

## Article

# Genome Characterization, Comparison and Phylogenetic Analysis of Complete Mitochondrial Genome of *Evolvulus alsinoides* Reveals Highly Rearranged Gene Order in *Solanales*

Pattayampadam Ramakrishnan Shidhi <sup>1,\*</sup>, Vadakkemukadiyil Chellappan Biju <sup>1</sup>, Sasi Anu <sup>1</sup> , Chandrasekharan Laila Vipin <sup>1</sup>, Kumar Raveendran Deelip <sup>2</sup> and Sukumaran Nair Achuthsankar <sup>1</sup>

<sup>1</sup> Department of Computational Biology and Bioinformatics, University of Kerala, Thiruvananthapuram 695581, India; bijuvcdcb@keralauniversity.ac.in (V.C.B.); anu24sasi@gmail.com (S.A.); vipincl1993@gmail.com (C.L.V.); achuthsankar@keralauniversity.ac.in (S.N.A.)

<sup>2</sup> Campus Computing Facility (CCF) at the Central Laboratory for Instrumentation and Facilitation, University of Kerala, Thiruvananthapuram 695581, India; deelip@keralauniversity.ac.in

\* Correspondence: shidhidcb@keralauniversity.ac.in or shidhibio@gmail.com; Tel.: +91-9895484530

**Abstract:** Mitogenome sequencing provides an understanding of the evolutionary mechanism of mitogenome formation, mechanisms driving plant gene order, genome structure, and migration sequences. Data on the mitochondrial genome for family Convolvulaceae members is lacking. *E. alsinoides*, also known as shankhpushpi, is an important medicinal plant under the family Convolvulaceae, widely used in the Ayurvedic system of medicine. We identified the mitogenome of *E. alsinoides* using the Illumina mate-pair sequencing platform, and annotated using bioinformatics approaches in the present study. The mitogenome of *E. alsinoides* was 344184 bp in length and comprised 46 unique coding genes, including 31 protein-coding genes (PCGs), 12 tRNA genes, and 3 rRNA genes. The secondary structure of tRNAs shows that all the tRNAs can be folded into canonical clover-leaf secondary structures, except three *trnW*, *trnG*, and *trnC*. Measurement of the skewness of the nucleotide composition showed that the AT and GC skew is positive, indicating higher A's and G's in the mitogenome of *E. alsinoides*. The Ka/Ks ratios of 11 protein-coding genes (*atp1*, *ccmC*, *cob*, *cox1*, *rps19*, *rps12*, *nad3*, *nad9*, *atp9*, *rpl5*, *nad4L*) were <1, indicating that these genes were under purifying selection. Synteny and gene order analysis were performed to identify homologous genes among the related species. Synteny blocks representing nine genes (*nad9*, *nad2*, *ccmFc*, *nad1*, *nad4*, *nad5*, *matR*, *cox1*, *nad7*) were observed in all the species of *Solanales*. Gene order comparison showed that a high level of gene rearrangement has occurred among all the species of *Solanales*. The mitogenome data obtained in the present study could be used as the *Convolvulaceae* family representative for future studies, as there is no complex taxonomic history associated with this plant.

**Keywords:** *E. alsinoides*; Convolvulaceae; *Solanales*; mitogenome; protein-coding genes



**Citation:** Shidhi, P.R.; Biju, V.C.; Anu, S.; Vipin, C.L.; Deelip, K.R.; Achuthsankar, S.N. Genome Characterization, Comparison and Phylogenetic Analysis of Complete Mitochondrial Genome of *Evolvulus alsinoides* Reveals Highly Rearranged Gene Order in *Solanales*. *Life* **2021**, *11*, 769. <https://doi.org/10.3390/life11080769>

Academic Editor: Francois Lefort

Received: 13 June 2021

Accepted: 27 July 2021

Published: 30 July 2021

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



**Copyright:** © 2021 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

Plastid, mitochondrial, and nuclear genomes constitute the total genome of a plant cell. There are around 1000 copies of chloroplast DNA, 100 copies of mitochondrial DNA (mtDNA), and 2 copies of nuclear DNA in a typical *Arabidopsis* leaf cell [1]. In comparison to the conserved chloroplast genome, mitogenomes are unique to each plant, and their analysis will help understand the evolutionary patterns better. Additionally, since mitogenomes contain phylogenetic details, they can reconstruct robust phylogenies in various taxa [2,3]. The mitogenomes show remarkable variation among land plants, and their size is usually larger than plastid genomes. Despite their relatively large size, mitogenomes contain fewer genes than their plastid counterparts [4,5]. Plant mitogenomes are remarkably different from animal mitogenomes [6]. It shows a distinctive 200-fold size divergence, ranging from 66kb in *V. scurruloideum* to 11.3 Mb in *S. conica*, due to the expansion primarily from repeat

sequence, intronic regions, and the incorporation of plastid and nuclear DNA through intracellular gene transfer (IGT) or foreign mtDNA via horizontal gene transfer (HGT) [7,8]. The accumulation of these repeat sequences in vascular plants causes active recombination and serves as a source of rearrangements, leading to the generation of alternative structure, even within a single individual [9]. These rearrangements are facilitated by homologs of bacterial DNA repair proteins, such as RecA and MutS [10]. Due to rearrangements, plant mtDNA is referred to as multipartite, interconverting between a master circle and subgenomic circles or minicircles [11]. Plant mtDNA evolves much more slowly in sequence than animal mtDNA does, considering its structural heterogeneity and gene sequences have meager base substitution rates. This could be because of an active homologous recombination mechanism, part of the repair system that potentially copy-corrects mutations by gene conversion [6]. Also, plant mitogenomes are not restricted to a single origin of replication. In *Arabidopsis thaliana* and *Nicotiana tabacum*, two organellar DNA polymerases are dual-targeted to both mitochondria and plastids, termed Pol1A and Pol1B [12,13].

An essential element of mitochondrial genetics is RNA editing, which is implicated in various plant developmental processes, including organelle biogenesis, environmental adaptability, and signal transduction [14,15]. RNA editing involves converting cytosine to uracil, forming a corrective mechanism for overcoming potentially deleterious mutations [16,17]. The gene family that is responsible for the RNA edits in land plants is the nuclear-encoded pentatricopeptide repeat (PPR) [18]. The nucleotide changes that are created by the editing events mainly occur at the first and second codon positions, leading to amino acid changes that appear to be conserved along with the evolution. These events are suggested to provide a control point at the initiation of translation [19]. Genome sequence data help to understand the degree of variation in editing between recently diverged lineages and the effect of interspecific hybridization on the patterns of editing. Such variations contribute to adaptation and population-level divergence [20]. Even though there is a high gene content conservation between close relatives, their gene order may differ significantly due to the frequent DNA repair via homologous recombination (HR) and non-homologous end-joining (NHEJ). Also, in some cases, there exists a remarkable difference in the shared sequence between closely related plants. So, sufficient sequence data are needed for the efficient comparative study of different phylogenetic groups of plants [21]. More than 100 angiosperm mitogenomes have been sequenced (<http://www.ncbi.nlm.nih.gov/genome/organelle/>, accessed on 9 October 2020), which show a remarkable variation in genome size, content, and structure. However, in angiosperms, sequencing and analysis of mitogenomes are more difficult, due to variations in the non-coding regions and low conservation across species [22]. With recent advances in DNA sequencing technology and efficient bioinformatics tools, there has been rapid growth in plant mitogenome projects and high-quality mitogenome assemblies. *A. thaliana* ecotype C24 mtDNA of 367 kb size was the first angiosperm mitogenome that was entirely sequenced. It codes for 32 proteins, 3rRNAs, and 22tRNAs [23,24]. Recent research has leveraged the growing availability of paired-end libraries and mate-pair libraries to complete mitogenome assemblies.

The mitochondrial genome of no species under the Convolvulaceae family has been sequenced and characterized yet. The family Convolvulaceae includes 85 genera and about 2800 species [25]. The mentioned family has several plants with medicinal properties. *E. alsinoides* is a short, hairy, procumbent, diffuse perennial herb with a small woody branched rootstalk, and is native to South America and widely distributed in tropical and subtropical regions of the world [26,27]. It is included in various traditional medicines, especially in ayurvedic formulations, such as Brahma rasayana, Jeevanyadi ghrita, Brahmi ghrita, Vachadi ghrita, Naladi ghrita, and Agastya rasayana [28]. Root and stem extracts of the plant are used to treat dysentery and depression, while the leaves are used to treat asthma and mental disturbances [29]. The various phytochemicals that are present in the plant possess pharmacological activities such as anxiolytic, anti-amnesic, tranquilizing, anti-depressant, anti-stress, antioxidant, antidiabetic, and anti-ulcer properties [30]. Since

this plant has no complex taxonomic history, the mitogenome of this plant can be used as the first representative from the family *Convolvulaceae* [27,31]. So, the objective of the present study is to sequence and assemble the complete mitogenome of *E. alsinoides*. The complete genome was annotated to analyze its gene content, genome composition, repetitive sequences, RNA editing sites, synonymous (Ks) and nonsynonymous (Ka) substitution rates, codon usage profile, phylogenetic analysis, gene arrangement, synteny analysis, and a complete comparison with all the available mitogenome of *Solanales* was performed. In the future, this will be a valuable resource for further studies on mitogenome in the family *Convolvulaceae*, as it is the first sequenced mitogenome in the family.

## 2. Materials and Methods

### 2.1. Plant Material, Library Preparation and Genome Sequencing

Fresh leaves of a single *E. alsinoides* plant (specimen voucher no. KUBH9910) were collected from Kerala, southern state in India. Total genomic DNA was isolated from the fresh leaves using CTAB (cetyl trimethyl ammonium bromide) method [32]. The sample's purity was evaluated using nanodrop (Nanodrop Technologies, Wilmington, DE, US), and the sample was quantified using the Qubit system (Thermo Fisher Scientific, Waltham, MA, USA). Then, 1 µg high-quality genomic DNA was used for mate-pair library preparation. Libraries were constructed using the Illumina Nextera mate-pair library prep kit (FC-132-1001) (Illumina, San Diego, CA, USA) using a standard protocol. The library quality and quantity were assessed using TapeStation 4200 (Agilent Technologies, Santa Clara, CA, USA). Illumina mate-pair library was prepared with an average insert size of 10 kb and was sequenced using an Illumina HiSeqX—2 × 150 bp PE platform (Illumina, San Diego, CA, USA). The raw reads used for the study are available in the NCBI Sequence Read Archive (SRA) under the SRA accession number: PRJNA551197.

