in comparison with both previously sequenced conspecific accessions and other publicly available plastomes, thus providing an assessment of both intraspecific and interspecific genetic variation in *Plantago* plastomes. Additionally, molecular evolutionary analyses indicated that eleven protein-coding genes involved in different plastid functions in *Plantago* plastomes underwent positive selection, suggesting they might have contributed to enhancing species' adaptation during the evolutionary history of *Plantago*. While the most variable mutational hotspots in *Plantago* plastomes were not suitable for the development of species-specific molecular markers, species-specific polymorphisms could discriminate *P. lanceolata* from its closest relatives. Taken together, these results highlight the potential of plastome sequencing for the development of molecular markers to improve the identification of species with relevance in herbal products.

Keywords: plantains; plastome; indels; medicinal plants; mutational hotspots; phylogeny; Plantaginaceae; substitutions; *Plantago*; molecular markers



Citation: Mehmood, F.; Li, M.; Bertolli, A.; Prosser, F.; Varotto, C. Comparative Plastomics of Plantains (*Plantago*, Plantaginaceae) as a Tool for the Development of Species-Specific DNA Barcodes. *Plants* **2024**, *13*, 2691. <https://doi.org/10.3390/plants13192691>

Academic Editor: Khalid Meksem

Received: 22 July 2024

Revised: 13 September 2024

Accepted: 18 September 2024

Published: 25 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Plantago is a cosmopolitan genus that has diversified into over 250 species, which are usually anemophilous herbs or rarely subshrubs, perennial or annual, and concentrated in temperate and high-elevation tropical regions [1–3]. The taxonomy of *Plantago* is notoriously difficult, owing to the relatively few morphological characters available and its evolutionary trend toward morphological reduction and simplification [4,5]. Traditionally, several morphological features, including trichomes and seeds, as well as chemotaxonomic analyses [6–8], have been employed in attempts to identify and classify the species. However, none of these methods have yielded a conclusive result that is considered satisfactory, not even with 91 mainly morphological and embryological characters [2]. Furthermore, the existence of polyploidy [9,10], hybridization [10], and reticulate evolution [11] adds more difficulty to the understanding of evolutionary relationships in *Plantago*. Not surprisingly, therefore, early attempts aimed at determining the phylogenetic relationships among *Plantago* species with either morphological or molecular data from a limited number of loci provided strong support for the sister relationship between *Plantago* and *Littorella*, but the infrageneric relationships were still not fully resolved [2,11]. Only recently have phylogenomic studies adopting whole-chloroplast genome sequencing succeeded in the

full resolution of the subgeneric structure, resulting in the recognition of four highly supported subgenera in the genus *Plantago*: *Bougueria* (Decne.) Rahn (previously regarded as a distinct monotypic genus; [2,12]), *Coronopus* (Lam. & DC.) Rahn, *Plantago* and *Psyllium* (Mill.) Harms & Reiche [4,13].

Some *Plantago* species hold significant value in the nutraceutical and pharmaceutical sectors because of the mucilaginous substance, psyllium, obtained from the seed husk. This product is recognized for its use as a functional food and dietary supplement, particularly for enhancing intestinal health [14,15]. In some regions, various *Plantago* species are used as ingredients in salads, soups, or baked products [16,17]. Additionally, certain species are utilized in animal feed to enhance health and reduce the reliance on antibiotics [18].

P. lanceolata is a common forb, native to Europe and Western Asia, with a cosmopolitan distribution and high adaptability [19–22]. The aerial parts of *P. lanceolata* exhibit properties that include wound-healing, anti-inflammatory, antibacterial, diuretic, and anti-asthmatic effects [23]. The consumption of juice extracted from *P. lanceolata*, when combined with wine or honey, is believed to provide relief from gout symptoms. Additionally, the crushed leaves of the plant, when mixed with salt, have been traditionally employed to treat arthritis symptoms [24]. At present, *P. lanceolata* is utilized in the treatment of mouth, throat, and upper respiratory tract conditions, as well as for topical application in skin diseases [25]. Many products, including teas and syrups, containing *P. lanceolata* have been introduced to the market, with a significant presence in Europe. The herb is recognized as safe for use in the pharmacopeias of several countries [26].

The market for herbal products has become increasingly prevalent around the world in recent decades, with the worldwide market for medicinal plants projected to reach a value of USD 5 trillion by 2050, with Europe being a significant driving force behind this growth [27,28]. Plantain (*Plantago* species) products, however, still show qualitative variability through the herbal medicinal value chain, affecting their impact and safety. A notable instance of plant misidentification, for instance, was documented in a study that detailed the cases of two patients who presented to hospital emergency rooms with symptoms such as palpitations, vomiting, nausea, and chest pressure, among other symptoms [29]. The investigation revealed that both these patients consumed raw material labeled as plantain (genus *Plantago*) that had been contaminated with *Digitalis lanata* due to misidentification in the field. The quality of manufactured herbal products is notoriously variable on a global scale, with significant inconsistencies in the quality of these products observed across different regions and suppliers. To address this issue, consistent and rigorous analytical approaches are necessary to accurately identify and monitor the quality of herbal products throughout the entire value chain [30]. The herbal medicine industry has considered DNA barcoding as a method that can be consistently applied in quality control over manufactured products and to identify medicinal materials to protect consumers from dishonest suppliers. In addition, this method can also be used to identify toxic herbal materials in life-threatening situations, prevent poisoning, and improve control procedures of herbal drug substances [3].

The structure and sequence of the chloroplast genome (or plastome) can be utilized to generate molecular markers that can be used in DNA barcoding [31]. Plastomes have been termed “super barcodes”, due to their comparatively conserved organization, gene content, adequate level of nucleotide substitution in protein-encoding genes, and uniparental inheritance, which make them excellent sources of genetic information for phylogenetic reconstruction and species identification at diverse taxonomic levels [32–35]. Currently, whole-chloroplast genomes are available from the National Centre for Biotechnology Information (NCBI) Organelle Genome Resources (accessed on 3 June 2024) and TreeBase (<https://www.treebase.org/treebase-web/home.html>; accessed on 3 June 2024) for approximately 110 of the about 250 *Plantago* species. While for about 21% of the species, multiple plastomes are available (for instance, in the case of *P. lanceolata*, for which five plastome sequences have been deposited), the intraspecific genetic variation for most species has not been determined. In the subgenus *Coronopus*, seven species have fully sequenced

plastomes, including two accessions of *P. maritima* [4,13]. In the *Psyllium* subgenus, the plastomes of 11 species have been sequenced, among which two (*P. lanceolata* and *Plantago lagopus*) belong to section *Lanceifolia*, and one (*P. atrata*) from the sister section *Montana* [13]. Former phylogenies based on ITS and *trnL-F* sequence data retrieved with good statistical support the narrow endemic *Plantago leiopetala* from the islands of Madeira and Porto Santo as a sister to *P. lanceolata* [12,36]. In an unpublished study with higher taxonomic sampling, however, this relationship was only weakly supported [37]. Interestingly, in the latter study, a highly supported clade of European and Mediterranean species formed by *P. lanceolata*, *P. leiopetala*, *P. plantago loeflingii*, *P. argentea*, and *Plantago altissima* was recovered, although the topology among these species lacked significant support [37]. In the published version of the work, only *P. argentea* was included in the analyses and five subgenera were recognized [5].

Here, we aimed to explore in greater depth the intra- and interspecific diversity of the chloroplast genomes of some of the species most closely related to *P. lanceolata* and to apply this knowledge to develop highly robust and reliable barcoding markers for this important plantain species. With this goal in mind, we sequenced, assembled, and annotated the complete plastomes of single accessions of *P. argentea*, *P. atrata*, *P. lanceolata*, and *P. maritima* co-occurring in the same region of the Eastern Alps. By comparing the newly sequenced plastome accessions with those present in public databases, in addition to presenting a novel assessment of the evolutionary patterns shaping the selective landscape of *Plantago* plastomes and reassessing the tempo of the genus evolution, we addressed the placement of the novel *P. argentea* plastome in the *Plantago* phylogeny and particularly with respect to *P. lanceolata*.

2. Results

2.1. Organization and Characteristics of *Plantago* Plastid Genomes

Our comparative analysis confirmed that *Plantago* species have similar plastid genome structures (Figure 1A and Table 1).

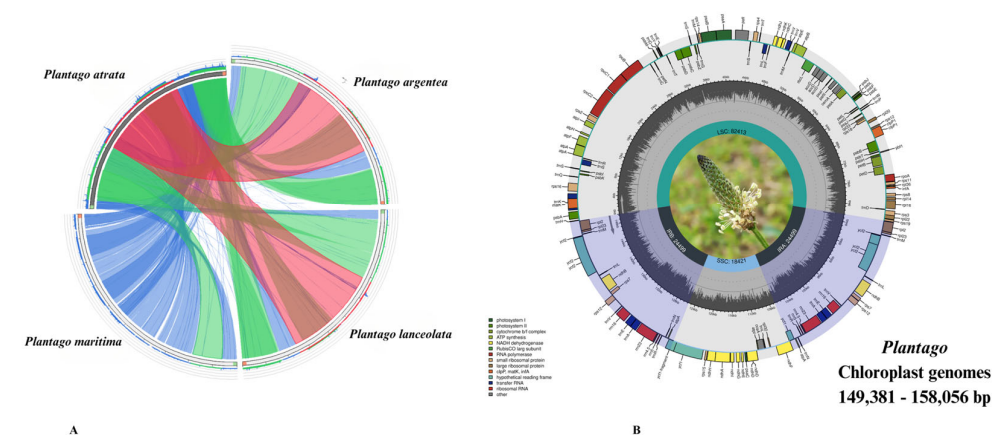


Figure 1. (A) Structural comparison of the four studied *Plantago* plastomes showing a high level of synteny and the lack of large rearrangements. The start and end points of the sequences are marked by green and orange blocks. The colored blocks outside the sequences refer to the score/max bit score ratio, with green ≤ 0.50 , orange ≤ 0.75 , and red > 0.75 . Blue blocks and chords represent the inverted repeats (IRs). (B) Consensus circular genome map of four studied *Plantago* plastomes. Genes drawn inside the circle are transcribed counterclockwise and those outside are clockwise. Different colors indicate the genes belonging to various functional groups. GC and AT content of the genome are plotted in light grey and dark, respectively, in the inner circle. Large single copy (LSC), inverted repeat A (IRa), and inverted repeat B (IRb) highlighted with color and small single copy (SSC) are shown in the circular diagram.

Table 1. Comparison and general features of chloroplast genomes of *P. argentea*, *P. atrata*, *P. lanceolata*, and *P. maritima*.

Characteristics	<i>P. argentea</i>	<i>P. atrata</i>	<i>P. lanceolata</i>	<i>P. maritima</i>	
Accession No.	PP541855	PP541856	PP541857	PP5541858	
Coverage depth	103×	50×	160×	80×	
Size (bp)	150,092	149,381	149,832	158,056	
LSC length (bp)	82,489	82,680	82,413	81,909	
SSC length (bp)	18,421	17,671	18,421	8665	
IR length (bp)	24,591	24,515	24,499	33,741	
Number of genes	133	133	133	137	
Protein-coding genes	86	86	86	90	
tRNA genes	38	38	38	39	
rRNA genes	8	8	8	8	
Duplicate genes	18	18	18	22	
GC content	Total (%)	38.3%	38.3%	38.4%	38.6%
	LSC (%)	36.5%	36.5%	36.6%	37%
	SSC (%)	32.5%	32.3%	32.6%	32.8%
	IR (%)	43.5%	43.5%	43.6%	41.4%
	CDS (%)	38.5%	38.5%	38.5%	38.7%
	rRNA (%)	55.2%	55.2%	55.2%	55.2%
	tRNA (%)	52.8%	52.8%	52.9%	53.8%
	All gene %	40.2%	40.2%	40.2%	40.4%
Protein coding part (CDS) (%bp)	51.7%	51.6%	51.7%	53.9%	
All gene (%bp)	72.5%	72.4%	72.6%	73%	
Non-coding region (%bp)	27.5%	27.6%	27.4%	27%	

All four *Plantago* plastomes exhibited a typical quadripartite structure containing a pair of inverted repeats (IR) regions (24,499–33,741 bp), an LSC region (81,909–82,680 bp) and an SSC region (8665–18,421 bp). The full-length variation between these genomes was about 8.6 kb (genome size: 149,381–158,056 bp). The assembled plastomes of *P. argentea*, *P. atrata*, *P. lanceolata*, and *P. maritima* had average coverage depths of 103×, 50×, 160×, and 80×, respectively. The total guanosine-cytosine (GC) content of the *de novo* assembled *Plantago* plastid genomes was 38.3%, as was that in the previously sequenced species. The GC content of the IR region was higher (43.6%) than the LSC (36.6%) and SSC (32.6%) regions, which could have been due to the occurrence of rRNA genes, which are known to contain GC-rich regions [38,39]. The plastomes of the *de novo* assembled *Plantago* accessions had 133–137 unique genes, whereas 18–22 genes were duplicated in the IR (Table 2 and Figure 1B).

