UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Botucatu

UNESP - UNIVERSIDADE ESTADUAL PAULISTA

CÂMPUS DE BOTUCATU

INSTITUTO DE BIOCIÊNCIAS

GENÔMICA ORGANELAR E EVOLUÇÃO DE *GENLISEA* E *UTRICULARIA* (LENTIBULARIACEAE)

SAURA RODRIGUES DA SILVA

Tese apresentada ao Instituto de Biociências, Câmpus de Botucatu, UNESP, para obtenção do título de Doutor em Ciências Biológicas (Botânica)

BOTUCATU - SP
- 2018 -

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

UNESP - UNIVERSIDADE ESTADUAL PAULISTA

CÂMPUS DE BOTUCATU

INSTITUTO DE BIOCIÊNCIAS

GENÔMICA ORGANELAR E EVOLUÇÃO DE *GENLISEA* E *UTRICULARIA* (LENTIBULARIACEAE)

SAURA RODRIGUES DA SILVA

PROF. DR. VITOR FERNANDES OLIVEIRA DE MIRANDA
ORIENTADOR
PROF. DR. ALESSANDRO DE MELLO VARANI
Coorientador

Tese apresentada ao Instituto de Biociências, Câmpus de Botucatu, UNESP, para obtenção do título de Doutor em Ciências Biológicas (Botânica)

BOTUCATU - SP
- 2018 -

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉCNICA DE AQUISIÇÃO E TRATAMENTO DA INFORMAÇÃO
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CAMPUS DE BOTUCATU - UNESP

Silva, Saura Rodrigues.
Genômica organelar e evolução de *Genlisea* e *Utricularia* (Lentibulariaceae) / Saura Rodrigues da Silva. – 2018.

Tese (doutorado) – Universidade Estadual Paulista, Instituto de Biociências de Botucatu, 2018.
Orientador: Vitor Fernandes Oliveira de Miranda
Co-orientador: Alessandro de Mello Varani
Assunto CAPES:

1. Sistemática Vegetal

CDD 581.1

Palavras-chave: *Utricularia*; *Genlisea*; genômica de organelas; ndhs; Evolução de organelas.

Dedico esta tese a minha mãe que sempre foi a minha maior incentivadora, amiga e meu porto seguro em todos os momentos.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

AGRADECIMENTOS

Escrever uma tese é tarefa árdua, repleta de percalços e desafios. No entanto, tive o privilégio de compartilhar estes momentos com adorados amigos e familiares que tornaram o desenvolvimento deste trabalho pleno de realizações e alegrias.

Gostaria imensamente de agradecer ao meu orientador, Dr. Vitor Fernandes Oliveira de Miranda, que desde a minha graduação, há 9 anos, ainda na Universidade de Mogi das Cruzes, me abriu uma das mais importantes oportunidades de minha vida que foi justamente desenvolver o meu trabalho com botânica, pela qual, desde a iniciação científica, sou absolutamente apaixonada. Sinto-me privilegiada por ser sua orientanda e por ter recebido todo apoio científico, contando com a sua competência exemplar e muito mais ainda pela grande amizade, dedicação e companheirismo. Por acreditar em cada coisa maluca que me vinha à cabeça, me desafiar e me apoiar visando o meu aperfeiçoamento e incentivando-me a almejar objetivos cada vez maiores.

Ao meu coorientador Dr. Alessandro de Melo Varani e ao Dr. Daniel Guariz Pinheiro por me auxiliarem nas análises e pelas discussões extremamente frutíferas. Sou extremamente grata pela contribuição e fortalecimento das discussões científicas por meio das leituras e observações críticas a respeito dos manuscritos aqui apresentados.

Ao Dr. Todd Michael e ao Dr. Elliott Meer, tendo em vista que não seria possível realizar grande parte desse trabalho sem as suas inestimáveis contribuições científicas e técnicas.

À minha mãe, Márcia Rejane Rodrigues, como a pessoa mais importante da minha vida, que sempre me ensinou a ter paixão pelo conhecimento, a viver com garra, responsabilidade, independência, a nunca desistir dos objetivos, a lutar, e independentemente dos obstáculos e percalços, nunca perder a alegria e sempre celebrar as conquistas.

Ao meu namorado, André Assis de Melo Neto, que ao longo desses 5 anos, sempre com muito amor e carinho, esteve ao meu lado em todos os acontecimentos e me compreendeu, mesmo nos momentos mais difíceis dessa jornada, sempre me incentivando a continuar o meu aprimoramento. Me proporcionou equilíbrio com toda sua calma e inteligência e, mesmo quando estava triste ou cansada, trouxe-me alegria e cores em forma de minhas músicas preferidas ao tocá-las divinamente em seu piano. Minha maior felicidade foi tê-lo escolhido como companheiro de vida.

Aos meus familiares, que me apoiaram e foram sempre compreensivos, principalmente durante os momentos de ausência em que estive envolvida na execução desse projeto.

Aos meus colegas de Laboratório de Sistemática Vegetal, Cristine Gobbo Menezes, Dasmiliá Arozarena, Fernanda Gomes Rodrigues, Giovanni Astuti, Guilherme Camara Seber,

Néstor Marulanda, Yani Aranguren, Yoannis Domingues Rodrigues, e aos colegas do Laboratório de Bioinformática, Luciano Kishi, Luis Teheran, Maria Fernanda, Michelli G. Funnicelli, Rafael Correia, Wellington Omori, pelas frutíferas discussões e sugestões e companheirismo.

Aos colegas Nilber Silva e Bruno Garcia por ajudarem durante as coletas e sempre estarem dispostos a mandar mais amostras caso fossem necessárias.

Aos professores Dr. Lorenzo Peruzzi (Universidade de Pisa) e Dra. Ana Paula de Moraes (UFABC) que contribuíram grandemente nas análises relacionadas a citogenética.

Aos meus grandes amigos que sempre estiveram ao meu lado acompanhando as alegrias e os dramas diários desde o princípio dessa jornada, Talyta Schartmann Ribeiro de Souza, Elias Matarazzo, Helen Penha, Gabriela Fernandes, Caroline Bartoli e Beatriz Yonamine.

Aos colegas do Instituto de Biociências de Botucatu, Felipe Girotto, Camila Zanetti, Ricardo Tozin, Angélica Lino Rodrigues, Carol que, apesar dos poucos momentos de convívio, foram sempre solícitos diante de eventuais necessidades.

À coordenadora Dra. Carmen Boaro por sempre nos incentivar a melhorar nossa formação e a nos ensinar a não pensarmos somente em nós mesmos, mas trabalharmos como equipe juntos com o programa de Botânica e assim alcançarmos notas de excelência dentro do nosso programa de pós-graduação.

Aos funcionários da seção técnica de pós-graduação pelo apoio técnico e atenção.

Ao Instituto de Biociências da UNESP de Botucatu, à CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) e a FAPESP pelo apoio financeiro e, principalmente pelo interesse neste estudo.

A UNESP/FCAV de Jaboticabal pela infraestrutura.

A todos que de forma direta ou indiretamente contribuíram para a realização.

ÍNDICE

SILVA, S.R. **GENÔMICA ORGANELAR E EVOLUÇÃO DE GENLISEA E UTRICULARIA (LENTIBULARIACEAE)** 2018. 397P. TESE (DOUTORADO) – INSTITUTO DE BIOCIÊNCIAS, UNESP - UNIVERSIDADE ESTADUAL PAULISTA, BOTUCATU.

RESUMO - *Utricularia* e *Genlisea* são gêneros irmãos de plantas carnívoras da família Lentibulariaceae. Possuem aproximadamente 260 espécies representadas em diversas formas de vida. Para o Brasil foram catalogadas 82 espécies, das quais 27 são consideradas endêmicas. Além de dispor das armadilhas carnívoras mais complexas entre plantas, algumas de suas espécies apresentam os menores genomas e as maiores taxas de mutações entre as angiospermas relatadas até o momento. A respeito de seus genomas organelares, os estudos são pífios. Neste contexto, há a necessidade de se investigar como são os genomas organelares, suas estruturas, seus genes e como se deu a evolução das organelas nos gêneros. Portanto este estudo teve como objetivo, a partir de sequenciamento de nova geração e montagem de genomas, estudar e comparar os genomas organelares de *Utricularia* e *Genlisea*. Neste âmbito, foram montados e sequenciados os cloroplastos das espécies *Utricularia foliosa*, *U. reniformis*, *G. aurea*, *G. filiformis*, *G. pygmaea*, *G. repens* e *G. tuberosa*, e o genoma mitocondrial de *U. reniformis*. Os resultados obtidos revelaram que possivelmente há relação entre forma de vida e presença de genes *ndhs* nos gêneros, em razão de que para as espécies terrestres há deleção e "pseudogenização" de genes *ndhs*, já as espécies aquáticas detêm todo repertório de *ndhs* intacto. A partir das evidências encontradas, foi possível constatar transferência horizontal de genes, inclusive de genes *ndhs,* em mitocôndrias.

Palavras-chave: Lentibulariaceae, genômica organelar, filogenômica, *Utricularia*, *Genlisea*

ABSTRACT – *Utricularia* and *Genlisea* are sister genera in the carnivorous family Lentibulariaceae. There are aprproximately 260 species representing diverse life forms. For Brasil there are 82 species, 27 considered endemic. At the moment, besides having the most complex carnivorous traps between all plants, some of its species have miniature genomes and the highest mutational rates among angiosperms. There are few studies regarding its organellar genome. In this context, it is necessary to investigate how are these organellar genomes, its structure, genes, and how evolutionary forces govern these organelles in the different genera. Therefore, the aim of this study is to study and compare the organellar genomes of *Utricularia* and *Genlisea*, using next generation sequencing and genome assembly. In this context, chloroplasts of the species *Utricularia foliosa*, *U. reniformis*, *Genlisea aurea*, *G. filiformis*, *G. pygmaea*, *G. repens* and *G. tuberosa*, and the mitochondrial genome of *U. reniformis* were assembled and sequenced. The results show that possibly there is a connection between life form and the presence of *ndhs* genes in the genera, since for the terrestrial species there are *ndhs* genes that are deleted and pseudogenization, in contrary to the aquatic species which have all intact *ndhs* repertoir. Concerning the evidences, it was possible to verify horizontal transfer of *ndhs* and other genes as there are chloroplasts genes in the mitochondria.

Keywords: Lentibulariaceae, organellar genomics, phylogenomic, *Utricularia, Genlisea.*

INTRODUÇÃO GERAL

1. Plantas carnívoras

As plantas carnívoras são plantas ímpares, que por meio de modificações foliares, produção de substâncias diversas e complexos mecanismos de captura podem usufruir de artrópodes e outros pequenos animais para a manutenção do seu próprio desenvolvimento.

Para uma planta ser considerada carnívora é necessário que esta possua um conjunto de características designadas como "síndrome de carnivoria" (Givnish et al., 1984), que compreende plantas com a capacidade de atrair, capturar, digerir e absorver nutrientes que provêm de presas. Dentro desse contexto, há cerca de 600 espécies de plantas carnívoras, distribuídas em 20 gêneros, 12 famílias em 5 ordens distintas (Givnish, 2015). Dentre elas, são conhecidas como "snap traps" as espécies dos gêneros *Aldrovandra* e *Dionaea* (Figura 1A); "pitcher plants" as de *Brocchinia*, *Catopsis*, *Cephalotus*, *Darlingtonia*, *Heliamphora*, *Nepenthes* (Figura 1B) e *Sarracenia*; "papel pega-moscas" as de *Drosera* (Figura 1C), *Drosophyllum*, *Pinguicula*, *Roridula, Philcoxia, Byblis* e *Triphyophyllum* e as "corckscrew plants" e "bladderworts" que são as *Genlisea* e *Utricularia*, respectivamente (Givnish, 2015; McPherson, 2010).

Apesar de a carnivoria ser comum entre esses grupos, sabe-se pelas histórias filogenéticas que essa síndrome surgiu em, pelo menos, cinco ordens na história evolutiva das angiospermas (The Angiosperm Phylogeny Group, 2016). Em face disso, houve clara convergência adaptativa para a carnivoria.

O interesse por estas plantas peculiares se reflete por uma vasta literatura popular e principalmente científica, sobre morfologia e funcionamento das armadilhas (*e.g.* Alcalá, et al., 2010; Cameron et al., 2002; Reifenrath et al., 2006; Westermeier et al., 2017), produção de enzimas digestivas (*e.g.* Morohoshi et al., 2011; Płachno et al., 2006), atração de presas (*e.g.* Bennett & Ellison, 2009; Plachno et al., 2008; Płachno et al., 2007) e espectro de presas capturadas (*e.g.* Bauer et

al., 2009; Peroutka et al., 2008; Sanabria-Aranda et al., 2006), assim como aspectos da evolução da carnivoria em angiospermas (Givnish, 2015; Jobson et al., 2004; Müller et al., 2004) e sua ecologia e relação com ambientes pobres em nutrientes (Ellison & Adamec, 2011; Pavlovič & Saganová, 2015).

Figura 1. A. Armadilhas de *Dionaea* do tipo" snap-traps"; **B.** "pitcher plants" de *Nepenthes*; **C.** "Papel pega-moscas" de espécies do gênero *Drosera* (Fonte: https://www.mnn.com/your-home/organic-farming-gardening).

2. Família Lentibulariaceae

A carnivoria em Lamiales surge pela presença de tricomas glandulares, presentes em quase todos os membros da família, que teve função alterada da secreção de substâncias para também absorvê-las (Müller et al., 2004). Lentibulariaceae, que pertence a esta ordem (The Angiosperm Phylogeny Group, 2016), é a maior entre as famílias de plantas carnívoras, compreendendo cerca de 360 espécies em três gêneros: *Pinguicula* L., *Genlisea* A.St.-Hil. e *Utricularia* L. (Casper, 1966; Fromm-Trinta, 1981; Givnish, 2015; Taylor, 1989).

Estudos filogenéticos feitos a partir de caracteres morfológicos e moleculares indicam a monofilia da família Lentibulariaceae (Jobson & Albert, 2002; Jobson et al., 2003; Müller et al., 2004; Müller & Borsch, 2005; Silva et al., 2018) e posicionam o gênero *Pinguicula* como grupo irmão do clado *Genlisea-Utricularia*. Evidências apontam que os ancestrais de *Genlisea-Utricularia* detinham as folhas rosetadas de *Pinguicula* que sofreram o processo de epiascidiase (Juniper et al., 1989) dando origem às folhas em forma de "Y" invertido em *Genlisea* e às pequenas vesículas em *Utricularia*. Já o

ancestral de Lentibulariaceae possuía raízes primárias que, após o processo de germinação, provavelmente eram reduzidas. Enquanto que em *Pinguicula* as raízes foram mantidas, no ancestral de *Genlisea* e *Utricularia* esses órgãos foram totalmente perdidos e a suas funções foram atribuídas às armadilhas e ao estolão (Müller et al., 2004).

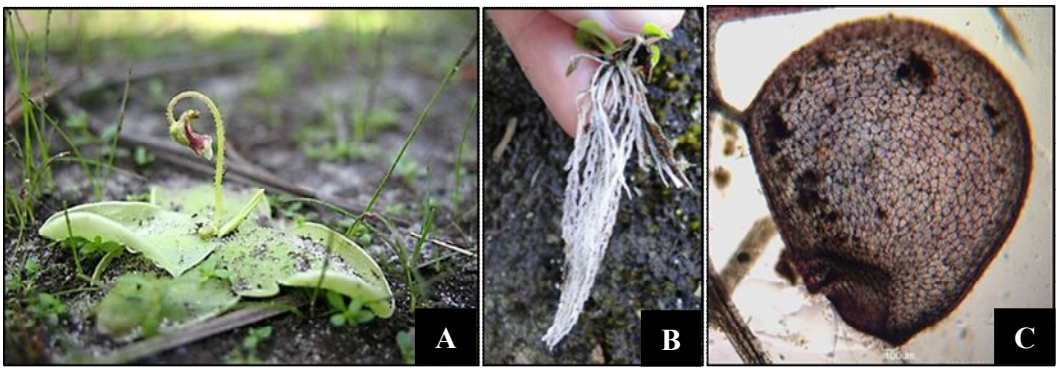


Figura 2. Tipos de armadilhas na família Lentibulariaceae **A.** *Pinguicula albida* C.Wright ex Griseb., de armadilhas pegajosas como "papel pega-moscas"; **B.** *Genlisea violacea* A.St.-Hil., com armadilhas na forma de "Y" invertido; **C.** *Utricularia foliosa* L., pequenas vesículas chamadas de "utrículos" (Fotos de nosso grupo – LSV/FCAV/UNESP.)

3. Os gêneros: *Pinguicula* L., *Genlisea* A.St.-Hil. e *Utricularia* L.

As espécies de *Pinguicula* L., também chamadas de "butterworts", são pequenas plantas herbáceas (de 1 cm a aproximadamente 30 cm de diâmetro) perenes ou raramente anuais de folhas em forma de rosetas que funcionam como "papel pega-moscas" similares aos dos gêneros de plantas carnívoras *Drosera* e *Byblis.* Entretanto, no que concerne a filogenia do gênero, outros aspectos sugerem *Pinguicula* como origem ancestral com clado irmão *Utricularia-Genlisea* se considerarmos a natureza fisiológica de seus tricomas glandulares, simetria zigomorfa das flores, anatomia do cálice, e filogenia baseada em diversos caracteres moleculares (Legendre, 2000).

Há aproximadamente 100 espécies de *Pinguicula* que podem variar na cor e forma das flores e folhas (Givnish, 2015; McPherson, 2010; Figura 3). É interessante ressaltar que, entre os gêneros da família a qual pertence, é o menos estudado e, assim, há diversas dúvidas sobre a circunscrição taxonômica de várias espécies. O gênero é informalmente dividido em três grupos de acordo com o

tipo de desenvolvimento foliar e *habitat*: heterófilas de ambientes temperados (~26 spp.), heterófilas de ambiente tropical (~41 spp.) e as homófilas (~34 spp.). Tem distribuição principalmente no Hemisfério Norte, contudo há espécies que colonizam alguns países da América Central, Caribe e México (Casper, 1966; Legendre, 2000; McPherson, 2010).

Figura 3. A. *Pinguicula filifolia* C.Wright ex Griseb. **B.** *Pinguicula cubensis* Urquiola & Casper **C.** *Pinguicula lignicola* Barnh. (Fotos de nosso grupo – LSV/FCAV/UNESP.)

O gênero *Genlisea* A.St.-Hil., das chamadas de "corckscrew plants", compreende cerca de 30 espécies distribuídas principalmente nas regiões tropicais e subtropicais da América do Sul, Central, regiões do continente africano e em Madagascar (Fleischmann, 2012). De acordo com classificação em subgênero, as espécies do subgênero *Genlisea* possuem fruto com deiscência circuncisa, enquanto que as do subgênero *Tayloria* (Fromm-Trinta, 1981) têm deiscência espiralada. A maioria das espécies se concentra principalmente no Brasil, principalmente do subgênero *Tayloria*, com cerca de 17 espécies, sendo que 10 são endêmicas (Miranda et al., 2018).

São ervas rosetadas, de flores zigomorfas, similares às de *Utricularia*, podendo ser facilmente confundidas (Figura 4A), exceto por seu cálice pentâmero. São terrestres, heterófilas, que têm lâminas foliares conspícuas, achatadas e fotossintetizantes de superfície do solo e folhas subterrâneas modificadas para a captura de presas, ancoragem e adesão ao subsolo (Figura 2B). Estas estruturas têm forma de "Y" invertido e geralmente se dispõem de maneira alternada às folhas fotossintetizantes (Figura 4B). Cada armadilha possui estruturas chamadas de "braços" que se organizam como fitas

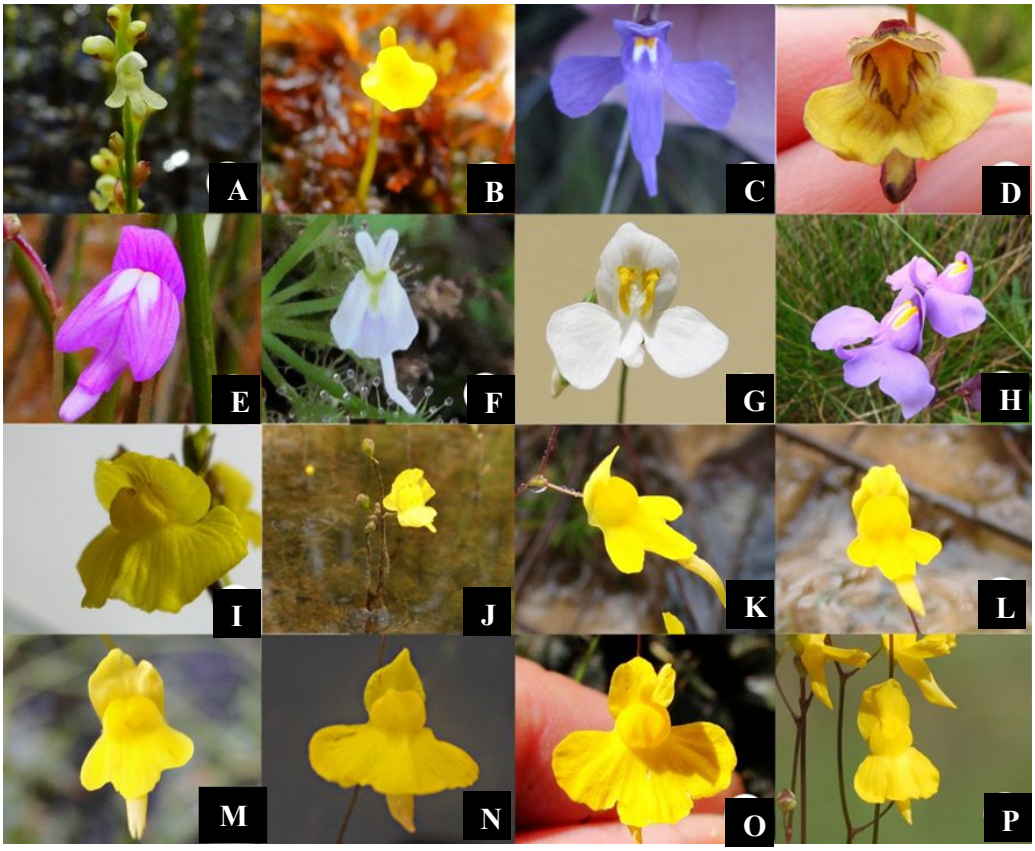
helicoidais torcidas, com fendas (Figura 4D e 4E). Dentro da armadilha, principalmente na região das fendas, há tricomas direcionados para o interior da estrutura e em direção a uma região intumescida chamada de "ampola" (Figura 4C) criando, assim, barreira física em que a presa consegue facilmente entrar na armadilha, mas não consegue sair. A ampola é uma cavidade para a digestão de presas, onde são secretadas enzimas para degradação e absorção de nutrientes provenientes das presas (Fleischmann, 2012).

Figura 4. A. Flor de *Genlisea violacea.* **B.** Exsicata de folhas de *Genlisea repens.* **C.** Região da "ampola" de *Genlisea*. **D.** Região dos braços helicoidais, a figura **E.** Mostra detalhe dos tricomas nas fendas da armadilha. (Fonte: Foto A e B do nosso grupo de pesquisa; C, D e E retiradas de Rutishauser, 2016.)

O gênero *Utricularia* L., ou "bladderworts", possui aproximadamente 250 espécies representadas em diversas formas de vida: terrícolas, aquáticas livres, aquáticas afixadas, reofíticas, epifíticas e litofíticas (Figura 5; Guisande *et al.*, 2004; Taylor, 1989). Para o Brasil foram catalogadas 65 espécies, das quais 16 são consideradas endêmicas (Miranda *et al.*, 2018).

De acordo com a proposta taxonômica de Taylor (1989), o gênero *Utricularia* está subdivido em dois subgêneros: *Polypompholyx* (Lehm.) P.Taylor (com duas seções) e *Utricularia* (com 33

seções). Para o tratamento, o autor se baseou principalmente na morfologia vegetativa e relacionada ao cálice e corola, tendo sido criadas para alguns casos seções monotípicas devido à existência de espécies muito distintas. Já outros grupos, como a seção *Utricularia*, foram arranjados principalmente pelo padrão morfológico do utrículo e pela forma de vida aquática. O gênero também apresenta diverso polimorfismo estrutural intra e interespecífico, e espécies bastante similares morfologicamente, sendo muitas vezes erroneamente identificadas mesmo por especialistas. Estudos filogenéticos, feitos a partir de sequências de DNA indicam que as relações propostas por Taylor (1989) são bastante curadas. Contudo, há classificações infragenéricas que necessitam revisão taxonômica.



Figuras 5. Espécies do gênero *Utricularia.* **A.** *Utricularia neottioides* A.St.-Hil. & Girard; **B.** *U. nana* A.St.-Hil. & Girard; **C.** *U. amethystina* Salzm. ex A.St.-Hil. & Girard; **D.** *U. hispida* Lam.; **E.** *U. cucullata* A.St.-Hil. & Girard; **F.** *U sandersonii* Oliver; **G.** *U. nephrophylla* Benj.; **H.** *U. reniformis* A.St.-Hil.; **I.** *U. foliosa* L.; **J.** *U. gibba* L.; **K.** *U nigrescens* Sylvén; **L.** *U. pusilla* Vahl; **M.** *U. flaccida* A.DC.; **N.** *U. triloba* Benj.; **O.** *U. subulata* L.; **P.** *U. nervosa* G.Weber ex Benj. (Fotos de nosso grupo – LSV/FCAV/UNESP.)

Dentro desse contexto, há seções, como *Psyllosperma-Foliosa* e *Iperua-Orchidioides* que, por filogenia molecular, indicam parafilia entre elas. Assim, se faz necessária a investigação filogenética para determinar se as espécies devem ser incluídas em uma única seção, que de acordo com o código de nomenclatura botânica deve ser o nome mais antigo, ou desmembrá-la e realizar um rearranjo de espécies para respeitar a monofilia das seções.

Existe ainda a questão taxonômica de determinadas espécies estarem arranjadas de forma duvidosa na classificação infragenérica, como o caso das espécies *Utricularia olivacea* e *U. flaccida*, que, de acordo com evidências encontradas em estudos baseados em DNA (Muller et al., 2004; Silva et al., 2018), ocasionam a parafilia das seções nas quais são arranjadas (Taylor, 1989) (Figura 6).

4. Genômica em Lentibulariaceae

As Lentibulariaceae têm sido consideradas plantas modelo não somente para se estudarem os processos biológicos relacionados à carnivoria, mas também para investigar processos que envolvem a expansão e contração genômica (Albert et al., 2010). Apresentam espécies com os menores genomas conhecidos (Leushkin et al., 2013), sendo *Utricularia gibba* a menor angiosperma com genoma sequenciado (Ibarra-Laclette et al., 2013; Lan et al., 2017), menor do que o da planta-modelo *Arabidopsis thaliana* (135 - 157 Mb) (Bennett & Leitch, 2005).

Se por um lado o genoma de *Utricularia gibba* é pequeno, por outro lado detém todo repertório gênico de plantas. Assim, a miniaturização do genoma pode ser atribuída à contração de sequências intergênicas e à presença de poucos elementos repetitivos. Enquanto que, para angiospermas, a quantidade de elementos repetitivos fica por volta de 10-60%, para *U. gibba* é de somente 3%. Outro fator relevante é a pouca quantidade de elementos de transposição (cerca de 569). Dentro desse contexto, os retroelementos são raros e somam somente cerca de 2,5% do genoma. Tais considerações apontam que na espécie decorreram vários eventos de duplicação gênica (WGD) seguidos de fragmentação (Ibarra-Laclette et al., 2013). Ao mesmo tempo, foi relatado que

Utricularia e *Genlisea* apresentam alta taxa evolutiva molecular, que pode estar relacionada a uma seleção positiva (Jobson et al., 2003; Wicke et al., 2014). Baseado no tamanho do genoma reduzido, a variabilidade cromossômica e na alta taxa de evolução de nucleotídeos, recentemente foi proposto que *Utricularia* tem mecanismos ativos para remover as regiões do DNA que são danificadas devido às espécies reativas de oxigênio (ROS) causadas pela carnivoria (Albert et al., 2010).

A partir de uma perspectiva filogenética, a expansão e contração genômica ocorreu em diversas linhagens de Lentibulariaceae. O ancestral hipotético do grupo têm genoma estimado de 414Mb (Veleba et al., 2014), um genoma pequeno se considerados os genomas de espécies de Lamiales. A partir dessa perspectiva, o gênero *Pinguicula* possui genomas que estão em discreta expansão genômica. Enquanto *Utricularia* e *Genlisea* sofreram drástica miniaturização independente nos clados da *U.* sect. *Foliosa,* sect. *Vesiculina,* sect. *Utricularia, G.* sect. *Genlisea* e sect. Recurvatae (Figura 7).

Diante do exposto, aliado ao fato de serem plantas no geral de pequeno porte e de fácil cultivo, fica evidente que *Utricularia* pode servir como modelo importante para compreender a estrutura e a evolução dos genomas de plantas, principalmente das angiospermas.

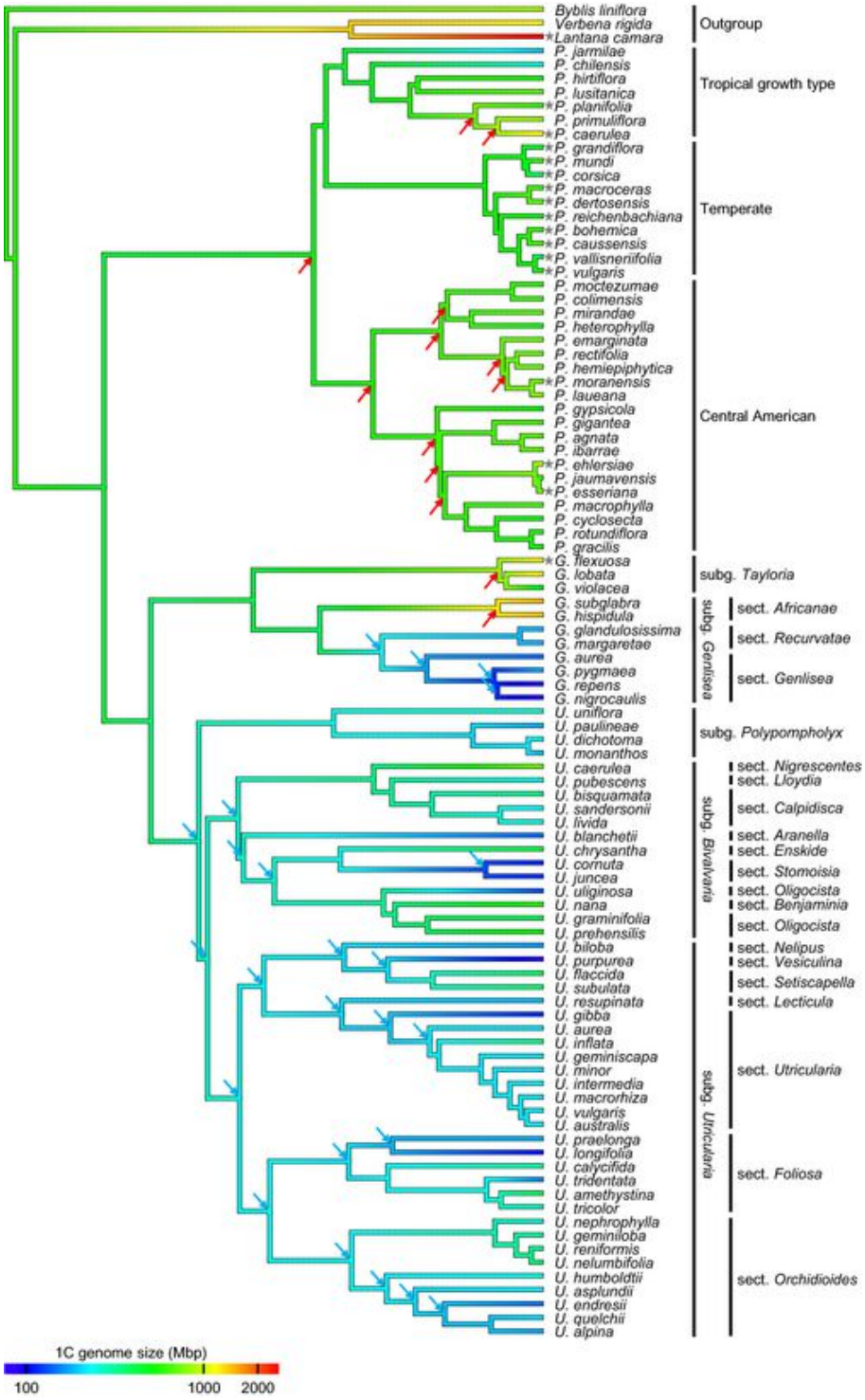


Figura 7. Reconstrução ancestral ("character tracing") do tamanho dos genomas de Lentibulariaceae. As setas azuis denotam miniaturização de genomas e setas vermelhas a expansão. Espécies em que possivelmente ocorreu poliploidia recente estão marcadas por estrelas cinzas (Imagem retirada de Veleba et al., 2014).

5. Organelas

A análise de genomas de cloroplastos e mitocôndrias é valiosa fonte de informações para reconstrução da história evolutiva de plantas. Tem sido usada em numerosos estudos de filogenia em diversos níveis taxonômicos e identificação de plantas. Vale ressaltar ainda que as informações genéticas podem ser usadas como caracteres em nível de nucleotídeos como na forma de caracteres discretos, ou seja, codificando em matrizes a estrutura do genoma, por exemplo, em presença e ausência de genes, e acrescentar a informação para aumentar a robustez das análises filogenéticas.

Nos últimos anos, o sequenciamento de organelas tem se tornado comum, principalmente devido o maior acesso e barateamento dos custos de sequenciamento de nova geração (Straub et al., 2012). De acordo com a base de dados "Organelle Genome Resources" (Wolfsberg et al, 2001) disponível no sítio do NCBI ("National Center for Biotechnology Information"), há 1.805 genomas cloroplastidiais e 223 genomas mitocondriais.

Os métodos principais para obtenção de genomas de organelas incluem: (1) isolamento de organelas por meio de métodos, como gradiente de Percoll (e.g. Dong et al., 2013; Figura 8A e 8B); e a (2) separação de organelas e núcleo *in silico* (e.g. Wang & Messing, 2011; Figura 9). O isolamento de organelas geralmente tem como quesito, grandes quantidades de tecido vegetal que geralmente não estão disponíveis para todas as amostras. Outro aspecto a ser levado em conta é o isolamento de organelas de plantas pequenas, as quais seriam necessários vários indivíduos para fazê-lo, porém, esse processo aumenta as chances de se obterem montagens erradas, principalmente se considerado processos de transferência horizontal de genes e heteroplasmia. Em face disso, o método mais comumente empregado é a separação por métodos *in silico* (Garaycochea et al., 2015).

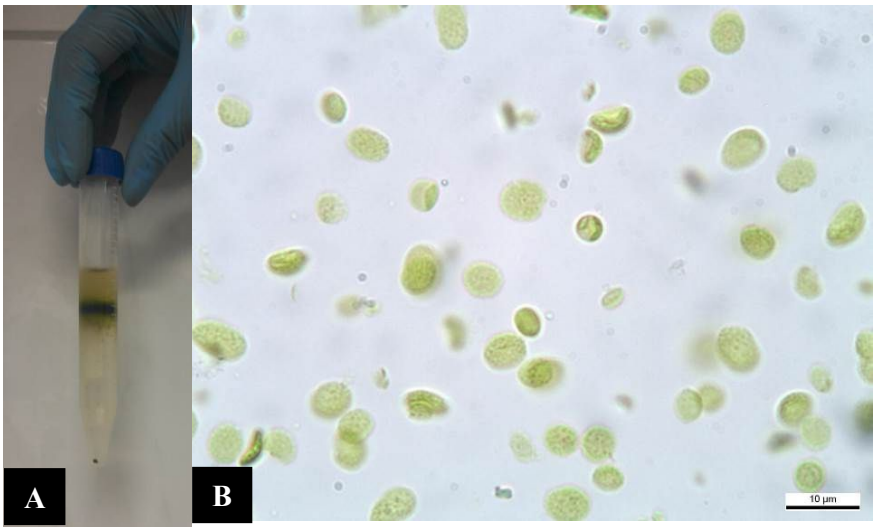


Figura 8. A. Separação de cloroplastos por meio de gradiente de Percoll da espécie *Utricularia foliosa*; **B.** Imagem de microscópio de luz de cloroplastos isolados de *U. foliosa.*

Figura 9. Exemplo de *pipeline* para a montagem de cloroplastos a partir do sequenciamento de DNA total amostra de Lemnoideae (Wang & Messing, 2011).

Os genomas cloroplastidiais são facilmente montados, pois são bastante conservados entre grupos, além disso, são haploides e de herança uniparental. Outro fator relevante é o elevado número de cópias do genoma por célula, que facilita a sua obtenção mesmo a partir de sequenciamento de baixa profundidade. Por outro lado, os genomas mitocondriais podem ser extremamente complexos, mesmo em espécies de relacionamento filogenético próximo, pois sofrem constantes eventos de recombinações devido a repetições em seu mtDNA (Sloan, 2013; Figura 10A), além de apresentarem elementos de transposição e transferência horizontal de genes entre o núcleo e cpDNA (Leister, 2005; Figura 10B).

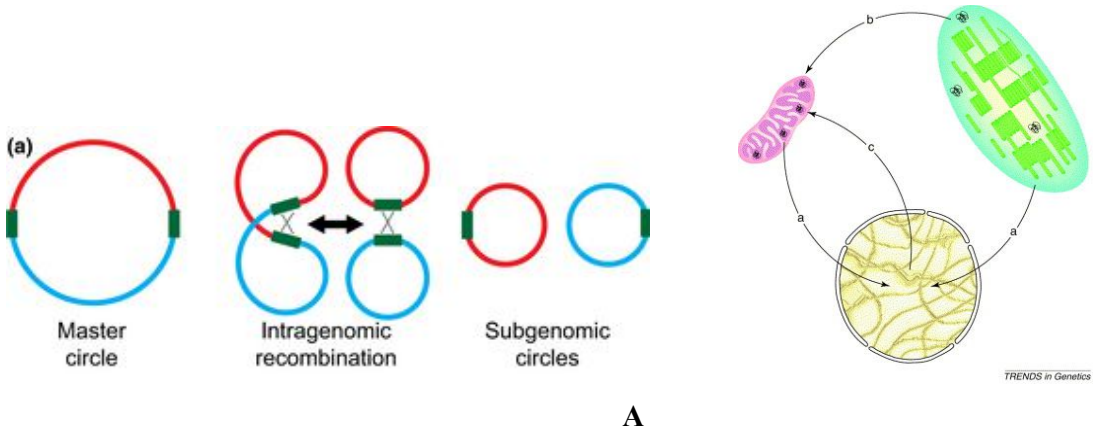


Figura 10. A. Representação das recombinações dos mtDNAs em plantas e formação de subcírculos genômicos (Sloan, 2013). **B.** Representação de transferência horizontal entre genes. (a) Organela para núcleo; (b) cloroplastos para mitocôndria; (c) Núcleo para mitocôndria. (Leister, 2005)

5.1 Organelas em Lentibulariaceae

Até a execução do presente trabalho, poucos estudos haviam sido feitos para investigar as características presentes nas organelas de Lentibulariaceae. Ibarra-Laclette et al. (2013) realizou o primeiro estudo de sequenciamento e montagem de genomas da espécie *Utricularia gibba.* Entretanto, o objetivo principal era a montagem do genoma nuclear, assim, somente com Wicke *et al.* (2014) que foram realizadas as primeiras análises com organelas de Lentibulariaceae, nos quais foram sequenciados plastomas de três espécies dos três gêneros da família: *Pinguicula ehlersiae*, *U. macrorhiza* e *Genlisea margaretae*.

Os cpDNAs de Lentibulariaceae detêm estrutura quadripartida típica, como ocorre para a maioria das angiospermas: duas regiões com genes repetidos, sendo um invertido ao outro (*Inverted Repeats*; IRs) separados por uma região grande de cópia única (*Large Single Copy*; LSC) e outra região pequena de cópia única (*Small Single Copy*; SSC) (Wicke et al., 2014).

Uma das características genômicas que mais receberam ênfase no trabalho, foi a deleção em *Genlisea* e *Pinguicula* de alguns genes NAD(P)H desidrogenase que não possui função exata reconhecida, contudo, é sabido que é responsável por auxiliar a fotossíntese principalmente sob condições de pouca luz e baixa concentração de CO_2 (Wicke *et al.*, 2014). De acordo com os autores, a perda desses genes em *Genlisea* e *Pinguicula* ocorreu em dois eventos independentes na história evolutiva da família Lentibulariaceae, já que na espécie de *Utricularia* analisada não há evidência de deleção ou pseudogenes. Contudo, apesar da relação entre a deleção de genes NAD(P)H e a hipótese filogenética estabelecida por Wicke *et al.* (2014), sabe-se que *Utricularia macrorhiza* é espécie que pertence a grupo filogeneticamente recente segundo a história contada por diversos estudos (Jobson et al. 2003; Müller e Borsch 2005; 2006; Silva et al. 2018).

Considerando tais colocações, são necessários novos estudos a fim de descobrir novas evidências sobre como se deu a evolução dos genomas organelares tanto em relação à família quanto ao gênero, visto que a presença dos genes NAD(P)H pode ser uma condição derivada no gênero *Utricularia* e que o ancestral comum entre *Genlisea* e *Utricularia*, assim como para espécies mais distantes filogeneticamente, podem revelar a deleção de genes NAD(P)H como em *Genlisea margaretae*.

Vale ressaltar ainda que, até o presente trabalho, somente havia em banco de dados públicos o mtDNA parcial (*draft*) de *Utricularia gibba*. Em razão desta realidade, a presente tese propõe a primeira montagem e análise comparativa de organelas de Lentibulariaceae.

