PLOS ONE

RESEARCH ARTICLE

Phylogenomics indicates the "living fossil" *Isoetes* diversified in the Cenozoic

Daniel Wood[1¤]*, Guillaume Besnard[2], David J. Beerling[1], Colin P. Osborne[1], Pascal-Antoine Christin[1]

1 Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield, United Kingdom, **2** CNRS, Université de Toulouse, IRD, UMR 5174, EDB (Laboratoire Évolution & Diversité Biologique), Toulouse, France

¤ Current address: Molecular Ecology and Fisheries Genetics Laboratory, Bangor University, Bangor, Gwynedd, United Kingdom
* daniel.wood@bangor.ac.uk

Citation: Wood D, Besnard G, Beerling DJ, Osborne CP, Christin P-A (2020) Phylogenomics indicates the "living fossil" *Isoetes* diversified in the Cenozoic. PLoS ONE 15(6): e0227525. https://doi.org/10.1371/journal.pone.0227525

Editor: William Oki Wong, Indiana University Bloomington, UNITED STATES

Received: December 7, 2019

Accepted: May 14, 2020

Published: June 18, 2020

Data Availability Statement: All relevant data are either a) within the paper and its Supporting Information files, or b) BioProject PRJNA548050 on the NCBI database.

Funding: PAC is supported by a Royal Society University Research Fellowships (number URF120119). https://royalsociety.org/grants-schemes-awards/grants/university-research/. DW was funded by a University of Sheffield Faculty Scholarship, 149440211, https://www.sheffield.ac.uk/. The funders had no role in study design, data

Abstract

The fossil record provides an invaluable insight into the temporal origins of extant lineages of organisms. However, establishing the relationships between fossils and extant lineages can be difficult in groups with low rates of morphological change over time. Molecular dating can potentially circumvent this issue by allowing distant fossils to act as calibration points, but rate variation across large evolutionary scales can bias such analyses. In this study, we apply multiple dating methods to genome-wide datasets to infer the origin of extant species of *Isoetes*, a group of mostly aquatic and semi-aquatic isoetalean lycopsids, which closely resemble fossil forms dating back to the Triassic. Rate variation observed in chloroplast genomes hampers accurate dating, but genome-wide nuclear markers place the origin of extant diversity within this group in the mid-Paleogene, 45–60 million years ago. Our genomic analyses coupled with a careful evaluation of the fossil record indicate that despite resembling forms from the Triassic, extant *Isoetes* species do not represent the remnants of an ancient and widespread group, but instead have spread around the globe in the relatively recent past.

Introduction

Determining the evolutionary relationships and divergence times between lineages is crucial for understanding the processes that generate diversity and evolutionary novelty [1–3]. Fossils provide a glimpse of the past, by preserving the anatomical features of organisms that existed millions or hundreds of millions of years ago. The fossil record is however very incomplete and often needs to be combined with analyses of extant diversity to infer periods of diversification and extinction of different lineages [4]. A fossil assigned to a lineage of organisms based on shared morphology provides a minimal age for the group, and can therefore help date evolutionary events. Putative causal factors can then be inferred for these events, such as the Chicxulub meteorite impact and the disappearance of the non-avian dinosaurs [5], or the coincident radiation of angiosperm and insect lineages [6]. In many cases, however, fossils

collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

assignable to particular extant lineages of organisms are unavailable, because of a lack of readily fossilisable tissues (e.g. jellyfish) or because the organisms live in environments that do not favour fossilization (e.g. cacti). In addition, morphological traits preserved in fossils may not vary sufficiently to distinguish multiple extant lineages, preventing a precise assignment of the fossils [7–9]. When this pattern of conservation concerns a large number of morphological traits, the extant species are referred to as "living fossils", a category that includes the coelacanth, cycads, sturgeons, platypus and lungfish that closely resemble fossils from the Mesozoic [10–14]. Clearly a lack of morphological change does not preclude changes in traits poorly represented in the fossil record, such as biochemical or behavioural changes–nevertheless, their unusually conserved morphology through time has long attracted the interest of biologists [15, 16]. This morphological stasis is often associated with decline–with current distributions of "living fossil" taxa interpreted as the remnants of larger ancestral ranges [17, 18], and extant species being the last members of ancient lineages diverging long in the past [19, 20]. However, this morphological uniformity and the resulting difficulties in fossil assignment mean that these hypotheses are difficult to test from the fossil record alone.

Analyses of DNA sequences over the last few decades have resolved the phylogenetic relationships between many extant lineages, and large numbers of selectively neutral changes in the genomes allowed inferring accurate phylogenies even for the most morphologically uniform organisms [21, 22]. Molecular divergence in parts of the tree with informative fossils can then be used to time-calibrate molecular changes in the rest of the tree, allowing inference of divergence times of groups of organisms lacking an appropriate fossil record [23–26]. This molecular dating technique, alongside other methods, has been used to investigate the evolutionary dynamics of some "living fossil" groups. In some cases, some of the expected features of "living fossil" groups are found, such as ancient within-group divergence, extant distributions resulting from continental drift tens of millions of years ago, low levels of genetic diversity and small ranges (e.g. coelacanths, the Cupressaceae, horshoe crabs [17, 27, 28]). In other groups such as birchirs [29], tadpole shrimp [30], cycads [13], bryophytes [31] and *Ginkgo* [26], however, extant diversity originated more recently than their conserved morphology would suggest, indicating complex evolutionary dynamics for some "living fossil" taxa. Age estimates from molecular dating techniques remain however sensitive to the treatment of fossil data, variability in the rates of nucleotide substitutions between molecular markers and species, and the correct alignment of nucleotide markers [24, 32–37]. These problems are exacerbated when the only available calibration points are distant from the group of interest, as is by definition the case for "living fossils" [28, 38–41]. Each possible source of error therefore needs to be isolated and carefully considered.

The lycopod genus *Isoetes* exemplifies many of the problems of "living fossil" taxa. The genus has long been of interest due to its status as the last lineage of the isoetalean lycopods. This group, known from at least the late Devonian, dominated terrestrial floras in the Carboniferous [42]. The extant *Isoetes* genus is a small herbaceous aquatic or semi-aquatic plant, generally lacking a stem and consisting of a number of stiff leaves atop a woody corm [43]. It demonstrates a number of unusual features such as roots comparable to fossil stigmarian rootlets [44] and aquatic Crassulacean Acid Metabolism (CAM) [45]. Fossils resembling the *Isoetes* growth form are found in the Triassic onwards, although their exact affinities and relationships to *Isoetes* are unclear [46–48]. A variety of morphological features (such as sunken sporangia, an elaboration of the basal part of the ligule into a glossopodium, and a velum or labium covering the sporangium) that characterise extant *Isoetes* appear at this time, although no single fossil displays all of these features [49]. The appearance of *Isoetites rolandii* in the Jurassic represents the earliest clear example of a isoetalean lycopsid containing all the major features uniting modern *Isoetes*, including the loss of both vegetative leaves and an elongating stem [49,

50], although elongated-stem forms such as *Nathorstiana* persisted until the Early Cretaceous [51]. Fossils of plants presenting the modern *Isoetes* growth form (e.g. *I. horridus*) are subsequently found from the Early Cretaceous and into the Tertiary [46, 48]. Within extant *Isoetes* lineages, a number of reductions from three to two lobed corms have occurred [52], along with transitions to a variety of habitats from ephemeral pools to oligotrophic lakes [43]. In summary, the overall morphology of *Isoetes* appears to have persisted virtually unchanged since at least the Jurassic, and the general growth habit in the lineage is potentially as old as the Triassic.

The close resemblance of fossil taxa to modern *Isoetes* suggests the extant species could be the remnants of a very ancient genus. However, establishing the relationship between these fossils and modern *Isoetes* has proven difficult due to the highly conserved morphology of the genus [18, 42, 48, 50]. The more than 200 extant *Isoetes* species have a global distribution, yet display very little morphological variation–features such as spore morphology, corm lobation and habitat are currently used to distinguish extant species, but many of these features are homoplastic or variable within species [43, 50, 52–54]. Morphology and the fossil record alone therefore provide limited insights into the relationships among extant and fossil species of *Isoetes*, restricting our ability to understand the temporal origins of extant *Isoetes* species diversity.

The relationships between extant species of *Isoetes* have been inferred using molecular phylogenetics [18, 55, 56], but attempts at linking fossils and extant species have not always been successful. Taylor and Hickey [43] hypothesised based on shared leaf morphology that a small group of South American species and fossil *Isoetes* represented the earliest split within the genus, but molecular phylogenetics falsified this hypothesis for extant species [18, 55]. Recent molecular dating studies suggest an origin of extant *Isoetes* species diversity in the Triassic to Jurassic, with species distributions consistent with the breakup of the Gondwana supercontinent [18, 57, 58]. These studies were, however, based on a limited number of markers, mainly from chloroplast genomes, where high rate variation can make dating estimates especially dependent on molecular clock model assumptions [24]. This is a potentially significant source of error given the ancient divergence between *Isoetes* and its sister group *Selaginella*, resulting in a large genetic distance between the nearest calibration point and our node of interest [48, 59]. We therefore decided to re-evaluate the divergence times within *Isoetes* using a combination of phylogenomic methods capturing markers spread across the genomes of numerous land plants.

In this study, we generate transcriptomes and genomic datasets for multiple *Isoetes* species and apply multiple molecular dating approaches to estimate the time to the most recent common ancestor of extant *Isoetes* based on nuclear and plastid genomes. Our results shed new light on the age and evolutionary dynamics of this "living fossil" lineage, and show how careful integration of large genomic datasets can help analyses of groups with a poorly informative fossil record.

Materials and methods

Ethics statement

Live plants were collected from Cwm Idwal, UK, with permission from Natural Resources Wales, the Snowdonia National Park Authority, and the landowners (National Trust). No permit was required for this collection and no protected species were sampled. DNA was acquired from preserved modern specimens permanently deposited at Kew Gardens, UK–specimen numbers are available in S1 Table.

General approach

In this study, we selected six *Isoetes* for generating genome-wide DNA datasets–*I. coromandelina* (this specimen is referred to as *I. coromandelina* sensu lato to reflect the taxonomic complexity of this species–see Pantil and Rajput [60]), *I. humilor*, *I. elatior*, *I. nuttallii*, *I. lacustris* and *I. andicola*. These were selected to capture the deeper divergence events within this group based on previous molecular studies [18]. Analyses of nuclear ribosomal DNA available for a large number of *Isoetes* confirmed that the last common ancestor of the selected species likely corresponds to the last common ancestor of extant *Isoetes*, and the low branch length variability throughout the genus suggests the sequenced species represent a good sample of evolutionary rates within the genus (Fig 1). Sparse taxon sampling has been shown to significantly affect

