Chloroplast genomes of key species shed light on the evolution of the ancient genus Isoetes

Jovani B. S. Pereira1,2*, Ana Maria Giulietti1,3, Eder S. Pires4, Marcelce Laux4, Maurício T. C. Watanabe1, Renato R. M. Oliveira4, Santelmo Vasconcelos4, and Guilherme Oliveira4

1Biodiversity and Ecosystems Services Group, Instituto Tecnológico Vale, Belém, Pará, Brazil
2Instituto de Botânica de São Paulo, São Paulo, Brazil
3Depo. Botânica, Universidade Estadual de Feira de Santa, Feira de Santana, Bahia, Brazil
4Environmental Genomics Group, Instituto Tecnológico Vale, Belém, Pará, Brazil
*Author for correspondence. E-mail: jovanibio@gmail.com

Received 26 September 2019; Accepted 23 September 2020; Article first published online 29 September 2020

Abstract Although phylogenetic studies have revealed major clades, the deepest relationships in Isoetes remain unresolved. The use of next-generation sequencing provides enormous amounts of gene sequences, which allows not only clarification of the basal relationships but also rapid radiations. Plastomes of six key Isoetes species were annotated, revealing a total of 129 or 130 genes, depending on the species. Our phylogenomic analyses comprising representatives of all major clades yielded well-supported nodes and identical topologies using maximum likelihood and Bayesian inference. The phylogenetic reconstructions detangled the deep relationships in Isoetes and illuminated the more recent radiations in the genus. A basal dichotomy was found that grouped Isoetes spp. from Brazil and South Africa into a clade sister to the remaining Isoetes groups. Interestingly, I. andicola was found to be sister to the North American species complex. Genomic trait mapping analysis showed that the missing introns in the atpF and clpP genes were well conserved in two major clades. The absence of trnK-UUU was observed in the Brazilian tropical species and in I. velata. Among lycoptyes, the gene trnR-CCG was missing only in I. eludens. In general, genomic traits such as the presence or absence of internal stop codons, a tRNA, and an intron were revealed to be conserved within groups, suggesting that these genomic traits might reveal vital information about the evolution of the genus. This study will contribute to understanding the diversification of Isoetes and the establishment of a better framework to address the evolutionary history of the genus.

Key words: internal stop codons, Isoetes, lycophytes, next-generation sequencing, phylogenomic, RNA editing, tRNA.

1 Introduction

Recovering the phylogenetic relationships among species is one of the prime goals of systematic and evolutionary biology (Cracraft & Donoghue, 2004). It enriches our understanding of how genes, genomes, and species evolve. Through phylogenetics, we learn not only how species came to be the way they are today but also general principles that enable us to predict how they will change in the future (Yates et al., 2004). Phylogenetic inference is being revolutionized with the emergence of next-generation sequencing (NGS), which provides unprecedented opportunities to address fundamental evolutionary questions even for more complex groups that present intriguing evolutionary histories (Soltis et al., 2013), such as the ancient genus Isoetes.

Isoetes is a heterosporous lycophyte that has long fascinated botanists and palaeobotanists alike due to its distinctive morphology and long evolutionary history, which appears to have few parallels among the extant vascular plants (Moran, 2004; Taylor et al., 2009). The genus is the only living survivor of the Isoetales, which arose in the Devonian (Pigg, 1992), and the most ancient species of Isoetes is known from the Late Jurassic (Pigg, 2001). Morphologically, the genus is quite simple and can readily be distinguished from any other group of vascular plants by its narrow leaves containing four air chambers, a single sunken adaxial sporangium covered by a velum and the presence of sporangial trabeculae (Taylor et al., 2016).

The genus is globally distributed and comprises approximately 250 species (Troia et al., 2016). Most of them are

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
narrow endemics occurring as aquatic or ephemeral terrestrial plants (Pfeiffer, 1922; Figs. 1A, 1B).

Among the living plant groups, the genus Selaginella P. Beauv. is the closest relative of Isoetes (Kenrick & Crane, 1997). However, although the sister relationship between Isoetes and Selaginella is supported by morphological and molecular data (Kenrick & Crane, 1997; Pryer et al., 2001), the divergence between these groups dates back to the Devonian (Klaus et al., 2017), resulting in considerable genetic divergence between them (Schuettpelz & Hoot, 2006).

The initial morphological classification of Isoetes divided the genus into a small, putatively relictual grade (I. subgenus Euphyllum) (Hickey, 1990) and a diverse and derived clade (I. subgenus Isoetes; Hickey, 1986). Notably, the subgenus Isoetes was recognized into two sections: I. section Coromandelina, restricted to the Indian subcontinent and the cosmopolitan I. sect. Isoetes (Taylor & Hickey, 1992). However, the classification based on two subgenera, one with two sections, was not supported by molecular studies (Rydin & Wikström, 2002).

Molecular phylogenetic studies (Hoot et al., 2006; Pereira et al., 2017) showed six major clades in Isoetes, which appear to be geographically structured as follows: the Gondwanan core, South African, Northern Hemisphere, Australasian, Mediterranean, and American clades. However, difficulties in identifying phylogenetic rooting have led to uncertainties about the deep relationships in Isoetes (Schuettpelz & Hoot, 2006).

Outgroup rooting can be challenging when the nearest sister group is distantly related and when there is relatively little variation within the ingroup (Huelsenbeck et al., 2002). A previous molecular phylogenetic study based on plastid rbcl gene sequences revealed a basal split of Isoetes into two major clades (Rydin & Wikström, 2002). Whereas increasing the number of molecular markers from one to three (rbcl, the plastid noncoding atpB-rbcl intergenic spacer, and the internal transcribed spacer of nuclear ribosomal DNA [nrITS]) and phylogenetic groups in the analyses showed a basal unresolved trichotomy in Isoetes (Schuettpelz & Hoot, 2006).

The use of NGS technology provides an enormous amount of gene sequence data (Soltis et al., 2013), which not only allows clarification of basal relationships but also resolves rapid radiations that would be intractable with a smaller number of genes (e.g., Jansen et al., 2007; Moore et al., 2010). Such analysis may contribute to understanding the diversification of Isoetes and provide a better framework for investigating the evolutionary history of the genus.