Out of these 133–137 genes, 86–90 were protein-encoding, 38–39 were tRNA, and 8 rRNA genes were similarly present in all *Plantago* species. The IR regions contained 18–22 duplicated genes and out of these 7–11 were protein-encoding, 4 were rRNA, and 7 were tRNA genes. In all four studied *Plantago* species, the *ycf3* protein-coding gene contains two introns. Within the subgenus *Psyllium* (*P. argentea*, *P. atrata*, and *P. lanceolata*), the genes *rps16*, *atpF*, *rpoC1*, *clpP*, *petB*, *petD*, *rpl16*, *rpl2*, *ndhB*, *rps12*, and *ndhA* each contain one intron. However, in *P. maritima* (subgenus *Coronopus*), the only difference observed is the loss of introns in the *rpl2* and *clpP* genes as compared to the subgenus *Psyllium*. The *rps12* gene was trans-spliced with its 5'-end exons located in the LSC, while its two 3'-end exons were found in the IR. The tRNAs (52.8%) and rRNAs (55.2%) showed the highest GC content.

Hydrophobic amino acids were abundant, while acidic amino acids were present in the smallest amount in the plastid genomes of all four studied *Plantago* species. These amino acids were coded by adenine-thymine (AT)-rich sequences in all species (Figure 2A). The RSCU and frequency of amino acids were also analyzed, which revealed that leucine (Leu) is the most abundant, while cysteine (Cys) was the rarest amino acid in all four studied *Plantago* plastid genomes (Figure 2B). The codon usage also revealed a shift toward a higher number of codons having A/T at the third position (Figure 2C and Table S1).

Table 2. Genes of chloroplast genomes of *P. argentea*, *P. atrata*, *P. lanceolata*, and *P. maritima*.

Category for Gene	Group of Gene	Name of Gene				
Photosynthesis-related genes	Photosystem I	<i>psaA</i>	<i>psaB</i>	<i>psaC</i>	<i>psaI</i>	<i>psaJ</i>
		<i>psbA</i>	<i>psbB</i>	<i>psbC</i>	<i>psbD</i>	<i>psbE</i>
	Photosystem II	<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>
	Cytochrome b/f complex	<i>petN</i>	<i>petA</i>	<i>petL</i>	<i>petG</i>	<i>petB</i> *
		<i>petD</i> *				
	ATP synthase	<i>atpI</i>	<i>atpH</i>	<i>atpA</i>	<i>atpF</i> *	<i>atpE</i>
		<i>atpB</i>				
	Assembly/stability of photosystem I	<i>ycf3</i> **	<i>ycf4</i>			
	NADPH dehydrogenase	<i>ndhB</i> ^a	<i>ndhH</i>	<i>ndhA</i> *	<i>ndhI</i>	<i>ndhG</i>
		<i>ndhJ</i>	<i>ndhE</i>	<i>ndhF</i>	<i>ndhC</i>	<i>ndhK</i>
		<i>ndhD</i>				
Rubisco	<i>rbcL</i>					
Transcription- and translation-related genes RNA genes	Transcription Small subunit of ribosome	<i>rpoA</i>	<i>rpoC2</i>	<i>rpoC1</i> *	<i>rpoB</i>	<i>rps16</i> *
		<i>rps7</i> ^a	<i>rps15</i>	<i>rps19</i>	<i>rps3</i>	<i>rps8</i>
		<i>rps14</i>	<i>rps11</i>	<i>rps12</i> ^{a,*}	<i>rps18</i>	<i>rps4</i>
		<i>rps2</i>				
	Large subunit of ribosome	<i>rpl2</i> ^{a,*}	<i>rpl23</i> ^a	<i>rpl32</i>	<i>rpl22</i>	<i>rpl14</i>
		<i>rpl33</i>	<i>rpl36</i>	<i>rpl20</i>	<i>rpl16</i> *	
	Ribosomal RNA	<i>rrn16</i> ^a	<i>rrn4.5</i> ^a	<i>rrn5</i> ^a	<i>rrn23</i> ^a	
	Translation initiation factor	<i>infA</i>				
	Transfer RNA	<i>trnV-GAC</i> ^a	<i>trnI-CAU</i> ^a	<i>trnA-UGC</i> ^{a,*}	<i>trnN-GUU</i> ^a	<i>trnP-UGG</i>
		<i>trnW-CCA</i>	<i>trnV-UAC</i> *	<i>trnL-UAA</i> *	<i>trnF-GAA</i>	<i>trnR-ACG</i> ^a
		<i>trnT-UGU</i>	<i>trnG-UCC</i> *	<i>trnT-GGU</i>	<i>trnR-UCU</i>	<i>trnE-UUC</i>
		<i>trnY-GUA</i>	<i>trnD-GUC</i>	<i>trnC-GCA</i>	<i>trnS-GCU</i>	<i>trnH-GUG</i>
		<i>trnK-UUU</i> *	<i>trnQ-UUG</i>	<i>trnM-CAU</i>	<i>trnG-GCC</i>	<i>trnS-UGA</i>
		<i>trnS-GGA</i>	<i>trnF-GAA</i>	<i>trnM-CAU</i>	<i>trnL-CAA</i> ^a	<i>trnL-UAG</i>
		<i>trnI-GAU</i> ^a				
	Other genes	RNA processing	<i>matK</i>			
	Carbon metabolism	<i>cemA</i>				
Fatty acid synthesis	<i>accD</i>					
Proteolysis	<i>clpP</i> *					
C-type cytochrome synthesis gene	<i>ccsA</i>					
Component of TIC complex	<i>ycf1</i> ^a					
Hypothetical proteins	<i>ycf2</i> ^a					

* Gene with one intron, ** Gene with two introns, ^a Gene with two copies.

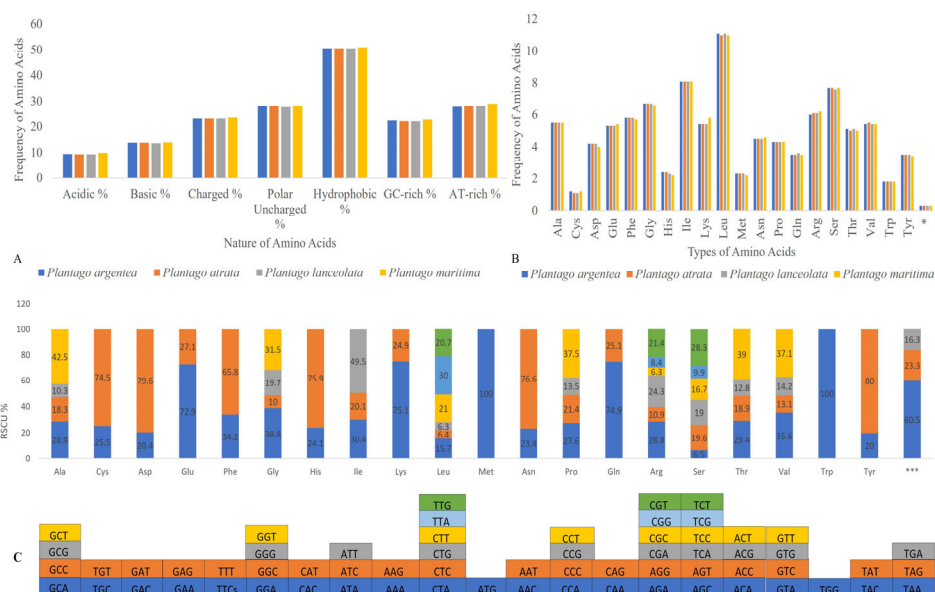


Figure 2. (A) Assessment of amino acid groups and (B) amino acid frequency comparison among *Plantago* species. (C) The codon content and RSCU value of 20 amino acids and stop codons in all protein-coding genes in the plastomes of *Plantago* species. The color of the histogram in (C) is consistent with the color of codons in the same panel. * and ** indicate the end of the protein and the stop codon, respectively.

2.2. Divergence Hotspots in *Plantago* Plastomes

Our comparison showed that all four studied *Plantago* chloroplast genomes had similar nucleotide compositions in all structural (LSC, SSC, and IR) and coding regions, which extended even to IGSs (Table S2). The number of substitutions ranged between 3119 (*P. atrata*) and 6358 (*P. maritima*), while substitution types were shared among species (Table 3). The most frequently occurring mutations were A/G and C/T conversions, compared with other SNPs (Table 3).

Table 3. Comparison of substitutions in *Plantago* species.

Types	<i>P. argentea</i>	<i>P. atrata</i>	<i>P. lanceolata</i>	<i>P. maritima</i>
A/G	954	888	983	1650
C/T	998	924	1002	1616
A/C	374	329	369	897
C/G	207	158	204	506
G/T	410	389	407	980
A/T	434	431	428	709
Total	3377	3119	3393	6358
LSC	2526	2422	2538	5127
SSC	720	572	713	616
IR	131	125	142	615

Plantago ovata was used as a reference for SNP detection.

The ratio of transitions (Ts) and transversions (Tv) in the plastid genomes ranged from 1.09 to 1.44 in the LSC and between 0.90 and 1.36 in the SSC, while varying from 0.8 to 1.22 in the IR region (Table S3). In general, the Ts were more frequent in four studied *Plantago* species, in line with observations in other plant species [27,39]. We found a huge difference in substitutions between the species of subgenera *Psyllium* (*P. argentea*, *P. atrata*,

and *P. lanceolata*) and *Coronopus* (*P. maritima*). InDels were also examined, using DnaSP in all regions of the plastid genomes. The number of InDels ranged from 517 (*P. atrata*) to 687 (*P. maritima*). They were mostly located in the LSC and sparsely found in the SSC, whereas the IRs contained only a few InDels (Table 4).

Table 4. Distribution of InDels (per bp) in *Plantago* chloroplast genomes.

	<i>P. argentea</i>	InDel Sites (bp)	InDel Average Length
LSC	411	4490	10.92
SSC	73	577	7.90
IR	39	410	10.51
	<i>P. atrata</i>	InDel sites (bp)	InDel average length
LSC	419	4271	10.19
SSC	66	1181	17.89
IR	32	346	10.81
	<i>P. lanceolata</i>	InDel sites (bp)	InDel average length
LSC	416	4366	10.49
SSC	78	607	7.782
IR	38	420	11.05
	<i>P. maritima</i>	InDel sites (bp)	InDel average length
LSC	580	9090	15.67
SSC	50	10,263	205.26
IR	57	2269	39.80

P. ovata was used as a reference for InDel detection.

This may be because IRs are more conserved in plastid genomes and evolve under concerted evolution, while the LSC and SSC regions are more prone to substitutions [40]. We also found a difference in InDels specifically in the IR between the species of subgenera *Psyllium* (*P. argentea*, *P. atrata*, and *P. lanceolata*) and *Coronopus* (*P. maritima*). By considering all positions with single- or multinucleotide variations as SNPs, 16,247 SNPs were identified, corresponding to a mean SNP frequency of 16.247 SNPs/kb in *Plantago* species while InDels showed a mean frequency of 2.259/kb. Additionally, we also compared the three *Plantago* genomes that were already published for *P. lanceolata* (MW877582.1), *P. atrata* (MW877580.1), and *P. maritima* (KR297244.1) with our sequenced genomes. We found a significant difference in the number of SNPs and indels between the previously published genomes and our newly sequenced genomes (Table S4). The intraspecific diversity was further examined among the multiple accessions of *Plantago* plastomes for each species. *Plantago aristata* showed the lowest rate of substitutions per base (0.00001) indicating very low intraspecific diversity and *Plantago rigida* showed the highest rate of substitutions per base (0.00230), suggesting higher intraspecific diversity (Table S5; Figure S1). Furthermore, a total of 185 species-specific SNPs for *P. lanceolata* were also identified, which could serve as a basis for the design of molecular markers potentially able to distinguish *P. lanceolata* from other species within the *Plantago* genus (Table S6; see Section 2.7 below).

The InDels and SNP mutational events in the plastid genome showed uneven distributions and clustered as ‘hotspots’ [41,42]. More polymorphism was shown in the IGS regions (average nucleotide diversity $\pi = 0.0961$) than in the gene’s regions ($\pi = 0.0219$) and intron (average nucleotide diversity $\pi = 0.0155$). Among the *Plantago* species, the values ranged from 0.0006 (*ndhB*) to 0.0432 (*ycf1* region) (Figure 3A). We selected 20 highly polymorphic regions for marker development that all belonged to IGS regions based on the analysis of mutation rates of the complete chloroplast genome sequences (Table 5).

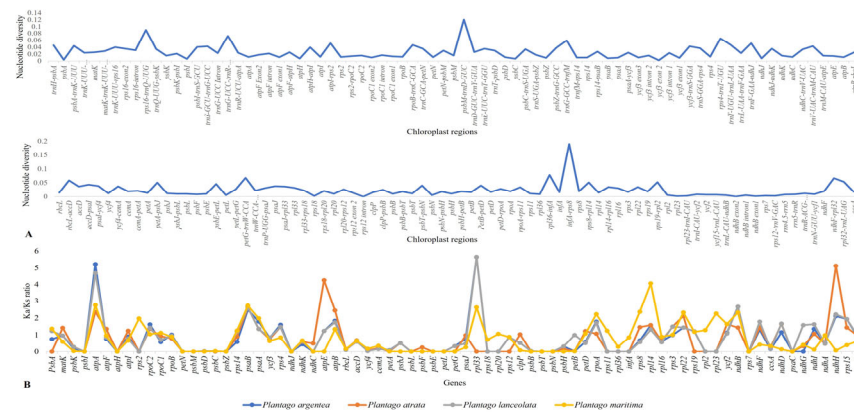


Figure 3. Polymorphism level and Ka/Ks ratios of different plastome regions: (A) Average π value for each coding and intergenic region of the 4 studied *Plantago* plastomes. (B) Ratio of Ka and Ks substitutions in 75 protein-coding genes of the plastomes of the four *Plantago* species.

