ACTA ABP BIOCHIMICA POLONICA

Vol. 68, No 2/2021
277–286
https://doi.org/10.18388/abp.2020_5555

Regular paper

Analysis of the complete mitochondrial genome sequence of the resurrection plant *Haberlea rhodopensis*

Vesselin Baev[1,2*], Zdravka Ivanova[1*✉], Galina Yahubyan[1,2], Valentina Toneva[1,2], Elena Apostolova[1,2], Georgi Minkov[1,2] and Ivan Minkov[1,3]

[1]Institute of Molecular Biology and Biotechnology, Plovdiv, Bulgaria; [2]Department of Plant Physiology and Molecular Biology, University of Plovdiv, Plovdiv, Bulgaria; [3]Centre of Plant Systems Biology and Biotechnology, Plovdiv, Bulgaria

***Haberlea rhodopensis* is a paleolithic tertiary relict species that belongs to the unique group of resurrection plants sharing remarkable tolerance to desiccation. When exposed to severe drought stress, this species shows an ability to maintain structural integrity of its deactivated photosynthetic apparatus, which easily reactivates upon rehydration. In addition to its homoiochlorophyllous nature, the resurrection capability of *H. rhodopensis* is of particular importance to the global climate change mitigation. In this study, we sequenced, assembled, and analyzed the mitochondrial (mt) genome of *H. rhodopensis* for the first time. The master circle has a typical circular structure of 484 138 bp in length with a 44.1% GC content in total. The mt genome of *H. rhodopensis* contains 59 genes in total, including 35 protein-coding, 21 tRNAs, and 3 rRNAs genes. 7 tandem repeats and 85 simple sequence repeats (SSRs) are distributed throughout the mt genome. The alignment of 20 plant mt genomes confirms the phylogenetic position of *H. rhodopensis* in the Lamiales order. Our comprehensive analysis of the complete mt genome of *H. rhodopensis* is a significant addition to the limited database of organelle genomes of resurrection species. Comparative and phylogenetic analysis provides valuable information for a better understanding of mitochondrial molecular evolution in plants.**

Keywords: *Haberlea rhodopensis*, resurrection plants, mitochondrial genome, genome assembly

Received: 28 November, 2020; revised: 09 March, 2021; accepted: 26 March, 2021; available on-line: 12 May, 2021

✉e-mail: zivanova@plantgene.eu
Acknowledgements of Financial Support: This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 739582 (Project PlantaSYST).
Availability of data and material: Complete mtDNA of *Haberlea rhodopensis* is deposited in the NCBI GeneBank database under accession number MH757117.
*Joint first authors
Abbreviations: cp, chloroplast; CV, Composition Vector; mt genome, plant mitochondrial genome; SSRs, simple sequence repeats; WGS, whole-genome sequencing

INTRODUCTION

Over the last several years, many plant genomes have been sequenced and assembled by the next-generation sequencing (NGS) technologies. Due to their small size, conserved gene order, and content, chloroplast (cp) genomes were sequenced more frequently than mitochondrial genomes. In comparison to the cp genomes, the mt genome molecules proved to be difficult to assemble because of their variable structure (Bi *et al.*, 2016).

Mitochondria are essential organelles in plants, often called the energy factory of the cell. They possess their own genome which is often used for comparative and evolutionary studies. A recent study highlights mt genes as significant markers for resolving relationships among genera, families, and higher rank taxa across angiosperms. It was observed that the low substitution rates of mt genes in comparison to cp genes make them very useful in the reconstruction of ancient phylogenetic relationships (Qiu *et al.*, 2010). Phylogenetic trees represent the true relationship between conserved core genome sequences and are used to resolve taxonomic grouping among different species. Most mt genomes are extremely large and complex when compared to those of animals and fungi, which generally show a stable and conservative mode of evolution (Li *et al.*, 2009). Land plants' mitochondrial genomes exhibit some features, such as a significant size expansion, frequent gene loss and gene transfer to the nucleus, RNA editing, genomic rearrangement, and replacement of some tRNAs by their cp counterparts (Knoop, 2004), that highlight their evolutionary dynamics. Mitochondrial genomes vary significantly in structure, size, and gene order (Bi *et al.*, 2016). Much of these variations occur even between members of the same family or genera (Alverson *et al.*, 2010). These main features contrast with animal mtDNAs which are structurally conserved, relatively small in size, and have very fast nucleotide substitution rates (Ballard *et al.*, 2004).

Previous studies revealed that the nucleotide substitution rates between animals and plants differ due to the fact that the plant and animal kingdoms diverged about 1000 million years ago and their patterns of evolution have become different. Also, researchers observed significantly different rates in nucleotide substitution rates among plants. Comparison between plant mitochondrial (mt), chloroplast (cp) and nuclear (n) DNA sequences revealed slower substitution rates in the mitochondrial DNA, which is maybe due to a lower mutation rate. In contrast to mammalian mtDNA, where mitochondrial sequences evolve 5 times faster than nDNA, mtDNA sequences in angiosperms evolve at least 5 times more slowly than nDNA (Wolfe *et al.*, 1987).

Plant mt genomes contain various repeat sequences, such as tandem repeat sequences, simple repeat sequences, and large repeats (Alverson *et al.*, 2010, 2011). In many studies it was observed that mt repeats contain valuable genetic information and are regarded as components in intramolecular recombination (Chang *et al.*, 2013). Methylated sites are linked to tandem repeats in the maize mt genome (Clifton, 2004). Simple sequence

repeats play an important role in the evolution of plant mt genomes and are responsible for structural variations and size variability of plant mt genomes (Chang *et al.*, 2013). Some genes with large repeats may have multiple copies. Recombination across large direct repeats may divide mt genomes into pairs of subgenomic molecules, whereas inverted repeats may generate isomeric circles (Handa, 2003; Clifton, 2004; Chang *et al.*, 2013).

The gene content of the plant mt genomes is highly variable. The number of protein-coding genes in the mt genomes varies from 3 to 67, whereas the number of tRNA genes varies from 0 to 27 (Adams & Palmer, 2003). In the course of evolution, many mt genes originally found in plant mt genomes have been lost during transfer to the nucleus (Lei *et al.*, 2013). For example, the *sdh2, rps9, rps11,* and *rps16* genes were lost in the evolution of plant mt genomes. The protein-coding genes: *rps12, sdh3,* and *sdh4* were lost in monocots, while *rps2* was lost in dicots (Zhang *et al.*, 2012).

RNA editing, a post-transcriptional process of changing nucleotide sequence of any RNA molecule, challenged the basic concept of molecular biology that the primary RNA sequence reflects the sequence of DNA from which it is transcribed. The changes encompass insertions and deletions of uridine residues, and a conversion of a cytidine to uridine within the RNA molecule. RNA editing affects the transcripts of protein-coding genes, non-coding transcribed regions, structural RNAs and intron sequences, and serves as a buffer for less preferred mutations in the coding sequences (Covello & Gray, 1989).