One interesting characteristic observed in the organellar genomes of heterosporous lycophytes (i.e., Isoetes and Selaginella) is RNA editing, which occurs in all major land plant lineages (Schallenberg-Rüdinger & Knoop, 2016) but
reaches extreme numbers of sites edited in these two groups (Grewe et al., 2009; Hecht et al., 2011). RNA editing is a form of nucleotide sequence alteration that occurs at the transcriptional level (Oldenkott et al., 2014; Schallenberg-Rüdinger & Knoop, 2016; Villarreal et al., 2018). It converts cytidines to uridines (C-to-U or canonical RNA editing) or uridines to cytidines (U-to-C or so-called reverse editing) in the primary transcript before translation (Schallenberg-Rüdinger & Knoop, 2016). RNA editing is reported to occur in different parts of the genome, such as protein-coding regions, rRNA, tRNA, and introns (Schallenberg-Rüdinger & Knoop, 2016). However, in protein-coding genes, it generally means restoring codons for conserved amino acids, generating start/stop codons or, often, removing internal stop codons (Schallenberg-Rüdinger & Knoop, 2016). Although RNA editing is not restricted to coding regions, it may be in these genomic areas where its most fundamental role takes place by maintaining the essential functions of the gene and potentially impacting the phenotype and evolution of the species. RNA editing is discontinuously distributed among the major taxonomic groups, suggesting that it has evolved independently several times (Chateigner-Boutin & Small, 2011). However, it remains unknown how RNA editing varies within a single genus, raising questions about the selective pressures acting to maintain RNA editing have yet to be resolved entirely using a phylogenetic framework.

In this study, we assembled chloroplast genomes for six species from distinct phylogenetic and taxonomic groups in Isoetes and then compared their overall structures as a contribution to understand the evolution of the chloroplast genome in land plants. We also reconstructed phylogenetic relationships using species from major phylogenetic and taxonomic groups and almost all continents where Isoetes occurs. The aim was to establish deep relationships and provide a first global overview of the most recent radiation events in Isoetes. Finally, we mapped the distribution of genomic traits on the phylogenetic tree to investigate their evolution and congruence with the phylogenetic groups.

2 Material and Methods

2.1 Taxon sampling, DNA extraction, library preparation, and sequencing

Total genomic DNA was extracted from fresh leaves of Isoetes gigantea U. Weber according to the CTAB protocol described by Weising et al. (2005). Whole genomic DNA was isolated from silica-dried tissue (for I. echinospora, I. malinverniana, I. eludens, and I. velata) and herbarium material for I. andicola (Amstutz) L.D.Gómez, using the Qiagen DNeasy Plant Mini Kit (Qiagen, Maryland, USA) by following the manufacturer’s protocol (Table 1).

The illumina NextSeq 500 platform was used for sequencing the six species of Isoetes. Briefly, paired-end libraries (150 bp) were constructed from ~50 ng of genomic DNA. Samples were subjected to a step of enzymatic and random fragmentation in which the DNA was simultaneously fragmented and bound to adapters using the QXT SureSelect kit (Agilent Technologies, Santa Clara, USA) according to the manufacturer’s instructions. The fragmented DNA was purified and subjected to an amplification step using primers complementary to the adapters and bound indexes. After libraries were quantified using the Qubit® 3.0 Fluorimeter (Life Technologies, USA) and fragments were checked for size in the 2100 Bioanalyzer using the High sensitivity DNA kit (Agilent Technologies). Then, libraries were diluted to loading concentration, pooled, and denatured. The sequencing run was performed using the NextSeq 500 v2 kit high-output (300 cycles).

2.2 Plastome assembly and annotation

The de novo assembly was carried out using NOVOPlasty version 2.6.3 (Dierckxsens et al., 2017). The reads with base quality Phred<20 and length<100 bp were trimmed, and the remaining reads with more than 20% low-quality bases (Phred<20) were filtered out using Fastx-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). The config files were set as follows: insert size 300, read length 150, type chloro, genome range 120k-200k, k-mer 39 and “seed input” with psbC and rbcL. Sequences of the same or closely related species from GenBank (NCBI accession numbers: MF565804 and MF56766 [I. eludens]; MF565766 and MF565804 [I. gigantea]; DQ294251 [I. velata]; GU191333 [I. andicola, and I. echinospora]; MF565789, AF404500 [I. malinverniana]). Annotations of extended contigs were carried out using GeSeq (Tillich et al., 2017), a tool for accurate annotation of organelle genomes. For tRNA loci, we further used TRNAscan-SE v2.0 (Lowe & Cahn, 2016) as implemented in GeSeq (Tillich et al., 2017). We used Geneious R11 (Kearse et al., 2012) to validate the assembled and annotated plastomes by read mapping (the “Mapping to Reference” function with “Highest Sensitivity/Slow” settings) using the plastomes of I. flaccida, I. nuttallii, and I. cangae as references (NCBI accession numbers: GU191333, NC_038073, and MG019394, respectively). The open reading frames (ORFs) for all coding

| Species  | Locality          | Voucher (herbarium acronym) | Genbank accession number |
|----------|-------------------|-----------------------------|--------------------------|
| Isoetes  |                   |                             |                          |
| I. gigantea | Minas Gerais, Brazil | Pereira 1086 (MG) | MK804478                |
| I. velata | Spain             | Stützbel s.n. (BOCH)       | MK804477                 |
| I. eludens| South Africa, North Cape | Smith 1140 (K)     | MK804475                 |
| I. andicola | La Libertad, Peru     | Weigend 7982 (B)    | MK804473                 |
| I. echinospora | Cornwall, United Kingdom | Chase 22740 (K) | MK804474                 |
| I. malinverniana | Italy             | Cook s.n. (Z/2T) | MK804476                 |
genes were manually checked and adjusted when necessary using Artemis v.16 (Carver et al., 2012).