8. Objetivos gerais

O objetivo foi investigar o genoma organelar de espécies de Lentibulariaceae para avaliar a estrutura das organelas e potencial para a filogenia das espécies.

8.1 Objetivos específicos

O capítulo 1 intitulado "The chloroplast genome of *Utricularia reniformis* sheds light on the evolution of the *ndh* gene complex of terrestrial carnivorous plants from the Lentibulariaceae family" teve como objetivo a montagem e análise de cloroplastos de *Utricularia reniformis* para propor hipóteses para a pseudogenização e deleção de genes *ndhs*.

O capítulo 2 intitulado "The complete chloroplast genome sequence of the leafy bladderwort, *Utricularia foliosa* L. (Lentibulariaceae) está apresentado aqui como um "Technical Note", como fora concebido, portanto o trabalho teve como objetivo descrever o cloroplasto de *Utricularia foliosa* e apresentar filogenia para verificar a posição da espécie dentro do contexto filogenômico de cpDNAs de Lentibulariaceae.

O capítulo 3 intitulado "The mitochondrial genome of the terrestrial carnivorous plant *Utricularia reniformis* (Lentibulariaceae): Structure, comparative analysis and evolutionary landmarks" teve como objetivo investigar os genes e estrutura mitocondrial a partir da primeira montagem de genoma mitocondrial feita para a planta carnívora do gênero *Utricularia*.

O capítulo 4 intitulado "Comparative genomic analysis of *Genlisea* (corkscrew plants – Lentibulariaceae) chloroplast genomes reveals an increasing loss of the *ndh* genes" teve como objetivo descrição detalhada e sob um contexto evolutivo propor hipóteses para a fragmentação, pseudogenização e deleção de genes *ndhs* dentro do gênero *Genlisea*.

REFERÊNCIAS BIBLIOGRÁFICAS

ALBERT, V. A., JOBSON, R. W., MICHAEL, T. P., & TAYLOR, D. J. The carnivorous bladderwort (*Utricularia*, Lentibulariaceae): a system inflates. **Journal of Experimental Botany**, *61*(1), 5–9, 2010. https://doi.org/10.1093/jxb/erp349

ALCALÁ, R. E., MARIANO, N. A., OSUNA, F., & ABARCA, C. A. An experimental test of the defensive role of sticky traps in the carnivorous plant *Pinguicula moranensis* (Lentibulariaceae). **Oikos**, *119*(5), 891–895, 2010. https://doi.org/10.1111/j.1600-0706.2009.18110.x

BAUER, U., WILLMES, C., & FEDERLE, W. Effect of pitcher age on trapping efficiency and natural prey capture in carnivorous *Nepenthes rafflesiana* plants. **Annals of Botany**, *103*(8), 1219–1226, 2009. https://doi.org/10.1093/aob/mcp065

BENNETT, K. F., & ELLISON, A. M. Nectar, not colour, may lure insects to their death. **Biology Letters**, *5*(4), 469–472, 2009. https://doi.org/10.1098/rsbl.2009.0161

BENNETT, M. D., & LEITCH, I. J. Nuclear DNA amounts in angiosperms: progress, problems and prospects. **Annals of Botany**, *95*(1), 45–90, 2005. https://doi.org/10.1093/aob/mci003

CAMERON, K. M., WURDACK, K. J., & JOBSON, R. W. Molecular evidence for the common origin of snap-traps among carnivorous plants. **American Journal of Botany,** 89(9), 1503–1509, 2002. https://doi.org/10.3732/ajb.89.9.1503

CASPER, S. J. Monographie der Gattung *Pinguicula* L. **Bibl. Bot.,** 127/128, 1–209, 1966.

DONG W, XU C, LI C, SU J, ZUO Y, SHI S, et al. *ycf*1, the most promising plastid DNA barcode of land plants. **Scientific Reports** 2015:1–5, 2015. doi:10.1038/srep08348

ELLISON, A. M., & ADAMEC, L. Ecophysiological traits of terrestrial and aquatic carnivorous plants: are the costs and benefits the same? **Oikos**, 120(11), 1721–1731, 2011. https://doi.org/10.1111/j.1600-0706.2011.19604.x

FLEISCHMANN, A. **A Monograph of the Genus *Genlisea***. Dorset: Redfern Natural History Productions Ltd, 2012.

FROMM-TRINTA, E. Revisão do gênero *Genlisea* St.- Hil. no Brasil. **Boletim do Museu Nacional**, (61), 1–29, 1981.

GARAYCOCHEA, S., SPERANZA, P., & ALVAREZ-VALIN, F. A strategy to recover a high-quality, complete plastid sequence from low-coverage whole-genome sequencing. **Applications in Plant Sciences**, 3(10), 2015. https://doi.org/10.3732/apps.1500022

GIVNISH, T. J. New evidence on the origin of carnivorous plants. **Proceedings of the National Academy of Sciences**, 112(1), 10–11, 2015. https://doi.org/10.1073/pnas.1422278112

GIVNISH, T. J., BURKHARDT, E. L., HAPPEL, R. E., & WEINTRAUB, J. D. Carnivory in the Bromeliad Brocchinia reducta, with a Cost/Benefit Model for the General Restriction of Carnivorous Plants to Sunny, Moist, Nutrient-Poor Habitats. **The American Naturalist**, 124(4), 479–497, 1984.

GUISANDE, C., LORENCIO, C. G., SOSSA, C. E. A., & ESCOBAR, S. R. D. Bladderworts. **Functional Plant Science and Biotechnology**, 1, 58–68, 2004.

IBARRA-LACLETTE, E., LYONS, E., HERNÁNDEZ-GUZMÁN, G., PÉREZ-TORRES, C. A., CARRETERO-PAULET, L., CHANG, T.-H.,et al. HERRERA-ESTRELLA, L. Architecture and evolution of a minute plant genome. **Nature**, 498(7452), 94–98. https://doi.org/10.1038/nature12132, 2013.

JOBSON, R. W., & ALBERT, V. A. Molecular Rates Parallel Diversification Contrasts between Carnivorous Plant Sister Lineages. **Cladistics,** 18(2), 127–136, 2002. https://doi.org/10.1111/j.1096-0031.2002.tb00145.x

JOBSON, R. W., NIELSEN, R., LAAKKONEN, L., WIKSTRÖM, M., & ALBERT, V. A. Adaptive evolution of cytochrome c oxidase: Infrastructure for a carnivorous plant radiation. **Proceedings of the National Academy of Sciences**, 101(52), 18064–18068, 2004. https://doi.org/10.1073/pnas.0408092101

JOBSON, R. W., PLAYFORD, J., CAMERON, K. M., & ALBERT, V. A. Molecular Phylogenetics of Lentibulariaceae Inferred from Plastid rps16 Intron and *trn*L-F DNA Sequences: Implications for Character Evolution and Biogeography. **Systematic Botany**, 28(1), 157–171, 2003. https://doi.org/10.1043/0363-6445-28.1.157

JUNIPER B, ROBINS R, JOEL D. **The Carnivirorous Plants**. Academic Press, London; 1989.

LAN, T., RENNER, T., IBARRA-LACLETTE, E., FARR, K. M., CHANG, T.-H., CERVANTES-PÉREZ, S. A., ALBERT, V. A. Long-read sequencing uncovers the adaptive topography of a

carnivorous plant genome. **Proceedings of the National Academy of Sciences of the United States of America**, 114(22), E4435–E4441, 2017. https://doi.org/10.1073/pnas.1702072114

LAN, T.; RENNER, T.; IBARRA-LACLETTE, E.; FARR, K. M.; CHANG, T-H; CERVANTES-PÉREZ, S.A.; et al. Long-read sequencing uncovers the adaptive topography of a carnivorous plant genome**. Proc Natl Acad Sci**. 201702072, 2017. doi:10.1073/pnas.1702072114

LEGENDRE, L. The genus *Pinguicula* L. (Lentibulariaceae): an overview. **Acta Botanica Gallica**, *147*(1), 77–95, 2000. https://doi.org/10.1080/12538078.2000.10515837

LEISTER, D. Origin, evolution and genetic effects of nuclear insertions of organelle DNA. **Trends in Genetics**, *21*(12), 655–663, 2005. https://doi.org/10.1016/j.tig.2005.09.004

LEUSHKIN, E. V., SUTORMIN, R. A., NABIEVA, E. R., PENIN, A. A., KONDRASHOV, A. S., & LOGACHEVA, M. D. The miniature genome of a carnivorous plant *Genlisea aurea* contains a low number of genes and short non-coding sequences. **BMC Genomics**, *14*(1), 476, 2013. https://doi.org/10.1186/1471-2164-14-476

MCPHERSON, S. **Carnivorous Plants and Their Habitats***: Volume Two*. (A. Fleischmann & A. Robinson, Orgs.). Poole: Redfern Natural History Productions Ltd. 2010.

MIRANDA, V.F.O.; MENEZES, C.G.; SILVA, S.R.; DÍAZ, Y.C.A.; RIVADAVIA, F. Lentibulariaceae in **Lista de Espécies da Flora do Brasil**. Jard. Botânico do Rio Janeiro 12 Jan 2018 Available from: http://floradobrasil.jbrj.gov.br/reflora/floradobrasil/FB146.

MOROHOSHI, T., OIKAWA, M., SATO, S., KIKUCHI, N., KATO, N., & IKEDA, T. Isolation and characterization of novel lipases from a metagenomic library of the microbial community in the pitcher fluid of the carnivorous plant *Nepenthes hybrida*. **Journal of Bioscience and Bioengineering**, *112*(4), 315–320, 2011. https://doi.org/10.1016/j.jbiosc.2011.06.010

MÜLLER, K., BORSCH, T., LEGENDRE, L., POREMBSKI, S., THEISEN, I., & BARTHLOTT, W. Evolution of Carnivory in Lentibulariaceae and the Lamiales. **Plant Biology**, *6*(4), 477–490, 2004. https://doi.org/10.1055/s-2004-817909

MÜLLER, K. F., & BORSCH, T. Phylogenetics of *Utricularia* (Lentibulariaceae) and molecular evolution of the trnK intron in a lineage with high substitutional rates. **Plant Systematics and Evolution**, *250*(1–2), 39–67, 2005. https://doi.org/10.1007/s00606-004-0224-1

MÜLLER, K. F., BORSCH, T., LEGENDRE, L., POREMBSKI, S., & BARTHLOTT, W. Recent Progress in Understanding the Evolution of Carnivorous Lentibulariaceae (Lamiales). **Plant Biology**, *8*(6), 748–757, 2006. https://doi.org/10.1055/s-2006-924706

PAVLOVIČ, A., & SAGANOVÁ, M. A novel insight into the cost-benefit model for the evolution of botanical carnivory. **Annals of Botany**, *115*(7), 1075–1092, 2015. https://doi.org/10.1093/aob/mcv050

PEROUTKA, M., ADLASSNIG, W., VOLGGER, M., LENDL, T., URL, W. G., & LICHTSCHEIDL, I. K. *Utricularia*: a vegetarian carnivorous plant? **Plant Ecology,** *199*(2), 153–162, 2008. https://doi.org/10.1007/s11258-008-9420-3

PŁACHNO, B. J., ADAMEC, L., LICHTSCHEIDL, I. K., PEROUTKA, M., ADLASSNIG, W., & VRBA, J. Fluorescence labelling of phosphatase activity in digestive glands of carnivorous plants. **Plant Biology** *(Stuttgart, Germany)*, *8*(6), 813–820, 2006. https://doi.org/10.1055/s-2006-924177

PŁACHNO, B. J., KOZIERADZKA-KISZKURNO, M., & ŚWIĄTEK, P. Functional Utrastructure of *Genlisea* (Lentibulariaceae) Digestive Hairs. **Annals of Botany**, *100*(2), 195–203, 2007. https://doi.org/10.1093/aob/mcm109

PLACHNO, B. J., KOZIERADZKA-KISZKURNO, M., SWIATEK, P., & DARNOWSKI, D. W. Prey attraction in carnivorous *Genlisea* [Lentibulariaceae]. **Acta Biologica Cracoviensia**. *Series Botanica*, *50*(2), 87–94, 2008.

REIFENRATH, K., THEISEN, I., SCHNITZLER, J., POREMBSKI, S., & BARTHLOTT, W. Trap architecture in carnivorous *Utricularia* (Lentibulariaceae). *Flora* - **Morphology, Distribution, Functional Ecology of Plants**, *201*(8), 597–605, 2006. https://doi.org/10.1016/j.flora.2005.12.004,

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. **Annals of Botany** *117*, 811–832, 2016.

SANABRIA-ARANDA, L., GONZÁLEZ-BERMÚDEZ, A., TORRES, N. N., GUISANDE, C., MANJARRÉS-HERNÁNDEZ, A., VALOYES-VALOIS, V., DUQUE, S. R. Predation by the tropical plant *Utricularia foliosa*. **Freshwater Biology**, *51*(11), 1999–2008, 2006. https://doi.org/10.1111/j.1365-2427.2006.01638.x

SILVA, S. R., GIBSON, R., ADAMEC, L., DOMÍNGUEZ, Y., & MIRANDA, V. F. O. Molecular phylogeny of bladderworts: A wide approach of *Utricularia* (Lentibulariaceae) species relationships based on six plastidial and nuclear DNA sequences. **Molecular Phylogenetics and Evolution**, *118*, 244–264, 2018. https://doi.org/10.1016/j.ympev.2017.10.010

SLOAN, D. B. One ring to rule them all? Genome sequencing provides new insights into the 'master circle' model of plant mitochondrial DNA structure. **New Phytologist**, *200*(4), 978–985, 2013. https://doi.org/10.1111/nph.12395

STRAUB, S. C. K., PARKS, M., WEITEMIER, K., FISHBEIN, M., CRONN, R. C., & LISTON, A. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. **American Journal of Botany**, *99*(2), 349–364, 2012. https://doi.org/10.3732/ajb.1100335

TAYLOR, P. **The genus Utricularia: a taxonomic monograph.** London: Royal Botanic Gardens, 1989.

THE ANGIOSPERM PHYLOGENY GROUP**.** An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. **Botanical Journal of the Linnean Society**, *181*(1), 1–20, 2016. https://doi.org/10.1111/boj.12385

VELEBA, A.; BUREŠ, P.; ADAMEC, L.; ŠMARDA, P.; LIPNEROVÁ, I.; HOROVÁ, L. Genome size and genomic GC content evolution in the miniature genome-sized family Lentibulariaceae. New Phytol, 203: 22–28, 2004. doi:10.1111/nph.12790

WANG, W., & MESSING, J. High-Throughput Sequencing of Three Lemnoideae (Duckweeds) Chloroplast Genomes from Total DNA. **PLoS ONE**, *6*(9), e24670, 2011. https://doi.org/10.1371/journal.pone.0024670

WESTERMEIER, A. S., FLEISCHMANN, A., MÜLLER, K., SCHÄFERHOFF, B., RUBACH, C., SPECK, T., & POPPINGA, S. Trap diversity and character evolution in carnivorous bladderworts (*Utricularia*, Lentibulariaceae). **Scientific Reports**, *7*(1), 12052, 2017. https://doi.org/10.1038/s41598-017-12324-4

WICKE, S., SCHÄFERHOFF, B., DEPAMPHILIS, C. W., & MÜLLER, K. F. Disproportional Plastome-Wide Increase of Substitution Rates and Relaxed Purifying Selection in Genes of Carnivorous Lentibulariaceae. **Molecular Biology and Evolution**, *31*(3), 529–545, 2014. https://doi.org/10.1093/molbev/mst261

WOLFSBERG, T. G., SCHAFER, S., TATUSOV, R. L., & TATUSOVA, T. A. (2001). Organelle genome resources at NCBI. **Trends in Biochemical Sciences,** *26*(3), 199–203, 2011. https://doi.org/10.1016/S0968-0004(00)01773-4

CAPÍTULO 1

The chloroplast genome of *Utricularia reniformis* sheds light on the evolution of the *ndh* gene complex of terrestrial carnivorous plants from the Lentibulariaceae family

Artigo publicado

Silva, S.R., Diaz, Y.C.A., Penha, H.A., Pinheiro, D.G., Fernandes, C.C., Miranda, V.F.O., Michael, T.P., and Varani, A.M. (2016). 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 11, e0165176.

The chloroplast genome of *Utricularia reniformis* sheds light on the evolution of the *ndh* gene complex of terrestrial carnivorous plants from the Lentibulariaceae family

Saura R. Silva[1], Yani C.A. Diaz[2], Helen Alves Penha[3], Daniel G. Pinheiro[3], Camila C. Fernandes[3], Vitor F.O. Miranda[2*], Todd P. Michael[4] and Alessandro M. Varani[3 *]

[1] Instituto de Biociências, UNESP - Univ Estadual Paulista, Câmpus Botucatu, São Paulo, Brazil

[2] Departamento de Biologia Aplicada à Agropecuária, Faculdade de Ciências Agrárias e Veterinárias, UNESP - Univ Estadual Paulista, Câmpus Jaboticabal, São Paulo, Brazil

[3] Departamento de Tecnologia, Faculdade de Ciências Agrárias e Veterinárias, UNESP - Univ Estadual Paulista, Câmpus Jaboticabal, São Paulo, Brazil

[4] Ibis Bioscience, Computational Genomics, Carlsbad, California, USA.

* Corresponding authors

amvarani@fcav.unesp.br (AMV) and vmiranda@fcav.unesp.br (VFOM)

Abstract

Lentibulariaceae is the richest family of carnivorous plants spanning three genera including *Pinguicula*, *Genlisea*, and *Utricularia*. *Utricularia* is globally distributed, and unlike *Pinguicula* and *Genlisea,* has both aquatic and terrestrial forms. In this study we present the analysis of the chloroplast (cp) genome of the terrestrial *Utricularia reniformis*. *U. reniformis* has a standard cp genome of 139,725bp, encoding a gene repertoire similar to essentially all photosynthetic organisms. However, an exclusive combination of losses and pseudogenization of the plastid NAD(P)H-dehydrogenase (ndh) gene complex were observed. Comparisons among aquatic and terrestrial forms of *Pinguicula*, *Genlisea*, and *Utricularia* indicate that, whereas the aquatic forms retained functional copies of the eleven *ndh* genes, these have been lost or truncated in terrestrial forms, suggesting that the ndh function may be dispensable in terrestrial Lentibulariaceae. Phylogenetic scenarios of the ndh gene loss and recovery among *Pinguicula*, *Genlisea*, and *Utricularia*, to the ancestral Lentibulariaceae cladeare proposed. Interestingly, RNAseq analysis evidenced that *U. reniformis* cp genes are transcribed, including the truncated *ndh* genes, suggesting that these are not completely inactivated. In addition, potential novel RNA-editing sites were identified in at least six *U. reniformis* cp genes, while none were identified in the truncated ndh genes. Moreover, phylogenomic analyses support that Lentibulariaceae is monophyletic, belonging to the higher core Lamiales clade, corroborating the hypothesis that the first *Utricularia* lineage emerged in terrestrial habitats and then evolved to epiphytic and aquatic forms. Furthermore, several truncated cp genes were found interspersed with *U. reniformis* mitochondrial and nuclear genome scaffolds, indicating that as observed in other smaller plant genomes, such as Arabidopsis thaliana, and the related and carnivorous *Genlisea nigrocaulis* and *G. hispidula*, the endosymbiotic gene transfer may also shape the *U. reniformis* genome in a similar fashion. Overall the comparative analysis of the *U. reniformis* cp genome provides new insight into the ndh genes and cp genome evolution of carnivorous plants from Lentibulariaceae family.

Introduction

Carnivorous plants from the genus *Utricularia* (Lentibulariaceae) are distributed worldwide and are comprised of approximately 235 species occurring across every continent except the poles, some arid regions and oceanic islands [1]. They are highly specialized plants with modified leaves (traps) for capturing prey [2,3], and there are diverse life forms, such as aquatic, terrestrial, epiphytic, and reophytic forms [1]. In addition, the genus has some of the smallest nuclear genomes across the angiosperms, ranging from 77 to 561 megabases (Mb; 1C), surpassed only by the genus *Genlisea* with species possessing the smallest genomes [4]. Interestingly, angiosperms present extremes in genome sizes, from around 61 Mb of *Genlisea* to 150,000 Mb of one of the largest plant genomes known, the monocot *Paris japonica* [5,6]. From this perspective, Lentibulariaceae also provides outstanding candidates for model plants, with flexible genomes from around 61 to 1,500 Mb [4,7]; it is thus an important group to address evolutionary, genomic and phylogenetic as well as functional questions. For instance, it is well known that polyploidy, the amount of repetitive DNA such as transposable elements and other repeats, and whole genome duplications (WGD), along with other mechanisms such as small-scale genome duplications or fractionation/gene death rates are the key drivers of genome size differences [6,8,9].

Utricularia reniformis A.St.-Hil. is a terrestrial bladderwort endemic to the Brazilian Atlantic Forest and restricted to the mountaintops of the southeastern coast of Brazil [1,10]. All known populations are limited to elevations above 700 m, and live in wet grasslands, wet rocks, and in Bromeliaceae leaf axes [11]. Little is known about the *U. reniformis* genome structure. However, previous studies based on microsatellite markers have shown that different populations of *U. reniformis* may have tetraploid genomes with high levels of heterozygosity [10]. Recently, the aquatic *U. gibba* was sequenced, revealing that its 82 Mb genome encodes a typical number of genes for a plant at 28,500 but that the genome has gone through several whole genome

duplications and reductions. In addition, it was reported that the 152kb chloroplast (cp) genome is similar to most other angiosperms [9]. Furthermore, additional studies have focused on genome contraction [12,13] and comparative transcriptomics comparing the different organs and tissues in the aquatic *U. gibba* genome [14]. In order to better understand the evolutionary dynamics of a terrestrial form, the *Utricularia reniformis* A.St.-Hil. genome and transcriptome of different tissues are being sequenced and analyzed by our research group.

Four complete cp genomes of the Lentibulariaceae have been published [9,15]. These are *U. gibba* and *U. macrorhiza*, which are found in aquatic environments as submersed plants, and *Genlisea margaretae* and *Pinguicula ehlersiae*, which are terrestrial species. These four cp genomes are composed of large and small single copy (LSC and SSC) regions and two inverted repeats (IRs), which is the typical circular cp genome quadripartite structure (LSC-IR-SSC-IR). In addition, independent deletions and p*seudogenization* of subunits of the NAD(P)H dehydrogenase complex (*ndh*), altered proportions of repeats, increase of substitution rates on coding regions and microstructural changes were observed as important landmarks of these cp genomes [15].

The *ndh* complex is composed of eleven genes (*ndhA*, *ndhB*, *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhK*, and *ndhJ*), and together with the nuclear genes *nhdL*, *ndhM*, *ndhN*, and *ndhO* encode the thylakoid NAD(P)H dehydrogenase complex. The complex is involved in the electron transfer from NAD(P)H to plastoquinone, protecting the plant cell against photo-oxidative stress, and maintaining optimal rates of cyclic photophosphorylation [16,17]. Interestingly, the *ndh* genes are related to land adaptation and photosynthesis, whereas small changes in any of the *ndh* genes significantly decrease the photosynthesis rate [18]. The *ndh* loss is mainly associated with heterotrophic (parasitic) plants and those endemic to an underwater environment [16]. Indeed, the plants found in submersed aquatic environments receive low levels of light, ultraviolet radiation and have specific limitations that require a number of adaptations, and therefore are under different selective pressures than terrestrial plants [19]. However, the *ndh* genes may be dispensable to the

plant. Functional studies suggest that depending on the environmental condition and stress, alternative metabolic pathways might surpass the absence of the *ndh* genes, indicating that these genes may not be central for photosynthesis [19].

However, other questions about the evolution of the Lentibulariaceae cp genomes still remain. For instance, is lateral gene transfer of cp DNA to the nuclear and/or mitochondrial genomes occurring very frequently as it does in other angiosperms [20]. This process is termed endosymbiotic gene transfer [21], and has not been fully addressed in the Lentibulariaceae cp genomes. Furthermore, previous studies suggested that there is a close relationship between carnivorous and parasitic plants at the level of the cp genome [15,22]. Therefore, sequenced cp genomes from this order are still necessary to further understand this relationship, and also to shed light on the systematic relationships in the order Lamiales.

To provide additional insights into the plastid genomes and the driving forces related with plastid NAD(P)H dehydrogenase complex genes loss and pseudogenization in terrestrial carnivorous plants, we have sequenced the cp genome of the *U. reniformis,* and compared it to the other cp genomes among other Lamiales, with a specific focus on the carnivorous plants from the family Lentibulariaceae. We have also analyzed the gene content and expression by RNAseq, RNA editing, repeats and structure of the *U. reniformis* plastid genome.

Material and Methods

Plant Sampling

The *Utricularia reniformis* A.St.-Hil. plant samples were collected in the fall of 2015 near the Serra do Mar Atlantic Forest, located in Salesópolis Municipality, São Paulo State, Brazil (Geographic Location: -23.5047, -45.9018, 961 m a.s.l.) and deposited in the JABU Herbarium of São Paulo State University (voucher V.F.O de Miranda et al., 1670 – JABU). 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 [23].

Chloroplast Sequencing and Assembly

The DNA was extracted following the QIAGEN DNeasy Plant Maxi Kit extraction protocol (QIAGEN). The whole-genome shotgun sequencing was performed using the Illumina MiSeq technology with a paired-end library of 2x300bp and insert size of ~600 bp. The library construction followed the Illumina Nextera XT Preparation Guide (Illumina, USA). The DNA was tagmented (tagged and fragmented) by the Nextera XT transposome. The Nextera XT transposome simultaneously fragments the input DNA and adds adapter sequences to the ends. The tagmented DNA is amplified via a limited-cycle PCR program. The PCR step also adds index 1 (i7) and index 2 (i5) to sequences required for cluster formation. After that, a PCR clean-up was performed by AMPure XP beads to purify the library DNA, and provides a size selection step that removes very short library fragments from the population. A total of 40M paired-end reads were generated and used for the cp genome assembly. Furthermore, in our ongoing studies of the *Utricularia reniformis* genome, we also sequenced 160M mate-pair reads (2x100 bp) with an average of 3,500 bp insert size using the Illumina HiScanSQ technology. The library construction followed the Nextera mate pair gel free protocol (Illumina, USA). The mate-pair set of reads was used in this study to confirm the cp assembly. Poor quality sequences (phred < 24), contaminants, adapters, and sequences with less than 50bp were removed using seqyclean software (https://github.com/ibest/seqyclean), leaving 36M (2x300 bp, paired-end) and 150M (2x100 bp, mate-pair) high quality reads.

The cp genome assembly was conducted in three steps. First, the trimmed reads were mapped back to *Utricularia gibba*, *U. macrorhiza, Genlisea margaretae*, and *Pinguicula ehlersiae* cp genomes (Accession numbers on S1 Table) with bowtie2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml) [24] using default parameters. In the second step, the resulting potential cp reads were then assembled separately with SPAdes v3.7.1

(http://bioinf.spbau.ru/spades) [25], and by iterative (mapping) assembly with MITObim (https://github.com/chrishah/MITObim) [26], using *U. gibba*, *U. macrorhiza, G. margaretae* and *P. ehlersiae* cp genomes as references. In the third step, the cp genome was manually reconstructed using the SPAdes and MITObim results. The mate-pair reads were mapped back to the cp genome with bowtie2, with the --*very-sensitive* parameter, to confirm the assembly.

Annotation and Comparative Analysis of the Chloroplast Genomes

The cp genome was annotated using DOGMA (Dual Organellar GenoMe Annotator - http://dogma.ccbb.utexas.edu/)[26] coupled with Prodigal v2.6.2 (http://prodigal.ornl.gov/) [27] and Blast (https://blast.ncbi.nlm.nih.gov) [28] for additional gene location and refinements. The Aragorn software package (http://mbio-serv2.mbioekol.lu.se/ARAGORN/) [29] was used for tRNAs validation and intron identification. Corrections of start and stop codons, and annotation curation were made with Artemis genome browser (http://www.sanger.ac.uk/science/tools/artemis) [30]. In this study the potential pseudogenes were defined by Blast comparative analysis with the use of at least one of the following criteria: (a) presence of at least one stop codon in-frame with the predicted coding region; (b) absence of start and/or stop-codon; (c) frameshift; (d) lacking of at least 20% of the coding region when compared to the respective coding region of related species. For gene assignments the Uniprot (http://www.uniprot.org/) [31] and InterProScan (https://github.com/ebi-pf-team/interproscan) [32] databases were used. The codon and amino acid usage were calculated using CodonW v1.4.4 (http://codonw.sourceforge.net). The circular gene maps were made with OGDRAW (OrganellarGenome DRAW - http://ogdraw.mpimp-golm.mpg.de/) [33]. Comparative analysis was carried out with Interactivenn (http://www.interactivenn.net/) [34], and Blast.

The annotated sequence and the raw reads for the *U. reniformis* chloroplast genome have been deposited in the GenBank database under accession number [GenBank: KT336489 and SRR3277235, respectively] (BioProject PRJNA290588).

Microsatellite and other repeats analysis

Microsatellite analyses were carried out with the MISA software package (http://pgrc.ipk-gatersleben.de/misa/) [35], with thresholds of seven repeats for mononucleotide SSRs, four repeats for di- and trinucleotide SSRs, and three repeats for tetra-, penta- and hexanucleotide SSRs. Maximal number of bases interrupting 2 SSRs in a compound microsatellite was set to 100 bp. Direct and palindromic repeats were determined with REPuter (https://bibiserv2.cebitec.uni-bielefeld.de/reputer) [36], with a minimal size of ≥ 30 bp, sequence identity ≥ 90% (hamming distance of 3).

Phylogenomic and phylogenetic analyses: Maximum Likelihood and Bayesian Inference

A total of 47 coding sequences from different chloroplasts were considered (S1 Table, A-B). The alignment of sequences was performed using MAFFT (http://mafft.cbrc.jp/alignment/software/) [37] with default parameters.

The phylogenetic analysis of the plastomes utilized Oleaceae as outgroup. For the probabilistic analysis, the best evolutionary models (best-of-fit) were chosen using ModelTest 3.7 (http://www.molecularevolution.org/software/phylogenetics/modeltest) [38]. Thus, the best-of-fit DNA model was evaluated for each data matrix with the corrected Akaike information criterion [39,40] to estimate the parameters. Maximum likelihood (ML) and Bayesian analyses were performed to estimate the phylogenetic hypothesis for each data matrix. The ML analyses were run with RAxML (http://sco.h-its.org/exelixis/web/software/raxml/index.html) [41]. Prior to the probabilistic analyses, the Akaike information criterion was used to compare the fit to the data of different models as implemented in ModelTest, resulting in the selection of GTR+GAMMA+I as the best-of-fit model. Thus, the GTR+GAMMA+I model was also employed and bootstrapping was applied with 10,000 pseudoreplicates. Bayesian analyses were performed with MrBayes software

version 3.2.5 (http://mrbayes.sourceforge.net/) [42] for each data set with $9x10^6$ generations sampled for each 100 generations, using the default parameters. For each analysis, two runs (nruns=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 (except the one with optimizations of ancestral states) were drawn with the program TreeGraph2 beta version 2.0.52-347 (http://treegraph.bioinfweb.info/) [43].

The phylogenetic analyses of the evolution of *ndh* genes was carried out using a matrix of presence/absence (pseudogenes and frame-shifts were regarded as absences), and the plastome phylogenomic tree (described above) was considered for tracing the presence/absence of *ndh* genes with the Mesquite software version 3.04 (http://mesquiteproject.org). The cloudgram was produced by DensiTree version 2.1 (https://www.cs.auckland.ac.nz/~remco/DensiTree/) [44] based on 18,000 trees sampled with Bayesian inference (using the same parameters as described above for phylogenomic analyses of plastomes) but with the *matK* gene. The *matK* data set was produced by fifty-five sequences from NCBI and also by sequences produced by this study (S1 Table, C-D). The sequences produced in this study were made with 1R-KIM and 3F-KIM primers, following the PCR protocol and procedures recommended by the CBOL Plant Working Group [45].

RNA-Seq and RNA-edit analyses

Three different plant tissues were used for RNA-seq analysis; fresh leaves, stolons and utricules. The tissues were pooled in three replicates and the total RNA (including the rRNA) was extracted using the PureLink RNA Mini Kit (Thermo Fisher Scientific), according to the manufacturer's 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) $\geq$ 7.0 were used for the sequencing. The cDNA libraries were sequenced on the Ion Proton System generating a yield of 180M of reads with an average read length of 200bp,

representing the nuclear and organellar cDNAs. Poor quality sequences (phred < 20), adapters, bacterial contaminants such as photoautotrophic bacteria, and sequences with less than 20bp were removed using prinseq lite v0.20.4 [46]. Two different approaches were used to distinguish potential nuclear/mitochondrial transcripts from authentic plastid transcripts. First, filtered reads were mapped back to the assembled *U. reniformis* plastome with bowtie2, with the *–very-sensitive* and *--end-to-end* parameters. The libraries of all three tissues were pooled and this generated a total of 1,632,156 cp related reads with an average length of 170bp for downstream analysis. Second, the RNAseq reads were mapped with CLC Genomics Workbench v9 (QIAGEN Aarhus, Denmark - http://www.clcbio.com) 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 cp genes were considered for the RNAseq read mapping and transcription abundance by RPKM (Reads Per Kilobase Million) normalization, whereas only unique read mapping were considered. In addition, the intronic regions of intron-containing genes were also considered for the identification of spliced exons.

The RNA-editing analyses were conducted with TopHat2 (https://ccb.jhu.edu/software/tophat/index.shtml) [47] against the pooled reads using the following parameters: no coverage search (--no-coverage-search), filtering multiple mapped reads such as low complexity or repetitive reads (--prefiltered-multihits), reads spanned by junctions with a minimum of 10 bases (--min-anchor-length 10) or with a maximum of 1 base mismatch in the anchor region (--splice-mismatches 1), to align trans-spliced genes, fusion search was activated (--fusion-search) with the minimum distance of 10,000 (--fusion-min-dist 10,000). The final alignment was inspected for C to U or U to C nucleotide substitutions by a custom perl script, with the use of the following parameters: editing site with a minimum coverage of 10x and phred quality score of ≥25. In addition, the PREP-Cp tool (http://prep.unl.edu/) [48] was used with default parameters to predict additional RNA editing sites.

The cp RNA-seq reads used in this study have been deposited in the GenBank database under accession number [GenBank: SRP072162] (BioProject PRJNA290588).

Results

High-quality assembly of the chloroplast genome sequence

A total of 1,259,272 (paired-end, 2x300bp), and 2,087,893 (mate-pair, 2x100bp) chloroplast (cp) related reads were filtered from the raw reads generated by the Illumina MiSeq and HiScanSQ platforms, respectively. These reads shows an average phred value above 30. A total of 1,029,442 (81.7%) paired-reads were assembled to the *U. reniformis* cp contigs. This resulted in the assembly of the entire LSC region as well as part of the IRs and the SSC, in three different contigs. The plastid supercontig was manually closed by iterative read mapping and Blast searches. This resulted in a circular sequence with 139,725 bp, with a GC content of 38.15% and average coverage of 3,300x (maximum coverage peak of 8,563x, and standard deviation of 1,725). No high-quality discrepancies were observed, thus indicating a high quality cp genome assembly. A total of 1,751,896 (84%) mate-pair reads were mapped in pairs to the assembled cp genome according to the average insert size of the library (~3,500bp), thus confirming the cp genome assembly. In addition, the mate-pairs read mapping confirmed the assembly of the LSC/IRa, LSC/IRb, and IRa/SSC/IRb boundaries, and therefore the circularity of the plastid supercontig.

Interestingly, the remaining paired-end reads (2x300bp; 229,830; 18.3%) were assembled into contigs containing fragments of incomplete or truncated cp genes (at least 28 contigs encoding 48 truncated cp genes, which span a total of 23,606kbp - S2 Table). Interestingly, these truncated genes are present as full-length gene copies in the cp genome. It is worth noting that the boundaries of some of these contigs do not have similarity to the assembled *U. reniformis* cp genome, nor the other Lentibulariaceae cp genomes analyzed. Indeed, during the ongoing genome sequence of the *Utricularia reniformis* A.St.-Hil., we have noted that these cp-derived sequences are interspersed within *U. reniformis* mitochondrial (mt) and nuclear genome scaffolds (data not shown). Therefore,

this finding suggests that these contigs do not belong to the cp genome sequence itself, and most likely represents distinct endosymbiotic gene transfer events. In addition to these 28 contigs, we have noted an additional number of truncated cp genes interspersed among the *U. reniformis* mt genome. For instance, truncated copies of the *ndhJ*, *ndhK* and *ndhC* genes, which are absent in the cp genome (see below), are present in the mt genome. These findings strongly support that gene transfer events between the organelles and nuclear DNA occurred and may shaped the *U. reniformis* genome.

Organization and gene content of the *U. reniformis* plastid genome

Utricularia reniformis circular cp genome size is 139,725 bp in length, showing the typical quadripartite structure found in most land plants. The IRs spans 24,064bp each (34%), whereas the small single copy (SSC), and large-single-copy region (LSC) span 12,661bp (10%) and 78,936bp (56%), respectively (Fig 1). The *ndhJ*, *ndhK* and *ndhC,* which are commonly located on t*he* LSC in angiosperms, are lost in the *U. reniformis* cp genome (Fig 1). However, with that exception, the *U. reniformis* cp LSC is collinear in gene content and arrangement to the respective region of the closely related species, *U. gibba*, *U. macrorhiza, G. margaretae* and *P. ehlersiae and to* other angiosperms, such as *Arabidopsis thaliana* and *Nicotiana tabacum.* The GC content of the LSC and SSC regions are 36% and 31.75%, respectively, whereas that of the IR regions is 43%. As observed in other land plants, such as from *Cynara humilis*, a species of family Asteraceae, the higher GC content in the IRs is due to the GC rich rRNA genes *rrn16*, *rrn23*, *rrn4*.5, and *rrn5* [49].

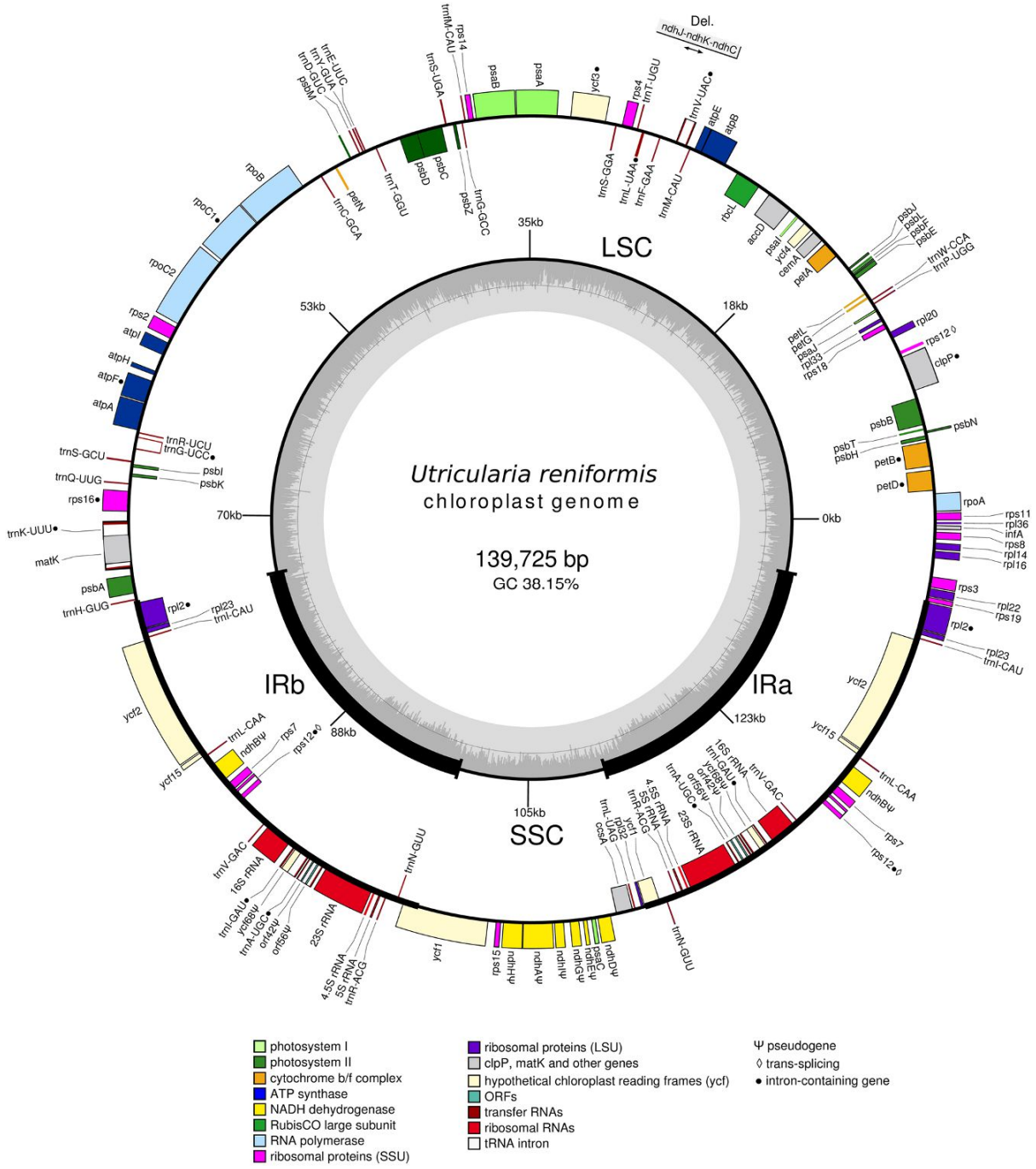


Figure 1. Genomic map of the *Utricularia reniformis* cp genome. Genes shown on the outside of the map are transcribed clockwise, whereas genes on the inside are transcribed counter-clockwise. Genes are color coded by their function in the legend.