Table 5. Mutational hotspots among *Plantago* species.

ID	Region	Nucleotide Diversity	InDel Diversity	No. of Mutation	Region Length
1	<i>infA-rps8</i>	0.18967	0.00955	230	333
2	<i>psbM-trnD-GUC</i>	0.12037	0.00361	91	1059
3	<i>rps16-trnQ-UUG</i>	0.08976	0.0119	117	224
4	<i>rpl36-infA</i>	0.07831	0.01124	12	246
5	<i>trnG-UCC-trnR-UCU</i>	0.07237	0.0122	21	1584
6	<i>ndhD-psaC</i>	0.06915	0.0101	13	636
7	<i>petG-trnW-CCA</i>	0.0678	0.0056	14	345
8	<i>ndhF-rpl32</i>	0.06681	0.00564	84	226
9	<i>rps4-trnT-UGU</i>	0.06469	0.00918	34	814
10	<i>trnG-GCC-trnfM</i>	0.05894	0.00877	18	713
11	<i>rbcL-accD</i>	0.05836	0.00534	108	356
12	<i>rpl32-trnL-UAG</i>	0.05348	0.00514	58	183
13	<i>rps19-rpl2</i>	0.05238	0	6	333
14	<i>rps8-rpl14</i>	0.05121	0.0036	16	111
15	<i>atpI-rps2</i>	0.05109	0	20	120
16	<i>trnF-GAA-ndhJ</i>	0.051	0.00711	28	765
17	<i>petA-psbJ</i>	0.05045	0.0106	78	691
18	<i>rpoB-trnC-GCA</i>	0.04623	0.0078	89	195
19	<i>trnT-UGU-trnL-UAA</i>	0.04614	0.00638	61	107
20	<i>psbE-petL</i>	0.04558	0.00654	63	1524

We further investigated the Ka and Ks substitutions and their ratio (Ka/Ks) by 75 protein-coding genes in the four studied *Plantago* species (Table S7 and Figure 3B).

The average Ka/Ks ratio for 75 protein genes analyzed in the four chloroplast genomes was 1.1118. There were significant differences in the evolutionary rates among these species; 19 of 75 protein genes had positively selected sites, which had Ka/Ks > 1, suggesting that these genes might have been subjected to positive selection during evolution. Especially four genes (*atpA*, *psaB*, *ndhB*, *ndhH*) had Ka/Ks ratios higher than 2.0 and these may be the most probable candidate genes for adaptive evolution. We selected these 19 genes

for further analysis using BUSTED, FUBAR, and MEME. We found evidence of episodic diversifying selection only on the *rpoB* gene by using BUSTED. Next, we implemented FUBAR and MEME to detect rare sites that might be under positive selection. The results indicated that several codons exhibited evidence of positive selection in the following genes: *psaA*, *atpF*, and *rps4* (1 codon); *rpl33*, *rpoB*, *rps3*, and *atpB* (2 codons); and *rpl22* (8), *rpoC2*, and *ycf2* (9 codons) (Table S8). Most of the genes showed a relatively slow evolutionary divergence, indicating the conserved nature of the protein-coding genes that are found in the plastomes. Plastid genes are mostly subjected to purifying selection, and the low Ka/Ks ratio is due to the conservation of the functions of the photosynthetic apparatus.

2.3. Repeat Structure and Analyses

Repeats in the plastid genome are useful in evolutionary studies and play crucial roles in genome arrangement, plant breeding, and linkage map construction [43,44]. We performed a microsatellite analysis that revealed shared microsatellite loci ranging from 424 (*P. atrata*) to 459 (*P. lanceolata*). Most SSRs were mononucleotide stretches followed by trinucleotide and dinucleotide repeats. In these groups, A/T motifs were highly abundant in mononucleotides, and AT/TA motifs were frequently observed among dinucleotide SSRs. The mononucleotide SSR motifs varied from 7- to 17-unit repeats; in dinucleotide SSRs, the motif change was from 4- to 6-unit repeats, whereas other types of SSRs were present mainly in 3–5-unit repeats. Most SSRs occurred in the LSC, followed by SSC and IR (Figure 4; Table S9).

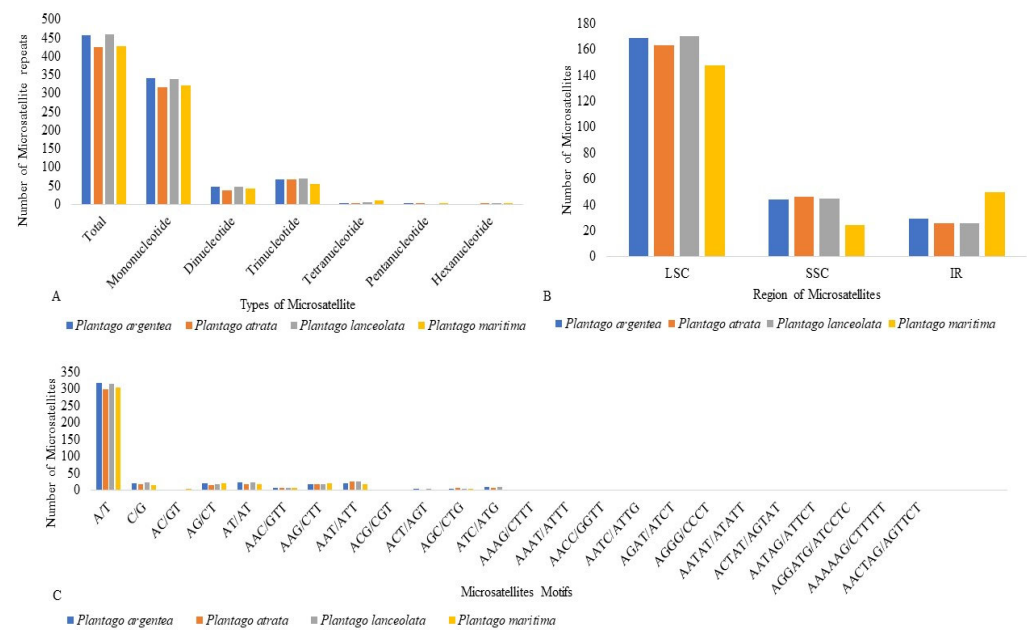


Figure 4. Comparative analysis of microsatellite repeats among four studied *Plantago* species: (A) Total number of microsatellites and their classification according to the number of repeat units. (B) The distribution of microsatellites among structural regions of the plastome. (C) Repeat unit composition of four studied *Plantago* microsatellites.

We found slightly more SSRs in the IR of *P. maritima* (subgenus *Coronopus*) as compared to the species of the subgenus *Psyllium* probably due to the expansion of the IR. REPuter was also employed to locate further oligonucleotide repeats in all four studied *Plantago* species. Overall, 88 oligonucleotide repeats were found in *P. maritima* (subgenus *Coronopus*), which is significantly higher as compared to *P. argentea* (30), *P. atrata* (23), and *P. lanceolata* (27), all of which belongs to the subgenus *Psyllium*. The forward (F) and palindromic (P) repeats were present in all species, while no reverse (R) and complement (C) repeats were found in any of the *Plantago* plastomes analyzed. The oligonucleotide repeats were variable in size (30–90 bp) and a large fraction of the repeats were in the IR, with a net prevalence in

the IGS for *P. maritima*, while with a more balanced distribution between CDS and IGS for the three species of the subgenus *Psyllium* (Figure 5 and Table S10).

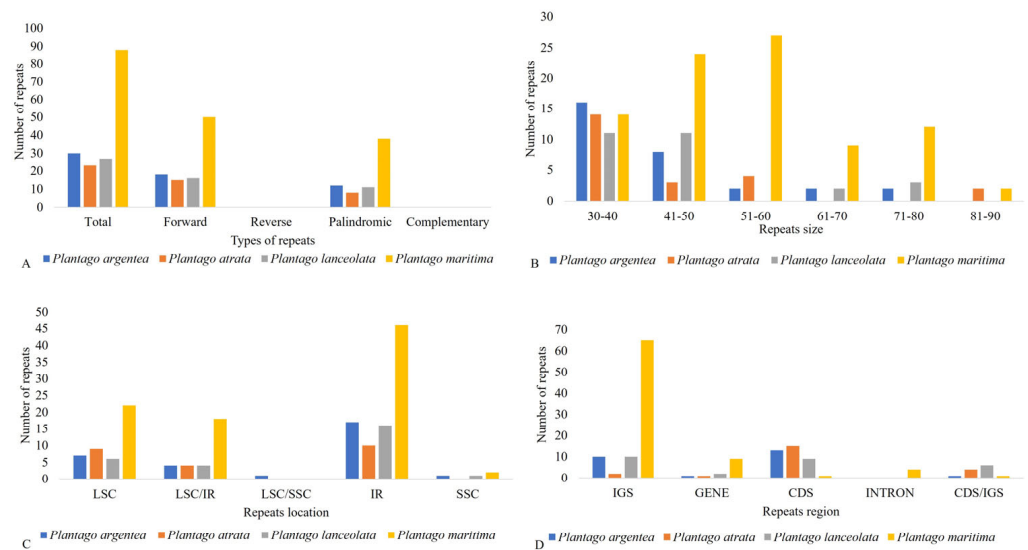


Figure 5. (A) Total number of oligonucleotides repeat among the four studied *Plantago* species and their distribution according to specific characteristics. (B) The distribution of repeats in size ranges. (C) The number of repeats grouped according to their location in each structural region. (D) The distribution of repeats in intergenic spacer regions (IGS), genes, coding DNA sequences (CDS), and introns and their proportionate occurrence.

Moreover, 45 tandem repeats were detected in *P. maritima* (subgenus *Coronopus*), which is slightly higher as compared to *P. argentea* (33), *P. atrata* (36), and *P. lanceolata* (34), all of which belong to the subgenus *Psyllium*. The tandem repeats were variable in size (22–330 bp) and a large fraction of the repeats were in the IGS, followed by CDS and intronic regions (Figure 6 and Table S11).

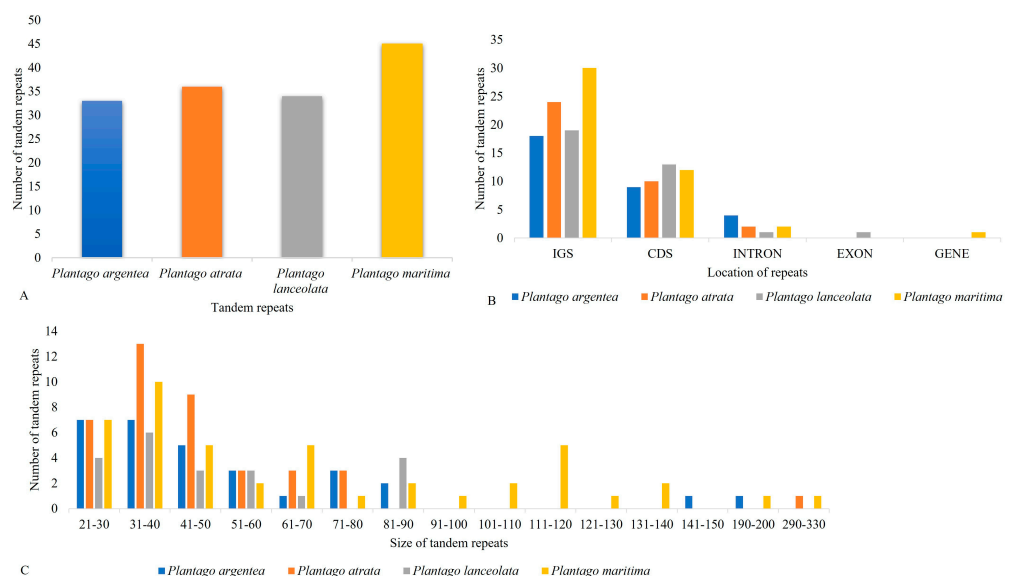


Figure 6. Assessment of tandem repeats: (A) Total number of tandem repeats and (B) their distribution among functional regions of the plastome. (C) Tandem repeat number, size, and distribution.

2.4. Comparative Plastomics and Inverted Repeat Boundaries

The plastid genome of land plants has a conserved quadripartite structure, but diversity exists at the junction sites of the major structural regions of the genome. The size range

of LSC, SSC, and IR varies among the plastid genomes of the species, which may cause alterations in several genes, leading to deletion, duplication, or functional pseudogenization at the junction sites [45]. To investigate such events, we compared the JL (LSC/IR) and JS (IR/SSC) junction sites of 10 *Plantago* plastid genomes (Figure 7).

