

RESEARCH ARTICLE

The mitochondrial genome of the terrestrial carnivorous plant *Utricularia reniformis* (Lentibulariaceae): Structure, comparative analysis and evolutionary landmarks

Saura R. Silva¹✉, Danilo O. Alvarenga²✉, Yani Aranguren³✉, Helen A. Penha², Camila C. Fernandes², Daniel G. Pinheiro², Marcos T. Oliveira², Todd P. Michael⁴✉, Vitor F. O. Miranda³*, Alessandro M. Varani²*



1 Departamento de Botânica, Instituto de Biociências, Universidade Estadual Paulista (UNESP), Botucatu, São Paulo, Brazil, **2** Departamento de Tecnologia, Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista (Unesp), Jaboticabal, São Paulo, Brazil, **3** Departamento de Biologia Aplicada à Agropecuária, Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista (Unesp), Jaboticabal, São Paulo, Brazil, **4** Computational Genomics, Ibis Bioscience, Carlsbad, CA, United States of America

✉ These authors contributed equally to this work.

✉a Current address: Universidad Simón Bolívar, Barranquilla, Colômbia

✉b Current address: J. Craig Venter Institute, La Jolla, CA, United States of America

* amvarani@fcav.unesp.br (AV); vmiranda@fcav.unesp.br (VM)

OPEN ACCESS

Citation: Silva SR, Alvarenga DO, Aranguren Y, Penha HA, Fernandes CC, Pinheiro DG, et al. (2017) The mitochondrial genome of the terrestrial carnivorous plant *Utricularia reniformis* (Lentibulariaceae): Structure, comparative analysis and evolutionary landmarks. PLoS ONE 12(7): e0180484. <https://doi.org/10.1371/journal.pone.0180484>

Editor: Zhong-Hua Chen, University of Western Sydney, AUSTRALIA

Received: April 10, 2017

Accepted: May 13, 2017

Published: July 19, 2017

Copyright: © 2017 Silva et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The annotated sequences and raw reads of the *Utricularia reniformis* mitochondrial genome have been deposited in the GenBank database under accession numbers [GenBank: KY774314, SRX2646180, SRX2646130 and SRX2646131] (BioProject PRJNA290588).

Funding: This study was funded by Fundação de Amparo do Estado de São Paulo, FAPESP (13/

Abstract

The carnivorous plants of the family Lentibulariaceae have attained recent attention not only because of their interesting lifestyle, but also because of their dynamic nuclear genome size. Lentibulariaceae genomes span an order of magnitude and include species with the smallest genomes in angiosperms, making them a powerful system to study the mechanisms of genome expansion and contraction. However, little is known about mitochondrial DNA (mtDNA) sequences of this family, and the evolutionary forces that shape this organellar genome. Here we report the sequencing and assembly of the complete mtDNA from the endemic terrestrial Brazilian species *Utricularia reniformis*. The 857,234bp master circle mitochondrial genome encodes 70 transcriptionally active genes (42 protein-coding, 25 tRNAs and 3 rRNAs), covering up to 7% of the mtDNA. A *ltrA*-like protein related to splicing and mobility and a LAGLIDADG homing endonuclease have been identified in intronic regions, suggesting particular mechanisms of genome maintenance. RNA-seq analysis identified properties with putative diverse and important roles in genome regulation and evolution: 1) 672kbp (78%) of the mtDNA is covered by full-length reads; 2) most of the 243kbp intergenic regions exhibit transcripts; and 3) at least 69 novel RNA editing sites in the protein-coding genes. Additional genomic features are hypothetical ORFs (48%), chloroplast insertions, including truncated plastid genes that have been lost from the chloroplast DNA (5%), repeats (5%), relics of transposable elements mostly related to LTR retrotransposons (5%), and truncated mitovirus sequences (0.4%). Phylogenetic analysis based on 32 different Lamiales mitochondrial genomes corroborate that Lentibulariaceae is a monophyletic group. In summary, the *U. reniformis* mtDNA represents the eighth largest plant mtDNA

25164-6). Ibis Biosciences, an Abbott Laboratories company, provided support to TM in the form of salary. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section.

Competing interests: TM is an employee of Ibis Biosciences, an Abbott Laboratories company. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

described to date, shedding light on the genomic trends and evolutionary characteristics and phylogenetic history of the family Lentibulariaceae.

Introduction

Carnivorous plants have highly specialized morphological and physiological features adapted to uptake nutrients from captured prey as an alternative source of nutrients, thus supplementing the deficiency that comes from oligotrophic soils [1,2]. These plants are mostly found in low vegetation, from sandy to granitic soils, in water bodies and even in small flooded films, which are harsh conditions for most plants, but well tolerated by carnivorous plants [3,4]. This wide range of habitats is accompanied by a number of life forms and nutrient uptake mechanisms associated with the prey trap itself and with the trap microbiome [5–10]. For the family Lentibulariaceae, their peculiar morphology encompasses structures that do not always follow the traditional morphological classification, with well-defined leaf, stem and root organs [11]. For example, species from the genera *Utricularia* and *Genlisea* absorb nutrients through their leaves, phylloclades and/or utricles (traps) and lack roots [3,12,13].

Although the genetic architecture of several carnivorous plants is yet to be elucidated, the aquatic bladderwort *Utricularia gibba* has recently been considered as an interesting model plant, since it represents a specialized group from the family Lentibulariaceae, with species that have the smallest nuclear genomes among angiosperms known to date at 101Mbp [14]. Interestingly, its organellar genomes maintain typical sizes and features, such as gene content, genomic recombination, insertion of foreign DNA, and RNA editing [15], which are shared with the chloroplast DNA (cpDNA) of the terrestrial *U. reniformis*, although loss and pseudogenization of the NAD(P)H-dehydrogenase genes have been observed in this case [16].

U. reniformis A.St.-Hil. is endemic to Brazil, growing as a terrestrial plant in wet grasslands, and in contrast to *U. gibba*, exhibits a larger and polyploid genome with high levels of heterozygosity [17]. However, mitochondrial DNA (mtDNA) sequence information is scarce for carnivorous plants, including *Utricularia*, most likely due to the complex features commonly found in this angiosperm mitochondrial genome. Recently, the partial mtDNA of *U. gibba* was deciphered by third-generation genome sequencing approaches, indicating a genome size of 283,823 bp [14], but no further analyses were performed. Nonetheless, it has been proposed that *Utricularia* has significantly higher nucleotide substitution rates in several mtDNA *loci* [18,19], and that this may be related to the increased respiration rates and, consequently, increased production of reactive oxygen species (ROS) which accompanies its carnivorous habit [5]. This phenomenon could have contributed to the rapid morphological evolution of the terrestrial, epiphytic, reophytic, and aquatic forms observed for this group, since the high abundance of ROS can lead to an accumulation of nucleotide substitutions in all genomic compartments (mitochondrial, chloroplast and nuclear) [19,20]. Therefore, besides the conserved processes of plant ATP production and synthesis of amino acids, vitamins, and lipids, mitochondrial function in the Lentibulariaceae species may also have significantly influenced the genome evolution, maybe contributing to the diverse bodyplan organizations and habitat adaptations.

Herein, we describe the sequencing and assembly of the first complete mtDNA sequence from the species *Utricularia reniformis*, using a combination of paired-end and mate-pair short read sequences. Annotation, comparative genomics and phylogenomics indicated that *U. reniformis* mtDNA retains common features often found in angiosperm mtDNA, providing

useful insights into the genomic trends and evolutionary characteristics and phylogenetic history of the family Lentibulariaceae.

Material and methods

Plant sampling

U. reniformis samples were collected in the fall of 2015 in the Serra do Mar Atlantic Forest reserve (Geographic Location: 23°31'315"S– 45°53'53"O, 781m a.s.l), located in the Municipality of Salesópolis, State of São Paulo, Brazil, and deposited in the Herbarium JABU of the São Paulo State University (voucher V.F.O de Miranda et al., 1725). No permission for collecting was necessary, as the sample was not collected in protected areas and *U. reniformis* is not a threatened species according to the global IUCN (The IUCN Red List of Threatened Species - <http://www.iucnredlist.org>) and the Brazilian List of Threatened Plant Species.

Mitochondrial sequencing and assembly

Total DNA was extracted following the QIAGEN DNeasy Plant Maxi Kit extraction protocol (QIAGEN). Whole-genome shotgun sequencing was performed using the Illumina MiSeq platform with a paired-ends (PE) library of 2x300bp and an average insert size of ~600 bp. Library construction followed the Illumina Nextera XT Preparation Guide (Illumina, USA). A total of 40M PE reads were generated. Furthermore, an additional set of 160M mate-paired (MP) reads (2x100 bp) with an average insert size of ~3,500 bp (fragment sizes varying from 1kbp to up to 9kbp) were generated using Illumina HiScanSQ platform. Low quality sequences (Phred scores < 24), contaminants, adapters, and sequences with less than 50bp were removed using *Platanus_trim* [21], leaving 36M (PE) and 150M (MP) high quality reads for the mtDNA assembly.

The assembly was conducted in seven steps, described below:

1. Trimmed PE and MP reads with full-length matches to the *U. reniformis* chloroplast genome [16] were discarded with bowtie2 v2.2.9 using *—very-sensitive* and *—end-to-end* parameters [22];
2. Filtered PE reads were assembled with CLC Genomics Workbench v9 (QIAGEN Aarhus, Denmark— <http://www.clcbio.com>), and the coverage of each assembled sequence was estimated for identification of abnormal coverage peaks, allowing the identification of potential mitochondrial-derived regions. It is expected that the coverage depth of a plant mitochondrial genome assembly is relatively constant across the genome, with peaks corresponding to plastid and duplicated regions;
3. Potential mitochondrial contigs were baited by mapping against plant mitochondrial genes commonly found in angiosperms [23] with BLAST v2.2.26 [24];
4. PE and MP reads generated from the (b) step were mapped back to the retrieved contigs from the (c) step with bowtie2. This resulted in a set of high-quality and filtered mitochondrial reads;
5. To avoid misassemblies and incorrect contig linking due to the presence of repeats or dynamic and multipartite structures commonly found in the angiosperm mitochondrial genomes [25], only the PE reads were assembled with SPAdes v3.9.0 [26] with default parameters, and the assembly graph was inspected with Bandage [27]. The MP reads were used in the next steps for resolution of repeats and master circle assembly;

6. Each assembled contig from the previous step was extended independently by iterative (mapping) assembly with MITObim v1.9 [28], allowing the identification of repeated sequences and possible connections between the contigs. During this process, the joining of contigs and scaffold construction were based on sequence similarity of terminal regions of each contig with a minimum overlap of 100 bp and >99% identity;
7. To validate the joining of contigs and for repeat resolution, the MP reads were mapped back to the extended sequences from the (f) step with bowtie2, and the assembly paths were inspected using a custom Perl program. To guarantee the correct assembly of each long repeated sequence (>300bp), three approaches were used: (1) depth-coverage analysis, in which a higher coverage is expected in the repeated regions than is observed in non-repeated regions; (2) at least 50 PE and 10 MP reads supporting the anchoring of each repeat to their respective genome location; (3) individual mapping and assembly of each repeated sequence to their respective anchoring borders, where the assembled repeats had to be concordant with at least two different assembly software (SPAdes and Platanus 1.2.4). This method ensured a higher confidence assembly of repeats longer than 300bp.

The master circle was manually constructed by analyzing the longest assembly path, composed of all the contigs including the repeats, and with support of MP read mapping across the entire sequence with the use of the CLC Genomics Workbench v9 (minimum of 10 different MP reads for each contig joining). The remaining gaps were closed with GapCloser v1.12 from SOAPdenovo2 package v2.04 [29]. The average coverage depth was estimated with bowtie2 with—*very-sensitive* and—*end-to-end* parameters and samtools *depth* [30].

Annotation and analysis of the mitochondrial genome

The mtDNA was annotated using MITOFY (Annotation of Plant Mitochondrial Genomes) [31] coupled with Prodigal v2.6.2 [32] using the standard genetic code, ARAGORN [33], and BLAST for additional gene location refinements. Corrections of start and stop codons, intron acceptor and donor sites, and annotation curation were performed with Artemis genome browser 16.0.0 [34]. For gene assignments, the Blast2GO tool [35] was used. Potential plastid-like sequences were identified with BLASTn and DOGMA (Dual Organellar GenoMe Annotator) [36]. Identification of potential mitovirus-derived sequences was carried out by tBLASTn searches against the available mitovirus RNA-directed RNA polymerase protein sequences from the Uniprot database [37]. Putative transposable elements were identified with RepeatMasker open-4.0.5 (<http://www.repeatmasker.org>), using the Viridiplantae dataset from the Repbase database version 20150807 [38]. Group I and II introns were detected with the RNA-weasel tool [39]. Potential truncated pseudogenes were defined by BLAST comparative analysis with the use of at least one of the following criteria:

1. presence of at least one stop codon in-frame within the predicted coding region;
2. absence of start and/or stop-codon;
3. frameshift;
4. lacking of at least 20% of the coding region when compared to the respective coding region of closely related species.