In plant organelles, RNA editing sites were found in the coding regions of mRNA, introns, and non-translated regions. The majority of post-transcriptional modifications include U-to-C, A-to-I, and the RNA-editing of C-to-U was identified in most of the angiosperms (Gray *et al.*, 1999). The process of site editing can generate an initiation or termination codon, but in most cases it generates an internal codon with functional relevance (Handa, 2003). Mt and cp RNA editing in plants is essential for the normal functioning of their translation and respiration activity, and is beneficial for understanding gene expression.

Resurrection plants are a group of flora that thanks to unique survival mechanisms evolved over time can survive extreme water shortages for years. Because chloroplasts and mitochondria play an irreplaceable role in stress sensing and responses, studying their genomes is an important prerequisite for understanding their desiccation tolerance. Recently, the cp genomes of the two representatives of resurrection plants – *Boea hygrometrica* (Zhang *et al.*, 2012) and *H. rhodopensis* (Ivanova *et al.*, 2017), both belonging to Gesneriaceae, were sequenced and annotated; this was later followed by the mt genome of *B. hygrometrica*. Here, we report the sequencing data, assembly, and annotation of the mt genome of *H. rhodopensis,* and provide a comparative and phylogenetic analysis that contributes to a better understanding of mitochondrial molecular evolution in plants.

MATERIALS AND METHODS

Plant material and sequencing

Plant material from *H. rhodopensis* was collected from the northeast Rhodopi Mountain, Bulgaria (location42.1'N24.52'E). Total DNA was extracted from the leaf tissue with a DNeasy Plant Mini Kit (QIAGEN), according to the manufacturer's instructions. The quality and quantity of DNA were checked with an Epoch microplate spectrophotometer and an agarose gel. Library preparation and sequencing by HiSeqX Illumina technology were performed at Macrogen (Seoul, South Korea) by Illumina standard protocol. The isolated DNA was used to generate reads with a 150 bp paired-end data library and insert size of 350 bp. DNA sequencing is generated in a total of 2 × 366909885 reads.

Genome assembly and annotation

De novo assembly of the *H. rhodopensis* mt genome (mtDNA) was performed by applying NOVOPlasty (https://github.com/ndierckx/NOVOPlasty), a sole de novo assembler. A seed-and-extend algorithm was used which assembles organelle genomes from whole-genome sequencing (WGS) data, starting from a related or distant single seed sequence with kmer 39 (Dierckxsens *et al.*, 2016). According to the manual of this assembler, we used the total DNA sequencing reads (including nuclear, mt, and cp reads) instead of mapping them to the reference genome and filtering mitochondrial reads. The mtDNA nucleotide sequence of the *cox1* gene from the closely related species *B. hygrometrica* (JN107812) was used as a seed sequence in the process of genome assembly. This was subsequently elongated, resulting in one contig and a successfully circularized genome.

Annotation of the *H. rhodopensis* mt genome was achieved using MITOFY (https://vcru.wisc.edu/cgi-bin/mitofy/mitofy.cgi) with manual start and stop codon correction and validation *via* comparing to the mt genes of previously annotated genomes. The mt gene nomenclature was based on that of published land plant mt genomes available in the NCBI database. Transfer RNA genes (tRNA) were identified by MITOFY and validated by the tRNAscan-SE program (http://lowelab.ucsc.edu/tRNAscan-SE/) with default settings (Schattner *et al.*, 2005). The mt genome circular representation was generated by OrganelllarGenomeDraw (OGDRAW) (Lohse *et al.*, 2013), and the complete mtDNA sequence was deposited in the NCBI GeneBank database under accession number MH757117.

Repeat structures

Mt genome tandem repeats were identified using the Tandem Repeats Finder software with default settings (Benson, 1999). Additionally, we analysed the distribution of simple sequence repeats (SSRs) with the MISA web-based server application (http://pgrc.ipk-gatersleben.de/misa/), with the following settings: 10 repeats for mono-, 5 for di-, and 4 for tri-, tetra-, penta- and hexanucleotide repeat patterns (Liu *et al.*, 2013).

Phylogenetic analysis

We performed a phylogenetic analysis with CVTree3 and a whole-genome based phylogenetic analysis without sequence alignment using a Composition Vector (CV) approach (Qi *et al.*, 2004; Zuo & Hao, 2015). This approach is successfully used in other studies of viruses, fungi, and plastids (Zuo & Hao, 2015), and has demonstrated its applicability in phylogenetic studies using vertebrate mitochondrial genomes.

We used the mt genome of *H. rhodopensis* and 19 genomes from other species (*Boea hygrometrica, Salvia miltiorrhiza, Ajuga reptans, Cucumis sativus, Cucurbita pepo, Ginkgo biloba, Hyoscyamus niger, Populus tremula, Vitis vinifera, Zea perennis, Zea mays, Brassica juncea, Brassica napus, Nicotiana*

tabacum, Oryza sativa, Salix purpurea, Olea europea, Spinacia oleracea, Cannabis sativa), obtained from the NCBI Organelle Genome Resources Web site (https://www.ncbi.nlm.nih.gov/genome/organelle/). 20 homologous protein-coding genes (*atp1, atp4 atp6, atp8, atp9, cob, cox1, cox2, cox3, rps3, rps4, nad1, nad2, nad3, nad4, nad4L, nad5, nad6, nad7, nad9*) were extracted from the genomes represented above in order to generate a phylogenetic tree and estimate evolutionary relationships among the taxa.

Analysis of RNA Editing

We used the web-based software platform PREP (Predictive RNA Editor for Plants) (http://prep.unl.edu/) to identify possible RNA-editing sites in the protein-coding genes of the *H. rhodopensis* mt genome. PREP sites is based on the evolutionary principle that the process of site editing increases protein conservation among species (Mower, 2005). For this analysis, we used the PREP software with default settings and a cut-off value of C=0.2. To determine RNA-editing sites in the *H.rhodopnesis* mt genome, we used a set of land plant protein-coding genes included in the PREP software. Additionally, we performed RNA editing analysis of three closely related genomes from the *Lamiales* order (*O. europaea, S. miltiorrhiza,* and *B. hygrometrica*), and used the results to compare the number of detected RNA sites between these genomes and mt genome of *H. rhodopensis*.

RESULTS AND DISCUSSION

Genome features

The complete mt genome of *H. rhodopensis* is a 484 138 bp long circular DNA molecule (Fig. 1). The summary of the *H. rhodopensis* mt genome features is presented in Table 1. Base composition (27.7% A, 22.1% C, 22.0% G, 28.2% T) is typical for previously published mt genomes from the Lamiales order.