### 2.2. Mitochondrial Genome Assembly

The raw reads quality check was performed using FastQC (v0.10.0) (available online: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>, accessed on 28 January 2020). Preceding to the genome assembly, reads containing adaptor sequences were removed using Cutadapt v1.8 [33], and the quality trimming of the reads was performed using Sickle v1.33 (available online: <https://github.com/najoshi/sickle>, accessed on 9 February 2020). The mate-pair high-quality reads were mapped to the previously published mitogenomes of land plants from NCBI organelle genome resources to filter the mitochondrial reads using BWA v0.7.12-r103 [34]. The filtered reads were de novo assembled using SPAdes-3.13.0 [35]. The resulting pre-assembled contigs were extended and scaffolded using SSPACE v2.0 [36]. The quality of assembly was evaluated using QUAST [37]. The initial assembly generated 4 contigs; three small gaps found in the intergenic region with sizes 4, 7, 5, respectively, were filled with reference genomes of *E. alsinoides* (PRJNA551197) and closely related species to get the complete genome. Gap filling was performed after comparing the mitogenome of *E. alsinoides* and 9 species of the order *Solanales* using BLAST analysis [38] (Supplementary Table S9) [38]. The final assembly was validated by mapping the mitochondrial genome assembly with the Illumina mate-pair reads using BWA v0.7.12-r103 [34].

### 2.3. Mitochondrial Genome Annotation

Preliminary mitogenome annotation was performed using GSeq webserver [39] (available online: <https://chlorobox.mpimp-golm.mpg.de/geseq.html>, accessed on 25 February 2020). Few of the mitochondrial features were not properly annotated using GSeq, hence MITOS [40] (available online: <http://mitos.bioinf.uni-leipzig.de/index.py>, accessed on 6 March 2020) was also used to annotate the genome. A BLAST analysis was performed using the mitogenome as a query and NCBI organelle genome as a database to identify PCGs and rRNAs to validate the above prediction. The genes coordinates, start, and stop position of the genes were manually corrected. The tRNA genes were annotated

using tRNAscan-SE v2.0 [41,42] (available online: <http://lowelab.ucsc.edu/tRNAscan-SE/>, accessed on 10 March 2020) with default settings. OrganellarGenomeDRAW [43] (OGDRAW) (available online: <https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>, accessed on 14 March 2020) program was used to draw the circular map for the genome. The complete mitogenome sequence of *E. alsinoides* was deposited under the GenBank accession number MT081228. The skewness of the nucleotide compositions was measured according to the following formulas: AT skew  $[(A - T)/(A + T)]$  and GC skew  $[(G - C)/(G + C)]$  [44,45].

#### 2.4. Repetitive Sequence Analysis

Forward, reverse, palindrome, and complementary repeats were identified using REPuter v2.74 [46] (available online: <https://bibiserv.cebitec.uni-bielefeld.de/reputer/>, accessed on 23 July 2020), with minimum repeat size taken as 20 bp. Simple sequence repeats SSRs were detected using MISA [47] (available online: <https://webblast.ipk-gatersleben.de/misa/>, accessed on 8 August 2020). A comparison was performed with 9 species of *Solanales*. The results obtained by various programs were manually consolidated, and redundancy was removed.

#### 2.5. RNA Editing Site Prediction

RNA editing site present in the PCGs of *E. alsinoides* mitogenome were predicted using PREP-mt [48] (available online <http://prep.unl.edu/>, accessed on 20 September 2020) and a comparison was made with the 9 selected species of *Solanales*. The cut-off value  $C = 0.6$  was used in the process for accurate prediction.

#### 2.6. Synonymous and Nonsynonymous Substitution Ratio

The 31 PCGs of *E. alsinoides* and 9 species of the order *Solanales* were aligned separately using MAFFT v7 [49] (available online: <https://mafft.cbrc.jp/alignment/server/>, accessed on 3 October 2020) and translated to protein sequence. Nonsynonymous ( $K_a$ ) and synonymous ( $K_s$ ) substitution ratio ( $K_a/K_s$ ) of the mitochondrial PCGs of *E. alsinoides* were predicted using DnaSP v5.10.01 [50,51] to identify the genes that are under selection pressure. Statistical analysis of  $K_a/K_s$  ratio of 19 protein coding genes were carried out by performing one-way Anova using GraphPad prism v8.0.2. (GraphPad Software, San Diego, CA, USA). Codon usage analysis was performed using CodonW v1.4.4 (available online: <http://codonw.sourceforge.net/>, accessed on 19 October 2020).

#### 2.7. Gene Arrangement and Synteny Analysis

Mitogenome organizations of all the species were detected by using MAUVE multiple genome alignments [52]. The gene location and arrangement were verified by the linear genome map generated by OGDRAW. The synteny analysis was performed using MAUVE 2.3.1 [53] software. CREx [54] was used for pairwise comparisons of the gene order. MLGO web server [55] (<http://www.geneorder.org/server.php>, accessed on 7 November 2020) was used to infer a phylogeny from gene order data.

#### 2.8. Phylogenetic Analysis

For phylogenetic analysis, the following two families were included within the order *Solanales*: *Convolvulaceae* and *Solanaceae*. All the mitogenomes under the order *Solanales*, available in the NCBI organelle genome, were selected for phylogenetic analysis. A total of 11 species (Supplementary Table S1) including 8 *Solanales* (3 *Nicotiana* species, 2 *Solanum* species, *C. annuum*, *H. niger*, and *P. orientalis*), 2 *Convolvulaceae* (*I. nil* and *E. alsinoides*), and 1 *Brassicaceae* (*A. thaliana*) as the outgroup were selected for constructing the phylogenetic tree. Maximum likelihood (ML) and Bayesian inference (BI) methods were employed to construct a phylogenetic tree. The mitogenome sequences were retrieved from the NCBI database (<https://www.ncbi.nlm.nih.gov/genbank/>, accessed on 14 December 2020). Nineteen shared PCGs were aligned separately using MAFFT v7 [49]. Gblock

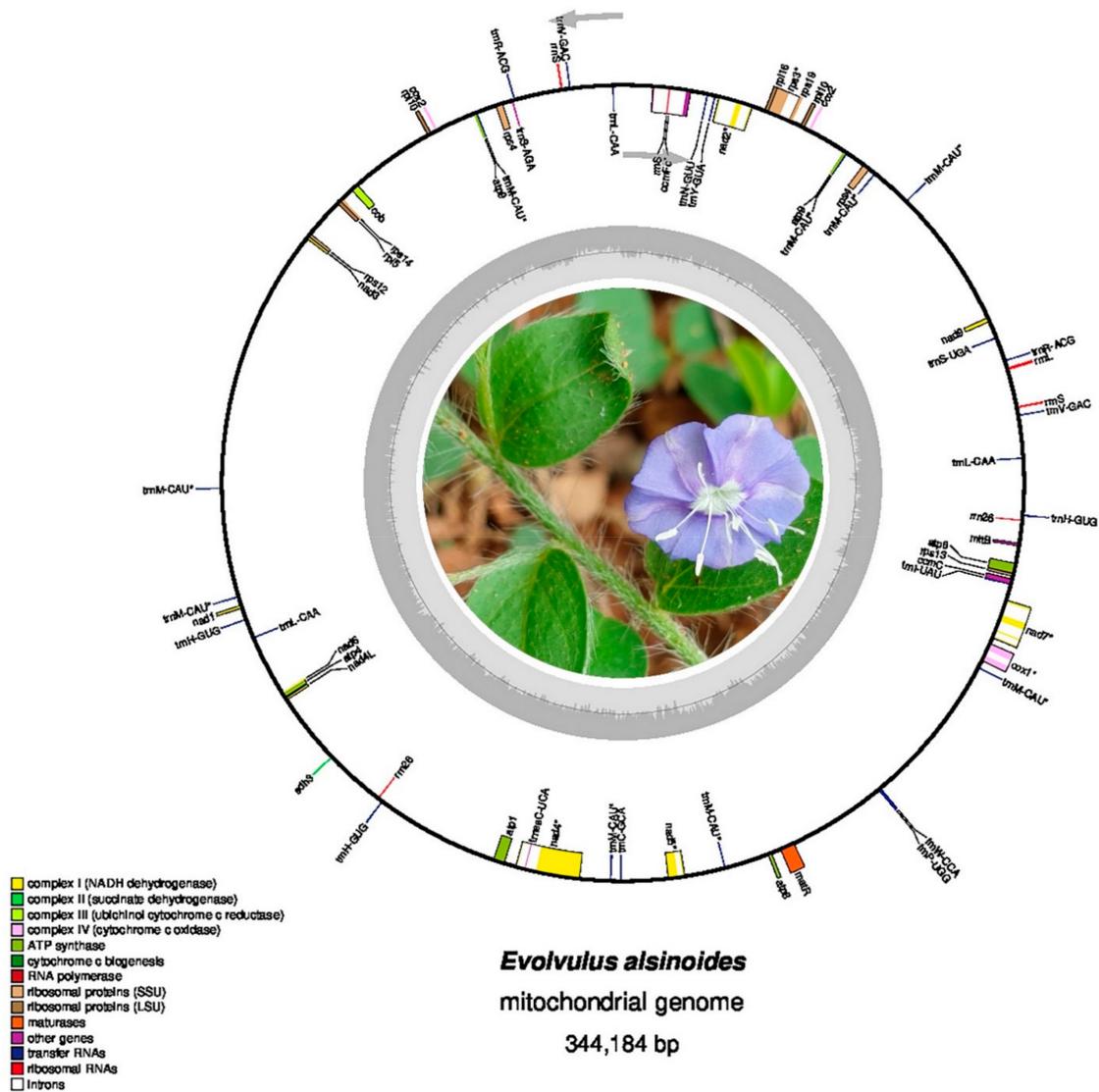
v.0.19b [56] was used to discard the hypervariable regions and poorly aligned nucleotides from the alignment using default parameters. All the individual gene alignments were concatenated to generate a super gene alignment. The best partitioning scheme for ML and BI phylogenetic analysis was determined using PartitionFinder2 [57]. For an additional confirmation Modeltest-NG [58] was performed to identify the best nucleotide substitution model. The model selection was based on Akaike's information criteria (AIC) and Bayesian information criteria (BIC) with a greedy search algorithm and branch lengths linked.

Partitioned  $1_N2_N$  and  $1_N2_N3_{RY}$  sequences were used for constructing the phylogenetic tree using ML and BI methods. The ML phylogenetic tree was constructed using RAxML v8.2.12 [59] with a GTR + G + I model of evolution using 1000 bootstrap replicates. Bayesian inference (BI) was performed using MrBayes v3.2.6 [60] with the model determined by the PartitionFinder2. Bayesian analysis was executed for two million generations, with sampling done at every 100 generations. Metropolis-coupled Markov chain Monte Carlo (MCMCMC) method was used with four chains (one cold chain and three heat chains). The first 5000 sampled trees were discarded as burn-in, and the remaining 15,000 sampled trees were used to construct the consensus tree and posterior probabilities. Two independent runs were used to provide additional confirmation of convergence of posterior probability distribution. The convergence of run was assessed by the log-likelihood (lnL) values and effective sample size (ESS) using Tracer v1.7.2 [61]. The ESS values above 200 were considered as good indicators of convergence [62]. The long branch length of *E. alsinoides* was analyzed by constructing the phylogenetic tree using the following different approaches: (a) after excluding long branch taxa; (b) after excluding outgroup taxa; (c) after eliminating third codon positions partitioning; (d) sampling more taxa [63–67].