A circular gene map was drawn with OGDRAW (OrganellarGenome DRAW) [40]. Regions repeated within the mitochondrial genome, and with high similarity between *Utricularia gibba* draft mtDNA and the *U. reniformis* chloroplast genome were detected with

BLASTn with the following parameters: e-value cutoff of 1^{-10} and at least 90% sequence identity. Comparative circular maps were generated with ClicO FS [41] and Circus v0.64 [42].

The annotated sequences and raw reads of the *Utricularia reniformis* mitochondrial genome have been deposited in the GenBank database under accession numbers [GenBank: KY774314, SRX2646130 and SRX2646131] (BioProject PRJNA290588).

Phylogenetic analyses

The concatenated alignment of the *atp1*, *cox1*, *matR*, *nad5*, *rps3* genes from 32 different species from the Lamiales order (S1 Table) was performed using MAFFT v7.123b [43] with default parameters. For the probabilistic analysis, the best evolutionary models (best-fit) were tested using ModelTest 3.7 [44]. Thus, the best-fit DNA model was evaluated for the combined dataset with the corrected Akaike information criterion [45,46]. Maximum likelihood (ML) and Bayesian inferences were performed to estimate the phylogenetic hypothesis for the dataset. ML analyses were run with RAxML v8 [47]. For the ML analyses the GTR+GAMMA+I model was selected with ModelTest, and 10,000 bootstrap pseudoreplicates were applied. Bayesian inferences were performed with MrBayes software version 3.2.5 [48] with 5×10^5 generations sampled for each 100 generations, using the default parameters. For each analysis, two runs (nrns = 2) with four chains (nchains = 4) were performed beginning from random trees. Initial samples were discarded after reaching stationary (estimated at 25% of the trees). Cladograms were drawn with TreeGraph2 v2.11.1–654 beta [49].

RNA-seq and RNA-edit analyses

Three different organs from plants from the same natural population in field were frozen in liquid N_2 and used for RNA-seq analysis: fresh leaves, stolons and utricles. The tissues were pooled in three replicates and total RNA was extracted using the PureLink RNA Mini Kit (Thermo Fisher Scientific), according to the manufacturer protocol. DNase I (Thermo Fisher Scientific) was used to remove any genomic DNA contamination. The extracted RNA was evaluated using an Agilent 2100 Bioanalyzer (Agilent Technologies) and a Qubit 2.0 Fluorometer (Invitrogen). Only samples having an RNA integrity number (RIN) ≥ 7.0 were used for the sequencing. cDNA libraries were sequenced on the Ion Proton system (Life Technologies) generating 180M reads with an average read length of 200bp. Low quality sequences (Phred < 20), bacterial contaminants, adapters, and sequences with less than 20bp were removed using prinseq_lite v0.20.4 [50].

To distinguish potential nuclear/plastid-like transcripts from potential authentic mitochondrial transcripts, two different approaches were used. First, filtered RNA reads were mapped back to the *Utricularia reniformis* mtDNA with bowtie2, with the—*very-sensitive* and—*end-to-end* parameters, and only full-length matches were considered. Second, the selected reads from the first step were mapped to traditional mtDNA genes with CLC Genomics Workbench v9 using the following parameters: mismatch cost of 3, insertion cost of 3, deletion cost of 3, minimal alignment coverage of 90% (Length fraction) and similarity fraction of >98%. The RNA-seq read mapping and transcription abundance were evaluated by RPKM (Reads Per Kilobase Million) normalization, whereas only unique read mappings were considered. In addition, intronic regions of intron-containing genes were also considered for the identification of spliced exons.

RNA-editing analyses based on the transcriptome data were carried out according to a previously proposed methodology [16]. In addition, the PREP-Mt tool [51] was used with default parameters to predict additional RNA editing sites. The mt RNA-seq reads used in this study have been deposited in the GenBank database under accession number [GenBank: SRX2646180] (BioProject PRJNA290588).

Results

Assembly of the *U. reniformis* mitochondrial genome

A total of 1,787,363 and 178,224 high-quality PE and MP mitochondrial reads were filtered from the raw reads generated by the Illumina MiSeq and HiScanSQ platforms, respectively. Approximately 830kbp (excluding repeated regions) were assembled into 7 contigs, with N50 length of 230kbp. The assembly graph analysis supports a complex scenario where 13 nodes, 8 edges, and 5 connected components lead to 14 dead ends, representing 53.85% of the assembled genome (Table 1 and S1 Fig). To investigate the dead ends and to determine the master circle molecule, the assembled contigs were independently extended by iterative read mapping. This analysis, together with the PE and MP read mapping, provided several contig connections that allowed the construction of different, inter-connected scaffolds, suggesting that a diverse set of alternative structures may occur *in vivo*. For instance, distinct smaller circular, short linear and branched structures were detected depending on the path taken to complete assembly. Interestingly, two repeated regions, spanning ~25kbp (LIR; long inverted repeat) and 3.2kbp (SDR; small direct repeat), consistently appeared during the contig extension process. Individual assembly of each repeat to their flanking borders resulted in their anchoring in the mtDNA

Table 1. Assembly summary statistics and validation of *Utricularia reniformis* mtDNA master circle (MC) genome.

Number of mtDNA-related paired-end reads (2x300bp ~600bp)	1,787,363
Number of mtDNA-related mate-paired reads (2x100bp ~3kbp)	178,224
Total assembled size (excluding repeats > 500bp)(bp)	831,638bp
PE assembly statistics	
- Nodes	13
- Edges	8
- Dead ends	14 (53.85%)
- Connected components	5
- Contigs	7
- Longest contig size	335,336bp
- Shortest contig size	40,798bp
- Average contig size	118,634bp
- N50	230,405
- L50	2
Master circle size (bp)	857,234
Master circle size GC%	43.98
- Average coverage	
- Paired-end	956x (st.dev. 265)
- Mate-pairs	40.98x (st.dev. 20.9)
- Long Repeat (LR) coverage (25kb)	1,342 (st.dev. 200)
- Small Repeat (SR) coverage (3.2kb)	1,201 (st. dev. 240)
- Mapped reads in pairs	
-paired-end reads	1,762,585 (98.61%)
-mate-paired reads	177,316 (99.49%)
- uncalled MC bases (Ns)	9 (0.0010%)
RNAseq mapping	
- Total Number of Reads (single-end ~200bp)	1,213,898
- Mitochondrial genome covered (bp)	672,561 (78.45%)
- Average coverage	83.9x (st. dev. 566.4)

<https://doi.org/10.1371/journal.pone.0180484.t001>

sequence, supporting that these repeats are in fact present. In addition, the MP read mapping analysis supported the LIR assembly, with both flanking borders completely and concordantly anchored to the assembly (S2 Fig). As expected, these repeated regions showed a constant and higher than the estimated average coverage on non-repeated regions (Table 1). However, the coverage is not twice as high, which can be explained by our method to bait and assemble the mitochondrial genome, the sequencing technology bias, and, as expected for plant mtDNA, the presence of several and different alternative linear and circular structures, that do not exist in equal stoichiometric frequencies, leading to a biased coverage estimation analysis. These findings strongly suggest that, as observed for other plants [52,53], the *U. reniformis* mtDNA is composed of multipartite structures, with repeat-mediated recombination processes acting as key drivers of structural variation.

Manual examination of the MP read mapping identified the most parsimonious master circle (MC) structure containing all the mtDNA related sequences. This resulted in a MC of 857kbp, with a GC content of 43.98% and an average coverage of 956X (+/- 265), with few uncalled bases (0.0010%). In addition, a total of 1,762,585 (98.61%) and 177,316 (99.49%) PE and MP reads, respectively, were mapped in pairs at the expected distance and orientation across the entire MC genome (Table 1), and no abnormal variation of coverage were observed in the non-repeated regions, thus supporting the MC assembly. According to the GenBank Organelles Database (<https://www.ncbi.nlm.nih.gov/genome/organelle>), among the 256 plant mtDNA genomes completely sequenced to date, the *U. reniformis* mtDNA represents the eighth largest, and the biggest one in the Lamiales order. The largest mitochondrial genome belongs to *Silene noctiflora* and *S. conica* (Caryophyllales), exhibiting a multichromosomal mtDNA genome, with ~59 and ~128 chromosomes ranging from 6.7 to 11.3Mb, respectively [54], followed by *Corchorus capsularis* and *C. olitorius* (Malvales) with 1.9Mbp and 1.8Mbp, respectively, *Cucumis sativus* and *C. pepo* (Cucurbitales) with 1.6Mbp and 982kbp, respectively, and *Welwitschia mirabilis* (Welwitschiales, the tree tumbo gymnosperm) with 978kbp.

The *U. reniformis* mtDNA content and organization

Utricularia reniformis presents a typical plant mitochondrial genome (Fig 1 and Table 2). The mtDNA encodes 70 mitochondrial genes, including 42 protein-coding, 25 tRNAs and 3 rRNAs, and an additional truncated copy of *rrna5* (Table 3). Two identical copies of the genes *ccmC*, *rpl2* and *trnL-CAA* were identified in the LIR regions, and two copies of the *trnT-GGT* gene were found in the SDR regions (Fig 1). As observed in other angiosperm mtDNA [23], the *U. reniformis* mtDNA does not contain a complete tRNA set, indicating that some functional tRNAs are imported from the cytoplasm for proper intra-mitochondrial translation. In addition, other genes related to the translation process, such as the ribosomal proteins *rps1*, *rps7*, *rps8*, *rps11* and *rpl6*, are absent, whereas *rps2*, *rps19* and *rpl16* appear to be pseudo or truncated genes.

Nineteen group II introns were found across the mtDNA sequence (Table 2), with one of them being of particular interest because it contains an ORF encoding a reverse transcriptase domain-containing protein (IPR000477 and cd01651 RT_G2_intron) similar to that of the *ltrA* gene. *ltrA* is a multifunctional protein that promotes splicing and mobility that was originally identified in *Lactococcus lactis* [55]. The reverse transcriptase domain of the *ltrA* gene is 49%, 50% and 51% similar to the 18S ribosomal RNA intron1, *atpA* intron1, and *cob* intron3, respectively, of the white spruce (*Picea glauca*) mitochondrial genome, indicating their putative role for the splicing of these genes. We also found seven trans-spliced genes, including *nad1*. Interestingly, the maturase-coding *matR* gene is located between exon1 and exon4 of *nad1*, and this syntenic feature is conserved in the *U. gibba* mtDNA.