There are a total of 136 predicted coding regions, 90 of which are single copy (68 CDS and 22 tRNAs), and 46 of which are duplicated in the IRs (22 CDS, 16 tRNAs and 8 rRNAs) (Table 1). In addition to the predicted coding regions, 14 pseudogenes, 6 of which are single copies, mostly located on the SSC and related with *ndh* complex genes, were identified (Fig 1 and Table 1). A total of 22,601 codons represent the coding repertoire of the protein coding regions (Table 2). The most prevalent codon encodes for leucine (2,326 – 5.99%), whereas the least is cysteine (243 – 1.07%). Only four coding regions have alternative start codons, these are the *rpl*16 (AUC), *rps*19 (GUG) and *ycf*15 (GUG). All 3 of the stop codons are present, with UAA being the most frequently used (UAA 51%, UAG 27% and UGA 21%). *T*he predicted tRNA genes enable the *U. reniformis* cp genome to decode all amino acids, but not all 61 codons (29 out 64 codons; *Table 2)*. A similar codon distribution is also observed with the related species *U. gibba*, *U. macrorhiza, G. margaretae* and *P. ehlersiae*.

Table 1. List of genes encoded by the *Utricularia reniformis* chloroplast (cp) genome.

Photosystem I	*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*
Protosystem II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*
Cytochrome b/f complex	*petA*, *petB*•, *petD*•, *petG*, *petL*, *petN*
ATP synthase	*atpA*, *atpB*, *atpE*, *atpF*•, *atpH*, *atpI*
NADH dehydrogenase	*ndhA* Ψ, *ndhB* Ψ (2x), *ndhD* Ψ, *ndhE* Ψ, *ndhG* Ψ, *ndhH* Ψ, *ndhI* Ψ
RubisCO large subunit	*rbcL*
RNA polymerase	*rpoA*, *rpoB*, *rpoC1*•, *rpoC2*
Ribosomal proteins (SSU)	*rps2*, *rps3*, *rps4*, *rps7* (2x), *rps8*, *rps11*, *rps12* ◊•(2x), *rps14*, *rps15*, *rps16* •, *rps18*, *rps19*
Ribosomal proteins (LSU)	*rpl2* (2x)•, *rpl14*, *rpl16*, *rpl20*, *rpl22*, *rpl23* (2x), *rpl32*, *rpl33*, *rpl36*
Other genes	*ccsA*, *clpP* •, *matK*, *accD*, *cemA*, *infA*
hypothetical chloroplast reading frames	*ycf1* (2x), *ycf2* (2x), *ycf3* •, *ycf4*, *ycf15* (2x), *ycf68* Ψ(2x), *orf42* Ψ(2x), *orf56* Ψ(2x)
Transfer RNAs	*trnA*-UGC (2x)•, *trnC*-GCA, *trnD*-GUC, *trnE*-UUC, *trnF*-GAA, *trnG*-GCC, *trnG*-UCC•, *trnH*-GUG, *trnI*-CAU (2x), *trnI*-GAU (2x)•, *trnK*-UUU•, *trnL*-CAA (2x), *trnL*-UAA•, *trnL*-UAG, *trnM*-CAU, *trnN*-GUU (2x), *trnP*-UGG, *trnQ*-UUG, *trnR*-ACG (2x), *trnR*-UCU, *trnS*-GCU, *trnS*-GGA, *trnS*-UGA, *trnT*-GGU, *trnT*-UGU, *trnV*-GAC (2x), *trnV*-UAC•, *trnW*-CCA, *trnY*-GUA, *trnfM*-CAU
Ribosomal RNAs	*rrn4.5* (2x), *rrn5* (2x), *rrn16* (2x), *rrn23* (2x)

Ψ pseudogene
◊ trans-splicing
• intron-containing gene

Table 2. Codon usage table. Codon usage and codon–anticodon recognition pattern for tRNA in *Utricularia reniformis* cp genome.

Amino acid	Codon	No.	RSCU (a)	%(b)	tRNA	Amino acid	Codon	No.	RSCU (a)	%(b)	tRNA
Ala	GCU	479	1.64	40.9	-	Pro	CCU	339	1.42	35.6	-
	GCC	201	0.69	17.2	-		CCC	214	0.90	22.4	-
	GCA	346	1.18	29.6	trnA-UGC •		CCA	240	1.01	25.2	trnP-UGG
	GCG	143	0.49	12.3	-		CCG	159	0.67	16.8	-
Cys	UGU	181	1.49	74.4	-	Gln	CAA	629	1.51	75.3	trnQ-UUG
	UGC	62	0.51	25.6	trnC-GCA		CAG	206	0.49	24.7	-
Asp	GAU	702	1.60	79.9	-	Arg	CGU	301	1.24	20.7	trnR-ACG
	GAC	176	0.40	20.1	trnD-GUC		CGC	115	0.47	7.9	-
Glu	GAA	878	1.50	74.9	trnE-UUC		CGA	338	1.39	23.3	-
	GAG	294	0.50	25.1	-		CGG	128	0.53	8.7	-

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

Phe	UUU	806	1.31	65.4	-		AGA	407	1.68	27.9	trnR-UCU
	UUC	426	0.69	35.6	trnF-GAA		AGG	168	0.69	11.5	-
Gly	GGU	483	1.27	31.7	-	Ser	UCU	475	1.67	27.8	-
	GGC	184	0.48	12.2	trnG-GCC		UCC	275	0.97	16.1	trnS-GGA
	GGA	555	1.46	36.4	trnG-UCC •		UCA	320	1.12	18.7	trnS-UGA
	GGG	300	0.79	19.7	-		UCG	198	0.70	11.6	-
His	CAU	420	1.49	74.6	-		AGU	345	1.21	20.2	-
	CAC	142	0.51	25.4	trnH-GUG		AGC	95	0.33	5.6	trnS-GCU
Ile	AUU	916	1.50	49.9	-	Thr	ACU	464	1.54	38.3	-
	AUC	369	0.60	20.1	trnI-GAU		ACC	234	0.77	19.4	trnT-GGU
	AUA	548	0.90	30	trnI-CAU •		ACA	366	1.21	30.3	trnT-UGU
Lys	AAA	990	1.48	73.9	trnK-UUU •		ACG	144	0.48	12	-
	AAG	348	0.52	26.1	-	Val	GUU	457	1.46	36.5	-
Leu	UUA	704	1.82	30.3	trnL-UAA •		GUC	163	0.52	13.1	trnV-GAC
	UUG	493	1.27	21.2	trnL-CAA		GUA	466	1.49	37.2	trnV-UAC •
	CUU	501	1.29	21.5	-		GUG	166	0.53	13.2	-
	CUC	148	0.38	6.3	-	Tyr	UAU	642	1.52	76	-
	CUA	312	0.80	13.4	trnL-UAG		UAC	202	0.48	24	trnY-GUA
	CUG	168	0.43	7.3	-	Trp	UGG	384	1.00	100	trnW-CCA
Met (START)	AUG	480	1.00	100	trnf(f)M-CAU	Stop	UGA	16	0.63	21	-
Asn	AAU	843	1.49	74.6	-		UAA	39	1.54	51	-
	AAC	287	0.51	25.4	trnN-GUU		UAG	21	0.83	27	-

(a) Relative Synonymous Codon Usage
(b) Codon frequency (in %) per amino acid
•Intron-containing tRNA genes

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

Other features are related with the genes *atpF*, *rpl2*, *rps16*, *rpoC1*, *rps16, petB* and *petD* which each contain one intron, and *clpP* and *ycf3* which each contain two introns. The introns of all protein-coding genes share the same splicing mechanism as Group II introns. The *rps12* gene is transpliced, the 5' end exon is located in the LSC region and the 3' exon and intron are duplicated and located in the IR regions. This is frequently observed in other angiosperms as well [50,51]. The splicing was also evidenced by mapping of the RNA-seq reads on spliced exons including those from transpliced genes *rps12*. Conversely, six tRNAs contain introns. In the IRs, the *trnI*-GAU intron includes the *ycf68* pseudogene, and *trn*A-UGC intron includes the *orf42* and *orf56* pseudogenes. On the LSC, *trn*G-UCC, *trn*V-UAC and *trn*L-UAA each contain one intron, and the *trn*K-UUU intron includes the *matK* gene. The *matK* encodes for a maturase which is not able to promote intron mobility due to the absence of the reverse transcriptase domain, and the *trnL*-UAA contains a group I intron. This genomic organization is highly conserved in the *U. gibba*, *U. macrorhiza, G. margaretae* and *P. ehlersiae*.

In summary, 68,969 bp (49.4%) of the cp genome is made up of coding DNA, whereas 57,159 bp (40.9%) 2,766 bp (2%), and 9,044 bp (6.4%) corresponds to CDS, tRNA and rRNAs, respectively. Introns represent 16,811 bp (12%), pseudogenes 6,453 bp (4.6%), and the remaining 47,492 bp (34%) of the genome is made up of non-coding and intergenic spacers.

Transcription evidence of the plastidial genes

Due to its endosymbiotic origin, the chloroplast retains a prokaryotic biochemistry, with its own gene expression machinery. However, the expression of cp genes in many land plants requires different nuclear-encoded proteins, mostly to bind transcripts and regulate translation [52]. Conversely, transcription of some genes is regulated by light-dark cycles or nutrient availability, such as those from photosystem I and II, and *rbcL*, and the results presented here show that *U. reniformis* follows this trend. Moreover, the chloroplast RNAs shows a poly(A) tail, which may

target the cp RNAs for rapid decay [53]. In order to explore the *U. reniformis* cp genome expression, cDNA libraries from three different tissues (leaves, stolon, and utricules) were created and sequenced (RNAseq).

The libraries were merged, and a total of 113,224 (7%, out from 1,632,156) reads were uniquely mapped to the annotated cp genes, and the expression profiles of the cp genes were investigated. Overall, >95% of the coding regions and pseudogenes have at least one read of coverage (average base coverage of 183x and median of 33x). The remaining reads (1,087,926; 65.5%) map to the cp rRNAs, such as the 23S rRNA, and also to the tRNAs, and 431,006 (27.5%) reads are non-specific matches. These results show that the cp genome is transcriptionally active, and that all predicted genes, including the pseudogenes are transcribed at some level (Fig 2 and S3 Table). The transcription of genes from photosystem I and II are highly expressed compared to the other genes. In particular the *psaJ* and *psbA* genes show the highest expression levels as measured in RPKM (323,421 and 437,421, respectively*; S3 Table),* and were removed from Fig 2 for brevity and clarity. Moreover, the genes *psaC*, *psaI*, *psbB*, *psbC*, *petB*, *atpA*, *atpH*, *rpl2*, *clpP*, and *rbcL* show expression near or above 20,000. The least expressed gene is the *petG*, which encodes the cytochrome b6/f complex subunit V, with expression value of 77, and only one RNAseq read mapped (covering >95% of *petG* with 99% of identity). Other genes with low RPKM expression values are: *psbF* (15 transcripts), *psbI* (11 transcripts), *psbJ* (12 transcripts), *psbK* (7 transcripts), *psbT* (6 transcripts), *petL* (2 transcripts), *rpoB* (409 transcripts), *rpoC1* (246 transcripts), *rpoC2* (566 transcripts), *rpl23* (11 transcripts), *rpl32* (5 transcripts), *ycf15* (11 transcripts) and *ycf2* (575 transcripts) (More details in S3 Table).

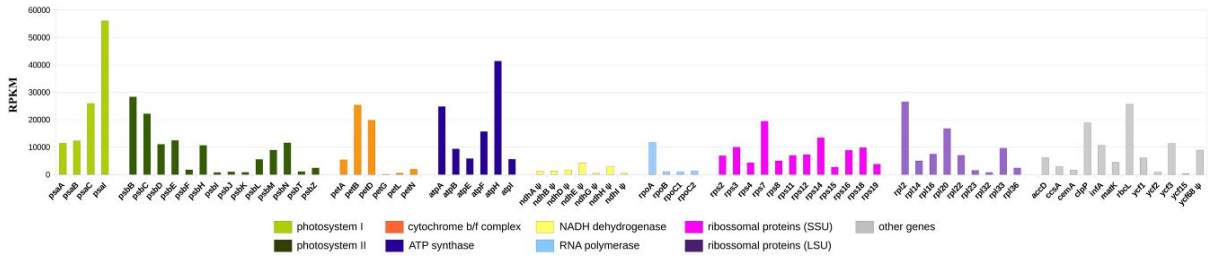


Figure 2. Expression profile of the *Utricularia reniformis* cp genes. The expression values are normalized in Reads Per Kilobase per Million mapped reads (RPKM) on the Y-axis. The *psaJ* and *psbA* genes show the highest expression levels as measured in RPKM (323,750 and 437,421, respectively) and were suppressed from the figure for clarity.

Another remarkable feature observed is related to the transcriptional activity of pseudogenes. Interestingly, all *ndh* complex genes have some sort of mutation, which lead to stop codons in-frame, frameshift and deletions (detailed analysis in the next sections). However, RNAseq reads were mapped to each of the ndh genes (≥98% identity considering 100% of the read length), thus suggesting that these pseudogenes are transcribed. Moreover the ycf68, orf42 and orf56 pseudogenes also show expression values in RPKM. The ycf68 gene has a frameshift and one stop codon in-frame, whereas orf42 has one stop codon in-frame, and orf56 has a frameshift. The ycf68 gene encodes a functional protein present in grasses, gymnosperm and Nymphaeales, whereas orf42 and orf56 are commonly found in the chloroplast genomes of the other species [54,55]. However, orf42 and orf56 are related to mitochondrial genes [55], showing a considerable sequence similarity, and therefore the expression results observed in both genes should not be fully considered due to non-specific and misleading alignments, and therefore they are not shown in Fig 2. Comparative analysis did not detect copies of the ndh, orf42, orf56 and ycf68 genes in nuclear or mitochondrial scaffolds, suggesting that these are really lost from *U. reniformis*, and thus indicating that these represent real transcripts from a pseudogene template. Interestingly, previous studies indicate that truncated, transcribed molecules may exist in the chloroplast [56], and may support

these findings. However, whether these transcripts encode stable RNAs, or even if they are translated to functional proteins is yet to be established, and their roles remain unknown.

Overall, these results indicate that essential cp genes are expressed. However, whether these pseudogenes are not completely inactivated yet or if they encoding regulatory RNAs, or if the photosystem is impaired or not, due to the loss and pseudogenization of *ndh* complex, and if some sort of *ndh* functional replacement could or is occurring remains unknown.

Identification and prediction of RNA editing sites

RNA editing is a post-transcriptional modification that normally changes a cytosine (C) to a uracil (U) or U to C nucleotides, producing transcripts that are different from their DNA template [57]. These modifications can alter the amino-acid sequence of protein, and can also introduce new start and/or stop codons [57]. At least 44 potential editing sites were identified using PREP-Cp and RNAseq analysis. These are all in the coding region, whereas six editing sites were exclusively detected by RNAseq data corresponding to potential novel editing sites (Table 3 and Table 4). The editing level from RNAseq data was inferred from the C versus U ratio of the transcripts derived from the respective loci. Among the RNAseq detected sites, three editing sites alters the first and the second nucleotide of a codon, and one editing site alters the third nucleotide of a codon. Except for *rpoA*, all editing sites lead to non-synonymous substitutions. Four sites located in *rpoB*, *rps14*, *petB* and *rps2* are the most frequent editing sites (> 81%), whereas the remaining sites located in *atpA*, *psaB* and *rpoA* were edited at low frequency (11 to 20%) (Table 3). Moreover, the PREP-Cp predicted 37 additional sites, whereas one site located on the *rpoB* gene was confirmed by RNAseq data (Table 3). All predicted sites by PREP-Cp lead to non-synonymous substitutions. Overall, except for the six editing site detected by RNAseq, the RNA editing sites are quite similar to those observed in other angiosperms [58]. Therefore, these results suggest that the prediction tools may fail to identify authentic RNA editing sites, and that RNAseq data can be used preferentially.

Table 3. RNA editing pattern in *Utricularia reniformis* cp genome identified by RNAseq reads alignment.

Gene	Genome (cp)		Codon		Codon position	Amino acid		Editing level
	Position	Strand	from	to		from	to	U/C (%)
atpA	129,471	+	CAC	UAC	1	H	Y	15
psaB	102,612	-	CCU	UCU	1	P	S	20
petB	68,312	-	CCA	UCA	1	P	S	99
rpoA *	65,985	+	AAC	AAU	3	N	N	11
rpoB †	115,658	+	UCU	UUU	2	S	F	81
rps2	124,300	+	ACA	AUA	2	T	I	99
rps14	103,606	+	CCA	UCA	2	P	S	85

* synonymous substitution
† Also predicted by PREP-Cp

Table 4. RNA editing pattern in *Utricularia reniformis* cp genome predicted by PREP-Cp.

Gene	Genome (cp)		Codon		Codon position	Amino acid		Editing Score
	Position	Strand	from	to		from	to	(PREP-Cp)
accD	86,311	-	ACG	AUG	2	T	M	1.0
	86,391	-	CGG	UGG	1	R	W	1.0
	86,562	-	CCU	UCU	1	P	S	1.0
	87,043	-	UCG	UUG	2	S	L	0.8
atpI	125,155	+	UUT	CUU	1	P	L	1.0
	125,752	+	UCA	UUA	2	S	L	1.0
ccsA	35,170	-	GCC	GUV	2	A	V	1.0
matK	136,847	+	UCT	UAU	3	H	Y	1.0
	137,456	+	CAU	UAU	1	H	Y	1.0
	137,579	+	CCG	UCG	1	P	S	0.86
petB	68,692	-	CGG	UGG	1	R	W	1.0
	68,885	-	CCA	CUA	2	P	L	1.0
psaI	84,202	-	CCU	UCU	1	P	S	1.0
psbB	70,971	-	CGU	UGU	1	R	C	1.0
	71,044	-	GCG	GUG	2	A	V	0.86
rpl2 ‡	1,364 60,409	- +	GCG	GUG	2	A	V	0.86
rpl20	75,977	+	UCA	UUA	2	S	L	0.86
rpoA	65,590	+	UCA	UUA	2	S	L	1.0
rpoB	113,701	+	CUU	UUU	1	L	F	1.0
	113,996	+	UCU	UUU	2	S	F	1.0
	114,131	+	UCA	UUA	2	S	L	0.86
	114,209	+	UCA	UUA	2	S	L	1.0
	114,575	+	ACG	AUG	2	T	M	0.86
	115,219	+	CUC	UUC	1	L	F	1.0

†	115,657	+	UCU	UUU	2	S	F	1.0
	116,084	+	UCA	UUA	2	S	L	0.86
rpoC1	119,318	+	CGU	UUA	1	R	C	0.86
rpoC2	120,709	+	CUU	UUU	1	L	F	0.86
	121,870	+	CUU	UUU	1	L	F	1.0
	122,038	+	CUU	UUU	1	L	F	1.0
	122,170	+	CGG	UGG	1	R	W	1.0
	123,590	+	UCA	UUA	2	S	L	0.86
rps14	103,537	+	UCA	UUA	2	S	L	1.0
	103,606	+	CCA	CUA	2	P	L	1.0
rps16	134,633	+	UCA	UUA	2	S	L	1.0
rps2	124,189	+	ACC	AUC	2	T	I	1.0
	124,414	+	UCA	UUA	2	S	L	1.0

‡ The *rpl2* is duplicated in the IRs
† Also detected by RNAseq read mapping

Several plastids transcripts require C to U editing, whereas *ndh* genes contain about 50% of the editing sites of angiosperm plastid transcripts [18]. However, the RNA edition was also observed in pseudogenes, such as the *ndhB* [58]. Interestingly none of the editing sites were associated with transcripts that align with *ndh* genes loci, supporting the idea that despite the fact that RNAseq reads align to these genes they may be indeed correspond to non-functional genes.

Microsatellite and other repeats

It was previously described that Lentibulariaceae plastomes carry a large number of chloroplast microsatellite (cpSSR) and small number of repeats longer than 60 bp [15]. The *U. reniformis* plastome follows this trend, having at least 331 cpSSR ranging from 7 to 179 bp (Fig 3). Among those, mono and di repeats were the most common, representing 86% (284 cpSSRs) and 4% (13 cpSSRs), respectively. No pentanucleotides or hexanucleotides repeats were found, and low frequencies of tri-, and tetra repeats were observed (Fig 3). Among the 284 mononucleotide repeats, only 16 C/G type repeats were found, with all other repeats belonging to the A/T type. Repeat number of mononucleotide motifs ranged from seven (48%) to 15. Furthermore, at least 55 cpSSRs mononucleotide repeats with a length of at least 10 bp were detected. It is noteworthy that the AT-rich di repeats are commonly found in others carnivorous plant plastomes [15]. In general, these

results are also quite similar to those observed in *Tanaecium tetragonolobum*, from the family Bignoniaceae, where 347 cpSSR were identified [59]. The distribution of the mononucleotide, di-repeats, tri-repeat and compound/polymorphic cpSSRs are shown in Table 5, and indicate a similar distribution of cpSSRs to that present in the coding regions of *U. gibba*, *U. macrorhiza, G. margaretae* and *P. ehlersiae*. However, the majority of the cpSSR are located in non-coding/intergenic regions, accounting for up to 197 (60%) occurrences, whereas 38 were polymorphic, representing good regions for the development of cpSSRs molecular markers for population studies and to estimate the relationship between different Lentibulariaceae.

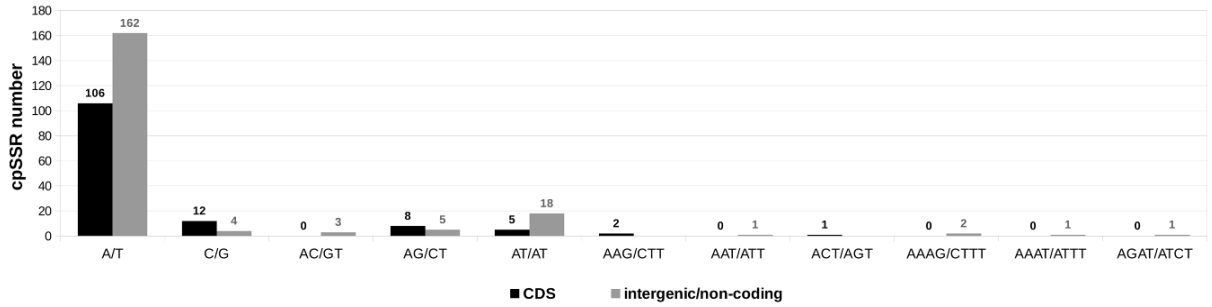


Figure 3. Frequency of SSR motifs found in *Utricularia reniformis* cp coding and intergenic regions, taking into account sequence complementarities.

Table 5. Distribution of the cpSSRs present in the U. reniformis cp coding regions.

cpSSR type	Genes	Pseudogenes
mono-repeats	*accD*, *atpB*, *atpF*, *ccsA*, *clpP*, *matK*, *petA*,*psaA*, *psaI*, *psbB*, *psbC*, *psbF*, *psbK*, *psbT*, *rpl14*, *rpl20*, *rpl2*, *rpl22*, *rpoA*, *rpoB*, *rpoC1*, *rpoC2*, *rps11*, *rps3*, *rps8*, *ycf1*, and *ycf2*	*ycf68*, *orf56*, *ndhA*, *ndhB*, *ndhD*,*ndhI* and *ndhG*
di-repeats	*petA* (AT), *psaA* (TC), *rpoB* (TA), *rpoC2* (AT), and *ycf2* (TA)	*ndhB* (AG)
tri-repeat (TTC)	*psbC*	
compound/polymorphic	*ccsA*, *psaA*, *rpl32*, *rpoC2*, *rps4*, *rps7*, *rps14*, *rps15*, *rps18*, *rps19*, *ycf1*, and *ycf2*	*ndhA* and *ndhI*

A plethora of forward and palindromic repeats were also identified in the *U. reniformis* cp genome (Table 6). A total of 24 pairs of repeats (30 bp or longer, and up to 58 bp) were identified. These repeats are spread out over the LSC (42%), IRs (50%) and SSC (8%), and no introns were found to contain repeated elements. These repeats are found predominantly in coding regions (58%), which are not commonly found in other angiosperm lineages [51], but are frequent in other Lentibulariaceae [15]. This may indicate that the Lentibulariaceae cp coding regions are repeat hotspots acting as a source of recombination and rearrangements. For instance, two palindromic repeats were identified within the genes *psaC*, *ndhD*Ψ and *ccsA,* suggesting a potential role during the pseudogenization process of *ndhD*. In addition, the region containing *psaC*, *ndhD* and *ccsA* in *U. reniformis*, *U. gibba*, *U. macrorhiza*, *P. ehlersiae* and *G. margaretae* is quite variable in terms of nucleotide and gene composition, and order (see below). Palindromic repeats located on the LSC are identified within the genes *accD*, *rbcL*, *ycf3*, *psaA*, *psaJ*, *rpl2*, *rpl33*, *rps14*, *rps19* and *psaB*, and the majority of repeats located on the IRs are associated with the gene *ycf2*. In summary, the number and distribution of these sequences vary from one species to another. However, comparative analyses indicate that this repeat repertoire is quite similar to those previously observed in the terrestrial forms *G. margaretae* and *P. ehlersieae.*

Table 6. Sequence repeats in the cp genome of *Utricularia reniformis.* Type, location and size (in bp) of repeated elements (IGS, Intergenic spacer).

Type	Location	Size (bp)
Palindrome	IGS: *ndhD*Ψ-*ccsA*	58
Palindrome	IGS: *accD-rbcL*	44
Palindrome	*rps19* | IGS:*trnH*-GUG-*rpl2*	37
Palindrome	IGS: *ycf3-psaA*	36
Palindrome	*trnS*-GGA-*trnS*-GCU	30
Forward	*psaA-psaB*	35
Forward	IGS: *rps12*_end – *trnV*-GAC	31
Palindrome	IGS: *rps12*_end – *trnV*-GAC | IGS: *trnV*-GAC-*rps7*	31
Palindrome	IGS: *rps12*_end – *trnV*-GAC | IGS: *trnV*-GAC-*rps7*	31
Forward	IGS: *trnV*-GAC-*rps7*	31
Palindrome	IGS: *rpl33-psaJ*	33
Palindrome	IGS: *rps14 – trn(f)M*-CAU	30
Forward	*trnS*-UGA-*trnS*-GCU	32
Forward	*ycf2*	31
Palindrome	*ycf2-ycf2*	31
Palindrome	*ycf2-ycf2*	31

Forward	*ycf2*	31
Forward	*ycf2*	30
Palindrome	*ycf2*-*ycf2*	30
Palindrome	*ycf2*-*ycf2*	30
Palindrome	IGS:*psaC*-*ndhD*Ψ	30
Forward	*ycf2*	30
Palindrome	*trnS*-GGA | IGS: *trnS*-UGA – *psbC*	30
Forward	*trnG*-GCC-*trnG*-UCC	30

Comparative analysis among other Lentibulariaceae

Terrestrial and aquatic forms show distinct SSC organization

The plastomes of the sequenced Lentibulariaceae are highly conserved in terms of gene synteny. Fig 4 shows comparisons of the SSC region and the IR-LSC and IR-SSC boundaries. In spite of the high level of synteny noted on the LSC/IRb and IRa/LSC junctions, the IRb/SSC, SSC/IRa boundaries and the SSC region itself show distinct organization. Gene deletions, rearrangements, expansions and contractions in the SSC and IR/SSC boundaries of these plants are markedly noted.In the SSC, only the *rpl*32, *trn*L, *css*A, *psa*C and rpl15 genes are conserved among the analyzed plants. Moreover, a distinct repertoire for the ndh gene complex is observed in each plastome. Indeed, previous studies have indicated that several mutational hotspots were found in the entire SSC as well as the region around the *ndh* genes of *U. macrorhiza*, *G. margaretae* and *P. ehlersiae [15]*. This may be one of the evolutionary mechanisms related to the ndh and *ycf1* pseudogenization process.

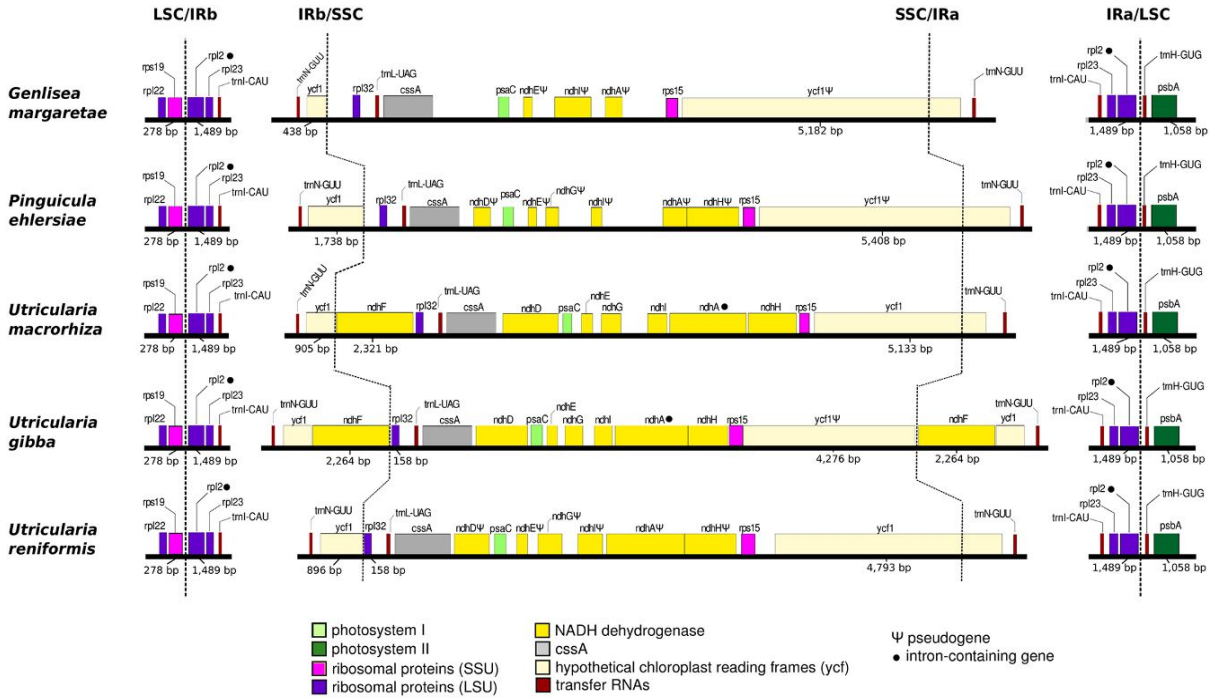


Figure 4. Comparison of the boundaries of the LSC, SSC, and IR regions in the currently available cp genomes of Lentibulariaceae.

The *ndhF* gene is exclusively present in aquatic species, and only *U. gibba* carries two copies of the *ndh*F gene, delimiting the IR/SSC boundaries. It is worth noting that the *ndhF* genes, which are commonly located on the IR/SSC boundaries in other angiosperms plastomes, are lost in the other Lentibulariaceae terrestrial taxa. The loss of the *ndh*F is not an exclusive feature of the terrestrial forms, since this gene can be either present or absent in angiosperms [51,60]. However *ndhF* is often found in Coniferophyta, Filicophyta, Ginkgophyta, Gnetophyta, Lycophyta, Psilophyta and Sphenophyta plastomes [61,62]. Interestingly, the *ndhF* loss may be related to shifts in the position of the junction of the IR and SSC regions in Orchidaceae [62]. Indeed these shifts may lead to losses due to recombination, as observed in the Lentibulariaceae (Fig 4).

For all the other species, two copies of the *ycf*1 gene, one larger and other smaller, delimits the IR/SSC boundaries. Interestingly, assuming that there have been no major errors in genome assembly, the *ycf*1 gene shows different sizes among the analyzed plants, indicating that together

with *ndh*F, these genes may be used as a potential hotspot for the study of the evolution of the IR/SSC junction in the Lentibulariaceae. For instance, in *G. margaretae*, *P. ehlersiae* and *U. gibba*, the larger copy of *ycf1* is a pseudogene, whereas all smaller copies are intact genes. In addition, *ycf1* is often associated with many rearrangements in other angiosperms [63], and this is indeed observed within the Lentibulariaceae plastomes. Therefore, as observed in other land plants, during the course of evolution, the Lentibulariaceae plastomes displayed rearrangements, deletions and gene loss. Overall, these finding also suggests a correlation of the plant life style with plastome genomic structure. For instance, whereas aquatic forms of Lentibulariaceae tend to maintain larger SSC regions, retaining the *ndh* complex genes intact, the terrestrial forms have smaller SSC regions and have lost *ndh* genes.

The Lentibulariaceae plastome gene repertoire varies mainly in the *ndh* genes

The gene repertoires of the plastome of the sequenced carnivorous plants (*U. gibba*, *U. macrorhiza*, *G. margaretae*, and *P. ehlersiae*) are quite similar (Fig 5). The Lentibulariaceae cp core gene repertoire is composed of 69 genes, mostly involved in photosynthesis, energy metabolism, and other housekeeping functions (Fig 5, center). Indeed, this result is similar to the core cp genome of essentially all photosynthetic organisms [52]. The main differences are related to a combination of losses and pseudogenization of *ndh* genes among the five plastomes. It is worth noting that the *ndh* gene loss/pseudogenization observed in terrestrial forms is clearly derived from independent events (Table 7), which is corroborated by previous studies [15]. In *Genlisea*, the genes *ndh*C, D, F, G, H, J, and K were lost from the plastome, and the genes *ndh*A, B, E, and I are truncated, whereas the *ndh*A, D, E, G, H, I, J, and K retain insertion/deletions and frame shifts in *Pinguicula* [15]. However, *U. reniformis* shows a different pattern, in that the genes *ndh*C, F, J and K were lost from the plastome, and the genes *ndh*A, B, DE, G, H and I reside as truncated pseudogenes. Previous studies indicate that the ndh gene loss in *Genlisea* and *Pinguicula* occurred two times independently within Lentibulariaceae [15]. Interestingly, considering only carnivorous

plants, the *ndh* gene complex spans to up to 16kb in aquatic forms, representing about of 10% of the plastome, whereas terrestrial forms vary from 5kb to 10kb (3-7% - Table 7).

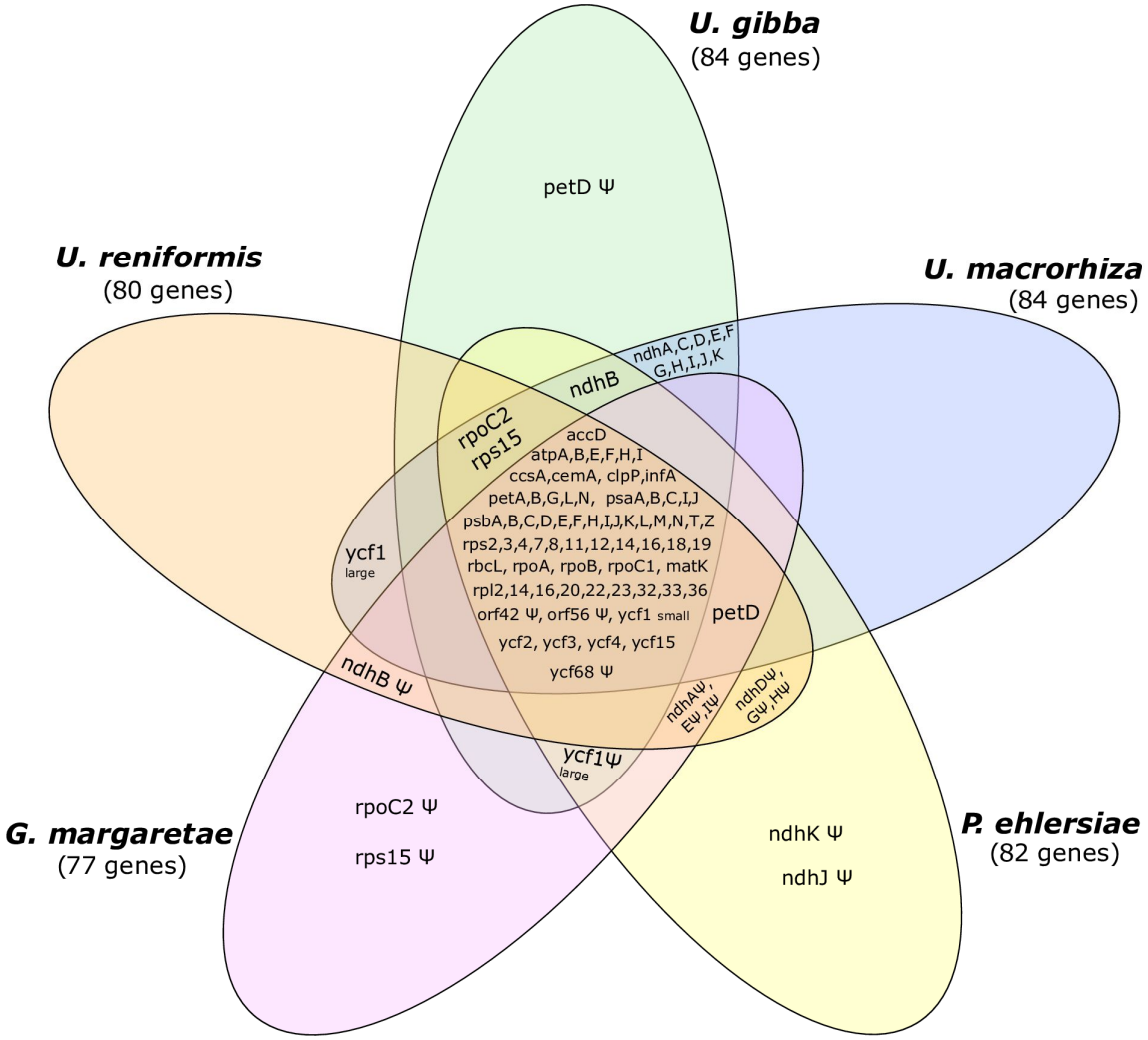


Figure 5. Venn diagram showing the full complement of genes present in the sequenced Lentibulariaceae cp genomes (pan genome). The tRNAs and rRNAs are not included. The numbers below each species represent the unique genes used in the comparison.

Table 7. Distribution and comparisons of the eleven genes in the ndh complex, encoded in aquatic and terrestrial cp genomes from carnivorous plants and the ancestral lineages Sesamum indicum, *Tanaecium tetragonolobum* and *Andrographis paniculata*. Black box represents that the given ndh gene is present and intact. Gray box indicates truncated *ndh* gene according to the legend. White box indicate that the given ndh gene is absent in the cp genome.

	Sesamum indicum **(terrestrial)**	***Tanaecium tetragonolobum*****(terrestrial)**	***Andrographis paniculata*****(terrestrial)**	***Pinguicula ehlersiae*****(terrestrial)**	***Genlisea margaretae*****(terrestrial)**	***Utricularia reniformis*****(terrestrial)**	***Utricularia macrorhiza*****(aquatic)**	***Utricularia gibba*****(aquatic)**
ndhA **•SSC**	2,171bp	2,162bp	2,069bp	1,811bp ■ □▲	197bp ▲	1,640bp ■ □	2,190bp	2,153bp
ndhB **• *IR**	2,211bp	2,211bp	2,21bp	2.211bp	2,121bp ■ □	1,080bp ■ ▲	2,211bp	2,211bp
ndhC**LSC**	362bp	362bp	362bp				362bp	362bp
ndhD**SSC**	1,502bp	1,148bp ▲	1,454bp	622bp □ ▲		737bp ■ ▲	1,526bp	1,526bp
ndhE**SSC**	305bp	305bp	305bp	309bp □	188bp ■ ▲	233bp ■ ▲	305bp	305bp
ndhF**SSC**	2,255bp	2,231bp	2,23bp				2,261bp	2,261bp (*)
ndhG**SSC**	530bp	530bp	530bp	520bp □		509bp ■	530bp	530bp
ndhH**SSC**	1,18bp	1,181bp	1,18bp	1,131bp □		1,085bp ■ ▲	1,187bp	1,181bp
ndhI**SSC**	506bp	506bp	506bp	514bp □	469bp □ ▲	520bp □	530bp	524bp
ndhJ**LSC**	476bp	476bp	476bp	434bp ■ ▲			476bp	476bp
ndhK**LSC**	701bp	677bp	677bp	410bp ▲			644bp	677bp
	14,411bp (9.4%)	14,000bp (9.1%)	14,213bp (9.4%)	10,173bp (6.9%)	5,096bp (3.6%)	6,884bp (4.9%)	14,433bp (9.4%)	16,678bp (10.3%)

• Intron-containing gene
■ Stop codon in-frame
□ Frameshift
▲Missing fragment of coding region (incomplete gene)
* *two gene copies*

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

Phylogenomic and phylogenetic analysis

The loss and recovery of the *ndh* gene complex among the Lentibulariaceae

Moreover, adding the *U. reniformis* plastome to more complete phylogenetic analyses, and tracing the possible evolutionary hypotheses according to the presence (functional gene) or absence (truncated or gene in frame-shift) of *ndh* genes, a more comprehensible evolutionary scenario can be determined. The phylogenetic history resulting from the cp phylogenomic analyses (for further discussion see below), indicates that Lentibulariaceae is a monophyletic family with the Pinguicula clade as a sister group of the *Genlisea-Utricularia* clade; the topology is also corroborated by previous studies [64,65]. By tracing the evolution of ndh genes in the trees, it was determined that the presence of functional *ndh*A, *ndh*C, *ndh*E, *ndh*F, *ndh*G, *ndh*H, *ndh*I, *ndh*J, and *ndh*K is plesiomorphic, but were lost by pseudogenization or frame-shift in the ancestors of the Lentibulariaceae clade (Fig 6A). However, those ndh genes were recovered in a reversion process for the aquatic group, formed by *Utricularia gibba* and *U. macrorhiza*. A similar situation may have occurred with the *ndh*B gene, but this gene is functional for *Pinguicula* (Fig 6B). Even assuming the presence of this gene in the ancestral lineage, two possibilities can be explored by optimizing the transformations in the tree, and having both accepted with the parsimony approach (ACCTRAN and DELTRAN - [66]) since both hypotheses assume the same number of transformations (two in this case). Therefore, it is possible to accept the alternative scenario of the loss of ndhB gene from the ancestors of the *Genlisea-Utricularia* clade, or independent (parallel) losses for *Genlisea* and *Utricularia reniformis* (Fig 6B). For the same reasons, it is necessary to assume two alternative histories for the *ndh*D gene as well. The *ndh*D is functional for the aquatic species, *Utricularia gibba* and *U. macrorhiza*. It exists as a pseudogene for *Pinguicula* and *U. reniformis*, but as a frame-shift for *Genlisea* (Table 7). After surveying all Lamiales taxa represented in the phylogenomic analyses (Fig 6), we found that most species present a functional version of *ndh*D (*Tanaecium tetragonolobum* as an exception), thus it is reasonable to suppose this

state as plesiomorphic. Assuming the topology as presented in Fig 6C, the *ndh*D gene can be lost independently more than once. In the present scenario, *ndh*D was lost for the Tanaecium and Lentibulariaceae clades, and afterwards reverted in the ancestors of the aquatic species *Utricularia gibba* and *U. macrorhiza* (Fig 6C), similar to the other *ndh* genes (Fig 6A and Fig 6B). There is also the possibility that *ndh*D was lost in the ancestors of the *Tanaecium-Andrographis-Pinguicula-Genlisea-Utricularia* clade, and reverted twice for the *Andrographis* and *U. gibba* - *U. macrorhiza* clades. Both of these hypotheses for the *ndh*D gene assume three events, thus both are plausible.