2.3 Alignment and phylogenetic analyses
Individual protein-coding sequence (CDS) of 63 genes found in 21 species of Isoetes, and two species of both Selaginella and Huperzia were extracted from complete chloroplast genomes using Artemis v.16 (Carver et al., 2012). CDSs were individually aligned using MAFFT v7.388 (Katoh & Standley, 2013) with the gap penalty = 1.53 and offset value = 0.123 parameters. Subsequently, the individual matrices were concatenated into a single supermatrix, and then, the best evolutionary model of nucleotide substitution was calculated using JModelTest (Darriba et al., 2012). Because the plastid genome is considered to behave as a single locus (Ruhfel et al., 2014), we used GTR + GAMMA for both the Akaike information criterion (AIC) and Bayesian information criterion (BIC) as the best evolutionary model in phylogenetic analyses of the concatenated dataset. Phylogenetic analyses were carried out using Bayesian inference (BI) and maximum likelihood (ML). BI was performed using MrBayes v.3.2 (Ronquist et al., 2012) in CIPRES Science Gateway V 3.3 (Miller et al., 2015). The analyses were performed using two runs, with 100 million generations each, sampling every 1000 generations and 25% burn-in. The convergence of the runs was analyzed using Tracer 1.7 (Rambaut et al., 2018), an effective sample size (ESS) above 200 was considered adequate for run convergence. For ML, the data set was analyzed using RAxML v.7.4.2 (Stamatakis, 2006). The ML analyses were performed using the rapid bootstrap method and 1,000 bootstrap replicate trees were constructed. We took the species identification of the corresponding plastomes from the GenBank as such (Table S1). The vouchers of the respective newly sequenced species included in this study were identified by the authors (Table 1).

2.4 Evolutionary pattern of genomic traits
Seventeen traits were scored for character mapping (Table 2). Genomic traits were chosen based on differences previously found in distinct plastomes published by Karol et al. (2010), Nunes et al. (2018), Schafran et al. (2018b), and Feng et al. (2019). These genomic traits can be classified into three groups: (i) internal stop codon; (ii) presence or absence of a tRNA (trnK-UUU and trnR-CGG); (iii) deletion of an intron (in clpP and atpF). For this analysis, only one accession per species was used, and because we did not find intraspecific plastome differences for the analyzed traits, the data set was restricted to the accessions sampled. In total, our data comprise information from 25 living species (21 from Isoetes and two from each Selaginella and Huperzia). Because organelle genome annotation may provide ambiguous information about gene start and stop codons, requiring additional manual adjustment, we decided to map only internal stop codons, which likely require RNA editing (see Section 4).

We used Artemis v.16 (Carver et al., 2012) to verify the presence of stop codons in the interior of each CDS. Analyses of protein sequences for C-to-U transitions were carried out using Pfam 32.0 (El-Gebali et al., 2019). Nonsynonymous to synonymous rate ratios (dN/dS ratios) of protein-coding genes was estimated to address if purifying selection still works on them. dN/dS ratios were calculated using the CODEML program in PAML4.9i (Yang, 2007). Multiple sequence alignments of proteins and their corresponding nucleotide sequences were converted into codon alignments using PAL2NAL (Suyama et al., 2006). PAML analyses were run under model Mo and F3X44 coding frequency.

The matrix of presence/absence of trnK-UUU and trnR-CGG, as well as the deletions of single introns in clpP and atpF, was obtained after aligning the genomes, as previously described, and the annotation was visualized using Geneious R11 (Kearse et al., 2012). The ML analyses were run as described above, and character states were then optimized onto the ML tree under a maximum parsimony criterion using Mesquite v.3.11 (Maddison & Maddison, 2018).

3 Results
3.1 Plastome structure, gene content, gene order, and internal stop codons
The newly generated chloroplast genome lengths varied from 143 439 to 145 422 bp. Isoetes gigantea had the smallest, whereas Isoetes malinverniana possessed the largest (Fig. 2; Table 3). All six genomes contained two copies of the inverted repeat (IR) regions, each ranging from 12 815 in I. gigantea to 13 217 bp in I. malinverniana (Fig. 2; Table 3). The IRs were separated by a large single copy (LSC) region ranging from 90 252 in I. velata to 91 851 bp in I. andicola and a small single copy (SSC) region ranging from 26 739 in I. eludens to 27 758 bp in I. velata (Fig. 3; Table 3). The overall guanine-cytosine (GC) content was similar and ranged from 37.6% in I. gigantea to 38.1% in both I. velata and I. eludens (Table 3).

A total of 129 (I. eludens, I. gigantea, and I. velata) and 130 genes (I. andicola, I. echinospora and, I. malinverniana) were annotated in the sampled species (Table S2). The

| Table 2 Genomic traits mapped on the phylogenetic tree |
|-------------------------------------------------------|
| Internal stop codon                                   |
| rps3                                                 |
| rph14                                                |
| rpoA                                                 |
| matK                                                 |
| atpI                                                 |
| rps2                                                 |
| rps1                                                 |
| rpoB                                                 |
| ndhH                                                 |
| ccsA                                                 |
| rpl21                                                |
| ndf                                                  |
| ycf2                                                 |
| trRNA absence/presence                               |
| trnK-UUU                                             |
| trnR-CGG                                             |
| Intron deletion                                      |
| clpP                                                 |
| atpF                                                 |
region \textit{trnK-UUU} was observed in all sampled species, with the exception of \textit{I. gigantea} and \textit{I. velata}. The region \textit{trnR-CCG} was absent in only \textit{I. eludens} (Table S2).

In our newly described plastomes, the \textit{clpP} gene contained only one intron in \textit{I. gigantea}, similar to what was observed for \textit{I. cangae} and \textit{I. serracarajensis} (NCBI accession numbers: MG019394 and NC_039393, respectively; Fig. 4). All remaining \textit{Isoetes} spp. showed two introns in \textit{clpP} (Fig. 4). \textit{atpF} showed a single intron in all the generated plastomes with the exception of \textit{I. velata}, as in its close relative \textit{I. nuttallii} (NCBI accession number: NC_038073; Fig. 4). \textit{accD}, \textit{infA}, \textit{rps2}, and \textit{rps16} were found as pseudogenes in all the newly generated plastomes (Table S2) as well as in all remaining \textit{Isoetes} spp.