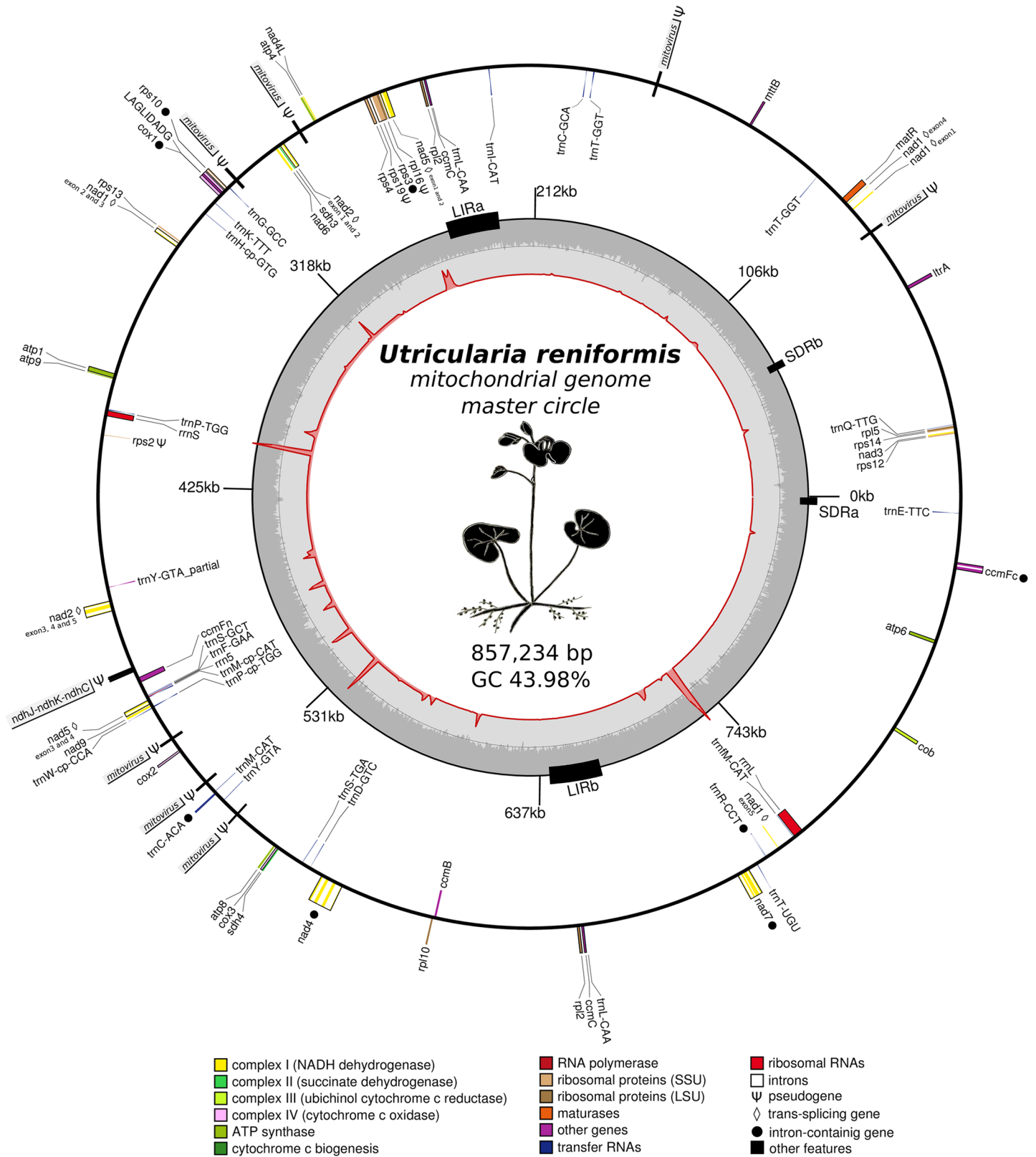


Fig 1. Genomic map of the *Utricularia reniformis* mtDNA genome. The inner red circle illustrates the transcript depth of coverage for *U. reniformis* mtDNA, whereas the peaks represent the most covered regions by RNAseq reads. The second level circle in gray scale represents the GC% distribution

across the mtDNA. The Large inverted repeats (LIR) and Small direct repeats (SDR) are shown as black boxes. The mitochondrial genes are shown in the outer circle, whereas genes shown on the outside of the map are transcribed clockwise, and the genes on the inside are transcribed counter-clockwise. Genes are color coded by their function in the legend, whereas partial mitovirus derived sequences are shown in gray boxes. The *ndhJ-ndhK-ndhC* loci which is deleted from the *U. reniformis* plastid genome is shown in the gray box.

<https://doi.org/10.1371/journal.pone.0180484.g001>

Group I introns have also been identified in *Utricularia reniformis* mtDNA genes. The software ARAGORN detected that the 193bp-long *trnR*-CCT gene contains a group I intron located in the anticodon loop, thus most likely not interrupting the overall tRNA structure. Although the tRNA splicing machinery and mechanism in plant cells are currently unclear [56], splicing would be indispensable for the maturation of this tRNA. Interestingly, the group I intron found in the *cox1* gene was identified encoding a LAGLIDADG endonuclease. It is noteworthy that the draft mtDNA of *U. gibba* Scaffold00369 (KC997779) presents an identical *cox1* organization containing exon1, LAGLIDADG, and exon2, with 97%, 97% and 100%

Table 2. Main characteristics and features of *Utricularia reniformis* mtDNA genome.

<i>U. reniformis</i> Mitochondrial Genomic Features	Number	Length (bp)	% of MT Genome
A—common mitochondrial genes	70	61,207	7,14
- mtDNA coding regions	42	54,079	6.31
- group I intron	1	892	0,1
- group II intron (cis/trans-spliced)	19	17,543	2,04
- pseudogenes	3	711	0,08
- mtDNA tRNAs	25	1,929	0,23
-group I intron	2	671	0,07
- mtDNA rRNAs	3	5,219	0,61
B—ORFs (from 89bp to 665bp; average 192bp)	2,149	414,236	48.32
- no transcription evidence supported by RNAseq	944	156,725	18.29
- with more than 150pb (average 206bp)	484	101,237	11.81
- transcription evidence supported by RNAseq (≥ 5 reads)	1,205	257,471	30.03
- with more than 150bp (average 237bp)	935	223,352	26.05
C—mitovirus derived (RNA-dependent RNA polymerase)	7	3,800	0,44
D—chloroplast-derived regions (from <i>U. reniformis</i> plastome)	262	44,431	5,18
- identified cpDNA genes and truncated pseudogenes	71	28,586	3.33
- intact cpDNA genes	2	531	0.62
- pseudogenes/truncated/partial	69	28,055	3.28
E—repeated regions (excluding A, B, C and D)	883	49,000	5,72
F—Transposable Elements (TEs) related regions § (excluding A, B, C, D and E)	125	41,165	4.8
- Retroelements (Class I)			
-LINEs	6	1,550	0.18
-LTR elements	119	39,615	4.62
-Ty1/Copia	60	20,958	2.44
-Gypsy/DIRS1	36	12,954	1.51
Summary			
- Annotated and characterized regions, and potential hypothetical/chimeric orfs	3,496	613,839	71.61
- excluding potential hypothetical/chimeric orfs	1,351	201,147	23.46
- including potential hypothetical/chimeric orfs with transcription evidence	2,552	457,074	53.32
- Rest of the genome (intergenic spacer regions and non-characterized regions)	-	243,395	28.39
- Mapped by RNAseq (unique mapping– 99% identity and 100% coverage)	-	178,814	21

§ Only fragments of TEs were identified

<https://doi.org/10.1371/journal.pone.0180484.t002>

Table 3. List of the traditional mitochondrial, chloroplast derived, and additional genes encoded by the *Utricularia reniformis* mtDNA genome.

Genes of Mitochondrial Origin	
Complex I (NADH dehydrogenase)	<i>nad1</i> •◇, <i>nad2</i> •◇, <i>nad3</i> , <i>nad4</i> •, <i>nad4L</i> , <i>nad5</i> •◇, <i>nad6</i> , <i>nad7</i> •, <i>nad9</i>
Complex II (succinate dehydrogenase)	<i>sdh3</i> , <i>sdh4</i>
Complex III (ubichinol cytochrome c reductase)	<i>cob</i>
Complex IV (cytochrome c oxydase)	<i>cox1</i> •, <i>cox2</i> , <i>cox3</i>
Complex V (ATP synthase)	<i>atp1</i> , <i>atp4</i> , <i>atp6</i> , <i>atp8</i> , <i>atp9</i>
Cytochrome c biogenesis	<i>ccmB</i> , <i>ccmC</i> (2x), <i>ccmF</i> •, <i>ccmFn</i>
Ribosomal proteins (SSU)	<i>rps2</i> Ψ, <i>rps3</i> •, <i>rps4</i> , <i>rps10</i> •, <i>rps12</i> , <i>rps13</i> , <i>rps14</i> , <i>rps19</i> Ψ
Ribosomal proteins (LSU)	<i>rpl2</i> (2x), <i>rpl5</i> , <i>rpl10</i> , <i>rpl16</i> Ψ
Maturases	<i>matR</i>
Transport membrane proteins	<i>mttB</i>
Transfer RNAs	<i>trnC</i> -GCA, <i>trnD</i> -GTC, <i>trnE</i> -TTC, <i>trnF</i> -GAA, <i>trnG</i> -GCC, <i>trnH</i> -cp-GTG, <i>trnI</i> -CAT, <i>trnK</i> -TTT, <i>trnL</i> -CAA (2x), <i>trnM</i> -CAT, <i>trnM</i> -cp-CAT, <i>trnP</i> -TGG, <i>trnP</i> -cp-TGG, <i>trnQ</i> -TTG, <i>trnR</i> -CCT•, <i>trnS</i> -GCT, <i>trnS</i> -TGA, <i>trnT</i> -GGT (2x), <i>trnT</i> -TGT, <i>trnV</i> -cp-TAC•, <i>trnW</i> -cp-CCA, <i>trnY</i> -GTA, <i>trnY</i> -GTA Ψ
Ribosomal RNAs	<i>rrn5</i> , <i>rrn5</i> Ψ, <i>rrnS</i> , <i>rrnL</i>
Others	LAGLIDADG endonuclease (intron region of <i>cox1</i> gene), Group II intron-encoded protein <i>ltrA</i>
Genes of Chloroplast Origin	
Photosystem I	<i>psaA</i> Ψ, <i>psaB</i> Ψ (2x), <i>psaC</i> Ψ
Protosystem II	<i>psbA</i> Ψ, <i>psbB</i> Ψ (2x), <i>psbC</i> Ψ (2x), <i>psbD</i> Ψ, <i>psbE</i> Ψ
Cytochrome b/f complex	<i>petD</i> Ψ, <i>petN</i> Ψ
ATP synthase	<i>atpA</i> Ψ, <i>atpB</i> Ψ (2x), <i>atpE</i> Ψ (2x), <i>atpF</i> Ψ
NADH dehydrogenase	<i>ndhA</i> Ψ (2x), <i>ndhB</i> Ψ, <i>ndhC</i> Ψ, <i>ndhD</i> Ψ, <i>ndhG</i> Ψ, <i>ndhH</i> Ψ (2x), <i>ndhJ</i> Ψ (2x), <i>ndhK</i> Ψ, <i>ndhN</i> Ψ
RubisCO large subunit	<i>rbcL</i> Ψ
RNA polymerase	<i>rpoA</i> Ψ, <i>rpoB</i> Ψ (4x), <i>rpoC1</i> Ψ, <i>rpoC2</i> Ψ (4x)
Ribosomal proteins (SSU)	<i>rps4</i> Ψ, <i>rps8</i> Ψ, <i>rps11</i> , <i>rps12</i> Ψ, <i>rps19</i> Ψ
Ribosomal proteins (LSU)	<i>rpl2</i> Ψ (2x), <i>rpl14</i> Ψ, <i>rpl23</i> Ψ, <i>rpl36</i>
Other genes	<i>accD</i> Ψ, <i>clpP</i> Ψ (2x), <i>infA</i> Ψ, <i>matK</i> Ψ (2x)
hypothetical chloroplast reading frames	<i>ycf2</i> Ψ (2x), <i>ycf15</i> Ψ, <i>orf56</i> Ψ
Ribosomal RNAs	<i>rrnA16</i> Ψ, <i>rrn23</i> Ψ (4x)
Transfer RNAs	<i>trnL</i> -cp-TAA Ψ (2x), <i>trnK</i> -cp-TTT Ψ, <i>trnS</i> -cp-GCT Ψ, <i>trnE</i> -cp-TTC Ψ
Additional Pseudogenes with Assigned Function	Mitovirus RNA-dependent RNA polymerase Ψ (7x), DNA polymerase type B, organellar and viral Ψ (2x) DNA-directed RNA polymerase subunit beta Ψ (2x), DNA-dependent RNA polymerase

Ψ pseudogene
 ◇ trans-splicing
 • intron-containing gene

<https://doi.org/10.1371/journal.pone.0180484.t003>

identical amino acid residues, respectively, when compared to the *U. reniformis* sequences, which supports intron acquisition from a common ancestor. Moreover, the *cox1* gene of both mtDNAs presented eleven, out of twelve, of the putative positively selected motifs (Phe/Lys-

164 motif is absent), in which accumulation of nucleotide substitutions, including the most important motif Cys-113-Cys-114, have been previously reported [18].

Utricularia reniformis mtDNA features

Repeats and similarities to *U. gibba* mtDNA. A total of 883 regions, ranging from 37bp to 25,125bp and corresponding to 49kbp, are repeated across *U. reniformis* mtDNA (Table 2 and Fig 2A). In addition to a large inverted repeat (LIR) and a small direct repeat (SDR) regions, at least seven additional repeated regions span more than 100 bp. These repeated regions could be involved with repeat-mediated homologous recombination that can generate sub-genomic circles or other alternative conformations. For instance, putative intramolecular recombination between the LIRa and LIRb and SDRa and SDRb repeats could generate at least four alternative MC conformations, including a small 70 kbp sub-circle, which may be involved in a putative direct-repeat-mediated deletion of a region containing the *rps12*, *rps14*, *rpl5*, *nad3* and *trnQ-TTG* genes (Fig 2B). According to a dot plot and AMOS software package [57] analysis, the *U. gibba* mtDNA sequence (downloaded from the CoGe OrganismView database; https://genomevolution.org/coge/GenomeInfo.pl?gid=29027-unitig_87) can be circularized into a single molecule of 270,037 bp. Although most of the *U. reniformis* traditional mitochondrial genes show high level of identity (~93.95%) with their homologues in *U. gibba*, these genomes exhibit highly repetitive content and essentially no conservation in synteny (Fig 2C), indicating that *U. gibba* has a significantly different mtDNA and that divergent evolutionary forces are acting in the intronic and non-traditional mitochondrial coding regions of both genomes.