### 3. Results and Discussion

#### 3.1. Mitogenome Structure, Organization and Composition

A total of 328,334,850 raw reads were generated after the sequencing, and 133,876,340 high-quality reads were obtained after quality trimming. Out of this, 5,097,201 reads were mapped to the mitochondrial origin. The primary de novo assembly using SPAdes consisted of 3578 contigs with GC% of 39.9% and N50 of 2363 bp. Of these primary contigs, extension and scaffolding using SSPACE resulted in 756 contigs with GC% of 39.6% and N50 of 3326 (Supplementary Table S1). After the gap-filling step, the *E. alsinoides* mitochondrial genome was assembled into a single, circular molecule with a length of 344,184 bp (Figure 1). The nucleotide composition of the whole mitogenome was A: 28.28%, C: 21.73%, G: 21.81%, and T: 28.18%. The AT and GC contents of the mitogenome were found to be 56.46% and 43.54%, respectively. The AT (0.001631254) skew and GC (0.001835083) skew were positive, indicating higher A's and G's in the mitochondrial genome of *E. alsinoides*. The mitogenome encoded 46 unique genes, including 31 PCGs, 12 tRNA genes, and 3 rRNA genes (Table 1, Supplementary Table S2). The complete mitogenome sequence of *E. alsinoides* has been deposited in the GenBank database with the accession number MT081228.



**Figure 1.** The complete Mitogenome map of *E. alsinoides*. Genes represented on the outside of the circle are transcribed counter-clockwise, whereas those inside are transcribed clockwise. The genes are represented as color bars in the circle based on their function. The inner dark grey and light grey circle indicated the GC and AT content of the mitochondrial genome.

**Table 1.** Genes present in the mitochondrial genome of *E. alsinoides*.

Group of Genes	Gene Names
Complex I (NADH dehydrogenase)	nad1, nad2 <sup>a,c</sup> , nad3, nad4 <sup>a</sup> , nad4l, nad5 <sup>a</sup> , nad6, nad7 <sup>a,d</sup> , nad9
Complex II (succinate dehydrogenase)	sdh3
Complex III (ubiquinol cytochrome c reductase)	cob
Complex IV (cytochrome c oxidase)	cox1 <sup>a</sup> , cox2 <sup>b(2)</sup>
Complex V (ATP synthase)	atp1, atp4, atp6, atp8, atp9 <sup>b(2)</sup>
Cytochrome c biogenesis	ccmc, ccmfc <sup>a</sup>
Ribosomal proteins (SSU)	rps3 <sup>a</sup> , rps4 <sup>b(2)</sup> , rps12, rps13, rps14, rps19
Ribosomal proteins (LSU)	rpl5, rpl10 <sup>b(2)</sup> , rpl16
Maturases	matr
Transport membrane protein	mttB
Ribosomal RNAs	rrn26 <sup>b(2)</sup> , rrnS <sup>b(3)</sup> , rrnL <sup>b</sup>
Transfer RNAs	trnC <sup>a</sup> , trnG <sup>a</sup> , trnH <sup>b(4)</sup> , trnL <sup>b(3)</sup> trnM <sup>b(8)</sup> , trnN, trnP, trnR <sup>b(2)</sup> , trnS, trnV <sup>b(2)</sup> , trnW <sup>a</sup> , trnY

<sup>a</sup>—intron-containing genes, <sup>b</sup>—duplicated copies of the genes, number of duplicates are represented in the bracket, <sup>c</sup>—gene containing two introns, <sup>d</sup>—gene containing three introns. Apex notation indicates the number of copies.

### 3.2. Protein-Coding Genes

Among the PCGs, four PCGs (*atp9*, *cox2*, *rps4*, *rpl10*) were duplicated in the mitochondrial genome of *E. alsinoides*. The number of introns varied among the PCGs. The genes *nad2*, *nad4*, *nad5*, *nad7*, *cox1*, *rps3*, and *ccmFc* exhibited a single intron (Table 1), *nad2* possessed two introns, and *nad7* exhibited three introns. The length of the protein-coding regions ranged from 101 to 4506 (Supplementary Table S2). The entire length of PCGs was 26,935 bp among the whole mitogenome. The nucleotide composition of the mitogenome's protein-coding region was A: 27.54%, C: 20.68%, G: 21.28%, and T: 30.50%. The AT and GC content of the mitochondrial protein-coding genes were found to be 58.04% and 41.96%, respectively. The GC (0.014156786) skew was positive, indicating a higher G content than C, whereas the AT (−0.050853963) skew was negative, indicating a higher T content than A among PCGs of the mitochondrial genome.

### 3.3. Transfer and Ribosomal RNA Genes

A total of 12 unique tRNAs were identified in the mitogenome (Supplementary Table S3), and, among these, five tRNAs (*trnH*, *trnL*, *trnM*, *trnR* and *trnV*) were found in multiple copies. Nine copies of *trnM*, three copies of *trnH* and *trnL*, and two copies of *trnR* and *trnV* genes were found in the mitogenome. Three tRNAs (*trnW*, *trnG*, *trnC*) exhibited an intron in their gene. The total size of the tRNA was 1962 bp, with PCGs size ranging from 64 bp to 86 bp (Supplementary Table S3). The secondary structure of the tRNAs showed that most of the tRNAs could be folded into canonical clover-leaf secondary structures, except three (one each for *trnM*, *trnY* and *trnI*). The total size of the rRNA genes was 724 bp, which was composed of duplicated copies of *rrn26*, triplicated copies of *rrnS*, and a single copy of *rrnL* (Table 1, Supplementary Table S2).

### 3.4. Comparative Analysis of Mitogenomes of Solanales

The *E. alsinoides* genome in the present study is an important resource to understand the basic genome features and gene arrangement in the family *Convolvulaceae*. A comparison of the *E. alsinoides* genome with other species of *Solanales* showed that the size of the genome varies among all the species, but the number of PCGs do not vary. *Solanum pennellii* has the highest number of PCGs, and *Ipomoea nil* has the lowest number of PCGs (Table 2). The AT and GC content in the genome range from 54 to 56% and 43 to 45%, respectively (Table 2). The *E. alsinoides* genome has the highest AT% and it has the lowest GC% among all the species (Table 2). *S. pennellii* and *Physochlaina orientalis* have negative AT and GC skew values. *I. nil* and the *Solanum lycopersicum* mitochondrial genome have positive AT skew and negative GC skew values, whereas *Hyoscyamus niger* has negative AT skew and positive GC skew values. Many species (*E. alsinoides*, *Nicotiana Sps.* and *C. annum*) show positive AT and GC skews in the mitochondrial genome, indicating that these genomes have higher As and Gs in their genome. The genome comparison of *E. alsinoides* with *Solanales* exhibited 19 shared genes (Supplementary Table S4).

**Table 2.** Genome comparison of *Solanales*.

Species Name	Genome Size	Total Gene Content	No of PCGs	rRNA	tRNA	AT%	GC%	AT Skew	GC Skew
<i>E. alsinoides</i>	344184	67	35	6	26	56.46	43.54	0.0016	0.0018
<i>I. nil</i>	265768	53	30	3	20	55.55	44.45	0.0075	−0.0012
<i>N. tabaccum</i>	430597	62	35	4	23	55.04	44.96	0.0072	0.0057
<i>N. attenuata</i>	394341	68	40	4	24	55.01	45.05	0.0020	0.0086
<i>N. sylverstris</i>	430597	64	37	4	23	55.04	44.96	0.0072	0.0057
<i>S. lycopersicum</i>	446257	41	39	5	27	54.93	45.07	0.0025	−0.0005
<i>S. pennellii</i>	423596	70	41	5	24	55.01	44.99	−0.0059	−0.0050
<i>C. annum</i>	511530	61	33	3	25	55.48	44.52	0.0069	0.0019
<i>H. niger</i>	501401	70	38	4	28	54.82	45.18	−0.0015	0.0003
<i>P. orientalis</i>	684857	65	37	4	24	55.19	44.81	−0.0001	−0.8216

### 3.5. Repetitive Sequence Analysis

The majority of angiosperm possess different forms of repeats in their mitogenome. Microsatellites or SSRs are short stretches of nucleotide motifs that repeatedly occur in the mitogenome. They have been widely used for species identification, genetic diversity studies, and evolutionary analysis, due to their high polymorphism level among the related species [68,69]. A total of 29 repeats were present in the mitogenome of *E. alsinoides*, including 21 mono-, 4 di-, and 4 tri- repeats. The mono-repeat A/T was highly predominant in the mitogenome, compared to the other types of repeats. The size of the repeats was in the range of 5–14 base pairs. SSR comparison between the mitogenome of *E. alsinoides* and nine species of the order *Solanales* showed that *S.lycopersicum* and *P.orientalis* have the highest repeats. *E. alsinoides* have the least number of repeats among all the species, compared to *S.lycopersicum* and *P.orientalis* (Figure 2A). The length of repeats in all the species of *Solanales* range from 5 bp to 22 bp. The species *P.orientalis* shows the largest repeat size, and *I. nil* has the smallest size. Only mono-, di- and tri- repeats are present in all the species of the order *Solanales* (Supplementary Table S5). The A/T mono-repeats are abundantly present in all the mitogenome of *Solanales*. Among the order *Solanales*, the number of SSRs are comparatively less than other angiosperms, and the repeat size is also less. In addition to the SSRs, a total of 199 repeats were identified in the mitogenome of *E. alsinoides*, which include forward repeats (49), reverse repeat (50), palindromic repeat (50), and complementary repeats (50) (Figure 2B). Based on the comparative analysis among the species of *Solanales*, the total number of repeats present in all the species range from 196 to 200 (Supplementary Table S6). The variation among forward, reverse, palindromic and complementary repeats among the species of *Solanales* are represented in Figure 2B. The size of repeats is almost similar in *N. attenuate*, *S. pennellii* and *E. alsinoides*. The size of forward and palindromic repeats is larger in all the species of *Solanales*. The size of the repeats ranges from 15 bp to 9300 bp. The size of the forward and palindromic repeats was >100 in all the species.