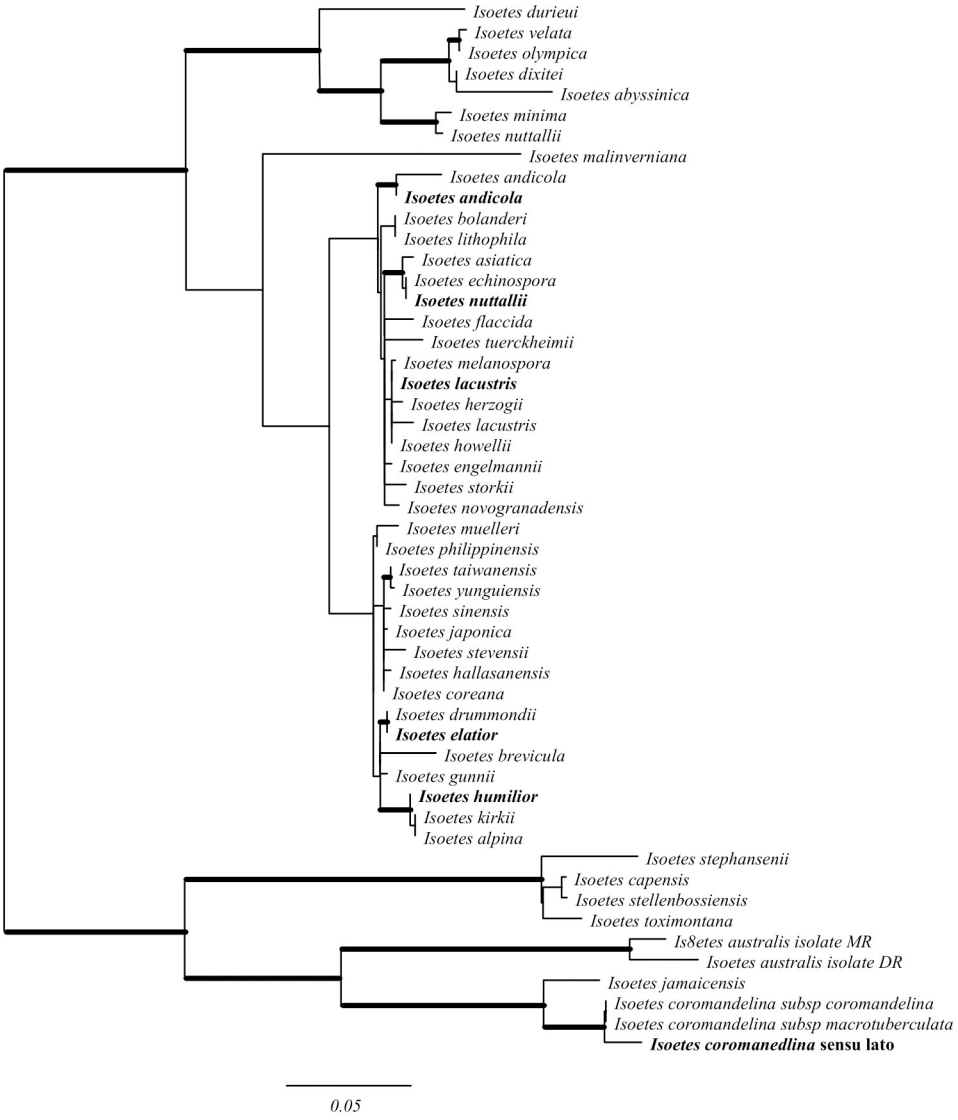


Fig 1. Maximum likelihood phylogram of *Isoetes* nuclear ribosomal internal transcribed spacer. Branch lengths are proportional to the expected number of substitutions per site, with scale bar representing 0.05 substitutions per site. Branches in bold have bootstrap support values greater than 90. Species in bold represent data generated in this study; nuclear ribosomal internal transcribed spacers for other *Isoetes* are from those used in Larsén and Rydin [18].

https://doi.org/10.1371/journal.pone.0227525.g001

estimated dates in some molecular dating studies [61, 62], although not in every case [63, 64]. Rate heterogeneity likely plays an important role in the effect of sparse taxon sampling on the accuracy of molecular dating, with high levels of rate heterogeneity demanding more sampling [62]. The relatively low levels of rate heterogeneity within *Isoetes* (Fig 1) suggest it is a suitable group to perform molecular dating with a relatively small number of taxa. To further investigate the impact of our sampling scheme, we reanalysed the dataset of Larsén and Rydin [18], which contains 45 *Isoetes* species including all the major clades identified by previous studies of *Isoetes* [56, 65]. This dataset was reanalysed using the same constraints and BEAST settings as Larsén and Rydin [18], but with the 45 *Isoetes* species used reduced to the 6 closest relatives of our chosen species (*I. asiatica*, *I. coromandelina* sensu lato, *I. drummondii*, *I. echinospora*, *I. kirkii* and *I. storkii*). This resulted in an estimated crown date of *Isoetes* of 153.4 Ma (53.9–277.3 95% CI), only a 7% increase compared to the full species sampling. This indicates limited taxon sampling should not substantially alter the estimation of the *Isoetes* crown node date.

DNA from these species was then used to generate genome-wide datasets, and different genome partitions were analysed in isolation to get accurate estimates of divergence times. Herbarium specimens represent a useful source of DNA, particularly for globally distributed, hard-to-access groups such as *Isoetes* [66, 67]. Low-coverage whole-genome scans can be applied to these samples, and will yield high coverage for genomic fractions present as multiple copies, such as the organellar genomes [68]. However, highly variable evolutionary rates in chloroplast markers have been reported from seed plants [69, 70], which potentially affect the results of dating methods that differ in their assumptions of rate heterogeneity [24].

Previous studies of the chloroplast marker *rbcL* indicate much higher rates of sequence evolution in *Selaginella* than in *Isoetes* [18, 70, 71]. Nuclear markers can be more useful for molecular dating if they show less variation in rates among branches [24] as suggested from large scale embryophyte phylotranscriptomics [72, 73]. Genome skimming can provide nuclear sequences, but low coverage makes *de novo* assembly difficult. However, the sequencing reads can be mapped to a reference dataset, providing phylogenetically informative characters [74, 75]. A reference genome is available for *Selaginella*, but it is too distant to allow accurate mapping of reads from *Isoetes*. Transcriptomes provide high coverage of expressed protein-encoding genes, which represent regions of the genome allowing read mapping across distinct species [74, 75]. We consequently decided to generate and assemble a transcriptome for a single *Isoetes* species, which was used as a reference to map reads from low-coverage whole-genome sequencing datasets obtained from the other *Isoetes* species sampled from herbarium collections. The sequencing data were used to obtain chloroplast and nuclear alignments for five *Isoetes* species as well as a number of other land plants (mosses, ferns, lycopods, gymnosperms and angiosperms) sequenced in other studies. The phylogenetic breadth of the datasets allowed the incorporation of fossil evidence providing calibration points spread across the tree.

Sequence acquisition

Live *Isoetes lacustris* were sampled from Cwm Idwal, Wales and maintained at the University of Sheffield in 40 x 30 x 25 cm transparent plastic containers, with a substrate of sand to a depth of 5 cm, and the containers filled to the top with deionised water. These were placed in a Conviron growth chamber with a 12-h day/night cycle, 495 μmol m^2s^{-1} light, temperature at 20˚C during the day and 18˚C at night, and CO_2 at 400 ppm for six days. To maximise the number of transcripts retrieved, leaves from three individuals were sampled 3 hours after dark and 3 hours after light and stored immediately in liquid nitrogen. We also generated a transcriptome for *Littorella uniflora* (Plantaginaceae), another species of aquatic plant that shares

aquatic CAM photosynthesis [76]. Individuals from this species were also sampled from Cwm Idwal and were grown under a variety of conditions before sampling their leaves as described above. Dried specimens were deposited in the Sheffield University Herbarium (*I. lacustris*–DW1, *L. uniflora*–DW2).

RNA was extracted from the sampled leaves using the RNeasy® Plant Mini Kit (Qiagen), following the manufacturer protocol, with the addition of on-column DNase I digestion (Qiagen RNase-Free DNase Set). We then added 2.5 μl SUPERase-In™ RNase inhibitor (Invitrogen) to 50 μl of extracted RNA to stabilise it. RNA was quantified using a gel electrophoresis, RNA 6000 Nano chips (Aligent) in an Aligent 2100 Bioanalyser, and a Nanodrop 8000. Samples were then prepared for Illumina sequencing using the TruSeq® RNA Sample Prep Kit v2 (Illumina). Paired-end sequencing was performed on an Illumina HiSeq 2500 platform available at Sheffield Diagnostic Genetics Service in rapid mode for 100 cycles, with 24 libraries pooled per lane of flow cell (other samples were from the same or different projects).

DNA from herbarium specimens of five *Isoetes* species were acquired from the DNA Bank from the Royal Botanical Gardens, Kew (S1 Table). This was supplemented with one silica gel dried leaf each of *I. lacustris* and *L. uniflora* collected from the field as described previously. Whole genome sequencing of these seven samples was performed at the Genotoul from the University of Toulouse, using previously described protocols [74, 77]. Each sample was sequenced on a 24th of a lane of a flow cell, with other samples from various projects. Raw sequencing reads were cleaned using NGS QC toolkit v2.3.3 [78] by removing adapter sequences, reads with ambiguous bases and reads with less than 80% of positions with a quality score above 20. Low quality bases ($q<20$) were removed from the 3' end of remaining reads. Species identity and branch length variability within the genus were assessed by assembling the nuclear ribosomal internal transcribed spacer (nrITS) using NOVOPlasty 2.5.9 [79]. The assembled sequences were aligned to nrITS sequences used in Larsén and Rydin (2015) [18] using MAFFT v7.164 [80]. A phylogeny for this marker was then produced using RAxML v8.2.11 [81], with a GTR + G + I substitution model, identified as the best-fit substitution model through hierarchical likelihood ratio tests (Fig 1).

Chloroplast data matrix

Cleaned reads from *Isoetes* and *Littorella* corresponding to the chloroplast genomes were assembled using NOVOPlasty, with a 39-bp kmer and a seed sequence of the *I. flaccida* chloroplast genome [71]. In cases where a circular chloroplast genome was not produced, contigs were aligned to the *I. flaccida* chloroplast genome using blastalign [82], and reads corresponding to regions of the reference chloroplast genome not covered by the contigs were used as seed sequences to assemble new contigs. All contigs were subsequently realigned to the reference genome, and overlapping contigs were merged. Chloroplast genome assemblies from 24 additional species representing the major embryophyte taxa, including two *Selaginella* species, were downloaded from NCBI database [83–101] (S2 Table).

Chloroplast protein-coding genes were identified from all chloroplast genomes using DOGMA [102] and coding sequences were extracted using TransDecoder v2.1.0 [103]. A total of 64 genes were identified and aligned by predicted amino acids using t-coffee [104] and MAFFT. Gene alignments were manually inspected and trimmed using AliView [105]. Twelve of them (*clpP*, *cysaA*, *psi_psbT*, *rpl16*, *rpl21*, *rps15*, *ycf1*, *ycf2*, *ycf3*, *ycf10*, *ycf66*, *ycf68)* were discarded either due to poor homology or alignment difficulties, or because sequences were obtained from less than 10 of the 33 chloroplast genomes analysed. The remaining 52 chloroplast genes were concatenated, producing a 55,542 bp matrix, with 33,582 polymorphic and 25,501 parsimony informative sites. A maximum likelihood phylogeny was generated using

RAxML, with a GTR + G + I model of sequence evolution, determined to be the best-fit model using hierarchical likelihood ratio tests. The same matrix was later used for molecular dating.

Nuclear data matrices

Cleaned RNAseq reads of *I. lacustris* were assembled using Trinity v2.3.2 [103], resulting in 285,613 contigs with an average length of 689 bp. A similar procedure yielded 159,920 contigs for *L. uniflora*, with an average length of 769bp. For each species, the longest open reading frames (ORFs) were extracted using TransDecoder, and for each unigene the contig with the longest ORF was used to build a reference dataset. Cleaned reads from the whole-genome sequencing for each of the *Isoetes* species were then separately mapped to this reference dataset using bowtie2 v2.3.2 [106] in local mode to avoid excluding reads overlapping exon/intron boundaries. Alignments with MAPQ quality below 20 were excluded using SAMtools v1.5 [107]. The SAMtools mpileup utility was then used to generate for each species consensus sequences from the reads mapping to each *I. lacustris* transcript.