The *H. rhodopensis* mt genome is comprised of 59 unique genes, including 35 protein-coding genes, 21 tRNA genes and 3 rRNAgenes (Table 2). In the group of protein-coding genes, 5 encode subunits for ATP synthase (*atp1, atp4, atp6, atp8, atp9*), 9 – subunits of NADH dehydrogenase (*nad1, nad2, nad3, nad4, nad4L, nad5, nad6, nad7, nad9*), 1 – a subunit of succinate dehydrogenase (*sdh4*), 1 – a subunit of ubiquinol cytochrome *c* reductase (cob), 3 – subunits of cytochrome *c* oxidase (*cox1, cox2, cox3*), 6 – small ribosomal subunits SSU (*rps3, rps4, rps10, rps12, rps13, rps14*), 4 – large ribosomal subunits (*rpl2, rpl5, rpl10, rpl16*), 1 – maturase (*mat*R), 1 – Sec-Y independent transporter *(mttB*), and 4 – subunits for biogenesis of cytochrome *c* (*ccmB, ccmC, ccmFc, ccmFn*). Seven protein-coding genes in the mt genome of *H. rhodopensis* contain introns (*cox1, nad1, nad2, nad4, nad5, nad7, rps3*) (Table 3). The presence of introns was not detected in the genes for tRNA *(trnS-GGA, trnS-UGA, trnS-GCU, trnM-CAU, trnY-GTA, trnI-CAU, trnL-CAA, trnP-UGG, trnD-GUC, trnG-GCC, trnM-CAU, trnF-GAA, trnF-GAA, trnF-CAA, trnW-CCA, trnK-UUU, trnM-CAU, trnN-GUU, trnQ-UUG, trnE-UUC, trnC-GCA)*. The positions, lengths and directions of protein-coding genes are presented in Table 4. According to a previous study examining the distribution of introns in 24 selected plant mt genomes from various taxa (Chlorophyta, Charophyta, Bryophyte, Pteridophyte, Gymnosperms, Monocotyledon, Dicotyledon), most genes do not contain introns in selected species (Xu *et al.*, 2015). The percentage of

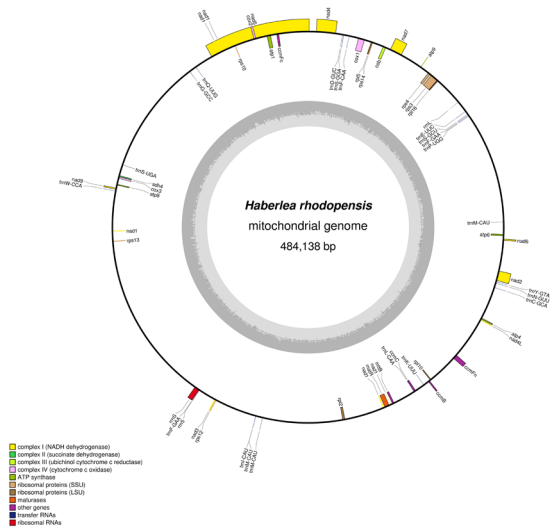

Figure 1. The complete circular mitochondrial genome of *H. rhodopensis*.
Genomic features on transcriptionally clockwise and counter clockwise strands are drawn on the inside and outside of the circle, respectively. Genes belonging to different functional groups are color-coded. GC content is depicted on the inner circle in grey colour.

intronless genes was in the range of 63.2% to 100%, and the intron number varied from 1 to 10 even between species within the same taxonomy level. In addition, we compared the intron content between *H. rhodopensis* and its closest representative (*B. hygrometrica*), also belonging to the Gesneriaceae family. The intron number is 16 and 18 in *H. rhodopensis* and *B. hygrometrica*, respectively. For example *cox2, ccmFc* and *sdh3* are intronless in *H. rhodopensis*, while the corresponding orthologs contain from 1 to 2 introns in *B. hygrometrica*. In contrast, *nad7* is intronless in *B. hygrometrica*, while it contains 3 introns in *H. rhodopensis* (Table 3). This study also provides information on the start and stop codons available in the mt genomes of the studied species and finds that the most common stop codons are TAA, TAG and TGA. Also, the presence of atypical stop codons, such as CAA (present in some species of bryophytes, pteridophyte and gymnosperm), CGA (in the vascular plants), GGT (P. Laevis), AAA and AAT (O. Sativa) is detected in selected species. The lengths of protein-coding genes in H. rhodopensis vary from 225 bp (*atp9*) to 1950 bp (*mat*R). Most of the protein-coding genes start with an AUG codon, except

Table 1. Summary of the complete H.rhodopensis mitochondrial genome

Total mt genome size	484138 bp
Number of genes	59
Number of protein-coding genes	35
tRNA genes	21
rRNA genes	3
A content	27.7%
C content	22.1%
G content	22.0%
T content	28.2%
GC content	44.1%

Table 2. List of genes encoded by *H. rhodopensis* mitochondrial genome

Group of genes	Name of genes		
Complex I (NADH dehydrogenase)	*nad1*	*nad2*	*nad3*
	nad4	*nad4L*	*nad5*
	nad6	*nad7*	*nad9*
Complex II (succinate dehydrogenase)	*sdh4*		
Complex III (ubichinol cytochrome *c* reductase)	*cob*		
Complex IV (cytochrome oxidase)	*cox1*	*cox2*	*cox3*
ATP synthase	*atp1*	*atp4*	*atp6*
	atp8	*atp9*	
Ribosomal proteins (SSU)	*rps3*	*rps4*	*rps10*
	rps12	*rps13*	*rps14*
Ribosomal proteins (LSU)	*rpl10*	*rpl16*	*rpl2*
	rpl5		
Maturases	*matR*		
Other genes	*ccmB*	*ccmC*	*ccmFn*
	ccmFc		
Ribosomal RNAs	*rrn5*	*rrnS*	*rrnL*
Transfer RNAs	*trnC-GCA*	*trnD-GUC*	trnE-UUC
	trnF-CAA	*trnF-GAA(x2)*	*trnG-GCC*
	trnI-CAU	*trnK-UUU*	*trnL-CAA*
	trnM-CAU(x3)	*trnN-GUU*	*trnP-UGG*
	trnQ-UUG	*trnS-GCU*	*trnS-GGA*
	trnS-UGA	*trnW-CCA*	*trnY-GTA*

for *rps4* and *mttB* which use UUG for a start codon – a feature reported in the study of other mt genomes (Wei *et al.*, 2016). 13 genes (*atp1, atp9, ccmB, ccmC, ccmFc, ccmFn, cob, cox3, nad4, nad5, rps13, rps12 and sdh4*) use UGA as a stop codon, 8 genes (*rps3, rps14, nad7, mat*R, *mttB, cox2, atp4 and apt6*) use UAG, and 14 genes (*atp8, cox1, nad1, nad2, nad3, nad4l, nad6, nad9, rpl10, rpsl16, rpsl2, rpl5, rps10* and *rps4*) use UAA.