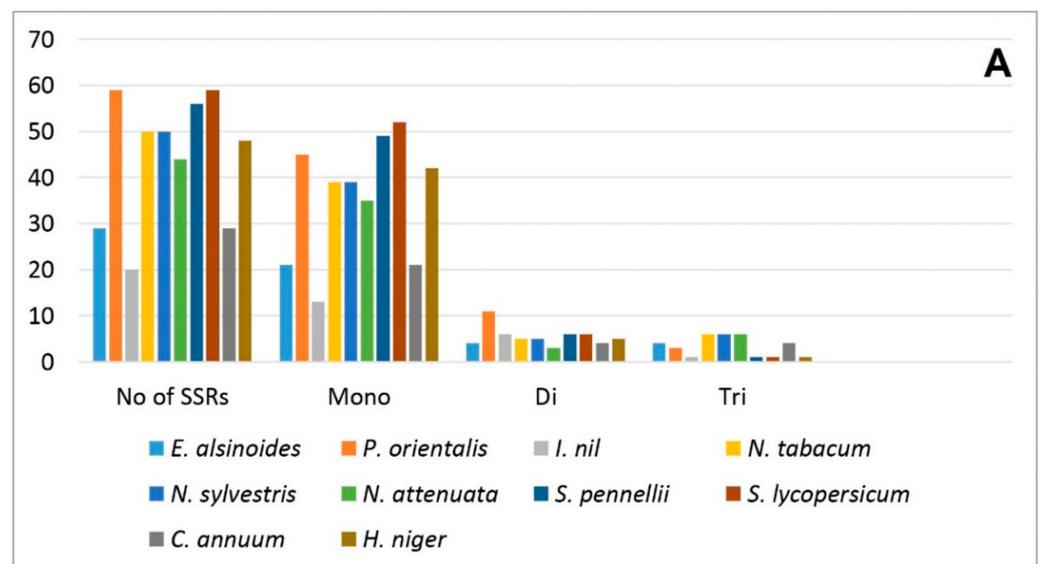
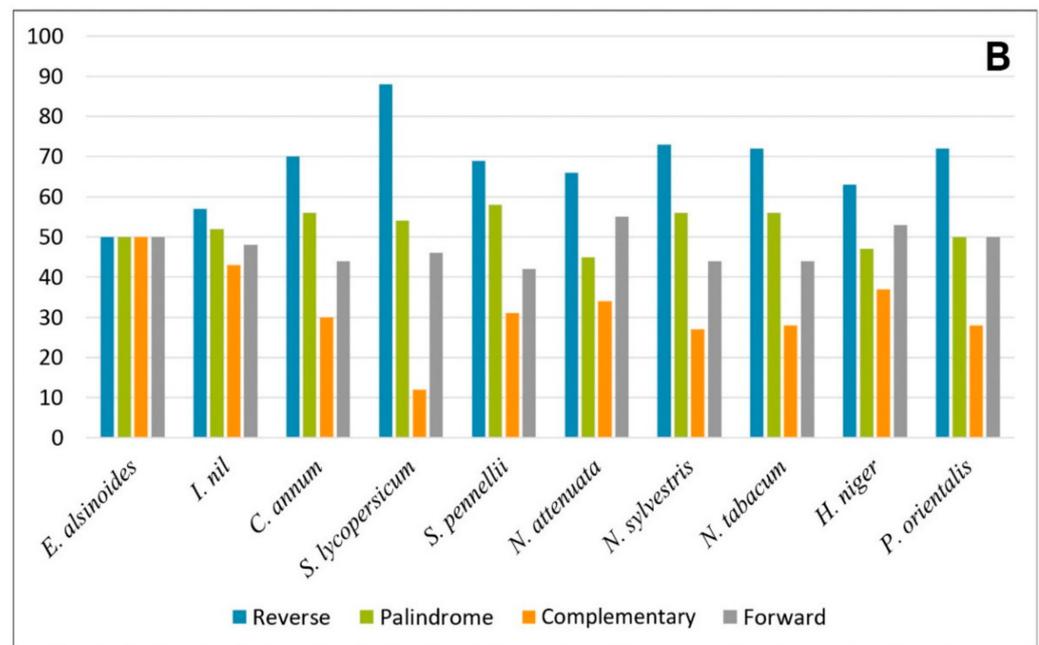


Figure 2. Cont.

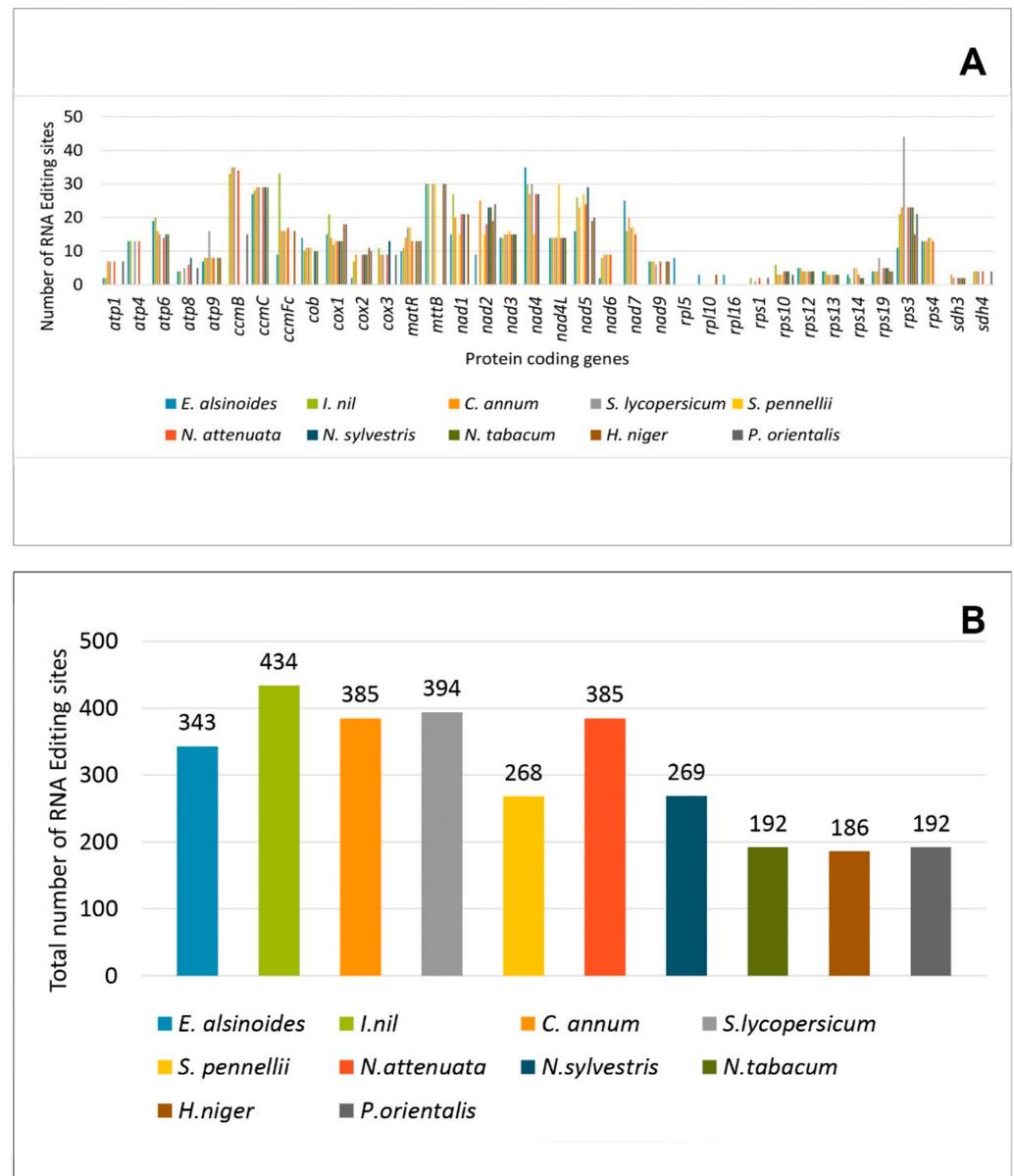


**Figure 2.** Repetitive sequences of all the species. (A) The total number of SSR repeats and different repeats present in each species was represented in the graph. The *x*-axis represents the number and type of SSR and *y*-axis represents the number of repeats. (B) Forward repeats, reverse repeats, palindromic repeats and complementary repeats among all the species of *Solanales*. The *x*-axis represents the name of species, and *y*-axis represents the number of repeats.

### 3.6. RNA Editing Site Prediction

RNA editing is necessary for gene expression in the mitochondrial and chloroplast genomes of all the angiosperms [70–73]. A total of 343 RNA editing sites have been identified in the 31 PCGs of the *E. alsinoides* mitogenome. Among the 343 RNA editing sites, 27.7% (102) were present in the first position of the codon, and 70.3% (368) were present in the second position, and none of the RNA editing sites were present in the third position of the codon. Most RNA editing sites in the third position did not alter the amino acid encoded by the codon. In the selected species of *Solanales* used for present study, all the RNA editing sites are cytidine (C) to uridine (U) transitions. Cytidine (C) to uridine (U) transitions were found in the RNA editing sites of the mitogenome [74]. The most frequently used amino acid replacement that occurred in the RNA editing site of *E. alsinoides* appeared to be leucine. RNA editing sites have been reported in all PCGs except *sdh3*. The PCGs *nad4*, *ccmC*, and *mttB* (27–35 substitutions) have the higher number of RNA editing sites, while *atp1*, *nad6*, *cox2* (2 substitutions) genes have a lesser number of RNA editing sites (Figure 3A). When comparing the RNA editing sites of all the *Solanales*, the *rps3* gene (44 substitutions) of *S. lycopersicum* showed the highest number of RNA editing sites. The *ccmB* gene of *I. nil*, *C. annuum*, *S. lycopersicum*, and *N. attenuata* have 32–35 substitutions for RNA editing sites (Supplementary Table S7). The *ccmC* and *nad4* genes of *E. alsinoides*, *I. nil*, *C. annuum*, *S. lycopersicum*, *Nicotiana attenuata*, and *Nicotiana sylvestris* have the highest number of (27–35) substitutions for RNA editing sites. The *mttB* gene has 30 substitutions for *E. alsinoides*, *I. nil*, *S. lycopersicum*, *S. pennellii*, *H. niger*, and *P. orientalis*. The RNA editing sites cause an alteration in the initiation and termination of the codons. The RNA editing sites that are present in some PCGs of *Solanales* are conserved, except a few genes such as the ribosomal proteins *rpl* and *rsp*. Based on the comparison, 434 RNA editing sites were found in *I. nil*, which encodes the most RNA editing sites, while *H. niger* (186) has the minimum number of RNA editing sites (Figure 3B). The number of RNA editing sites varies among all the species of *Solanales*. The PCGs in the *E. alsinoides* *rps10*, *rps3*, and *nad2* use ACG as their initiation codons, but this changes to AUG during

RNA editing. The same RNA modification occurs in various PCGs among all the species of *Solanales* that are considered for the present study.