Open reading frames. The largest portion of the *Utricularia reniformis* mtDNA (414kbp or 48%) is composed of 2,149 ORFs ranging from 89bp to 665bp (average 192bp) (Table 2). Blast searches identified that 1,898 of these have no similarity to any sequence from the Viridiplantae in the NCBI database, and 152 had hits to hypothetical proteins in other plant genomes, mostly exhibiting resemblance to partial sequences of organellar genes, such as the DNA and RNA polymerases, likely derived from mitochondrial plasmids [23], retrotransposons or nuclear genes (Table 4). Interestingly, a large number of unknown, hypothetical ORFs with putative function are transcribed, whereas some exhibit signal peptide and transmembrane domains (Table 4 and in more details in S2 Table), suggesting that these putative proteins may be exported to participate in cell signaling or inserted into the mitochondrial membranes. Therefore, several ORFs should produce peptides, whereas others are may be recombination remnants.

Sequences of plastid origin. Integrated plastid sequences generally correspond to from 1% to up to 12% of an angiosperm mtDNA, and are indeed widespread in seed plants [23,58]. Plastid-like sequences were located in intergenic regions corresponding to a least 44kbp (5%) of *Utricularia reniformis* mtDNA (Table 2 and Fig 2D), corroborating our previous observations of the occurrence of extensive lateral gene transfer between the organelles [16]. These insertions are spread in fragments of up to 1.5kbp and represent 31% of the *U. reniformis* plastid genome. In addition to numerous intergenic spacer regions, two intact plastid genes, *rps11* and *rpl36*, and 69 truncated pseudogenes and tRNAs were identified (Table 3). We have previously reported that the plastid NAD(P)H-dehydrogenase complex *ndhJ-ndhK-ndhC* gene locus, which is absent in the *U. reniformis* plastid genome, is present as truncated pseudogenes in the mtDNA [16]. Indeed, the *U. reniformis* mtDNA *ndhJ-ndhK-ndhC* locus shows ~87% of nucleotide similarity to the homologous region in *U. gibba* and *U. macrorhiza* cpDNA (Fig 3), although these genes in *U. reniformis* mtDNA should not be functional.

Transposable elements. The *U. reniformis* mtDNA contains 125 fragments of transposable element-related sequences from different families, accounting for up to 41kbp (4.8%) of

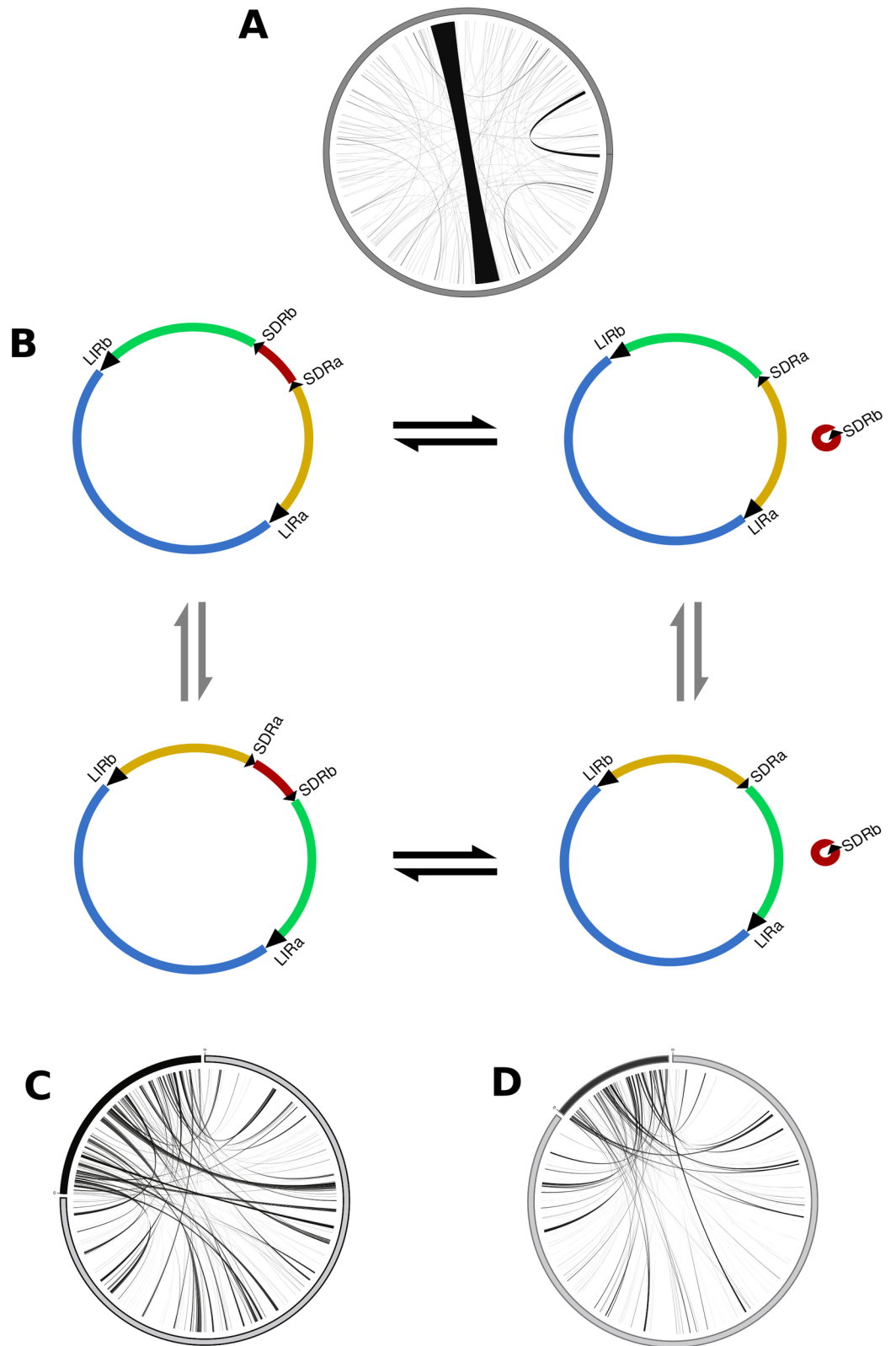


Fig 2. Repeats and alternative master circle structures. (A) Repeats across *U. reniformis* mtDNA; (B) Putative alternative conformations of *U. reniformis* mtDNA based on repeat-mediated intramolecular recombination mechanism; (C) Shared regions among *U. reniformis* mtDNA and *U. gibba* mtDNA; (D) repeats between *U. reniformis* cpDNA against the mtDNA.

<https://doi.org/10.1371/journal.pone.0180484.g002>

Table 4. Main characteristics of the unknown (no hits) and hypothetical open reading frames identified in *U. reniformis* mtDNA.

ORFs	Number	Signal Peptide + Transmembrane Domain	Only Signal Peptide	Only Transmembrane Domain
Unknown (no hits)	1,898	21	187	254
- transcribed	1,018	17	98	135
- non transcribed	880	4	89	119
Hypothetical	152	2	8	15
- transcribed	111	1	8	14
- non transcribed	41	1	0	1
Putative Function	99	0	4	9
- transcribed	76	0	3	6
- non transcribed	23	0	1	3

<https://doi.org/10.1371/journal.pone.0180484.t004>

the genome (Table 2), the majority of which being Ty1/Copia (21kbp) and Gypsy/DIRS1 (13kbp). Most are located in intergenic regions, and no complete elements were identified, indicating that these sequences represent relics of ancient events of lateral transfer from the nucleus. Using the same methodology to track TE fragments, the circular mtDNA of *U. gibba* exhibited a TE load of 10,452bp (3.9%), with Ty1/Copia (6.5kbp) and Gypsy/DIRs (3.1kbp)

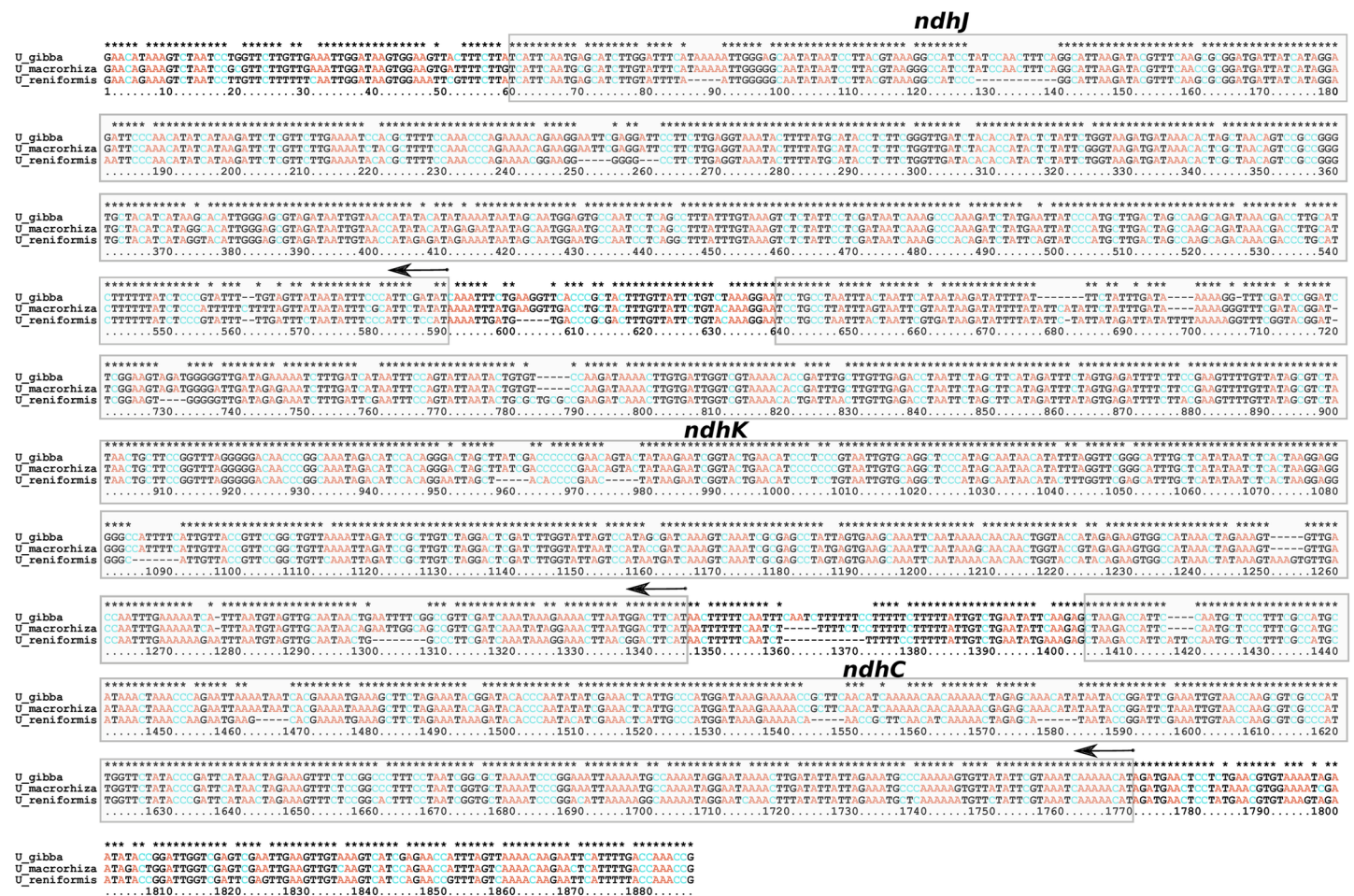


Fig 3. *U. reniformis* mtDNA *ndhJ-ndhK-ndhC* loci alignment against the respective loci in the cpDNA of *U. gibba* and *U. macrorhiza*.

<https://doi.org/10.1371/journal.pone.0180484.g003>

the major representatives. Indeed, the presence of relics of TEs is often observed in angiosperm and eudicot mtDNA [23], suggesting a putative role for these elements in shaping the mitochondrial genome structure and evolution.

Mitovirus derived sequences. At least 3.8kbp (0.4%) of the mtDNA correspond to up to 7 regions related to partial mitovirus sequences. Interestingly, more than 3 unique full-length RNAseq reads were mapped in each region (S3 Table), suggesting that these truncated mitovirus regions are still transcribed, but whether they encode regulatory or functional proteins is currently unclear. The mitovirus sequences belong to the Narnaviridae family and are the simplest, unencapsidated viruses, ranging from 2.3 to 3.6 kbp and encoding only a single RNA dependent RNA polymerase protein (RdRp) [59,60]. The Narnaviridae family is widespread among filamentous fungi, in particular phytopathogenic fungi [60]. Therefore, it is believed that the mitoviruses originated from horizontal gene transfer (HGT) from plant pathogenic fungi [59]. The RdRp regions identified in *U. reniformis* mtDNA share identity to the PFAM (<http://pfam.xfam.org>) family PF05919, present in several mtDNA of species from the Viridiplantae, including *Arabidopsis thaliana*, which contains complete mitovirus copies in both nuclear and mitochondrial genomes. To date, we are unable to determine whether complete mitovirus sequences are present in *U. reniformis* nuclear genome assembly. Interestingly, in *A. thaliana*, only the mitochondrial copy is expressed [59].