Analysis of tandem repeats and SSRs

Tandem repeats

Tandem repeats are short lengths of DNA repeated multiple times within the genome. They play an important role in genome rearrangement and recombination (Cavalier-Smith, 2002), and are widely used in phylogenetic and comparative analyses (Nie *et al.*, 2012). Using the Tandem Repeat Finder software, 7 tandem repeats were detected in the mt genome of *H. rhodopensis* (Table 5). The length of TR ranges from 18 to 24 bp. Two of the repeats were observed in a protein-coding region, while others are distributed in non-coding regions.

SRRs

SSRs, or microsatellites, are short DNA motifs that are usually repeated 5–50 times and commonly observed throughout the mt genomes (Chen *et al.*, 2006). They are regarded as molecular markers and widely used in population genetics (Doorduin *et al.*, 2011). Using MISA, 85 SSRs were identified in the *H. rhodopensis* mt genome (Table 6). 42 of them have mononucleotides, 25 – di-nucleotides, and 18 – tri-nucleotides (Table 6). Tetra-, penta- and hexa-nucleotides were not found in the specified setting. Many mononucleotide repeats are comprised of A/T.10 di-nucleotides

Table 3. Intron containing genes of the mt genomes of *H. rhodopensis* and *B. hygrometrica*

Gene name	*H. rhodopesnsis*	*B. hygrometrica*
cox1	1	1
cox2	–	1
nad1	4	1
nad2	2	4
nad4	3	3
nad5	2	4
nad7	3	–
rps3	1	1
ccmFc	–	2
sdh3	–	1

Table 4. Protein coding genes profile and organization – position, direction, start and stop codons

Name	Type	Minimum	Maximum	Length	Direction	Start codon	Stop codon
atp1	CDS	135.857	137.383	1.527	reverse	ATG	ACT
atp4	CDS	446.477	447.07	594	forward	ATG	TAG
atp6	CDS	480.685	481.401	717	reverse	ATG	ATC
atp8	CDS	224.947	225.426	480	reverse	ATG	ATT
atp9	CDS	73.66	73.884	225	forward	ATG	TGA
ccmB	CDS	414.1	414.72	621	forward	ATG	TGA
ccmC	CDS	407.252	408.004	753	reverse	ATG	ACT
ccmFc	CDS	132.609	133.493	885	reverse	ATG	ACT
ccmFn	CDS	427.596	429.323	1.728	forward	ATG	TGA
cob	CDS	89.532	90.725	1.194	reverse	ATG	ACT
cox1	CDS	98.421	100.957	1.584	reverse	ATG	ATT
cox2	CDS	141.755	142.588	834	forward	ATG	TAG
cox3	CDS	221.861	222.658	798	reverse	ATG	ACT
matR	CDS	394.144	396.093	1.95	reverse	ATG	ATC
mttB	CDS	397.435	398.235	801	reverse	AAC	ATC
nad1 mixed	CDS	155.599	396.822	1.128	mixed	ATG	ATT
nad2	CDS	462.917	466.823	1.158	forward	ATG	TAA
nad3	CDS	324.766	325.122	357	forward	ATG	TAA
nad4	CDS	110.119	117.813	1.647	forward	ATG	TGA
nad4L	CDS	445.995	446.297	303	forward	ATG	TAA
nad5	CDS	120.585	160.66	1.59	forward	ATG	TGA
nad6	CDS	478.915	479.532	618	forward	ATG	TAA
nad7	CDS	82.83	87.622	1.275	forward	ATG	TAG
nad9	CDS	226.512	227.084	573	forward	ATG	TAA
rpl10	CDS	415.035	415.523	489	reverse	ATG	ATT
rpl16	CDS	65.037	>65471	>435	reverse	CAC	ATT
rpl2	CDS	376.782	377.768	987	reverse	ATG	ATT
rpl5	CDS	95.209	95.763	555	reverse	ATG	ATT
rps10	CDS	150.291	150.62	330	reverse	ATG	ATT
rps12	CDS	325.171	325.548	378	forward	ATG	TGA
rps13	CDS	247.477	247.827	351	reverse	ATG	ACT
rps14	CDS	94.905	95.207	303	reverse	ATG	ATC
rps3	CDS	65.443	68.608	1.683	reverse	ATG	ATC
rps4	CDS	69.077	70.072	996	reverse	AAC	ATT
sdh4	CDS	221.406	221.933	528	reverse	ATG	ACT

are composed primarily of AT/AT and the rest of SRRs have a high content of A/T. These observations confirm other studies that polyA and polyT repeats are found in the mt genomes (Kuntal *et al.*, 2012).

Comparison with other mt genomes

Plant mt genomes are larger than animal mt genomes and they vary significantly in size, gene order, and content (Alverson *et al.*, 2011). We selected 15 mt plant genomes to compare genome features and detect variability between them and the mt genome of *H. rhodopensis* (Table 7).

CG content, one of significant compositional genome features, differs slightly among the selected genomes. The range is from 43.3% (*B. hygrometrica*) (Zhang *et al.*, 2012) to 50.4% (*Ginkgo biloba*).

The size of selected mt genomes ranges from 219 Kb (*Brassica juncea*) to 773 KB (*Vitis vinifera*). The smallest number of genes (51) is observed in *Vigna angularis*, and the largest (163) in *Nicotiana tabacum*. A significant tRNA gene number variability is detected as well. This

Table 5. Distribution of the repeat sequences in the *H. rhodopensis* mt genome

No	Size(bp)	Location	Repeat
1	24	IGS(trnM-CAU-trnP-UGG)	TTTGTCCAAGCCACTTCTTTTT (x 2.5)
2	21	IGS(trnM-CAU-trnP-UGG)	TTTTTGTCCAAGTCACTTCTT (x 3.2)
3	18	IGS(trnM-CAU-trnP-UGG)	CGATATTGATGCTAGTGA (x 3.3)
4	21	IGS(trnM-CAU-trnP-UGG)	TTCCTTTCAAGCTACTACCAA (x 2)
5	21	nad5	CTTCTTCATTCAATCAGAACT (x 1.9)
6	18	nad5	TTTCTGTTCATGTTTTT (x 2)
7	21	IGS(rps12-trnM-CAU)	TTTCTCACTCCATATATACTT (x 2.1)

varies from 15 in *Cannabis sativa* to 27 in *B. hygrometrica* and *Oryza sativa*. However, a relatively stable content of rRNA genes is found among the selected genomes. Most of them contain 3 rRNA genes, whereas 6 and 5 genes were observed in the genomes of *O. sativa and S. miltiorrhiza*, respectively (Table 8).