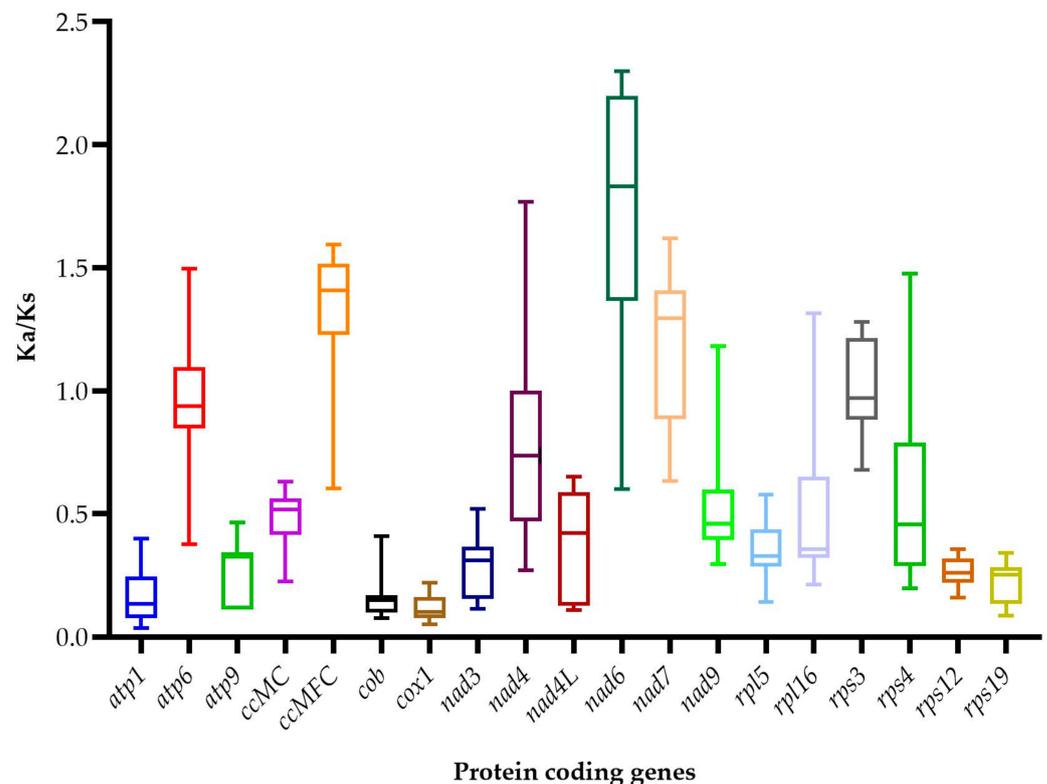


**Figure 3.** RNA editing sites of all the species. (A) RNA editing sites are present in the PCGs. The *x*-axis represents the PCGs in all the species, and *y*-axis represents the number of RNA editing sites. (B) Total number of RNA editing sites present in each species. The *x*-axis represents the number of species, and *y*-axis represents the total number of RNA editing sites.

### 3.7. Synonymous and Nonsynonymous Substitution Ratio

The ratio of nonsynonymous and synonymous substitutions ( $K_a/K_s$ ) were calculated in order to detect the selection pressure among 19 shared PCGs in the mitogenome of *E. alsinoides* and the species of the order *Solanales*. The selective pressure on PCGs evolution is evaluated by comparing the rate of nonsynonymous ( $K_a$ ) and synonymous nucleotide substitutions ( $K_s$ ) [75,76]. The  $K_a/K_s$  ratio was calculated to detect the selection pressure among 19 shared PCGs in the mitogenome of *E. alsinoides* and the species of the order *Solanales*. The  $K_a/K_s$  ratio  $>1$  shows that the nonsynonymous substitutions exceed synonymous substitutions, which implies that the sequences are less conserved throughout evolution [77]. Here, among the 19 genes, 8 genes (*atp6*, *nad6*, *ccmFC*, *nad4*, *rpl16*, *rps3*,

*rps4*, *nad7*) have a  $K_a/K_s$  value of  $>1$  (Figure 4), and they were under positive selection, suggesting that some selective advantage had emerged during evolution, and they were fixed in the population [78]. Eleven genes (*atp1*, *ccmC*, *cob*, *cox1*, *rps19*, *rps12*, *nad3*, *nad9*, *atp9*, *rpl5*, *nad4L*) have a  $K_a/K_s$  ratio  $<1$ ; this indicates that the synonymous substitutions were tolerated compared to nonsynonymous substitution, implying that the sequences are highly conserved. The results showed that the average  $K_a/K_s$  ratio ranges from 0.102 in *cox1* to 0.516 in *ccmC*. The order of the genes are as follows: *cox1*  $<$  *atp1*  $<$  *cob*  $<$  *rps19*  $<$  *rps12*  $<$  *nad3*  $<$  *atp9*  $<$  *rpl5*  $<$  *nad4L*  $<$  *nad9*  $<$  *ccmC*. These genes were well conserved and were under purifying selection. When  $K_a/K_s$  is equal to one, the number of synonymous and nonsynonymous substitutions are equal [77]. One-way ANOVA (Supplementary Table S4) was performed in order to evaluate the statistical significance ( $p < 0.05$ ) of negative selection, using *rps3* as the control group, since most of the mean values fall near one. The result showed that all the genes with a  $K_a/K_s$  ratio of  $<1$  were highly significant, except *nad9*. The lowest  $K_a/K_s$  ratio was for the *cox1* gene, which signifies lesser amino acid changes; hence, it could be used as a molecular marker for species identification and phylogenetic studies. Several studies have explored the potential of the *cox1* gene used as a molecular marker in plants [79,80].



**Figure 4.** Box plot for pairwise divergence nonsynonymous and synonymous ratio ( $K_a/K_s$ ) for 19 shared PCGs of all the *Solanales* mitogenomes.

### 3.8. Codon Usage Analysis

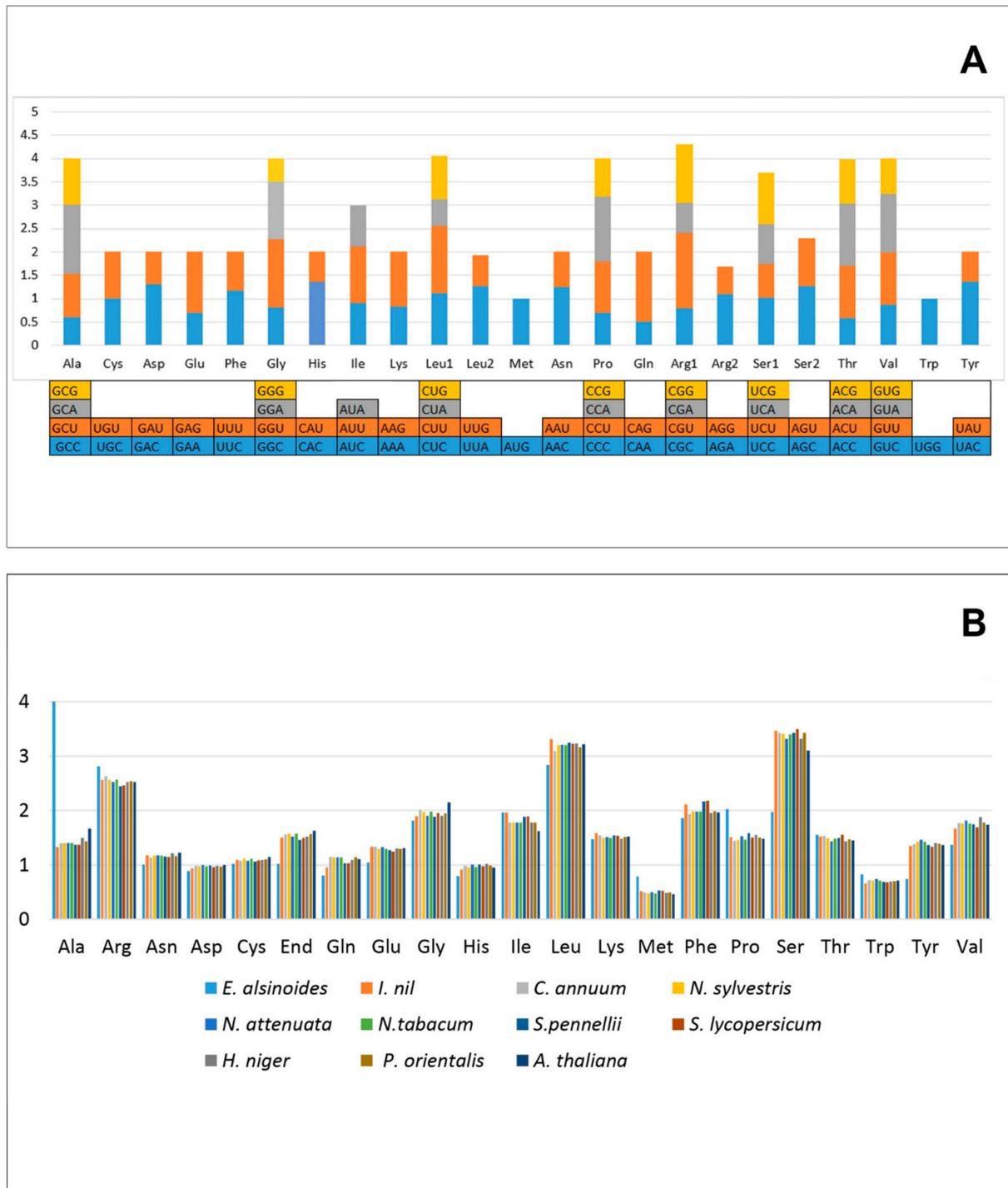
In the mitogenome of *E. alsinoides*, most PCGs have AUG as the start codon and UAA as the stop codons, with a few exceptions in the PCGs *atp6*, *ccmF*, and *sdh3* (Figure 5). The codon usage analysis shows that AGA (Arg) and GCU (Ala) are the most frequently used codons, and GGC (Gly) and CAG (Gln) are the least frequently used codons in the PCGs of the mitogenome. The codon usage bias was measured by calculating the relative synonymous codon usage (RSCU) [81]. The relative synonymous codon usage (RSCU) analysis revealed that all the codons are present in the PCGs. A total of 9023 codons were present in 31 unique PCGs of the *E. alsinoides* mitogenome. The RSCU values  $> 1$  indicate high codon usage bias, and the RSCU values  $< 1$  indicate less codon usage bias in the

mitogenome PCGs. The RSCU values of the NNU and NNA codons are higher than 1.0, except for two codons, Ile (AUA) and Leu (CUA). The RSCU value of 29 codons in the PCGs was >1, which indicates a strong A or T bias in the third position of the codon in the PCGs of the *E. alsinoides* mitogenomes, which is commonly observed in all the studied plant mitogenomes [82]. The effective number of codons (ENC) was calculated for 31 PCGs of *E. alsinoides*, to understand the codon usage analysis. The ENC values were in the range from 37.6 (*sdh3*) to 60.6 (*atp4*), the majority of which were greater than 40, indicating a weak codon bias in the PCGs of the *E. alsinoides* mitogenome. Comparing *E. alsinoides* with the order *Solanales* revealed that most amino acids are similar in all the species, except a few (Ala, Arg, Gln, Glu, Met, Ser, and Tyr). The percentage of the amino acids Ala, Arg, Met, and Trp was high, whereas the amino acids Gln, Glu, and Tyr were less in the *E. alsinoides* mitogenome than other species among *Solanales*. The presence of the amino acid Met was high in the *E. alsinoides* mitogenome. Met is the start codon for most PCGs; hence, alternate start codons are less than other species of *Solanales*.