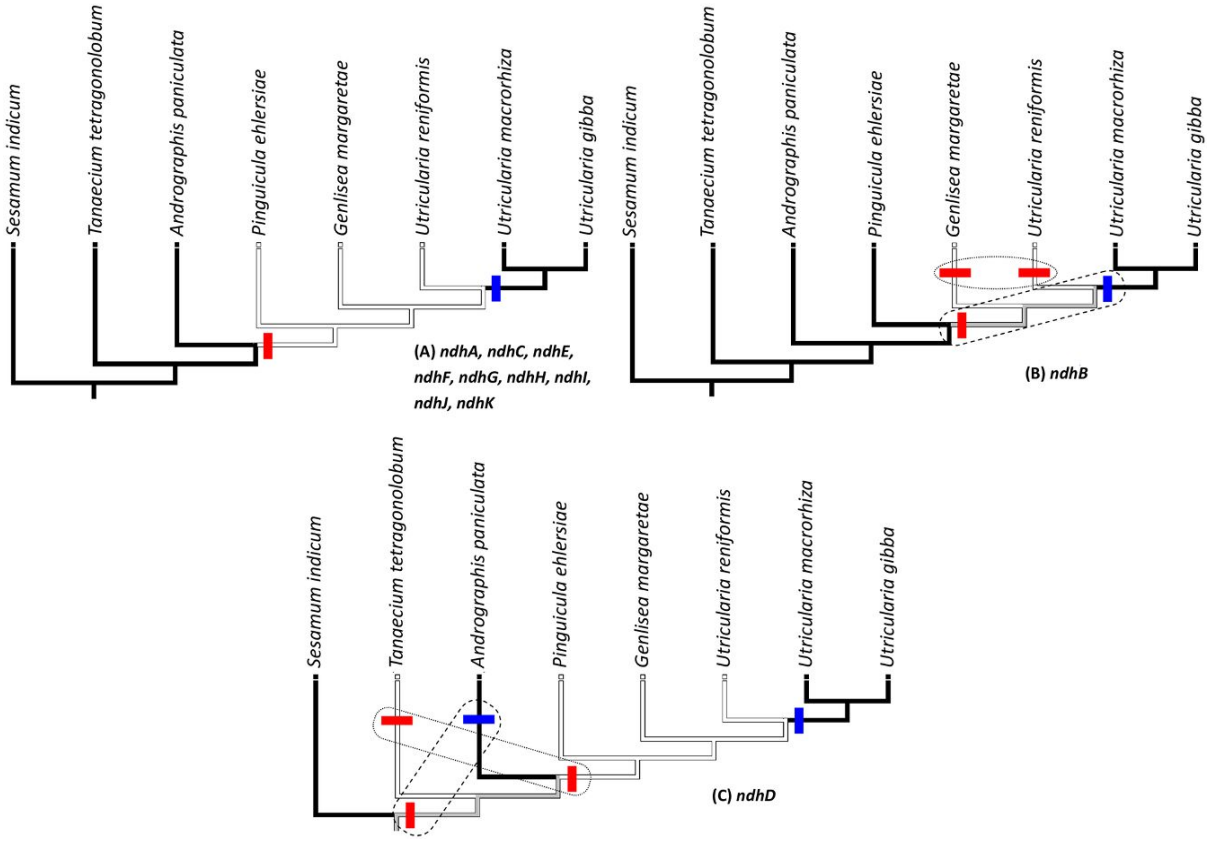


Figure 6. Phylogenetic history of *ndh* genes. (A) represents the only scenario for the *ndh*A, *ndh*C, *ndh*E, *ndh*F, *ndh*G, *ndh*H, *ndh*I, *ndh*J, and *ndh*K genes evolution. **(B)** represents the phylogenetic history and the two possible scenarios for the *ndh*B gene and **(C)** for the *ndh*D gene (blue bars indicate the arising of the functional genes and red bars their loss).

Relationships with the Lamiales order

The relationships between the families in the order Lamiales are only partially resolved [67]. Despite the attempts based on several plastomes and mitochondrial genes [67,68] to identify the sister family of Lentibulariaceae, this issue still remains unclear. Therefore, phylogenomic analyses based on whole plastomes might contribute significantly to the elucidation of the systematic relationships inside of this order. With the aim to identify the phylogenetic position of *U. reniformis* and Lentibulariaceae among the other families of Lamiales, plastomes from 23 different taxa were compared. The phylogenomic analyses were constructed using 47 genes from the Lamiales plastomes. All Lamiales with less than 47 genes were excluded from our analysis, including parasitic plant cp genomes (truncated/pseudogenes and tRNAs were not considered). Both maximum likelihood (ML) and Bayesian analyses recovered the same tree topology with high support values (Fig 7), and agree with a previous study based on the rapidly evolving genes *trnK*/*matK*, *trn*L-F and *rps16* [68]. Our results indicate that Lentibulariaceae is monophyletic (100 maximum likelihood bootstrap – BML; 100 Bayesian posterior probability - PP) and *Pinguicula* is a sister clade of the *Genlisea*-*Utricularia* clade (100 BML; 100 PP), corroborating previous studies [65,69]. In addition, that *U. reniformis* is a sister to the *U. gibba*-*U. macrorhiza* clade, is also well supported (100 BML; 100 PP). According to our results based on *matK*, and also corroborated by previous studies [64,65], the ancestral lineage of Lentibulariaceae was possibly terrestrial, with life forms adapted for this environment developed by most species of *Pinguicula* and *Genlisea* (Fig 8). The alternative life forms present in *Utricularia* species [1,70] are thus derived from the ancestral terrestrial state, representing the occupation of different environments and the consequent diversification of body forms in an adaptive response to several ecological niches. Very specialized alternative life forms were developed by *Utricularia* lineages, for instance the rare reophytes, which inhabit rapids, cascades and streams at flood level during torrential conditions. For *Utricularia* this life form is represented only by the four species *U. neottioides*, *U. oliveriana*, *U. rigida*, and *U. tetraloba* [1], which have at least two independent phylogenetic origins (Fig 8, black circles 4 and

4'). Most aquatic species are represented by *Utricularia* sect. *Utricularia* (Taylor 1989), despite this life form having arisen at least twice (Fig 8, blue circles 2 and 2'), with species found in both the Northern and Southern Hemispheres (including *U. gibba* and *U. macrorhiza*). *Utricularia reniformis* is nested within a very specialized clade (Fig 8, green circle 3), with species adapted for terrestrial (the case of *U. reniformis*) and also epiphytic life forms, represented by species from sections *Iperua* and *Orchidioides*. The sister species of *U. reniformis* is *U. nelumbifolia* ([69]; this study), a rare endemic Brazilian *Utricularia*, very similar morphologically to *U. reniformis*, but which lives among the leaves of the giant bromeliads found on inselbergs [71]. These results also support positioning the Lentibulariaceae close to the Acanthaceae, Bignoniaceae, Pedaliaceae, Orobanchaceae, and Lamiaceae (Fig 7), composing the higher core of the Lamiales group previously proposed [68].

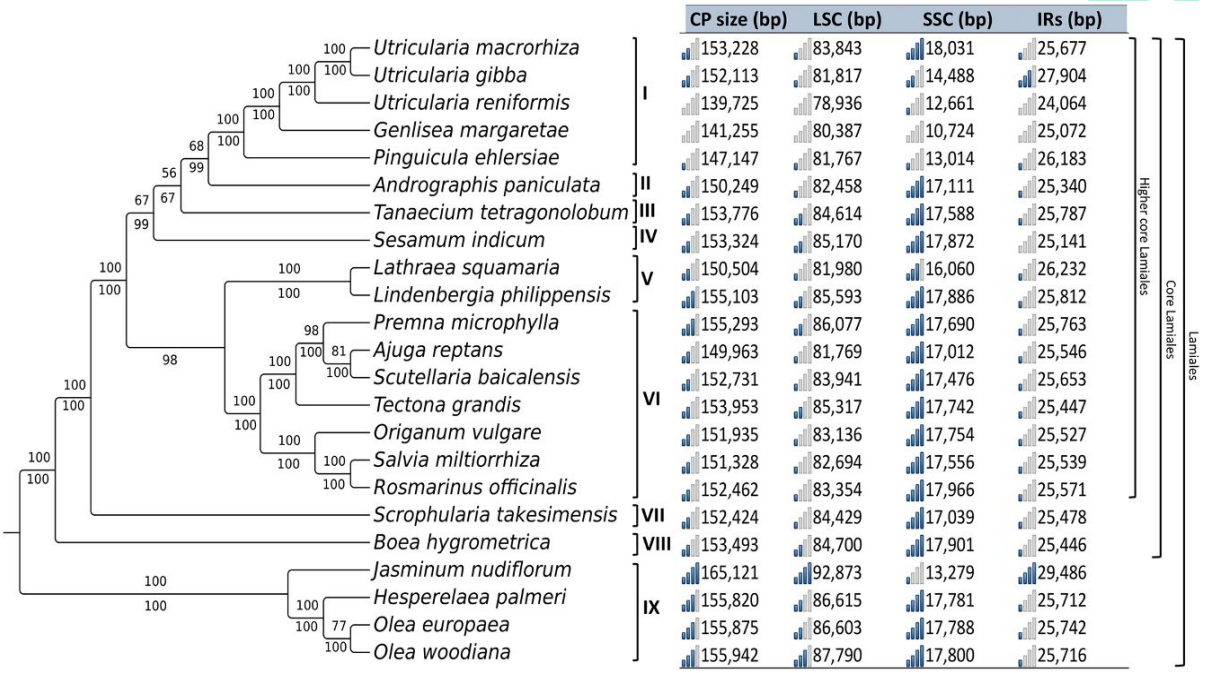


Figure 7. Phylogenomic analysis based on 23 complete Lamiales cp genomes. The phylogenetic position of *Utricularia* reniformis was inferred by maximum likelihood and Bayesian analyses. Numbers above and below the lines indicate the maximum likelihood bootstrap values and Bayesian posterior probabilities values > 50% for each clade, respectively. The table on the right indicates the genome features in base pairs (chloroplast genome length, LSC, SSC and IRs regions). The histograms located to the left of each feature (CP size, LSC, SSC and IRs), graphically illustrate the size distribution for each feature. Sub-titles: **I** Lentibulariaceae; **II** Acanthaceae; **III** Bignoniaceae; **IV** Pedaliaceae; **V** Orobanchaceae ; **VI** Lamiaceae; **VII** Scrophulariaceae; **VIII** Gesneriaceae, and **IX** Oleaceae.

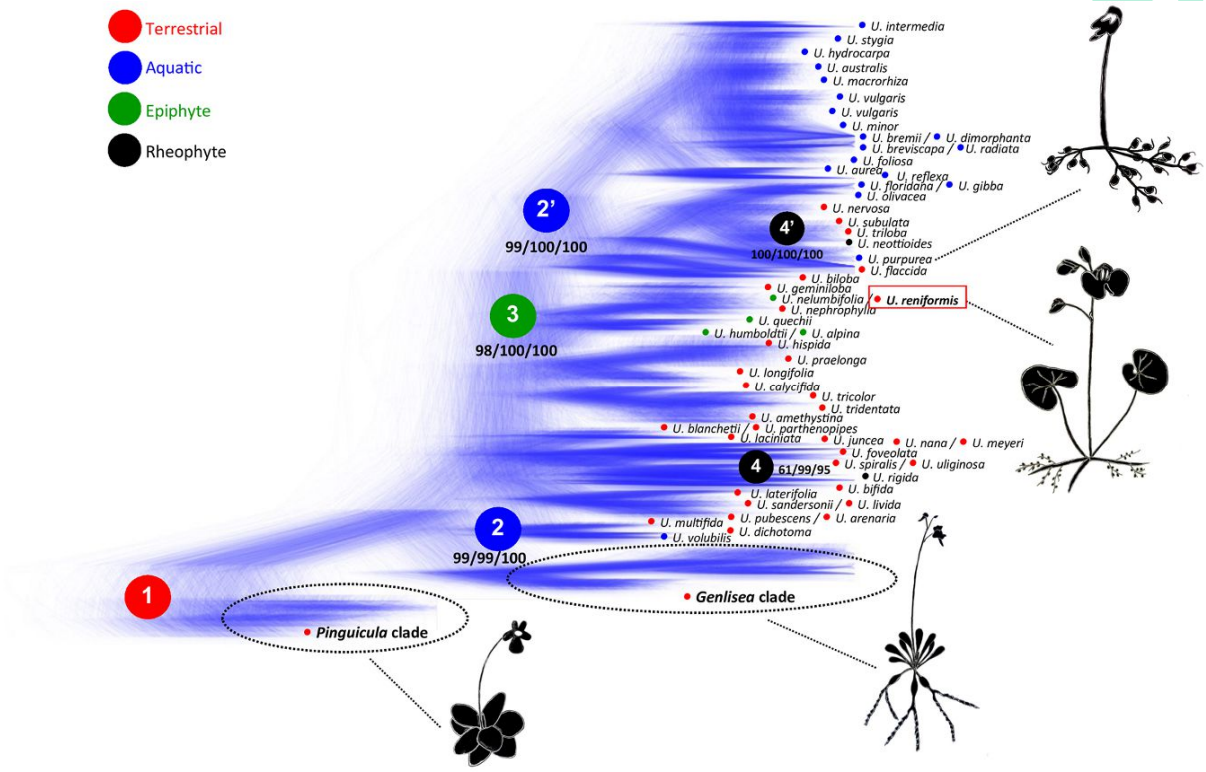


Figure 8. Cloudgram (Bayesian inference) of Lentibulariaceae from 18,000 Bayesian trees based on *mat*K cp gene. The red circle 1 indicates the terrestrial ancestral lineage for the Lentibulariaceae family. The blue circles, 2 and 2', represent the independent radiations to the aquatic habitat of *Utricularia* lineages. The green circle 3 indicates the possible ancestor of the epiphytic species from the plesiomorphic terrestrial life form, and the black circles, 4 and 4', represent the independent origins for the rare reophytic life form for *Utricularia* lineages. Numbers below clades represent the support (maximum parsimony bootstrap/ maximum likelihood bootstrap/ posterior probabilities based on Bayesian inference). Higher color densities represent higher levels of certainty represented by congruent trees (from the 18,000 trees).

In general, the order Lamiales maintains the quadripartite structure, except for the parasitic family of Orobanchaceae that was not included in our analysis due to the small genome size (45kb to 120kb) and number of genes (21 to 42) [22]. In general the SSC and IR are 17 and 25kb long, respectively and only a few differences in genome size are observed (Fig 7). The biggest cp genome belongs to the family Oleaceae. The *Jasminum nudiflorum* cp genome is 165 kb with an IR and LSC expansion followed by a contraction in the SSC. This is also observed in *Schwalbea americana* from the Orobanchaceae family, which is 160kb long. In addition, SSC contraction and IR expansion were observed in the *Lathraea squamaria* from the Orobanchaceae family. However, major differences were noted in the Lentibulariaceae; SSC contraction was mainly observed in *Utricularia-Genlisea* and *Pinguicula* (Fig 7). This suggests that this family is under different selective pressures, resulting in dynamic plastome structures. It is noteworthy that only the Lentibulariaceae are carnivorous, suggesting their carnivorous syndrome may impact metabolism and photosynthesis.

Discussion

The *U. reniformis* plastome contribution to the study of the evolution of terrestrial carnivorous plants from the Lentibulariaceae family

We sequenced the cp genome of *Utricularia reniformis* and compared it against other carnivorous plants from the Lentibulariaceae family. This study revealed that the 139kbp cp genome of *U. reniformis* is quite similar to the cp genome of *U. gibba*, *U. macrorhiza*, *G. aurea*, and *P. ehlersiae* in terms of gene synteny, repeats and cpSSRs content; whereas the main differences are located on the SSC region and the *ndh* genes repertoire (Fig 3 and Fig 5 and Tables 1,2,5 and 6). In spite of the similarity of the gene repertoire of the *U. reniformis* cp genome to essentially all photosynthetic organisms, comparative genomics analysis corroborated previous studies [15], which show that whereas aquatic forms maintain the complete *ndh* gene complex

composed of eleven genes, the terrestrial forms have shown a number of losses of the *ndh* genes, and these losses are exclusive for each species (Table 7). In addition, the proposed phylogenetic history of *ndh* genes shown in Fig 6, suggests that independent *ndh* losses occurred during the course of the evolution of the genera; whether other terrestrial *Utricularia*, *Genlisea*, and *Pinguicula* species have also lost the ndh gene set, and whether the *ndh* pseudogenes found in the terrestrial forms were lost recently, remains to be investigated. Indeed, this is an important question to be explored in future work. Moreover, phylogenomic analysis supports that the family Lentibulariaceae is monophyletic, belonging to the higher core of the Lamiales clade, and thus corroborating the hypothesis that the first *Utricularia* lineage emerged in terrestrial habitats and then evolved to epiphytic and aquatic forms, as shown by the Fig 8.

Furthermore, the transcriptome analysis by RNAseq approach indicate that mostly cp genes are transcribed (Fig 2, and S3 Table), whereas even the truncated *ndh* genes, *orf*42, *orf*56 and *ycf*68 show some level of transcription. In addition to the previous observation that truncated transcribed molecules may exist in the chloroplast [56], this finding supports that the pervasive transcription, which is commonly found in bacterial genomes, may also occur in cp genomes, thus suggesting that these transcripts have an important role in gene regulation and genome evolution, as previous discussed elsewhere [72]. However, further studies are necessary to uncover the potential role of these transcripts. In addition, this study also shed some light on the RNA editing in cp genomes, with novel editing site being uncovered (Table 3), suggesting that the methodology used in this study represent a powerful tool to identify novel RNA editing sites.

Endosymbiotic cp gene transfer to the nuclear and mitochondrial genome of *U. reniformis*

It was well known that during the course of evolution cp genes can be transferred to the nucleus, and their protein products can be reimported into the organelle lumen by the action of a

transit peptide [20]. This indeed is a very widespread phenomenon in nature [73]. In addition, gene transfers from organelles often lead to functional replacement of host genes in a process called endosymbiotic gene replacement [21]. For instance, a large chunk of endosymbiotic cp genome was observed on chromosome 10 of rice, which contains a recent 33 kb insertion of cp DNA in addition to a 131 kb insertion representing nearly the entire plastid genome [74]. Furthermore, endosymbiotic gene transfer is also observed in smaller plant genomes, such as *A. thaliana* [75]. In our ongoing analysis of the *U. reniformis* genome we have noticed the presence of an endosymbiotic gene transfer of truncated cp genes to the nuclear and mitochondrial genomes. For instance, a total of 26kbp of cp-derived sequences were assembled in 28 contigs and mapped to mt and nuclear DNA assembled scaffolds (S2 Table). In addition to these contig-derived regions, during the ongoing assembly of the *U. reniformis* mt genome, we have found a truncated copy of the *ndh*J, *ndh*K and *ndh*C genes (data not shown). Interestingly the *ndh*J, *ndh*K and *ndh*C genes are absent from the *U. reniformis* cp genome (shown in details in the Fig 1 and Table 1). Indeed, this suggests that an ancient lineage of *U. reniformis* had these genes in their cp genome, which were subsequently transferred to the mt genome by an endosymbiotic event. However, due to evolutionary pressures, yet to be established, the *ndh*J, *ndh*K and *ndh*C genes were decayed from the cp copies, and remained as relics in the mt genome. These observations also suggest that during the course of the evolution of the ndh complex in U. reniformis, endosymbiotic gene replacement events from the mt *ndh*J, *ndh*K and *ndh*C copies, may have occurred. Further investigation is needed based on the sequencing of new species, and the presence and absence of the of *ndh* genes of others carnivorous plant cp and mt genomes. Therefore, the endosymbiotic gene transfer events are shaping the *U. reniformis* nuclear and mitochondrial genomes.

A detailed analysis of the *U. reniformis* nuclear and mt genomes, including functional annotation and comparative genomics showing the endosymbiotic transfer eventsbetween the organelles and the nuclear DNA are in progress.

The impact of the carnivorous syndrome in the *ndh* complex genes evolution

It was previously observed that lineages that have lost photosynthetic function, trend toward reduced cp genome size [52]. The analysis of the cp genomes of the terrestrial forms, *Utricularia reniformis*, *Genlisea margaretae* and *Pinguicula ehlersiae*, may support this observation. However, *U. reniformis*, *G. margaretae* and *P. ehlersiae* are all photosynthetic organisms that lack the *ndh* genes, which apparently does not affect the fitness of these plants. Indeed, previous studies suggested that the ndh function might be dispensable under favorable growth conditions [19], suggesting that the carnivorous syndrome may act in favor of the functional *ndh* absence. Interestingly, the *ndh* gene loss or pseudogenization is relatively rare among the Viridiplantae clade [17,19]. However, it seems that the ndh genes were not essential during plant evolution, and their loss may also be related to early events leading to parasitic behavior [18]. In addition, the *ndh* genes are related to photosynthetic response to environmental stress, indicating its participation in the transition to terrestrial habitats [18,19]. However the first lineages of the Lentibulariaceae emerged in a terrestrial habitat, and then evolved to aquatic environments, suggesting that the evolutive history of the ndh genes among the Lentibulariaceae followed an opposite direction. For instance, we propose that the plastid ndh genes present in the aquatic forms *U. gibba* and *U. macrorhiza* were recovered in a reversion process, and that the ndh function may be dispensable in terrestrial forms (Fig 6).

In order to explain this genomic trend related to the loss of the *ndh* genes observed in the terrestrial Lentibulariaceae, a hypothesis posits that carnivorous plants are metabolically similar to parasitic plants in that they use organic carbon obtained through their prey, or their host for parasitic plants, to overcome environmental stress [15]. In addition, different and variable levels of nutritional stress to the plant may occur in aquatic, terrestrial, epiphytic and reophytic forms. This hypothesis is quite interesting, since it suggests that nutritional stress, which is a common feature of

the carnivorous plants [2], can impact molecular and biochemistry characteristics shaping the cp genome. Moreover, over an evolutionary time scale these differences can lead to morphological changes, such as the adaptation to aquatic or land environments, and thus supporting the ndh gene repertoire differences observed among these species (Fig 8). However this hypothesis is yet to be established.

Conversely, it has been proposed that under relaxed selection, unequal efficiency of DNA repair, and high levels of mutagenic reactive oxygen species (ROS), the genome architecture of the Lentibulariaceae may also have been shaped in a fashion similar as those observed in the parasitic plants [9,14,15]. Indeed, a previous study has shown a genome-level convergence between carnivorous and parasitic plants [22]. Moreover, the *ndh* loci have accumulated several nucleotide substitutions and repeats [15], which may have resulted in the loss and pseudogenization process observed in *U. reniformis*, *G. margaretae* and *P. ehlersiae*. Indeed the sequence repeats are located mostly in the coding regions, and this is particularly noted with *ndh*DΨ gene in *U. reniformis* (Table 6). However the sequencing of additional terrestrial and aquatic forms is necessary to corroborate the role of the sequence repeats with the pseudogenization process.

Overall, we propose that sequencing of additional cp and nuclear genomes from other individuals and species from the Lentibulariaceae family will shed light on the relationships between the rearrangement and loss of *ndh* genes, life style (aquatic, terrestrial, epiphytic and reophytic) and endosymbiotic gene transfer of cpDNA. Indeed, a recent study has shown that the endosymbiotic gene transfer has also occurred in other carnivorous plants, such as the *Genlisea* nigrocaulis and *G. hispidula* genomes [76], thus suggesting that this may be an evolutionary trend. Whether the *ndh* genes loss in terrestrial forms and the endosymbiotic gene transfer is an evolutionary trend of this group, which is leading to biochemistry and plastome-scale convergence with the parasitic plants, remains as an important question to be answered in the near future.

Acknowledgments

We would like to thank Dr. Cristine G. Menezes for the plant sample preparation, and all the members of the Plant Systematics Laboratory, UNESP-FCAV. We also would like to thank Dr. Lubomír Adamec who kindly provided the *Utricularia* samples used in the cloudogram analysis, and to the anonymous reviewers for their contributions to this manuscript.

References

1. Taylor P. Genus Utricularia: a taxonomic monograph. 2nd edition. Royal Botanic Gardens, Kew; 1989.

2. Juniper B, Robins R, Joel D. The Carnivirorous Plants. Academic Press, London; 1989.

3. Reifenrath K, Theisen I, Schnitzler J, Porembski S, Barthlott W. Trap architecture in carnivorous Utricularia (Lentibulariaceae). Flora - Morphol Distrib Funct Ecol Plants. 2006;201: 597–605. doi:10.1016/j.flora.2005.12.004

4. Fleischmann A, Michael TP, Rivadavia F, Sousa A, Wang W, Temsch EM, et al. Evolution of genome size and chromosome number in the carnivorous plant genus Genlisea (Lentibulariaceae), with a new estimate of the minimum genome size in angiosperms. Ann Bot. 2014;114: 1651–1663. doi:10.1093/aob/mcu189

5. Pellicer J, Fay MF, Leitch IJ. The largest eukaryotic genome of them all? Bot J Linn Soc. 2010;164: 10–15. doi:10.1111/j.1095-8339.2010.01072.x

6. Michael TP. Plant genome size variation: bloating and purging DNA. Brief Funct Genomics. 2014;13: 308–317. doi:10.1093/bfgp/elu005

7. Greilhuber J, Borsch T, Müller K, Worberg A, Porembski S, Barthlott W. Smallest angiosperm genomes found in lentibulariaceae, with chromosomes of bacterial size. Plant Biol Stuttg Ger. 2006;8: 770–777. doi:10.1055/s-2006-924101

8. Kelly LJ, Renny-Byfield S, Pellicer J, Macas J, Novák P, Neumann P, et al. Analysis of the giant genomes of Fritillaria (Liliaceae) indicates that a lack of DNA removal characterizes extreme expansions in genome size. New Phytol. 2015;208: 596–607. doi:10.1111/nph.13471

9. 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; doi:10.1038/nature12132

10. 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. doi:10.3732/ajb.1200080

11. Clivati D, Cordeiro GD, Płachno BJ, de Miranda VFO. Reproductive biology and pollination of Utricularia reniformis A.St.-Hil. (Lentibulariaceae). Plant Biol. 2014;16: 677–682. doi:10.1111/plb.12091

12. Carretero-Paulet L, Librado P, Chang T-H, Ibarra-Laclette E, Herrera-Estrella L, Rozas J, et al. High Gene Family Turnover Rates and Gene Space Adaptation in the Compact Genome of the Carnivorous Plant Utricularia gibba. Mol Biol Evol. 2015;32: 1284–1295. doi:10.1093/molbev/msv020

13. Carretero-Paulet L, Chang T-H, Librado P, Ibarra-Laclette E, Herrera-Estrella L, Rozas J, et al. Genome-wide analysis of adaptive molecular evolution in the carnivorous plant Utricularia gibba. Genome Biol Evol. 2015;7: 444–456. doi:10.1093/gbe/evu288

14. 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. doi:10.1186/1471-2229-11-101

15. Wicke S, Schäferhoff B, dePamphilis CW, Müller KF. Disproportional Plastome-Wide Increase of Substitution Rates and Relaxed Purifying Selection in Genes of Carnivorous Lentibulariaceae. Mol Biol Evol. 2014;31: 529–545. doi:10.1093/molbev/mst261

16. Braukmann TWA, Kuzmina M, Stefanović S. Loss of all plastid ndh genes in Gnetales and conifers: extent and evolutionary significance for the seed plant phylogeny. Curr Genet. 2009;55: 323–337. doi:10.1007/s00294-009-0249-7

17. Peredo EL, King UM, Les DH. The Plastid Genome of Najas flexilis: Adaptation to Submersed Environments Is Accompanied by the Complete Loss of the NDH Complex in an Aquatic Angiosperm. PLoS ONE. 2013;8. doi:10.1371/journal.pone.0068591

18. Martín M, Sabater B. Plastid ndh genes in plant evolution. Plant Physiol Biochem PPB Société Fr Physiol Végétale. 2010;48: 636–645. doi:10.1016/j.plaphy.2010.04.009

19. Ruhlman TA, Chang W-J, Chen JJ, Huang Y-T, Chan M-T, Zhang J, et al. NDH expression marks major transitions in plant evolution and reveals coordinate intracellular gene loss. BMC Plant Biol. 2015;15: 100. doi:10.1186/s12870-015-0484-7

20. Martin W. Gene transfer from organelles to the nucleus: Frequent and in big chunks. Proc Natl Acad Sci. 2003;100: 8612–8614. doi:10.1073/pnas.1633606100

21. Brown JR. Ancient horizontal gene transfer. Nat Rev Genet. 2003;4: 121–132. doi:10.1038/nrg1000

22. Wicke S, Müller KF, de Pamphilis CW, Quandt D, Wickett NJ, Zhang Y, et al. Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. Plant Cell. 2013;25: 3711–3725. doi:10.1105/tpc.113.113373

23. Miranda VFO, Borges RAX, Hering RLO, Monteiro NP, Santos-Filho LAF. Lentibulariaceae. Livro Vermelho da Flora do Brasil. 1st ed. Instituto de Pesquisas Jardim Botânico do Rio de Janeiro; 2013. pp. 614–615. Available: http://dspace.jbrj.gov.br/jspui/handle/doc/26

24. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9: 357–359. doi:10.1038/nmeth.1923

25. 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. doi:10.1089/cmb.2012.0021

26. 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. doi:10.1093/nar/gkt371

27. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. Bioinforma Oxf Engl. 2004;20: 3252–3255. doi:10.1093/bioinformatics/bth352

28. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10: 421. doi:10.1186/1471-2105-10-421

29. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 2004;32: 11–16. doi:10.1093/nar/gkh152

30. 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. doi:10.1093/bioinformatics/btr703

31. Magrane M, Consortium U. UniProt Knowledgebase: a hub of integrated protein data. Database. 2011;2011: bar009-bar009. doi:10.1093/database/bar009

32. Mulder N, Apweiler R. InterPro and InterProScan: tools for protein sequence classification and comparison. Methods Mol Biol Clifton NJ. 2007;396: 59–70.

33. 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. doi:10.1093/nar/gkt289

34. Heberle H, Meirelles GV, da Silva FR, Telles GP, Minghim R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. BMC Bioinformatics. 2015;16: 169. doi:10.1186/s12859-015-0611-3

35. Thiel T, Michalek W, Varshney RK, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). TAG Theor Appl Genet Theor Angew Genet. 2003;106: 411–422. doi:10.1007/s00122-002-1031-0

36. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 2001;29: 4633–4642.

37. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30: 772–780. doi:10.1093/molbev/mst010

38. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. Bioinforma Oxf Engl. 1998;14: 817–818.

39. Akaike H. A new look at the statistical model identification. IEEE Trans Autom Control. 1974;19: 716–723. doi:10.1109/TAC.1974.1100705

40. Burnham KP, Anderson DR. Multimodel Inference Understanding AIC and BIC in Model Selection. Sociol Methods Res. 2004;33: 261–304. doi:10.1177/0049124104268644

41. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinforma Oxf Engl. 2014;30: 1312–1313. doi:10.1093/bioinformatics/btu033

42. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinforma Oxf Engl. 2001;17: 754–755.

43. Stöver BC, Müller KF. TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses. BMC Bioinformatics. 2010;11: 7. doi:10.1186/1471-2105-11-7

44. Bouckaert RR. DensiTree: making sense of sets of phylogenetic trees. Bioinforma Oxf Engl. 2010;26: 1372–1373. doi:10.1093/bioinformatics/btq110

45. CBOL Plant Working Group. A DNA barcode for land plants. Proc Natl Acad Sci U S A. 2009;106: 12794–12797. doi:10.1073/pnas.0905845106

46. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinforma Oxf Engl. 2011;27: 863–864. doi:10.1093/bioinformatics/btr026

47. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14: R36. doi:10.1186/gb-2013-14-4-r36

48. Mower JP. PREP-Mt: predictive RNA editor for plant mitochondrial genes. BMC Bioinformatics. 2005;6: 96. doi:10.1186/1471-2105-6-96

49. Curci PL, De Paola D, Danzi D, Vendramin GG, Sonnante G. Complete Chloroplast Genome of the Multifunctional Crop Globe Artichoke and Comparison with Other Asteraceae. PLoS ONE. 2015;10: e0120589. doi:10.1371/journal.pone.0120589

50. Yi D-K, Kim K-J. Complete Chloroplast Genome Sequences of Important Oilseed Crop Sesamum indicum L. PLoS ONE. 2012;7: e35872. doi:10.1371/journal.pone.0035872

51. Yao X, Tang P, Li Z, Li D, Liu Y, Huang H. The First Complete Chloroplast Genome Sequences in Actinidiaceae: Genome Structure and Comparative Analysis. PloS One. 2015;10: e0129347. doi:10.1371/journal.pone.0129347

52. Barbrook AC, Howe CJ, Kurniawan DP, Tarr SJ. Organization and expression of organellar genomes. Philos Trans R Soc B Biol Sci. 2010;365: 785–797. doi:10.1098/rstb.2009.0250

53. Lisitsky I, Klaff P, Schuster G. Addition of destabilizing poly(A)-rich sequences to endonuclease cleavage sites during the degradation of chloroplast mRNA. Proc Natl Acad Sci. 1996;93: 13398–13403.

54. Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boore JL, et al. Comparative chloroplast genomics: analyses including new sequences from the angiosperms Nuphar advena and Ranunculus macranthus. BMC Genomics. 2007;8: 174. doi:10.1186/1471-2164-8-174

55. Do HDK, Kim JS, Kim J-H. Comparative genomics of four Liliales families inferred from the complete chloroplast genome sequence of Veratrum patulum O. Loes. (Melanthiaceae). Gene. 2013;530: 229–235. doi:10.1016/j.gene.2013.07.100

56. Schuster G, Lisitsky I, Klaff P. Polyadenylation and Degradation of mRNA in the Chloroplast. Plant Physiol. 1999;120: 937–944. doi:10.1104/pp.120.4.937

57. Lin C-P, Ko C-Y, Kuo C-I, Liu M-S, Schafleitner R, Chen L-FO. Transcriptional Slippage and RNA Editing Increase the Diversity of Transcripts in Chloroplasts: Insight from Deep Sequencing of Vigna radiata Genome and Transcriptome. PloS One. 2015;10: e0129396. doi:10.1371/journal.pone.0129396

58. Zeng W-H, Liao S-C, Chang C-C. Identification of RNA Editing Sites in Chloroplast Transcripts of Phalaenopsis aphrodite and Comparative Analysis with Those of Other Seed Plants. Plant Cell Physiol. 2007;48: 362–368. doi:10.1093/pcp/pcl058

59. Nazareno AG, Carlsen M, Lohmann LG. Complete Chloroplast Genome of Tanaecium tetragonolobum: The First Bignoniaceae Plastome. PloS One. 2015;10: e0129930. doi:10.1371/journal.pone.0129930

60. Zhang T, Fang Y, Wang X, Deng X, Zhang X, Hu S, et al. The Complete Chloroplast and Mitochondrial Genome Sequences of Boea hygrometrica: Insights into the Evolution of Plant Organellar Genomes. PLoS ONE. 2012;7: e30531. doi:10.1371/journal.pone.0030531

61. Neyland R, Urbatsch LE. The ndhF chloroplast gene detected in all vascular plant divisions. Planta. 1996;200: 273–277.

62. Kim HT, Kim JS, Moore MJ, Neubig KM, Williams NH, Whitten WM, et al. Seven New Complete Plastome Sequences Reveal Rampant Independent Loss of the ndh Gene Family across Orchids and Associated Instability of the Inverted Repeat/Small Single-Copy Region Boundaries. PLoS ONE. 2015;10: e0142215. doi:10.1371/journal.pone.0142215

63. Li R, Ma P-F, Wen J, Yi T-S. Complete Sequencing of Five Araliaceae Chloroplast Genomes and the Phylogenetic Implications. PLoS ONE. 2013;8: e78568. doi:10.1371/journal.pone.0078568

64. Müller K, Borsch T, Legendre L, Porembski S, Theisen I, Barthlott W. Evolution of carnivory in Lentibulariaceae and the Lamiales. Plant Biol Stuttg Ger. 2004;6: 477–490. doi:10.1055/s-2004-817909

65. Müller KF, Borsch T, Legendre L, Porembski S, Barthlott W. Recent progress in understanding the evolution of carnivorous lentibulariaceae (lamiales). Plant Biol Stuttg Ger. 2006;8: 748–757. doi:10.1055/s-2006-924706

66. Agnarsson I, Miller JA. Is ACCTRAN better than DELTRAN? Cladistics. 2008;24: 1032–1038. doi:10.1111/j.1096-0031.2008.00229.x

67. Refulio-Rodriguez NF, Olmstead RG. Phylogeny of Lamiidae. Am J Bot. 2014;101: 287–299. doi:10.3732/ajb.1300394

68. Schäferhoff B, Fleischmann A, Fischer E, Albach DC, Borsch T, Heubl G, et al. Towards resolving Lamiales relationships: insights from rapidly evolving chloroplast sequences. BMC Evol Biol. 2010;10: 352. doi:10.1186/1471-2148-10-352

69. Jobson RW, Playford J, Cameron KM, Albert VA. Molecular Phylogenetics of Lentibulariaceae Inferred from Plastid rps16 Intron and trnL-F DNA Sequences: Implications for Character Evolution and Biogeography. Syst Bot. 2003;28: 157–171.

70. Raynal-Roques A, Jérémie J. Biologie diversity in the genus Utricularia (Lentibulariaceae). Acta Bot Gallica. 2005;152: 177–186. doi:10.1080/12538078.2005.10515468

71. Płachno BJ, Stpiczyńska M, Davies KL, Świątek P, de Miranda VFO. Floral ultrastructure of two Brazilian aquatic-epiphytic bladderworts: Utricularia cornigera Studnička and U. nelumbifolia Gardner (Lentibulariaceae). Protoplasma. 2016; doi:10.1007/s00709-016-0956-0

72. Wade JT, Grainger DC. Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. Nat Rev Microbiol. 2014;12: 647–653. doi:10.1038/nrmicro3316

73. Bock R, Timmis JN. Reconstructing evolution: Gene transfer from plastids to the nucleus. BioEssays. 2008;30: 556–566. doi:10.1002/bies.20761

74. Rice Chromosome 10 Sequencing Consortium. In-depth view of structure, activity, and evolution of rice chromosome 10. Science. 2003;300: 1566–1569. doi:10.1126/science.1083523

75. Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature. 2000;408: 796–815. doi:10.1038/35048692

76. Vu GTH, Schmutzer T, Bull F, Cao HX, Fuchs J, Tran TD, et al. Comparative Genome Analysis Reveals Divergent Genome Size Evolution in a Carnivorous Plant Genus. Plant Genome. 2015;8: 0. doi:10.3835/plantgenome2015.04.0021

Supporting Information

S1 Table. Phylogenomic and phylogenetic analysis. (A) List of the chloroplast genomes used in the phylogenomics analysis with their respective GenBank accession number. **(B)** List of chloroplast genes used in the phylogenomics analysis. **(C)** The *mat*K genes used in the phylogenetic analysis with their respective GenBank accession number. **(D)** The *mat*K genes generated in this study and used in the phylogenetic analysis with their respective GenBank accession number.