Mitochondrial-related genes are transcribed together with newly identified ORFs and inter-genic regions

A total of 1,2Mbp of RNA-seq reads were mapped to the *Utricularia reniformis* mtDNA, confirming that all traditional mitochondrial genes are transcriptionally active. Additionally, 78% (672kbp) of the mtDNA is covered by at least one RNA-seq read, indicating that the newly identified ORFs and portions of the intergenic spacer regions are also transcribed (Table 1). It is noteworthy that from 243kbp of the intergenic spacer regions, a total of 178kbp have at least one RNA-seq read mapped (Table 2), indicating unexpected transcription of a large portion of the mtDNA, or that genome-length poly-cystronic transcripts are produced, as is the case for animal mtDNA [61]. Fig 1 (inner red circle) shows the RNA-seq read mapping depth of *U. reniformis* mtDNA, indicating, as expected, that *rrnS* and *rrnL* are the sites for most transcriptional activity. However, an additional peak located near a mitovirus related region at genomic position 531kbp (Fig 1; eight forth five o'clock) is associated with a partial retrotransposon derived region. This finding suggests that relics of retrotransposons are still transcribed. However, it is unlikely that this fragment was actively transposed, since it carries only a partial sequence of a retrotransposon without terminal repeats. The transcription pattern of this region and all other TE-related should be interpreted with caution due to the difficulty in determining the source of this transcript (nuclear genomic retroelement or real organellar). The transcription pattern of all traditional mitochondrial genes is shown in Fig 4, showing that *atp1* and *atp9* are the most expressed genes, and that *ccmB*, *ccmFn* and *ltrA* are the least expressed ones. Even the truncated pseudogenes *rps19Ψ*, *rpl16Ψ* and *rps2Ψ* present unique mapped RNAseq reads, indicating that they are transcribed ($\geq 98\%$ identity considering 100% of the read length) (S3 Table). This appears to be a common feature of *U. reniformis* organelles, since the transcription evidence of truncated *ndh* pseudogenes were also observed in the *U. reniformis* cpDNA [16].

Mitochondrial RNA editing analysis revealed 69 novel sites

The PREP-Mt tool and the mapped RNA-seq data identified, respectively, at least 385 and 147 RNA editing sites, 69 of which corresponding to novel editing sites in traditional protein-

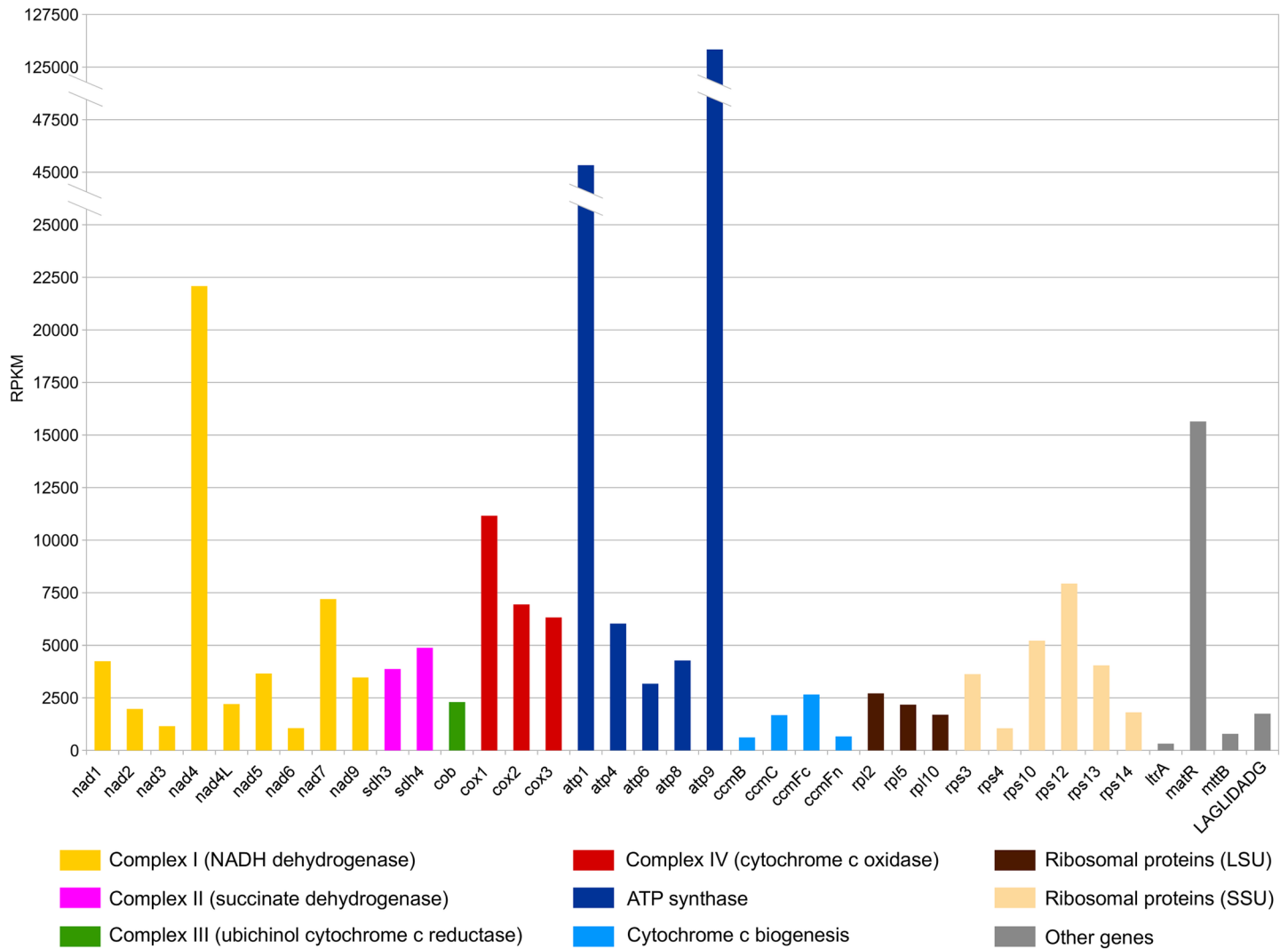


Fig 4. Transcription patterns of the traditional mitochondrial genes by Reads Per Kilobase Million (RPKM).

<https://doi.org/10.1371/journal.pone.0180484.g004>

coding genes of the *U. reniformis* mtDNA (Table 5). *nad1*, *nad4* and *nad7* were the most edited genes, indicating a common process for mitochondrial complex I genes. Interestingly, CGA (R) to UGA editing was detected in exon2 of the *rps3* gene in about 85% of the mapped RNA-seq reads. This change causes a premature stop codon, located 1,220 nucleotides upstream of the predicted stop codon, thus interrupting at least 73% of *rps3* exon2. This finding suggests a balanced production of a long and a short isoform for the *rps3*-encoded protein. Moreover, six editing sites were identified in the *cox1* gene, two of which cause non-synonymous substitutions at the positively selected amino acid motifs Ser/Phe-78 and Pro/Leu-194 [18] at the rate of 91% for S→F (position 321,375) and 13% for the P→L (position 321,723).

Phylogenetics analysis based on mtDNA supports Lentibulariaceae as a monophyletic group

The use of mitochondrial genes for phylogenetic purposes has been broadly discussed as these genes are more conserved than genes found in chloroplast genomes, and despite the fact that

Table 5. The 69 novel RNA editing sites identified in the protein-coding genes of *U. reniformis* mtDNA. For a complete list of the 147 editing sites identified by the RNAseq approach please refer to [S4 Table](#).

Gene	mtDNA		Codon		Codon position	Amino acid		Editing level (%)
	Position	Strand	From	To		From	To	
<i>nad1</i> (4 sites)	99048	+(exon 1)	CUA	UUA	1	L	L	18
	99201	+(exon 1)	CCG	UCG	1	P	S	88
	344481	+(exon 3)	CCC	CCU	3	P	P	33
	726995	-(exon 5)	AAC	AAU	3	F	N	63
<i>nad2</i> (5 sites)	297295	-(exon 2)	CCA	CUA	2	P	L	92
	463531	+(exon 4)	UUC	UUU	3	F	F	18
	463653	+(exon 4)	CCG	CUG	2	P	L	81
	465022	+(exon 5)	UCC	UUC	2	S	F	70
	465030	+(exon 5)	CUC	UUC	1	L	F	64
<i>nad3</i> (1 site)	20810	-	CCA	CUA	2	P	L	92
<i>nad4</i> (3 sites)	578975	+(exon 3)	CUA	UUA	1	L	L	36
	579324	+(exon 3)	CCA	CUA	2	P	L	94
	581798	+(exon 4)	ACC	ACU	3	T	T	93
<i>nad5</i> (5 sites)	260779	-(exon 2)	GCC	GCU	3	A	A	29
	260950	-(exon 2)	AUC	AUU	3	I	I	20
	261311	-(exon 2)	GCC	GUC	2	A	V	75
	261673	-(exon 2)	ACC	ACU	3	T	T	25
	261677	-(exon 2)	CCG	CUG	2	P	L	92
<i>nad6</i> (2 sites)	300266	-	CCA	CUA	2	P	L	93
	300273	-	CCC	UCC	1	P	S	90
<i>nad7</i> (13 sites)	711488	+(exon 1)	AAC	AAU	3	N	N	13
	712855	+(exon 2)	UCU	UUU	2	S	F	70
	713691	+(exon 3)	CGU	UGU	1	R	C	100
	713819	+(exon 3)	GUC	GUU	3	V	V	24
	713906	+(exon 3)	UCC	UCU	3	S	S	12
	714114	+(exon 3)	CUU	UUU	1	L	F	91
	715347	+(exon 4)	CCU	CUU	2	P	L	72
	715366	+(exon 4)	UCC	UCU	3	S	S	61
	715453	+(exon 4)	CCC	CCU	3	P	P	88
	715506	+(exon 4)	UCU	UUU	2	S	F	85
	715522	+(exon 4)	CAC	CAU	3	H	H	72
	715527	+(exon 4)	CCA	CUA	2	P	L	98
	715569	+(exon 4)	UCU	UUU	2	S	F	100
<i>nad9</i> (3 sites)	497201	+	UCU	UUU	2	S	F	73
	497282	+	CCA	CUA	2	P	L	47
	497708	+	UCU	UUU	2	S	F	78
<i>cob</i> (2 sites)	779528	+	UCU	UUU	2	S	F	92
	779794	+	CUA	UUA	1	L	L	86
<i>cox1</i> (2 sites)	321763	+(exon 1)	ACC	ACU	3	T	T	89
	323181	+(exon 2)	CUG	UUG	1	L	L	96
<i>cox2</i> (1 site)	514300	+	CUA	UUA	1	L	L	84
<i>cox3</i> (1 site)	558125	+	UUC	UUU	3	F	F	67

(Continued)