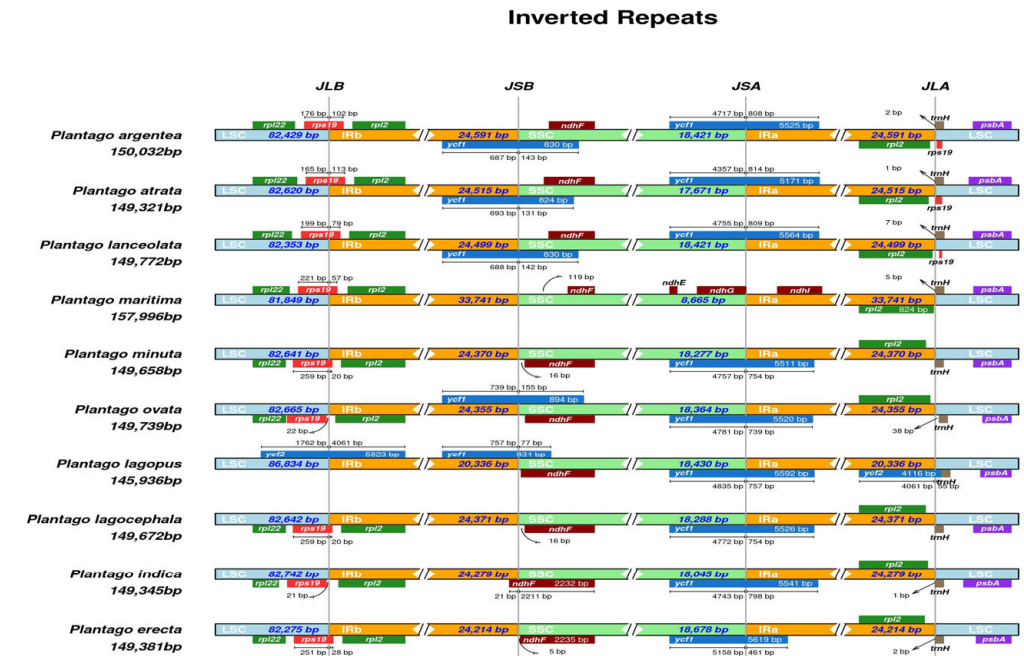


Figure 7. Schematic representation of junction sites in the plastomes of different *Plantago* species belonging to subgenera *Psyllium* and *Coronopus* (accession numbers listed in Table S12). The junction between LSC and IR is indicated as JL, while the junction between IR and SSC is indicated as JS. Genes above and below the different plastome regions are, respectively, in forward and reverse orientation. The number of bases in each region is reported for genes at the boundaries.

The resemblance at junctions revealed the close relatedness among the three *Plantago* species of the subgenus *Psyllium*, where no obvious amplification or contraction events were detected. In *P. maritima* (subgenus *Coronopus*), however, the IRs significantly increased in size up to 9.15 kb as compared to other *Plantago* species of the subgenus *Psyllium*. Moreover, repeat expansion occurred in the SSC, resulting in the transfer of five former SSC genes (*ndhI*, *ndhA*, *ndhH*, *rps15*, *ycf1*) into the IR regions. Additionally, an important variation among the *Plantago* plastomes IR regions was a small-scale inversion related to the *ycf1* gene, which was detected only in the *P. maritima* plastome (Figure S2).

The *rps19* gene was found at the junction site of JLB (LSC/IRb), and a portion of this gene (20–113 bp) was copied in the IRa in all studied *Plantago* genomes except *P. ovata* and *Plantago indica*, while the *rpl2* gene was completely duplicated in IRs. The *ycf1* gene was situated at the IRb/SSC boundary, and a portion of this gene (77–155 bp) was copied in the SSC region. Still at this boundary, the *ndhF* gene was entirely present in the SSC region in all *Plantago* genomes except in *P. indica* (21 bp), where it extended across the IRb region. At the SSC/IRa boundary, the *ycf1* gene was consistently found in all studied *Plantago* plastid genomes and spanned 461–814 bp across the IRa boundary. The *rps19* pseudogene was exclusively identified in *P. argentea*, *P. atrata*, and *P. lanceolata*.

2.5. Putative RNA-Editing Sites

RNA editing can modify the DNA-encoded sequencing of transcribed RNA by adding, deleting, or modifying the nucleotides [46]. RNA editing aids in creating transcripts and maintaining protein diversity [47]. To examine the RNA editing in four studied *Plantago* species, we predicted putative sites in the plastid genomes, using PREPACT, which revealed

67 putative sites in 10 genes of *P. atrata*, while 75, 76, and 95 editing sites were found in 11 genes of *P. argentea*, *P. lanceolata*, and *P. maritima*, respectively.

Most of these RNA-editing sites were found in *rpoC1* (20), *rpcC2* (16), and *ndhA* (12). All four studied *Plantago* species had high levels of conversion from threonine (Thr) to isoleucine (Ile) [22.6% (*P. argentea*), 20.8% (*P. atrata*), and 23.6% (*P. lanceolata*)], followed by proline (Pro) to serine (Ser) [14.28%, 17.94%, and 16.21%, respectively]. By contrast, in *P. maritima* high levels of conversion for proline (Pro) to leucine (Leu) (22.1%) were found followed by proline (Pro) to serine (Ser) (16.8%). Out of all the putative RNA-editing sites detected 52 (69.3%), 48 (71.6%), 51 (67.1%), and 61 (64.2%) codons substituted on the second nucleotide in *P. argentea*, *P. atrata*, *P. lanceolata*, and *P. maritima*, respectively. Many amino acids were converted from Thr to Ile and changes at these sites assisted in the formation of hydrophobic amino acids, e.g., valine (Val), leucine (Leu), and phenylalanine (Phe) (Table S13).

2.6. Phylogenetic Analysis

Maximum likelihood phylogenetic reconstruction for 45 *Plantago* species was carried out based on alignments of complete chloroplast genomes (accession numbers listed in Table S13). The analysis was based on a 213,372 bp alignment using the best-fit model TVM + F + R7 and resulted in a phylogenetic tree supported by high bootstrap values (Figure 8).

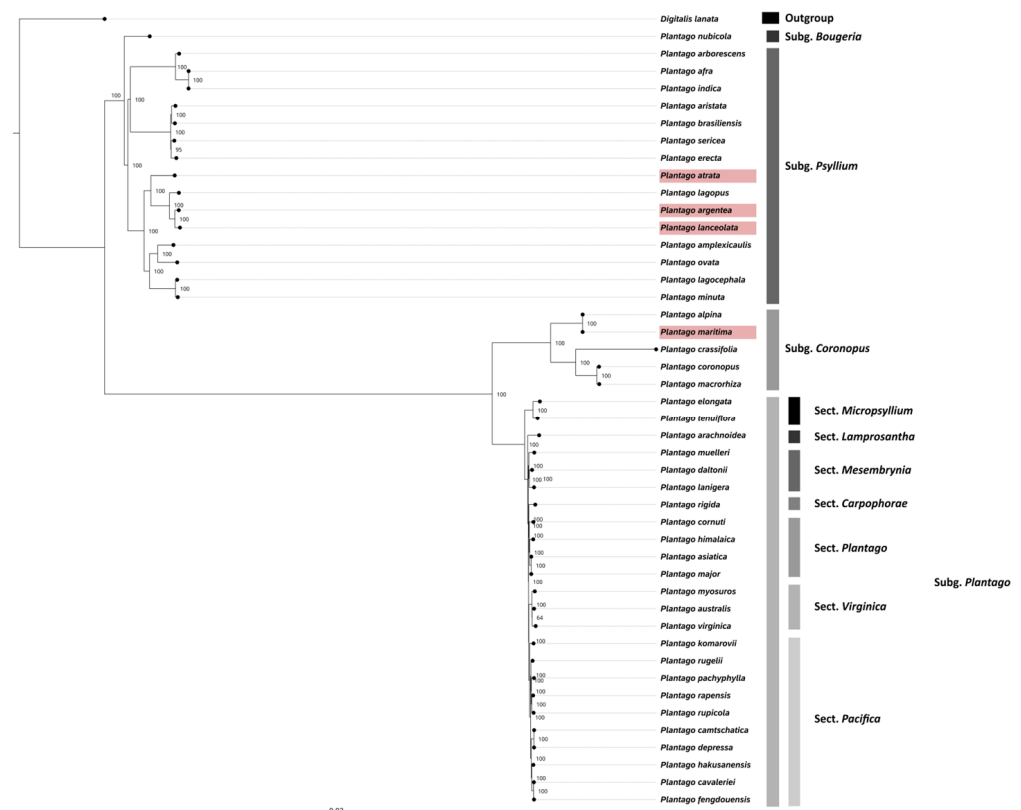


Figure 8. Maximum likelihood phylogenetic reconstruction of 45 *Plantago* species based on fully sequenced plastomes.

Among the species we sequenced, *P. maritima* was a more divergent species as compared to *P. atrata*, *P. argentea*, and *P. lanceolata*. The latter two species were very closely related and formed a single clade with high bootstrap (100%) support. The whole-plastome tree topology obtained here also indicated that the *Plantago* genus is monophyletic with strong support (BS = 100%) and the subgenus *Bougueria* is sister to the remainder subgenera of *Plantago*. Moreover, within the subgenus, *Plantago* sect. *Micropsyllium* is sister to the re-

remainder of subg. *Plantago* (BS = 100%). The divergence times of *Plantago* species analyzed in the current study were estimated using a relaxed uncorrelated clock implemented in BEAST. The results show that the divergence of *Aragoa* and *Plantago* occurred about 8.43 million years ago (Mya) in the Late Miocene, and the diversification of the subgenus *Psyllium* and subgenus *Coronopus* occurred at 3.62 and 3.38 Mya, respectively. Furthermore, *P. argentea* and *P. lanceolata* began to differentiate from the common ancestor into two distinct lineages at 0.3 Mya (Figure S3).

2.7. Development of a *P. lanceolata*-Specific Molecular Marker

Given the relevance of *P. lanceolata* for the herbal market and its cosmopolitan distribution, we set out to develop a PCR assay that could reliably discriminate this plantain species from all the others for which whole-plastome sequences are available. Due to the close phylogenetic relatedness, as well as the very high sequence similarity of *P. argentea*, *P. atrata*, *P. lagopus*, and *P. lanceolata*, we did not find strong candidates among the top 20 mutational hotspots identified above for the *Plantago* genus. We, therefore, manually curated the design of primer pairs in the regions with nucleotide polymorphisms distinguishing the *P. lanceolata* plastome from all other published *Plantago* plastomes. This resulted in a primer pair (PlaLan_1F + PlaLan_2F) whose forward primer was designed in the *rpoC2* CDS, and the reverse primer in the *rpoC2-rpoC1* IGS (Table S14). Each primer 3' end corresponds to a *P. lanceolata*-specific SNP (Table S14; Figure 9). In addition, a control primer pair (Pla_CTRL_F + Pla_CTRL_R) was designed in a region of the *rbcL* gene, which is fully conserved among *Plantago* species. In addition to *P. lanceolata*, the genomic DNA of *P. atrata* and *P. argentea* were used as templates in PCR reactions using either primer combination. According to expectations, a strong amplification band of about 1850 bp was obtained specifically for *P. lanceolata*, but not for *P. atrata* and *P. argentea*. On the other hand, the positive control amplification provided a single band of about 400 bp for all three species (Figure 9), demonstrating that the lack of amplification in *P. atrata* and *P. argentea* is not due to poor DNA quality. Taken together, these results confirm that the PlaLan_1F + PlaLan_2F primer combination reliably discriminates *P. lanceolata* from its closest relatives.

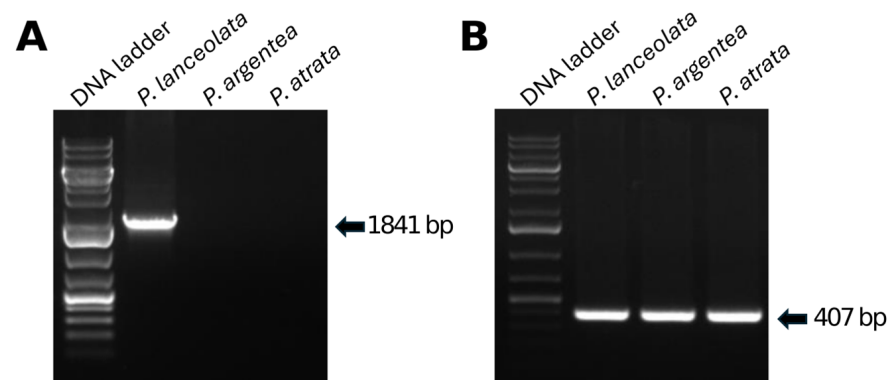


Figure 9. Amplification patterns with plastome markers: (A) The amplicon obtained with primer combination PlaLan_1F + PlaLan_2F is specific for *P. lanceolata*. (B) The amplicon obtained with the Pla_CTRL_F + Pla_CTRL_R primer combination amplifies from three *Plantago* species.

3. Discussion

Whole-plastome sequencing has revolutionized many fundamental and applied fields of plant biology. In this study, we applied whole-plastome sequencing of four accessions of *Plantago* to support the development of species-specific molecular markers for *P. lanceolata*. From an applicative point of view, in fact, the complete chloroplast genome has been proposed as a potential candidate for the next generation of DNA barcodes in plants [38]. Previously, it was demonstrated that the identification of highly variable regions by comparative plastomics could provide insights into the loci that could be used in DNA barcoding [39,40]. The hotspot regions identified in this study could in principle be useful

in such DNA barcoding investigations at the generic level for species identification [41,42]. The extreme phylogenetic relatedness and sequence identity of *P. lanceolata*, especially to *P. argentea*, *P. atrata*, and *P. lagopus*, however, made arduous the development of a suitable barcoding marker even from these 20 mutational hotspot regions. By leveraging on the availability of the whole plastome for a high number of *Plantago* species and of multiple plastomes for *P. lanceolata* as well as some other species, however, we could develop a highly selective primer in the *rpoC1-rpoC2* region, which, despite being less polymorphic, hosted two SNPs that are highly diagnostic for *P. lanceolata*. While intraspecific variation in plastome sequences is known to occur [43–45], a minority of studies have included it as a validation of the markers designed. This is particularly relevant for the *Plantago* genus, which has accelerated plastome evolution [46]. Indeed, while the accessions of *P. lanceolata*, *P. atrata*, and *P. maritima* that we re-sequenced in our study, as expected, clustered with high support with conspecific sequences previously published, they show substitution levels in the range observed for other *Plantago* species (Table S5). However, none of the intraspecific substitutions affected the region selected for primer design, thus further confirming the robustness of the developed marker for the identification of the target species. These results highlight the importance of tailoring the design of barcoding primers to the specific set of species and the question at hand, as mutational hotspots may not be providing optimal discrimination power for very closely related species.