A. Plastomes considered in the phylogenomics analysis

Species	Family	Genbank acession #
Andrographis paniculata	Acanthaceae	KF150644
Tanaecium tetragonolobum	Bignoniaceae	KR534325
Boea hygrometrica	Gesneriaceae	JN107811
Ajuga reptans	Lamiaceae	NC_023102
Origanum vulgare	Lamiaceae	JX880022
Premna microphylla	Lamiaceae	KM981744
Rosmarinus officinalis	Lamiaceae	KR232566
Salvia miltiorrhiza	Lamiaceae	JX312195
Scutellaria baicalensis	Lamiaceae	KR233163
Tectona grandis	Lamiaceae	HF567869
Genlisea margaretae	Lentibulariaceae	NC_025652
Pinguicula ehlersiae	Lentibulariaceae	NC_023463
Utricularia gibba	Lentibulariaceae	NC_021449
Utricularia macrorhiza	Lentibulariaceae	NC_025653
Hesperelaea palmeri	Oleaceae	LN515489
Jasminum nudiflorum	Oleaceae	DQ673255
Olea europaea	Oleaceae	FN997650
Olea woodiana	Oleaceae	FN998901
Lathraea squamaria	Orobanchaceae	KM652488
Lindenbergia philippensis	Orobanchaceae	HG530133
Sesamum indicum	Pedaliaceae	NC_016433
Scrophularia takesimensis	Scrophulariaceae	KM590983

B. Chloroplast genes considered for the phylogenomic analysis

atpA, atpB, atpE, atpF, atpH, atpI, clpP, matK, petA, petG, petI, petN, psaA, psaB, psaC, psaI, psaJ, psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, rbcL, rpl2, rpl14, rpl20, rpl23, rpl36, rpoB, rpoC1, rps2, rps4, rps7, rps14, rps18, ycf2, ycf3, ycf4

C. *mat*K accession numbers from GenBank Database used in the cloudgram analysis

DQ010661, DQ010653, AF531782, AF531786, AF531814 ,FN641691. FN641690, FN641689, FN641695, FN641694, FN641714, FN641717, FN641716, FN641711, AF531838, AF531832, NC021449, AF531821, AF531840, AF531837, AF531851, AF531849, AF531828, AF531850, AF531830, FN773562, AF531827, JN894431, JN894028, AF531844, KC584950, AF531839, JN896195, AF531835, AF531831, AF531823, AF531847, AF531833, JN894029, JN894054,

JN966728, JN894027, AF531846, AF531822, AF531836, AF531845, AF531825, AF531829, AF531843, AF531834, AF531824, AF531848, AF531826, AF531842, AF531841

D. *mat*K accession numbers generated in this study used in the cloudgram analysis

KX604173, KX604174, KX604175, KX604176, KX604177, KX604178, KX604179, KX604180, KX604181, KX604182, KX604183, KX604184, KX604185, KX604186, KX604187, KX604188, KX604189, KX604190, KX604191, KX604192, KX604193, KX604194, KX604195, KX604196, KX604197, KX604198, KX604199, KX604200, KX604201, KX604202, KX604203, KX604204, KX604205, KX604206, KX604207, KX604208, KX604209, KX604210, KX604211, KX604212, KX604213, KX604214, KX604215, KX604216, KX604217, KX604218, KX604219, KX604220, KX604221, KX604222, KX604223, KX604224, KX604225, KX604226, KX604227, KX604228, KX604229, KX604230, KX604231

S2 Table. The remaining paired-end reads (2x300bp; 229,830; 18.3%) which were assembled into contigs containing fragments of incomplete or truncated cp genes. The distribution, number of these contigs, truncated gene content and alignment to the *Utricularia reniformis* cp genome is shown.

							Coordinates on *U. reniformis* cp genome		
%substitutions	%deletions	%insertions	CP related contig (SPAdes assembly)	Start	Stop	Orientation	Start	Stop	Truncated cp gene
2.51	6.47	0	NODE_114_length_1943_cov_260.237_ID_42614	12	490	+	127324	127833	
3.65	4.92	0.49	NODE_114_length_1943_cov_260.237_ID_42614	501	1924	+	128247	129733	trnG-UCC, atpF, atpA
0.8	0	0	NODE_1310_length_622_cov_13.402_ID_196895	1	622	+	69549	70170	trnG-UCC, psbH, psbN
2.89	0	0	NODE_1374_length_607_cov_12.175_ID_62322	1	588	-	47138	46551	trnG-UCC
3.33	4.9	0.85	NODE_142_length_1821_cov_102.445_ID_39855	32	1805	+	122275	124120	trnG-UCC, rpoC2
1.24	4.97	0	NODE_1603_length_563_cov_270.567_ID_65498	1	563	-	1570	980	trnG-UCC, rpl2
3.38	2.69	1.94	NODE_168_length_1664_cov_287.092_ID_40316	34	1630	-	88540	86932	trnG-UCC, rbcL
0.96	1.54	0	NODE_1822_length_521_cov_161.962_ID_69672	1	521	+	65640	66168	trnG-UCC, rps11, rpoA
2.81	5.18	1.08	NODE_1835_length_518_cov_82.6394_ID_69838	12	474	+	33478	33959	trnG-UCC, psaC
3.54	5.82	1.01	NODE_2398_length_439_cov_78.9038_ID_75199	26	420	+	79382	79795	trnG-UCC, psbE, psbF
0.61	0	0	NODE_2664_length_400_cov_77.2711_ID_78001	1	164	+	105865	106028	
0.41	0	0	NODE_2664_length_400_cov_77.2711_ID_78001	158	400	+	106103	106345	trnG-UCC, ycf2
0.67	0.67	1.67	NODE_304_length_1274_cov_244.282_ID_58408	1	300	-	95486	95190	
0	0	0	NODE_304_length_1274_cov_244.282_ID_58408	291	513	-	95009	94787	
1.64	2.6	1.23	NODE_304_length_1274_cov_244.282_ID_58408	514	1244	-	94126	93386	trnG-UCC, rps4,trnL-UAA
0	0	0	NODE_3728_length_284_cov_215.694_ID_98587	23	70	-	12169	12122	
0.9	0	0	NODE_3728_length_284_cov_215.694_ID_98587	63	284	+	12196	12417	trnG-UCC
3.01	4.47	0	NODE_383_length_1130_cov_9.67597_ID_42832	35	1130	+	83284	84428	trnG-UCC, ycf4, psal
3.96	2.81	0.99	NODE_404_length_1098_cov_194.262_ID_43050	23	628	+	118150	118766	
1.9	3.38	0.84	NODE_404_length_1098_cov_194.262_ID_43050	625	1098	+	118795	119280	trnG-UCC, rpoC1
0	5.77	0	NODE_4062_length_260_cov_170.459_ID_197095	1	260	-	8544	8270	ycf2
0.39	0	0	NODE_4384_length_254_cov_5.61417_ID_114979	1	254	+	12163	12416	trnG-UCC
0	0.41	0	NODE_4505_length_245_cov_59.7542_ID_164079	1	245	+	23856	24101	ycf1
0.46	0	0	NODE_4981_length_217_cov_5.11111_ID_196685	1	217	-	56191	55975	ycf2
2.25	5.25	0	NODE_534_length_972_cov_260.49_ID_45655	40	972	-	100093	99112	psaA
2.6	3.51	0	NODE_66_length_2507_cov_185.713_ID_39476	28	682	-	58418	57741	
2.72	8.15	0	NODE_66_length_2507_cov_185.713_ID_39476	685	1089	+	3084	3521	
2.7	5.54	0.71	NODE_66_length_2507_cov_185.713_ID_39476	1080	2487	+	3790	5265	ycf2
4.59	3.61	0	NODE_698_length_844_cov_63.954_ID_49489	16	320	-	61121	60806	
2.57	5.94	0.79	NODE_698_length_844_cov_63.954_ID_49489	327	831	-	60602	60072	rpl22, rps19, rpl2
0.72	2.51	0.84	NODE_705_length_838_cov_129.097_ID_49689	1	838	+	68628	69479	petB
4.48	1.92	0	NODE_707_length_837_cov_113.123_ID_49719	27	807	-	137020	136225	matK
2.41	4.42	1.34	NODE_796_length_781_cov_266.352_ID_51715	35	781	+	89810	90579	atpB
2.08	1.56	0	NODE_798_length_781_cov_18.1147_ID_51777	13	781	-	55585	54805	ycf2
1.74	3.49	0.13	NODE_878_length_746_cov_140.459_ID_53007	1	746	-	68150	67380	trnG-UCC
4.43	6.5	0.28	NODE_937_length_724_cov_84.4154_ID_54287	1	723	+	107430	108197	psbD
1.42	6.02	0	NODE_967_length_716_cov_198.027_ID_54971	1	565	-	54247	53649	
0	0	0	NODE_967_length_716_cov_198.027_ID_54971	590	716	-	53569	53443	ycf2

S3 Table. RNAseq experiment table, showing the expression profile of all chloroplast related genes of *Utricularia reniformis*.

Name	TPM	RPKM	Gene length	Unique gene reads	Total gene reads
accD	3630.39	5293.81	1473	886	886
atpA	18130.66	26437.99	1524	4578	4578
atpB	6898.44	10059.25	1497	1711	1711
atpE	3693.45	5385.76	402	246	246
atpF	9679.59	14114.71	588	943	943
atpH	30497.12	44470.66	246	1243	1243
atpI	3691.15	5382.4	744	455	455
ccsA	1910.97	2786.56	957	303	303
cemA	992.76	1447.64	687	113	113
clpP	13184.44	19225.44	591	1291	1291
infA	7789.58	11358.7	234	302	302
matK	2647.63	3860.75	1500	658	658
ndhA ψ	778.32	1134.94	1644	212	212
ndhB ψ	612.47	893.1	1084	110	110
ndhD ψ	924.16	1347.6	738	113	113
ndhE ψ	2218.22	3234.6	234	86	86

ndhG ψ	376.49	549	513	32	32
ndhH ψ	2084.13	3039.06	1086	375	375
ndhI ψ	275.91	402.34	525	24	24
orf42 ψ	33037.14	48174.51	114	624	624
orf56 ψ	11769.48	17162.17	120	234	234
petA	3653.97	5328.19	963	583	583
petB	18041.7	26308.26	648	1937	1937
petD	11034.09	16089.82	483	883	883
petG	52.94	77.2	114	1	1
petL	125.74	183.36	96	2	2
petN	1207.13	1760.22	90	18	18
psaA	7506.37	10945.72	2253	2802	2802
psaB	9024.71	13159.76	2205	3297	3297
psaC	22032.51	32127.64	246	898	898
psaI	18324.4	26720.49	111	337	337
psaJ	222021.85	323750.55	135	4966	4966
psbA	299974.91	437421.11	1059	52633	52633
psbB	19996.24	29158.37	1527	5059	5059
psbC	16120.49	23506.77	1422	3797	3798
psbD	7513.28	10955.81	1062	1322	1322
psbE	9197.15	13411.22	252	384	384
psbF	754.45	1100.14	120	15	15
psbH	7226.88	10538.17	228	273	273
psbI	598.13	872.18	111	11	11
psbJ	588.84	858.65	123	12	12
psbK	227.15	331.22	186	7	7
psbL	3043.61	4438.17	117	59	59
psbM	5460.81	7962.91	105	95	95
psbN	7910.34	11534.79	132	173	173
psbT	335.31	488.95	108	6	6
psbZ	1309.32	1909.24	189	41	41
rbcL	18439.4	26888.2	1434	4381	4381
rpl14	3467.63	5056.47	369	212	212
rpl16	5360.11	7816.07	411	365	365
rpl20	3472.11	5063.01	372	214	214
rpl22	4236.93	6178.26	453	318	318
rpl23	1005.94	1466.85	282	47	47
rpl2	16095.02	23469.63	828	2208	2208
rpl32	189.8	276.76	159	5	5
rpl33	6966.5	10158.5	201	232	232
rpl36	1535.38	2238.88	114	29	29
rpoA	8381.81	12222.29	993	1379	1379
rpoB	761.91	1111	3240	409	409
rpoC1	720.41	1050.5	2061	246	246
rpoC2	833.01	1214.69	4101	566	566
rps11	5181.67	7555.87	417	358	358
rps12	3504.56	5110.32	372	216	216
rps14	8246.71	12025.28	303	414	414
rps15	1786.05	2604.41	294	87	87
rps16	6080.84	8867.04	267	269	269
rps18	5883.79	8579.7	318	310	310
rps19	1514.32	2208.16	279	70	70
rps2	4903.95	7150.9	720	585	585
rps3	6758.8	9855.62	651	729	729

rps4	2539.75	3703.44	606	255	255
rps7	13244.86	19313.55	468	1027	1027
rps8	2399.35	3498.71	405	161	161
ycf1_small	40.37	58.87	897	2	6
ycf15	232.95	339.69	285	11	11
ycf1_large	3408.11	4969.67	4794	2314	2707
ycf2	509.17	742.46	6816	575	575
ycf3	7929.16	11562.25	510	670	670
ycf68 ψ	6264.83	9135.33	395	410	410

PLOS | ONE

RESEARCH ARTICLE

The Chloroplast Genome of *Utricularia reniformis* Sheds Light on the Evolution of the *ndh* Gene Complex of Terrestrial Carnivorous Plants from the Lentibulariaceae Family

Saura R. Silva[1], Yani C. A. Diaz[2], Helen Alves Penha[3], Daniel G. Pinheiro[3], Camila C. Fernandes[3], Vitor F. O. Miranda[2]*, Todd P. Michael[4], Alessandro M. Varani[3]*

1 Instituto de Biociências, UNESP - Univ Estadual Paulista, Câmpus Botucatu, São Paulo, Brazil, **2** Departamento de Biologia Aplicada à Agropecuária, Faculdade de Ciências Agrárias e Veterinárias, UNESP - Univ Estadual Paulista, Câmpus Jaboticabal, São Paulo, Brazil, **3** Departamento de Tecnologia, Faculdade de Ciências Agrárias e Veterinárias, UNESP - Univ Estadual Paulista, Câmpus Jaboticabal, São Paulo, Brazil, **4** Ibis Bioscience, Computational Genomics, Carlsbad, California, United States of America

* amvarani@fcav.unesp.br (AMV); vmiranda@fcav.unesp.br (VFOM)

OPEN ACCESS

Citation: Silva SR, Diaz YCA, Penha HA, Pinheiro DG, Fernandes CC, Miranda VFO, et al. (2016) 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 11(10): e0165176. doi:10.1371/journal.pone.0165176

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

Received: July 18, 2016

Accepted: October 8, 2016

Published: October 20, 2016

Data Availability Statement: The annotated sequence, and the raw DNAseq and RNAseq reads for the U. reniformis chloroplast genome has been deposited in the GenBank database under accession number [GenBank: KT336489, SRR3277235 and SRP072162 respectively] (BioProject PRJNA290588). All sequences generated by this projects are already public on the NCBI repository (BioProject PRJNA290588).

Funding: This study was funded by São Paulo Research Foundation/Brazil, FAPESP [grant

Abstract

Lentibulariaceae is the richest family of carnivorous plants spanning three genera including *Pinguicula*, *Genlisea*, and *Utricularia*. *Utricularia* is globally distributed, and, unlike *Pinguicula* and *Genlisea*, has both aquatic and terrestrial forms. In this study we present the analysis of the chloroplast (cp) genome of the terrestrial *Utricularia reniformis*. *U. reniformis* has a standard cp genome of 139,725bp, encoding a gene repertoire similar to essentially all photosynthetic organisms. However, an exclusive combination of losses and pseudogenization of the plastid NAD(P)H-dehydrogenase (*ndh*) gene complex were observed. Comparisons among aquatic and terrestrial forms of *Pinguicula*, *Genlisea*, and *Utricularia* indicate that, whereas the aquatic forms retained functional copies of the eleven *ndh* genes, these have been lost or truncated in terrestrial forms, suggesting that the *ndh* function may be dispensable in terrestrial Lentibulariaceae. Phylogenetic scenarios of the *ndh* gene loss and recovery among *Pinguicula*, *Genlisea*, and *Utricularia* to the ancestral Lentibulariaceae cladeare proposed. Interestingly, RNAseq analysis evidenced that *U. reniformis* cp genes are transcribed, including the truncated *ndh* genes, suggesting that these are not completely inactivated. In addition, potential novel RNA-editing sites were identified in at least six *U. reniformis* cp genes, while none were identified in the truncated *ndh* genes. Moreover, phylogenomic analyses support that Lentibulariaceae is monophyletic, belonging to the higher core Lamiales clade, corroborating the hypothesis that the first *Utricularia* lineage emerged in terrestrial habitats and then evolved to epiphytic and aquatic forms. Furthermore, several truncated cp genes were found interspersed with *U. reniformis* mitochondrial and nuclear genome scaffolds, indicating that as observed in other smaller plant genomes, such as *Arabidopsis thaliana*, and the related and carnivorous *Genlisea*

CAPÍTULO 2

The complete chloroplast genome sequence of the leafy bladderwort, *Utricularia foliosa* L. (Lentibulariaceae)

Nota técnica publicada

Silva, S.R., Pinheiro, D.G., Meer, E.J., Michael, T.P., Varani, A.M., and Miranda, V.F.O. (2016). The complete chloroplast genome sequence of the leafy bladderwort, *Utricularia foliosa* L. (Lentibulariaceae). Conservation Genetics Resources 1–4.

The complete chloroplast genome sequence of the leafy bladderwort, *Utricularia foliosa* L. (Lentibulariaceae)

Saura R. Silva[1], Daniel G. Pinheiro[2], Elliott J. Meer[3],
Todd P. Michael[3], Alessandro M. Varani[2] and Vitor F.O. Miranda[1,4]

[1] Instituto de Biociências, UNESP - Univ Estadual Paulista, Câmpus Botucatu, São Paulo, Brazil.
[2] Departamento de Tecnologia, Faculdade de Ciências Agrárias e Veterinárias, UNESP – Univ Estadual Paulista, Câmpus Jaboticabal, São Paulo, Brazil.
[3] Ibis Bioscience, Computational Genomics, Carlsbad, CA USA.
[4] Departamento de Biologia Aplicada à Agropecuária, Faculdade de Ciências Agrárias e Veterinárias, UNESP - Univ Estadual Paulista, Câmpus Jaboticabal, São Paulo, Brazil.

Authors for correspondence:
Alessandro M. Varani (amvarani@fcav.unesp.br); Vitor F.O. Miranda (vmiranda@fcav.unesp.br)

ABSTRACT

Utricularia foliosa L., commonly known as leafy bladderwort, is a widespread aquatic carnivorous plant from the Lentibulariaceae family. The species of the *Utricularia* and *Genlisea* genera are known to possess the smallest nuclear genomes across angiosperms. This study reveals that *U. foliosa* have a typical chloroplast genome of 150,851 bp in length, comprised of a large single-copy region (LSC) of 82,720 bp, a small single-copy region (SSC) of 17,481 bp, and inverted repeat regions (IRs) of 25,325 bp. A total of 139 genes, 88 of which are protein coding genes, 37 tRNA genes, 8 rRNA genes and 6 pseudogenes were identified. All plastid NAD(P)H-dehydrogenase genes are present as intact copies. Phylogenetic analyses indicate that *U. foliosa* is closely related to other suspended or affixed aquatic species belonging to the *Utricularia* sect. *Utricularia*.

Keywords: carnivorous plant, Lentibulariaceae, bladderwort, plastome, cpDNA

Utricularia foliosa L. (leafy bladderwort, Lentibulariaceae) is a perennial free floating aquatic plant (Taylor 1989) that grows in lakes, ponds and swamps. The genus *Utricularia*, which also includes *Genlisea*, has increasingly gained researchers attention due to the dynamic size of its nuclear DNA content (Albert et al. 2010; Veleba et al. 2014). Despite its widespread distribution in the Americas, Africa and Madagascar, only a few studies of *U. foliosa* have been conducted. These studies have focused mainly on prey composition (Solís-Parra and Críales-Hernández 2016), bladder respiration and photosynthesis (Adamec 2006). Only a single phylogenetic analysis, with one molecular marker, has been used to determine its phylogenetic position within the *Utricularia* genus (Silva et al. 2016). *U. foliosa* can bloom throughout the year, and is found in slowly flowing, shallow to deep water bodies, such as lakes, marshes and rivers (Taylor 1989). These habitats, where *U. foliosa* is commonly found, are suffering rapid environmental loss due to negative anthropic impacts such as drainage and conversion to urban or agricultural activities.

Herein, we report the first complete chloroplast genome of *Utricularia foliosa*. The annotated *U. foliosa* genome has been deposited in the public database, GenBank, with the accession number: KY025562 (BioProject: PRJNA350159; BioSample: SAMN05933770).

Total genomic DNA was extracted from silica gel dried inflorescences of *Utricularia foliosa,* collected from ponds in the Tietê River in Mogi das Cruzes (São Paulo, Brazil; lat. -23.532294, long. -46.202648, 765 m a.s.l.) using a modified CTAB method (Doyle and Doyle 1987). Herbarium voucher (V.F.O.de Miranda et al. 2070) is deposited at the JABU Herbarium at Universidade Estadual Paulista (UNESP/ FCAV).

Genomic sequencing was performed on the Illumina MiSeq Platform (Illumina, San Diego, CA), resulting in 2,933,837 paired-end reads (2x300bp). The reads were quality trimmed using Platanus_trim v1.0.7 (Kajitani et al. 2014) and aligned to the reference *Utricularia gibba* chloroplast genome (Genbank accession number: NC_021449) using Bowtie 2 v.2.2.3 (Langmead and Salzberg 2012). The matched chloroplast reads (135,879 reads) were *de novo* assembled with SPAdes 3.9 (Bankevich et al. 2012) and ambiguous regions were picked out to extend the length using an iteration method with MITObim v.1.8 (Hahn et al. 2013). Quality filtered reads were mapped back to the chloroplast genome using Bowtie2 (98.95% overall alignment rate) to confirm assembly accuracy quality and repeat region junctions. The gene annotation was performed using DOGMA (http://evogen.jgi-psf.org/dogma/) (Wyman et al. 2004) and confirmed by BLASTn searches on NCBI nucleotide collection database (nt). tRNA genes were annotated using DOGMA and ARAGORN v1.2 (Laslett and Canback 2004). The genome map was generated using OGDRAW v. 1.2 (http://ogdraw.mpimp-golm.mpg.de/) (Lohse et al 2013) followed by manual modifications.

The phylogenetic position of *Utricularia foliosa* was inferred using 53 genes of the previously published Lentibulariaceae chloroplast genomes, with the *Tectona grandis* (Lamiaceae), *Sesamum indicum* (Pedaliaceae) and *Tanaecium tetragonolobum* (Bignoniaceae) plastomes used as outgroups.

The sequences were aligned using MAFFT v.7 (Katoh and Standley 2013) and probabilistic phylogeny was conducted using Mr. Bayes v.3 for Bayesian inference (Ronquist and Huelsenbeck 2003) with $5x10^5$ generations, using the default parameters, and RAxML for maximum likelihood (Stamatakis 2014) using the default parameters with bootstrap support of 10^4 pseudoreplicates. The evolutionary model used was GTR+G, tested a priori using jModelTest (Darriba et al. 2012).

The chloroplast genome of *Utricularia foliosa* is a typical quadripartite structure with a length of 150,851 bp, which contained inverted repeats (IR) of 25,325 bp separated by a large single-copy (LSC) and a small single-copy (SSC) of 82,720 bp and 17,481 bp, respectively. The cpDNA contains 139 genes, comprising 88 protein-coding genes, 8 ribosomal RNA genes, 37 tRNA genes and 6 pseudogenes. Among the annotated genes, 15 of them contain one intron (*atp*F, *pet*B, *pet*D, *rpl*2, *rpl*16, *rpo*C1, *rps*12, *trn*A-UGC, *trn*G-UCC, *trn*I-GAU, *trn*L-UAA, *trn*V-UAC, *trn*K-UUU, *ndh*B, *ndh*A), and *ycf*3 and *clp*P genes contain two introns. The overall GC content of *U. foliosa* cp genome is 37.32%, while it was 43% for IRs, 35.20% for LSC and 30.83% for SSC regions (Fig. 1). Interestingly, all the eleven subunits of the plastid NAD(P)H-dehydrogenase (*ndh*) genes are present as intact copies, corroborating previous studies which suggested that the plastid ndh complete gene set may exist only in aquatic forms of carnivorous plants in the Lentibulariaceae family (Silva et al. 2016).

Previous phylogenetic studies have shown that Lentibulariaceae and *Utricularia* are monophyletic groups, supported by molecular and morphological characteristics (Müller et al. 2004; Silva et al. 2016). Our phylogenetic analysis strongly supports the positioning of *U. foliosa* nested within the known related aquatic species *U. gibba* and *U. macrorhiza* (*Utricularia* sect. *Utricularia*) (Fig. 2).

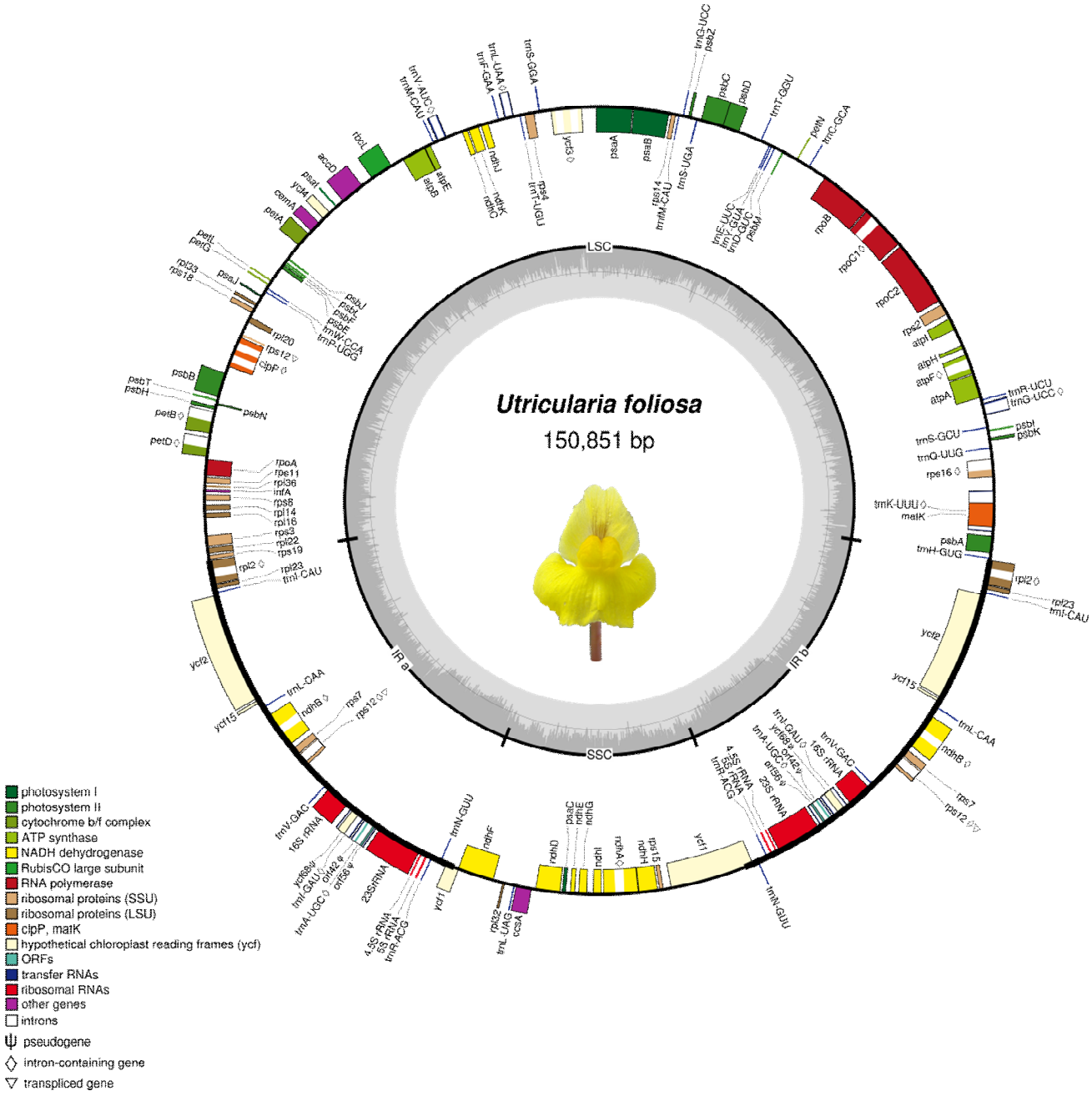


Fig. 1 Gene map of the *Utricularia foliosa* chloroplast genome. Genes belonging to different functional groups are shown in different colors. The ♦ and ψ indicate genes with intron(s) and pseudogene, respectively.

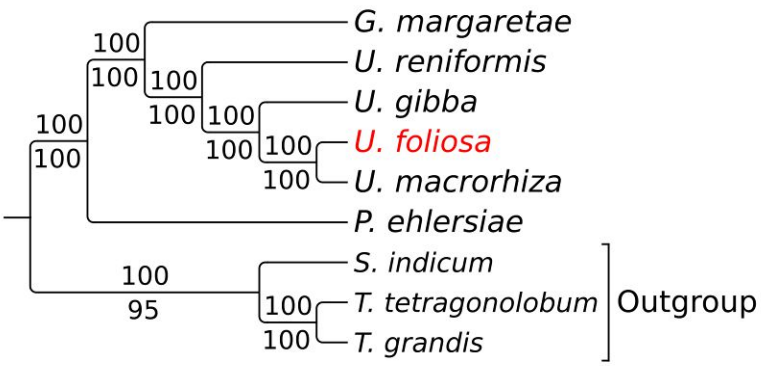


Fig. 2 Bayesian Inference phylogram based on 53 genes of complete chloroplast genome sequences deposited in public databases. Numbers below the nodes are bootstrap values for Maximum Likelihood analysis and above the nodes are posterior probability values. Accession Numbers: *Genlisea margaretae* NC_025652, *Pinguicula ehlersiae* NC_023463, *Sesamum indicum* JN_637766, *Tanaecium tetragonolobum* NC_027955, *Tectona grandis* NC_020098, *Utricularia gibba* NC_021449, *U. reniformis* NC_029719, *U. macrorhiza* NC_025653.

Acknowledgements

The first author was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). VFOM thanks for the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the fellowship (Bolsa de Produtividade - Proc. # 309040/2014-0).

REFERENCES

Adamec L (2006), Respiration and Photosynthesis of Bladders and Leaves of Aquatic *Utricularia* Species. Plant Biology, 8: 765–769. doi:10.1055/s-2006-924540

Albert VA, Jobson RW, Michael TP, Taylor DJ (2010) The carnivorous bladderwort (Utricularia, Lentibulariaceae): a system inflates. J Exp Bot 61:5-9. doi:10.1093/jxb/erp349.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing J. Comput. Biol. 19:455–477. doi: 10.1089/cmb.2012.0021.

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9:772. doi:10.1038/nmeth.2109.

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull, 19:11-15.

Hahn C, Bachmann L, Chevreux B (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads - a baiting and iterative mapping approach. Nucleic Acids Res. 41(13):e129. doi: 10.1093/nar/gkt371.

Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, Yabana M, Harada M, Nagayasu E, Maruyama H, Kohara Y, Fujiyama A, Hayashi T, Itoh T, "Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads". Genome Res. 2014 Aug;24(8):1384-95. doi: 10.1101/gr.170720.113.

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. doi: 10.1093/molbev/mst010.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357-359. doi: 10.1038/nmeth.1923.

Laslett D, Canback B (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32:11–16. doi: 10.1093/nar/gkh152.

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. doi: 10.1093/nar/gkt289.

Müller K, Borsch T, Legendre L, Porembski S, Theisen I, Barthlott W. (2004) Evolution of carnivory in Lentibulariaceae and the Lamiales. Plant Biology. 6: 1-14. doi:10.1055/s-2004-817909.

Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574. doi: 10.1093/bioinformatics/btg180.

Silva SR, Diaz YCA, Penha HA, Pinheiro DG, Fernandes CC, Miranda VFO, Michael TP, Varani, AM. (2016) 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, in press. doi: 10.1371/journal.pone.0165176

Solís-Parra, J, Críales-Hernández, M (2016). Capture and selectivity of zooplankton by *Utricularia foliosa* (Lentibulariaceae) in the Ciénaga de Paredes, Santander, Colombia. J Trop Biol Cons 64:1297-1310. doi: 10.15517/rbt.v64i3.21387.

Stamatakis A. (2014) RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. Bioinformatics 30:1312-1313. doi: 10.1093/bioinformatics/btu033

Taylor P (1989) The Genus *Utricularia*: A Taxonomic Monograph. Kew Bulletin Additional Series XIV. Kew Royal Botanic Gardens. London.

Veleba A, Bureš P, Adamec L, Šmarda P, Lipnerová I, Horová L (2014) Genome size and genomic GC content evolution in the miniature genome-sized family Lentibulariaceae. New Phytol, 203: 22–28. doi:10.1111/nph.12790

Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20:3252-3255. doi:10.1093/bioinformatics/bth352.

Conservation Genet Resour (2017) 9:213–216
DOI 10.1007/s12686-016-0653-5

TECHNICAL NOTE

The complete chloroplast genome sequence of the leafy bladderwort, *Utricularia foliosa* L. (Lentibulariaceae)

Saura R. Silva[1] · Daniel G. Pinheiro[2] · Elliott J. Meer[3] · Todd P. Michael[3] · Alessandro M. Varani[2] · Vitor F. O. Miranda[1,4]

Received: 28 October 2016 / Accepted: 15 November 2016 / Published online: 18 November 2016

Abstract *Utricularia foliosa* L., commonly known as leafy bladderwort, is a widespread aquatic carnivorous plant from the Lentibulariaceae family. The species of the *Utricularia* and *Genlisea* genera are known to possess the smallest nuclear genomes across angiosperms. This study reveals that *U. foliosa* have a typical chloroplast genome of 150,851 bp in length, comprised of a large single-copy region (LSC) of 82,720 bp, a small single-copy region (SSC) of 17,481 bp, and inverted repeat regions (IRs) of 25,325 bp. A total of 139 genes, 88 of which are protein coding genes, 37 tRNA genes, eight rRNA genes and six pseudogenes were identified. All plastid NAD(P)H-dehydrogenase genes are present as intact copies. Phylogenetic analyses indicate that *U. foliosa* is closely related to other suspended or affixed aquatic species belonging to the *Utricularia* sect. *Utricularia*.

Keywords Carnivorous plant · Lentibulariaceae · Bladderwort · Plastome · CpDNA

✉ Alessandro M. Varani
amvarani@fcav.unesp.br

✉ Vitor F. O. Miranda
vmiranda@fcav.unesp.br

[1] Instituto de Biociências, UNESP - Univ Estadual Paulista, Câmpus Botucatu, São Paulo, Brazil

[2] Departamento de Tecnologia, Faculdade de Ciências Agrárias e Veterinárias, UNESP – Univ Estadual Paulista, Câmpus Jaboticabal, São Paulo, Brazil

[3] Ibis Bioscience, Computational Genomics, Carlsbad, CA, USA

[4] Departamento de Biologia Aplicada à Agropecuária, Faculdade de Ciências Agrárias e Veterinárias, UNESP - Univ Estadual Paulista, Câmpus Jaboticabal, São Paulo, Brazil

Utricularia foliosa L. (leafy bladderwort, Lentibulariaceae) is a perennial free floating aquatic plant (Taylor 1989) that grows in lakes, ponds and swamps. The genus *Utricularia*, which also includes *Genlisea*, has increasingly gained researchers attention due to the dynamic size of its nuclear DNA content (Albert et al. 2010; Veleba et al. 2014). Despite its widespread distribution in the Americas, Africa and Madagascar, only a few studies of *U. foliosa* have been conducted. These studies have focused mainly on prey composition (Solís-Parra and Críales-Hernández 2016), bladder respiration and photosynthesis (Adamec 2006). Only a single phylogenetic analysis, with one molecular marker, has been used to determine its phylogenetic position within the *Utricularia* genus (Silva et al. 2016). *U. foliosa* can bloom throughout the year, and is found in slowly flowing, shallow to deep water bodies, such as lakes, marshes and rivers (Taylor 1989). These habitats, where *U. foliosa* is commonly found, are suffering rapid environmental loss due to negative anthropic impacts such as drainage and conversion to urban or agricultural activities.

Herein, we report the first complete chloroplast genome of *U. foliosa*. The annotated *U. foliosa* genome has been deposited in the public database, GenBank, with the accession number: KY025562 (BioProject: PRJNA350159; BioSample: SAMN05933770).

Total genomic DNA was extracted from silica gel dried inflorescences of *Utricularia foliosa*, collected from ponds in the Tietê River in Mogi das Cruzes (São Paulo, Brazil; lat. −23.532294, long. −46.202648, 765 m a.s.l.) using a modified CTAB method (Doyle and Doyle 1987). Herbarium voucher (V.F.O.de Miranda et al. 2070) is deposited at the JABU Herbarium at Universidade Estadual Paulista (UNESP/ FCAV).

Genomic sequencing was performed on the Illumina MiSeq Platform (Illumina, San Diego, CA), resulting in

CAPÍTULO 3

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

Artigo publicado:

Silva, S.R., Alvarenga, D.O., Aranguren, Y., Penha, H.A., Fernandes, C.C., Pinheiro, D.G., Oliveira, M.T., Michael, T.P., Miranda, V.F.O., and Varani, A.M. (2017). The mitochondrial genome of the terrestrial carnivorous plant *Utricularia reniformis* (Lentibulariaceae): Structure, comparative analysis and evolutionary landmarks. PLOS ONE 12, e0180484.

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

Saura R. Silva[1§], Danillo O. Alvarenga[3§], Yani Aranguren [2,#a], Helen A. Penha[3], Camila C. Fernandes[3], Daniel G. Pinheiro[3], Marcos T. Oliveira[3], Todd P. Michael[4,#b], Vitor F.O. Miranda[2*] and Alessandro M. Varani[3*]

[1] Departamento de Botânica, Instituto de Biociências, Universidade Estadual Paulista (UNESP), Botucatu, São Paulo, Brazil

[2] Departamento de Biologia Aplicada à Agropecuária, Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista (Unesp), Jaboticabal, São Paulo, Brazil

[3] Departamento de Tecnologia, 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

[#a] Current address: Universidad Simón Bolívar, Barranquilla, Colômbia

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

* Corresponding authors

E-mail: amvarani@fcav.unesp.br (AV) and vmiranda@fcav.unesp.br (VM)

[§]These authors contributed equally to this work

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 transcriptionaly 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 described to date, shedding light on the genomic trends and evolutionary characteristics and

phylogenetic history of the family Lentibulariaceae.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

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, 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:

(**a**) 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];

(**b**) 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;

(**c**) Potential mitochondrial contigs were baited by mapping against plant mitochondrial genes commonly found in angiosperms [23] with BLAST v2.2.26 [24];

(**d**) 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;

(**e**) 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;

(**f**) 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;

(**g**) 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 RNAweasel tool [39]. Potential truncated pseudogenes were defined by BLAST comparative analysis with the use of at least one of the following criteria:

(**a**) presence of at least one stop codon in-frame within the predicted coding region;

(**b**) absence of start and/or stop-codon;

(**c**) frameshift;

(**d**) 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 (nruns=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) $\geq$ 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 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.

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)

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.

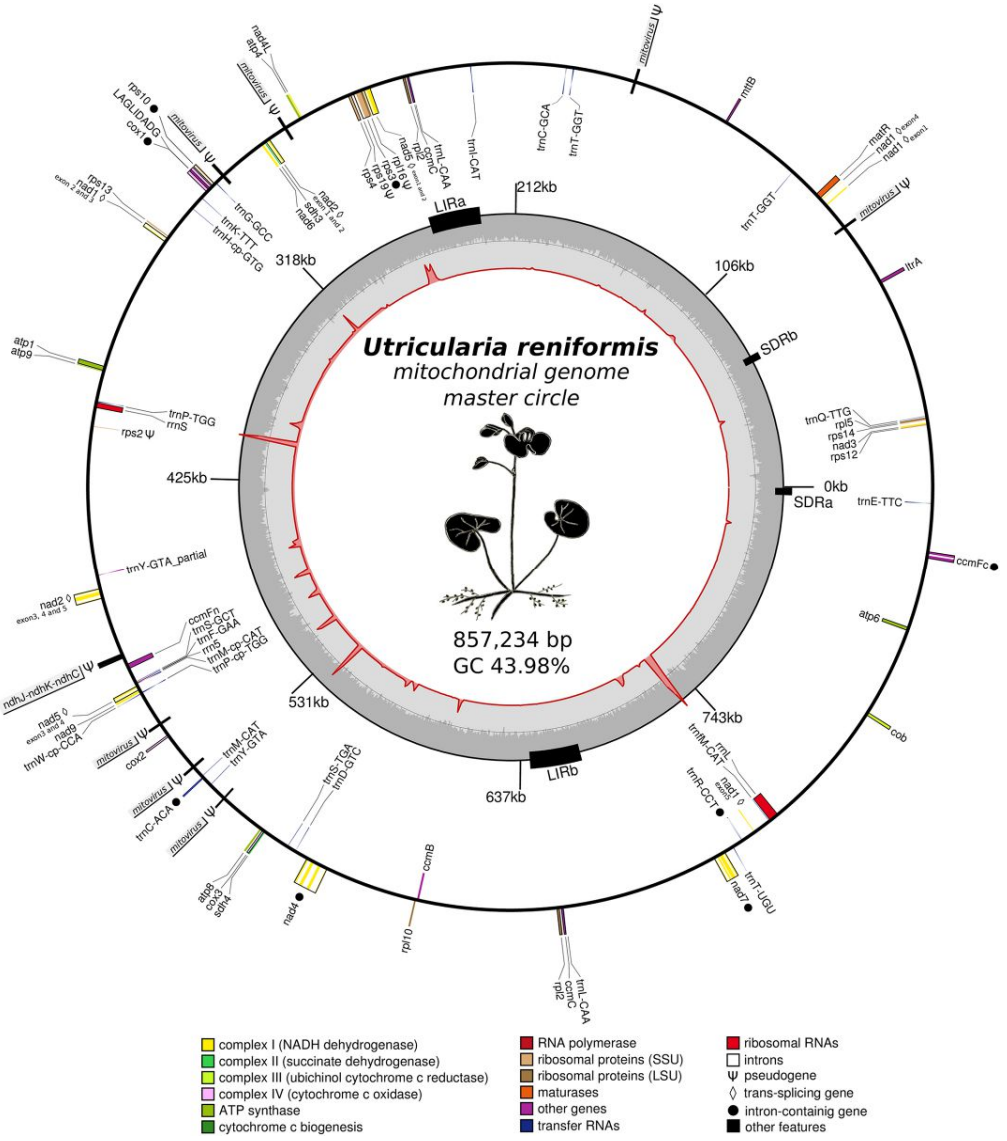


Fig 1. Genomic map of the *Utricularia reniformis* mtDNA genome. The inner red circle illustrates the transcript depth of coverage for *U. reniformis* mtDNA, whereas the peaks represent the most covered regions by RNAseq reads. The second level circle in gray scale represents the GC% distribution across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

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

***U. reniformis* 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 *U. reniformis* 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

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

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

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

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)	*nad1*•◊, *nad2*•◊, *nad3*, *nad4*•, *nad4L*, *nad5*•◊, *nad6*, *nad7*•, *nad9*
Complex II (succinate dehydrogenase)	*sdh3*, *sdh4*
Complex III (ubichinol cytochrome c reductase)	*cob*
Complex IV (cytochrome c oxydase)	*cox1*•, *cox2*, *cox3*
Complex V (ATP synthase)	*atp1*, *atp4*, *atp6*, *atp8*, *atp9*
Cytochrome c biogenesis	*ccmB*, *ccmC* (2x), *ccmFc*•, *ccmFn*
Ribosomal proteins (SSU)	*rps2* Ψ, *rps3*•, *rps4*, *rps10*•, *rps12*, *rps13*, *rps14*, *rps19* Ψ
Ribosomal proteins (LSU)	*rpl2* (2x), *rpl5*, *rpl10*, *rpl16* Ψ
Maturases	*matR*
Transport membrane proteins	*mttB*
Transfer RNAs	*trnC*-GCA, *trnD*-GTC, *trnE*-TTC, *trnF*-GAA, *trnG*-GCC, *trnH*-cp-GTG, *trnI*-CAT, *trnK*-TTT, *trnL*-CAA (2x), *trnfM*-CAT, *trnM*-CAT, *trnM*-cp-CAT, *trnP*-TGG, *trnP*-cp-TGG, *trnQ*-TTG, *trnR*-CCT•, *trnS*-GCT, *trnS*-TGA, *trnT*-GGT (2x), *trnT*-TGT, *trnV*-cp-TAC•, *trnW*-cp-CCA, *trnY*-GTA, *trnY*-GTA Ψ
Ribosomal RNAs	*rrn5*, *rrn5* Ψ, *rrnS*, *rrnL*
Others	LAGLIDADG endonuclease (intron region of *cox1* gene), Group II intron-encoded protein *ltrA*
Genes of Chloroplast Origin	

Photosystem I	*psaA* Ψ, *psaB* Ψ (2x), *psaC* Ψ
Protosystem II	*psbA* Ψ, *psbB* Ψ (2x), *psbC* Ψ (2x), *psbD* Ψ, *psbE* Ψ
Cytochrome b/f complex	*petD* Ψ, *petN* Ψ
ATP synthase	*atpA* Ψ, *atpB* Ψ (2x), *atpE* Ψ (2x), *atpF* Ψ
NADH dehydrogenase	*ndhA* Ψ (2x), *ndhB* Ψ, *ndhC* Ψ, *ndhD* Ψ, *ndhG* Ψ, *ndhH* Ψ (2x), *ndhJ* Ψ (2x), *ndhK* Ψ, *ndhN* Ψ
RubisCO large subunit	*rbcL* Ψ
RNA polymerase	*rpoA* Ψ, *rpoB* Ψ (4x), *rpoC1* Ψ, *rpoC2* Ψ (4x)
Ribosomal proteins (SSU)	*rps4* Ψ, *rps8* Ψ, *rps11*, *rps12* Ψ, *rps19* Ψ
Ribosomal proteins (LSU)	*rpl2* Ψ (2x), *rpl14* Ψ, *rpl23* Ψ, *rpl36*
Other genes	*accD* Ψ, *clpP* Ψ (2x), *infA* Ψ, *matK* Ψ (2x)
hypothetical chloroplast reading frames	*ycf2* Ψ (2x), *ycf15* Ψ, *orf56* Ψ
Ribosomal RNAs	*rrna16* Ψ, *rrn23* Ψ (4x)
Transfer RNAs	*trnL*-cp-TAA Ψ (2x), *trnK*-cp-TTT Ψ, *trnS*-cp-GCT Ψ, *trnE*-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

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

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.