Special attention was paid to the comparison of the mt genomes of *H. rhodopensis* and *B. hygrometrica* – the other representative of the Gesneriaceae family belonging to the resurrection plants. The mt genome size of B. *hygrometrica* is 510 KB, which is slightly larger than the mt genome of *H. rhodopensis*, while the two mtDNAs have a similar base composition. Upon comparison between these two genomes, it is evident that although 7 protein-coding genes (*mttB, rpl10, rpl2, rpl5, rps10, rps14, sdh4*) and a *trnI-CAU* tRNA gene are observed in the *H. rhodopensis* mt genome, they are absent in the *B. hygrometrica* genome. 2 protein-coding genes (*rps7 and sdh3*), 4 tRNA genes (*trnR-ACG, trnT-UGU, trnH-GUG, trnW-CCA and trnV-GAC*) and 4 hypothetical proteins (*orf1, orf2, orf3 and orf4*) are present in *B. hygrometrica* mt genome, but are not present in *H. rhodopensis* mt genome. Interestingly, a previous study found that succinate dehydrogenase genes were usually lost in angiosperms, and losses of *sdh4* were predominant in the monocots while no losses were detected in basal angiosperms (Adams *et al.*, 2001). However, our comparison revealed that the *sdh*3 gene exists in the mt genomes of *B. hygrometrica, N. tabacum, G. biloba, O. europea* and *V. vinifera*, while *sdh4* exists in the genomes of *H.rhodopensis, S. miltiorrhiza (pseudo gene), G. biloba, S. purpurea, C. sativa, O. europea and V. vinifera*. The whole set of RNA genes *(trnA, trnC, trnD, trnE, trnF, trnG, trnH, trnI, trnK, trnL, trnM, trnN, trnP, trnQ, trn*R, *trnS, trnT, trnV, trnW, trnY* and *trnfM*) is required for the protein synthesis of mt plant genomes. However, a large number of RNA genes is either lost or deactivated during the evolution of plant mt genomes (Dietrich *et al.*, 1996). We observed that three RNA genes (*trnV, trnT, trn*R) were lost in a large number of plant mt genomes: the *trnV* gene was lost in *B. napus, V. angularis, N. tabacum, B. jun-*

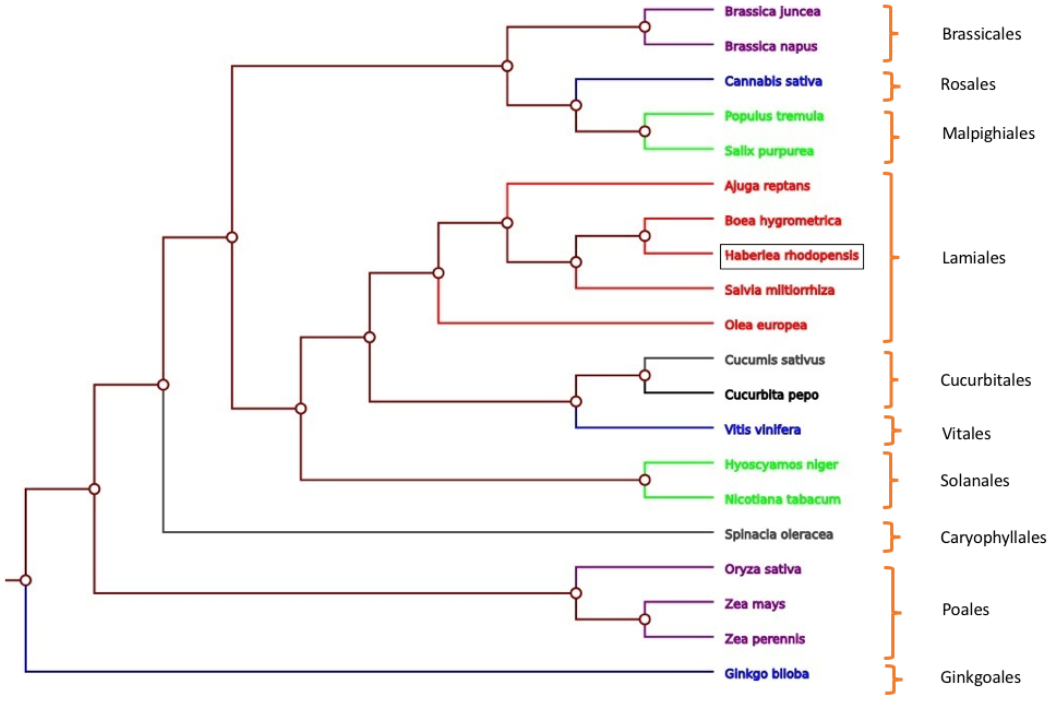

Figure 2. The phylogenetic tree was build based on 20 conserved genes of 20 represented mt genomes.
The plant order of analysed genomes is depicted by different colours.

Table 6. Cumulative SSR frequency and corresponding primer pairs in *H. rhodopensis*.
SSR search parameters: 1-10; 2-5; 3-4; 4-4; 5-4; 6-4 where 1, 2, 3, 4, 5, 6 indicate the mono- di-, tri-, tetra-, penta- and hexa- nucleotide repeats

Total number of sequences examined:	1
Total size of examined sequences (bp):	484138
Total number of identified SSRs:	85
Number of SSR containing sequences:	1
Number of sequences containing more than 1 SSR:	1
Number of SSRs present in compound formation:	6
SSR Type	Frequency
A	20
T	22
AG	5
AT	4
CG	1
GA	5
TA	6
TC	3
TG	1
AAG	1
AGA	1
ATA	3
ATG	1
ATT	1
CAT	1
CTT	4
GAA	1
GGA	1
TAA	1
TAT	1
TTC	2
A/T	42
AC/GT	1
AG/CT	13
AT/AT	10
CG/CG	1
AAG/CTT	9
AAT/ATT	6
AGC/CCT	1
ATC/ATG	2
SSR Type – types of detected SSRs in *H. rhodopensis* genome	
Frequency – frequency of identified SSR motifs	

*cea, O. sativa, D. carota, G. biloba, C. sativa, S. oleracea and H. rhodopensis.*The *trnT* gene was lost in *B. napus, V. angularis, N. tabacum, B. juncea, O. sativa, D. carota, S. miltiorrhiza, G. biloba, S. purpurea, C. sativa, S. oleracea, Z. perennis* and *H. rhodopensis*, while the *trnR* gene was lost in *B. napus, V. angularis, N. tabacum, B. juncea, D. carota, S. miltiorrhiza, C. sativa, S. oleraceae, O. europeae* and *H. rhodopensis*.