Gene duplication and losses are common in nuclear genomes–polyploidy is common in *Isoete*s, including in *I. lacustris* which has previously been identified as a decaploid [53]. Therefore, a combined reciprocal best blast and phylogenetic approach was adopted to identify groups of co-orthologs covering *I. lacustris* and the other land plants. Families of homologous ORFs generated by the method of Vilella et al. [108] were downloaded from EnsemblPlants. In total, 4,516 homolog families highly conserved among land plants (containing at least one sequence from *Physcomitrella patens*, *Selaginella moellendorffii*, *Amborella trichopoda*, *Oryza sativa*, *Arabidopsis thaliana* and *Theobroma cacao*) were used for subsequent ortholog identification.

Transcriptome and coding sequence data from seven additional species representing different embryophyte groups were retrieved from the literature [109–115] (S3 Table) and ORFs were extracted. Reciprocal best protein BLAST searches assigned ORFs of *I. lacustris*, *L. uniflora* and the additional embryophyte species to homolog families, with a minimum match length of 50 amino acids and e-value of 10^{-7}. The expanded homolog families were then aligned according to their protein sequences using MAFFT, and phylogenies were constructed using RAxML and the GTR + G + I model, which fits most genes and is therefore appropriate for constructing large numbers of gene trees [116–118]. The longest sequence of each monospecific clade of sequences belonging to the same species was identified using custom scripts to remove transcripts representing the same gene or genes that duplicated after the divergence from all other species. These sequences were then realigned and a new phylogeny was inferred. Sets of 1:1 orthologs were then identified as clades containing exactly one gene per species, resulting in 30,258 groups of co-orthologs. Of these, 2,165 contained more than nine species, including *I. lacustris*, *S. moellendorffii* and either *P. patens* or *Ceratodon purpurea*, which were needed to use some of the fossil calibration points. By restricting our analysis to these 1:1 orthologs, we eliminate the possibility of non-orthologous genes resulting from gene or genome duplications being considered as orthologs. These 2,165 orthogroups were realigned, and consensus sequences of the genome skimming data were added to the alignments. Only the 782 orthogroups containing sequences for *I. coromandelina* sensu lato, which is necessary to capture the earliest split among extant species of *Isoetes* [18] (Fig 1), were considered further. New phylogenetic trees were inferred from these datasets, and genes failing to recover the monophyly of the vascular plants, Isoetopsida (*Isoetes* plus *Selaginella*, [119]) or *Isoetes* were considered phylogenetically uninformative and excluded. The remaining 292 datasets were deemed suitable for the phylogenetic problem addressed here, and were used for molecular dating. A phylogenetic tree was inferred separately for each of these markers, and a maximum

likelihood phylogeny was also inferred using the 694,437 bp concatenated alignment, which was 41.14% complete with 443,864 polymorphic and 316,350 parsimony informative sites.

Calibration points

Time-calibrated trees were inferred from the different markers using the same set of calibration points. To date the crown node of extant *Isoetes*, a fossil constraining the crown node of extant *Isoetes* would be ideal–such a constraint would require a fossil containing a synapomorphy from one of the two descendant branches of this node. Previous studies have identified a geographically diverse group of *Isoetes* including *I. coromandelina* sensu lato as the outgroup to the rest of the *Isoetes* [18, 56, 65]. Whilst the *I. coromandelina* complex itself contains a number of features initially thought to identify this as diverging earliest from other extant members of the group, its presence within "Clade A" identifies these features as derived [18, 55]. No morphological features reliably distinguishing Clade A and the rest of the *Isoetes* appear to exist [18, 55]. Therefore, no fossil will contain features distinguishing these two groups, so fossils cannot provide a hard minimum age for this node. Within the *Isoetes* crown group, no morphological features clearly divide different clades [18, 42, 58]. A number of features vary between taxa and clades, such as corm lobation and glossopodium structure, but these are either not widely characterised across the genus or show multiple transitions within clades [52, 120]. The fossil record therefore does not allow implementing hard minimum ages for nodes within *Isoetes*.

The nearest node to the *Isoetes* crown node for which reliable synapomorphies are available is the crown node of the Isoetopsida (*Isoetes* plus *Selaginella*). The Isoeptopsida are a well supported clade appearing in the Devonian, containing synapomorphies such as a heterospory and a ligule [119, 121]. Isoetalean lycopsid trees are considered to form a clade within the Isoetales, being more closely related to *Isoetes* than *Selaginella*. This assessment is based on shared synapomorphies including bipolar growth from a shoot like "rhizomorph" structure and secondary woody tissue [48]. Arborescent lycopsids are known from the Frasnian [122, 123] (382.7–372.2 Ma), although the rhizomorph root structure could not be identified in these early fossils. However, discovery of a putatively homosporous arborescent lycopsid (the Isoetales are heterosporous) suggests that arborescence could be a convergent phenotype within the lycopods [124]. As multiple examples of isoetalean arborescent lycopsids, including rhizomorphs, are known from Famennian strata [125, 126] (358.9 to 372.2 Ma), a minimal age of 358 Ma was implemented using a uniform distribution between 358 and 485 Ma.

A maximum age constraint of the crown node of all land plants was set based on the appearance of cryptospores in the fossil record. These abundant spores are considered a likely synapomorphy of early land plants [127]. Their appearance in the fossil record is therefore likely to occur soon after the origins of land plants, making them appropriate for setting a maximum age for land plants [18, 128]. The earliest unequivocal cryptospores are found in the early Middle Ordovician [129] (473–471 Ma). However, pre-Middle Ordovician terrestrial sediments are rare [130], and as no unequivocal cryptospores are found in pre-Ordovician rocks [128, 131], the beginning of the Ordovician (485 Ma) was used as a conservative upper limit for the age of land plants. Whilst other molecular dating studies have estimated the age for this node to be comparable or older than this date (e.g.[132–134]), these are based on comparable fossil evidence but with soft maxima assigned to this node, allowing older age estimates than the hard maximum approach used in the present study. This maximum age was used to constrain the crown node of the liverworts plus the rest of vascular plants in the chloroplast dataset, and the crown node of the bryophytes plus vascular plants in the nuclear dataset. The minimum age of the same node in both cases was constrained by the early vascular plant macrofossil, *Baragwanathia longifolia* from the Ludlow epoch in the Silurian at 421 Ma [135–137].

This is of a similar age to other putative vascular plant fossils such as *Cooksonia* in the Wenlock epoch [138–140].

For the chloroplast dataset, trees were rooted by constraining each of the liverworts and the rest of the land plants to be monophyletic [136]. For the nuclear dataset, which only contained bryophytes and vascular plants, the tree was rooted by enforcing the monophyly of each of these two groups. Details of fossil calibrations as outlined in Parham et al. [141] are outlined in S4 Table.

Molecular dating software

Molecular dating was performed using r8s [142] and BEAST [143], two commonly used relaxed-clock methods that differ in their general approach and the strategy used to assign rates to internal branches of the phylogeny. r8s implements a semiparametric method that uses a penalised likelihood approach to assign rates among branches [144]. The smoothing parameter, which determines the extent to which rates vary among branches, is determined for each dataset using an empirical approach [142]. The method takes a phylogram as input, assumes no uncertainty in topology, and uses a simplified model of nucleotide substitution. BEAST implements a highly parametrised Bayesian method that samples trees generated from nucleotide data using an explicit model of sequence evolution [145]. When using the relaxed molecular clocks implemented in BEAST, rates are uncorrelated across the tree, but an overall distribution of rates is assumed, with the mean and standard deviation inferred from the data.

For r8s, version 1.81 was used, with the "TN" algorithm and additive penalty function. Cross validation was performed for a range of smoothing parameters from 10^{-2} to 10^{6}, increasing by a power of $10^{0.5}$ each time, and the best smoothing parameter was used for molecular dating. Confidence intervals were obtained by generating 100 bootstrap pseudo-replicates using seqboot [146] and obtaining branch lengths for each of these using RaxML (GTR+G+I) while constraining the trees to the topology generated by the original dataset. These trees were then individually dated using r8s, providing a distribution of ages across the pseudo-replicates. This approach was used to date the chloroplast dataset, the concatenated nuclear dataset, as well as individual nuclear markers.

For BEAST, version v1.8.4 was used. A lognormal relaxed clock was adopted with a GTR + G + I model of nucleotide substitution with four rate categories and a birth-death speciation prior. For the concatenated chloroplast markers, four independent analyses were run for at least 20,000,000 generations and appropriate burn-in periods (at least 10%) were assigned by inspection of the traces using Tracer v1.6 [143]. For individual nuclear genes, BEAST was run for 3,000,000 generations (based on observing convergence times with a subset of genes) with a burn-in period of 50%. Dating the concatenated nuclear dataset was computationally too intensive with this approach. We therefore randomly subsampled 55,743bp (approximately the length of the chloroplast alignment) from the 694,437bp nuclear alignment eight times, and analysed these subsamples with BEAST. The same parameters as the individual nuclear genes were used, with the exception that BEAST was run twice for 10,000,000 generations for each alignment, with a burn-in period of 10%, with convergence verified using Tracer. These were combined using logcombiner and treeannotator to produce a maximum clade credibility tree. For comparison we performed r8s on these alignments as described previously.

Results

Phylogenetic reconstruction using *Isoetes* nrITS sequences

The maximum likelihood phylogeny based on nrITS broadly agrees with phylogenies previously inferred for the group (Fig 1) [18], and confirms that the species sampled for genomic

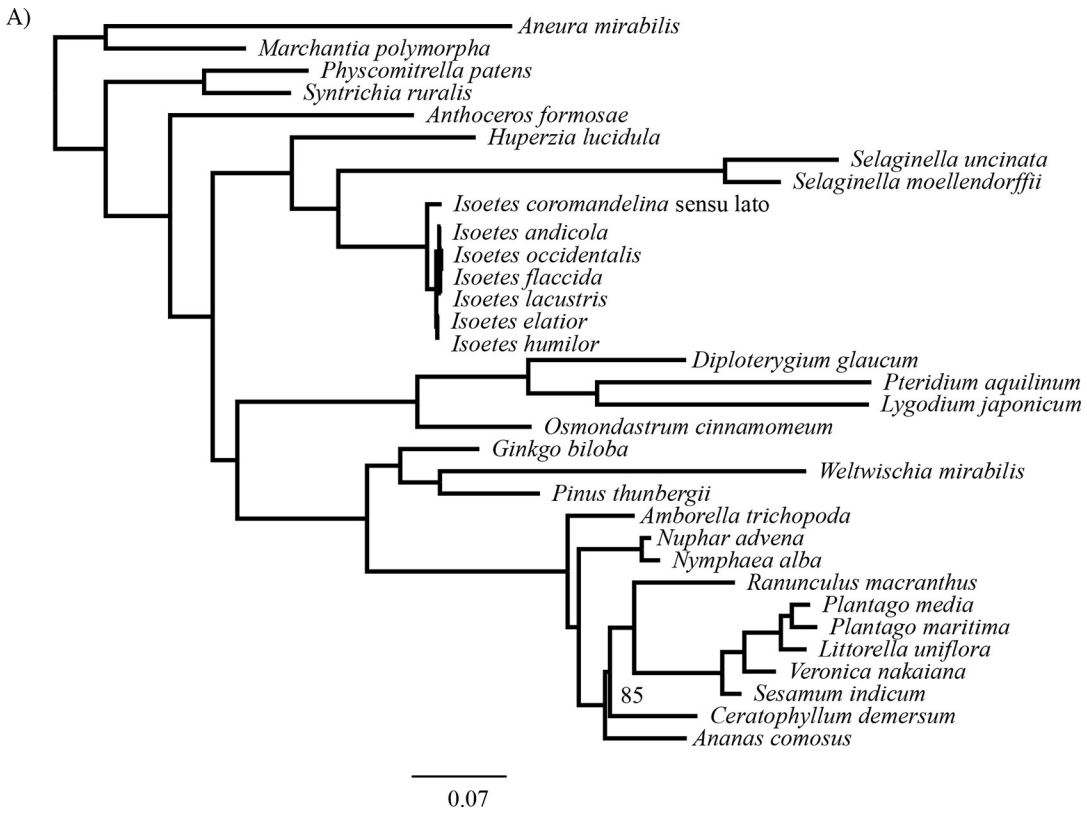
analyses encapsulate the first divergence within the *Isoetes* species for which molecular data is available (Fig 1). These six samples are moreover representative of the group in terms of branch lengths, and therefore evolutionary rates (Fig 1). The samples sequenced in this study generally cluster with other individuals from the same species sequenced previously, with the exception of *Isoetes nuttallii*. The newly sequenced sample of *I. nuttallii* groups with *I. asiatica* and *I. echinospora*, disagreeing with the topologies found in other studies [18, 55]. The sample in this study was collected on Knight Island in the Prince William Sound, Alaska, USA, outside the known range of *I. nuttallii* which extends up to British Columbia at its northernmost [147]. *I. asiatica*, *I. echinospora* and *I. occidentalis* and their interspecific hybrids occur throughout Alaska [148]. The underwater growth habit and large megaspores in this specimen are inconsistent with previously identified *I. nuttallii* specimens that are typically emergent [148, 149]. These features are consistent with the hexaploid *I. occidentalis*, a close relative of *I. asiatica* that has large megaspores, an underwater growth habit and is native to the area [43, 56, 149]. Furthermore, two duplicates of this collection have been subsequently redesignated as *I. occidentalis*. We therefore refer to this specimen as *I. occidentalis* throughout the remainder of this manuscript and associated data files.