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% 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.

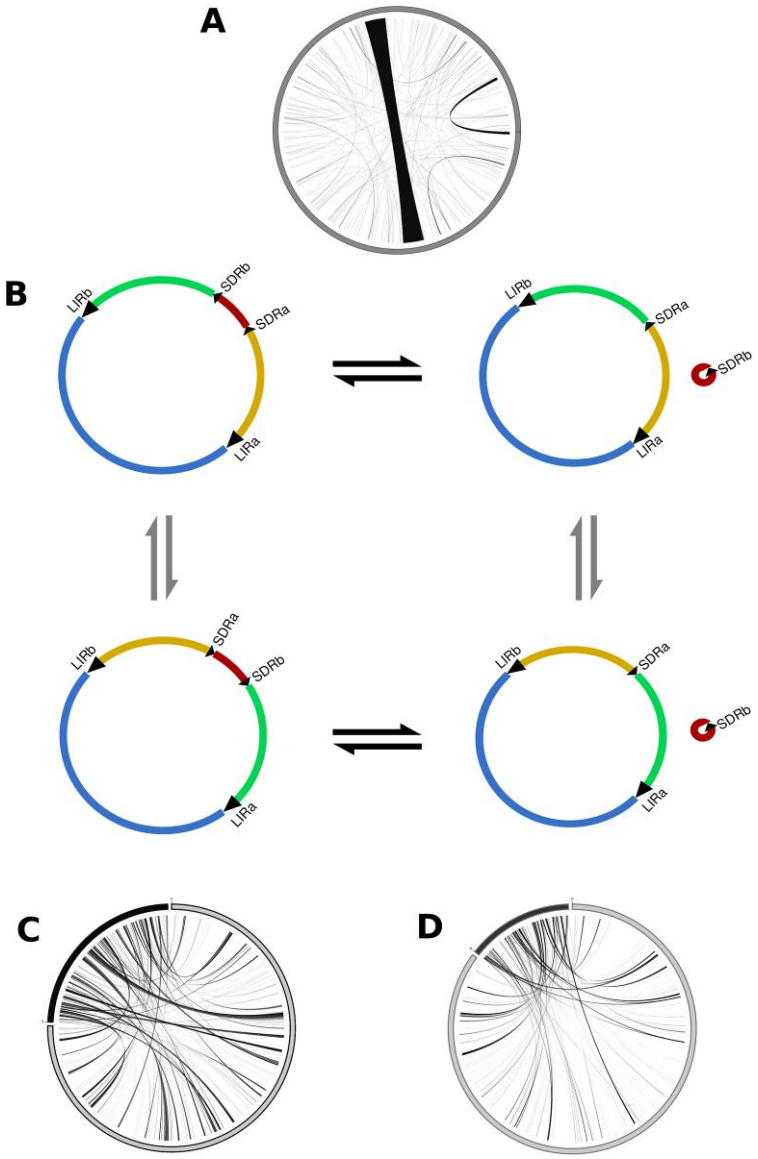


Fig 2. Repeats and alternative master circle structures. (A) Repeats across *U. reniformis* mtDNA; (B) Putative alternative conformations of *U. reniformis* mtDNA based on repeat-mediated intramolecular recombination mechanism; (C) Shared regions among *U. reniformis* mtDNA and *U. gibba* mtDNA; (D)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

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.

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

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

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.

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

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 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

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

circular mtDNA of *U. gibba* exhibited a TE load of 10,452bp (3.9%), with Ty1/Copia (6.5kbp) and Gypsy/DIRs (3.1kbp) 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 mitoviruses 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].

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

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 *rps19* Ψ, *rpl16* Ψ and *rps2* Ψ present unique mapped RNAseq reads, indicating that they are transcribed (≥ 98% identity considering 100% of the read length) (S3 Table). This appears to be a common feature of *U. reniformis* organelles, since the

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

transcription evidence of truncated *ndh* pseudogenes were also observed in the *U. reniformis* cpDNA [16].

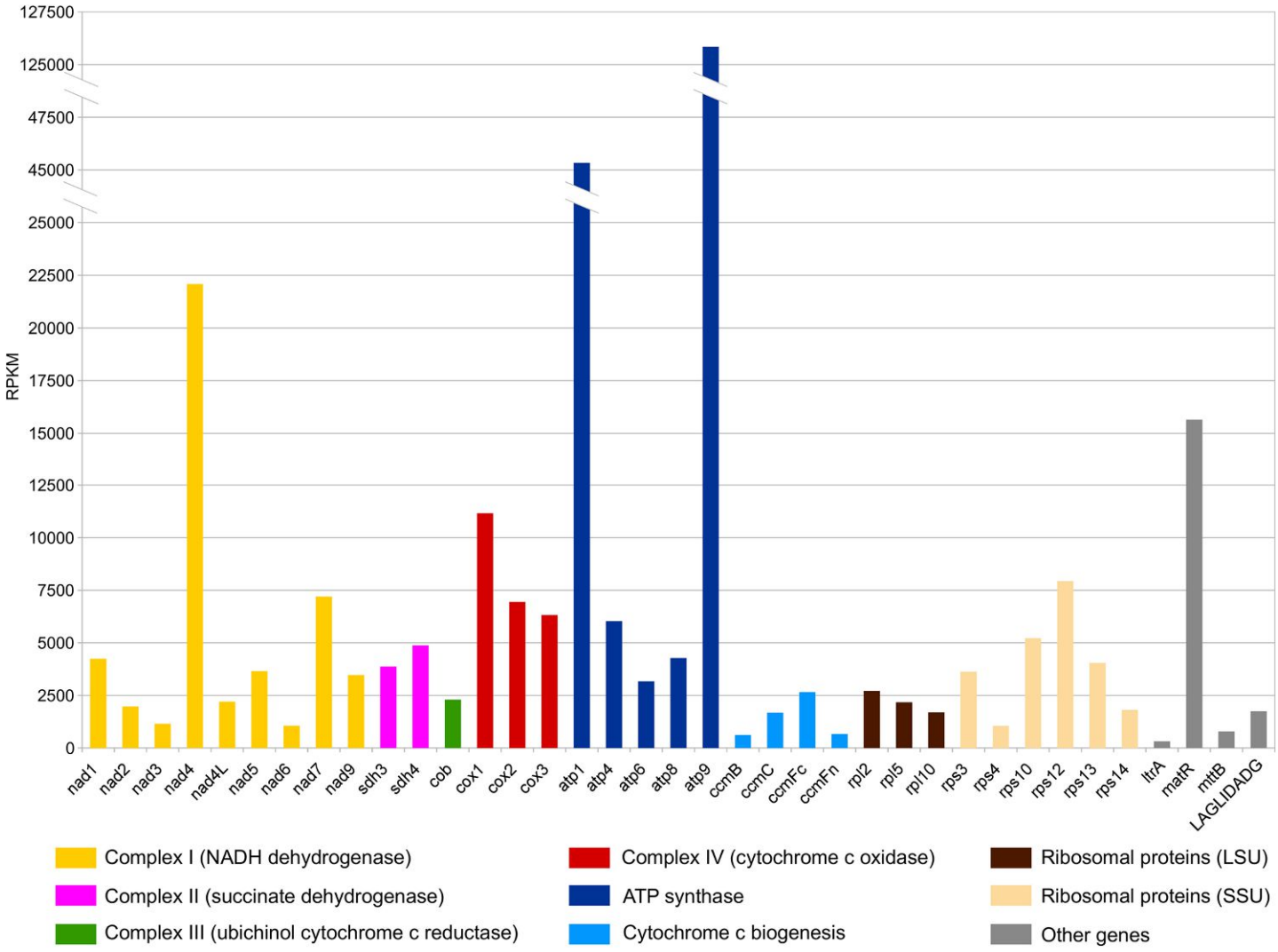


Fig 4. Transcription patterns of the traditional mitochondrial genes by Reads Per Kilobase Million (RPKM).

Mitochondrial RNA editing analysis revealed 69 novel sites

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

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-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 *rps*3 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).

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	
nad1 (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
nad2 (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
nad3	20810	-	CCA	CUA	2	P	L	92

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

(1 site)								
nad4 (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
nad5 (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
nad6 (2 sites)	300266	-	CCA	CUA	2	P	L	93
	300273	-	CCC	UCC	1	P	S	90
nad7 (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
nad9 (3 sites)	497201	+	UCU	UUU	2	S	F	73
	497282	+	CCA	CUA	2	P	L	47
	497708	+	UCU	UUU	2	S	F	78
cob (2 sites)	779528	+	UCU	UUU	2	S	F	92
	779794	+	CUA	UUA	1	L	L	86
cox1 (2 sites)	321763	+ (exon 1)	ACC	ACU	3	T	T	89
	323181	+ (exon 2)	CUG	UUG	1	L	L	96
cox2 (1 site)	514300	+	CUA	UUA	1	L	L	84
cox3 (1 site)	558125	+	UUC	UUU	3	F	F	67
atp4 (2 sites)	285091	+	CCC	CUC	2	P	L	87
	285114	+	CCG	UCG	1	P	S	95
atp6 (2 sites)	810361	-	UCC	UCU	3	S	S	48
	810993	-	CGU	UGU	1	R	C	64
atp8 (1 site)	556640	+	UGC	UGU	3	C	C	84
ccmFc (3 sites)	834747	+ (exon 1)	CAU	UAU	1	H	Y	63
	834901	+ (exon 1)	CCA	CUA	2	P	L	77

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

	836320	+ (exon 2)	CUA	UUA	1	L	L	32
ccmFn (1 site)	488236	-	ACC	AUC	2	T	I	61
rpl5 (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
rpl10 (3 sites)	611859	+	ACC	ACU	3	T	T	35
	611934	+	CGG	UGG	1	R	W	41
	612084	+	UAC	UAU	3	Y	Y	40
matR (7 sites)	102277	+	UCC	UCU	3	S	S	96
	103757	+	CGC	UCG	1	R	S	16
	103771	+	UAC	UAU	3	Y	Y	75
rps3 (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
rps4 (1 site)	267470	-	UCC	UCU	3	S	S	56
rps12 (1 site)	20376	-	CCC	CCU	3	P	P	11
rps14 (2 sites)	21929	-	UCC	UCU	3	S	S	21
	21974	-	CCC	CCU	3	P	P	17

Phylogenetics analysis based on mtDNA supports Lentibulariaceae as a

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

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 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.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

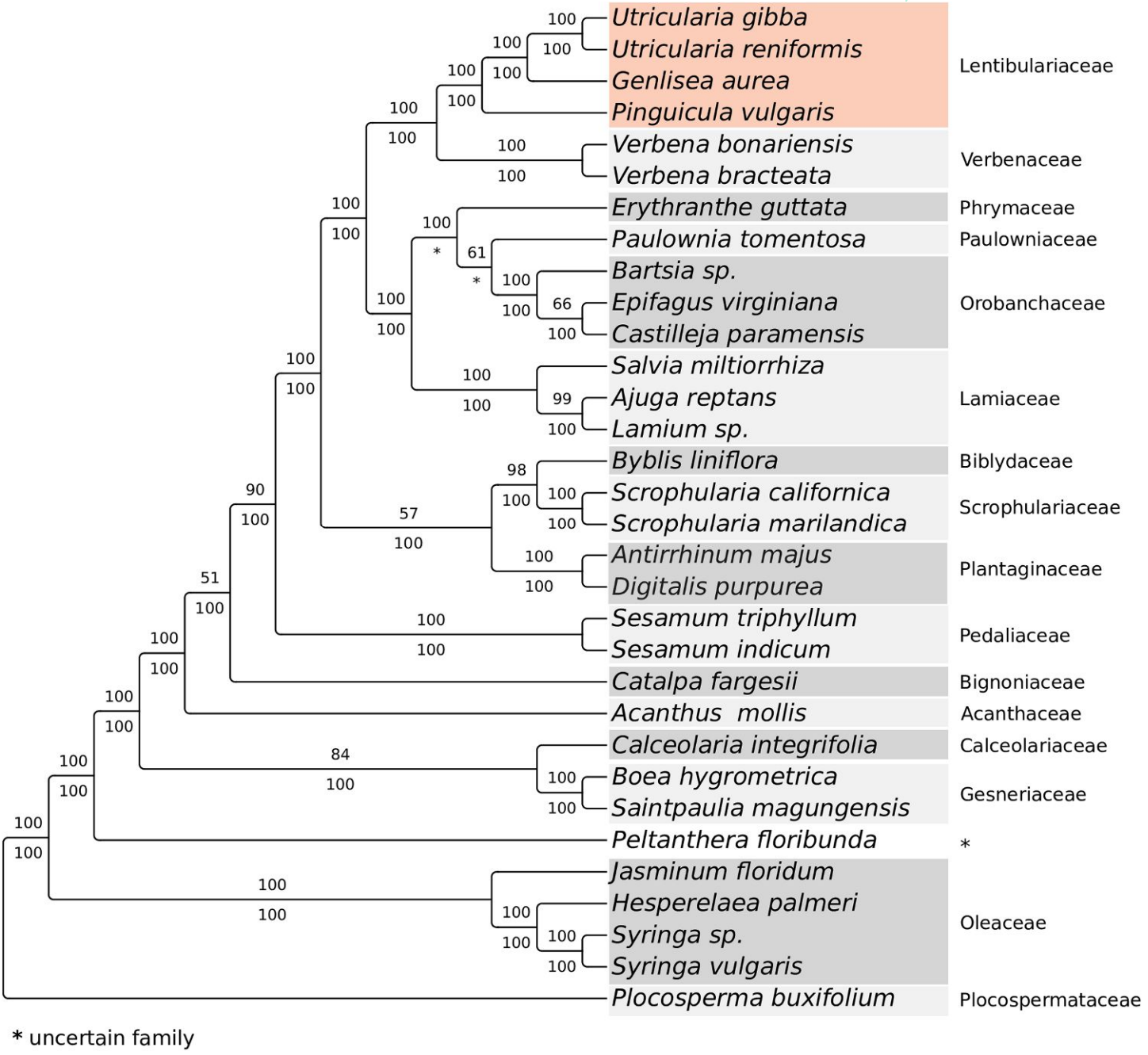


Fig 5. Phylogenetic analysis based on the mitochondrial genes *atp1*, *cox1*, *matR*, *nad5*, *rps3* from 32 different species from the Lamiales order.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

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

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

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*

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

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 polyplodization 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.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

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.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

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

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

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.

Acknowledgements

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

References

1. Adamec L. Mineral nutrition of carnivorous plants: A review. Bot Rev. 1997;63: 273–299.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

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. doi:10.1093/jxb/ern179

3. Juniper B, Robins R, Joel D. The Carnivirorous 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. doi:10.1093/jxb/erp349

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. doi:10.1371/journal.pone.0148979

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. doi: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. doi:10.1007/s00248-011-9832-9

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. doi: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. doi:10.1016/j.syapm.2015.05.006

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. doi:10.1093/aob/mcv172

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

12. Adlassnig W, Peroutka M, Lambers H, Lichtscheidl IK. The Roots of Carnivorous Plants. Plant Soil. 2005;274: 127–140. doi: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. doi: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. doi:10.1073/pnas.1702072114

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; doi:10.1038/nature12132

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. doi:10.1371/journal.pone.0165176

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. doi:10.3732/ajb.1200080

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. doi:10.1073/pnas.0408092101

19. Jobson RW, Albert VA. Molecular Rates Parallel Diversification Contrasts between Carnivorous Plant Sister Lineages. Cladistics. 2002;18: 127–136. doi: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

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

bladderwort (Utricularia), a carnivorous plant with a minimal genome. BMC Plant Biol. 2011;11: 101. doi:10.1186/1471-2229-11-101

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. doi:10.1101/gr.170720.113
22. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9: 357–359. doi:10.1038/nmeth.1923
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. doi:10.1186/1471-2105-10-421
25. Gualberto JM, Mileshina D, Wallet C, Niazi AK, Weber-Lotfi F, Dietrich A. The plant mitochondrial genome: dynamics and maintenance. Biochimie. 2014;100: 107–120. doi:10.1016/j.biochi.2013.09.016
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. doi:10.1089/cmb.2012.0021
27. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics. 2015;31: 3350–3352. doi:10.1093/bioinformatics/btv383
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. doi:10.1093/nar/gkt371

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

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. doi:10.1186/2047-217X-1-18

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. doi:10.1093/bioinformatics/btp352

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. doi:10.1093/molbev/msq029

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. doi:10.1186/1471-2105-11-119

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. doi:10.1093/nar/gkh152

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. doi:10.1093/bioinformatics/btr703

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. doi:10.1093/bioinformatics/bti610

36. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. Bioinforma Oxf Engl. 2004;20: 3252–3255. doi:10.1093/bioinformatics/bth352

37. Magrane M, Consortium U. UniProt Knowledgebase: a hub of integrated protein data. Database. 2011;2011: bar009–bar009. doi:10.1093/database/bar009

38. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 2015;6: 11. doi:10.1186/s13100-015-0041-9

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

39. Lang BF, Laforest M-J, Burger G. Mitochondrial introns: a critical view. Trends Genet. 2007;23: 119–125. doi:10.1016/j.tig.2007.01.006

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. doi:10.1093/nar/gkt289

41. Cheong W-H, Tan Y-C, Yap S-J, Ng K-P. ClicO FS: an interactive web-based service of Circos. Bioinforma Oxf Engl. 2015;31: 3685–3687. doi:10.1093/bioinformatics/btv433

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. doi:10.1101/gr.092759.109

43. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30: 772–780. doi:10.1093/molbev/mst010

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. doi: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. doi: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. doi:10.1093/bioinformatics/btu033

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. doi:10.1186/1471-2105-11-7

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

50. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinforma Oxf Engl. 2011;27: 863–864. doi:10.1093/bioinformatics/btr026

51. Mower JP. PREP-Mt: predictive RNA editor for plant mitochondrial genes. BMC Bioinformatics. 2005;6: 96. doi:10.1186/1471-2105-6-96

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. doi:10.1186/1471-2229-12-61

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. doi:10.1186/1471-2164-15-405

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. doi:10.1371/journal.pbio.1001241

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.

56. Yoshihisa T. Handling tRNA introns, archaeal way and eukaryotic way. Front Genet. 2014;5. doi:10.3389/fgene.2014.00213

57. Treangen TJ, Sommer DD, Angly FE, Koren S, Pop M. Next Generation Sequence Assembly with AMOS. Curr Protoc Bioinforma Ed Board Andreas Baxevanis Al. 2011;CHAPTER: Unit11.8. doi:10.1002/0471250953.bi1108s33

58. Wang D, Rousseau-Gueutin M, Timmis JN. Plastid Sequences Contribute to Some Plant Mitochondrial Genes. Mol Biol Evol. 2012;29: 1707–1711. doi:10.1093/molbev/mss016

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

59. Bruenn JA, Warner BE, Yerramsetty P. Widespread mitovirus sequences in plant genomes. PeerJ. 2015;3: e876. doi:10.7717/peerj.876

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

61. Falkenberg M, Larsson N-G, Gustafsson CM. DNA replication and transcription in mammalian mitochondria. Annu Rev Biochem. 2007;76: 679–699. doi:10.1146/annurev.biochem.76.060305.152028

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. doi:10.1007/s00294-004-0522-8

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. doi:10.1371/journal.pgen.1003265

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. doi: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. doi: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. doi:10.1111/boj.12285

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

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. doi:10.1093/gbe/evv244

69. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 2016; gkw955. doi:10.1093/nar/gkw955

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. doi:10.1093/molbev/msn129

71. Oldenburg DJ, Bendich AJ. Size and Structure of Replicating Mitochondrial DNA in Cultured Tobacco Cells. Plant Cell. 1996;8: 447–461. doi:10.1105/tpc.8.3.447

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. doi:10.1111/nph.12395

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. doi:10.1073/pnas.1421397112

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. doi:10.1186/1471-2148-11-277

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 CoxI gene of Peperomia. J Mol Evol. 1995;41: 563–572.

76. Clivati D, Cordeiro GD, Płachno BJ, de Miranda VFO. Reproductive biology and pollination of Utricularia reniformis A.St.-Hil. (Lentibulariaceae). Plant Biol. 2014;16: 677–682. doi:10.1111/plb.12091

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

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. doi:10.1038/srep09040

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. doi:10.1186/1471-2164-12-279

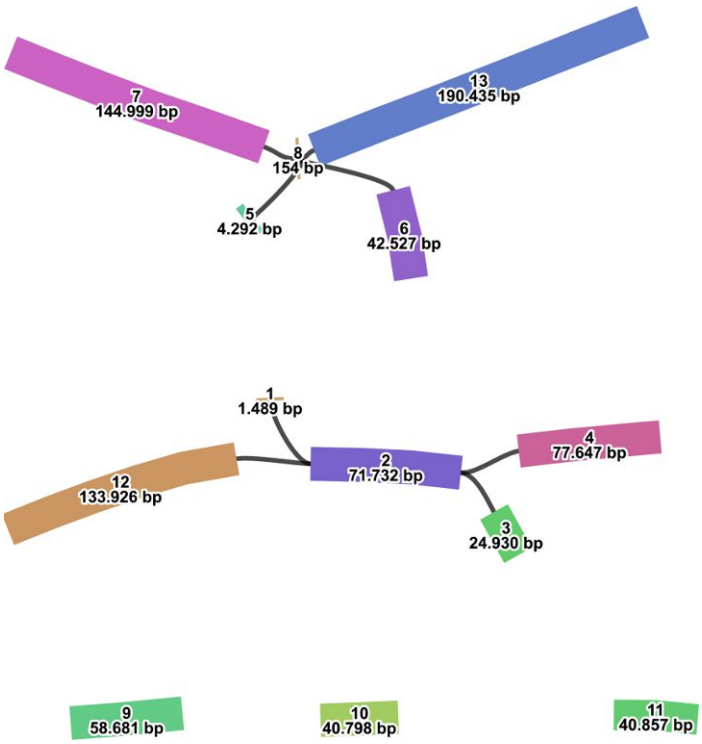
80. Grimes BT, Sisay AK, Carroll HD, Cahoon AB. Deep sequencing of the tobacco mitochondrial transcriptome reveals expressed ORFs and numerous editing sites outside coding regions. BMC Genomics. 2014;15: 31. doi:10.1186/1471-2164-15-31

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

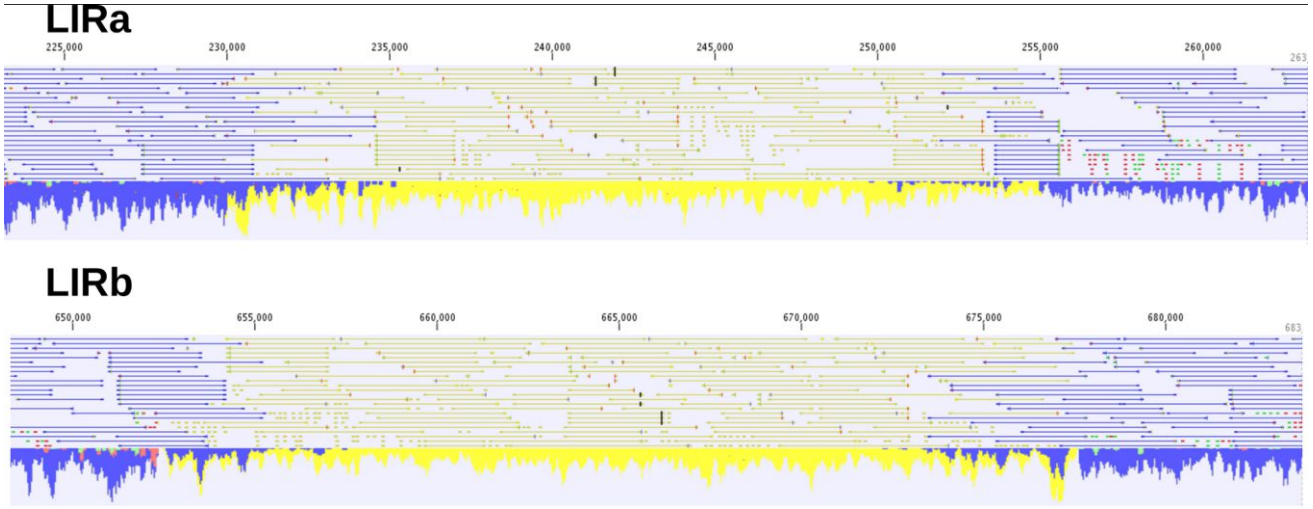
82. Stone JD, Storchova H. The application of RNA-seq to the comprehensive analysis of plant mitochondrial transcriptomes. Mol Genet Genomics MGG. 2015;290: 1–9. doi:10.1007/s00438-014-0905-6

Supplementary information

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br



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.



Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

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 is shown as peaks located in the bottom of each figure, whereas blue represent the repeat borders and yellow the repeat itself.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

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

Species		Author	Family	Accession numbers				
				atp1	*cox1*	*matR*	*nad5*	*rps3*
Utricularia	*gibba*	L.	Lentibulariaceae	KC997783	AY600097	KC997781	KC997780	KC997780
Utricularia	*reniformis*	A.St.-Hil.	Lentibulariaceae					
Pinguicula	*vulgaris*	L.	Lentibulariaceae	GU351090	********	GU351279	GU351489	GU351773
Genlisea	*aurea*	A.St.-Hil.	Lentibulariaceae	KE526711	AY600115	********	KE526711	********
Verbena	*bonariensis*	L.	Verbenaceae	AY741828	********	GU351323	GU351542	GU351849
Verbena	*bracteata*	Cav. ex Lag. & Rodr.	Verbenaceae	HQ385211	********	HQ384798	HQ384808	HQ384946
Erythranthe	*guttata*	(Fisch. DC.) G.L.Nesom	Scrophulariaceae	NC018041	NC018041	NC018041	NC018041	NC018041
Epifagus	*virginiana*	(L.) W.P.C.Barton	Orobanchaceae	EU281004	EU281078	EU281125	********	********
Paulownia	*tomentosa*	Steud.	Scrophulariaceae	AY741826	AJ247592	********	********	HQ384943
Ajuga	*reptans*	L.	Lamiaceae	NC023103	NC023103	NC023103	NC023103	NC023103
Lamium	sp.		Lamiaceae	DQ401312	AJ223428	DQ401385	DQ406871	GU351711
Salvia	*miltiorrhiza*	Bunge	Lamiaceae	NC023209	NC023209	NC023209	NC023209	NC023209
Antirrhinum	*majus*	L.	Scrophulariaceae	GU350962	********	GU351152	GU351339	GU351566
Scrophularia	*marilandica*	L.	Scrophulariaceae	GU351113	********	GU351301	********	GU351808
Scrophularia	*californica*	Cham. & Schltdl.	Scrophulariaceae	HQ385209	********	HQ384796	********	HQ384969
Byblis	*liniflora*	Salisb.	Byblidaceae	GU350976	AY600112	GU351167	GU351356	GU351593
Sesamum	*triphyllum*	Welw. ex Asch.	Pedaliaceae	GU351114	********	GU351302	GU351513	GU351809
Sesamum	*indicum*	L.	Pedaliaceae	AY741827	AJ247598	********	********	HQ384956
Catalpa	*fargesii*	Bureau	Bignoniaceae	********	AJ223411	GU351174	GU351364	GU351606
Acanthus	*mollis*	L.	Acanthaceae	GU350950	********	DQ110297	GU351329	GU351549
Saintpaulia	*magungensis*	E.P.Roberts	Gesneriaceae	GU351107	********	GU351296	GU351507	GU351797
Boea	*hygrometrica*	(Bunge) R.Br.	Gesneriaceae	NC016741	NC016741	NC016741	NC016741	NC016741
Calceolaria	*integrifolia*	L.	Scrophulariaceae	GU350977	********	GU351168	GU351357	GU351595
Syringa	*vulgaris*	L.	Oleaceae	AY741821	********	HQ384794	HQ384811	HQ384983

Syringa	sp.		Oleaceae	********	********	GU351312	GU351525	GU351829
Jasminum	*floridum*	Bunge	Oleaceae	EU280978	EU281051	EU281112	********	********
Plocosperma	*buxifolium*	Benth.	Loganiaceae	HQ385208	********	HQ384793	HQ384807	HQ384988
Hesperelaea	*palmeri*	A.Gray	Oleaceae	NC031323	NC031323	NC031323	NC031323	NC031323
Castilleja	*paramensis*	F.González & Pabón-Mora	Orobanchaceae	NC031806	NC031806	NC031806	NC031806	NC031806
Bartsia	sp.	Benth.	Orobanchaceae	KP940487	KP940490	KP940485	KP940485	KP940486
Digitalis	*purpurea*	L.	Scrophulariaceae	EU280962	AJ223415	EU281103	********	********
Peltanthera	*floribunda*	Benth.	Peltanthraceae	GU351080	********	GU351271	GU351480	HQ384981

******** *missing data*

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

S2 Table. List of all identified open reading frames and detected signal peptide and transmembrane domains.

Tabela localizada no link: https://doi.org/10.1371/journal.pone.0180484.s004

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

Tabela localizada no link: https://doi.org/10.1371/journal.pone.0180484.s005

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

Tabela localizada no link: https://doi.org/10.1371/journal.pone.0180484.s006

PLOS ONE

RESEARCH ARTICLE

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

Saura R. Silva[1☯], Danillo O. Alvarenga[2☯], Yani Aranguren[3¤a], Helen A. Penha[2], Camila C. Fernandes[2], Daniel G. Pinheiro[2], Marcos T. Oliveira[2], Todd P. Michael[4¤b], Vitor F. O. Miranda[3]*, Alessandro M. Varani[2]*

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

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 transcriptionaly 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

CAPÍTULO 4

Comparative genomic analysis of *Genlisea* (corkscrew plants – Lentibulariaceae) chloroplast genomes reveals an increasing loss of the *ndh* genes

Artigo publicado

Silva, S.R., Michael, T.P., Meer, E.J., Pinheiro, D.G., Varani, A.M., and Miranda, V.F.O. (2018). Comparative genomic analysis of Genlisea (corkscrew plants—Lentibulariaceae) chloroplast genomes reveals an increasing loss of the ndh genes. PLOS ONE 13, e0190321.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

Comparative genomic analysis of *Genlisea* (corkscrew plants – Lentibulariaceae) chloroplast genomes reveals an increasing loss of the *ndh* genes

Saura R. Silva[1], Todd P. Michael[2], Elliott J. Meer[3], Daniel G. Pinheiro[4], Alessandro M. Varani[4]*, Vitor F.O. Miranda[5]*

[1.]Universidade Estadual Paulista (Unesp), Botucatu, Instituto de Biociências, São Paulo, Brazil.

[2.] J. Craig Venter Institute, La Jolla, CA, United States of America.

[3] 10X Genomics, Pleasanton, California, United States of America.

[4.]Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal, Departamento de Tecnologia, São Paulo, Brazil.

[5.]Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal, Departamento de Biologia Aplicada à Agropecuária, São Paulo, Brazil.

* Corresponding Authors

Email: vmiranda@fcav.unesp.br (VFOM); amvarani@fcav.unesp.br (AMV)

Short title: Comparative genomic analysis of *Genlisea* chloroplast genomes

Abstract

In the carnivorous plant family Lentibulariaceae, all three genome compartments (nuclear, chloroplast, and mitochondria) have some of the highest rates of nucleotide substitutions across angiosperms. While the genera *Genlisea* and *Utricularia* have the smallest known flowering plant nuclear genomes, the chloroplast genomes (cpDNA) are mostly structurally conserved except for deletion and/or pseudogenization of the NAD(P)H-dehydrogenase complex (*ndh*) genes known to be involved in stress conditions of low light or CO_2 concentrations. In order to determine how the cpDNA are changing, and to better understand the evolutionary history within the *Genlisea* genus, we sequenced, assembled and analyzed complete cpDNA from six species (*G. aurea, G. filiformis, G. pygmaea, G. repens, G. tuberosa* and *G. violacea*) together with the publicly available *G. margaretae* cpDNA. In general, the cpDNA structure among the analyzed *Genlisea* species is highly similar. However, we found that the plastidial *ndh* genes underwent a progressive process of degradation similar to the other terrestrial Lentibulariaceae cpDNA analyzed to date, but in contrast to the aquatic species. Contrary to current thinking that the terrestrial environment is a more stressful environment and thus requiring the *ndh* genes, we provide evidence that in the Lentibulariaceae the terrestrial forms have progressive loss while the aquatic forms have the eleven plastidial *ndh* genes intact. Therefore, the Lentibulariaceae system provides an important opportunity to understand the evolutionary forces that govern the transition to an aquatic environment and may provide insight into how plants manage water stress at a genome scale.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

Introduction

The carnivorous plant *Genlisea* has astonished scientists for many years. Charles Darwin was seduced by this "remarkable genus" which he described at the end of his book Insectivorous Plants [1]. The genus *Genlisea* A.St.-Hil. belongs to the carnivorous family Lentibulariaceae together with genera *Utricularia* and *Pinguicula* [2]. *Genlisea* encompass about 30 species that inhabit open areas with nutrient-poor soil distributed in tropical Africa and the Neotropics (eight of twenty species are endemic to Brazil) [3-6]. *Genlisea* are small, rootless, terrestrial herbs commonly known as "corkscrew plants" due to Y-shaped-underground leaves that are twisted helically and have the ability to capture, digest and absorb prey [7,8]. It is difficult to distinguish different species based solely on the vegetative forms due to *Genlisea* having a diverse set of intraspecific phenotypes. Despite Darwin's early interest however, *Genlisea* remains poorly studied due to cultivation challenges, and being found in isolated and remote habitats [9].

Genlisea and *Utricularia* have one of the highest nucleotide substitution rates across all three genome compartments (nucleus, chloroplast, mitochondria) in comparison to other angiosperms [10-12] with previous studies revealing that both genera have an exclusive mutation in the mitochondrial cytochrome *c* oxidase gene (*cox1*) [13]. These mutations lead to a proton pumping change and, during oxidative phosphorylation, cause electrons to leak into the mitochondria, generating reactive oxygen species (ROS). It is proposed that the ROS can damage DNA, which produces breaks in the double helix structure, leading to point mutations [14-16]. On an evolutionary timescale this potential increase in ROS could explain the high nucleotide substitution rate, the process of genome miniaturization [17], and a high diversification of morphological traits [14].

Previous systematic studies were carried out using morphological traits, mainly based on capsule dehiscence together with trap, pollen, flower characteristics [4,18,19] and molecular markers

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

from the three plastidial loci: *trnK/matK*, *rps16* and *trnQ-rps16*. Phylogenies based on these markers suggested two major groups within *Genlisea*: the subgenus *Genlisea,* comprising the sections *Genlisea*, *Africanae* and *Recurvatae*, and the subgenus *Tayloria*. However, due to the recent discovery of new species, unresolved clades and possible cryptic species, the evolutionary history of *Genlisea* requires further investigation [4,18].

Chloroplast genome (cpDNA) sequencing and analysis of different species provides a powerful tool to dissect out the evolutionary history of plant genera. The highly conserved structure and gene content of the cpDNA enable plant evolution and phylogeny studies [20]. Structural rearrangements, gene decay and loss are often observed in cpDNA and inform a plethora of evolutionary relationships among different taxa. For example, plastid gene loss in the most extreme cases is linked to lineages with heterotrophic nutrition, such as parasitic [21] and mycoheterotrophic plants [22].

One of the gene losses that occur in such plants is related to the NAD(P)H-dehydrogenase complex (*ndh*) genes. The *ndh* genes consist of eleven (11) subunits in the cpDNA (*ndhA*, *B*, *C*, *D*, *E*, *F*, *G*, *H*, *I*, *J* and *K*) that encodes, along with nuclear genes, the thylakoid NAD(P)H dehydrogenase complex [23]. This complex is involved in photosynthesis, the photosynthetic response and stress acclimation [24], and has been hypothesized to be related to the transition to terrestrial habitats [14,16]. The eleven *ndh* subunit genes are present in the aquatic Lentibulariaceae species, but are lost in the terrestrial *Utricularia* species, suggesting that the evolutive history of the *ndh* genes among the *Utricularia* lineages followed an opposite trend, and that the *ndh* function may be dispensable in terrestrial forms [25]. However the presence and absence of the *ndh* genes remain to be established in *Genlisea* species. Therefore, the *ndh* genes in the cpDNA can provide a valuable resource for the understanding of *Genlisea* evolution and how these genes can be associated to the habitats.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

To better understand the evolutionary history of the *Genlisea* genus and explore the role of *ndh* gene loss, we sequenced, assembled six chloroplast genomes and, together with the published *G. margaretae* cpDNA, carried out a full analysis. These seven *Genlisea* species represent both subgenera *Tayloria* (*G. violacea*) and *Genlisea (G. aurea, G. filiformis, G. pygmaea, G. repens, G. tuberosa* and *G. margaretae)*. We found that the chloroplast genome is highly similar across species, but unlike their aquatic relatives, in the terrestrial *Genlisea* species the *ndh* genes are deleted, fragmented or pseudogenized. These findings not only add to the understanding of terrestrial heterotrophic plants, and their cpDNA evolution, but also provide an important opportunity to understand the evolutionary forces that govern the transition to an aquatic environment at a genome scale.

Material and Methods

Plant samples, preparation and sequencing

Fresh photosynthetic leaves of *Genlisea* species were sampled from natural populations and also cultivated and stored in silica gel. Total DNA was extracted using modified CTAB protocol and concentration, integrity and purity was assessed using Nanodrop™ spectrophotometer (Thermo Scientific) and Agilent 2100 Bioanalyzer (Agilent Genomics). Herbarium vouchers are deposited at the Herbarium JABU at Universidade Estadual Paulista (UNESP/FCA; ICMBio/ MMA for collecting permits SISBIO #26938 and #48516) (S1 Table).

The paired-end libraries were prepared using Illumina library preparation manufacturer's protocol and genomic DNA was sequenced using Illumina Miseq Platform (Illumina, San Diego, CA).

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

The publicly available *Genlisea aurea* DNA sequencing data was obtained from raw genome database SRA (accession number SRR916071) that was previously used for nuclear genome assembly [26].

Assembly and annotation

The quality of raw reads was assessed by FastQC [27]. Removal of adapters from both ends and trimming to obtain high quality reads were performed using the Platanus_trim (v.1.0.7) [28] with Phred quality score of >30 and length cutoff of 150bp for 300bp reads, 100bp for 150bp reads, 80bp for 100bp reads and 50bp for 75bp reads (see S1 Table). In addition, to exclude nuclear and mitochondrial genomes, the *Genlisea* species chloroplast genome paired end reads were extracted by mapping all raw reads to the reference cpDNA *Utricularia gibba* (NC021449) with Bowtie2 (v.2.2.3) [29] (i.e. –very-sensitive-local with –N 1 modification). Then this selected set of reads was assembled using Spades (v.3.7.1) [30] software with default parameters. Uncertain regions, such as IR junctions, were picked out from published Lentibulariaceae species (*U. gibba* and *Genlisea margaretae* [NC025652.1]) to extend the length using iteration method with MITObim (v.1.8) [31]. As the assembly usually collapses the inverted repeats in one single contig, the IR region of some species were manually inverted and duplicated to integrate the whole chloroplast genome using BioEdit (v.7.2) [32]. High quality filtered reads were mapped back using Bowtie2 (i.e. –very-sensitive; end-to-end) in Geneious Pro (v.10.2.3) [33] to each assembled chloroplast genome to confirm assembly accuracy quality and repeat region junctions (S1 Table; S1 Fig).