### 3.9. Gene Arrangement and Synteny Analysis

The genome size, gene content, and gene order in the mitochondrial genome of higher plants vary significantly, and gene loss frequently occurs during evolution [83]. It is evident from the analysis that the gene arrangement is highly variable in the PCGs and tRNAs of the order *Solanales* (Figure 6A). A small gene block consisting of *nad5*, *nad4*, *nad2* was conserved in a few species of *Solanales* (Figure 6B). A total of nine synteny blocks, representing nine genes (*nad9*, *nad2*, *ccmFc*, *nad1*, *nad4*, *nad5*, *matR*, *cox1*, *nad7*), were observed in all the species. The two species *N. sylvestris* and *N. tabacum* have the same number of genes, and the gene arrangement is also similar in both the species. The gene blocks *nad5*, *nad4*, and *nad2* were present in the same gene arrangement in all the three *Nicotiana* species (*N. sylvestris*, *N. tabacum*, *N. attenuate*). *H. niger*, *C. annuum*, and *P. orientalis* show the presence of these gene blocks with gene rearrangement. The position and arrangement of genes varies among all the species that were considered for the study (Supplementary Figure S1). The species that belong to the same family, which are evolutionary related, show the same gene arrangement in their mitogenome [84]. In all other species, such as *S. pennellii*, *S. lycopersicum*, *I. nil* and *E. alsinoides*, there are only *nad5* and *nad4* gene blocks with gene rearrangement. The *nad2* gene is present in some other locations in the mitogenome. The species under the *Solanales* were not closely related, except *N. sylvestris* and *N. tabacum*, hence it could be concluded that a high level of gene rearrangement has occurred in all the species.

The pairwise comparison of the mitochondrial gene order in *Solanales* shows that *N. sylvestris*, *N. tabacum*, *N. attenuate*, and *H. niger* are related, and come under a single group. *C. annuum*, *P. orientalis* and *I. nil* together form another group. The next group of species consisted of *E. alsinoides*, *S. lycopersicum* and *S. pennellii*, which form a single entity (Supplementary Table S8). These results were almost similar by their position in the phylogenetic tree. The gene arrangement shows that the genes *nad4*, *cox1*, and *nad7* are present as blocks, and *rpl5* and *rps12* are present in reverse transposition, but at different locations in the genome of *E. alsinoides* and *S. lycopersicum*. The genes *cox1*, *nad7*, *ccmC*, and *atp6* are present as a block of two genes, but there was transposition of these genes in *E. alsinoides* and *S. pennellii*.



**Figure 5.** Codon usage of the mitochondrial PCGs. (A) Relative synonymous codon usage. Codons are plotted on the x-axis and represented in different colors. RSCU values are plotted on y-axis. (B) Relative synonymous codon usage for all the species. Codons are plotted on the x-axis and RSCU values are plotted on y-axis.



genus *Nicotiana* and *Hyoscamus*. *A. thaliana*, the outgroup that forms a separate branch from all the other species, belongs to the distinctly related family *Brassicaceae*. In the present BI analysis, the ESS values were greater than 200, which indicates good convergence and that the predicted tree was well determined. From the phylogenetic analysis, it is evident that *E. alsinoides* is distinctly related to all the other species of *Solanaceae*. It forms a separate clade in the phylogenetic tree with *I.nil*. *E. alsinoides* forms a long branch length in the constructed phylogenetic tree. There was no change in the constructed phylogenetic tree topology after adopting the methods to avoid long branch lengths. These results reveal that the long branch length of the species of *E. alsinoides* could be due to a lack of taxas in the family *Convolvulaceae*.

The phylogenetic tree was reconstructed using genome rearrangement and gene order data. It also provides an estimate of the ancestral genome. The reconstructed phylogenetic tree forms four clades; clade I and clade II consist of *E. alsinoides* and *I. nil*, which form two separate clades from the ancestor. *H. niger* and *P. orientalis* form clade III. Clade IV consists of *C. annuum*, *Solanum* species, and *Nicotiana* species, divided into two subclades (Figure 7C). From the reconstructed phylogenetic tree, it is evident that highly variable gene orders are present in the species *H. niger*, *P. orientalis*, *I. nil*, and *E. alsinoides*. *I. nil* and *E. alsinoides* belong to the same family *Convolvulaceae*, but they are not closely related genera. Hence, from the reconstructed phylogenetic tree, we can conclude that *E. alsinoides* belong to a separate group from the *Solanales*, and the gene order is also highly variable.

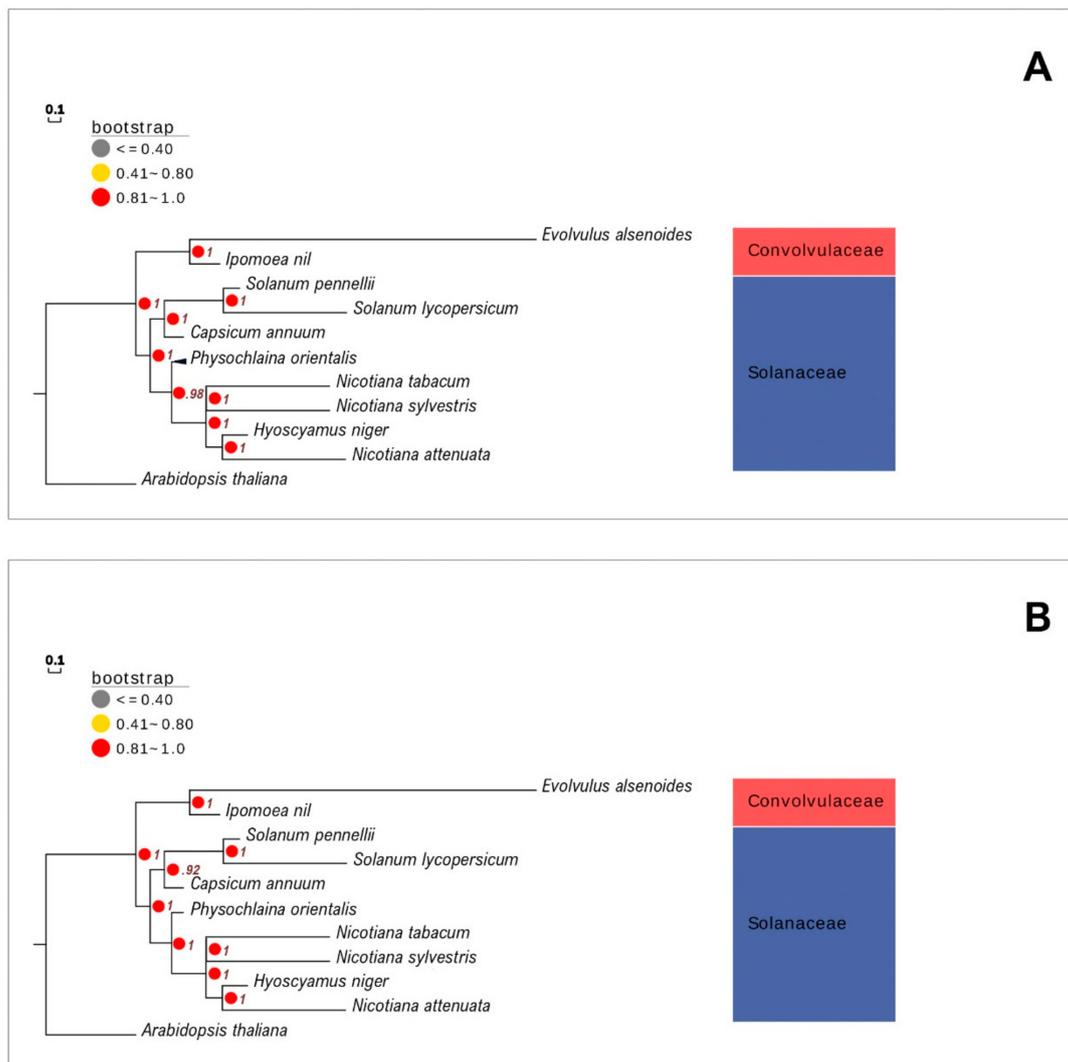
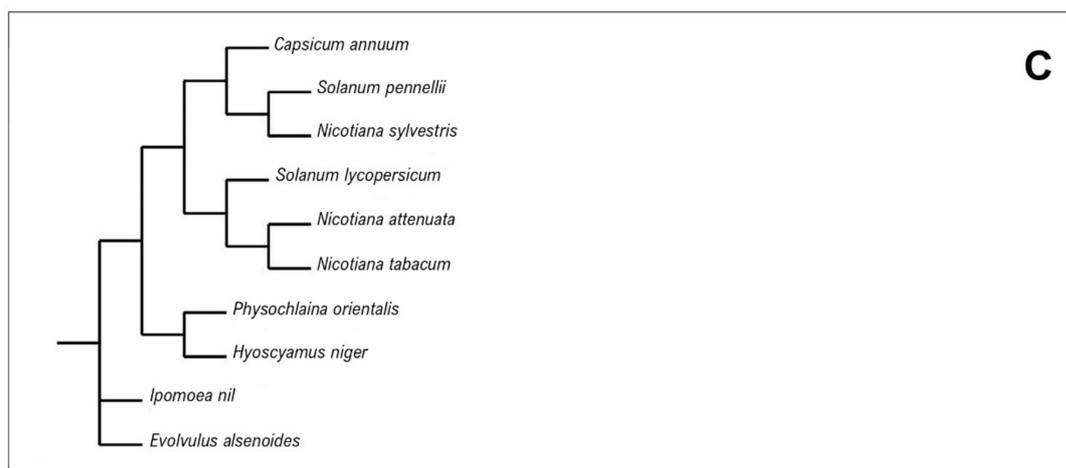


Figure 7. Cont.