Phylogenetic reconstruction and dating based on the chloroplast genome

The maximum likelihood phylogeny based on chloroplast markers recapitulated major land plant relationships and expected relationships within the *Isoetes* clade, with *I. coromandelina* sensu lato being sister to the rest of samples (Fig 2A). The tree was well resolved, with only the *Ceratophyllum*/eudicot split receiving less than 95% bootstrap support. Branch lengths were highly variable, particularly between *Isoetes* and *Selaginella*, with the latter having accumulated approximately 4.5 times more substitutions than *Isoetes* since their most recent common ancestor (Fig 2A).

Based on chloroplast markers, r8s estimated the age of the crown group of *Isoetes* at 24.2 Ma with an optimum smoothing parameter of 1000 identified by cross validation, and a 95% bootstrap confidence interval of 22.8–25.9 Ma (near the Oligocene-Miocene boundary; Table 1). Decreasing the value of the smoothing parameter resulted in an increased age of the *Isoetes* crown group, with a smoothing value of 0.01 giving a crown age of *Isoetes* of 219 Ma (Late Triassic; Fig 3). Whilst low smoothing values result in over-fitted models that perform poorly in cross validation, high levels of smoothing may produce rates that are nevertheless poor predictors of branch lengths in particular parts of the tree. For high smoothing values, the ratio of the effective rate (the branch length divided by the estimated time elapsed) to the rate assigned by the model was 0.33 for the stem branch of *Isoetes* (Fig 4), showing that the branch is significantly shorter than would be expected for the assigned rate and divergence time. On the other hand, the average ratio for the crown branch lengths remains near 1 for all smoothing values, indicating that the crown branches are close to the expected values for the assigned rates and divergence times (Fig 4).

For the same chloroplast markers, BEAST estimated the crown age of *Isoetes* at 23.2 Ma (95% HPD = 6.4–46.8 –; middle Oligocene; Table 1), similar to the value obtained with the optimum level of smoothing in r8s. Unlike in r8s, rates in BEAST can vary throughout the tree, but their distribution is assigned a priori–in this case a lognormal distribution. Rates in the maximum clade credibility tree accordingly follow a lognormal distribution—the log-transformed rates following a straight line on a quantile-quantile plot, indicating the rates are distributed lognormally (Fig 5). Notably, the *Isoetes* stem branch is assigned the lowest rate in the tree and the crown branches assigned rates closer to the average rates in the rest of the tree (Fig 5). While these rate assignments lead to a lognormal distribution that satisfies the priors, they result in a lower rate in the *Isoetes* stem branch compared to the crown branches.


A)
Aneura mirabilis
Marchantia polymorpha
Physcomitrella patens
Syntrichia ruralis
Anthoceros formosae
Huperzia lucidula
Selaginella uncinata
Selaginella moellendorffii
Isoetes coromandelina sensu lato
Isoetes andicola
Isoetes occidentalis
Isoetes flaccida
Isoetes lacustris
Isoetes elatior
Isoetes humilor
Diploterygium glaucum
Pteridium aquilinum
Lygodium japonicum
Osmondastrum cinnamomeum
Ginkgo biloba
Weltwischia mirabilis
Pinus thunbergii
Amborella trichopoda
Nuphar advena
Nymphaea alba
Ranunculus macranthus
Plantago media
Plantago maritima
Littorella uniflora
Veronica nakaiana
85
Sesamum indicum
Ceratophyllum demersum
Ananas comosus
0.07

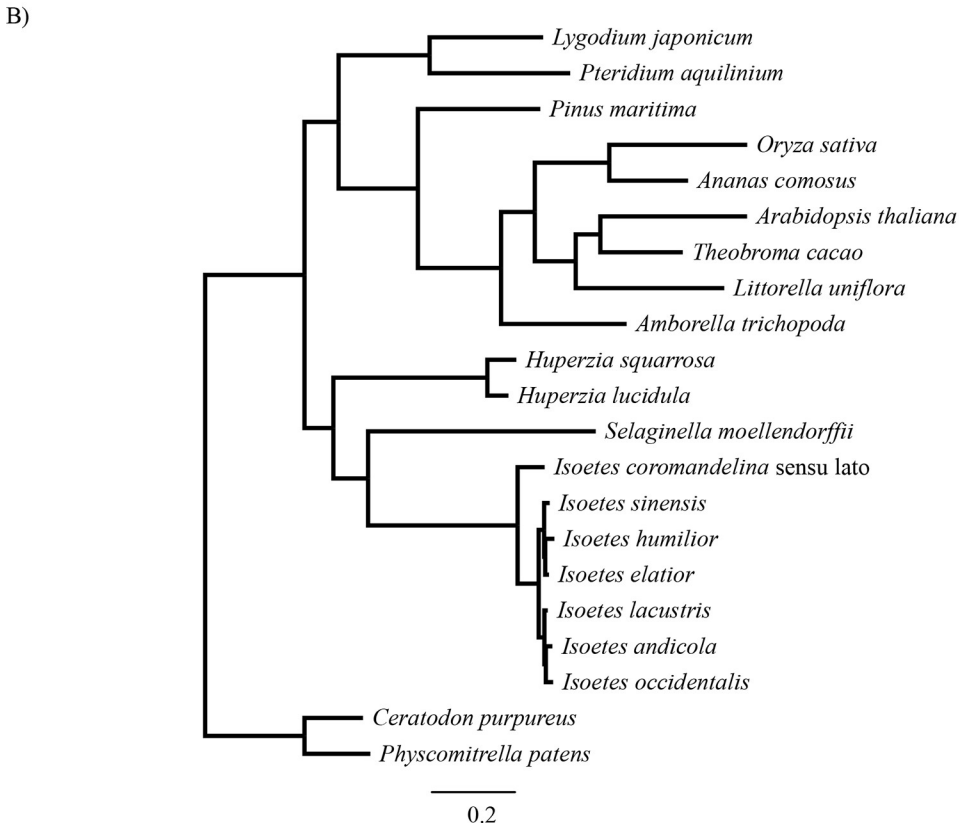

B)
Lygodium japonicum
Pteridium aquilinium
Pinus maritima
Oryza sativa
Ananas comosus
Arabidopsis thaliana
Theobroma cacao
Littorella uniflora
Amborella trichopoda
Huperzia squarrosa
Huperzia lucidula
Selaginella moellendorffii
Isoetes coromandelina sensu lato
Isoetes sinensis
Isoetes humilior
Isoetes elatior
Isoetes lacustris
Isoetes andicola
Isoetes occidentalis
Ceratodon purpureus
Physcomitrella patens
0.2

Fig 2. Maximum likelihood phylograms of concatenated chloroplast and nuclear markers. Phylograms are shown for a) concatenated chloroplast markers and b) concatenated nuclear markers. Branch lengths are proportional to the number of expected substitutions per site, with scale bar representing a) 0.07 and b) 0.2 substitutions per site. All bootstraps support values are 100 with the exception of the branch separating *Ananas comosus* from the clade containing *Ceratophylum demersum* in b), which has a support value of 85.

https://doi.org/10.1371/journal.pone.0227525.g002

For both r8s and BEAST, a date of 23–29 Ma (Oligocene) is obtained via the implicit or explicit inference of a decrease in the rate of evolution along the stem branch, with rates in the crown branches being more similar to those in the rest of the tree. This assumption results from the model, and is not necessarily correct, urging for independent evidence.

Phylogenetic reconstruction and dating based on nuclear markers

The concatenated nuclear phylogram also recapitulated major land plant relationships (Fig 2B). The topology of the *Isoetes* clade was consistent with that of the chloroplast phylogeny, with *I. coromandelina* again being sister to all other species. Despite overall longer branch lengths in the concatenated nuclear phylogeny, variation among groups was reduced. Particularly, the total branch lengths from the common ancestor of *Isoetes* and *Selaginella* were much more similar than in the chloroplast phylogeny, with *Selaginella* having accumulated approximately 1.25 times more mutations than *Isoetes* since their last common ancestor (Fig 2B). However, the ratio of the average crown branch length to stem length in the *Isoetes* lineage was very similar between the nuclear and chloroplast markers; approximately 5.8 for the chloroplast dataset and 5.6 for the nuclear dataset (Fig 2).

Dating of the concatenated matrix of nuclear markers in r8s gave an estimated crown node age of *Isoetes* at 58.9 Ma (Table 1), with an estimated stem node age of 358 Ma, at an optimum smoothing value of 0.1. Unlike with the chloroplast markers, the date of the *Isoetes* crown node was similar across all smoothing values tested (Fig 3). Increased smoothing values led to increases in the disparity between effective and assigned stem rates (Fig 4), although this was low compared with the concatenated chloroplast alignment (0.82 vs 0.33 for the stem branch for a smoothing value of 10^6). As with the chloroplast markers, the disparity between effective and assigned rates in crown branches was low across the range of smoothing values (Fig 4). The conservation of the effective rates in the stem and crown branches of *Isoetes* across a range of smoothing parameters indicates that the average rates predicted across the entire nuclear tree are a relatively good fit to both stem and crown branches of *Isoetes* (Fig 4). This suggests that stem and crown branches of *Isoetes* have similar rates, which is consistent with their highly similar length ratios between the chloroplast and nuclear trees (Fig 2).

Dating individual nuclear genes in r8s resulted in a wide range of optimum smoothing values (S1 Fig). Low smoothing values frequently resulted in gradient check failures, indicating a single optimum solution is not reached (S1 Fig). For genes reaching a single optimum, the median estimated crown date for *Isoetes* was 46.4 Ma with 95% of estimates between 16.1 and 85.8 Ma and 50% of results between 31.9 and 58.3 Ma (Table 1). Overall, the estimated dates form a unimodal distribution (S2 Fig). While low values of the smoothing parameter increased

Table 1. Estimates of *Isoetes* crown date.