### Table 3 Summary of six newly assembled plastomes of \textit{Isoetes}

| Species          | Total plastome length (bp) | IRa and IRb | LSC     | SSC     | GC (%) |
|------------------|----------------------------|-------------|---------|---------|--------|
| \textit{I. gigantea} | 143 439                   | 12 815      | 90 757  | 27 052  | 37.6   |
| \textit{I. velata}   | 144 412                   | 13 201      | 90 252  | 27 758  | 38.1   |
| \textit{I. eludens}  | 144 511                   | 13 108      | 91 556  | 26 739  | 38.1   |
| \textit{I. andicola} | 144 828                   | 13 084      | 91 851  | 26 809  | 37.9   |
| \textit{I. echinospora} | 145 211                  | 13 098      | 91 786  | 27 229  | 38.0   |
| \textit{I. malinverniana} | 145 422                  | 13 217      | 91 586  | 27 402  | 38.0   |

Lengths are in base pairs (bp). GC, guanine-cytosine; LSC, large single copy; SSC, small single copy.
**Fig. 3.** Inverted repeat regions of 11 Isoetes species. Boxes with green arrows indicate genes. The blocks in blue and brown represent the large single copy (LSC) and small single copy (SSC), respectively. Asterisks (*) indicate the newly sequenced plastomes. Plastome lengths are given in base pairs (bp).

**Fig. 4.** Alignment comparison of the clpP and atpF genes in 11 Isoetes species showing the deletion of clpP intron 2 (I. gigantea, I. cangae, and I. serracarajensis) and atpF intron 1 (I. velata, and I. nuttallii). Black lines between black bars indicate gap areas in the alignment. Green and yellow bars show high and low levels of identity among the compared samples, respectively. Asterisks (*) indicate the newly sequenced species.
3.2 Phylogenetic relationships

Maximum likelihood and Bayesian inference (BI) analyses yielded identical topologies (Fig. 5). Almost all nodes received very strong support values in the two analyses (BI, posterior probability value \( PP = 1.0 \); ML, bootstrap value \( BS > 95 \)), whereas only one node in the phylogenetic tree were weakly supported (\( PP < 0.95 \); \( BS < 70 \)). The phylogenomic analyses highly supported a basal dichotomy in \( Isoetes \) grouping \( I. cangae, I. gigantea, \) and \( I. serracarajensis \) from tropical region in Brazil in a clade along with \( I. eludens \) from South Africa (clade A; Fig. 5). \( Isoetes velata \) from the Mediterranean region in Europe and \( I. nuttallii \) from western North America were well supported as sister to the remaining species of clade B (Fig. 5). \( Isoetes malinverniana \) was well resolved as a sister to a group comprising species from eastern Asia (\( I. yunguiensis \)), South American Andes (\( I. andicola \)) and a species complex from North America. Interestingly, \( I. andicola \) was found as a sister to the North American species complex (\( PP = 1.0; BS = 100 \); Fig. 5).

3.3 \( dN/dS \) ratios and genomic trait mapping

Our estimation of \( dN/dS \) ratios using the ML method revealed values < 1 showing negative or purifying selection for all protein-coding genes (Table S3). Additionally, \( dN/dS \) ratios were similar among genes with and without internal stop codons (Fig. 5). Among protein-coding genes with internal stop codons, \( dN/dS \) ratios ranged from 0.2011 in \( ndhH \) to 0.61633 in \( ycf2 \). Genes without internal stop codon showed \( dN/dS \) ratios ranging from 0.02755 in \( psbA \) to 0.6751 in \( psbK \) (see Table S3).

Premature stop codons were observed in the CDS of 14 genes in \( Isoetes \) (Table S4), but they were absent for the same genes in \( Selaginella \) and \( Huperzia \) (Fig. 6; Table S4). The location of the internal stop codons was the same in all \( Isoetes \) spp. All genes presented only a single internal stop codon except for \( rpoC1 \) and \( rpoC2 \). The \( rpoC1 \) gene revealed two internal stop codons in all species, with the exceptions of \( I. gigantea, I. cangae, \) and \( I. serracarajensis \). All \( Isoetes \) spp. had three premature stop codons in \( rpoC2 \). The \( matK \) gene was found with an internal stop codon in all species, except in \( I. eludens \) and \( I. yunguiensis \) (Fig. 6; Table S4). In \( rps3 \), an internal stop codon was found in \( I. eludens, I. velata, I. malinverniana, \) and \( I. andicola \), which are from distinct clades, indicating that this trait may have evolved at least four times in the genus. An internal stop codon in \( rpl14 \) appeared at least three times in the genus. It evolved once in both the clade A and American clade and independently in \( I. velata \). The \( rpoA \) gene containing an internal stop codon was found in all \( Isoetes \) spp., except in \( I. yunguiensis \) and the

---

**Fig. 5.** Molecular phylogenetic tree of 21 \( Isoetes \) species and four outgroups based on coding sequences (CDSs) of 63 genes using Bayesian inference (BI) and maximum likelihood (ML). The numbers at each node are the BI posterior probability and ML bootstrap support value. Bars on the right indicate the hypothetical major clades found in previous studies and molecular regions used in these studies. The purple and blue bars correspond to clades A and B, respectively, of Rydin & Wikström (2002). The bar with three shades of blue corresponds to the \( Isoetes \) groups of the basal trichotomy in Schuettpelz & Hoot (2006). The six bar colors correspond to the South African, Gondwanan core, Northern Hemisphere, \( Isoetes \) malinverniana, Australasian and American clades (from top to bottom) of both Hoot et al. (2006) and Pereira et al. (2017). Finally, the pink bar represents the North American species complex identified by Schafran et al. (2018a).
Fig. 6. Mapping of 17 genomic traits on the phylogenetic tree using maximum parsimony optimization. Orange and purple branches on the cladogram (left) and boxes (right) represent the absence and presence of the trait, respectively. i, internal stop codon in the protein-coding region of the gene. clpP*2, clpP intron 2. atpF*, atpF intron.
clade, including *I. velata* and *I. nuttallii* (Northern Hemisphere clade). The internal stop codon in the *ndhH* gene evolved at least twice in the phylogeny of *Isoetes*: one in the Northern Hemisphere clade and another in *I. malinverniana*. Except for the monophyletic group comprising *I. yunguiensis* and the American clade, a ccsA gene containing an internal stop codon was found in all groups in *Isoetes*. Finally, *atp1, rpoC2, rpoB, rpl21, ndhF, and ycf2* each presented an internal stop codon in all *Isoetes* spp. except for *I. yunguiensis* (Fig. 6; Table S4).