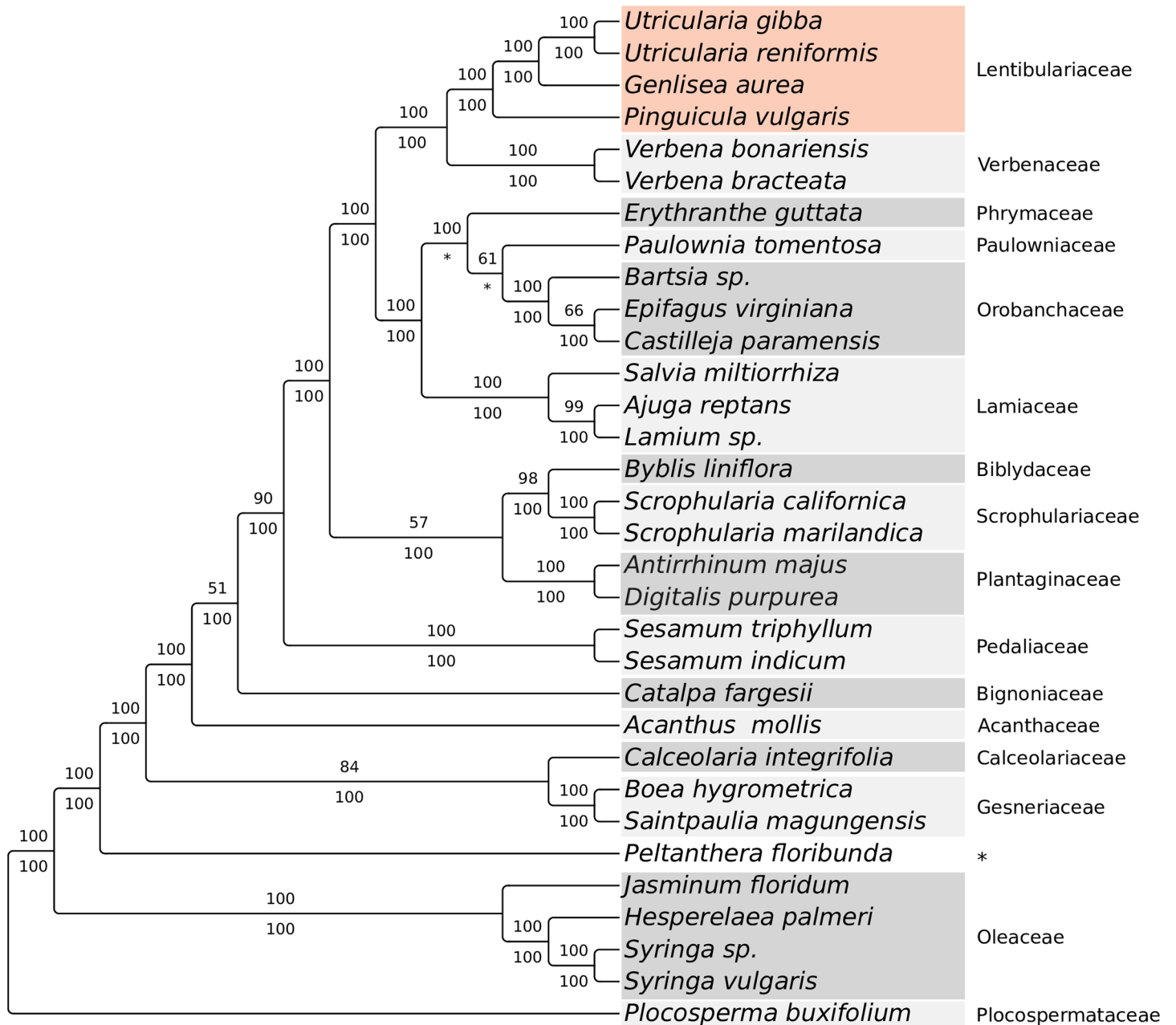
Table 5. (Continued)

Gene	mtDNA		Codon		Codon position	Amino acid		Editing level (%)
	Position	Strand	From	To		From	To	
<i>atp4</i> (2 sites)	285091	+	CCC	CUC	2	P	L	87
	285114	+	CCG	UCG	1	P	S	95
<i>atp6</i> (2 sites)	810361	-	UCC	UCU	3	S	S	48
	810993	-	CGU	UGU	1	R	C	64
<i>atp8</i> (1 site)	556640	+	UGC	UGU	3	C	C	84
<i>ccmFc</i> (3 sites)	834747	+(exon 1)	CAU	UAU	1	H	Y	63
	834901	+(exon 1)	CCA	CUA	2	P	L	77
	836320	+(exon 2)	CUA	UUA	1	L	L	32
<i>ccmFn</i> (1 site)	488236	-	ACC	AUC	2	T	I	61
<i>rpl5</i> (4 sites)	22103	-	CCG	UCG	1	P	S	16
	22105	-	CCU	CUU	2	P	L	77
	22173	-	AUC	AUU	3	I	I	13
	22389	-	UCC	UCU	3	S	S	31
<i>rpl10</i> (3 sites)	611859	+	ACC	ACU	3	T	T	35
	611934	+	CGG	UGG	1	R	W	41
	612084	+	UAC	UAU	3	Y	Y	40
<i>matR</i> (7 sites)	102277	+	UCC	UCU	3	S	S	96
	103757	+	CGC	UCG	1	R	S	16
	103771	+	UAC	UAU	3	Y	Y	75
<i>rps3</i> (4 sites)	263919	-(exon 2)	CCG	CUG	2	P	L	97
	263864	-(exon 2)	UCC	UCU	3	S	S	48
	264252	-(exon 2)	CCA	CUA	2	P	L	15
	264811	-(exon 2)	CGA	UGA	1	R	Stop	85
<i>rps4</i> (1 site)	267470	-	UCC	UCU	3	S	S	56
<i>rps12</i> (1 site)	20376	-	CCC	CCU	3	P	P	11
<i>rps14</i> (2 sites)	21929	-	UCC	UCU	3	S	S	21
	21974	-	CCC	CCU	3	P	P	17

<https://doi.org/10.1371/journal.pone.0180484.t005>

mitochondrial genome complexity could lead to biased tree reconstruction at interspecific level [62], many deeper clades in plant phylogeny are accepted because of mitochondrial gene information [63]. However, it is well known that the phylogeny of plant mitochondria may not accurately represent the organism's evolutionary history, mainly due to the frequent horizontal transfer of mitochondrial genes [64] (Fig 5).

As expected, our mitochondrial phylogenetic analysis shows that the family Lentibulariaceae is monophyletic and that *U. reniformis* is placed with *U. gibba* as sister group of *Genlisea aurea*, nested with *Pinguicula vulgaris* (Fig 5), as previous published phylogenies have shown [65,66]. Although the relationships among families of Lamiales are rather controversial (see [67]), the relations found in this study are consistent with the APG IV classification [66], except for the relations of Bignoniaceae, Pedaliaceae, Plantaginaceae, Scrophulariaceae and Biblydaceae. Missampling of taxa and the probable horizontal transfer of mitochondrial genes between divergent species could account for these discrepancies. Further studies are required to clarify the relations among these families.



* uncertain family

Fig 5. Phylogenetic analysis based on the mitochondrial genes *atp1*, *cox1*, *matR*, *nad5*, *rps3* from 32 different species from the Lamiales order.

<https://doi.org/10.1371/journal.pone.0180484.g005>

Discussion

The mtDNA master circle paradigm and multipartite structures in *U. reniformis*

Current methods and protocols to isolate and sequence plastid or mitochondrial DNAs are still laborious. To avoid these technical issues, whole-genome sequencing methods stand as one of the most reliable approaches to uncover plant organellar genomes. Using this approach,

the nuclear and organellar reads are mixed, making it mandatory to filter for authentic organellar reads prior to genome assembly. Although this would be expected to be accomplished fairly easily, since the organellar reads are overrepresented in deep whole-genome sequencing data [68], the occurrence of similar and horizontally transferred sequences among the organelles and the nuclear genome makes this procedure rather demanding from a bioinformatics point-of-view due to the occurrence of similar and horizontally transferred sequences between the organelles and the nuclear genome [28,69,70]. Even with good organellar read filters, genome assembly in the case of plant mtDNA is not straightforward, due to its complex physical organization. Therefore, the genomic MC is actually considered a representative format in which plant mitochondrial genomes are reported and used for comparative analysis and to reconstruct phylogeny [25].

The approach developed in this study successfully allowed the reconstruction of the *U. reniformis* mtDNA MC. However, it is well known that circular structures are difficult to observe or do not even exist *in vivo* [23,71,72]. Indeed, several complex alternative linear and circular genomic conformations, such as subgenomic circles, branched structures, head-to-tail concatamers and circularly permuted linear molecules are often found in attempts to observe the *in vivo* structure of plant mtDNA [23,25,72]. These alternative conformations are generally produced by intramolecular recombination driven by repeats. The presence of large repeats is commonly observed in angiosperm mtDNA, including related asterid species [23,53] and *U. reniformis*. Interestingly, these features are apparently absent in the *U. gibba* mtDNA, whereas all traditional mitochondrial genes are encoded in a circular molecule of 270kbp. Therefore, the results presented here suggest that repeat-mediated recombination processes may play a role in generating diverse *U. reniformis* mtDNA forms that alternate from the MC, although further investigation is necessary to show the presence, types and frequencies of these putative alternative genome conformations.

Possible TE expansion and whole genome duplication detected in *U. reniformis* nuclear genome

Angiosperm mtDNA normally range from 222kbp to 983kbp with GC content of 43–45% [23], and shares several features, such as size variation among different families, and even within species, presence of large amounts of unknown ORFs and non-coding DNA, repeated sequences, TEs, incorporation of cpDNA and nuclear sequences, and gain or loss of a number of chromosomes in mega-sized mitochondrial genomes [23,73]. Furthermore, trans-splicing, RNA editing, cytoplasmic male sterility (CMS) related genes, and partial copies of RdRp genes derived from uncapsidated mitoviruses are other common features often present in angiosperm mtDNA [23,59]. *Utricularia reniformis* mtDNA exhibits almost all of these characteristics, but when compared to *U. gibba*, genomic size stands out. Interestingly, *U. reniformis* and *U. gibba* mitochondrial genomes are 857 and 270kbp long, respectively, and both encode almost all traditional mitochondrial genes. An accumulation of TE fragments in both *Utricularia* mtDNA provides valuable insights into this size difference, suggesting a distinct evolutionary trajectory for both species. It is well established that retrotransposons and nuclear genes present in the mtDNA are derived from lateral transfer events between the cellular compartments [53,64], so the abundance of TEs in the *Utricularia* mtDNA may be a reflection of what is happening in the nucleus. In fact, the nuclear genome of *U. gibba* is 101Mbp in size, of which at least 8.9% corresponds to complete TEs [14], whereas *U. reniformis* exhibits a ~316Mbp-long nuclear genome. Previous studies support that *U. reniformis* may have a tetraploid nuclear genome with high levels of heterozygosity [17], corroborating our k-mer spectrum plot analysis that strongly suggests whole-genome duplication (WGD) and/or

polyploidization events in ~316Mbp-long and highly heterozygous genome. Therefore, we propose that TE propagation and extinction might be one of the possible mechanisms to explain genome expansion and contraction observed in both *Utricularia* species. Future work on *U. reniformis* nuclear genome and comparative analysis against *U. gibba* will definitely help us test this hypothesis.

The role of the *cox1* intron in *U. reniformis*

The *cox1* intron has been frequently acquired via horizontal transfer in angiosperms, whereas previous authors proposed that the *cox1* intron was originally acquired from fungi and laterally transferred several times during angiosperm evolution [74,75]. The intron-containing *cox1* gene is typically formed by two exons spanning 726bp and 858bp [70]; in *Utricularia reniformis*, exon1 is 721bp and exon2 is 864bp-long. The presence of an intact ORF for a LAGLIDADG endonuclease located in this intronic region, and its expression as detected by the RNA-seq data, are puzzling, since it was proposed that LAGLIDADG occurrence is associated with very low substitution rates in plant mtDNA [70]. However, previous studies indicate that *Utricularia* has significantly high rates of nucleotide substitutions in all three genome-bearing cellular compartments [5], possibly due to the increased respiratory rates associated with positive selection at the *cox1* locus [18,19]. We therefore speculate that the acquisition of a LAGLIDADG-coding intron in *Utricularia* mtDNA represents a recent event as previously proposed [70], with possible implications for the genome function and evolution, although the putative role of this endonuclease in keeping a low genomic substitution rate remains uncertain.

Several other group II introns were identified; one in particular encodes an *ltrA*-like protein, which may be related to splicing and mobility. The *ltrA* gene has not been previously described as a common feature of angiosperms mtDNA [23], even though relics of truncated copies of the *ltrA*-like gene are found in several other species, but not in the *U. gibba* mtDNA. These relics were not previously annotated in most of these mtDNAs, suggesting that *ltrA* is in fact an unnoticed common feature propagated during the course of the angiosperms evolution. Nonetheless, because *ltrA* is one of the least expressed genes, its putative mobility and splicing functions and role in the *U. reniformis* mtDNA evolution remain to be thoroughly investigated.

Large number of unknown ORFs and their putative roles in cytoplasmic male sterility

We identified at least 2,149 unknown ORFs, a large number of which exhibited transcripts and predicted proteins with signal peptides and/or transmembrane domains, suggesting important functional roles. It has been proposed that recombination events can give rise to novel ORFs that are often a combination of common mitochondrial genes and unknown ORFs, and that this phenomenon can be associated with the cytoplasmic male sterility (CMS), long reported for plant species [23]. The floral morphological aspects of *U. reniformis* suggest low reproductive success from cross-pollination in natural conditions [76], and a putative role for CMS is implied. Despite the nature and function of the CMS associated genes being poorly understood, it was previously cogitated that a malfunctioning ATPase or absence of *nad7* may be related [77]. However, the *nad7* gene is present in *U. reniformis* mtDNA, and truncated *atp*-related genes were not identified in association with unknown ORFs. We believe that further characterization of these ORFs and their putative products is warranted and will help establish the CMS genetic architecture in angiosperms.