Phylogenetic analysis

We analyzed plant mitochondrial phylogeny of 20 conserved homologous mt genes *(atp1, atp4, atp6, atp8, atp9, cob, cox1, cox2, cox3, rps3, rps4, nad1, nad2, nad3, nad4, nad4L, nad5, nad6, nad7, nad9*) from 19 representative plant mitochondrial genomes (*Boea hygrometrica, Salvia miltiorrhiza, Ajuga reptans,Cucumis sativus, Cucurbita pepo, Ginkgo biloba, Populus tremula, Vitis vinifera, Zea perennis, Zea mays, Brassica juncea, Brassica napus, Nicotiana tabacum, Oryza sativa, Salix purpurea, Olea europea, Spinacia oleracea, Cannabis sativa, Hyoscyamus niger*), belonging to 10 orders (*Brassicales, Rosales, Malpighiales, Lamiales, Cucurbitales, Vitales, Solanales, Caryophiallales, Poales*, and *Ginkgoales*) (Fig. 2). We constructed the phylogenetic tree including 20 species from 10 orders, of which 4 species belong to *Lamiales* – *A. reptans, B. hygrometrica, O. europea and S. miltiorrhiza*. The closest mt genome relative of *H. rhodopensis* is that of *B. hygrometrica*. Moreover, these two species share one of the most interesting and unique features – they can survive extreme drought. The conducted phylogenetic analysis and generated tree strongly support the closest relationship between *H. rhodopensis* and *B. hygrometrica*, as well as confirm that these two species belong to the *Gesneriaceae* family.

Analysis of RNA Editing

Using the PREP program, we analysed 35 protein-coding genes from the *H. rhodopensis* mt genome and we predicted 419 RNA editing sites inside them (Supplementary Table 1 at https://ojs.ptbioch.edu.pl/index.php/abp). We discovered that the NAD(H) complex contains 153 editing sites (36.51% of all predicted sites). 5 genes (*ccmB, ccmC, ccmFn, mttB* and *nad4*) encoded most of the RNA editing sites (156), while 4 genes (*atp1, atp8, atp9* and *rpl10*) encoded the fewest sites. Among the 419 editing sites, 102 (24.34%) were converted from Serine to Leucine, 88 (21%) were converted from Proline to Leucine and 63 (15%) were converted from Serine to Phenylalanine. The other 166 amino acids are converted between different types of amino acids. Also, we calculated the codon position where these changes happened. 117 (27.92%) of the RNA editing sites occurred in the first position of the codon, whereas 284 (67.78%) were in the second position.

To compare the RNA editing sites between the closely related mt genomes of the *Lamiales* order, we additionally analysed protein-coding genes from *B. hygrometrica*, *O. europaea,* and *S. miltiorrhiza* mt genomes. The number of predicted RNA editing sites is 431 for *S. miltiorrhiza* (Supplementary Table 2 at https://ojs.ptbioch.edu.pl/index.php/abp), 459 for *O. europaea* (Supplementary Table 3 at https://ojs.ptbioch.edu.pl/index.php/abp), and 389 for *B. hygrometrica* (Supplementray Table 4 at https://ojs.ptbioch.edu.pl/index.php/abp). We observed that the number of RNA editing sites under amino acid changes in the compared mt genomes is similar to that of the *H. rhodopensis* mt genome (Fig. 3). This analysis revealed that the NAD(H) complex in these three genomes contains a fair amount of RNA editing sites, i.e. 175, 164, and 170 for *O. europaea*, *S.*

Table 7. List of mitochondrial genomes used for the comparative analysis

No.	Taxon	Family	Order	GenBank Accession number
1	*Cannabis sativa*	Cannabaceae	Rosales	KU310670
2	*Spinacia oleracea*	Chenopodiaceae	Caryophyllales	KY768855
3	*Brassica juncea*	Brassicaceae	Brassicales	JF920288
4	*Brassica napus*	Brassicaceae	Brassicales	AP006444.1
5	*Nicotiana tabacum*	Solanaceae	Solanales	BA000042
6	*Zea perennis*	Poaceae	Poales	NC_008331
7	*Vitis vinifera*	Vitaceae	Vitales	NC_012119
8	*Oryza sativa*	Poaceae	Poales	JN861112
9	*Salix purpurea*	Salicaeae	Malpighiales	KU198635
10	*Daucus carota*	Apiales	Scandiceae	JQ248574
11	*Olea europea*	Oleaceae	Lamiales	MG372116
12	*Salvia miltiorrhiza*	Lamiaceae	Lamiales	KF177345
13	*Boea hygrometrica*	Gesneriaceae	Lamiales	JN107812
14	*Ginkgo biloba*	Ginkoaceae	Ginkoales	KM672373
15	*Vigna angularis*	Fabaceae	Fabales	AP012599

miltiorrhiza, and *B. hygrometrica*-, respectively, which is similar to what was observed in the *H. rhodopensis* mt genome.We detected that 5 genes (*ccmB, ccmC, ccmFn, mttB* and *nad4*) in *O. europaea* and *S. miltiorrhiza,* and 4 genes (*ccmB, ccmC, ccmFn* and *nad4*) in *B. hygrometrica* encoded most of RNA editing sites, i.e. 160, 153 and 128, respectively. Among all RNA editing sites, we detected that 110, 96 and 99 were converted from Serine to Leucine, 104, 80 and 91were converted from Proline to Leucine, and 68, 62, and 66 – from Serine to Phenylalanine, for *O. europaea*, *B. hygrometrica,* and *S. miltiorrhiza*, respectively. We calculated the position of the codon substitutions in these genomes and we observed that 136 (*S. miltiorrhiza*), 137 (*O. europaea*) and 115 (*B. hygrometrica*) sites occurred in the first position, whereas 281 (*S. miltiorrhiza*), 304 (*O. europaea*) and 261 (*B. hygrometrica*) occurred in the second.

In addition, we compared the RNA editing data of *O. sativa (Poales) and Brassica napus (Brassicales) (Maier et al., 1996) with representatives from the Lamiales order.* The most edited transcripts were *ccmC* with 36 editing sites, in *O. sativa* (Notsu *et al.*, 2002), and *ccmB* with 39 editing sites in *Brassica napus* (Handa, 2003). We compared these results to those of the Lamiales representatives analysed in our study, and found that, similar to *O. sativa* and *B. napus*, one of the most edited transcripts belongs to the cytochrome-*c* familly.

CONCLUSIONS

The main characteristics of the *H. rhodopensis Friv.* mitochondrial genome, including its variable size and structure, as well as fluctuating gene order and content, are consistent with the mitochondrial genomes of most higher plants. Its sequencing and annotation that were performed here constitute an important addition to the limited body of sequenced and analyzed mt genomes from the Gesneriaceae family, especially when it comes to resurrected plants. The comparative and phylogenetic analysis presented here provides a greater understanding of mt molecular evolution in higher plants, facilitating further study of gene organization and evolution in the Lamiales genus.