**Figure 7.** Original phylogenetic tree and reconstructed phylogenetic tree. (A) Phylogenetic tree based on 19 shared PCGs using maximum likelihood (ML) tree. (B) Phylogenetic tree based on 19 shared PCGs using Bayesian inference (BI). The values in the nodes represent bootstrap value. (C) Reconstructed phylogenetic tree according to the gene order.

#### 4. Conclusions

In summary, *E. alsinoides* is the first sequenced genome in the family Convolvulaceae. So far, only a few mitogenomes are available from the order Solanales. The *E. alsinoides* mitochondrial genome and its detailed analysis will provide a base for future investigations on angiosperms, specifically for the family Convolvulaceae. The complete mitogenome information that is presented in the present study would serve as a valuable genetic resource for further studies on population genetics, species identification, gene arrangement, and evolutionary studies.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/life11080769/s1>, Table S1: Assembly statistics of mitogenome, Table S2: Complete annotation details of mitogenome, Table S3: Details of tRNA, Table S4: List of shared genes and their statistical analysis of Ka/Ks ratio, Table S5: Repeat annotation using MISA, Table S6: Repeat annotation using repeter, Table S7: Prediction of RNA editing sites, Table S8: Gene rearrangement matrix, Table S9: Species details.

**Author Contributions:** Conceptualization, P.R.S. and S.N.A.; data curation, P.R.S., V.C.B., S.A., C.L.V. and K.R.D.; formal analysis, P.R.S., V.C.B., S.A., C.L.V. and K.R.D.; funding acquisition, S.N.A.; investigation, P.R.S., V.C.B., S.A., C.L.V., K.R.D.; methodology, P.R.S. and S.N.A.; project administration, S.N.A. and P.R.S.; software, P.R.S., V.C.B., S.A., C.L.V. and K.R.D.; supervision, S.N.A.; validation, P.R.S. and S.N.A.; visualization, P.R.S.; writing—original draft, P.R.S.; writing—review and editing, P.R.S., V.C.B., S.A., C.L.V., K.R.D. and S.N.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** We acknowledge the funds received from University of Kerala (Plan fund) and also the facility under SIUCEB project, DBT-BIF centre, MHRD-FAST (AiCADD center) in the Department of Computational Biology and Bioinformatics, University of Kerala.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** Our sincere gratitude to P.R. Sudhakaran and Oommen V Oommen for their valuable suggestions. We are grateful to the Campus Computing Facility (CCF) at the Central Laboratory for Instrumentation and Facilitation, University of Kerala for providing the HPC/GPU cluster facility to carry out this research work. We thank AgriGenome Labs, Cochin, Kerala, India, for performing Illumina sequencing.

**Conflicts of Interest:** The authors declare that there is no conflict of interest.

## References

1. Logan, D.C. The mitochondrial compartment. *J. Exp. Bot.* **2006**, *57*, 1225–1243. [[CrossRef](#)] [[PubMed](#)]
2. Bullerwell, C.E.; Gray, M.W. Evolution of the mitochondrial genome: Protist connections to animals, fungi and plants. *Curr. Opin. Microbiol.* **2004**, *7*, 528–534. [[CrossRef](#)] [[PubMed](#)]
3. Mower, J.P. Variation in protein gene and intron content among land plant mitogenomes. *Mitochondrion* **2020**, *53*, 203–213. [[CrossRef](#)]
4. Smith, D.R.; Keeling, P.J. Mitochondrial and plastid genome architecture: Reoccurring themes, but significant differences at the extremes. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 10177–10184. [[CrossRef](#)] [[PubMed](#)]
5. Fauron, C.; Allen, J.; Clifton, S.; Newton, K. Plant mitochondrial genomes. In *Molecular Biology and Biotechnology of Plant Organelles*; Springer: Berlin, Germany, 2004; pp. 151–177.
6. Gualberto, J.M.; Newton, K.J. Plant mitochondrial genomes: Dynamics and mechanisms of mutation. *Annu. Rev. Plant Biol.* **2017**, *68*, 225–252. [[CrossRef](#)] [[PubMed](#)]
7. Shi, Y.; Liu, Y.; Zhang, S.; Zou, R.; Tang, J.; Mu, W.; Peng, Y.; Dong, S. Assembly and comparative analysis of the complete mitochondrial genome sequence of *Sophora japonica* ‘Jinhuai’2’. *PLoS ONE* **2018**, *13*, e0202485. [[CrossRef](#)] [[PubMed](#)]
8. Gui, S.; Wu, Z.; Zhang, H.; Zheng, Y.; Zhu, Z.; Liang, D.; Ding, Y. The mitochondrial genome map of *Nelumbo nucifera* reveals ancient evolutionary features. *Sci. Rep.* **2016**, *6*, 30158. [[CrossRef](#)]
9. Chang, S.; Yang, T.; Du, T.; Huang, Y.; Chen, J.; Yan, J.; He, J.; Guan, R. Mitochondrial genome sequencing helps show the evolutionary mechanism of mitochondrial genome formation in *Brassica*. *BMC Genom.* **2011**, *12*, 497. [[CrossRef](#)]
10. Shedje, V.; Arrieta-Montiel, M.; Christensen, A.C.; Mackenzie, S.A. Plant mitochondrial recombination surveillance requires unusual RecA and MutS homologs. *Plant Cell* **2007**, *19*, 1251–1264. [[CrossRef](#)]
11. Bendich, A.J. Reaching for the ring: The study of mitochondrial genome structure. *Curr. Genet.* **1993**, *24*, 279–290. [[CrossRef](#)]
12. Ono, Y.; Sakai, A.; Takechi, K.; Takio, S.; Takusagawa, M.; Takano, H. NtPoll-like1 and NtPoll-like2, bacterial DNA polymerase I homologs isolated from BY-2 cultured tobacco [*Nicotiana tabacum*] cells, encode DNA polymerases engaged in DNA replication in both plastids and mitochondria. *Plant Cell Physiol.* **2007**, *48*, 1679–1692. [[CrossRef](#)]
13. Cupp, J.D.; Nielsen, B.L. Minireview: DNA replication in plant mitochondria. *Mitochondrion* **2014**, *19*, 231–237. [[CrossRef](#)] [[PubMed](#)]
14. Yan, J.; Zhang, Q.; Yin, P. RNA editing machinery in plant organelles. *Sci. China Life Sci.* **2018**, *61*, 162–169. [[CrossRef](#)] [[PubMed](#)]
15. Giudice, C.L.; Hernández, I.; Ceci, L.R.; Pesole, G.; Picardi, E. RNA editing in plants: A comprehensive survey of bioinformatics tools and databases. *Plant Physiol. Biochem.* **2019**, *137*, 53–61. [[CrossRef](#)]
16. Ichinose, M.; Sugita, M. RNA editing and its molecular mechanism in plant organelles. *Genes* **2017**, *8*, 5. [[CrossRef](#)]
17. Steinhäuser, S.; Beckert, S.; Capesius, I.; Malek, O.; Knoop, V. Plant mitochondrial RNA editing. *J. Mol. Evol.* **1999**, *48*, 303–312. [[CrossRef](#)]
18. Lurin, C.; Andrés, C.; Aubourg, S.; Bellaoui, M.; Bitton, F.; Bruyère, C.; Caboche, M.; Debast, C.; Gualberto, J.; Hoffmann, B. Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* **2004**, *16*, 2089–2103. [[CrossRef](#)]
19. Small, I.D.; Schallenberg-Rüdinger, M.; Takenaka, M.; Mireau, H.; Ostersetzer-Biran, O. Plant organellar RNA editing: What 30 years of research has revealed. *Plant J.* **2020**, *101*, 1040–1056. [[CrossRef](#)]
20. Kovar, L.; Nageswara-Rao, M.; Ortega-Rodriguez, S.; Dugas, D.V.; Straub, S.; Cronn, R.; Strickler, S.R.; Hughes, C.E.; Hanley, K.A.; Rodriguez, D.N. PacBio-based mitochondrial genome assembly of *Leucaena trichandra* (Leguminosae) and an intragenomic assessment of mitochondrial RNA editing. *Genome Biol. Evol.* **2018**, *10*, 2501–2517. [[CrossRef](#)]
21. Negruk, V. Mitochondrial genome sequence of the legume *Vicia faba*. *Front. Plant Sci.* **2013**, *4*, 128. [[CrossRef](#)]
22. Cardi, T.; Giegé, P.; Kahlau, S.; Scotti, N. Expression profiling of organellar genes. In *Genomics of Chloroplasts and Mitochondria*; Springer: Berlin, Germany, 2012; pp. 323–355.
23. Sloan, D.B.; Wu, Z.; Sharbrough, J. Correction of persistent errors in *Arabidopsis* reference mitochondrial genomes. *Plant Cell* **2018**, *30*, 525–527. [[CrossRef](#)]
24. Unsel, M.; Marienfeld, J.R.; Brandt, P.; Brennicke, A. The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nat. Genet.* **1997**, *15*, 57–61. [[CrossRef](#)] [[PubMed](#)]
25. Gupta, R.C. Meiotic studies of the Convolvulaceae Juss. from Indian Hot Desert. *Chromosom. Bot.* **2018**, *12*, 77–85.
26. Priya, T. Antimicrobial Activity of *Evovulus* Alisinoids (L) Extract with Different Organic Solvents in Pathogenic Bacteria and Fungal Species. *Int. J. Appl. Nat. Sci.* **2017**, *6*, 47–54.
27. Austin, D.F. *Evovulus alsinoides* (Convolvulaceae): An American herb in the old world. *J. Ethnopharmacol.* **2008**, *117*, 185–198. [[CrossRef](#)] [[PubMed](#)]
28. Madhavan, V.; Yoganasimhan, N.; Gurudeva, M.R. Pharmacognostical studies on Sankhapushpi (*Convolvulus microphyllum* Sieb. ex Spreng. and *Evovulus alsinoides* (L.) L. *Indian J. Tradit. Knowl.* **2008**, *7*, 529–541.
29. Rajakaruna, N.; Harris, C.S.; Towers, G.H.N. Antimicrobial activity of plants collected from serpentine outcrops in Sri Lanka. *Pharm. Biol.* **2002**, *40*, 235–244. [[CrossRef](#)]
30. Sethiya, N.K.; Nahata, A.; Singh, P.K.; Mishra, S.H. Neuropharmacological evaluation on four traditional herbs used as nervine tonic and commonly available as Shankpushpi in India. *J. Ayurveda Integr. Med.* **2019**, *10*, 25–31. [[CrossRef](#)]