Lateral gene transfer and potential functional replacement of plastid genes to the mtDNA

We also identified several pseudogenes of plastid origin in *Utricularia reniformis* mtDNA, a feature that is conserved in *U. gibba* [15]. Notably, the presence of the *ndhJ-ndhK-ndhC* locus, which is absent in the cpDNA, indicates a lateral transfer to the mtDNA, followed by a decay of the original copies in the cpDNA. Similar translocation of *ndh* genes from the cpDNA to the mtDNA was also observed in the Orchidaceae family, in particular in *Erycina pusilla* [78], suggesting that this type of event is more common than previously thought.

Moreover, the cpDNA *rps11* and *rpl36* genes appear complete, indicating that they may still be functional. However, the transcriptome profile of the plastid-like insertions was not inferred due to it being extremely difficult to determine the source of the transcripts (plastid, mitochondrial, or even nuclear). Previous studies indicated that the integrated plastid regions originate novel mitochondrial genes involved in maturation of mitochondrial mRNAs, and are therefore unrelated to their original plastidic functions [58]. The high number of truncated plastid genes may support this trend in *U. reniformis*, but their functional and evolutionary roles remain to be established.

Unexpected transcripts and RNA editing as important players in *U. reniformis* mtDNA evolution

The evidence of RNA transcripts for intergenic regions, newly identified ORFs, and partial fragments of retrotransposons in the mtDNA of *Utricularia reniformis* indicates an enrichment of unexpected genomic expression. Similar findings were also observed in other plant mtDNA, such as *Oryza sativa* and *Nicotiana tabacum* [79,80], supporting a common trend, and indicating that in-depth transcriptome analysis sheds additional light onto the mechanisms of mitochondrial genomic function and evolution. This analysis also revealed up to 532 RNA editing sites, a number consistent with the estimated number (roughly 500) for angiosperm organellar genomes [81]. Among these, 69 correspond to novel editing sites, including the site that leads to a premature stop codon in the *rps3* gene. RNA editing generating stop codons are also observed in the *atp9* and *rps10* genes of the *Rhazya stricta*, *Citrullus lanatus*, and *Cucurbita pepo* [31,53]. The fact that PREP-Mt detected more sites than the RNA-seq based approach may be related to the transcriptome conditions and tissues used, and to the high Phred quality values considered here to ensure data confidence. The RNA editing mechanism may be responsible for creating considerable polypeptide diversity in plant organelles, especially mitochondria, in which the genome shows such high structural plasticity [81]. We thus agree with previous work [82] that proposes the use of combined next-generation-sequencing approaches to unravel plant mitochondrial genomes and transcriptomes.

Conclusion

In this study, we characterized the first carnivorous plant and eighth largest mtDNA from the Brazilian endemic and terrestrial carnivorous *Utricularia reniformis*, providing several insights into the genomics trends and evolutionary characteristics and trajectory of the family Lentibulariaceae.

Supporting information

S1 Fig. Assembly graph of *Utricularia reniformis* mtDNA based on the paired-read (2x300bp) assembly, generated by the Bandage software. The assembled contigs (nodes, represented as colored bars) with multiple inputs and outputs, and dead ends; and the connections between those

contigs (edges, represented as black connectors) are shown.
(TIF)

S2 Fig. A paired-end (mate-pairs) mapping read-track displaying the reads and coverage generated by CLC Genomics Workbench v9.5.2 tool. Blue lines represent the paired reads located on the border of the repeat region; the yellow lines represent the paired reads located on each repeated region. Mismatches between the reads and reference are shown as narrow vertical traits. The read coverage are shown as peaks located in the bottom of each figure, whereas blue represent the repeat borders and yellow the repeat itself.

(TIF)

S1 Table. Mitochondrial genes (*atp1*, *cox1*, *matR*, *nad5*, *rps3*) from 32 different species from the Lamiales order used in the phylogenetic analysis.

(XLSX)

S2 Table. List of all identified open reading frames and detected signal peptide and trans-membrane domains.

(XLSX)

S3 Table. RNAseq analysis of all identified genes of *U. reniformis* mtDNA.

(XLSX)

S4 Table. The 147 RNA edit sites identified on the traditional mitochondrial coding-regions by RNAseq read mapping.

(XLSX)

Acknowledgments

We would like to thank Emily Mckinney for proofreading the manuscript.

Author Contributions

Conceptualization: Todd P. Michael, Vitor F. O. Miranda, Alessandro M. Varani.

Data curation: Saura R. Silva, Danillo O. Alvarenga, Alessandro M. Varani.

Formal analysis: Yani Aranguren, Helen A. Penha, Camila C. Fernandes, Daniel G. Pinheiro.

Funding acquisition: Alessandro M. Varani.

Investigation: Saura R. Silva, Danillo O. Alvarenga, Daniel G. Pinheiro, Marcos T. Oliveira, Todd P. Michael, Vitor F. O. Miranda, Alessandro M. Varani.

Methodology: Saura R. Silva, Danillo O. Alvarenga, Yani Aranguren, Helen A. Penha, Camila C. Fernandes, Todd P. Michael, Vitor F. O. Miranda, Alessandro M. Varani.

Project administration: Vitor F. O. Miranda, Alessandro M. Varani.

Resources: Yani Aranguren, Helen A. Penha, Camila C. Fernandes.

Software: Saura R. Silva, Danillo O. Alvarenga, Daniel G. Pinheiro, Todd P. Michael, Alessandro M. Varani.

Supervision: Marcos T. Oliveira, Todd P. Michael, Vitor F. O. Miranda, Alessandro M. Varani.

Validation: Marcos T. Oliveira, Todd P. Michael, Vitor F. O. Miranda, Alessandro M. Varani.

Visualization: Saura R. Silva, Danillo O. Alvarenga, Daniel G. Pinheiro, Alessandro M. Varani.

Writing – original draft: Saura R. Silva, Vitor F. O. Miranda, Alessandro M. Varani.

Writing – review & editing: Saura R. Silva, Danillo O. Alvarenga, Marcos T. Oliveira, Todd P. Michael, Vitor F. O. Miranda, Alessandro M. Varani.

References

1. Adamec L. Mineral nutrition of carnivorous plants: A review. *Bot Rev.* 1997; 63: 273–299. <https://doi.org/10.1007/BF02857953>
2. Ellison AM, Gotelli NJ. Energetics and the evolution of carnivorous plants—Darwin’s “most wonderful plants in the world.” *J Exp Bot.* 2009; 60: 19–42. <https://doi.org/10.1093/jxb/ern179> PMID: 19213724
3. Juniper B, Robins R, Joel D. *The Carnivorous Plants.* Academic Press, London; 1989.
4. Taylor P. *Genus Utricularia: a taxonomic monograph.* 2nd edition. Royal Botanic Gardens, Kew; 1989.
5. Albert VA, Jobson RW, Michael TP, Taylor DJ. The carnivorous bladderwort (*Utricularia*, *Lentibulariaceae*): a system inflates. *J Exp Bot.* 2010; 61: 5–9. <https://doi.org/10.1093/jxb/erp349> PMID: 20007200
6. Alcaraz LD, Martínez-Sánchez S, Torres I, Ibarra-Laclette E, Herrera-Estrella L. The Metagenome of *Utricularia gibba*’s Traps: Into the Microbial Input to a Carnivorous Plant. *PLOS ONE.* 2016; 11: e0148979. <https://doi.org/10.1371/journal.pone.0148979> PMID: 26859489
7. Caravieri FA, Ferreira AJ, Ferreira A, Clivati D, de Miranda VFO, Araújo WL. Bacterial community associated with traps of the carnivorous plants *Utricularia hydrocarpa* and *Genlisea filiformis*. *Aquat Bot.* 2014; 116: 8–12. <https://doi.org/10.1016/j.aquabot.2013.12.008>
8. Koopman MM, Carstens BC. The microbial Phyllogeography of the carnivorous plant *Sarracenia alata*. *Microb Ecol.* 2011; 61: 750–758. <https://doi.org/10.1007/s00248-011-9832-9> PMID: 21431933
9. Sirová D, Borovec J, Černá B, Rejmánková E, Adamec L, Vrba J. Microbial community development in the traps of aquatic *Utricularia* species. *Aquat Bot.* 2009; 90: 129–136. <https://doi.org/10.1016/j.aquabot.2008.07.007>
10. Takeuchi Y, Chaffron S, Salcher MM, Shimizu-Inatsugi R, Kobayashi MJ, Diway B, et al. Bacterial diversity and composition in the fluid of pitcher plants of the genus *Nepenthes*. *Syst Appl Microbiol.* 2015; 38: 330–339. <https://doi.org/10.1016/j.syapm.2015.05.006> PMID: 26138047
11. Rutishauser R. Evolution of unusual morphologies in *Lentibulariaceae* (bladderworts and allies) and *Podostemaceae* (river-weeds): a pictorial report at the interface of developmental biology and morphological diversification. *Ann Bot.* 2016; 117: 811–832. <https://doi.org/10.1093/aob/mcv172> PMID: 26589968
12. Adlassnig W, Peroutka M, Lambers H, Lichtscheidl IK. The Roots of Carnivorous Plants. *Plant Soil.* 2005; 274: 127–140. <https://doi.org/10.1007/s11104-004-2754-2>
13. Ellison AM, Adamec L. Ecophysiological traits of terrestrial and aquatic carnivorous plants: are the costs and benefits the same? *Oikos.* 2011; 120: 1721–1731. <https://doi.org/10.1111/j.1600-0706.2011.19604.x>
14. Lan T, Renner T, Ibarra-Laclette E, Farr KM, Chang T-H, Cervantes-Pérez SA, et al. Long-read sequencing uncovers the adaptive topography of a carnivorous plant genome. *Proc Natl Acad Sci.* 2017; 201702072. <https://doi.org/10.1073/pnas.1702072114> PMID: 28507139
15. Ibarra-Laclette E, Lyons E, Hernández-Guzmán G, Pérez-Torres CA, Carretero-Paulet L, Chang T-H, et al. Architecture and evolution of a minute plant genome. *Nature.* 2013; <https://doi.org/10.1038/nature12132> PMID: 23665961
16. Silva SR, Diaz YCA, Penha HA, Pinheiro DG, Fernandes CC, Miranda VFO, et al. The Chloroplast Genome of *Utricularia reniformis* Sheds Light on the Evolution of the *ndh* Gene Complex of Terrestrial Carnivorous Plants from the *Lentibulariaceae* Family. *PloS One.* 2016; 11: e0165176. <https://doi.org/10.1371/journal.pone.0165176> PMID: 27764252
17. Clivati D, Gitzendanner MA, Hilsdorf AWS, Araújo WL, Oliveira de Miranda VF. Microsatellite markers developed for *Utricularia reniformis* (*Lentibulariaceae*). *Am J Bot.* 2012; 99: e375–378. <https://doi.org/10.3732/ajb.1200080> PMID: 22947485
18. Jobson RW, Nielsen R, Laakkonen L, Wikström M, Albert VA. Adaptive evolution of cytochrome c oxidase: Infrastructure for a carnivorous plant radiation. *Proc Natl Acad Sci.* 2004; 101: 18064–18068. <https://doi.org/10.1073/pnas.0408092101> PMID: 15596720
19. Jobson RW, Albert VA. Molecular Rates Parallel Diversification Contrasts between Carnivorous Plant Sister Lineages. *Cladistics.* 2002; 18: 127–136. <https://doi.org/10.1006/clad.2001.0187>