Analysis	*Isoetes* crown age—BEAST (95% CI)	*Isoetes* crown age—r8s (95% CI)
Chloroplast concatenated markers	23.2 (6.4–46.8)	24.2 (22.8–25.9)
Nuclear concatenated markers	-	58.9 (57.3–60.1)
Nuclear individual markers	47.6 (24.1–90.8)	46.4 (16.1–85.8)
Nuclear concatenated subset	54.5 (27.9–85.2)	62.9 (54.0–62.9)

https://doi.org/10.1371/journal.pone.0227525.t001

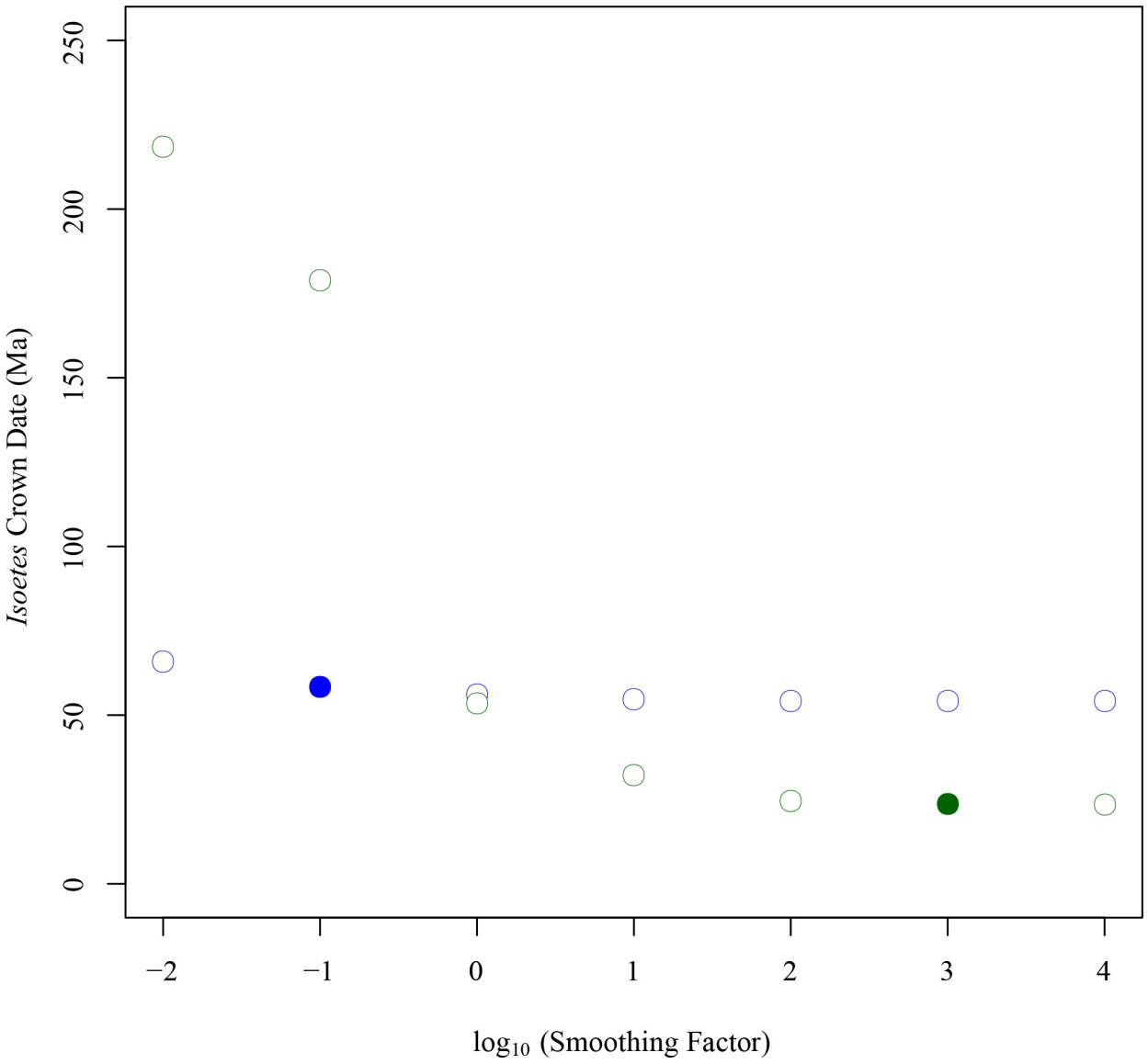


Fig 3. Effect of different smoothing factors on *Isoetes* crown date estimation in r8s. Estimated crown dates for *Isoetes* produced by r8s for concatenated chloroplast (green) and nuclear (blue) datasets are shown for a range of smoothing factors. The best fitting smoothing factor, as identified by cross validation, is highlighted for each dataset by a filled circle.

https://doi.org/10.1371/journal.pone.0227525.g003

the age estimates, all values above 10 yielded estimates centred around 50 Ma, similar to those based on the optimum smoothing values (S2 Fig). As with the chloroplast datasets, increasing smoothing values resulted in a decreased effective/assigned stem rate (S3 Fig). The disparities for the optimum smoothing values were again reduced compared to the chloroplast data (S3 Fig), indicating the globally optimum smoothing values for the individual nuclear markers fit the stem and crown branches of the *Isoetes* better than in the chloroplast dataset.

Dating individual genes using BEAST gave a median estimate of 47.6 Ma for the crown of *Isoetes*, with 95% of estimates between 24.1 and 90.1 Ma, and 50% between 39.2 and 58.6 Ma. The ages obtained for individual genes were highly correlated between r8s and BEAST (linear model, slope = 0.94, p-value < 0.001; $R^2 = 0.64$; Fig 6). Linear modelling suggested a significant

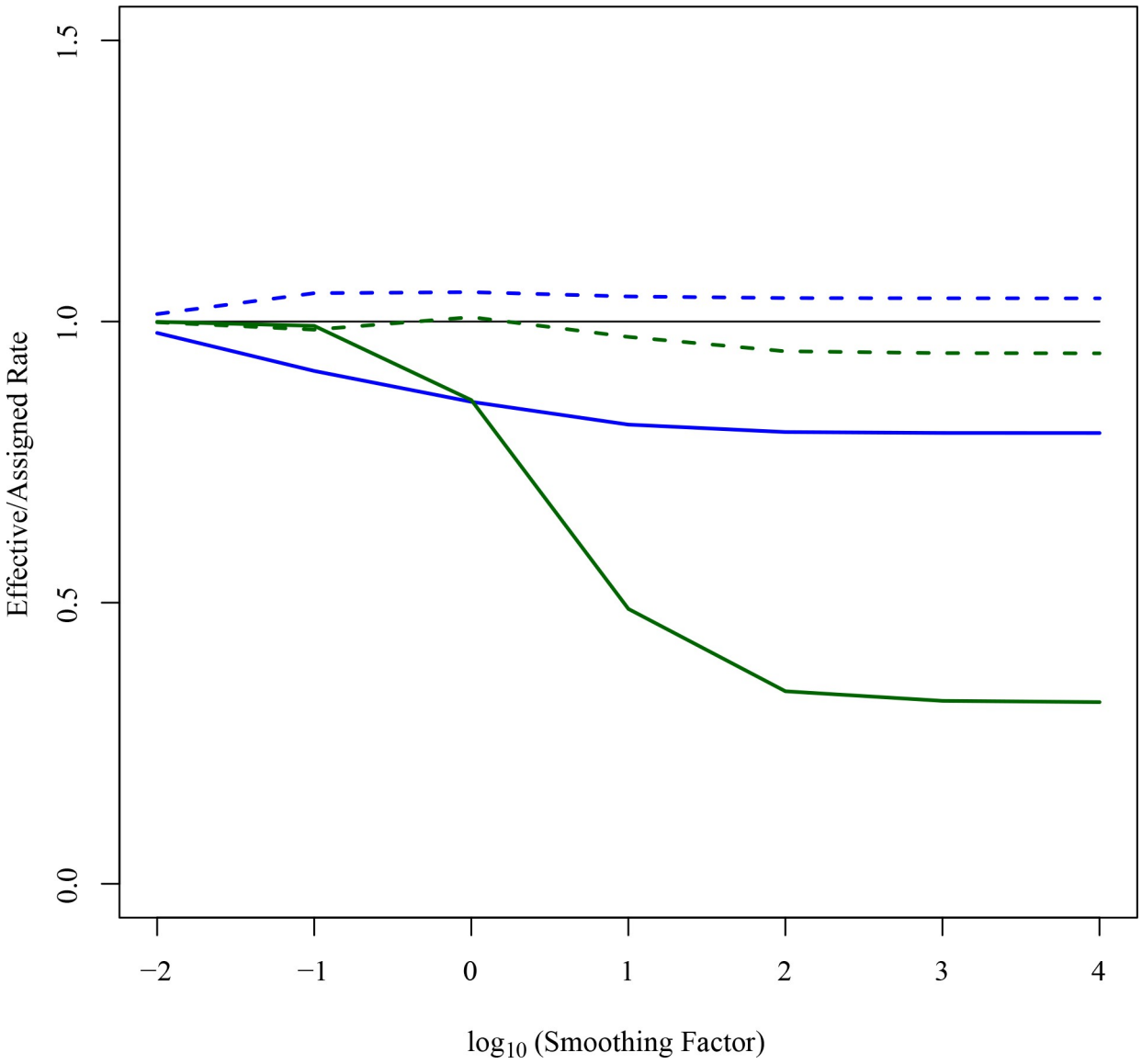


Fig 4. Rate assignment on crown and stem branches of *Isoetes* in r8s. The ratio of effective vs. assigned rates is shown for different smoothing factors in r8s for the stem branch of *Isoetes* (solid lines) and for the *Isoetes* crown branches (average; dashed lines), for the concatenated chloroplast (green) and nuclear (blue) datasets. Solid black line represents effective/assigned rate ratio of 1, for reference.

https://doi.org/10.1371/journal.pone.0227525.g004

but small effect of the percent completeness of the alignments on the estimate for the crown age of *Isoetes*, with a larger effect from the average completeness of *Isoetes* sequences (S5 Table). However, the adjusted R^2 for this latter effect was 0.059 for values obtained with r8s and 0.042 for those obtained with BEAST, indicating that the completeness of the alignment has relatively little impact on the estimated dates. BEAST dating of 55,743bp subsamples of the concatenated dataset gave a mean crown date for *Isoetes* as 54.5 Ma (mean 95% HPD 27.9–85.2Ma; Table 1; Fig 7). The eight individual subsamples gave very similar estimates of the mean crown date, with a standard deviation of 2.8Ma between the different subsamples. r8s gave a slightly older estimate, 62.9 Ma (mean 95% HPD 54.0–62.9; Table 1) with a standard deviation of 2.9Ma between the mean estimates of the different subsamples.