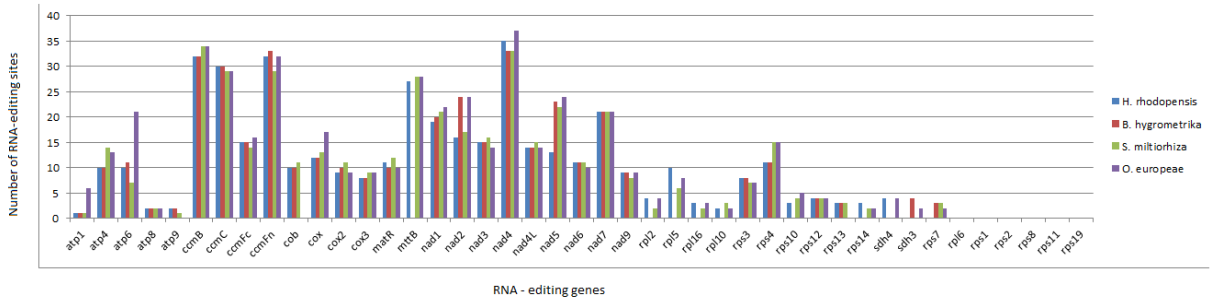

Figure 3. The distribution and comparison of RNA-editing sites in the protein-coding genes of *H. rhodopesnsis* mt genome and three closely related mt genome from the Lamiales order.

Table 8. Comparison of gene content among various plant mt genomes

No.	Taxon	Genes	CDS	tRNA	rRNA
1	*Cannabis sativa*	55	35	15	3
2	*Spinacia oleracea*	55	29	23	3
3	*Brassica juncea*	99	78	18	3
4	*Brassica napus*	106	79	17	3
5	*Nicotiana tabacum*	193	156	23	4
6	*Zea perennis*	59	32	17	3
7	*Vitis vinifera*	161	74	31	3
8	*Oryza sativa*	100	67	27	6
9	*Salix purpurea*	55	32	21	3
10	*Daucus carota*	122	90	25	4
11	*Olea europea*	71	41	26	3
12	*Salvia miltiorrhiza*	167	138	22	5
13	*Boea hygrometrica*	66	32	27	4
14	*Ginkgo biloba*	66	40	23	3
15	*Vigna angularis*	51	26	16	3

It is well known that higher resurrection plants maintain their energetic status during periods of dehydration and hibernation (Dinakar & Bartels, 2013; Gechev *et al.,* 2013). It is well established that Haberlea respiration remains energetically stable during leaf drying (Kimenov & Minkov, 1975; Minkov *et al.,* 1977). Undoubtedly, mitochondria play an important role in the process of rehydration and "resurrection" of these plants. These processes most likely differ in *H. rhodopensis Friv.* and other resurrecting plants (such as *B. hygrometrica*), owing to significant differences in their respective mt genes. These "resurrection" mechanisms, therefore, may have evolved independently of the evolution of their respective mt gene arrangements. Furthermore, the emergence of this process appears to be the result of regulation, rather than specific gene composition. High levels of NADPH and other phosphorylated sugars found in Haberlea's mitochondria during drought and recovery support this view. These sugars most likely maintain the plant's high energy status during the periods of stress (Bardarov *et al.,* 2018).

We can further speculate that convergent evolution in higher plants has produced "resurrection" mechanisms multiple times. Despite their advantage for individual organisms, however, many instances appear to have subsequently become vestigial due to their extremely complex, onerous maintenance, and dubious benefits for the overall population. Immortality of individual organisms, therefore, appears to involve more than their own self-interest. System complexity, energy balance, and evolutionary cost for the entire population all play an important role in shaping evolution. Clearly, immortality of the individual organisms has evolved many times before. Today, only its residual traces remain within the genomes, physiology and biochemistry of some plants. Regardless, these vestigial "resurrection" mechanisms can still be reactivated when the evolutionary pressure of colonizing new spaces outweighs other concerns.

REFERENCES

Adams KL, Rosenblueth M, Qiu YL, Palmer JD (2001) Multiple losses and transfers to the nucleus of two mitochondrial succinate dehydrogenase genes during angiosperm evolution. *Genetics* **158**: 1289–1300

Adams KL, Palmer JD (2003) Evolution of mitochondrial gene content: Gene loss and transfer to the nucleus. *Mol. Phylogenet. Evol.* **29**: 380–395. https://doi.org/10.1016/S1055-7903(03)00194-5

Alverson AJ, Wei X, Rice DW, Stern DB, Barry K, Palmer JD (2010) Insights into the evolution of mitochondrial genome size from complete sequences of citrullus lanatus and cucurbita pepo (Cucurbitaceae). *Mol. Biol. Evol.* **27**: 1436–1448. https://doi.org/10.1093/molbev/msq029

Alverson AJ, Zhuo S, Rice DW, Sloan DB, Palmer JD (2011) The mitochondrial genome of the legume vigna radiata and the analysis of recombination across short mitochondrial repeats. *PLoS One* **6**: https://doi.org/10.1371/journal.pone.0016404

Ballard J, William O, Whitlock M (2004) The incomplete natural history of mitochondria. *Mol. Ecol.* **13**: 729–744. https://doi.org/10.1046/j.1365-294X.2003.02063.x

Bardarov K, Apostolova E, Toneva V, Minkov I (2018) HILIC LC-MS metabolite profiling of Nucleotide phosphates, deoxyNucleotide phosphates and amino acids in *Haberlea rodopensis* Friv. leaves and roots-changes during drought and rehydration. *Resurrect. Plants Hope Crop drought Toler. ReHOPE, FEBS Adv. courses*

Benson G (1999) Seminar – Tandem repeats finder: a program to analyze DNA sequences. **27**: 573–580

Bi C, Paterson AH, Wang X, Xu Y, Wu D, Qu Y, Jiang A, Ye Q, Ye N (2016) Analysis of the complete mitochondrial genome sequence of the diploid cotton *Gossypium raimondii* by comparative genomics approaches. *Biomed Res. Int.* **2016**: https://doi.org/10.1155/2016/5040598

Cavalier-Smith T (2002) Chloroplast evolution: secondary symbiogenesis and multiple losses chloroplasts originated from cyanobacteria only. *Curr. Biol.* **12**: 62–64. https://doi.org/10.1016/s0960-9822(01)00675-3

Chang S, Wang Y, Lu J, Gai J, Li J, Chu P, Guan R, Zhao T (2013) The mitochondrial genome of soybean reveals complex genome structures and gene evolution at intercellular and phylogenetic levels. *PLoS One* **8**. https://doi.org/10.1371/journal.pone.0056502

Chen C, Zhou P, Choi YA, Huang S, Gmitter FG (2006) Mining and characterizing microsatellites from citrus ESTs. *Theor. Appl. Genet.* **112**: 1248–1257. https://doi.org/10.1007/s00122-006-0226-1

Clifton SW (2004) Sequence and comparative analysis of the maize NB mitochondrial genome. *Plant Physiol.* **136**: 3486–3503. https://doi.org/10.1104/pp.104.044602

Covello PS, Gray MW (1989) RNA Editing in plant mitochondria. *Nature* **341:** 662–666. https://doi.org/10.1038/341662a0

Dierckxsens N, Mardulyn P, Smits G (2016) NOVOPlasty: *De novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* **45**: e1. https://doi.org/10.1093/nar/gkw955