The comparative analyses of newly sequenced accessions of *P. lanceolata*, *P. atrata*, and *P. maritima*, as well as the novel plastome of *P. argentea*, with previously published sequences, confirmed most of the general features already determined for *Plantago* plastomes. In particular, the plastomes of many land plant lineages are evolving quite slowly in terms of sequence and structural organization compared to their mitochondrial or nuclear counterparts [47,48]. Among the hotspots for structural reorganizations inside plastomes are the IRs, which regularly are subject to expansion, contraction, or even complete loss. Such modifications have occurred independently numerous times throughout land plant evolution, frequently specific to orders and families, occasionally even to one or a few species within a genus [49,50]. In *Plantago*, however, we confirmed that the plastomes from the subgenus *Coronopus* show significant genomic change, including increased genome size, gene content, repeat structures, substitutions and indels, RNA-editing sites, and expansions and inversions of the inverted repeat that simultaneously reduced the size of the SSC as compared to the subgenus *Psyllium*. The size of subgenus *Coronopus* plastomes currently available in GenBank varied from 157 to 159 kb as compared to the subgenus *Psyllium* (145–150 kb). Similarly, the IR size of subgenus *Coronopus* plastomes varied from 33 to 34 kb as compared to the subgenus *Psyllium* (20–24 kb). The number of genes also varied between the currently available plastomes in the GenBank belonging to the subgenus *Coronopus* (137–138) as compared to the subgenus *Psyllium* (127–133). Furthermore, in *P. maritima* (subgenus *Coronopus*), a unique gain of a *trnL-CAG* gene via duplication and the diversification of the more common plastid gene *trnL-UAG* were observed. In *P. maritima* (subgenus *Coronopus*), additional synapomorphies include losses of the *rpl2* and *clpP* introns. Additionally, we observed that in the subgenus *Psyllium*, there were 18 (with 12 protein-coding and 6 tRNA genes) intron-containing genes as compared to the subgenus *Coronopus* where the number was 16 (with 10 protein-coding and 6 tRNA genes). Similar changes were also observed in other species of *Plantago* [13,51] and other genera such as *Silene* [52] and *Passiflora* [53]. Generally, these chloroplast characteristics and structural organizational variations provide strong complementary support for the phylogenetic trees produced from the whole-plastid nucleotide sequence, which are in line with the results reported by Mower et al., 2021 [13], regarding (1) the reassignment of *Bougueria nubicola* as a member of *Plantago* (as *Plantago nubicola*), which had been indicated (but with weak support) in previous studies [2,12,54]; and (2) a sister-group relationship between a monophyletic subg. *Coronopus* and a monophyletic subg. *Plantago*, a result that in previous studies had been either weakly recovered [12,55–57] or weakly contradicted [54,58]. Additionally, our phylogenomic analyses also showed with full support the close relatedness of *P. argentea* to

P. lanceolata, thus confirming the weakly supported relationship identified earlier with a limited number of markers [5].

Plants have evolved complex physiological and biochemical adaptations to adjust and adapt to different environmental stresses. We analyzed the patterns of synonymous (Ks) and non-synonymous (Ka) substitution of protein-coding genes, which are essential markers in evolutionary genetics for defining slow- and fast-evolving genes [59]. While substitution rate variations have been previously associated in comparative studies of *Plantago* and Plantaginaceae with plastome structural changes like inversions and IR expansions/contractions [13,46], to the best of our knowledge, no detailed analyses on the pattern of adaptive evolution have been carried out on plastidial genes of the genus. The Ka/Ks ratios in our analysis indicate, for the first time, signatures of changes in selective pressures in a total of 10 plastome genes, which had Ka/Ks values greater than one, possibly due to either relaxed purifying selection, increased positive selection, or a combination of the two. This finding was conclusively supported by an integrative analysis using the Fast Unconstrained Bayesian AppRoximation (FUBAR) and Mixed-Effects Model of Evolution (MEME) methods, which identified a specific set of positively selected codons within these genes. Among these genes, two are involved in chloroplast transcription (*rpoB*, *rpoC2*), four in chloroplast translation (*rps3*, *rps4*, *rpl22*, and *rpl33*), one in protein import (*ycf2*), and three in energy generation through photosynthesis (*psaA*, *atpB*, and *atpF*). Chloroplast gene expression has recently emerged as an important factor in plant responses to environmental stresses, ranging from chilling to high light, salinity, osmotic, and high-temperature stress [60]. Our results show that both phases of chloroplast gene expression, transcription, and translation, show signs of positive selection in the studied *Plantago* species: *rpoB* and *rpoC2* are two of the four plastid-encoded RNA polymerase (PEP) core subunits [61], while the ribosomal genes putatively under positive selection belong to both the small and the large subunit of the chloroplast ribosome [62]. Both *rpoB* and *rpoC2* were among the five positively selected plastome genes in the genus *Hosta* [63], suggesting that they could be implicated in climate change adaptation of the species that underwent rapid diversification in East Asia during the Miocene [64]. Based on previous results and in agreement with the molecular dating reported in this work [12] (Figure S3), during the late Miocene, the *Plantago* genus also underwent a diversification phase associated with the split between the clades encompassing the subgenera *Plantago/Coronopus* and *Psyllium/Bougueria*. It is, therefore, possible that the complex climatic oscillations that took place in the late Miocene and the beginning of the Pliocene [65,66] could have at least in part contributed to determining the selective patterns observed in *rpoB* and *rpoC2* genes of the studied *Plantago*. Among the ribosomal genes, knockout of *rpl33* has been shown to increase plant sensitivity specifically to cold stress, but not to heat or to low or high light [67], suggesting that this gene may have contributed to the adaptation of the studied *Plantago* species to the glacial phases of the Pleistocene glaciation, consistent with the dating results indicating that after about 3.5 MA, a second phase of *Plantago* subgeneric differentiation and speciation took place in both the major clades mentioned above. Contrary to former assumptions, in fact, recent molecular evidence suggests that Quaternary climatic oscillations probably were important drivers of many plant speciations and radiations [68]. Although not directly involved in chloroplast protein expression, *ycf2*, involved in the ATP-driven import of nuclear proteins through the inner chloroplast membrane [69], contributes to the correct functioning of chloroplast protein complexes. Like in the studied *Plantago*, *ycf2* has been found to bear signatures of positive selection in several other genera [70–73]. Among the photosynthetic genes, *atpF*, a subunit of the plastidial H⁺-ATP synthase responsible for electron transport and photophosphorylation during photosynthesis [74], has been associated with the adaptation to cold/drought stress in the *Quercus* species [75]. Intriguingly, *psaA*, a core subunit of photosystem I (PSI), has been associated with the adaptation to different light intensities in the genus *Oryza* [76] and PSI is known to play a relevant role in protection from photoinhibition by regulating photosynthetic cyclic electron flow [77]. As light stress is a common consequence of other

types of stress impairing photosynthesis, like light, low temperature, and drought [77], the repeated cold spells and the cycles of aridification of the climate during the late Miocene and throughout the Pleistocene discussed above may also have been among the causes of the positive selective signatures in the *psaA*, *atpB*, and *atpF* genes of the studied *Plantago* species.

4. Materials and Methods

4.1. Plant Materials, DNA Extraction, and Illumina Sequencing

Fresh lush green leaves of *Plantago* species were collected from the Trentino region, Italy (46.0666° N, 11.1166° E). The leaf samples were rinsed with 70% ethanol, and total genomic DNA was extracted with DNeasy Plant Mini Kit (Qiagen, Hilden, Germany). DNA quality and concentration were assessed by gel electrophoresis on a 1% agarose gel and using the Qubit 4 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA), respectively. A total of 300 ng of extracted genomic DNA for each species was independently sheared with Covaris S220 (Covaris Inc., Woburn, MA, USA) to the average size of 400 bp and used for Illumina-sequencing library preparation. Each library was prepared with TruSeq DNA sample preparation kits for paired-end sequencing (Illumina Inc., San Diego, CA, USA) similarly as previously described [74], pooled in equimolar ratio, and sequenced on an Illumina HiSeq2000 sequencer (CIBIO NGS Facility, Povo (TN), Italy).

4.2. Assembly, Annotation, and Visualization of *Plantago* Species Plastomes

The bowtie program [78] was used to identify and remove phiX174 contaminants using the parameters `-t -p 10 -un -S`. The raw sequencing read quality was checked with the FastQC v0.12.0 tool [79]. The reads were trimmed with FASTX-Toolkit v0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/; accessed on 16 April 2024), with parameters `-q 20 -p 90 -Q 33`, and adapters were eliminated by Trimmomatic v0.32 [80] using PE -phred33 ILLUMINACLIP: TruSeq3-PE-2.fa:2:30:10 LEADING: 3 TRAILING: 3 SLIDING-WINDOW: 4:15 MINLEN: 100 as settings. Then, FastQC was used once again to carry out the final reads quality check with default parameters. GetOrganelle 1.7.7.1 [81] was used for assembly using the parameters `-w 48 -R 25 -k 21,45,65,85`. Gene annotation was conducted using GeSeq [82] with a BLAT [83] search using 85% identity to annotate protein-coding genes, rRNAs, and tRNAs. CPGAVAS2 [84] was used with default parameters by selecting option 1 “43-plastome”. After automatic annotation, start/stop codons and the position of introns were further confirmed manually by visual inspection of the translated protein of each gene in Geneious R9.0.5 (<http://www.geneious.com>) and BLAST search using default settings with homologous genes of plastid genomes of Plantaginaceae. The tRNA genes were further verified by tRNAscan-SE v2.0 [85] with default settings using options: sequence source “Organellar tRNA”, search mode “Default”, genetic code “Universal”, and Cut-off score for reporting tRNAs “15”. ARAGORN v1.2.38 [86] was used with default parameters by selecting genetic code “Bacterial/Plant chloroplast” with maximum intron length of 3000 bp. Circular genome maps were drawn with Chloroplot software [84] (<https://irscope.shinyapps.io/Chloroplot/>; accessed on 16 April 2024) by uploading the GenBank (.gb) format. The average coverage depth of *Plantago* species plastid genomes was calculated with GetOrganelle 1.7.7.1 [81]. The novel *Plantago* plastid genomes were deposited in the National Center for Biotechnology Information (NCBI) under the following accession numbers: *P. argentea* (PP541855), *P. atrata* (PP541856), *P. lanceolata* (PP541857), and *P. maritima* (PP5541858).

4.3. Structural Features, Codon Usage, and Amino Acids Frequencies

Codon usage and amino-acid frequencies were analyzed by using Geneious R9.0.5 software (<http://www.geneious.com>). The intergeneric comparison was carried out to gain insight into differences and synteny that may exist between *Plantago* species. Circosletto [87] was used to compare structural features of *Plantago* chloroplast genomes using blastn search (e-value of $<1 \times 10^{-10}$) to create a Circos output.

4.4. SNPs, Indel, and Mutational Hotspots Detection with PCR Primer Designing

All de novo chloroplast genomes were aligned with multiple alignments using fast Fourier transform (MAFFT) 7.309 [88], using default parameters. Protein-encoding genes, intergenic spacer (IGS) regions, and introns were extracted to calculate the average number of nucleotide differences per site or nucleotide diversity (π) with a 100 bp window size as implemented in DnaSP v6 [89]. The substitution, transition (Ts), and transversion (Tv) rates were resolved from the MAFFT alignment, using *Plantago ovata* as a reference. Each structural element, including the LSC, SSC, and IR, was aligned individually to analyze SNPs and indel polymorphisms with Geneious and DnaSP, respectively. Furthermore, MEGA 11 [90] software was utilized to assess the intraspecific diversity of the complete chloroplast genome of *Plantago* species. Specific primers were designed with Primer3 [91], focusing on the conserved nucleotide sequences at both ends of mutational hotspots.

4.5. IRs Junction Positions and RNA Editing Sites Prediction

The junction sites of the IRs and their border positions were compared using all *Plantago* species using the default setting of the IRscope [92]. The Plant RNA Editing Prediction & Analysis Computer Tool (PREPACT) [93] was used to predict putative RNA editing sites using default settings with *Plantago ovata* as a reference.