The annotation of the chloroplast genomes were performed using Dual Organellar GenoMe Annotator (DOGMA) [34] with manual corrections for start and stop codons and intron boundaries by comparison to homologous genes from sequenced chloroplast of *Utricularia gibba, U. reniformis*

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

(NC029719.2) and *Genlisea margaretae*. The tRNA genes were also verified with ARAGORN [35] and tRNAscan-SE [36]. The codon usage was calculated using CodonW (v1.4.4) [37].The circular chloroplast genome maps were drawn using OrganellarGenome DRAW tool (OGDRAW) [38].

To determine whether a gene was a pseudogene, fragmented or deleted gene, Blastn and Blastx searches were performed using other chloroplasts as reference, such as *U. gibba*, and a pseudogene was characterized according to the absence of start and/or stop codon, frameshift and genes with more than 20% of the coding region in comparison to other related species. The genes that are considered as fragmented were any group of nucleotides that had at least >25bp and had correspondence to position and blastn and tblastx alignment with the complete gene.

Repeat identification

REPuter [39] was used to search both direct and palindrome sequences, with a minimum repeat size of > 30bp and a sequence identity greater than 90% (parameters: repfind –f –p –l 30 –h 3 –best 10,000). Microsatellites for mono-, di-, tri-, penta- and hexanucleotides were detected using the Perl script MISA [40]. The established parameters were performed according with Silva *et al.* [25].

Identity and variation analyses

The chloroplast genomes were aligned using MAFFT (v.7) [41] with FFT-NS-2 parameters and identity comparisons between chloroplasts were conducted with mVISTA program [42].

Average p-distances were calculated to determine genetic divergence between *Genlisea* species and the number of phylogenetically informative characters (PICs) for each plastome gene, intergenic spacers, introns and pseudogenes using PAUP (v.4b10) [43]. Nonparametric Spearman

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

test was used to test for correlation between PICs and average p-distances between sequences of *Genlisea* species.

Phylogenomic analyses

Phylogenetic analyses were performed to different partitions by using the whole chloroplast genome sequence, protein coding genes, intergenic spacers, LSC (Large Single Copy), SSC (Small Single Copy), IR (Inverted Repeat) and *ndh* genes. For *ndh* phylogenetic tree, pseudogenes and fragments of deleted genes of at least 25bp were considered (S2-S3 Tables). Previously published Lentibulariaceae chloroplast genomes were included (*Utricularia foliosa* [KY025562], *U. gibba* [NC021449], *U. macrorhiza* [NC025653], *U. reniformis* [NC029719.2] and *Pinguicula ehlersiae* [NC023463]) and *Tectona grandis* (Lamiaceae) [NC020098], *Sesamum indicum* (Pedaliaceae) [NC016433] and *Tanaecium tetranolobum* (Bignoniaceae) [NC027955] cpDNA used as outgroup.

The alignments were conducted using MAFFT (v.7) [41] and the evolutionary model (best-of-fit) that was most appropriate for all the data according with corrected Akaike Information Criterion (AICc), calculated using jModelTest [44].

Maximum parsimony criterion was performed using PAUP (v.4b10) [43] with heuristic searches of 2,000 replicates and bootstrap analysis with 1,000 pseudoreplicates, both using the tree bisection-reconnection branch swapping (TBR) and random addition of sequences. The probabilistic analysis was conducted using RAxML (v.8) [45] for maximum likelihood (ML) using the default parameters with bootstrap support of 1,000 pseudoreplicates and MrBayes (v.3) [46] for Bayesian inference with 5×10^5 generations with two runs and four chains following the substitution matrix assessed as mentioned above. Both analyses were performed on CIPRES Science Gateway website [47] and cladograms were edited with the program TreeGraph2 (beta v.2.0) [48].

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

In an attempt to test also the phylogenetic signal of *ndh* genes in *Genlisea* lineages, we created a matrix with 22 characters. The characters 1 to 11 we codified if each *ndh* gene (*ndhA, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ*, and *ndhK*) was absent (0) or present (1) and the characters 12 to 22 if each gene was pseudogenized (0), decayed (1) or complete (2) for each of the eleven *ndh* genes and carried out a parsimony analysis (S2-S4 Tables). The consensual tree (strict consensus) of most parsimonious trees was presented and evolution of *ndh* genes was traced using both matrix and chloroplast phylogenomic tree described above with PAUP (v.4b10) with ACCTRAN optimization [43].

Results

Genome content and organization of the six *Genlisea* chloroplast genomes

The cpDNA of *Genlisea* ranged from 140,010 bp (*G. aurea*) to the largest plastome of the sequenced species with 143,416 bp (*G. violacea*) (Fig 1, Table 1). All six chloroplast genomes display a quadripartite structure, which consists of a pair of inverted repeats (IR) separated by a Large Single Copy (LSC) and a Small Single Copy (SSC) region. The plastomes contain 103 unique genes, including 69 protein-coding genes, 30 tRNAs, 4 rRNAs and the average GC content was 38.57±0.08%. Fourteen genes contain a single intron, such as *atpF*, *petB, petD, rpl16, rpl2, rpoC1, rps12, rps16, trnA*-UGC, *trnG*-UCC, *trnI*-GAU, *trnK*-UUU, *trnL*-UAA and *trnV*-UAC, while *clpP* and *ycf3* have two introns. The *orf42, orf56*, and *ycf68* genes of the IR region are pseudogenes due to lack of start and/or stop codons (Table 2)

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

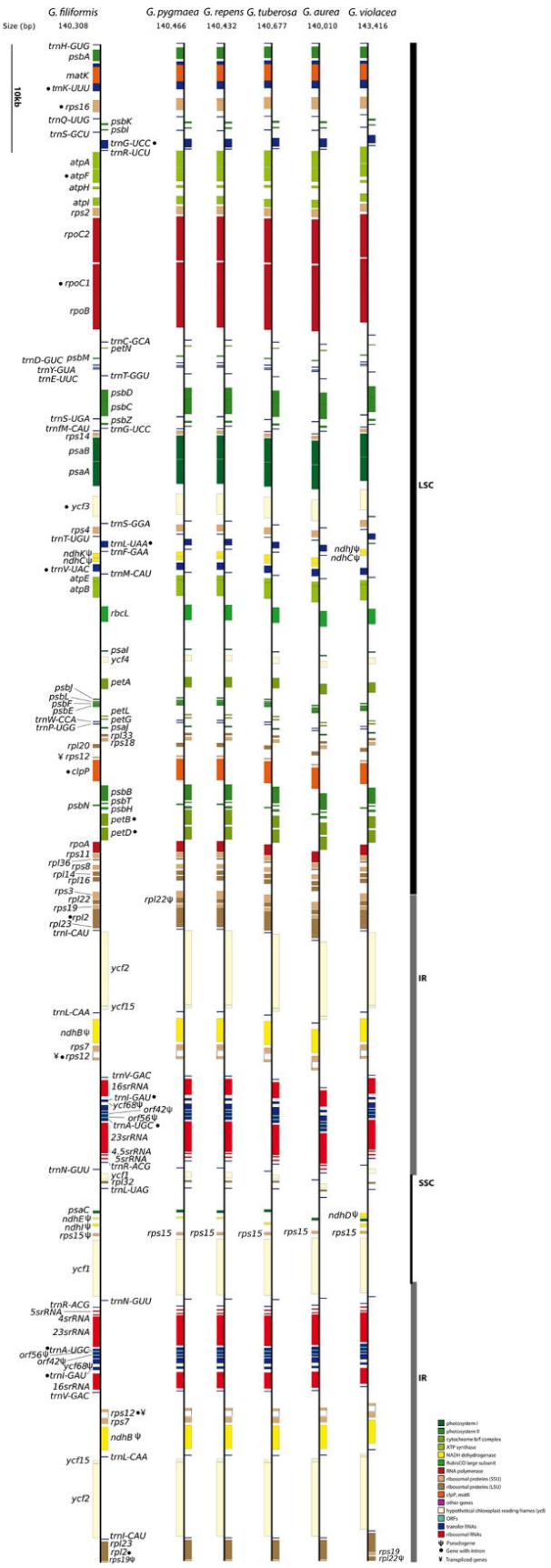


Fig 1. Physical chloroplast genome maps of six assembled *Genlisea* species. The chloroplast genome is showed with the genes colorized according to the functional classes for each species. The genes shown on the right side of each cpDNA map are transcribed clockwise, whereas gene on the left side are transcribed counter clockwise. The symbol Ψ after the gene name indicates that is a pseudogene, • the presence of introns and ¥ denotes transpliced genes. Large single copy (LSC), inverted repeats (IR) and single copy repeat (SSC) are represented by the black and grey bars.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

Table 1. Summary of assembly data for *Genlisea* plastomes (for details about sequencing data see S1 Table).

Species	cpDNA size (bp)	LSC size (bp)	SSC size (bp)	IRs size (bp)	GC content (%)	GenBank accession number
Genlisea aurea	140,010	80,653	9,419	24,969	38.5	MF593121
G. filiformis	140,308	79,754	10,316	25,119	38.7	MF593122
G. pygmaea	140,466	79,888	10,346	25,116	38.6	MF593123
G. repens	140,432	79,875	10,325	25,116	38.5	MF593124
G. tuberosa	140,677	80,347	10,462	24,934	38.5	MF593125
G. violacea	143,416	81,089	10,969	25,679	38.6	MF593126

Table 2. Genes in the six *Genlisea* chloroplast genomes (except *G. margaretae*).

Category of genes	Group of gene	Name of the gene
Self-replication	Ribosomal RNA genes (rRNAs)	4.5S rRNA (2x), 5S rRNA (2x), 16S rRNA (2x), 23S rRNA (2x)
	Transfer RNA genes (tRNAs)	trnH-GUG, trnK-UUU●, trnQ-UUG, trnS-GCU, trnG-UCC●, trnR-UCU, trnC-GCA, trnD-GUC, trnY-GUA, trnE-UUC, trnT-GGU, trnS-UGA, trnG-UCC●, trnfM-CAU, trnS-GGA, trnT-UGU, trnL-UAA●, trnF-GAA, trnV-UAC●, trnM-CAU, trnW-CCA, trnP-UGG, trnI-CAU, trnL-CAA (2x), trnV-GAC (2x), trnI-GAU ● (2x), trnA-UGC ● (2x), trnR-ACG (2x), trnN-GUU (2x), trnL-UAG
	Small subunit of ribosomal protein	rps2, rps3, rps4, rps7 (2x), rps8, rps11, rps12● (2x) ¥, rps14, rps15**, rps16●, rps18, rps19***
	Large subunit of ribosomal protein	rpl2● (2x), rpl14, rpl16●, rpl20, rpl22*, rpl23 (2x), rpl32, rpl33, rpl36
	RNA polymerase subunit	rpoA, rpoB, rpoC1●, rpoC2
Photosynthesis	NADH dehydrogenase	All are ψ or deleted (see Fig 1 and 4 for each *Genlisea* species)
	Photosystem I	psaA, psaB, psaC, psaI, psaJ, ycf3● ,ycf4
	Photosystem II	psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ
	Cytochrome b/f complex	petA, petB●, petD●, petG, petL, petN
	ATP synthase	atpA, atpB, atpE, atpF●, atpH, atpI
	Rubisco large subunit	rbcL
Other genes	Translation initiation factor	infA
	Maturase	matK
	Protease	clpP●
	Envelope membrane protein	cemA
	Subunit of acetyl-CoA-carboxylase	accD

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

	c-type cytochrome synthesis gene	ccsA
Unknown function	Conserved hypothetical protein	ycf1, ycf2 (2x), ycf15 (x2), ycf68 ψ (2x), orf56 ψ (2x), orf42 ψ (2x)

● Gene with intron; ψ Pseudogenes; ¥ Transpliced genes. * One of duplicated gene is partial in *G. violacea* and is pseudogene in *G. pygmaea*; **Pseudogene in *G. filiformis*; *** Duplicated gene in *G. violacea*.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

Overall, all the *Genlisea* cpDNAs are highly conserved in organization and structure (Fig 1), except for the *ndh* genes that are pseudogenized, fragmented or deleted in all *Genlisea* plastomes. In addition, *G. violacea* has slightly expanded IR/LSC boundary genes with the duplication of intact *rps19* gene and *rpl22* as pseudogene (see S2 Fig), and the *rps15* and *rpl22* are present as pseudogenes in *G. filiformis* and *G pygmaea*, respectively (Fig 1).

Repeats in the *Genlisea* plastomes

Repeats were divided in three categories: tandem, direct and palindromic (Fig 2). The great majority of the repeats across the chloroplast genomes were simple sequence repeats (SSRs) of lengths between 7 and 20 bp. An average of 210 repeats were detected in the six chloroplast genomes, 6.80% (69 repeats) of which are direct repeats, 5.80% (59 repeats) were palindromic repeats, and 87.40% (888 repeats) tandem repeats (Fig 2; S5 Table). Moreover, most of the repeats are located in the intergenic regions (39.40%), followed by coding (36.60%) and intronic regions (14.90%). Few repeats were found in tRNA, rRNA and pseudogenes regions (9.10%). The majority of microsatellites in all species are A/T mono- and dinucleotides. There are few tetra- and pentanucleotide and one hexanucleotide in *G. pygmaea*. Among all chloroplast genomes, 41 repeat regions (4%) were shared by all analyzed *Genlisea* species (S5 Table).

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

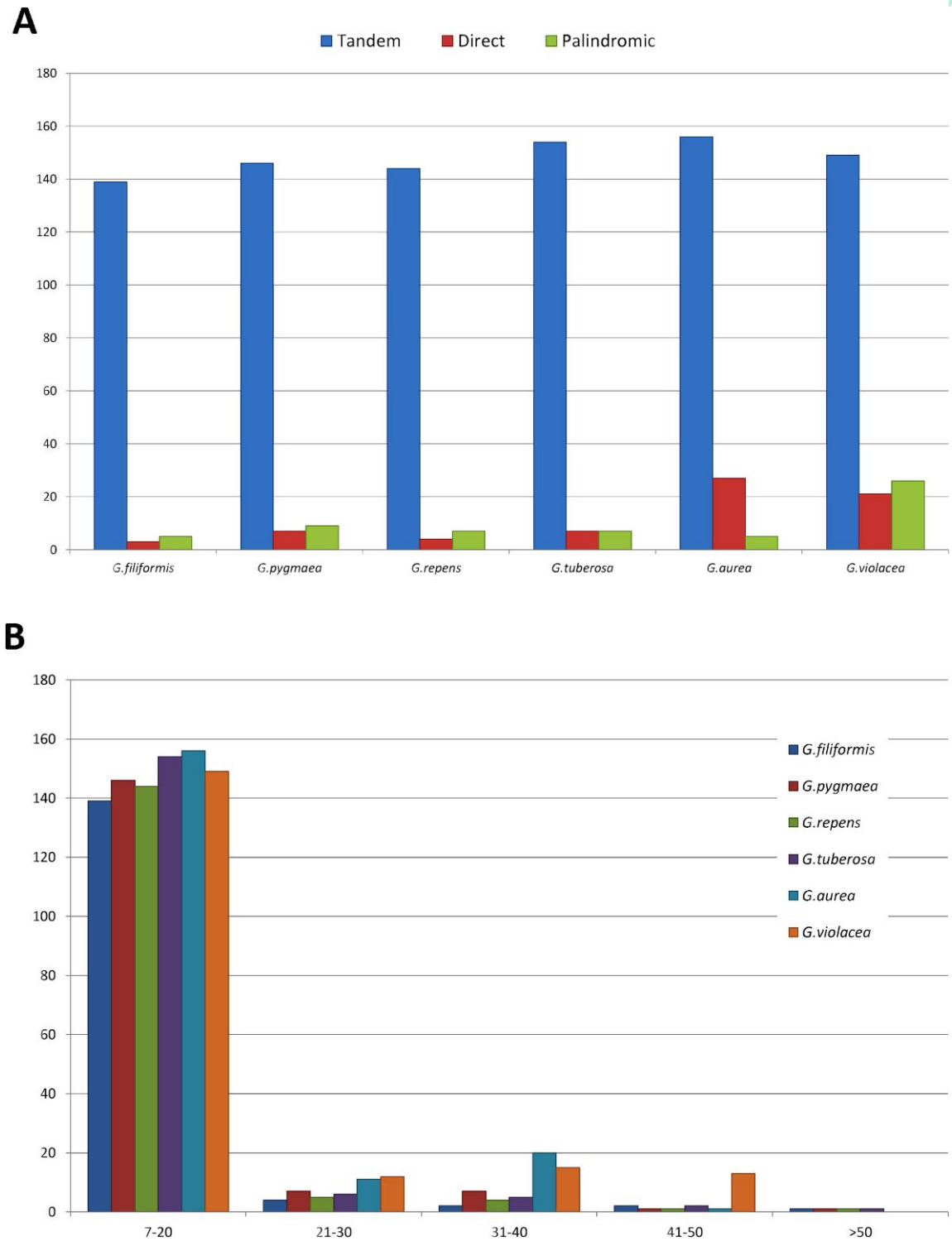


Fig 2. Analysis of repeats in *Genlisea* chloroplast genomes. (A) Quantity of tandem, direct and palindromic repeats of each species. (B) Quantity of repeats by length.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

Molecular markers identification

Genome wide comparison allowed the identification of genomic regions that could be used as possible phylogenetic markers to reconstruct the evolutionary history of the genus. A positive correlation between the percentage of variable sites, given by p-distance, and phylogenetically informative characters (PICs) (ρ=0.583, P<0.001; S3 Fig) were identified. Thus, the PICs of each coding and non-coding alignment region were used to identify potential regions for phylogenetics and population studies.

The divergence hotspot analysis given by p-distance and phylogenetically informative characters (S6 Table) revealed that the most informative regions for phylogenetic analyses were non-coding DNA regions such as intergenic spacers and introns (Fig 3; S4 Fig). Moreover, the p-distance between *Genlisea* and *Pinguicula* was 0.043, *Genlisea* and *Utricularia* 0.057 and between *Genlisea* species was of 0.032. The overall p-distance between *G. repens* and *G. pygmaea*, the most related species in this study, was 0.001. Phylogenetically informative characters suggest that the top ten regions with the greatest number of PICs are three genes (*ycf1*, *matK* and *rpoC2*), two introns (*rpl16-intron*, *trnK-intron*) and five intergenic regions (*trnK-rps16*, *rps12-clpP*, *petA-psbJ, rpl20-rps12*, *rps12-trnV*) (S4 Fig; S6 Table).

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

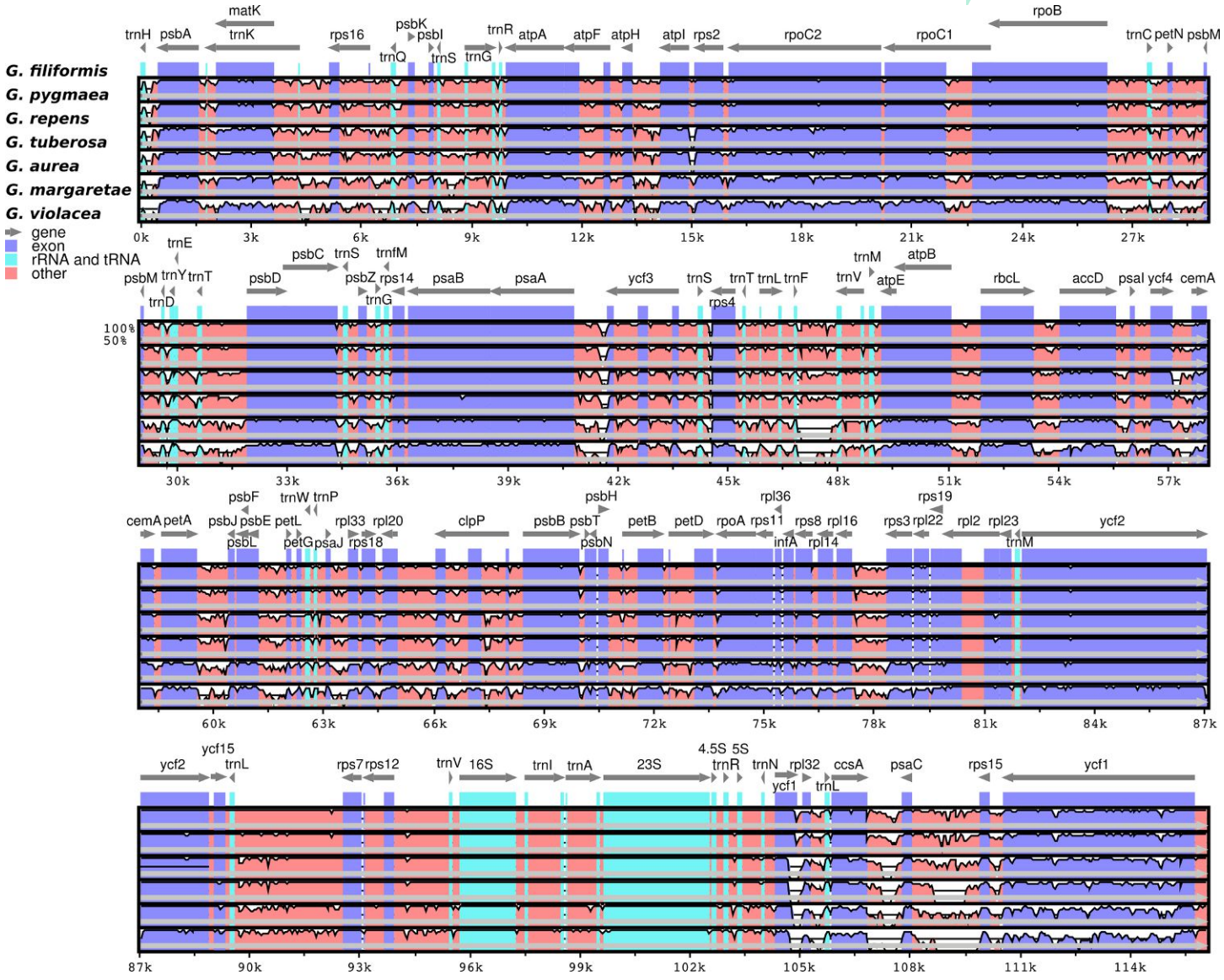


Fig 3. Sequence identity plots for the six assembled *Genlisea* species and previously published *G. margaretae*.

Phylogenomic analysis

Regarding the Lentibulariaceae, the topologies were totally congruent for all chloroplast dataset partitions (LSC, IR, SSC, coding regions, intergenic spacers and introns; S5 Fig). The whole chloroplast alignment resulted in 178,161 characters of which 21,687 are informative sites (Table 3). The most parsimony, Bayesian (BS) and maximum likelihood (ML) trees are highly congruent with

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

very high support (ML bootstraps and posterior probabilities mostly 100) and support Lentibulariaceae as a monophyletic group, and *Genlisea–Utricularia* as sister clade with *Pinguicula*. When all branch lengths for each cladogram are visualized, the IR tree depicts very short branches (S5 Fig), resulting from the lowest proportion of variable sites (9%; Table 3). These results support that the *Genlisea* genus is monophyletic and its topology follows previous phylogenetic studies [18]: subgenus *Tayloria* (represented by *G. violacea*) as a sister clade to subgenus *Genlisea* (*G. margaretae*, *G. filiformis*, *G. pygmaea*, *G. repens*, *G. tuberosa* and *G. aurea*) (Fig 4). Moreover, the phylogenetic analyses based on the *ndh* genes partition, which treated each nucleotide ordinarily as a character, reveals a topology totally congruent to the trees resulting from other partitions and whole plastomes (Fig 4; S5 Fig). Also, when the processes that could be involved in the *ndh* degeneration (pseudogenization and decay) were codified in a multistate character matrix (see S2-S4 Tables); the resultant tree (Fig 5B) was mostly congruent with the nucleotide-by-nucleotide tree (Fig 4-5A).

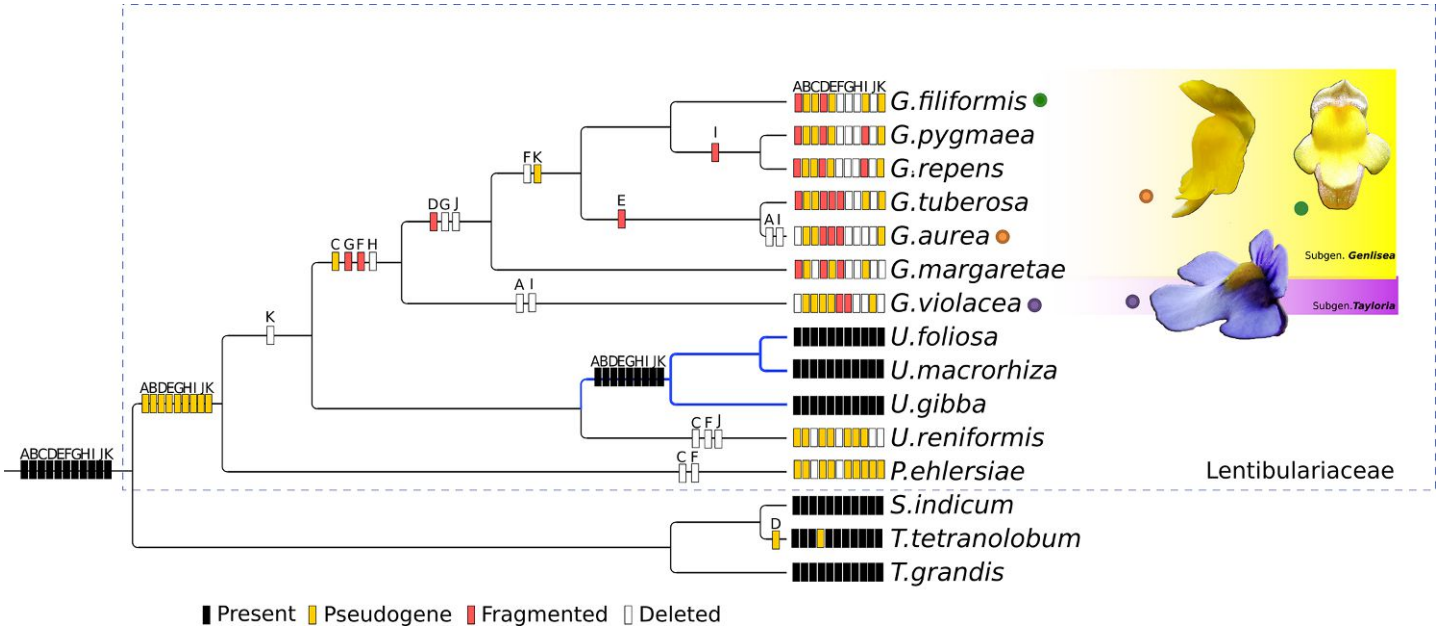


Fig 4. Phylogenomics of whole chloroplasts of *Genlisea* species and *ndh* genes evolution. The boxes indicate the *ndhA*, *ndhB*, *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ* and *ndhK* genes. Black boxes denote intact genes, yellow

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

boxes pseudogenized genes, red boxes fragmented and white boxes indicate deleted genes. Blue lines indicate the aquatic *Utricularia* species with complete *ndh* repertoire. Numbers of support values are all 100% for Bayesian inference, maximum likelihood and maximum parsimony bootstrap, except for outgroup clade *S. indicum* and *T. tetranolobum* with parsimony bootstrap value of 85%.

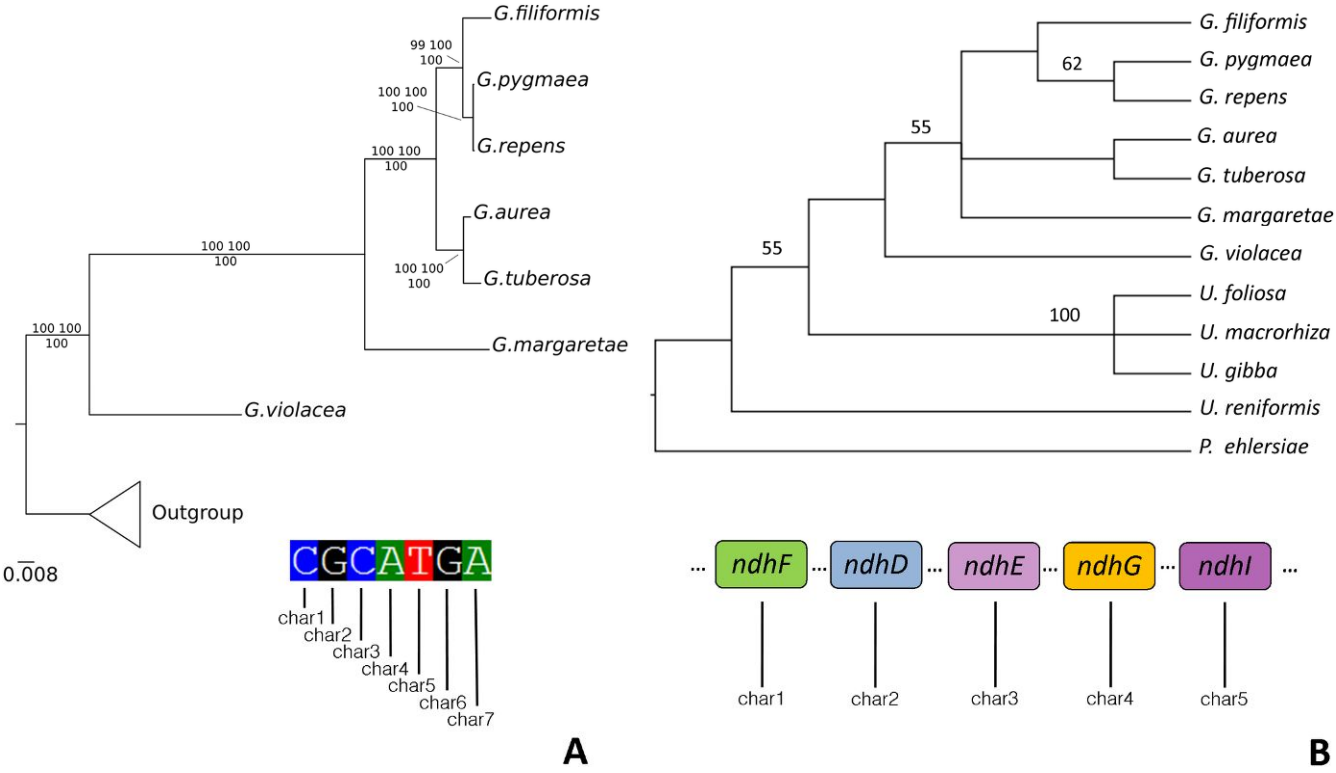


Fig 5. Phylogenetic hypothesis based on *ndh* sequences. **A.** Analyses based on *ndh* sites (nucleotide-by-nucleotide). In this analysis, each nucleotide was used as one character (e.g. char1, char2, char3) **B**. Strict consensus of the two most parsimonious trees (33 steps; IC= 0.70; IR = 0.78) based on the matrix codified for *ndh* patterns. In this analysis, each *ndh* gene was applied to two characters: one codified as absent (state 0) or present (state 1) (characters 1 to 11) and other codified as pseudogenized (state 0), decayed (state 1), complete (state 2) and inapplicable (state "-", when the gene is deleted) (characters 12 to 22). For details see S2, S3, and S4 Tables.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

1

Table 3. Datasets and phylogenetic statistics for each *Genlisea* cpDNA partition.

	Whole chloroplast	**LSC**	**SSC**	**IR**	**Protein coding**	**Intergenic spacers**	**Introns**	***ndh* genes**
Alignment (bp)	178,161	99,235	20,156	28,636	67,437	46,068	15,090	9,462
Overall GC content (%) – Only *Genlisea* species	38.5	36.4	30.5	43.5	40.4	32.5	36.1	35.7
Overall GC content (%) – *Genlisea* + outgroup	38.1	36.1	31.3	43.1	40.4	32	35.9	35.2
Variable sites (%)	40,427 (22%)	27,508 (27%)	7,753 (38%)	2,752 (9%)	12,817 (19%)	15,502 (33%)	4,057 (26%)	1,944 (20%)
Informative sites (%)	21,687 (12%)	15,218 (15%)	4,275 (21%)	1,140 (4%)	6,909 (10%)	8,616 (18%)	2,360 (15%)	535 (6%)
Consistency index (CI)	0.856	0.852	0.837	0.922	0.845	0.855	0.847	0.976
Retention index (RI)	0.875	0.876	0.847	0.919	0.868	0.875	0.881	0.948
Model of substitution (AICc)	GTR+G+I	GTR+G+I	TVM+G+I	TVM+G+I	GTR+G+I	TVM+G+I	GTR+G+I	TVM+G

Discussion

Chloroplast genomes are a powerful tool to understand the evolutionary forces acting on a species because their structure and sequence are highly constrained across flowering plants. The carnivorous plant family Lentibulariaceae has been shown to have a high rate of nucleotide substitution in all three genome compartments, including the chloroplast genome [10]. In this study we describe seven *Genlisea* cpDNA including both subgenera within carnivorous plant *Genlisea*: subgen. *Tayloria* (*G. violacea*) and subgen. *Genlisea* (*G. aurea, G. filiformis, G. pygmaea*, *G. repens*, *G. tuberosa* and *G. margaretae*).

The *Genlisea* cpDNA have typical quadripartite structure with a similar gene repertoire, as previously described for other Lentibulariaceae [25,49,50]. However, we do find that the *ndh* genes are deleted, fragmented or pseudogenized, which provides new insight into the evolutionary trajectory of *Genlisea* as well as the terrestrial forms of the Lentibulariaceae.

Even though cpDNAs are structurally conserved, changes in genome composition have been identified in many species of angiosperms [51] and also in some gnetophytes [52]. These variations are principally due to the expansion and contraction of IR and SSC regions [53] and gene loss and duplicated genes in IR/SC or IR/LSC boundaries [54]. Among the six cpDNAs described in this study and the previously published *Genlisea margaretae* cpDNA [50], *G. violacea* proved to be the most divergent from the other *Genlisea* species with possible IR expansion that includes duplication of *rps19* gene and partial duplication of *rpl22* gene. In addition, *G. filiformis* and *G. pygmaea* showed pseudogenization of *rps15* and *rpl22* genes, respectively. However, the absence of these genes is observed in other angiosperms.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

For instance, the *rpl22* gene was loss in several cpDNA, such as legumes [55,56], *Gossypium* [57], *Citrus* [58], *Castanea* [59], *Quercus* [60] and *Passiflora* species [61]. Moreover, some studies suggest that there is strong evidence that the *rpl22* gene has been transferred to the nucleus in some angiosperms [56,60].

The GC content among seed plant plastomes ranges between 34–40% and, comparing each cpDNA region, the SSC is the one with the lowest GC content [51]. For the *Genlisea* cpDNAs, we also found that the SSC had the lowest GC content (31.3%). One explanation for the SSC having the lowest GC content is that this region is susceptible to nucleotide substitutions, which is consistent with the high level of nucleotide variation (38%) we observed, compared to other cpDNA regions (Table 3).

The codon usage in *Genlisea* plastomes is similar to that reported for other Lentibulariaceae family cpDNA. Approximately 19,268 codons represent the coding repertoire of the protein coding regions (S7 Table). Codons frequency that ends with A and T have higher usage than G and C ending codons. For all plastomes the most frequent codon was Leucine (with approximately 1,989; 10.35%), whereas the least frequent was Cysteine (approximately 210–1.10%).

The identification of phylogenetically informative characters (PICs; including the parsimony informative characters) is an important procedure for evaluating characters with phylogenetic signal. Indeed, the PICs are represented by the synapomorphies [62,63] rather than nucleotide changes lacking phylogenetic signal. In this context, the results presented in this study support that the cpDNA is a powerful source of information for phylogenetic inferences. For *Genlisea*, two previous phylogenetic studies employed the cpDNA loci *trnK*/*matK* and *rps16* [4,18]. Our study suggests that other cpDNA regions (such as *ycf1*,

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

rpl20-rps12, *rpoC2*) have more PICs and consequently have higher phylogenetic signal than previously considered sequences used to assess phylogenies and populations studies.

According to the Consortium for the Barcode of Life's (CBOL), further studies are necessary to define the best DNA sequences for DNA barcoding of plants [64,65]. As many plants have poor resolution at the population level, previous studies have proposed using combinations of loci (as *matK, rbcL, trnH-psbA*), suggesting that no unique region exists [65,66]. However, a recent study suggested a single region in *ycf1* gene [67] could be used as a better barcode. Our PIC and divergence analysis corroborate usage of *ycf1* and/or *matK* for barcoding purposes, since *ycf1* is the first PIC classification and *matK* is the sixth (S6 Table).

Widely used in plant genotyping [68,69], SSRs are an important source of genetic variation that can be used for species discrimination, population structure and genetic diversity [69]. Similarly to our findings for *Genlisea* species, previous studies on cpDNA SSRs of Lentibulariaceae [70], reported that the chloroplast genomes have a large number of SSRs [25,50]; similarly, we find many SSRs across *Genlisea* species. Long repeats, represented by direct repeats and palindromic repeats can cause hairpin structures, which are associated with recombination, and can contribute significantly to rearranged gene order and addition of polymorphism [71,72]. In the evaluated *Genlisea* species, the long repeats were mainly found in non-coding regions, which is consistent with most angiosperms [73]. And, although long repeats are rare in Lentibulariaceae [50], both the smallest (*G. aurea*) and the largest chloroplast genomes (*G. violacea*) have a high number of direct repeats and palindromic repeats. In the *G. violacea* chloroplast genome, regions with palindromic repeats are found near the LSC/IR junctions, suggesting they could be contributing to IR expansion.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

In *Utricularia reniformis* [25], repeat hotspots seem to be associated with *ndh* gene degradation, since some repeats regions are close to *ndh* genes. However, in *Genlisea* the repeats are dispersed over the cpDNA indicating that, for this genus, there is no relationship between the repeats and *ndh* pseudogenization. This observation suggests that, unlike *Utricularia*, different evolutionary processes are acting in the *Genlisea ndh* loci.

Different dataset partitions (IR, LSC, SSC, coding regions, intergenic spacers and introns; S5 Fig), recovered the same tree topology for Lentibulariaceae with high clade support. Indeed, all datasets contained a considerable percentage of informative characters, thus phylogenetic signal can be found along the whole *Genlisea* cpDNA.

The eleven *ndh* genes present in all *Genlisea* species are pseudogenized, decayed or even deleted (Fig 5; S4 Table). *ndh* genes losses have been found a few times in other taxa and are attributed to heterotrophic plants [23], some conifers [52], orchids [74], and other species of Lentibulariaceae [25,49,50]. And even with the remarkable degradation of *ndh* genes, the nucleotide composition of *ndh* still provides sufficient signal for a phylogenetic analysis (Fig 4-5). As such, the topology of *ndh* phylogenetic tree reveals the cladogenetic separation of different subgenera (*Tayloria* and *Genlisea*) and resolution of all *Genlisea* species (Fig 5). While in some orchids [74] the *ndh* losses seem to have no relation with taxonomy, and environment where these species are found, in Lentibulariaceae the *ndh* genes appear to have been maintained in aquatic taxa [25,49,50].

When the *ndh* gene events (arisen, pseudogenization, decay or deletion) are traced in the total evidence (entire plastomes) phylogenetic analysis (Fig 4), we can verify a different scenario when comparing *Genlisea* lineages to *Pinguicula* and *Utricularia* lineages. The terrestrial taxa *Pinguicula* (represented by the *P. ehlersiae*) and *U. reniformis* have most *ndh*

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

genes as pseudogenized or deleted. Interestingly, the clade represented by the aquatic species of *Utricularia* (*U. foliosa*, *U. macrorhiza*, and *U. gibba*) has gained, probably as independent (or not) reversion events (Fig 4; clade denoted with blue lines), the almost entire *ndh* repertoire. We have previously shown that aquatic species of *Utricularia* have maintained and conserved *ndh* genes [25]. The *ndh* genes activity appears dispensable under favorable conditions, as pointed out by transcriptomic studies [75] and verified in knock-out mutants [76-78]. But, episodes of abiotic stress can impact terrestrial habitats and, according to Ruhlman *et al.* [75], appear to be the cause of retention of *ndh* genes. Nonetheless, our phylogenetic hypothesis shows that Lentibulariaceae follows an opposite trend, since terrestrial species of *Pinguicula*, *U. reniformis* and all seven *Genlisea* possess degenerated *ndh* genes and the aquatic species of *Utricularia*, on the other hand, display a conserved *ndh* repertoire. Moreover, it is important to emphasize that the aquatic environment also provides a stressful habitat for plants, since these habitats can present low carbon and light availability, anoxia, wave exposure, significant restrictions to sexual reproduction, and sometimes also osmotic stress and limited nutrient supply [79]. Thus, the complete recovery of all eleven genes for the aquatic *Utricularia* supports the hypothesis that the *ndh* genes are conserved in stressful habitats.

The trend of decay and deletion of the *ndh* genes, represented within the different lineages of *Genlisea* is remarkable. The *Genlisea* clade presented the highest concentration of fragmented and deleted *ndh* genes, when compared to *Utricularia* and *Pinguicula* species (Fig 4). In an attempt to phylogenetically test this tendency of *ndh* genes to degrade in *Genlisea* lineages, we codified the state (present, pseudogenized, decayed or deleted; S2-S4 Tables) for each of eleven plastidial *ndh* genes and carried out a parsimony analysis. The

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

consensual topology of both most parsimonious trees (Fig 5A-B) also supports this hypothesis when compared with the total evidence tree (Fig 4).