31. Bhatt, M.D.; Kunwar, R.M.; Busmann, R.W.; Paniagua-Zambrana, N.Y.; Turi, M.A. *Evolvulus alsinoides* (L.) L. Convolvulaceae. In *Ethnobotany of the Himalayas*; Kunwar, R.M., Sher, H., Busmann, R.W., Eds.; Springer: Cham, Switzerland, 2021; pp. 1–9.
32. Healey, A.; Furtado, A.; Cooper, T.; Henry, R.J. Protocol: A simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant Methods* **2014**, *10*, 21. [[CrossRef](#)]
33. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **2011**, *17*, 10–12. [[CrossRef](#)]
34. Li, H.; Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **2010**, *26*, 589–595. [[CrossRef](#)]
35. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, S.A.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [[CrossRef](#)]
36. Boetzer, M.; Pirovano, W. SSPACE-LongRead: Scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinform.* **2014**, *15*, 211. [[CrossRef](#)]
37. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [[CrossRef](#)] [[PubMed](#)]
38. Altschul, S.F. BLAST algorithm. *e LS* **2001**. [[CrossRef](#)]
39. Tillich, M.; Lehwar, 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](#)]
40. Bernt, M.; Donath, A.; Jühling, F.; Externbrink, F.; Florentz, C.; Fritzsche, G.; Pütz, J.; Middendorf, M.; Stadler, P.F. MITOS: Improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.* **2013**, *69*, 313–319. [[CrossRef](#)]
41. Chan, P.P.; Lowe, T.M. tRNAscan-SE: Searching for tRNA genes in genomic sequences. In *Gene Prediction*; Springer: Berlin, Germany, 2019; pp. 1–14.
42. 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](#)]
43. Lohse, M.; Drechsel, O.; Kahlau, S.; Bock, R. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res.* **2013**, *41*, W575–W581. [[CrossRef](#)]
44. Perna, N.T.; Kocher, T.D. Unequal base frequencies and the estimation of substitution rates. *Mol. Biol. Evol.* **1995**, *12*, 359.
45. Perna, N.T.; Kocher, T.D. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J. Mol. Evol.* **1995**, *41*, 353–358. [[CrossRef](#)]
46. Kurtz, S.; Choudhuri, J.V.; Ohlebusch, E.; Schleiermacher, C.; Stoye, J.; Giegerich, R. REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **2001**, *29*, 4633–4642. [[CrossRef](#)] [[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](#)]
48. Mower, J.P. PREP-Mt: Predictive RNA editor for plant mitochondrial genes. *BMC Bioinform.* **2005**, *6*, 96. [[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. Librado, P.; Rozas, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **2009**, *25*, 1451–1452. [[CrossRef](#)]
51. Rozas, J.; Sánchez-DelBarrio, J.C.; Messeguer, X.; Rozas, R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **2003**, *19*, 2496–2497. [[CrossRef](#)]
52. Darling, A.C.E.; Mau, B.; Blattner, F.R.; Perna, N.T. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **2004**, *14*, 1394–1403. [[CrossRef](#)]
53. Darling, A.E.; Mau, B.; Perna, N.T. progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* **2010**, *5*, e11147. [[CrossRef](#)] [[PubMed](#)]
54. Bernt, M.; Merkle, D.; Ramsch, K.; Fritzsche, G.; Perseke, M.; Bernhard, D.; Schlegel, M.; Stadler, P.F.; Middendorf, M. CREx: Inferring genomic rearrangements based on common intervals. *Bioinformatics* **2007**, *23*, 2957–2958. [[CrossRef](#)]
55. Hu, F.; Lin, Y.; Tang, J. MLGO: Phylogeny reconstruction and ancestral inference from gene-order data. *BMC Bioinform.* **2014**, *15*, 354. [[CrossRef](#)] [[PubMed](#)]
56. Talavera, G.; Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **2007**, *56*, 564–577. [[CrossRef](#)] [[PubMed](#)]
57. Lanfear, R.; Frandsen, P.B.; Wright, A.M.; Senfeld, T.; Calcott, B. PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* **2017**, *34*, 772–773. [[CrossRef](#)] [[PubMed](#)]
58. Darriba, D.; Posada, D.; Kozlov, A.M.; Stamatakis, A.; Morel, B.; Flouri, T. ModelTest-NG: A new and scalable tool for the selection of DNA and protein evolutionary models. *Mol. Biol. Evol.* **2020**, *37*, 291–294. [[CrossRef](#)] [[PubMed](#)]
59. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [[CrossRef](#)]
60. Ling, C.; Luo, A.; Gao, J. MrBayes 3.2. 6 on Tianhe-1A: A high performance and distributed implementation of phylogenetic analysis. In Proceedings of the 2016 IEEE 22nd International Conference on Parallel and Distributed Systems (ICPADS), Wuhan, China, 13–16 December 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1181–1186.

61. Rambaut, A.; Drummond, A.J.; Xie, D.; Baele, G.; Suchard, M.A. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **2018**, *67*, 901. [[CrossRef](#)]
62. Lanfear, R.; Hua, X.; Warren, D.L. Estimating the effective sample size of tree topologies from Bayesian phylogenetic analyses. *Genome Biol. Evol.* **2016**, *8*, 2319–2332. [[CrossRef](#)]
63. Stefanović, S.; Rice, D.W.; Palmer, J.D. Long branch attraction, taxon sampling, and the earliest angiosperms: Amborella or monocots? *BMC Evol. Biol.* **2004**, *4*, 1–19. [[CrossRef](#)]
64. Soltis, D.E.; Albert, V.A.; Savolainen, V.; Hilu, K.; Qiu, Y.-L.; Chase, M.W.; Farris, J.S.; Stefanović, S.; Rice, D.W.; Palmer, J.D. Genome-scale data, angiosperm relationships, and ‘ending incongruence’: A cautionary tale in phylogenetics. *Trends Plant Sci.* **2004**, *9*, 477–483. [[CrossRef](#)]
65. Ma, P.-F.; Guo, Z.-H.; Li, D.-Z. Rapid sequencing of the bamboo mitochondrial genome using Illumina technology and parallel episodic evolution of organelle genomes in grasses. *PLoS ONE* **2012**, *7*, e30297. [[CrossRef](#)]
66. Bergsten, J. A review of long-branch attraction. *Cladistics* **2005**, *21*, 163–193. [[CrossRef](#)]
67. Siddall, M.E.; Whiting, M.F. Long-branch attractions. *Cladistics* **1999**, *15*, 9–24. [[CrossRef](#)]
68. Ma, Q.; Li, S.; Bi, C.; Hao, Z.; Sun, C.; Ye, N. Complete chloroplast genome sequence of a major economic species, *Ziziphus jujuba* (Rhamnaceae). *Curr. Genet.* **2017**, *63*, 117–129. [[CrossRef](#)] [[PubMed](#)]
69. Huang, J.C.; Sun, M. Genetic diversity and relationships of sweetpotato and its wild relatives in *Ipomoea* series *Batatas* (Convolvulaceae) as revealed by inter-simple sequence repeat (ISSR) and restriction analysis of chloroplast DNA. *Theor. Appl. Genet.* **2000**, *100*, 1050–1060. [[CrossRef](#)]
70. Bock, R.; Khan, M.S. Taming plastids for a green future. *TRENDS Biotechnol.* **2004**, *22*, 311–318. [[CrossRef](#)]
71. Chen, H.; Deng, L.; Jiang, Y.; Lu, P.; Yu, J. RNA Editing Sites Exist in Protein-coding Genes in the Chloroplast Genome of *Cycas taitungensis*. *J. Integr. Plant Biol.* **2011**, *53*, 961–970. [[CrossRef](#)] [[PubMed](#)]
72. Raman, G.; Park, S. Analysis of the complete chloroplast genome of a medicinal plant, *Dianthus superbus* var. *longicalycinus*, from a comparative genomics perspective. *PLoS ONE* **2015**, *10*, e0141329. [[CrossRef](#)]
73. Wakasugi, T.; Hirose, T.; Horihata, M.; Tsudzuki, T.; Kössel, H.; Sugiura, M. Creation of a novel protein-coding region at the RNA level in black pine chloroplasts: The pattern of RNA editing in the gymnosperm chloroplast is different from that in angiosperms. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 8766–8770. [[CrossRef](#)]
74. Moreira, S.; Valach, M.; Aoulad-Aissa, M.; Otto, C.; Burger, G. Novel modes of RNA editing in mitochondria. *Nucleic Acids Res.* **2016**, *44*, 4907–4919. [[CrossRef](#)]
75. Li, W.-H.; Gojobori, T. Rapid evolution of goat and sheep globin genes following gene duplication. *Mol. Biol. Evol.* **1983**, *1*, 94–108. [[PubMed](#)]
76. Hughes, A.L.; Nei, M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **1988**, *335*, 167–170. [[CrossRef](#)] [[PubMed](#)]
77. Roth, C.; Liberles, D.A. A systematic search for positive selection in higher plants (Embryophytes). *BMC Plant Biol.* **2006**, *6*, 1–11. [[CrossRef](#)]
78. Pareek, A.; Mishra, D.; Rathi, D.; Verma, J.K.; Chakraborty, S.; Chakraborty, N. The small heat shock proteins, chaperonin 10, in plants: An evolutionary view and emerging functional diversity. *Environ. Exp. Bot.* **2021**, *182*, 104323. [[CrossRef](#)]
79. Robba, L.; Russell, S.J.; Barker, G.L.; Brodie, J. Assessing the use of the mitochondrial *cox1* marker for use in DNA barcoding of red algae (Rhodophyta). *Am. J. Bot.* **2006**, *93*, 1101–1108. [[CrossRef](#)]
80. Song, H.; Chen, Y.; Gibson, K.; Liu, S.; Yu, Z.; Chen, N. High genetic diversity of the harmful algal bloom species *Phaeocystis globosa* revealed using the molecular marker COX1. *Harmful Algae* **2021**, *107*, 102065. [[CrossRef](#)]
81. Sharp, P.M.; Cowe, E. Synonymous codon usage in *Saccharomyces cerevisiae*. *Yeast* **1991**, *7*, 657–678. [[CrossRef](#)]
82. Bi, C.; Lu, N.; Xu, Y.; He, C.; Lu, Z. Characterization and Analysis of the Mitochondrial Genome of Common Bean (*Phaseolus vulgaris*) by Comparative Genomic Approaches. *Int. J. Mol. Sci.* **2020**, *21*, 3778. [[CrossRef](#)]
83. Richardson, A.O.; Rice, D.W.; Young, G.J.; Alverson, A.J.; Palmer, J.D. The “fossilized” mitochondrial genome of *Liriodendron tulipifera*: Ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. *BMC Biol.* **2013**, *11*, 1–17. [[CrossRef](#)]
84. Weihe, A.; Börner, T. The mitochondrial genome of higher plants. *Proc. Plant Sci.* **1984**, *93*, 305–316.