4.6. Microsatellites, Oligonucleotide, and Tandem Repeat Analysis

Microsatellite repeats within the chloroplast genomes of *Plantago* species were detected by the MicroSAteLLite identification tool (MISA) [94] with the minimal repeat number: 7 for mononucleotide SSRs, 4 repeat units for dinucleotide SSRs, and 3 repeat units for tri-, tetra-, penta-, and hexanucleotide SSRs. The REPuter software [95] (<https://bibiserv.cebitec.uni-bielefeld.de/reputer>; accessed on 16 April 2024) with minimal repeats size set to 30 bp, hamming distance to 3, minimum similarity percentage of two repeats copies up to 90%, and maximum computed repeats numbers to 500 bp parameters was used for scanning and visualizing genomic repeats including forward (F), reverse (R), complementary (C), and palindromic (P) repeats. Tandem repeats were found using the Tandem Repeats Finder v4.09.1 [96] with default parameters.

4.7. Selective Pressure Analysis of *Plantago* Plastomes

The ratios of synonymous (Ks) and non-synonymous (Ka) substitutions for each extracted protein-encoding gene were calculated with DnaSP for all *Plantago* species, using *Plantago ovata* as a reference. The data were interpreted as $Ka/Ks > 1$, $Ka/Ks = 1$, and $Ka/Ks < 1$, representing positive, neutral, and purifying selection, respectively. Furthermore, we evaluated the impact of positive selection using additional codon models to estimate the rates of synonymous and nonsynonymous substitution. The signs of positive selection were further assessed using fast unconstrained Bayesian approximation (FUBAR) [97], the mixed effects model of evolution (MEME) [98] and (BUSTED) [99] as implemented in the DATAMONKEY web server [100]. Sites with cut-off values of PP > 0.9 in FUBAR were considered candidates to have evolved under positive selection. From all the analyses performed in DATAMONKEY, the most suited model of evolution for each dataset was selected as directly estimated on this web server. In addition, the mixed effects model of evolution (MEME), a branch-site method incorporated in the DATAMONKEY server, was used to test for both pervasive and episodic diversifying selection. MEME applies models with variable ω across lineages at individual sites, restricting ω to ≤ 1 in a proportion p of branches and unrestricted at a proportion $(1 - p)$ of branches per site.

4.8. Plastid Phylogenomic Analysis

We included all available *Plantago* plastid genome sequences in our analysis (Organelle Genome Resources of NCBI, accessed on 16 April 2024). We used *Digitalis lanata* Ehrh. as an outgroup to root our tree from the closet tribe Digitalideae to understand the evolutionary relationships within the genus. For phylogenetic analysis, the complete

chloroplast genomes were aligned using MAFFT, and then tree searches were performed in IQ-TREE 1.5.5 [101]. We used IQ-TREE to infer the best-fitting models of substitution for partitioning the matrix-combining multiple genes with the -TESTMERGEONLY and -AICc (Akaike information criterion corrected for small sample sizes) options in the built-in ModelFinder [102]. Maximum likelihood (ML) analyses were performed, using the ultrafast bootstrap approximation (UFBoot; [103] with 1000 replicates. The key idea behind UFBoot is to keep trees encountered during the ML-tree search for the original sequence alignment and to use them to evaluate the tree likelihoods for the bootstrap sequence alignment. UFBoot provides relatively unbiased bootstrap estimates under mild model misspecifications and reduces computing time while achieving more unbiased branch support than standard bootstrap [103]. The SH-like approximate likelihood ratio test (SH-aLRT) was also conducted together with UFBoot, while iTOL [104–106] was used for the visualization of a phylogenetic tree.

Relative divergence times were estimated for *Plantago* species analyzed in the current study by using BEAST v.1.10.4 [107], applying GTR + I + G rate substitution to the protein-coding plastid gene matrix. A birth and death tree prior and an uncorrelated relaxed clock model that allows rates to vary independently along branches were used [108] with all other parameters set to default. The median time split between *Plantago* and *Aragoa* species (mean = 7.1 Myr; standard deviation = 1.0) was used as a temporal constraint to calibrate the BEAST analyses derived from Rønsted et al. (2002) [12]. Uncertainty regarding this date was incorporated by assigning normal prior distributions to the calibration point [109,110]. Four independent BEAST runs were conducted with Markov Chain Monte Carlo (MCMC) samples based on 100 million generations, sampling every 1000 generations. Convergence of all parameters was assessed in Tracer 1.5 [111] and 10% of each chain was removed as burn-in. The Markov chains were combined in LogCombiner 1.7.2. [107] to calculate the maximum clade credibility tree.

5. Conclusions

The taxonomy of *Plantago* is notoriously difficult, mainly because its low morphological variation, reduced morphology, and lack of useful taxonomic characters have prevented a full understanding of the evolution and taxonomic delimitation of the genus and its species [11,12]. In these complicated groups, well-known plastid barcode regions (e.g., *trnH-psbA*, *matK*) might not have sufficient polymorphism and thus may not be able to provide species-specific information necessary for differentiation [112]. It has been shown that plastome-based “super barcoding” could overcome these difficulties and could differentiate species in difficult taxonomic groups [113]. Here, we investigated whether ‘super-barcoding’ based on plastid genomes could be applied to closely related *Plantago* species by comparing the complete plastome sequences of four species, among which the one of *P. argentea* was obtained for the first time. By comparing our sequences to previously published plastomes, we further identified sequence divergence hotspots and located repeat sequences and indels in the plastomes of *Plantago* species. These regions may provide a useful means of developing suitable molecular markers for species identification and DNA barcoding of *Plantago* medicinal products, although in the case of closely related species belonging to particular clades the use of diagnostic SNPs in less polymorphic regions may be very valuable. Hopefully, by demonstrating the usefulness of species-specific SNPs, our study will provide a solid basis for future studies and assist in the development of DNA barcoding markers for the clarification of the taxonomic identity of *Plantago* species in medicinal plant production. Such plastome-based “super barcoding” could be repeatable, reliable, and sensitive enough to discriminate similar *Plantago* species, as we experimentally demonstrated in the case of *P. lanceolata*.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/plants13192691/s1>, Figure S1: Multiple accessions phylogeny of our sequenced *Plantago* genomes with already published one; Figure S2: Small-scale inversion related to *ycf1* gene in the IR of the *P. maritima* chloroplast genome by using (A) OGDRAW and (B) Mauve alignment; Figure S3: Divergence time estimation and diversification of *Plantago* species using *Aragoa abietina* as an outgroup; Table S1: Codon usage in the chloroplast genome of *P. argentea*, *P. atrata*, *P. lanceolata*, and *P. maritima*; Table S2: Base composition in the chloroplast genome of *P. argentea*, *P. atrata*, *P. lanceolata*, and *P. maritima*; Table S3: The distribution and characteristics of SNPs found among *Plantago* plastid genomes; Table S4: Comparison of SNPs and Indels of our sequenced and already published *Plantago* species plastomes; Table S5: Intraspecific diversity between the multiple accession *Plantago* plastomes; Table S6: *P. lanceolata* species-specific SNPs; Table S7: Comparison of synonymous and non-synonymous substitution in *Plantago* species; Table S8: Positively selected sites in protein-coding genes and list of converted amino acids by using BUSTED, FUBAR, and MEME in *Plantago* species; Table S9: Microsatellites loci in the chloroplast genome of *P. argentea*, *P. atrata*, *P. lanceolata*, and *P. maritima*; Table S10: Oligonucleotide repeats in chloroplast genomes of *P. argentea*, *P. atrata*, *P. lanceolata*, and *P. maritima*; Table S11: Tandem Repeat sequences in the chloroplast genome of *P. argentea*, *P. atrata*, *P. lanceolata*, and *P. maritima*; Table S12: NCBI GenBank accession numbers used in this study; Table S13: Putative RNA editing sites composition in the chloroplast genome of *P. argentea*, *P. atrata*, *P. lanceolata*, and *P. maritima*; Table S14: Barcoding primers developed in this study.

Author Contributions: Conceptualization: F.M. and C.V.; Formal analysis: F.M., M.L. and C.V.; Investigation: F.M., M.L. and C.V.; Methodology: F.M., M.L. and C.V.; Project administration: F.M., M.L. and C.V.; Resources: M.L. and C.V.; Software: F.M., M.L. and C.V.; Supervision: C.V.; Validation: F.M., M.L. and C.V.; Visualization: F.M. and C.V.; Writing—original draft: F.M.; Writing—review and editing: F.M., M.L., A.B., F.P. and C.V. All authors have read and agreed to the published version of the manuscript.

Funding: This study was carried out within the National Biodiversity Future center and received funding from the European Union NextGenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR)—MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4—D.D. 1034 17/06/2022, CN00000033, CUPD43C22001280006). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

Data Availability Statement: The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author/s.

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

References

- Pilger, R.K.F. *Plantaginaceae*; H.R. Engelmann (J. Cramer): Weinheim, Germany, 1956; pp. 1–466.
- Rahn, K. A Phylogenetic Study of the Plantaginaceae. *Bot. J. Linn. Soc.* **1996**, *120*, 145–198. [[CrossRef](#)]
- Li, M.; Cao, H.; But, P.P.H.; Shaw, P.C. Identification of Herbal Medicinal Materials Using DNA Barcodes. *J. Syst. Evol.* **2011**, *49*, 271–283. [[CrossRef](#)]
- Hassemer, G.; Bruun-Lund, S.; Shipunov, A.B.; Briggs, B.G.; Meudt, H.M.; Rønsted, N. The Application of High-Throughput Sequencing for Taxonomy: The Case of *Plantago Subg. Plantago* (Plantaginaceae). *Mol. Phylogenet. Evol.* **2019**, *138*, 156–173. [[CrossRef](#)]
- Shipunov, A.; Fernández-Alonso, J.L.; Hassemer, G.; Alp, S.; Lee, H.J.; Pay, K. Molecular and Morphological Data Improve the Classification of Plantagineae (Lamiales). *Plants* **2021**, *10*, 2299. [[CrossRef](#)]
- Rønsted, N.; Göbel, E.; Franzyk, H.; Jensen, S.R.; Olsen, C.E. Chemotaxonomy of *Plantago*. Iridoid Glucosides and Caffeoyl Phenylethanoid Glycosides. *Phytochemistry* **2000**, *55*, 337–348. [[CrossRef](#)] [[PubMed](#)]
- Shehata, A.A.; Loutfy, M.H.A. On the Taxonomy of Plantaginaceae Juss. Sensu Lato: Evidence from SEM of the Seed Coat. *Turk. J. Botany* **2006**, *30*, 71–84.
- Andrzejewska-Golec, E. Ontogeny of Trichomes in Taxa of Genus *Plantago* L. Subgenus *Plantago*. *Acta Soc. Bot. Pol.* **1991**, *60*, 249–258. [[CrossRef](#)]
- Wong, C.; Murray, B.G. Variable Changes in Genome Size Associated with Different Polyploid Events in *Plantago* (Plantaginaceae). *J. Hered.* **2012**, *103*, 711–719. [[CrossRef](#)]
- Wong, C.; Murray, B.G. In Situ Hybridization with Genomic and rDNA Probes Reveals Complex Origins for Polyploid New Zealand Species of *Plantago* (Plantaginaceae). *New Zeal. J. Bot.* **2014**, *52*, 315–327. [[CrossRef](#)]
- Ishikawa, N.; Yokoyama, J.; Tsukaya, H. Molecular Evidence of Reticulate Evolution in the Subgenus *Plantago* (Plantaginaceae). *Am. J. Bot.* **2009**, *96*, 1627–1635. [[CrossRef](#)]