Among the lycophytes, the missing intron in *atpF* appears to occur only once in *Isoetes* in the ancestor of *I. nuttallii* and *I. velata* (Fig. 6; Table S4). *Isoetes gigantea, I. cangae,* and *I. serracarajensis* were found with only one intron in ctpP. The missing intron in ctpP was also found in *Selaginella uncinata* and *S. moellendorffii*. In contrast, the presence of two introns occurs in all remaining *Isoetes* spp. as well as in the genus *Huperzia*. Additionally, the absence of trnK-UUU was observed in the genus *Selaginella* and in *I. gigantea, I. cangae, I. serracarajensis,* and *I. velata*. The trnR-CCG tRNA was missing in only *I. eludens* among the lycophytes (Fig. 6; Tables S2, S4).

### 4 Discussion

#### 4.1 Plastome and genomic trait mapping

Despite its long evolutionary history, with the molecular data matching the age of the oldest known fossil of modern *Isoetes* from the Late Jurassic (Larsén & Rydin, 2016), the genus showed overwhelming evolutionary stasis in the plastome, with overall similar genome structure, gene content, gene order, and space between repeat regions (Fig. 2). The same pattern of plastome evolutionary stasis was found among *Huperzia* spp. (Wolf et al., 2005; Guo et al., 2016), making the genus *Selaginella* the only known group of lycophytes with a significant variation in structure and size (Mower et al., 2018).

Although the chloroplast genomes showed only slight differences in size, they appeared to reflect the phylogenetic relationships in the genus. For instance, the plastome of *I. gigantea* was 143,439 bp in size, similar to those of its close relatives, such as *I. cangae* and *I. serracarajensis* (Fig. 3). On the other hand, the chloroplast genome size of *I. gigantea* was approximately 1000 bp smaller than the plastome of *I. velata* (144,412 bp), which, in turn, was similar in size to that of its close relative *I. nuttallii* (Fig. 3). The plastome of *I. malinverniana* is the largest observed to date in *Isoetes*, with 145,422 bp, and slightly larger than those of its sister group comprising *I. yunguiensis* and the American clade (Fig. 3). In the American clade, plastome sizes ranged from 144,828 bp in *I. andicola* to 145,300 bp in *I. flaccida* (Karol et al., 2010).

The GC content of organelle DNA is positively correlated and directly connected to levels of organelle RNA editing in land plants (Smith, 2009). The proportion of GC content found in *Isoetes* (37.6–38%) is similar to those found by Mower et al. (2018) in Lycopodiaceae (35%–36.3%), Grewe et al. (2013) in basal fern groups (33.4%–42%), Chaw et al. (2018) in gymnosperms (34.24%–40.11%), and Yu et al. (2019) in liverwort groups (28.8%–40.6%), with the remarkable exception of *Haplomitrium* (44.4%). However, the plastomes of *Isoetes* spp. are substantially less rich in GC than those of their closest relative, *Selaginella* (52.3%–54.8%; Mower et al., 2018), which arguably indicates an extremely high level of RNA editing (Smith, 2009; Oldenkott et al., 2014).

*dN/dS* ratios have been extensively applied to the annotation of genomes under the simple premise that functionally protein coding gene evolve under purifying selection (*dN/dS* < 1), whereas dispensable or nonfunctional sites do it neutrally (*dN/dS* ≈ 1) (Andrieux & Arendales, 2014). In *Isoetes*, *dN/dS* ratios < 1 along with the presence of one internal stop codon in the CDS (more rarely, two or three, as in rpoC1 and rpoC2, respectively), which is located in the same position in all *Isoetes* lineages, indicates that these regions are undergoing RNA editing through U-to-C conversion, as found in ferns (see Wolf et al., 2003), but transcript sequences are required to prove RNA editing. The discontinuous phylogenetic distribution of RNA editing suggests that this feature has appeared independently several times among the major living groups (Chateigner-Boutin & Small, 2011) as well as among the major clades in *Isoetes* (Fig. 4). However, instead of being randomly distributed in the phylogenetic tree, these internal stop codons are well conserved within phylogenetic groups in *Isoetes*, reflecting a common inheritance of RNA editing (Fig. 4).

Remarkably, among the species studied herein, *I. yunguiensis* was the only *Isoetes* spp. in which RNA editing is probably not essential for the coding function ability of *atp1, rpoC2, rpoB, rpl21, ndhF,* and *ycf2*. The internal stop codon was found in these genes in all *Isoetes* spp., except in the case of *I. yunguiensis* (see Feng et al., 2019). In addition, these same genes revealed the absence of internal stop codons in *Selaginella* and *Huperzia*. Remarkably, among the species studied herein, *I. yunguiensis* was the only *Isoetes* spp. in which RNA editing is probably not essential for the coding function ability of *atp1, rpoC2, rpoB, rpl21, ndhF,* and *ycf2*. The internal stop codon was found in these genes in all *Isoetes* spp., except in *I. yunguiensis* (see Feng et al., 2019). In addition, these same genes revealed the absence of internal stop codons in *Selaginella* and *Huperzia*. The presence of the purifying selection on these protein coding regions may have independently contributed to the evolution of the coding functional ability without the assistance of RNA editing in *I. yunguiensis* within *Isoetes*.

By mapping the presence or absence of specific introns and genes on the phylogeny, we showed that they may either be specific to one single clade or have arisen independently in more than one clade and been subsequently lost in some species (Fig. 6). Notably, Nunes et al. (2018) showed that in *I. cangae* and *I. serracarajensis*, the ctpP gene contained only one intron, which is a genomic trait shared by *I. gigantea* and appears to be a synapomorphy for the clade. Among the lycophytes, the same pattern occurs in *S. moellendorffii* and *S. uncinata* (Fig. 6).

*Isoetes velata* and *I. nuttallii* are the only two species in lycophytes in which the intron of *atpF* is absent. These species are the only two species of the Northern Hemisphere clade included in our analyses, and the absence of *atpF* intron...
appears to represent an evolutionary novelty for this clade in lycophytes. The gene trnK-UUU is missing in Selaginella spp. (Mower et al., 2018) and appears to have been independently lost at least twice inIsoetes. This tRNA is absent in I. velata and in the clade including I. cangae, I. gigantea, and I. serracarajensis.