As in most Lentibulariaceae cpDNA, the loss of *ndh* genes does not seem to affect plant fitness despite the harsh environmental conditions common for the carnivorous habit [25,50]. However, as seen in the present study, it has been reported that in terrestrial species most of *ndh* genes were lost for Lentibulariaceae species (Fig 1).

Silva *et al.* [80] identified several pseudogenes of plastid origin in *U. reniformis* mtDNA. For instance, the presence of the *ndhJ-ndhK-ndhC* loci in the mtDNA supports the hypothesis of lateral transfer since these genes are absent in the cpDNA [80]. Similar translocation of *ndh* genes from the plastome to the mitochondrial genome was also suggested to the Epidendroideae orchid *Erycina pusilla* [81]. According to this study, other than the *ndh* genes could be transferred to mtDNA, since more than 76% of the cpDNA genome was transferred into the mtDNA genome of *E. pusilla* and the largest cpDNA insertion into the mtDNA genome in this species was 12kb.

In addition to the transfer of plastid genes to the mtDNA, cpDNA genes can also be transferred to the nuclear genome [82]. With the *G. aurea* nuclear genome published [26], we performed blastn and tblastn searches of all plastidial *ndh* genes subunits and none of these genes were also found in nuclear assembled scaffolds. However, one cannot discard the idea that these genes are present in the mtDNA. This hypothesis has to be further investigated since the mitochondrial genome is not available [26].

Studies have pointed out the function of *ndh* genes for modulating ROS in chloroplasts [23]. Plants with high expression of *ndh* genes also have an increasing concentration of ROS, which can lead to the cell death [83]. Assuming that terrestrial

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

environments are less stressful than aquatic ones [79], the presence of complete *ndh* repertoire is understandable for aquatic species of Lentibulariaceae, since these genes are important for ROS modulation in the presence of their high respiratory rates. However, only the aquatic Lentibulariaceae species of *Utricularia* have had their respiration rates measured [84]. More chloroplast genomes from the *Genlisea* and *Utricularia* lineages are required to test this hypothesis. But the oxidant activities of ROS are well known for DNA [14,85,16] and it is not difficult to suppose their deleterious action even in genomes from different compartments.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

Conclusions

Here we report the chloroplast genome of six *Genlisea* species of both subgenera: *Tayloria* and *Genlisea*. These genomes were compared with the previously published *G. margaretae* cpDNA, showing that they are very similar in content and have the same gene order and quadripartite structure. Phylogenomic analysis showed that using coding regions, non-coding regions and even decayed *ndh* sequences it is possible to obtain the evolutionary history with great congruence, recovering with high support the position of assessed taxa in *Genlisea* genus and Lentibulariaceae family. Importantly, we corroborate previous observations that distinct from the aquatic taxa of Lentibulariaceae, the terrestrial *Genlisea* chloroplast genomes showed a pseudogenization and a progressive degradation of *ndh* genes, as reported for other Lentibulariaceae. In summary, we propose that the Lentibulariaceae system provides an important opportunity to understand the evolutionary forces that govern the transition to an aquatic environment, and may provide insight into how plants manage water stress at a genome scale. These findings may have implications for engineering crop species for better water stress tolerance, both too much and too little water.

Acknowledgements

The authors thank Dr. T.S. Balbuena and Dr. J.A.M. de Souza for their careful reading of our manuscript and their helpful comments and suggestions.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

References

1. Darwin C. Insectivorous plants. D. Appleton and Company, London; 1899.
2. Jobson RW, Playford J, Cameron KM, Albert VA. Molecular Phylogenetics of Lentibulariaceae Inferred from Plastid *rps*16 Intron and *trn*L-F DNA Sequences: Implications for Character Evolution and Biogeography. Syst. Bot. 2003;28: 157–171.
3. BFG - The Brazil Flora Group. Growing knowledge: an overview of Seed Plant diversity in Brazil. Rodriguésia, 2015;66: 1085–1113.
4. Fleischmann A, Schäferhoff B, Heubl G, Rivadavia F, Barthlott W, Müller KF. Phylogenetics and character evolution in the carnivorous plant genus *Genlisea* A. St.-Hil. (Lentibulariaceae). Mol. Phylogenet. Evol. 2010;56: 768–83.
5. Fleischmann A, Rivadavia F, Gonella PM, Heubl G. A revision of *Genlisea* subgenus *Tayloria* (Lentibulariaceae). Syst. Bot. 2011;40: 1–40.
6. Miranda VFO, Menezes CG, Silva SR, Díaz YCA., Rivadavia F. Lentibulariaceae in Lista de Espécies da Flora do Brasil. Jard. Botânico do Rio Janeiro 12 Set 2015 Available from: http://floradobrasil.jbrj.gov.br/reflora/floradobrasil/FB146.
7. Barthlott W, Porembski, S, Fischer, E, Gemmel, B. First protozoa-trapping plant found. Nature 1998; 392:447.
8. Płachno BJ., Adamus K, Faber J, Kozłowski J. Feeding behaviour of carnivorous *Genlisea* plants in the laboratory. Acta Bot. Gall. 2005;152: 159–164.
9. Fleischmann A. Monograph of the Genus *Genlisea*. Poole, Dorset, England: Redfern Natural History Productions; 2012.
10. Jobson R, Albert VA. Molecular Rates Parallel Diversification Contrasts between Carnivorous Plant Sister Lineages. Cladistics 2002;18: 127–136.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

11. Müller K, Borsch T, Legendre L, Porembski S, Theisen I, Barthlott W. Evolution of Carnivory in Lentibulariaceae and the Lamiales. Plant Biol. 2004;6: 477-490.

12. Müller KF, Borsch T, Legendre L, Porembski S, Barthlott W. Recent progress in understanding the evolution of carnivorous Lentibulariaceae (Lamiales). Plant Biol. (Stuttg). 2006;8: 748–57.

13. 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. U.S.A. 2004;101: 18064–18068. doi:10.1073/pnas.0408092101.

14. Albert VA, Jobson RW, Michael TP, Taylor DJ. The carnivorous bladderwort (*Utricularia*, Lentibulariaceae): A system inflates. J. Exp. Bot. 2010;61: 5–9. doi:10.1093/jxb/erp349.

15. Ibarra-Laclette E, Albert VA, Pérez-Torres CA, Zamudio-Hernández F, Ortega-Estrada MDJ, 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. doi:10.1186/1471-2229-11-101.

16. Laakkonen L, Jobson RW, Albert VA. A new model for the evolution of carnivory in the bladderwort plant (*Utricularia*): Adaptive changes in cytochrome c oxidase (COX) provide respiratory power. Plant Biol. 2006;8: 758–764. doi:10.1055/s-2006-924459.

17. Greilhuber J, Borsch T, Müller K, Worberg A, Porembski S, Barthlott W. Smallest angiosperm genomes found in Lentibulariaceae, with chromosomes of bacterial size. Plant Biol. 2006;8: 770–777. doi:10.1055/s-2006-924101.

18. Fleischmann A, Michael TP, Rivadavia F, Sousa A, Wang W, Temsch EM, et al. Evolution of genome size and chromosome number in the carnivorous plant genus

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

Genlisea (Lentibulariaceae), with a new estimate of the minimum genome size in angiosperms. Ann. Bot. 2014;114: 1651–1663. doi:10.1093/aob/mcu189.

19. Fromm-Trinta E. Revisão das espécies do gênero *Genlisea* St.-Hil. -Lentibulariaceae- das regiões sudeste e sul do Brasil. Rodriguésia 1979;31: 17–139.
20. Daniell H, Lin C-S, Yu M, Chang W-J. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 2016;17: 134. doi:10.1186/s13059-016-1004-2.
21. Bungard RA. Photosynthetic evolution in parasitic plants: Insight from the chloroplast genome. BioEssays 2004;26: 235–247. doi:10.1002/bies.10405.
22. Schelkunov MI, Shtratnikova VY, Nuraliev MS, Selosse MA, Penin AA, Logacheva MD. Exploring the limits for reduction of plastid genomes: A case study of the mycoheterotrophic orchids *Epipogium aphyllum* and *Epipogium roseum*. Genome Biol. Evol. 2015;7: 1179–1191. doi:10.1093/gbe/evv019.
23. Martín M, Sabater B. Plastid *ndh* genes in plant evolution. Plant Physiol. Biochem. 2010;48: 636–645. doi:10.1016/j.plaphy.2010.04.009.
24. Peng L, Yamamoto H, Shikanai T. Structure and biogenesis of the chloroplast NAD(P)H dehydrogenase complex. Biochim. Biophys. Acta - Bioenerg. 2011;1807: 945–953. doi:10.1016/j.bbabio.2010.10.015.
25. 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: 1–29. doi:10.1371/journal.pone.0165176.
26. Leushkin EV, Sutormin RA, Nabieva ER, Penin AA, Kondrashov AS., Logacheva MD

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

The miniature genome of a carnivorous plant *Genlisea aurea* contains a low number of genes and short non-coding sequences. BMC Genomics 2013;14: 476. doi:10.1186/1471-2164-14-476.

27. Andrews S. FastQC: A quality control tool for high throughput sequence data. 2010. Available from: http://www.bioinformatics.babraham.ac.uk/projects/Fastqc/.
28. 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. doi:10.1101/gr.170720.113.
29. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9: 357–359. doi:10.1038/nmeth.1923.
30. 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. 2012;19: 455–477. doi:10.1089/cmb.2012.0021.
31. 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. doi:10.1093/nar/gkt371.
32. Hall T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 1999;41: 95–98.
33. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 2012;28: 1647–1649. doi:10.1093/bioinformatics/bts199.
34. Wyman SK, Jansen, RK, Boore, JL. Automatic annotation of organellar genomes with

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

DOGMA. Bioinformatics 2004;20: 3252–3255. doi:10.1093/bioinformatics/bth352.

35. Laslett D, Canback, B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 2004;32: 11–16. doi:10.1093/nar/gkh152.
36. Lowe TM, Eddy SR. TRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1996;25: 955–964. doi:10.1093/nar/25.5.0955.
37. Peden JF. CodonW. 2005 Available from: http://codonw.sourceforge.net/
38. 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: W575-81. doi:10.1093/nar/gkt289.
39. Kurtz S, Schleiermacher C. REPuter: Fast computation of maximal repeats in complete genomes. Bioinformatics 1999;15: 426–427. doi:10.1093/bioinformatics/15.5.426.
40. Thiel T, Michalek W, Varshney R, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor. Appl. Genet. 2003;106: 411–422. doi:10.1007/s00122-002-1031-0.
41. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7 : Improvements in Performance and Usability Article Fast Track. 2013;30: 772–780. doi:10.1093/molbev/mst010.
42. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: Computational tools for comparative genomics. Nucleic Acids Res. 2004;32: W273-W279. doi:10.1093/nar/gkh458.
43. Swofford DL. PAUP* Phylogenetic Analysis Using Parsimony * (and other methods). version 4.0. 2002. Sinauer Assoc. Sunderland, Massachusetts.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

44. Santorum JM, Darriba D, Taboada GL, Posada D. Jmodeltest.Org: Selection of Nucleotide Substitution Models on the Cloud. Bioinformatics, 2014;30: 1310–1311. doi:10.1093/bioinformatics/btu032.

45. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014;30: 1312–1313. doi:10.1093/bioinformatics/btu033.

46. Huelsenbeck JP, Ronquist F. MrBayes: Bayesian inference of phylogeny. Bioinformatics 2001;17: 754–755. doi:DOI: 10.1093/bioinformatics/17.8.754.

47. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. in 2010 Gateway Computing Environments Workshop, GCE 2010. 2010. doi:10.1109/GCE.2010.5676129.

48. Stöver BC, Müller KF TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. BMC Bioinformatics 2010;11: 7. doi:10.1186/1471-2105-11-7.

49. Silva SR, Pinheiro DG, Meer EJ, Michael TP, Varani AM, Miranda VFO. The complete chloroplast genome sequence of the leafy bladderwort, *Utricularia foliosa* L. (Lentibulariaceae). Conserv. Genet. Resour. 2017;9: 213–216. doi:10.1007/s12686-016-0653-5.

50. Wicke S, Schaferhoff B, dePamphilis CW, Muller KF Disproportional Plastome-Wide Increase of Substitution Rates and Relaxed Purifying Selection in Genes of Carnivorous Lentibulariaceae. Mol. Biol. Evol. 2014;31: 529–545. doi:10.1093/molbev/mst261.

51. Jansen RK, Ruhlman TA. Genomics of Chloroplasts and Mitochondria. In: Bock R, Knoop V. Advances in Photosynthesis and Respiration. Netherlands:Springer; 2012. 475.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

52. Wu CS, Wang YN, Hsu CY, Lin CP, Chaw SM. Loss of different inverted repeat copies from the chloroplast genomes of pinaceae and cupressophytes and influence of heterotachy on the evaluation of gymnosperm phylogeny. Genome Biol. Evol. 2011;3: 1284–1295. doi:10.1093/gbe/evr095.

53. Dugas DV, Hernandez D, Koenen EJM, Schwarz E, Straub S, Hughes CE, et al. Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions, and accelerated rate of evolution in *clp*P. Sci. Rep. 2015;5: 16958. doi:10.1038/srep16958.

54. Goulding SE, Olmstead RG, Morden CW, Wolfe KH. Ebb and flow of the chloroplast inverted repeat. Mol. Gen. Genet. 1996;252: 195–206. doi:10.1007/s004389670022.

55. Doyle JJ, Doyle JL, Palmer JD. Multiple Independent Losses of Two Genes and One Intron from Legume Chloroplast Genomes. Syst. Bot. 1995;20: 272–294. doi:10.2307/2419496.

56. Gantt JS, Baldauf SL, Calie PJ, Weeden NF, Palmer JD. Transfer of *rpl*22 to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. EMBO J. 1991;10: 3073–3078.

57. Lee S-B, Kaittanis C, Jansen RK, Hostetler JB, Tallon LJ, Town CD, et al. The complete chloroplast genome sequence of Gossypium hirsutum: organization and phylogenetic relationships to other angiosperms. BMC Genomics 2006;7: 61. doi:10.1186/1471-2164-7-61.

58. Bausher MG, Singh ND, Lee S-B, Jansen RK, Daniell H. The complete chloroplast genome sequence of *Citrus sinensis* (L.) Osbeck var "Ridge Pineapple": organization and phylogenetic relationships to other angiosperms. BMC Plant Biol. 2006;6: 21. doi:10.1186/1471-2229-6-21.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

59. Jansen RK, Saski C, Lee SB, Hansen AK, Daniell H. Complete plastid genome sequences of three rosids (*Castanea*, *Prunus*, *Theobroma*): Evidence for at least two independent transfers of *rpl*22 to the nucleus. Mol. Biol. Evol. 2011;28: 835–847. doi:10.1093/molbev/msq261.

60. Yang Y, Zhou T, Duan D, Yang J, Feng L, Zhao G. Comparative Analysis of the Complete Chloroplast Genomes of Five *Quercus* Species. Front. Plant Sci. 2016;7: 959. doi:10.3389/fpls.2016.00959.

61. Cauz-Santos LA, Munhoz CF, Rodde N, Cauet S, Santos AA, Penha HA, et al. The Chloroplast Genome of *Passiflora edulis* (Passifloraceae) Assembled from Long Sequence Reads: Structural Organization and Phylogenomic Studies in Malpighiales . Front. Plant Sci. 2017;8: 334. doi:10.3389/fpls.2017.00334.

62. Fitch WM. On the Problem of Discovering the Most Parsimonious Tree. The American Naturalist 1977;111: 223–57.

63. Saitou N, Nei M. The number of nucleotides required to determine the branching order of three species, with special reference to the human-chimpanzee-gorilla divergence. J. Mol. Evol. 1986;24: 189–204. doi:10.1007/BF02099966.

64. Hollingsworth, PM Refining the DNA barcode for land plants. Proc. Natl. Acad. Sci. U.S.A. 2011;108: 19451–2. doi:10.1073/pnas.1116812108.

65. CBOL Plant Working. A DNA Barcode for Land Plants. Proc. Natl. Acad. Sci. U.S.A. 2009;106: 12794–97. doi:10.1073/pnas.0905845106.

66. Pang X, Liu C, Shi L, Liu R, Liang D, Li H, et al. Utility of the *trn*H-*psb*A Intergenic Spacer Region and Its Combinations as Plant DNA Barcodes: A Meta-Analysis. PLoS One 2012;7: e48833. doi:10.1371/journal.pone.0048833.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

67. Dong W, Xu C, Li C, Su J, Zuo Y, Shi S, et al. *ycf*1, the most promising plastid DNA barcode of land plants. 2015:1–5. doi:10.1038/srep08348.

68. Batley J. Plant Genotyping. Methods Mol. Biol. 2015;1245: 101–18. doi:10.1007/978-1-4939-1966-6_8.

69. Zalapa JE, Cuevas H, Zhu H, Steffan S, Senalik D, Zeldin E, et al. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. Am. J. Bot. 2012;99: 193–208. doi:10.3732/ajb.1100394.

70. Clivati D, Gitzendanner MA, Hilsdorf AWS, Araújo WL, Miranda VFO. Microsatellite markers developed for *Utricularia reniformis* (Lentibulariaceae). Am. J. Bot. 2012;99: E375-378. doi:10.3732/ajb.1200080.

71. Maul JE, Lilly JW, Cui L, dePamphilis CW, Miller W, Harris EH, et al. The *Chlamydomonas reinhardtii* plastid chromosome: islands of genes in a sea of repeats. Plant Cell 2002;14: 2659–79. doi:10.1105/tpc.006155.present.

72. Quandt D, Müller K, Huttunen S. Characterisation of the Chloroplast DNA *psb*T-H Region and the Influence of Dyad Symmetrical Elements on Phylogenetic Reconstructions. Plant Biol. 2003;5: 400–410. doi:10.1055/s-2003-42715.

73. Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol. Biol. 2011;76: 273–97. doi:10.1007/s11103-011-9762-4.

74. Luo J, Hou BW, Niu ZT, Liu W, Xue QY, Ding XY. Comparative chloroplast genomes of photosynthetic orchids: Insights into evolution of the Orchidaceae and development of molecular markers for phylogenetic applications. PLoS One 2014;9. doi:10.1371/journal.pone.0099016.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

75. Ruhlman TA, Chang W-J, Chen JJW, Huang Y-T, Chan M-T, Zhang J, et al. NDH expression marks major transitions in plant evolution and reveals coordinate intracellular gene loss. BMC Plant Biol. 2015;15: 100. doi:10.1186/s12870-015-0484-7.

76. Rumeau D, Bécuwe-Linka N, Beyly A, Louwagie M, Garin J, Peltier G. New subunits NDH-M, -N, and -O, encoded by nuclear genes, are essential for plastid Ndh complex functioning in higher plants. Plant Cell 2005;17: 219–32. doi:10.1105/tpc.104.028282.

77. Ueda M, Kuniyoshi T, Yamamoto H, Sugimoto K, Ishizaki K, Kohchi T, et al. Composition and physiological function of the chloroplast NADH dehydrogenase-like complex in *Marchantia polymorpha*. Plant J. 2012;72: 683–693. doi:10.1111/j.1365-313X.2012.05115.x.

78. Yamori W, Sakata N, Suzuki Y, Shikanai T, Makino A. Cyclic electron flow around photosystem I via chloroplast NAD(P)H dehydrogenase (NDH) complex performs a significant physiological role during photosynthesis and plant growth at low temperature in rice. Plant J. 2011;68: 966–976. doi:10.1111/j.1365-313X.2011.04747.x.

79. Santamaría L. Why are most aquatic plants widely distributed? Dispersal, clonal growth and small-scale heterogeneity in a stressful environment. Acta Oecologica 2002;23: 137–154. doi:10.1016/S1146-609X(02)01146-3.

80. Silva SR, Alvarenga DO, Aranguren Y, Penha HA, Fernandes C, Pinheiro DG, et al. The mitochondrial genome of the terrestrial carnivorous plant *Utricularia reniformis* (Lentibulariaceae): Structure, comparative analysis and evolutionary landmarks. 2017; 1–26. doi:10.1371/journal.pone.0180484

81. Lin C-S, Chen JJW, Huang Y, Chan M, Daniell H, Chang W, et al. The location and translocation of *ndh* genes of chloroplast origin in the Orchidaceae family. Sci. Rep.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

2015;5: 9040. doi:10.1038/srep09040.

82. Sanderson MJ, Copetti D, Búrquez A, Bustamante E, Charboneau JLM, Eguiarte LE, et al. Exceptional reduction of the plastid genome of saguaro cactus (*Carnegiea gigantea*): Loss of the *ndh* gene suite and inverted repeat 1. Am. J. Bot. 2015;102: 1115–1127. doi:10.3732/ajb.1500184.

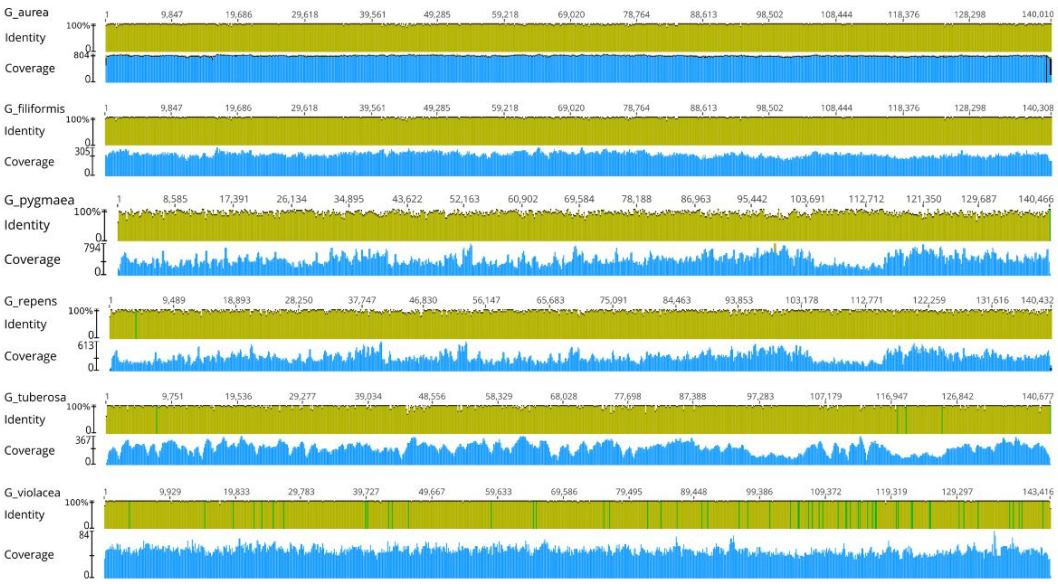
83. Petrov V, Hille J, Mueller-Roeber B, Gechev TS. ROS-mediated abiotic stress-induced programmed cell death in plants. Front. Plant Sci. 2015;6: 69. doi:10.3389/fpls.2015.00069.

84. Adamec L. Respiration and photosynthesis of bladders and leaves of aquatic *Utricularia* species. Plant Biology, 2006;8: 765–769. doi:10.1055/s-2006-924540.

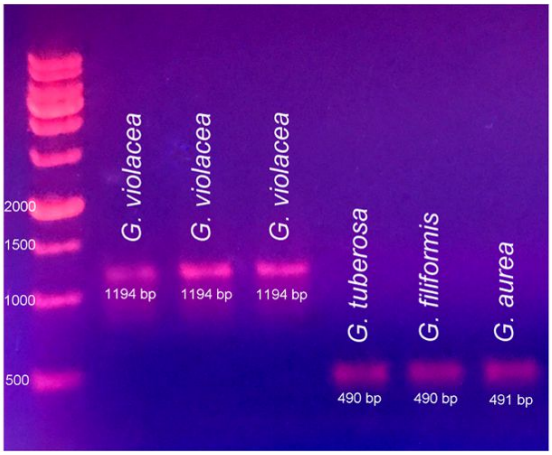
85. Ibarra-Laclette E, Albert VA, Herrera-Estrella A, Herrera-Estrella L. Is GC bias in the nuclear genome of the carnivorous plant *Utricularia* driven by ROS-based mutation and biased gene conversion? Plant Signal. Behav. 2011;6: 1631–1634. doi:10.4161/psb.6.11.17657.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

Supplementary information



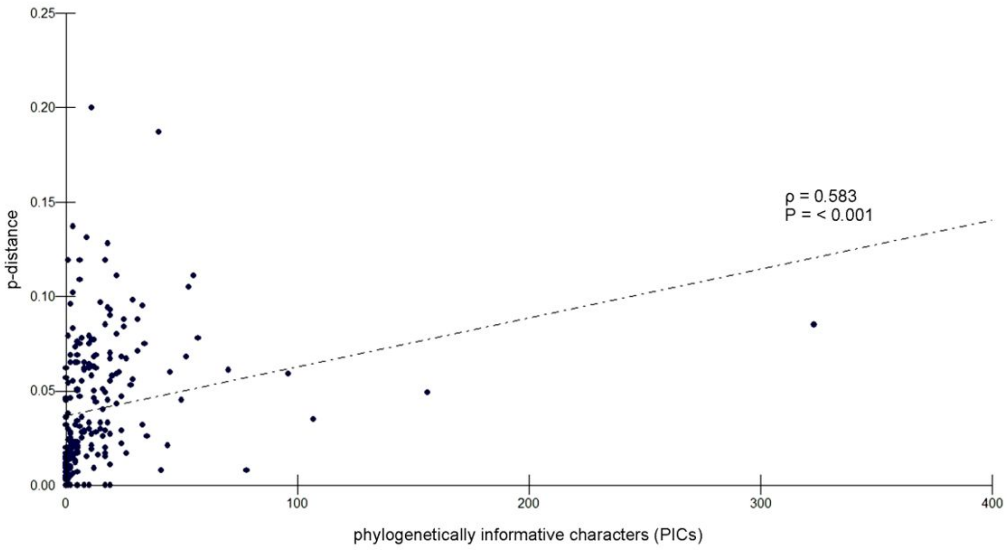
S1 Fig. Coverage and read identity plots for the reconstructed plastid genomes of *Genlisea* species. All quality-trimmed reads from sequencing data sets have been mapped back to the reconstructed plastid supercontig. The upper plot indicates the identity per site and the lower plot shows the coverage plot per species.



S2 Fig. Agarose gel electrophoresis (0.8%) of PCR products of the cpDNA IR/LSC boundary of *Genlisea violacea* (3 bioreplicates = 3 specimens), *G. aurea*, *G. filiformis*, and *G. tuberosa*. Note the product of *G.*

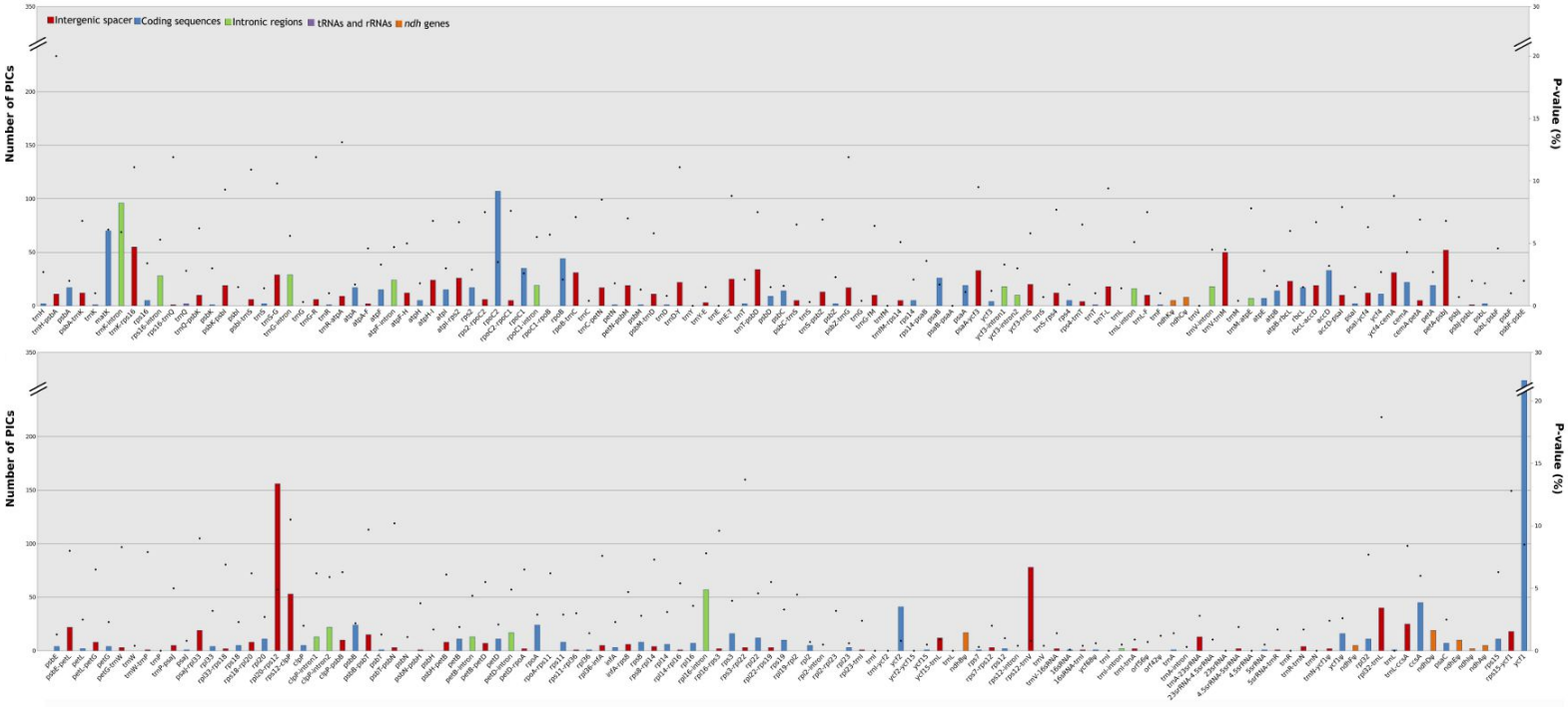
Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

violacea cpDNA that presents the duplication of *rps19* gene and *rpl22* as pseudogene (amplicon with 1,194 bp), while the other species present an expected product with ~490 bp. (Amplification reactions of the rpl2-trnH(GUG) marker were conducted in 25 µL of the solution containing 20 mM of $MgCl_2$, 100 mM of dNTPs, 10 mM of each primer, 1 U of Dream Taq Polymerase – Fermentas, and 50 ng of DNA template. The thermal profile for amplification was 1min at 94°C; 35 cycles of 40s at 94°C, 20s at 64°C, 90s at 72°C, and 5min of final extension at 72°C. Forward primer = 5'-AGT CGG ACA AGT GGG GAA TG-3'; reverse primer = 5'-GGA TGT GGC CAA GTG GAT CA-3').



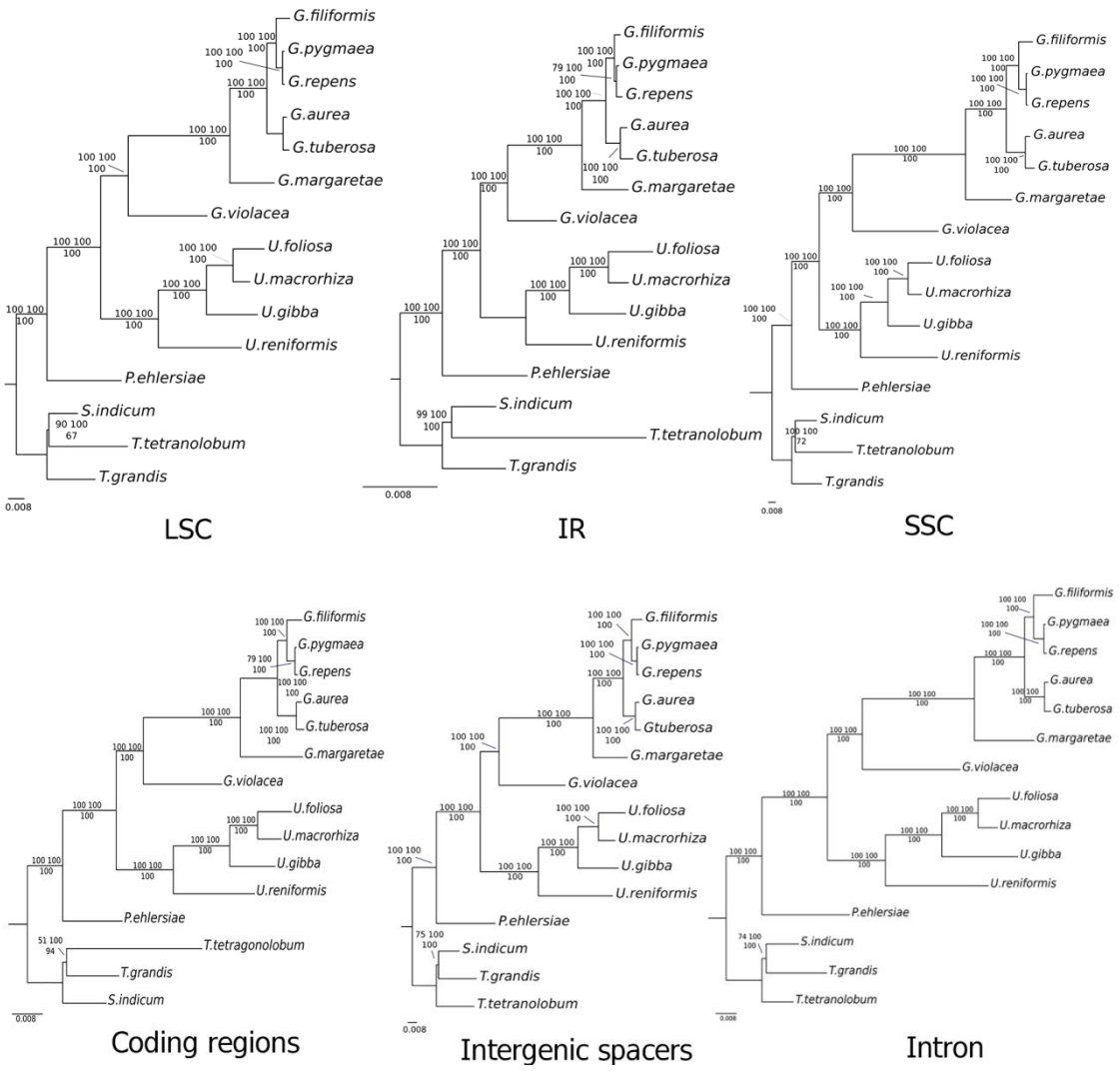
S3 Fig. Correlation between p-distance and phylogenetically informative characters (PICs). Statistics from Spearman correlation tests are given near the corresponding trend lines.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br



S4 Fig. Phylogenetically informative characters (PIC) and p-distance in *Genlisea* cpDNA based on alignment data. PIC values are represented as bars and cpDNA region is marked by colors. Black dots represent p-distance. Only PIC of *ndhs* were not calculated to avoid p-distance alignment artefact (see S6 Table).

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br



S5 Fig. Phylogenomic trees based on different datasets for *Genlisea* species. Numbers above are parsimony bootstrap (left), maximum likelihood bootstrap (right) and posterior probability values are represented below. Lamiales species were used as outgroup.

S1 Table. Summary of sequencing data for *Genlisea* species

Tabela pode ser encontrada no link: https://doi.org/10.1371/journal.pone.0190321.s006

S2 Table. Characters and states codified of *ndh* genes for Lentibulariaceae.

Tabela pode ser encontrada no link: https://doi.org/10.1371/journal.pone.0190321.s007

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

S3 Table. Matrix with codified characters of *ndh* genes for Lentibulariaceae. The characters were codified according the S2 Table. (G.= *Genlisea*; P.= *Pinguicula*; U.= *Utricularia*)

Tabela pode ser encontrada no link: https://doi.org/10.1371/journal.pone.0190321.s008

S4 Table. *ndh* genes length variation among *Genlisea* and *Utricularia gibba* species. Numbers within table refer to sequence length (bp). Colors refer to the state of character: white – deleted gene; yellow – pseudogenized; pink – decayed gene; grey – complete gene; n/a – absent.

Tabela pode ser encontrada no link: https://doi.org/10.1371/journal.pone.0190321.s009

S5 Table. Repeats (direct, palindromic and tandem) for each *Genlisea* species. F – Direct repeats; P – Palindromic repeats; T – Tandem repeats (inside parenthesis the repeated nucleotide). Common genes with repeats between the six species are highlighted with yellow background color in *G. aurea* table.

Tabela pode ser encontrada no link: https://doi.org/10.1371/journal.pone.0190321.s010

S6 Table. Phylogenetically informative characters (PIC) and p-distance of each gene for *Genlisea* species. Deleted *ndh* genes in all *Genlisea* species and boundaries between *ndh* pseudogenes are uncertain and were not included in this analysis (represented as n/a).

Tabela pode ser encontrada no link: https://doi.org/10.1371/journal.pone.0190321.s011

S7 Table. Codon usage and amino acid frequencies for *Genlisea* plastomes.

Tabela pode ser encontrada no link: https://doi.org/10.1371/journal.pone.0190321.s012

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"

PLOS ONE

RESEARCH ARTICLE

Comparative genomic analysis of *Genlisea* (corkscrew plants—Lentibulariaceae) chloroplast genomes reveals an increasing loss of the *ndh* genes

Saura R. Silva[1], Todd P. Michael[2], Elliott J. Meer[3], Daniel G. Pinheiro[4], Alessandro M. Varani[4]*, Vitor F. O. Miranda[5]*

1 Universidade Estadual Paulista (Unesp), Botucatu, Instituto de Biociências, São Paulo, Brazil, **2** J. Craig Venter Institute, La Jolla, CA, United States of America, **3** 10X Genomics, Pleasanton, California, United States of America, **4** Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal, Departamento de Tecnologia, São Paulo, Brazil, **5** Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal, Departamento de Biologia Aplicada à Agropecuária, São Paulo, Brazil

* vmiranda@fcav.unesp.br (VFOM); amvarani@fcav.unesp.br (AMV)

OPEN ACCESS

Citation: Silva SR, Michael TP, Meer EJ, Pinheiro DG, Varani AM, Miranda VFO (2018) Comparative genomic analysis of *Genlisea* (corkscrew plants—Lentibulariaceae) chloroplast genomes reveals an increasing loss of the *ndh* genes. PLoS ONE 13(1): e0190321. https://doi.org/10.1371/journal.pone.0190321

Editor: Xiu-Qing Li, Agriculture and Agri-Food Canada, CANADA

Received: September 18, 2017

Accepted: December 12, 2017

Published: January 2, 2018

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: SRS was supported with fellowship by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). VFOM and AMV thank the Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp – Proc. #2013/05144-0 and #2013/25164-6) and VFOM the Conselho Nacional de Desenvolvimento Científico e Tecnológico

Abstract

In the carnivorous plant family Lentibulariaceae, all three genome compartments (nuclear, chloroplast, and mitochondria) have some of the highest rates of nucleotide substitutions across angiosperms. While the genera *Genlisea* and *Utricularia* have the smallest known flowering plant nuclear genomes, the chloroplast genomes (cpDNA) are mostly structurally conserved except for deletion and/or pseudogenization of the NAD(P)H-dehydrogenase complex (*ndh*) genes known to be involved in stress conditions of low light or CO_2 concentrations. In order to determine how the cpDNA are changing, and to better understand the evolutionary history within the *Genlisea* genus, we sequenced, assembled and analyzed complete cpDNA from six species (*G. aurea*, *G. filiformis*, *G. pygmaea*, *G. repens*, *G. tuberosa* and *G. violacea*) together with the publicly available *G. margaretae* cpDNA. In general, the cpDNA structure among the analyzed *Genlisea* species is highly similar. However, we found that the plastidial *ndh* genes underwent a progressive process of degradation similar to the other terrestrial Lentibulariaceae cpDNA analyzed to date, but in contrast to the aquatic species. Contrary to current thinking that the terrestrial environment is a more stressful environment and thus requiring the *ndh* genes, we provide evidence that in the Lentibulariaceae the terrestrial forms have progressive loss while the aquatic forms have the eleven plastidial *ndh* genes intact. Therefore, the Lentibulariaceae system provides an important opportunity to understand the evolutionary forces that govern the transition to an aquatic environment and may provide insight into how plants manage water stress at a genome scale.

CONCLUSÕES FINAIS

A partir da pesquisa proposta na presente tese foi possível obter cloroplastos das espécies: *Utricularia reniformis*, *U. foliosa*, *Genlisea repens*, *G. pygmaea*, *G. filiformis*, *G. aurea*, *G. tuberosa*. Para mitocôndrias, foi possível montar e descrever o primeiro mtDNA de *Utricularia*, da espécie *Utricularia reniformis*.

De acordo com as evidências é possível observar que os cloroplastos de ambos os gêneros possuem estrutura quadripartida típica. Entretanto há deleção, fragmentação e pseudogenização de genes *ndhs*, bem como estes genes *ndhs,* nos cloroplastos de *Utricularia* e *Genlisea,* possivelmente têm relação com sua forma de vida. Posto que, somente as espécies aquáticas possuem o repertório de genes *ndhs* completos, em contraposiçao aos genes *ndhs* em espécies terrestres, que foram encontrados deletados, pseudogenizados e fragmentados até o presente trabalho.

Em adição, foi possível reconstruir a história evolutiva das espécies a partir de diferentes fontes de dados, como genes, regiões intergênicas, não codificantes e até mesmo a partir da codificação de caracteres em relação a presença e ausência de genes.

As evidências encontradas neste trabalho podem ter implicações futuras no entendimento sobre o processo de pseudogenização que ocorre em cloroplastos e implicações no entendimento sobre a relação entre genes e tolerância a estresse hídrico, em espécies de plantas, inclusive as agriculturáveis.

Instituto de Biociências – Departamento de Botânica
Distrito de Rubião Júnior s/n CEP 18618-000 Botucatu SP Brasil
Tel 14 3811 6265/6053 fax 14 3815 3744 botanica@ibb.unesp.br