Dietrich A, Small I, Cosset A, Weil JH, Maréchal-Drouard L (1996) Editing and import: Strategies for providing plant mitochondria with a complete set of functional transfer RNAs. *Biochimie* **78**: 518–529. https://doi.org/10.1016/0300-9084(96)84758-4

Dinakar C, Bartels D (2013) Desiccation tolerance in resurrection plants: New insights from transcriptome, proteome, and metabolome analysis. *Front. Plant Sci.* **4**: 1–14. https://doi.org/10.3389/fpls.2013.00482

Doorduin L, Gravendeel B, Lammers Y, Ariyurek Y, Chin-A-Woeng T, Vrieling K (2011) The complete chloroplast genome of 17 individuals of pest species *Jacobaea vulgaris*: SNPs, microsatellites and barcoding markers for population and phylogenetic studies. *DNA Res.* **18**: 93–105. https://doi.org/10.1093/dnares/dsr002

Gechev TS, Dinakar C, Benina M, Toneva V, Bartels D (2013) Molecular mechanisms of desiccation tolerance in resurrection plants. *Cell. Mol. Life Sci.* **69**: 3175–3186. https://doi.org/10.1007/s00018-012-1088-0

Gray MW, Burger G, Lang BF (1999) Mitochondrial evolution. *Science* **283**: 1476–1481. https://doi.org/10.1126/science.283.5407.1476

Handa H (2003) The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.): Comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. *Nucleic Acids Res.* **31**: 5907–5916. https://doi.org/10.1093/nar/gkg795

Ivanova Z, Sablok G, Daskalova E, Zahmanova G, Apostolova E, Yahubyan G, Baev V (2017) Chloroplast genome analysis of resurrection tertiary relict *Haberlea rhodopensis* highlights genes important for desiccation stress response. *Front. Plant Sci.* **8**: 1–15. https://doi.org/10.3389/fpls.2017.00204

Kimenov G, Minkov IN (1975) On the behavior of *Haberlea rhodopensis* Friv. and *Ramonda serbica* Panč. to the poikiloxerophytic type of plants. *C.R. Acad. Bulg. Sci.* **28**: 829

Knoop V (2004) The mitochondrial DNA of land plants: Peculiarities in phylogenetic perspective. *Curr. Genet.* **46**: 123–139. https://doi.org/10.1007/s00294-004-0522-8

Kuntal H, Sharma V, Daniell H (2012) Microsatellite analysis in organelle genomes of Chlorophyta. *Bioinformation* **8**: 255–259. https://doi.org/10.6026/97320630008255

Lei B, Li S, Liu G, Chen Z, Su A, Li P, Li Z, Hua J (2013) Evolution of mitochondrial gene content: Loss of genes, tRNAs and introns between *Gossypium harknessii* and other plants. *Plant Syst. Evol.* **299**: 1889–1897. https://doi.org/10.1007/s00606-013-0845-3

Li L, Wang B, Liu Y, Qiu YL (2009) The complete mitochondrial genome sequence of the hornwort megaceros aenigmaticus shows a mixed mode of conservative yet dynamic evolution in early land plant mitochondrial genomes. *J. Mol. Evol.* **68**: 665–678. https://doi.org/10.1007/s00239-009-9240-7

Liu G, Cao D, Li S, Su A, Geng J, Grover CE, Hu S, Hua J (2013) The complete mitochondrial genome of *Gossypium hirsutum* and evolutionary analysis of higher plant mitochondrial genomes. *PLoS One* **8**: 1–14. https://doi.org/10.1371/journal.pone.0069476

Lohse M, Drechsel O, Kahlau S, Bock R (2013) OrganellarGenomeDRAW – a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res.* **41**: 575–581. https://doi.org/10.1093/nar/gkt289

Maier R, Zeltz P, Kössel H, Bonnard G, Gualberto J, Grienenberger J (1996) RNA editing in plant mitochondria and chloroplasts. *Plant Mol Biol* **32**: 343–365. https://doi.org/10.1007/BF00039390

Minkov I, Kimenov G, Kalucheva I (1977) Structural characteristics of the chloroplasts of *Haberlea rhodopensis* Friv. upon drying and restoration. *C. R. Acad. Bulg. Sci.* **30**: 897

Mower JP (2005) PREP-Mt: Predictive RNA editor for plant mitochondrial genes. *BMC Bioinformatics* **6**: 1–15. https://doi.org/10.1186/1471-2105-6-96

Nie X, Lv S, Zhang Y, Du X, Wang L, Biradar SS, Tan X, Wan F, Weining S (2012) Complete chloroplast genome sequence of a major invasive species, crofton weed (*Ageratina adenophora*). *PLoS One* **7**: https://doi.org/10.1371/journal.pone.0036869

Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, Hirai A, Kadowaki K (2002) The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: Frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Mol. Genet. Genomics* **268**: 434–445. https://doi.org/10.1007/s00438-002-0767-1

Qi J, Wang B, Hao BI (2004) Whole proteome prokaryote phylogeny without sequence alignment: A K-string composition approach. *J. Mol. Evol.* **58**: 1–11. https://doi.org/10.1007/s00239-003-2493-7

Qiu YL, Li L, Wang B, Xue JY, Hendry TA, Li RQ, Brown JW, Liu Y, Hudson GT, Chen ZD (2010) Angiosperm phylogeny inferred from sequences of four mitochondrial genes. *J. Syst. Evol.* **48**: 391–425. https://doi.org/10.1111/j.1759-6831.2010.00097.x

Schattner P, Brooks AN, Lowe TM (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* **33**: 686–689. https://doi.org/10.1093/nar/gki366

Wei S, Wang X, Bi C, Xu Y, Wu D, Ye N (2016) Assembly and analysis of the complete *Salix purpurea* L. (Salicaceae) mitochondrial genome sequence. *Springerplus* **5**: https://doi.org/10.1186/s40064-016-3521-6

Wolfe KH, Li W-H, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. U. S. A.* **84**: 9054–9058. https://doi.org/10.1073/pnas.84.24.9054

Xu W, Xing T, Zhao M, Yin X, Xia G, Wang M (2015) Synonymous codon usage bias in plant mitochondrial genes is associated with intron number and mirrors species evolution. *PLoS One* **10**: 1–21. https://doi.org/10.1371/journal.pone.0131508

Zhang T, Fang Y, Wang X, Deng X, Zhang X, Hu S, Yu J (2012) The complete chloroplast and mitochondrial genome sequences of boea hygrometrica: Insights into the evolution of plant organellar genomes. *PLoS One* **7**: https://doi.org/10.1371/journal.pone.0030531

Zuo G, Hao B (2015) CVTree3 Web Server for whole-genome-based and alignment-free prokaryotic phylogeny and taxonomy. *Genomics, Proteomics Bioinforma.* **13**: 321–331. https://doi.org/10.1016/j.gpb.2015.08.004