Isoetes eludens was the only studied species to lack the trnR-CCG gene. A nonfunctional trnR-CCG gene was previously observed in the moss Physcomitrella patens (Sugiura & Sugita, 2004). However, among the lycophytes, trnR-CCG appears to be absent in only I. eludens (for comparison, see Mower et al., 2018), although additional studies are needed to confirm the absence of this gene in the closest relatives of this species.

4.2 Phylogenetic relationships

In the last few years, there has been an increase in the number of new chloroplast genomes published for Isoetes (e.g., Mower et al., 2018; Nunes et al., 2018; ScharGAN et al., 2018b; Feng et al., 2019). However, in this study, we expanded on these previous publications by including representatives from all continents where the genus occurs (except Oceania), and all morphological and phylogenetic Isoetes groups to establish the relationships among them.

The backbone of the phylogenetic reconstruction and the clades agree with previous studies by Hoot et al. (2006), Larsén & Rydin (2016), and Pereira et al. (2017). Clade A comprises two other monophyletic groups: the Gondwanan core and South African (sensu Hoot et al., 2006; Pereira et al., 2017) clades. Clade B includes I. malinverniana and representatives of the Northern Hemisphere (I. velata and I. nuttallii), Australasian (I. yunguiensis), and American clades (I. andicola, I. echinospora, and several other species from North America) (sensu Hoot et al., 2006; Pereira et al., 2017).

Phylogenetic analyses based on plastid-genome-scale data have been used to resolve several enigmatic deep relationships within angiosperms and ferns (e.g., Jansen et al., 2007; Moore et al., 2010; Wei et al., 2017). In Isoetes, establishing deep relationships has been partially hampered by the difficulty of finding an adequate outgroup to root the phylogenetic tree and the small quantity of genomic data available (Rydin & Wikström, 2002; Schuettpelz & Hoot, 2006). Selaginella is the closest living relative to Isoetes, but the ancient divergence between these lineages (Klaus et al., 2017; Pereira et al., 2017) and the relative lack of variation within the ingroup in comparison to the outgroup (Schuettpelz & Hoot, 2006) have caused difficulties in rooting the phylogeny. Our phylogenomic analyses using coding regions of 64 genes, sampling representatives of all major groups, and including Selaginella and Huperzia as outgroups revealed a strongly supported basal dichotomy in the phylogenetic tree, resolving the deepest relationships in Isoetes (clades A and B).

In addition to resolving the deep relationships, our analyses shed light on the more recent radiations in the genus. Isoetes andicola, once hypothesized to represent a separate lineage and assigned to the genus Stylites (Amstutz, 1957), was shown as a sister to the North American species complex. Isoetes echinospora, a species with circumboreal distribution (Taylor et al., 1994), was found to be a sister to I. valida. Our data also corroborated Pereira et al. (2017), showing I. gigantea, a member of I. subg. Euphyllym (sensu Hickey, 1990) with hypothetical plesiomorphic morphological characters associated with Isoëtites fossils, grouping with I. cangae and I. serracarajensis in the Gondwanan core clade.

The increasing capacity to provide new plastid genomes with NGS is revolutionizing phylogenomic and evolutionary studies (Soltis et al., 2013). The data provided here allowed us to reconstruct robust phylogenetic inferences, building a necessary path toward a better understanding of evolutionary patterns in Isoetes.

In conclusion, this study not only detangled the deep relationships in Isoetes but also shed light on the more recent radiations in the genus. The reference-guided assembly showed a high level of evolutionary stasis in the plastome. On the other hand, the presence or absence of genomic traits such as internal stop codons, tRNA (trnK-UUU and trnR-CCG), and deletion of an intron (in clpP and atpF) appear to be conserved within groups, showing that these genome traits might reveal vital information about the evolution of Isoetes. One important aspect of this type of data is that changes that first appear to be autapomorphic or even synapomorphic in small datasets might later emerge as well-established key synapomorphies as additional species are included in the analyses.

Acknowledgements

The authors thank the ICMBio for authorizing collecting living materials in Brazilian national parks to JBSP (35897-4) and the K and SANBI herbaria for providing DNA samples. We are also grateful to Michael Kessler for providing silica-dried samples. The authors thank the following institution for financially supporting this study: Vale (Projeto Isoetes RBR500603.86), CAPES (88887.130628/2016-00), GCRF - RCUK (BB/P027849/1 - CABANA) to GO. GO is a CNPq fellow (307479/2016-1). Additionally, this study was supported through a postdoctoral fellowship from Capes/ITV (88887.130616/2016) to JBSP. We also thank Harald Schneider and two other anonymous reviewers for their helpful comments on the manuscript. The authors declare no conflict of interest.

References

Amstutz E. 1957. Stylites, a new genus of Isoetaceae. Annals of the Missouri Botanical Garden 44: 121–123.
Andrieux LO, Arenales DT. 2014. Whole-genome identification of neutrally evolving pseudogenes using the evolutionary measure dN/dS. Methods in Molecular Biology 1167: 75–85.
Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. 2012. Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics 28: 464–469.
Chateauigner-Boutin A-L, Small I. 2011. Organellar RNA editing. Wiley Interdisciplinary Reviews: RNA 2: 493–506.

J. Syst. Evol. 59(3): 429–441, 2021

www.jse.ac.cn
Chloroplast genomes of Isoetes 439

Chaw S-M, Wu C-S, Sudianto E. 2018. Evolution of gymnosperm plastid genomes. In: Chaw SH, Jansen RK eds. Plastid genome evolution. London: Elsevier. 195–222.

Cracraft J, Donoghue M. 2004. Assembling the tree of life. Oxford: Oxford University Press.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. JModelTest 2: More models, new heuristics and parallel computing. Nature Methods 9: 772.

Diercks N, Mardulpyn S, Smits G. 2017. NOVOPlasty: De novo assembly of organelle genomes from whole genome data. Nucleic Acids Research 45: e18.

El-Gebali S, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELH, Shish L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. 2019. The Pfam protein families database in 2019. Nucleic Acids Research 47: D427–D432.

Feng T, Xiao Q-Y, Liang Q, Li WJ, Yu Y. 2019. The complete chloroplast genome of Isoetes yunguiensis (Isoetaeae), a rare and endangered fern endemic to China. Mitochondrial DNA Part B 4: 563–564.