12. Rønsted, N.; Chase, M.W.; Albach, D.C.; Bello, M.A. Phylogenetic Relationships within *Plantago* (Plantaginaceae): Evidence from Nuclear Ribosomal ITS and Plastid *trnL-F* Sequence Data. *Bot. J. Linn. Soc.* **2002**, *139*, 323–338. [[CrossRef](#)]
13. Mower, J.P.; Guo, W.; Partha, R.; Fan, W.; Levsen, N.; Wolff, K.; Nugent, J.M.; Pabón-Mora, N.; González, F. Plastomes from Tribe Plantagineae (Plantaginaceae) Reveal Infrageneric Structural Synapormorphies and Localized Hypermutation for *Plantago* and Functional Loss of *ndh* Genes from *Littorella*. *Mol. Phylogenet. Evol.* **2021**, *162*, 107217. [[CrossRef](#)] [[PubMed](#)]
14. Samuelson, A.B. The Traditional Uses, Chemical Constituents and Biological Activities of *Plantago major* L. A Review. *J. Ethnopharmacol.* **2000**, *71*, 1–21. [[CrossRef](#)] [[PubMed](#)]
15. Yu, L.L.; Lutterodt, H.; Cheng, Z. Beneficial Health Properties of *Psyllium* and Approaches to Improve Its Functionalities. *Adv. Food Nutr. Res.* **2008**, *55*, 193–220.
16. Pieroni, A.; Nebel, S.; Quave, C.; Münz, H.; Heinrich, M. Ethnopharmacology of Liakra: Traditional Weedy Vegetables of the Arbëreshë of the Vulture Area in Southern Italy. *J. Ethnopharmacol.* **2002**, *81*, 165–185. [[CrossRef](#)] [[PubMed](#)]
17. Heimler, D.; Isolani, L.; Vignolini, P.; Tombelli, S.; Romani, A. Polyphenol Content and Antioxidative Activity in Some Species of Freshly Consumed Salads. *J. Agric. Food Chem.* **2007**, *55*, 1724–1729. [[CrossRef](#)]
18. Tamura, Y.; Nishibe, S. Changes in the Concentrations of Bioactive Compounds in Plantain Leaves. *J. Agric. Food Chem.* **2002**, *50*, 2514–2518. [[CrossRef](#)]
19. Alexander, J.M.; van Kleunen, M.; Ghezzi, R.; Edwards, P.J. Different Genetic Clines in Response to Temperature across the Native and Introduced Ranges of a Global Plant Invader. *J. Ecol.* **2012**, *100*, 771–781. [[CrossRef](#)]
20. Penczykowski, R.M.; Sieg, R.D. *Plantago* spp. as Models for Studying the Ecology and Evolution of Species Interactions across Environmental Gradients. *Am. Nat.* **2021**, *198*, 158–176. [[CrossRef](#)]
21. Skinner, R.H.; Stewart, A. V Narrow-leaf Plantain (*Plantago lanceolata* L.) Selection for Increased Freezing Tolerance. *Crop Sci.* **2014**, *54*, 1238–1242. [[CrossRef](#)]
22. Smith, A.L.; Hodkinson, T.R.; Villellas, J.; Catford, J.A.; Csergő, A.M.; Blomberg, S.P.; Crone, E.E.; Ehrlén, J.; Garcia, M.B.; Laine, A.-L. Global Gene Flow Releases Invasive Plants from Environmental Constraints on Genetic Diversity. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 4218–4227. [[CrossRef](#)] [[PubMed](#)]
23. Fons, F.; Gargadennec, A.; Gueiffier, A.; Roussel, J.L.; Andary, C. Effects of Cinnamic Acid on Polyphenol Production in *Plantago lanceolata*. *Phytochemistry* **1998**, *49*, 697–702. [[CrossRef](#)]
24. Adams, M.; Berset, C.; Kessler, M.; Hamburger, M. Medicinal Herbs for the Treatment of Rheumatic Disorders—A Survey of European Herbals from the 16th and 17th Century. *J. Ethnopharmacol.* **2009**, *121*, 343–359. [[CrossRef](#)] [[PubMed](#)]
25. Oloumi, M.M.; Vosough, D.; Derakhshanfar, A.; Nematollahi, M.H. The Healing Potential of *Plantago lanceolata* Ointment on Collagenase-Induced Tendinitis in Burros (*Equus asinus*). *J. Equine Vet. Sci.* **2011**, *31*, 470–474. [[CrossRef](#)]
26. Blumenthal, M.; Busse, W.R.; Goldberg, A.; Gruenwald, J.; Hall, T.; Riggins, C.W.; Rister, R.S.; Klein, S.; Rister, R.S. Plantain. *Compleat. Ger. Comm. E Monogr. Guid. Herb. Med. Am. Bot. Counc. Coop. Integr. Med. Commun. Bost.* **1998**, 107–123.
27. Shinwari, Z.K. Medicinal Plants Research in Pakistan. *J. Med. Plants Res.* **2010**, *4*, 161–176.
28. Tripathi, H.; Suresh, R.; Kumar, S.; Khan, F. International Trade in Medicinal and Aromatics Plants: A Case Study of Past 18 Years. *J. Med. Aromat. Plant Sci.* **2017**, *39*, 1–17. [[CrossRef](#)]
29. Slifman, N.R.; Obermeyer, W.R.; Aloï, B.K.; Musser, S.M.; Correll Jr, W.A.; Cichowicz, S.M.; Betz, J.M.; Love, L.A. Contamination of Botanical Dietary Supplements by *Digitalis lanata*. *N. Engl. J. Med.* **1998**, *339*, 806–811. [[CrossRef](#)]
30. Wang, H.; Chen, Y.; Wang, L.; Liu, Q.; Yang, S.; Wang, C. Advancing Herbal Medicine: Enhancing Product Quality and Safety through Robust Quality Control Practices. *Front. Pharmacol.* **2023**, *14*, 1265178. [[CrossRef](#)]
31. Jin, S.; Daniell, H. The Engineered Chloroplast Genome Just Got Smarter. *Trends Plant Sci.* **2015**, *20*, 622–640. [[CrossRef](#)]
32. Moore, M.J.; Bell, C.D.; Soltis, P.S.; Soltis, D.E. Using Plastid Genome-Scale Data to Resolve Enigmatic Relationships among Basal Angiosperms. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19363–19368. [[CrossRef](#)] [[PubMed](#)]
33. Ravi, V.; Khurana, J.P.; Tyagi, A.K.; Khurana, P. An Update on Chloroplast Genomes. *Plant Syst. Evol.* **2008**, *271*, 101–122. [[CrossRef](#)]
34. Mehmood, F.; Abdullah; Shahzadi, I.; Ahmed, I.; Waheed, M.T.; Mirza, B. Characterization of *Withania somnifera* Chloroplast Genome and Its Comparison with Other Selected Species of Solanaceae. *Genomics* **2020**, *112*, 1522–1530. [[CrossRef](#)]
35. Iram, S.; Hayat, M.Q.; Tahir, M.; Gul, A.; Abdullah; Ahmed, I. Chloroplast Genome Sequence of *Artemisia scoparia*: Comparative Analyses and Screening of Mutational Hotspots. *Plants* **2019**, *8*, 476. [[CrossRef](#)]
36. Serrano, H.C.; Cotrim, H.; Pinto, M.J.; Martins-Loução, M.A.; Branquinho, C. Metal Hyperaccumulation Patterns within *Plantago* Phylogeny (Plantaginaceae). *Plant Soil* **2017**, *411*, 227–241. [[CrossRef](#)]
37. Shipunov, A.; Fernández, A.J.L.; Hassemmer, G.; Alp, S.; Lee, H.J.; Pay, K. How to Map a Plantain: Phylogeny of the Diverse Plantagineae (Lamiales). *bioRxiv* **2020**.
38. Jiao, L.; Lu, Y.; He, T.; Li, J.; Yin, Y. A Strategy for Developing High-Resolution DNA Barcodes for Species Discrimination of Wood Specimens Using the Complete Chloroplast Genome of Three Pterocarpus Species. *Planta* **2019**, *250*, 95–104. [[CrossRef](#)] [[PubMed](#)]
39. Yan, M.; Dong, S.; Gong, Q.; Xu, Q.; Ge, Y. Comparative Chloroplast Genome Analysis of Four *Polygonatum* Species Insights into DNA Barcoding, Evolution, and Phylogeny. *Sci. Rep.* **2023**, *13*, 16495. [[CrossRef](#)]
40. Yao, R.; Guo, R.; Liu, Y.; Kou, Z.; Shi, B. Identification and Phylogenetic Analysis of the Genus *Syringa* Based on Chloroplast Genomic DNA Barcoding. *PLoS ONE* **2022**, *17*, e0271633. [[CrossRef](#)]

41. Tang, D.; Lin, Y.; Wei, F.; Quan, C.; Wei, K.; Wei, Y.; Cai, Z.; Kashif, M.H.; Miao, J. Characteristics and Comparative Analysis of *Mesona chinensis* Benth Chloroplast Genome Reveals DNA Barcode Regions for Species Identification. *Funct. Integr. Genom.* **2022**, *22*, 467–479. [[CrossRef](#)]
42. Park, I.; Yang, S.; Kim, W.J.; Noh, P.; Lee, H.O.; Moon, B.C. Authentication of Herbal Medicines *Dipsacus asper* and *Phlomis umbrosa* Using DNA Barcodes, Chloroplast Genome, and Sequence Characterized Amplified Region (SCAR) Marker. *Molecules* **2018**, *23*, 1748. [[CrossRef](#)] [[PubMed](#)]
43. Tong, W.; He, Q.; Wang, X.-Q.; Yoon, M.-Y.; Ra, W.-H.; Li, F.; Yu, J.; Oo, W.H.; Min, S.-K.; Choi, B.-W. A Chloroplast Variation Map Generated Using Whole Genome Re-Sequencing of Korean Landrace Rice Reveals Phylogenetic Relationships among *Oryza sativa* Subspecies. *Biol. J. Linn. Soc.* **2015**, *115*, 940–952. [[CrossRef](#)]
44. Zhao, Y.; Yin, J.; Guo, H.; Zhang, Y.; Xiao, W.; Sun, C.; Wu, J.; Qu, X.; Yu, J.; Wang, X. The Complete Chloroplast Genome Provides Insight into the Evolution and Polymorphism of *Panax ginseng*. *Front. Plant Sci.* **2015**, *5*, 696. [[CrossRef](#)] [[PubMed](#)]
45. Muraguri, S.; Xu, W.; Chapman, M.; Muchugi, A.; Oluwaniyi, A.; Oyebo, O.; Liu, A. Intraspecific Variation within Castor Bean (*Ricinus communis* L.) Based on Chloroplast Genomes. *Ind. Crops Prod.* **2020**, *155*, 112779. [[CrossRef](#)]
46. Zhu, A.; Guo, W.; Gupta, S.; Fan, W.; Mower, J.P. Evolutionary Dynamics of the Plastid Inverted Repeat: The Effects of Expansion, Contraction, and Loss on Substitution Rates. *New Phytol.* **2016**, *209*, 1747–1756. [[CrossRef](#)]
47. Mower, J.P.; Vickrey, T.L. Structural Diversity among Plastid Genomes of Land Plants. *Adv. Bot. Res.* **2018**, *85*, 263–292.
48. 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](#)]
49. Cao, J.; Wang, H.; Cao, Y.; Kan, S.; Li, J.; Liu, Y. Extreme Reconfiguration of Plastid Genomes in Papaveraceae: Rearrangements, Gene Loss, Pseudogenization, IR Expansion, and Repeats. *Int. J. Mol. Sci.* **2024**, *25*, 2278. [[CrossRef](#)]
50. Choi, I.-S.; Jansen, R.; Ruhlman, T. Lost and Found: Return of the Inverted Repeat in the Legume Clade Defined by Its Absence. *Genome Biol. Evol.* **2019**, *11*, 1321–1333. [[CrossRef](#)]
51. Asaf, S.; Khan, A.L.; Lubna, A.; Khan, A.; Khan, G.; Lee, I.-J.; Al-Harrasi, A. Expanded Inverted Repeat Region with Large Scale Inversion in the First Complete Plastid Genome Sequence of *Plantago ovata*. *Sci. Rep.* **2020**, *10*, 3881. [[CrossRef](#)]
52. Sloan, D.B.; Triant, D.A.; Forrester, N.J.; Bergner, L.M.; Wu, M.; Taylor, D.R. A Recurring Syndrome of Accelerated Plastid Genome Evolution in the Angiosperm Tribe Sileneae (Caryophyllaceae). *Mol. Phylogenet. Evol.* **2014**, *72*, 82–89. [[CrossRef](#)] [[PubMed](#)]
53. Weng, M.L.; Blazier, J.C.; Govindu, M.; Jansen, R.K. Reconstruction of the Ancestral Plastid Genome in Geraniaceae Reveals a Correlation between Genome Rearrangements, Repeats, and Nucleotide Substitution Rates. *Mol. Biol. Evol.* **2014**, *31*, 645–659. [[CrossRef](#)]
54. Hoggard, R.K.; Kores, P.J.; Molvray, M.; Hoggard, G.D.; Broughton, D.A. Molecular Systematics and Biogeography of the Amphibious Genus *Littorella* (Plantaginaceae). *Am. J. Bot.* **2003**, *90*, 429–435. [[CrossRef](#)] [[PubMed](#)]
55. Bello, M.A.; Chase, M.W.; Olmstead, R.G.; Rønsted, N.; Albach, D. The Páramo Endemic *Aragoa* Is the Sister Genus of *Plantago* (Plantaginaceae; Lamiales): Evidence from Plastid *rbcL* and Nuclear Ribosomal ITS Sequence Data. *Kew Bull.* **2002**, *57*, 585–597. [[CrossRef](#)]
56. Tay, M.L.; Meudt, H.M.; Garnock-Jones, P.J.; Ritchie, P.A. DNA Sequences from Three Genomes Reveal Multiple Long-Distance Dispersals and Non-Monophyly of Sections in Australasian *Plantago* (Plantaginaceae). *Aust. Syst. Bot.* **2010**, *23*, 47–68. [[CrossRef](#)]
57. Affenzeller, M.; Kadereit, J.W.; Comes, H.P. Parallel Bursts of Recent and Rapid Radiation in the Mediterranean and Eritreo-Arabian Biodiversity Hotspots as Revealed by *Globularia* and *Campylanthus* (Plantaginaceae). *J. Biogeogr.* **2018**, *45*, 552–566. [[CrossRef](#)]
58. Höpke, J.; Mucina, L.; Albach, D.C. Phylogenetic and Morphometric Analysis of *Plantago* Section *Coronopus* (Plantaginaceae). *Taxon* **2019**, *68*, 315–339. [[CrossRef](#)]
59. Kimura, M. Model of Effectively Neutral Mutations in Which Selective Constraint Is Incorporated. *Proc. Natl. Acad. Sci. USA* **1979**, *76*, 3440–3444. [[CrossRef](#)]
60. Zhang, Y.; Zhang, A.; Li, X.; Lu, C. The Role of Chloroplast Gene Expression in Plant Responses to Environmental Stress. *Int. J. Mol. Sci.* **2020**, *21*, 6082. [[CrossRef](#)]
61. Börner, T.; Aleynikova, A.Y.; Zubo, Y.O.; Kusnetsov, V.V. Chloroplast RNA Polymerases: Role in Chloroplast Biogenesis. *Biochim. Biophys. Acta (BBA)-Bioenerg.* **2015**, *1847*, 761–769. [[CrossRef](#)]
62. Tiller, N.; Bock, R. The Translational Apparatus of Plastids and Its Role in Plant Development. *Mol. Plant* **2014**, *7*, 1105–1120. [[CrossRef](#)]
63. Yang, J.; Choi, M.-J.; Kim, S.-H.; Choi, H.-J.; Kim, S.-C. Plastome Characterization and Phylogenomic Analysis Yield New Insights into the Evolutionary Relationships among the Species of the Subgenus *Bryocles* (Hosta; Asparagaceae) in East Asia. *Plants* **2021**, *10*, 1980. [[CrossRef](#)] [[PubMed](#)]
64. Yoo, M.-J.; Lee, B.-Y.; Kim, S.; Lim, C.E. Phylogenomics with Hyb-Seq Unravels Korean *Hosta* Evolution. *Front. Plant Sci.* **2021**, *12*, 645735. [[CrossRef](#)]
65. Fortelius, M.; Eronen, J.; Liu, L.; Pushkina, D.; Tesakov, A.; Vislobokova, I.; Zhang, Z. Late Miocene and Pliocene Large Land Mammals and Climatic Changes in Eurasia. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **2006**, *238*, 219–227. [[CrossRef](#)]
66. Graham, A. Middle Miocene through Pliocene North American Vegetational History: 16.3–1.6 Ma. In *Late Cretaceous and Cenozoic History of North American Vegetation (North of Mexico)*; Oxford University Press: Oxford, UK, 1999; ISBN 0197561128.