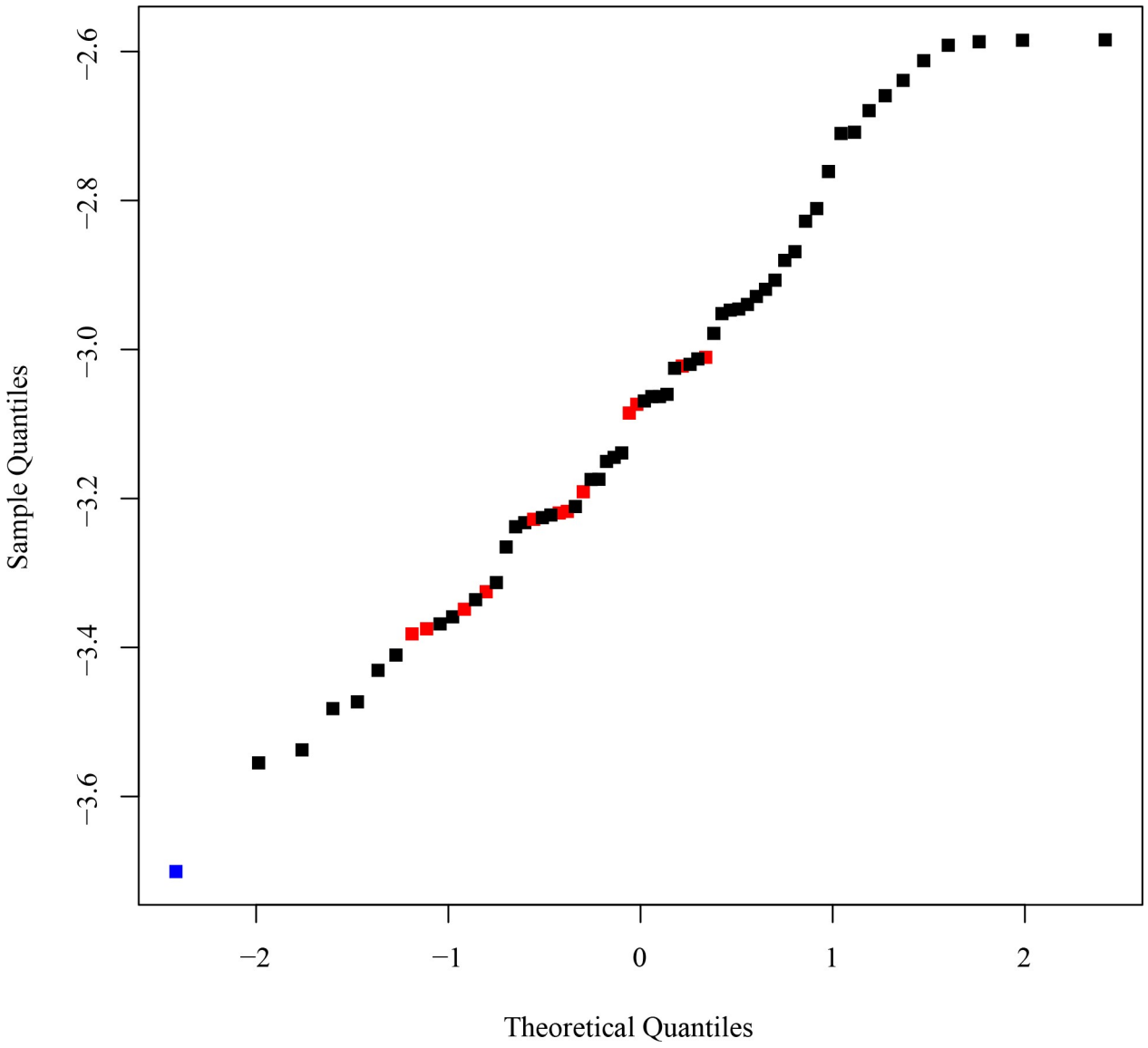


Fig 5. Quantile-quantile plot of BEAST rates for concatenated chloroplast markers. The quantile-quantile plot of $\log_{10}$ transformed branch rates is shown for the concatenated chloroplast dataset in BEAST. The values for the *Isoetes* stem branch (blue) and crown branches (red) are highlighted.

https://doi.org/10.1371/journal.pone.0227525.g005

Discussion

Nuclear analysis supports a recent origin of extant *Isoetes*

In this study, we used phylogenomics to estimate the age of *Isoetes*, a group of lycopods often interpreted as "living fossils". Using molecular dating with calibration points on deep branches of the land plant phylogeny, we found very different dates for the crown of *Isoetes* using the chloroplast and nuclear datasets, at 23–24 Ma (Oligocene) and 45–60 Ma (Paleocene and Eocene), respectively (Table 1). These differences are unlikely to be caused by the dating methods employed, since BEAST and r8s produced almost identical dates (Table 1; Fig 6), despite the very distinct ways in which these two programs deal with rate variation among branches. Subsets of 55,743bp (approximately the same size as the chloroplast alignment) of the concatenated nuclear alignment gave dates consistent with the other nuclear datasets,

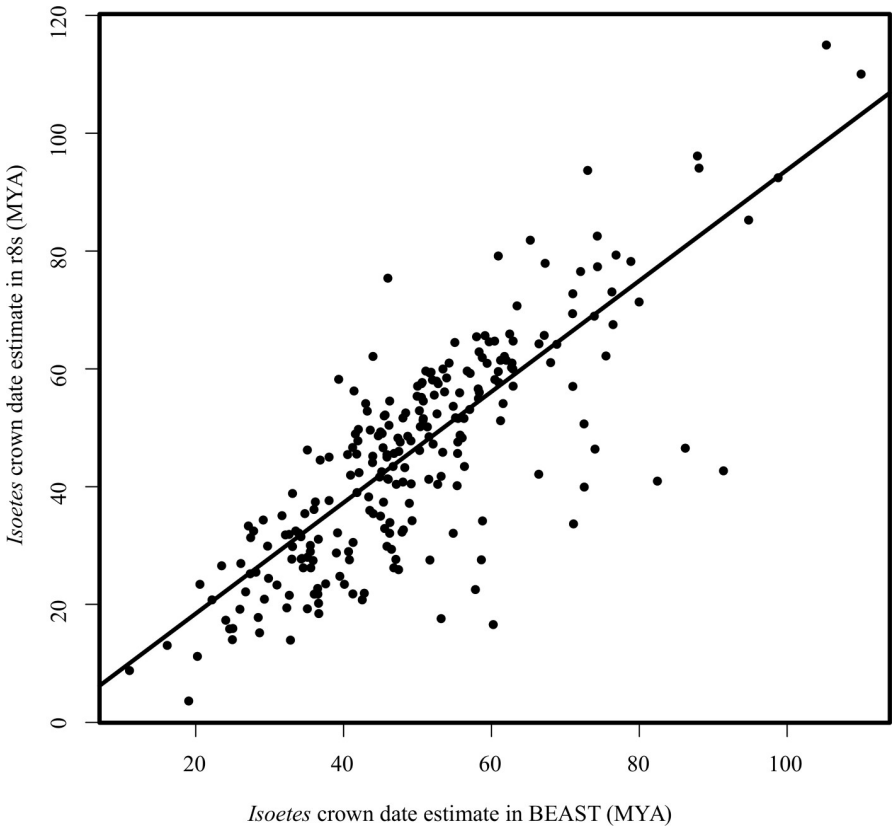


Fig 6. r8s versus BEAST *Isoetes* crown estimates for individual nuclear genes. The scatterplot shows the estimates of the *Isoetes* crown date in r8s and BEAST for each individual nuclear gene. Line represents output of linear model using lm() function in R v3.5.2.

https://doi.org/10.1371/journal.pone.0227525.g006

indicating alignment size was not the cause of this disparity either (Table 1). Instead, the incompatibilities between estimates based on nuclear and chloroplast datasets probably arise from differences in rate variation among branches. Branch lengths varied widely between *Selaginella* and *Isoetes* chloroplast markers, and throughout the chloroplast tree (Fig 2A), as previously reported [18, 69–71]. These high levels of variability make low levels of smoothing in r8s relatively poor fits to the data, as rates are poorly correlated between nearby branches on the tree, resulting in poor cross-validation scores. This results in a high smoothing value being identified as optimum for the chloroplast tree, effectively forcing a uniform rate on the tree that is determined by the average root-to-tip branch length. That, in turn, results in a high rate fitted to the *Isoetes* branch that is a poor match to its relatively short branch lengths overall (Fig 2A, Fig 4). The overall model likelihood in r8s is calculated as the sum of the log likelihood of each branch [142], meaning a date for the crown node will be assigned primarily to optimise the fit for the numerous crown branches as opposed to the single stem branch. For higher rates as smoothing values increase, younger *Isoetes* crown dates are therefore assigned. This results in a good fit for the short crown branches, but an increasingly worse fit for the stem branch (Fig 4), where the high rate and long temporal duration predict a long branch, which is not observed in the data. Similarly, in BEAST, the lognormal prior distribution results in a relatively low rate assignment on the stem branch compared to the crown branches, which leads to a better fit to the lognormal distribution across branches than if all crown branches had a low rate (Fig 5). We conclude that the high rate variability hampers accurate molecular dating

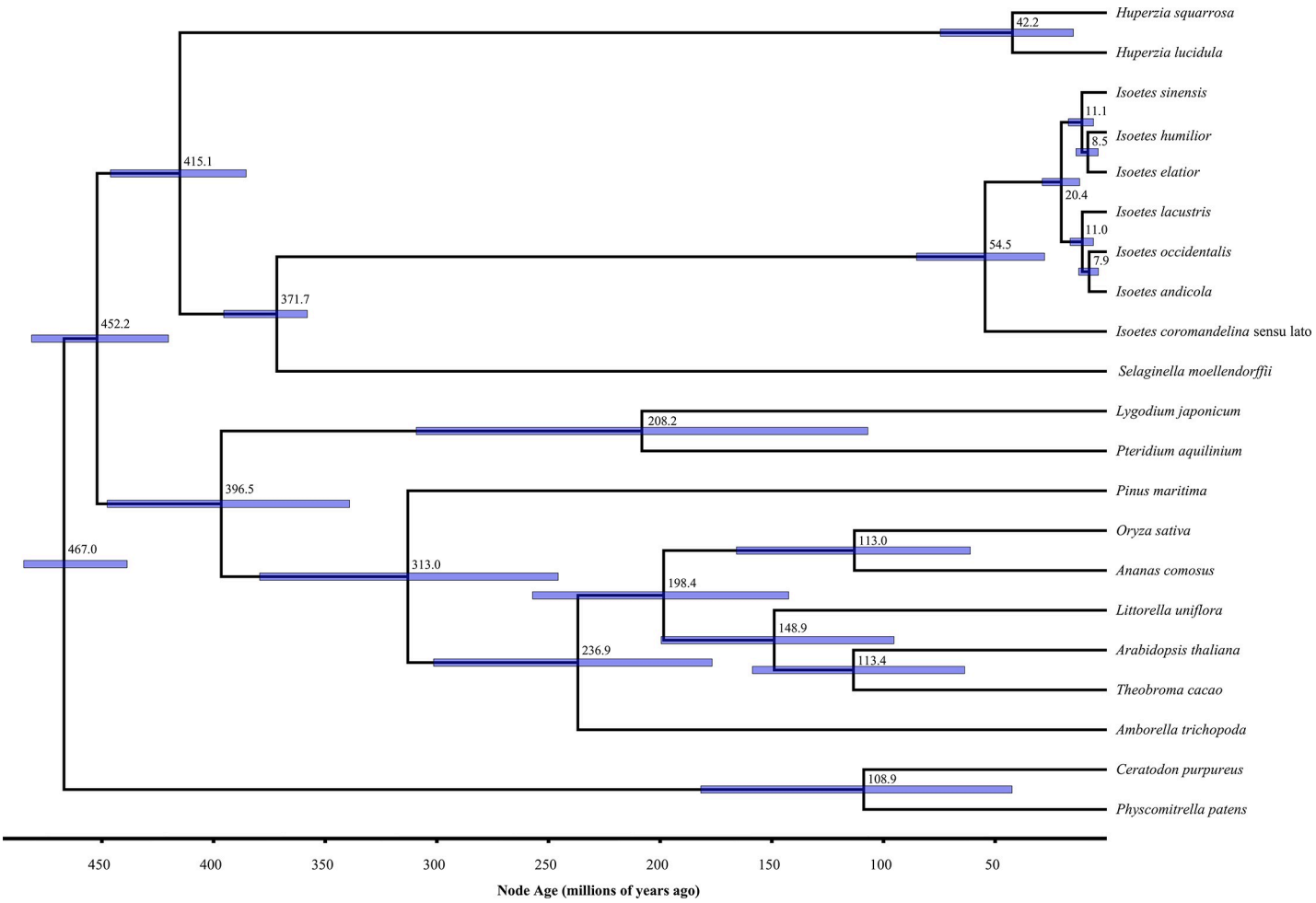


Fig 7. Maximum clade credibility tree for combined nuclear subsamples in BEAST. Node labels represent ages (Ma), blue bars represent 95% mean HPD intervals.

https://doi.org/10.1371/journal.pone.0227525.g007

using the chloroplast data. By contrast, the individual and concatenated nuclear datasets have a small disparity between estimated and effective rates, and the crown age estimate for *Isoetes* is consistent across genes (Figs 4 and S2), both in the concatenated versus individual datasets (Table 1) and between BEAST and r8s (Fig 6). The more consistent rates make the nuclear dataset more appropriate for estimating divergence times.