20. Ibarra-Laclette E, Albert VA, Pérez-Torres CA, Zamudio-Hernández F, Ortega-Estrada M de J, Herrera-Estrella A, et al. Transcriptomics and molecular evolutionary rate analysis of the bladderwort (*Utricularia*), a carnivorous plant with a minimal genome. *BMC Plant Biol.* 2011; 11: 101. <https://doi.org/10.1186/1471-2229-11-101> PMID: 21639913
21. Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, et al. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* 2014; 24: 1384–1395. <https://doi.org/10.1101/gr.170720.113> PMID: 24755901
22. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012; 9: 357–359. <https://doi.org/10.1038/nmeth.1923> PMID: 22388286
23. Mower JP, Sloan DB, Alverson AJ. Plant Mitochondrial Genome Diversity: The Genomics Revolution. In: Wendel JF, Greilhuber J, Dolezel J, Leitch IJ, editors. *Plant Genome Diversity Volume 1*. Springer Vienna; 2012. pp. 123–144. Available: http://link.springer.com/chapter/10.1007/978-3-7091-1130-7_9
24. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009; 10: 421. <https://doi.org/10.1186/1471-2105-10-421> PMID: 20003500
25. Gualberto JM, Milesina D, Wallet C, Niazi AK, Weber-Lotfi F, Dietrich A. The plant mitochondrial genome: dynamics and maintenance. *Biochimie.* 2014; 100: 107–120. <https://doi.org/10.1016/j.biochi.2013.09.016> PMID: 24075874
26. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol J Comput Mol Cell Biol.* 2012; 19: 455–477. <https://doi.org/10.1089/cmb.2012.0021> PMID: 22506599
27. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics.* 2015; 31: 3350–3352. <https://doi.org/10.1093/bioinformatics/btv383> PMID: 26099265
28. Hahn C, Bachmann L, Chevreux B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Res.* 2013; 41: e129. <https://doi.org/10.1093/nar/gkt371> PMID: 23661685
29. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience.* 2012; 1: 18. <https://doi.org/10.1186/2047-217X-1-18> PMID: 23587118
30. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinforma Oxf Engl.* 2009; 25: 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352> PMID: 19505943
31. Alverson AJ, Wei X, Rice DW, Stern DB, Barry K, Palmer JD. Insights into the Evolution of Mitochondrial Genome Size from Complete Sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Mol Biol Evol.* 2010; 27: 1436–1448. <https://doi.org/10.1093/molbev/msq029> PMID: 20118192
32. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics.* 2010; 11: 119. <https://doi.org/10.1186/1471-2105-11-119> PMID: 20211023
33. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 2004; 32: 11–16. <https://doi.org/10.1093/nar/gkh152> PMID: 14704338
34. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinforma Oxf Engl.* 2012; 28: 464–469. <https://doi.org/10.1093/bioinformatics/btr703> PMID: 22199388
35. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics.* 2005; 21: 3674–3676. <https://doi.org/10.1093/bioinformatics/bti610> PMID: 16081474
36. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. *Bioinforma Oxf Engl.* 2004; 20: 3252–3255. <https://doi.org/10.1093/bioinformatics/bth352> PMID: 15180927
37. Magrane M, Consortium U. UniProt Knowledgebase: a hub of integrated protein data. *Database.* 2011; 2011: bar009–bar009. <https://doi.org/10.1093/database/bar009> PMID: 21447597
38. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA.* 2015; 6: 11. <https://doi.org/10.1186/s13100-015-0041-9> PMID: 26045719
39. Lang BF, Laforest M-J, Burger G. Mitochondrial introns: a critical view. *Trends Genet.* 2007; 23: 119–125. <https://doi.org/10.1016/j.tig.2007.01.006> PMID: 17280737
40. 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: W575–581. <https://doi.org/10.1093/nar/gkt289> PMID: 23609545

41. Cheong W-H, Tan Y-C, Yap S-J, Ng K-P. ClcO FS: an interactive web-based service of Circos. *Bioinforma Oxf Engl*. 2015; 31: 3685–3687. <https://doi.org/10.1093/bioinformatics/btv433> PMID: 26227146
42. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome Res*. 2009; 19: 1639–1645. <https://doi.org/10.1101/gr.092759.109> PMID: 19541911
43. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013; 30: 772–780. <https://doi.org/10.1093/molbev/mst010> PMID: 23329690
44. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. *Bioinforma Oxf Engl*. 1998; 14: 817–818.
45. Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control*. 1974; 19: 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
46. Burnham KP, Anderson DR. Multimodel Inference Understanding AIC and BIC in Model Selection. *Sociol Methods Res*. 2004; 33: 261–304. <https://doi.org/10.1177/0049124104268644>
47. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinforma Oxf Engl*. 2014; 30: 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033> PMID: 24451623
48. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinforma Oxf Engl*. 2001; 17: 754–755.
49. Stöver BC, Müller KF. TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics*. 2010; 11: 7. <https://doi.org/10.1186/1471-2105-11-7> PMID: 20051126
50. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinforma Oxf Engl*. 2011; 27: 863–864. <https://doi.org/10.1093/bioinformatics/btr026> PMID: 21278185
51. Mower JP. PREP-Mt: predictive RNA editor for plant mitochondrial genes. *BMC Bioinformatics*. 2005; 6: 96. <https://doi.org/10.1186/1471-2105-6-96> PMID: 15826309
52. Iorizzo M, Senalik D, Szklarczyk M, Grzebelus D, Spooner D, Simon P. De novo assembly of the carrot mitochondrial genome using next generation sequencing of whole genomic DNA provides first evidence of DNA transfer into an angiosperm plastid genome. *BMC Plant Biol*. 2012; 12: 61. <https://doi.org/10.1186/1471-2229-12-61> PMID: 22548759
53. Park S, Ruhlman TA, Sabir JSM, Mutwakil MHZ, Baeshen MN, Sabir MJ, et al. Complete sequences of organelle genomes from the medicinal plant *Rhazya stricta* (Apocynaceae) and contrasting patterns of mitochondrial genome evolution across asterids. *BMC Genomics*. 2014; 15: 405. <https://doi.org/10.1186/1471-2164-15-405> PMID: 24884625
54. Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, et al. Rapid Evolution of Enormous, Multichromosomal Genomes in Flowering Plant Mitochondria with Exceptionally High Mutation Rates. *PLOS Biol*. 2012; 10: e1001241. <https://doi.org/10.1371/journal.pbio.1001241> PMID: 22272183
55. Mills DA, McKay LL, Dunny GM. Splicing of a group II intron involved in the conjugative transfer of pRS01 in lactococci. *J Bacteriol*. 1996; 178: 3531–3538. PMID: 8655550
56. Yoshihisa T. Handling tRNA introns, archaeal way and eukaryotic way. *Front Genet*. 2014; 5. <https://doi.org/10.3389/fgene.2014.00213> PMID: 25071838
57. Treangen TJ, Sommer DD, Angly FE, Koren S, Pop M. Next Generation Sequence Assembly with AMOS. *Curr Protoc Bioinforma Ed Board Andreas Baxeavanis AI*. 2011; CHAPTER: Unit11.8. <https://doi.org/10.1002/0471250953.bi1108s33> PMID: 21400694
58. Wang D, Rousseau-Gueutin M, Timmis JN. Plastid Sequences Contribute to Some Plant Mitochondrial Genes. *Mol Biol Evol*. 2012; 29: 1707–1711. <https://doi.org/10.1093/molbev/mss016> PMID: 22319165
59. Bruenn JA, Warner BE, Yerramsetty P. Widespread mitovirus sequences in plant genomes. *PeerJ*. 2015; 3: e876. <https://doi.org/10.7717/peerj.876> PMID: 25870770
60. Hillman BI, Cai G. Chapter Six—The Family *Narnaviridae*: Simplest of RNA Viruses. In: Ghabrial SA, editor. *Advances in Virus Research*. Academic Press; 2013. pp. 149–176. Available: <http://www.sciencedirect.com/science/article/pii/B9780123943156000064> <https://doi.org/10.1016/B978-0-12-394315-6.00006-4>
61. Falkenberg M, Larsson N-G, Gustafsson CM. DNA replication and transcription in mammalian mitochondria. *Annu Rev Biochem*. 2007; 76: 679–699. <https://doi.org/10.1146/annurev.biochem.76.060305.152028> PMID: 17408359
62. Palmer JD. Mitochondrial DNA in Plant Systematics: Applications and Limitations. In: Soltis PS, Soltis DE, Doyle JJ, editors. *Molecular Systematics of Plants*. Springer US; 1992. pp. 36–49. Available: http://link.springer.com/chapter/10.1007/978-1-4615-3276-7_3

63. Knoop V. The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. *Curr Genet.* 2004; 46: 123–139. <https://doi.org/10.1007/s00294-004-0522-8> PMID: 15300404
64. Xi Z, Wang Y, Bradley RK, Sugumaran M, Marx CJ, Rest JS, et al. Massive Mitochondrial Gene Transfer in a Parasitic Flowering Plant Clade. *PLOS Genet.* 2013; 9: e1003265. <https://doi.org/10.1371/journal.pgen.1003265> PMID: 23459037
65. The Angiosperm Phylogeny Group. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot J Linn Soc.* 2009; 161: 105–121. <https://doi.org/10.1111/j.1095-8339.2009.00996.x>
66. The Angiosperm Phylogeny Group. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot J Linn Soc.* 2016; 181: 1–20. <https://doi.org/10.1111/boj.12385>
67. Christenhusz MJM, Vorontsova MS, Fay MF, Chase MW. Results from an online survey of family delimitation in angiosperms and ferns: recommendations to the Angiosperm Phylogeny Group for thorny problems in plant classification. *Bot J Linn Soc.* 2015; 178: 501–528. <https://doi.org/10.1111/boj.12285>
68. Jackman SD, Warren RL, Gibb EA, Vandervalk BP, Mohamadi H, Chu J, et al. Organellar Genomes of White Spruce (*Picea glauca*): Assembly and Annotation. *Genome Biol Evol.* 2015; evv244. <https://doi.org/10.1093/gbe/evv244> PMID: 26645680
69. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 2016; gkw955. <https://doi.org/10.1093/nar/gkw955> PMID: 28204566
70. Sanchez-Puerta MV, Cho Y, Mower JP, Alverson AJ, Palmer JD. Frequent, Phylogenetically Local Horizontal Transfer of the *cox1* Group I Intron in Flowering Plant Mitochondria. *Mol Biol Evol.* 2008; 25: 1762–1777. <https://doi.org/10.1093/molbev/msn129> PMID: 18524785
71. Oldenburg DJ, Bendich AJ. Size and Structure of Replicating Mitochondrial DNA in Cultured Tobacco Cells. *Plant Cell.* 1996; 8: 447–461. <https://doi.org/10.1105/tpc.8.3.447> PMID: 12239390
72. Sloan DB. One ring to rule them all? Genome sequencing provides new insights into the “master circle” model of plant mitochondrial DNA structure. *New Phytol.* 2013; 200: 978–985. <https://doi.org/10.1111/nph.12395> PMID: 24712049
73. Wu Z, Cuthbert JM, Taylor DR, Sloan DB. The massive mitochondrial genome of the angiosperm *Silene noctiflora* is evolving by gain or loss of entire chromosomes. *Proc Natl Acad Sci.* 2015; 112: 10185–10191. <https://doi.org/10.1073/pnas.1421397112> PMID: 25944937
74. Sanchez-Puerta MV, Abbona CC, Zhuo S, Tepe EJ, Bohs L, Olmstead RG, et al. Multiple recent horizontal transfers of the *cox1* intron in Solanaceae and extended co-conversion of flanking exons. *BMC Evol Biol.* 2011; 11: 277. <https://doi.org/10.1186/1471-2148-11-277> PMID: 21943226
75. Vaughn JC, Mason MT, Sper-Whitis GL, Kuhlman P, Palmer JD. Fungal origin by horizontal transfer of a plant mitochondrial group I intron in the chimeric *Cox1* gene of *Peperomia*. *J Mol Evol.* 1995; 41: 563–572. PMID: 7490770
76. Clivati D, Cordeiro GD, Plachno BJ, de Miranda VFO. Reproductive biology and pollination of *Utricularia reniformis* A.St.-Hil. (Lentibulariaceae). *Plant Biol.* 2014; 16: 677–682. <https://doi.org/10.1111/plb.12091> PMID: 24834508
77. Carlsson J, Glimelius K. Cytoplasmic Male-Sterility and Nuclear Encoded Fertility Restoration. In: Kempken F, editor. *Plant Mitochondria*. Springer New York; 2011. pp. 469–491. Available: http://link.springer.com/chapter/10.1007/978-0-387-89781-3_18
78. Lin C-S, Chen JJW, Huang Y-T, Chan M-T, Daniell H, Chang W-J, et al. The location and translocation of *ndh* genes of chloroplast origin in the Orchidaceae family. *Sci Rep.* 2015; 5: 9040. <https://doi.org/10.1038/srep09040> PMID: 25761566
79. Fujii S, Toda T, Kikuchi S, Suzuki R, Yokoyama K, Tsuchida H, et al. Transcriptome map of plant mitochondria reveals islands of unexpected transcribed regions. *BMC Genomics.* 2011; 12: 279. <https://doi.org/10.1186/1471-2164-12-279> PMID: 21627843
80. Grimes BT, Sisay AK, Carroll HD, Cahoon AB. Deep sequencing of the tobacco mitochondrial transcriptome reveals expressed ORFs and numerous editing sites outside coding regions. *BMC Genomics.* 2014; 15: 31. <https://doi.org/10.1186/1471-2164-15-31> PMID: 24433288
81. Bruhs A, Kempken F. RNA Editing in Higher Plant Mitochondria. In: Kempken F, editor. *Plant Mitochondria*. Springer New York; 2011. pp. 157–175. Available: http://link.springer.com/chapter/10.1007/978-0-387-89781-3_7
82. Stone JD, Storchova H. The application of RNA-seq to the comprehensive analysis of plant mitochondrial transcriptomes. *Mol Genet Genomics MGG.* 2015; 290: 1–9. <https://doi.org/10.1007/s00438-014-0905-6> PMID: 25182379