Grewe F, Guo W, Gubbels EA, Hansen AK, Mower JP. 2013. Complete plastid genomes from Ophioglossum californicum, Psilotum nudum, and Equisetum hyemale reveal an ancestral land plant genome structure and resolve the position of Equisetales among monilophytes. BMC Evolutionary Biology 13: 8.

Grewe F, Viehsoever P, Weisshaar B, Knoop V. 2009. A trans-splicing group I intron and tRNA-hyperediting in the mitochondrial genome of the lycophyte Isoetes engelmannii. Nucleic Acids Research 37: 5093–5104.

Guo Z-Y, Zhang H-R, Shrestha N, Zhang X-C. 2016. Complete chloroplast genome of a valuable medicinal plant, Huperzia serrata (Lycopodiaceae), and comparison with its congener. Applications in Plant Sciences 4: 1600071.

Hecht J, Grewe F, Knoop V. 2011. Extreme RNA editing in coding islands and abundant microsatellites in repeat sequences of Selaginella moellendorffii mitochondria: the root of frequent plant mtDNA recombination in early tracheophytes. Genome Biology and Evolution 3: 344–358.

Hickey RJ. 1986. The early evolutionary and morphological diversity of Isoetes, with descriptions of two new Neotropical species. Systematic Botany 11: 309–321.

Hickey RJ. 1990. Studies of Neotropical Isoetes L. I. Euphyllum, a new subgenus. Annals of the Missouri Botanical Garden 77: 239–245.

Hoot SB, Taylor WC, Napier NS. 2006. Phylogeny and biogeography of Isoetes based on nuclear and plastid DNA. Systematic Botany 31: 449–460.

Huelsgenbeek JP, Bollback JP, Levine AM. 2002. Inferring the root of a phylogenetic tree. Systematic Biology 51: 32–43.

Jansen RK, Cai Z, Raubeson LA, Daniell H, Depamphilis CW, Leebens-Mack J, Müller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, Chumley TW, Lee S-B, Peery R, McNeal JR, Kuehl JV, Boone JL. 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proceedings of the National Academy of Sciences USA 104: 19369–19374.

Karol KG, Arumuganathan K, Boone JL, Duffy AM, Everett KDE, Hall JD, Hansen SK, Kuehl JV, Mandoli DF, Mishler BD, Olmstead RG, Renzaglia KS, Wolf PG. 2010. Complete plastome sequences of Equisetum arvense and Isoetes flaccida: implications for phylogeny and plastid genome evolution of early land plant lineages. BMC Evolutionary Biology 10: 321.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Molecular Biology and Evolution 30: 772–780.

Kearse M, Moir R, Wilson A, Stonex-Hayas S, Cheung M, Sturrock S, Bufton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Mointjes P, Drummond A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647–1649.

Kenrick P, Crane PR. 1997. The origin and early diversification of land plants: A cladistic study. Washington, DC: Smithsonian Inst. Press.

Klaus KV, Schulz C, Bauer D, Stützel T. 2017. Historical biogeography of the ancient lycophyte genus Selaginella: Early adaptation to xeric habitats on Pangea. Cladistics 33: 469–480. https://doi.org/10.1111/cia.12184

Larsen E, Rydin C. 2016. Disentangling the phylogeny of Isoetes (Isoetaeae), using nuclear and plastid data. International Journal of Plant Sciences 177: 157–174.

Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: Integrating search and context for analysis of transfer RNA genes. Nucleic Acids Research 44: W54–W57.

Maddison WP, Maddison DR. 2018. Mesquite: A modular system for evolutionary analysis, version 3.51. http://www.mesquiteproject.org. [accessed 13 July 2019].

Miller MA, Schwartz T, Pickett BE, He S, Klem EB, Scheuermann RH, Passarotti M, Kaufman S, O’Leary MA. 2015. A RESTful API for access to phylogenetic tools via the CIPRES Science Gateway. Evolutionary Bioinformatics Online 11: 43–48.

Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE. 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. Proceedings of the National Academy of Sciences USA 107: 4623–4628.

Moran RC. 2004. A natural history of ferns. Portland: Timber Press.

Mower JP, Ma P-F, Grewe F, Taylor A, Michael TP, VanBuren R, Qui Y-L. 2018. Lycophyte plastid genomics: Extreme variation in GC, gene and intron content and multiple inversions between a direct and inverted orientation of the rRNA repeat. The New Phytologist 222: 1061–1075.

Nunes GL, Oliveira RRM, Guimarães JTF, Giulietti AM, Caldeira C, Vasconcelos S, Pires E, Dias M, Watanabe M, Pereira J, Jafê R, Bandeira, Cinthia Helena MM, Carvalho-Filho N, da Silva EF, Rodrigues TM, Dos Santos, Fernando Marino Gomes, Fernandes T, Castilho A, Souza-Filho PWM, Imperatriz-Fonseca V, Siqueira JO, Alves R, Oliveira G. 2018. Quillworts from the Amazon: A multidisciplinary populational study on Isoetes serracarajensis and Isoetes cangae. PloS One 13: e0204147.

Oldenkott B, Yamaguchi K, Tsuji Y, Zhang H, Cardio AC, VanHerckhoven P, Gilmore SJ, Tanaka J, Williams MD, Hidalgo J, et al. 2012. The New York Botanical Garden’s Isoetes repository: A catalog of living species and their collections. Kew Bulletin 67: 5–16.

Oldenkott B, Yamaguchi K, Tsuji Y, Zhang H. 2012. The New York Botanical Garden’s Isoetes repository: A catalog of living species and their collections. Kew Bulletin 67: 5–16.
Pfeiffer NE. 1922. Monograph of the Isoetaceae. Annals of the Missouri Botanical Garden 9: 79–233.

Pigg KB. 1992. Evolution of Isoetalean lycopsids. Annals of the Missouri Botanical Garden 79: 589–612.

Pigg KB. 2001. Isoetalean lycopsid evolution: From the devonian to the present. American Fern Journal 91: 99–114.

Pryer KM, Schneider H, Smith AR, Crane PP, Hoot SB. 2006. Inferring the root of land plants. In: Jacquot JP, Gadal P eds. Exploring the plastid genome disparity of liverworts. New York, Oxford: Oxford University Press. 64–75.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Tracer 1.7. Systematic Biology 67: 901–904.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu S, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61: 539–542.

Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG. 2014. From algae to angiosperms-inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. BMC Evolutionary Biology 14: 23.

Rydin C, Wikström N. 2002. Phylogeny of Isoëtes (Lycopsida): Resolving basal relationships using rbcL sequences. Taxon 51: 83–89.

Schafran PW, Johnson G, Taylor WC, Zimmer EA, Musselman LJ. 2018a. Low-copy nuclear markers in Isoëtaceae (Isoëtaceae) identified with transcriptomes. Applications in Plant Sciences 6: e1142.

Schafran PW, Zimmer EA, Taylor WC, Musselman LJ. 2018b. A whole chloroplast genome phylogeny of diploid species of Isoëtes (Isoëtaceae, Lycopodiophyta) in the Southeastern United States. Castanea 83: 224–235.

Schallenberg-Rüdinger M, Knoop V. 2016. Coevolution of organelle RNA editing and nuclear specificity factors in early land plants. In: Jacquot JP, Gadalia P eds. Genomes and Evolution of Charophytes, Bryophytes, Lycophytes and Ferns. London, UK: Elsevier. 37–93. https://doi.org/10.1016/bs.abr.2016.01.002

Schuettpez E, Hoot SB. 2006. Inferring the root of Isoëtes: Exploring alternatives in the absence of an acceptable outgroup. Systematic Botany 31: 258–270.

Smith DR. 2009. Unparalleled GC content in the plastid DNA of Selaginella. Plant Molecular Biology 71: 627–639.

Soltis DE, Gitzendanner MA, Stull G, Chester M, Chanderbali A, Chamala S, Jordon-Thaden I, Soltis PS, Schnable PS, Barbazuk WB. 2015. The potential of genomics in plant systematics. Taxon 62: 886–898.

Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688–2690.

Sugita M, Sugasita M. 2004. Plastid transformation reveals that moss tRNA(Arg)-CCG is not essential for plastid function. The Plant Journal for Cell and Molecular Biology 40: 314–321.

Suyama M, Torretta D, Bork P. 2006. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Research 34: W609–W612.

Taylor TN, Taylor EL, Kings M. 2009. Paleobotany: The biology and evolution of fossil plants. London: Academic.

Taylor WC, Hickey RJ. 1992. Habitat, evolution, and speciation in Isoëtes. Annals of the Missouri Botanical Garden 79: 613–622.

Taylor WC, Luebke NT, Britton DM, Hickey RJ, Brunton DF. 1994. Isoëtaceae. In: Flora of North America Committee ed. Flora of North America and North of Mexico. 2. New York, Oxford: Oxford University Press. 64–75.

Taylor WC, Moran RC, Brunton DF. 2016. Isoëtaceae, the quillwort family. In: NaccI A ed. New manual of vascular plants. New York: The New York Botanical Garden Press. 1–9. https://doi.org/10.2113/93275471.015

Tillich M, Lehmpark W, Pelliizzer T, Ulbrich-Jones E, Fischer A, Bock R, Greiner S. 2017. GeSeq - versatile and accurate annotation of organelle genomes. Nucleic Acids Research 45: W6–W11.

Troia A, Pereira JB, Kim C, Taylor WC. 2016. The genus Isoetes (Isoëtaceae): A provisional checklist of the accepted and unresolved taxa. Phytotaxa 277: 101.

Villaereal AJC, Turmel M, Bourgoun-Couture M, Laroche J, Salazar Allen N, Li F-W, Cheng S, Renzaglia K, Lemieux C. 2018. Genome-wide organellar analyses from the hornwort Leiosporoceros dussii show low frequency of RNA editing. PLoS One 13: e0200491.

Wei R, Yan Y-H, Harris AJ, Kang J-S, Shen H, Xiang Q-P, Zhang X-C. 2017. Plastid phylogenomics resolve deep relationships among eupolyplid II ferns with rapid radiation and rate heterogeneity. Genome Biology and Evolution 9: 1646–1657.

Weising K, Nybom H, Wolf K, Kahl G. 2005. DNA fingerprinting in plants: Principles, methods, and applications. New York: CRC Press.

Wolf PG, Karol KG, Mandoli DF, Kuehli J, Arumuganathan K, Ellis MW, Mishler BD, Kelch DG, Olmstead RG, Boore JL. 2005. The first complete chloroplast genome sequence of a lycophyte, Huperzia lucidula (Lycopodiaceae). Gene 350: 117–128.

Wolf PG, Rowe CA, Sinclair RB, Hasebe M. 2003. Complete nucleotide sequence of the chloroplast genome from a Leptosporangiate fern, Adiantum capillus-veneris L. DNA Research 10: 59–65.

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24: 1586–1591.

Yates TL, Salazar-Bravo J, Dragoo JW. 2004. The importance of knowing the tree of life. In: Donoghue MJ, Cracraft J eds. Assembling the tree of life. New York: Oxford University Press. 7–17.

Yu Y, Liu H, Yang J, Ma W, Pressel S, Wu Y, Schneider H. 2019. Exploring the plastid genome disparity of liverworts. Journal of Systematics and Evolution 57: 382–394.

**Supplementary Material**

The following supplementary material is available online for this article at [http://onlinelibrary.wiley.com/doi/10.1111/jse.12693/supinfo](http://onlinelibrary.wiley.com/doi/10.1111/jse.12693/supinfo)

**Fig. S1.** Results of the non-synonymous to synonymous rate ratio (dN/dS) analyses.

A. Boxplots showing no significant variation of dN/dS ratios between protein-coding genes with and without internal stop codon. B. Histogram showing of dN/dS ratios of all shared chloroplast protein-coding genes in 25 species (21 Isoetes spp., 02 Selaginella spp., and 02 Huperzia spp.).

**Table S1.** List of species, locality, voucher and Genbank accession number of plastomes included in the phylogenetic analyses. Asterisks point six newly sequenced plastomes species.
Table S2. List of genes and pseudogenes found in six newly assembled chloroplasts in Isoetes with their respective copy numbers.

Table S3. The non-synonymous to synonymous rate ratio ($d_{N}/d_{S}$) of shared protein-coding genes in the plastome of Isoetes, Selaginella and Huperzia with their correspondent aligned sequence length in base pairs (bp).

Table S4. Matrix of genomic traits mapped on the phylogenetic tree of Isoetes.