We conclude, based on our nuclear genome-wide analyses, that the diversity of extant *Isoetes* most likely originated during the Paleogene, between 45 and 60 Ma (Paleocene-Eocene; Table 1), although an origin from the Late Cretaceous to the Micoene is within the 95% confidence interval (16–86Ma; Table 1). This conclusion sharply contrasts with previous estimates of the crown group *Isoetes* of 147–251 Ma (Triassic to Jurassic) [18, 57, 58]. Kim and Choi [57] used Triassic *I. beestonii* to provide a narrow lognormal prior with an offset of 245.5, a mean of 1.5 and standard deviation of 0.5Ma for the age of crown *Isoetes*, resulting in an estimate of the crown age at 251Ma. Pereira et al. [58] used Jurassic *I. rolandii* to provide a minimum age for the crown node of 145 Ma, resulting in an estimate of the crown age at 147Ma (145–154 95% CI). In both these studies, fossils are used to *a priori* strongly constrain the crown node of *Isoetes* to old ages. However, these fossils do not provide evidence that the split between "clade A" and the rest of *Isoetes* had occurred, as no synapomorphies are known from the extant

members of these clades that could be preserved in fossils. The differences between these previous studies and our own emphasise the impact of fossil calibrations on date estimates. The study of Lars00E9n and Rydin (2015) [18] used fossil calibrations consistent with those used in the present study, but nevertheless estimated a crown age of 147Ma [96–215 95% CI]. Their study was based on three markers, with only *rbcL* aligning with sequences outside of the genus. The markers that do not align outside *Isoete*s should not affect the crown node age estimate as they do not inform the ratio of crown to stem substitutions. However, including non-coding markers such as nrITS with much higher levels of substitution than *rbcL* may result in a young *Isoetes* crown node giving a poor model fit for a single partition. Indeed, maximum likelihood phylogenies of the full dataset have *Isoetes* crown branch lengths that are five times longer when compared to the same taxa only with *rbcL* (S6 Table). To investigate the effects of this imbalance, we reanalysed the dataset using BEAST with the same parameters as Larsén and Rydin (2015) using only *Isoetes* species for which all three markers were available, finding a similar crown age of 145.8 Ma [88–208.7 95% CI]. Removal of the non-coding markers available solely for *Isoetes* species results in a crown estimate for *Isoetes* of 40.5 [22.6–61.6 95% CI] Ma, comparable with the results of the present study (S7 Table; BEAST files used for reanalysis available as File S6). We conclude that calibration points and molecular data both strongly impact age estimates in the case of *Isoetes*.

The number of taxa sampled in this study is however lower than in these previous studies. Reduced taxon sampling has been shown to have an impact in some [61, 62], but not all [63, 64], cases, with high levels of rate heterogeneity likely requiring increased taxon sampling [62]. The relatively low levels of rate heterogeneity in *Isoetes* (Fig 1) indicate this is unlikely to affect our age estimates, and reduction of the taxon sampling in the comprehensive Larsén and Rydin [18] dataset by 87% only resulted in a 7% change in the estimated age of the *Isoetes* crown date (see Materials and Methods). Reanalysis of the entire Larsén and Rydin [18] dataset only with markers alignable outside *Isoetes* resulted in a similar age estimate to the present study, despite the significant differences in taxon sampling (S7 Table). These considerations suggest that rather than taxon sampling, the distribution of nucleotide data among groups explain the differences between our study and that of Larsén and Rydin [18]. Therefore, while using appropriate fossil calibrations is always critical, the choice of molecular data can also have a large impact on estimated dates.

Despite our improved molecular dataset and careful assignment of fossils, the long gap between the *Isoetes* crown node and the nearest available calibration points presents a challenge in appropriate rate assignment for any node-based dating approach. Total-evidence based approaches [150] may be able to leverage the rich Isoetalean lycopsid fossil record [42, 48] to inform estimates of the rate along this branch. Nevertheless, groups such as *Isoetes* represent a particular challenge for molecular dating, necessitating careful treatment of fossil and molecular data, and the modelling approaches that use these datasets to produce age estimates. Our approach generated nucleotide data that are homogeneously distributed among taxonomic groups, and the fossil evidence is used cautiously, even though this results in a great distance between the calibrated nodes and our node of interest, the crown node of *Isoetes*. These considerations allow disentangling the significant effects of methodological variation, rates of molecular evolution, and treatment of fossils on the molecular dating of a group of "living fossils".

Despite morphological stasis, *Isoetes* recently expanded

The relatively young age of the Isoetes crown node indicates that despite morphologically similar forms appearing in the Triassic [46–48], all modern *Isoetes* are descended from a single

lineage in the early Cenozoic. This indicates that the fossil *Isoetites* from the Jurassic, and morphologically similar plants from earlier epochs, are likely stem relatives of extant *Isoetes*. The results contrast with the expectation for "living fossil" taxa, that extant species members are the last members of once diverse lineages, diverging long in the past [19, 20]. This is consistent with a number of studies in some "living fossil" plant groups such as cycads [13], bryophytes [31] and *Ginkgo* [26] showing relatively recent origins of extant species of these groups, despite long periods of morphological stasis. It is important to note, however, that other groups fit with the traditional expectations for"living fossils" [17, 27, 28]. When compared to other spore-producing plants, the pattern identified in this study is similar to the high levels of diversification seen in ferns and Lycopodiaceae since the Cretaceous [132, 151, 152], but contrasts with steady patterns of diversification over time in *Selaginella* [153]. Within both "living fossil" taxa and spore-producing plants, there appears to be a variety of patterns of lineage-diversification through time.

The global distribution of extant *Isoetes* indicates that this lineage was able to successfully colonise the globe in a relatively short amount of time. This contradicts the conclusions of previous studies that, based on older estimates for the age of the *Isoetes* crown node, explain current distributions by vicariance due to continental drift [18, 57, 58]. Based on our age estimates, geographic disparities within several subclades of *Isoetes*–such as Larsen and Rydin's [18] Clade B containing Mediterranean, North American and Indian species–indicate that long distance dispersal events have been relatively common Cenozoic *Isoetes*. It is important to note that numerous geographic disparities would remain with older age estimates. For example, the closely related Indian and Australian clades in Larsen and Rydin's Clade E diverged less than 15 Ma despite the separation of these continents during the Jurassic [154]. Further studies of *Isoetes* dispersal rates and mechanisms, which are poorly understood, are required [18, 155–157]. It should however be noted that many relationships within the Isoetes are poorly supported, and rely on a small number of genetic markers and taxon samples–and indeed many additional cryptic species within current taxa may exist [158]. Further studies will be required to fully explain extant distributions of *Isoetes* species.

The rapid global spread of extant *Isoetes* strongly contrasts with the expectation that the distribution of "living fossil" taxa are the remnants of potentially larger ancestral ranges [17, 18]. Our results show that despite having undergone little morphological change for hundreds of millions of years, rather than being the declining remnants of a bygone era, the modern *Isoetes* species instead represent recent arrivals onto the world stage.

Conclusions

Using molecular dating based on genome-wide datasets and a careful evaluation of the fossil record, we estimated the origins of extant species diversity in *Isoetes*, showing that this group of plants probably diversified in the last 45–60 million years. These results suggest that *Isoetes*-like fossils dating back to the Triassic are stem relatives of extant *Isoetes* species, and that extant *Isoetes* distribution cannot be explained by vicariance from the breakup of Gondwana. Despite their morphological conservatism over hundreds of millions of years, extant *Isoetes* diversified and spread around the world in the relatively recent past, This indicates the morphological stasis of "living fossil" taxa does not preclude lineages of these taxa from diversifying and spreading all around the world.

Supporting information

S1 Fig. Optimum smoothing values for nuclear genes in r8s. Histogram of optimum smoothing values in r8s identified by cross validation for individual nuclear genes. Proportion

of genes for each smoothing value that fail gradient checks are highlighted in red.
(PDF)

S2 Fig. *Isoetes* crown dates for individual nuclear genes for different smoothing values in r8s. Histograms showing estimated *Isoetes* crown group dates for individual nuclear genes in r8s that pass gradient checks for a range of assigned smoothing values, and the histogram of estimates where each gene is assigned its optimum smoothing value based on cross validation (final panel).
(PDF)

S3 Fig. Effective/assigned rate ratios for individual nuclear genes in r8s. Histograms of the ratio of effective vs. assigned branch rates for the stem (red) and average value for crown (blue) branches of *Isoetes* for individual nuclear genes in r8s that passed gradient checks for a range of assigned smoothing values, and the histogram of estimates where each gene is assigned its optimum smoothing value based on cross validation (final panel). Median values are displayed in the top righthand corner of each panel.
(PDF)

S1 Table. Kew herbarium DNA specimens. Published with the permission of the Board of Trustees of the Royal Botanic Gardens, Kew.
(DOCX)

S2 Table. Chloroplast data sources.
(DOCX)

S3 Table. Transcriptome data sources. See main text for references.
(DOCX)

S4 Table. Fossil constraints used for molecular dating.
(DOCX)

S5 Table. Effects of individual nuclear gene properties on estimated dates. Linear models in R (using the lm function) are used to identify the relationship between a number of alignment properties for individual nuclear genes and the resultant predicted dates in r8s and BEAST. Significant p-values (<0.05) are highlighted in bold.
(DOCX)

S6 Table. Branch lengths of full vs reduced Larsen and Rydin (2015) dataset. Branch lengths from maximum likelihood phylogenies generated using RaxML (GTR+G+I model) for (i) all three markers used in the study, ii) the *atpB-rbcL* intergenic spacer removed, (iii) *nrITS* removed (iii) or iv) both the *atpB-rbcL* spacer and *nrITS* removed.
(DOCX)

S7 Table. Reanalysis of dataset of Larsén and Rydin (2015). The alignment from Larsén and Rydin (2015) was re-analysed using the same constraints and BEAST settings as the previous paper, with at least 3 independent runs reaching ESS > 100. The dataset contains *rbcL* sequences for *Isoetes* species and other Embryophyte groups, and additional highly variable sequences for *nrITS* and the atpB-rbcL intergenic spacer for *Isoetes* only. *Isoetes* species lacking an rbcL sequence were excluded from the analysis. The entire dataset (i) gave similar estimates of the *Isoetes* crown age to Larsén and Rydin, 2015, but removal of the *atpB-rbcL* intergenic spacer (ii) reduced ages for the Isoetes crown, and removal of either *nrITS* (iii) or both *Isoetes*-specific markers (iv) resulted in ages consistent with the present study.
(DOCX)

S1 File. Individual_nuclear_alignments. Folder containing fasta files for each of the individual nuclear alignments.
(GZ)

S2 File. Combined_nuclear_alignment. Folder containing fasta file for the combined nuclear alignment.
(GZ)

S3 File. Chloroplast_alignment. Folder containing fasta file for the chloroplast alignment.
(GZ)

S4 File. Chloroplast_phylogram. Folder containing nexus file for chloroplast phylogram.
(GZ)

S5 File. Nuclear_concatenated_phylogram. Folder containing nexus file for concatenated nuclear phylogram.
(GZ)

S6 File. LarsenRydin_reanalysis_BEAST. Folder containing BEAST files used to reanalyse the Larsén and Rydin (2015) dataset.
(GZ)

S1 Data.
(TXT)

S2 Data.
(FA)

Acknowledgments

We thank Hélène Holota for help with the sequencing, Luke T. Dunning, Jill K. Olofsson, Jose J. Moreno-Villano and Matheus E. Bianconi for advice in DNA sequencing and computational analyses, Charles Wellman for advice on phylogenetic assignment of cryptospores, Hannah Sewell for assistance in live plant collection and Dan Brunton for advice regarding the putative *I. nuttallii* specimen and comments on the manuscript. Catarina Rydin kindly provided the alignment from Larsén and Rydin (2015). PAC is supported by a Royal Society University Research Fellowships (number URF120119).