67. Rogalski, M.; Schottler, M.A.; Thiele, W.; Schulze, W.X.; Bock, R. *rpl33*, a Nonessential Plastid-Encoded Ribosomal Protein in Tobacco, Is Required under Cold Stress Conditions. *Plant Cell* **2008**, *20*, 2221–2237. [CrossRef]
68. Kadereit, J.W.; Abbott, R.J. Plant Speciation in the Quaternary. *Plant Ecol. Divers.* **2021**, *14*, 105–142. [CrossRef]
69. Kikuchi, S.; Asakura, Y.; Imai, M.; Nakahira, Y.; Kotani, Y.; Hashiguchi, Y.; Nakai, Y.; Takafuji, K.; Bédard, J.; Hirabayashi-Ishioka, Y. A Ycf2-FtsHi Heteromeric AAA-ATPase Complex Is Required for Chloroplast Protein Import. *Plant Cell* **2018**, *30*, 2677–2703. [CrossRef]
70. Henriquez, C.L.; Abdullah, I.; Carlsen, M.M.; Zuluaga, A.; Croat, T.B.; Mckain, M.R. Molecular Evolution of Chloroplast Genomes in Monsteroideae (Araceae). *Planta* **2020**, *251*, 72. [CrossRef] [PubMed]
71. She, R.; Zhao, P.; Zhou, H.; Yue, M.; Yan, F.; Hu, G.; Gao, X.; Zhang, S. Complete Chloroplast Genomes of Liliaceae (SI) Species: Comparative Genomic and Phylogenetic Analyses. *Nord. J. Bot.* **2020**, *38*, e02477. [CrossRef]
72. Zhao, S.; Gao, X.; Yu, X.; Yuan, T.; Zhang, G.; Liu, C.; Li, X.; Wei, P.; Li, X.; Liu, X. Comparative Analysis of Chloroplast Genome of *Meconopsis* (Papaveraceae) Provides Insights into Their Genomic Evolution and Adaptation to High Elevation. *Int. J. Mol. Sci.* **2024**, *25*, 2193. [CrossRef]
73. Corvalán, L.C.J.; Sobreiro, M.B.; Carvalho, L.R.; Dias, R.O.; Braga-Ferreira, R.S.; Targueta, C.P.; Silva-Neto, C.M.E.; Berton, B.W.; Pereira, A.M.S.; Diniz-Filho, J.A.F. Chloroplast Genome Assembly of *Serjania erecta* Raldk: Comparative Analysis Reveals Gene Number Variation and Selection in Protein-Coding Plastid Genes of Sapindaceae. *Front. Plant Sci.* **2023**, *14*, 1258794. [CrossRef] [PubMed]
74. Hahn, A.; Vonck, J.; Mills, D.J.; Meier, T.; Kühlbrandt, W. Structure, Mechanism, and Regulation of the Chloroplast ATP Synthase. *Science* **2018**, *360*, eaat4318. [CrossRef] [PubMed]
75. Yin, K.; Zhang, Y.; Li, Y.; Du, F.K. Different Natural Selection Pressures on the *atpF* Gene in Evergreen Sclerophyllous and Deciduous Oak Species: Evidence from Comparative Analysis of the Complete Chloroplast Genome of *Quercus aquifolioides* with Other Oak Species. *Int. J. Mol. Sci.* **2018**, *19*, 1042. [CrossRef]
76. Gao, L.-Z.; Liu, Y.-L.; Zhang, D.; Li, W.; Gao, J.; Liu, Y.; Li, K.; Shi, C.; Zhao, Y.; Zhao, Y.-J. Evolution of *Oryza* Chloroplast Genomes Promoted Adaptation to Diverse Ecological Habitats. *Commun. Biol.* **2019**, *2*, 278. [CrossRef] [PubMed]
77. Didaran, F.; Kordrostami, M.; Ghasemi-Soloklui, A.A.; Pashkovskiy, P.; Kreslavski, V.; Kuznetsov, V.; Allakhverdiev, S.I. The Mechanisms of Photoinhibition and Repair in Plants under High Light Conditions and Interplay with Abiotic Stressors. *J. Photochem. Photobiol. B Biol.* **2024**, *259*, 113004. [CrossRef]
78. Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S.L. Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome. *Genome Biol.* **2009**, *10*, R25. [CrossRef]
79. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010. Available online: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed on 16 April 2024).
80. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]
81. Jin, J.-J.; Yu, W.-B.; Yang, J.-B.; Song, Y.; DePamphilis, C.W.; Yi, T.-S.; Li, D.-Z. GetOrganelle: A Fast and Versatile Toolkit for Accurate de Novo Assembly of Organelle Genomes. *Genome Biol.* **2020**, *21*, 241. [CrossRef] [PubMed]
82. 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]
83. Kent, W.J. BLAT—The BLAST-like Alignment Tool. *Genome Res.* **2002**, *12*, 656–664. [CrossRef]
84. 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] [PubMed]
85. 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] [PubMed]
86. Laslett, D.; Canback, B. ARAGORN, a Program to Detect tRNA Genes and tmRNA Genes in Nucleotide Sequences. *Nucleic Acids Res.* **2004**, *32*, 11–16. [CrossRef] [PubMed]
87. Darzentas, N. Circoletto: Visualizing Sequence Similarity with Circos. *Bioinformatics* **2010**, *26*, 2620–2621. [CrossRef]
88. 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]
89. Rozas, J.; Ferrer-Mata, A.; Sanchez-DelBarrio, J.C.; Guirao-Rico, S.; Librado, P.; Ramos-Onsins, S.E.; Sanchez-Gracia, A. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol. Biol. Evol.* **2017**, *34*, 3299–3302. [CrossRef] [PubMed]
90. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027. [CrossRef]
91. Untergasser, A.; Cutcutache, I.; Koressaar, T.; Ye, J.; Faircloth, B.C.; Remm, M.; Rozen, S.G. Primer3—New Capabilities and Interfaces. *Nucleic Acids Res.* **2012**, *40*, e115. [CrossRef]
92. Amiroufsefi, A.; Hyvönen, J.; Poczai, P. IRscope: An Online Program to Visualize the Junction Sites of Chloroplast Genomes. *Bioinformatics* **2018**, *34*, 3030–3031. [CrossRef]
93. Lenz, H.; Knoop, V. PREPACT 2.0: Predicting C-to-U and U-to-C RNA Editing in Organelle Genome Sequences with Multiple References and Curated RNA Editing Annotation. *Bioinform. Biol. Insights* **2013**, *7*, BBI-S11059. [CrossRef]
94. 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]

95. Kurtz, S. REPuter: The Manifold Applications of Repeat Analysis on a Genomic Scale. *Nucleic Acids Res.* **2001**, *29*, 4633–4642. [[CrossRef](#)] [[PubMed](#)]
96. Benson, G. Tandem Repeats Finder: A Program to Analyze DNA Sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580. [[CrossRef](#)]
97. Murrell, B.; Moola, S.; Mabona, A.; Weighill, T.; Sheward, D.; Kosakovsky Pond, S.L.; Scheffler, K. FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. *Mol. Biol. Evol.* **2013**, *30*, 1196–1205. [[CrossRef](#)]
98. Murrell, B.; Wertheim, J.O.; Moola, S.; Weighill, T.; Scheffler, K.; Kosakovsky Pond, S.L. Detecting Individual Sites Subject to Episodic Diversifying Selection. *PLoS Genet.* **2012**, *8*, e1002764. [[CrossRef](#)] [[PubMed](#)]
99. Murrell, B.; Weaver, S.; Smith, M.D.; Wertheim, J.O.; Murrell, S.; Aylward, A.; Eren, K.; Pollner, T.; Martin, D.P.; Smith, D.M. Gene-Wide Identification of Episodic Selection. *Mol. Biol. Evol.* **2015**, *32*, 1365–1371. [[CrossRef](#)]
100. Delport, W.; Poon, A.F.Y.; Frost, S.D.W.; Kosakovsky Pond, S.L. Datamonkey 2010: A Suite of Phylogenetic Analysis Tools for Evolutionary Biology. *Bioinformatics* **2010**, *26*, 2455–2457. [[CrossRef](#)]
101. Trifinopoulos, J.; Nguyen, L.T.; von Haeseler, A.; Minh, B.Q. W-IQ-TREE: A Fast Online Phylogenetic Tool for Maximum Likelihood Analysis. *Nucleic Acids Res.* **2016**, *44*, W232–W235. [[CrossRef](#)]
102. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; von Haeseler, A.; Jermin, L.S. ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates. *Nat. Methods* **2017**, *14*, 587–589. [[CrossRef](#)]
103. Hoang, D.T.; Chernomor, O.; von Haeseler, A.; Minh, B.Q.; Vinh, L.S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **2018**, *35*, 518–522. [[CrossRef](#)]
104. Dereeper, A.; Guignon, V.; Blanc, G.; Audic, S.; Buffet, S.; Chevenet, F.; Dufayard, J.-F.; Guindon, S.; Lefort, V.; Lescot, M.; et al. Phylogeny.Fr: Robust Phylogenetic Analysis for the Non-Specialist. *Nucleic Acids Res.* **2008**, *36*, W465–W469. [[CrossRef](#)] [[PubMed](#)]
105. Lemoine, F.; Correia, D.; Lefort, V.; Doppelt-Azeroual, O.; Mareuil, F.; Cohen-Boulakia, S.; Gascuel, O. NGPhylogeny.Fr: New Generation Phylogenetic Services for Non-Specialists. *Nucleic Acids Res.* **2019**, *47*, W260–W265. [[CrossRef](#)] [[PubMed](#)]
106. Letunic, I.; Bork, P. Interactive Tree of Life (ITOL) v6: Recent Updates to the Phylogenetic Tree Display and Annotation Tool. *Nucleic Acids Res.* **2024**, *52*, W78–W82. [[CrossRef](#)]
107. Suchard, M.A.; Lemey, P.; Baele, G.; Ayres, D.L.; Drummond, A.J.; Rambaut, A. Bayesian Phylogenetic and Phylodynamic Data Integration Using BEAST 1.10. *Virus Evol.* **2018**, *4*, vey016. [[CrossRef](#)] [[PubMed](#)]
108. Drummond, A.J.; Ho, S.Y.W.; Phillips, M.J.; Rambaut, A. Relaxed Phylogenetics and Dating with Confidence. *PLoS Biol.* **2006**, *4*, e88. [[CrossRef](#)]
109. Couvreur, T.L.P.; Chatrou, L.W.; Sosef, M.S.M.; Richardson, J.E. Molecular Phylogenetics Reveal Multiple Tertiary Vicariance Origins of the African Rain Forest Trees. *BMC Biol.* **2008**, *6*, 54. [[CrossRef](#)]
110. Evans, M.; Aubriot, X.; Hearn, D.; Lanciaux, M.; Lavergne, S.; Cruaud, C.; Lowry, P.P.; Haevermans, T. Insights on the Evolution of Plant Succulence from a Remarkable Radiation in Madagascar (Euphorbia). *Syst. Biol.* **2014**, *63*, 698–711. [[CrossRef](#)]
111. Rambaut, A.; Suchard, M.A.; Xie, D.; Drummond, A.J. Tracer v1.6. Computer Program and Documentation Distributed by the Author. 2014. Available online: <http://beast.bio.ed.ac.uk/Tracer> (accessed on 16 April 2024).
112. Spooner, D.M. Dna Barcoding Will Frequently Fail in Complicated Groups: An Example in Wild Potatoes. *Am. J. Bot.* **2009**, *96*, 1177–1189. [[CrossRef](#)]
113. Ji, Y.; Yang, J.; Landis, J.B.; Wang, S.; Yang, Z.; Zhang, Y. Deciphering the Taxonomic Delimitation of *Ottelia acuminata* (Hydrocharitaceae) Using Complete Plastomes as Super-Barcodes. *Front. Plant Sci.* **2021**, *12*, 681270. [[CrossRef](#)]

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