Author Contributions

Conceptualization: Daniel Wood, David J. Beerling, Colin P. Osborne, Pascal-Antoine Christin.

Data curation: Daniel Wood.

Formal analysis: Daniel Wood.

Funding acquisition: Pascal-Antoine Christin.

Investigation: Daniel Wood, Guillaume Besnard, Pascal-Antoine Christin.

Methodology: Daniel Wood, Pascal-Antoine Christin.

Project administration: Daniel Wood, Pascal-Antoine Christin.

Resources: Guillaume Besnard, David J. Beerling, Colin P. Osborne, Pascal-Antoine Christin.

Supervision: David J. Beerling, Colin P. Osborne, Pascal-Antoine Christin.

Visualization: Daniel Wood.

Writing – original draft: Daniel Wood, Pascal-Antoine Christin.

Writing – review & editing: Daniel Wood, Guillaume Besnard, David J. Beerling, Colin P. Osborne, Pascal-Antoine Christin.

References

1. Bromham L, Penny D. The modern molecular clock. Nat. Rev. Genet. 2003; 4(3): 216–24. https://doi.org/10.1038/nrg1020 PMID: 12610526
2. Hipsley CA, Müller J. Beyond fossil calibrations: realities of molecular clock practices in evolutionary biology. Front. Genet. 2014; 5: 138. https://doi.org/10.3389/fgene.2014.00138 PMID: 24904638
3. Bromham L, Duchêne S, Hua X, Ritchie AM, Duchêne DA, Ho SYW. Bayesian molecular dating: opening up the black box. Biol. Rev. 2018; 93(2): 1165–91. https://doi.org/10.1111/brv.12390 PMID: 29243391
4. Sauquet H. A practical guide to molecular dating. Comptes Rendus Palevol. 2013; 12(6):355–67.
5. Renne PR, Deino AL, Hilgen FJ, Kuiper KF, Mark DF, Mitchell WS et al. Time scales of critical events around the Cretaceous-Paleogene boundary. Science 2013; 339(6120): 684–7. https://doi.org/10.1126/science.1230492 PMID: 23393261
6. Bronstein JL, Alarcón R, Geber M. The evolution of plant–insect mutualisms. New Phytol. 2006; 172 (3): 412–28. https://doi.org/10.1111/j.1469-8137.2006.01864.x PMID: 17083673
7. Patterson C. Significance of fossils in determining evolutionary relationships. Annu. Rev. Ecol. Syst. 1981; 12(1): 195–223.
8. Magallón S, Gómez-Acevedo S, Sánchez-Reyes LL, Hernández-Hernández T. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. New Phytol. 2015; 207(2): 437–53. https://doi.org/10.1111/nph.13264 PMID: 25615647
9. Poinar GO. Chapter Two—The geological record of parasitic nematode evolution. In: De Baets K, Littlewood DTJ, editors. Advances in Parasitology. Academic Press, Cambridge, Massachusetts; 2015. pp. 53–92.
10. Smith JL. A living fish of Mesozoic type. Nature 1939; 143(3620): 455–6.
11. Hilton EJ, Grande L. Review of the fossil record of sturgeons, family Acipenseridae (Actinopterygii: Acipenseriformes), from North America. J. Paleontol. 2006; 80(4): 672–83.
12. Rowe T, Rich TH, Vickers-Rich P, Springer M, Woodburne MO. The oldest platypus and its bearing on divergence timing of the platypus and echidna clades. Proc. Natl. Acad. Sci. 2008; 105(4): 1238–42. https://doi.org/10.1073/pnas.0706385105 PMID: 18216270
13. Nagalingum NS, Marshall CR, Quental TB, Rai HS, Little DP, Mathews S. Recent synchronous radiation of a living fossil. Science 2011; 334(6057): 796–9. https://doi.org/10.1126/science.1209926 PMID: 22021670
14. Kemp A, Cavin L, Guinot G. Evolutionary history of lungfishes with a new phylogeny of post-Devonian genera. Palaeogeogr. Palaeoclimatol. Palaeoecol. 2017; 471: 209–19.
15. Darwin C. On the origin of the species by means of natural selection: or, the preservation of favoured races in the struggle for life. London: John Murray; 1859.
16. Lidgard S, Hopkins M. Stasis. Oxford bibliographies on evolutionary biology. New York: Oxford University Press; 2015.
17. Mao K, Milne RI, Zhang L, Peng Y, Liu J, Thomas P et al. Distribution of living Cupressaceae reflects the breakup of Pangaea. Proc. Natl. Acad. Sci. USA 2012; 109(20): 7793–8. https://doi.org/10.1073/pnas.1114319109 PMID: 22550176
18. Larsén E, Rydin C. Disentangling the phylogeny of *Isoetes* (Isoetales), using nuclear and plastid data. Int. J. Plant Sci. 2015; 177(2): 157–74.
19. Werth AJ, Shear WA. The evolutionary truth about living fossils. Am. Sci. 2014; 102(6): 434.
20. Cavin L, Guinot G. Coelacanths as "almost living fossils". Front. Ecol. Evol. 2014; 2: 49.
21. Satler JD, Carstens BC, Hedin M. Multilocus species delimitation in a complex of morphologically conserved trapdoor spiders (Mygalomorphae, Antrodiaetidae, *Aliatypus*). Syst. Biol. 2013; 62(6): 805–23. https://doi.org/10.1093/sysbio/syt041 PMID: 23771888
22. Liu Z, Chen G, Zhu T, Zeng Z, Lyu Z, Wang J et al. Prevalence of cryptic species in morphologically uniform taxa–Fast speciation and evolutionary radiation in Asian frogs. Mol. Phylogenet. Evol. 2018; 127: 723–31. https://doi.org/10.1016/j.ympev.2018.06.020 PMID: 29920336

23. Arakaki M, Christin P-A, Nyffeler R, Lendel A, Eggli U, Ogburn RM et al. Contemporaneous and recent radiations of the world's major succulent plant lineages. Proc. Natl. Acad. Sci. U.S.A. 2011; 108(20): 8379–84. https://doi.org/10.1073/pnas.1100628108 PMID: 21536881

24. Christin P-A, Spriggs E, Osborne CP, Strömberg CAE, Salamin N, Edwards EJ. Molecular dating, evolutionary rates, and the age of the grasses. Syst. Biol. 2014; 63(2): 153–65. https://doi.org/10.1093/sysbio/syt072 PMID: 24287097

25. dos Reis M, Thawornwattana Y, Angelis K, Telford MJ, Donoghue PC, Yang Z. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. Curr. Biol. 2015; 25(22): 2939–50. https://doi.org/10.1016/j.cub.2015.09.066 PMID: 26603774

26. Hohmann N, Wolf EM, Rigault P, Zhou W, Kiefer M, Zhao Y et al. *Ginkgo biloba*'s footprint of dynamic Pleistocene history dates back only 390,000 years ago. BMC Genomics 2018; 19(1): 299. https://doi.org/10.1186/s12864-018-4673-2 PMID: 29703145

27. Sugeha HY, Pouyaud L, Hocdé R, Hismayasari IB, Gunaisah E, Widiarto SB, et al. A thirteen-million-year divergence between two lineages of Indonesian coelacanths. Sci, Rep. 2020; 10(1): 1–9.

28. Obst M, Faurby S, Bussarawit S, Funch P. Molecular phylogeny of extant horseshoe crabs (Xiphosura, Limulidae) indicates Paleogene diversification of Asian species. Mol. Phylogenet, Evol. 2012; 62 (1): 21–6.

29. Obst M, Faurby S, Bussarawit S, Funch P. Molecular phylogeny of extant horseshoe crabs (Xiphosura, Limulidae) indicates Paleogene diversification of Asian species. Mol. Phylogenet. Evol. 2012; 62 (1): 21–6. https://doi.org/10.1016/j.ympev.2011.08.025 PMID: 21939777

30. Vanschoenwinkel B, Pinceel T, Vanhove MP, Denis C, Jocque M, Timms BV, et al. Toward a global phylogeny of the "living fossil" crustacean order of the Notostraca. PloS One. 2012; 7(4).

31. Laenen B, Shaw B, Schneider H, Goffinet B, Paradis E, Désamoré A et al. Extant diversity of bryophytes emerged from successive post-Mesozoic diversification bursts. Nat. Commun. 2014; 5: 5134. https://doi.org/10.1038/ncomms6134 PMID: 25346115

32. Welch JJ, Bromham L. Molecular dating when rates vary. Trends Ecol. Evol. 2005; 20(6): 320–7. https://doi.org/10.1016/j.tree.2005.02.007 PMID: 16701388

33. Rutschmann F, Eriksson T, Salim KA, Conti E, Savolainen V. Assessing calibration uncertainty in molecular dating: The assignment of fossils to alternative calibration points. Syst. Biol. 2007; 56(4): 591–608. https://doi.org/10.1080/10635150701491156 PMID: 17654364

34. Goodall-Copestake WP, Harris DJ, Hollingsworth PM. The origin of a mega-diverse genus: dating *Begonia* (Begoniaceae) using alternative datasets, calibrations and relaxed clock methods. Bot. J. Linn. Soc. 2009; 159(3):363–80.

35. Ho SY, Phillips MJ. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Syst. Biol. 2009; 58(3): 367–80. https://doi.org/10.1093/sysbio/syp035 PMID: 20525591

36. Brandley MC, Wang Y, Guo X, Montes de Oca AN, Fería-Ortíz M, Hikida T, et al. Accommodating heterogenous rates of evolution in molecular divergence dating methods: An example using intercontinental dispersal of Plestiodon (Eumeces) lizards. Syst. Biol. 2010; 60(1): 3–15. https://doi.org/10.1093/sysbio/syq045 PMID: 20952756

37. Bromham L. Six impossible things before breakfast: Assumptions, models, and belief in molecular dating. Trends Ecol. Evol. 2019; 34(5):474–86. https://doi.org/10.1016/j.tree.2019.01.017 PMID: 30904189

38. van Tuinen M, Hedges SB. The effect of external and internal fossil calibrations on the avian evolutionary timescale. J. Paleontol. 2004; 78(1): 45–50.

39. Magallón S. Using fossils to break long branches in molecular dating: A comparison of relaxed clocks applied to the origin of angiosperms. Syst. Biol. 2010; 59(4): 384–99. https://doi.org/10.1093/sysbio/syq027 PMID: 20538759

40. Sauquet H, Ho SY, Gandolfo MA, Jordan GJ, Wilf P, Cantrill DJ et al. Testing the impact of calibration on molecular divergence times using a fossil-rich group: the case of *Nothofagus* (Fagales). Syst. Biol. 2011; 61(2): 289–313. https://doi.org/10.1093/sysbio/syr116 PMID: 22201158

41. Duchêne S, Lanfear R, Ho SY. The impact of calibration and clock-model choice on molecular estimates of divergence times. Mol. Phylogenet. Evol. 2014; 78: 277–89. https://doi.org/10.1016/j.ympev.2014.05.032 PMID: 24910154

42. Pigg KB. Evolution of Isoetalean lycopsids. Ann. Mo. Bot. Gard. 1992; 79: 589–612.

43. Taylor WC, Hickey RJ. Habitat, evolution, and speciation in Isoetes. Ann. Mo. Bot. Gard. 1992; 79: 613–22.

44. Hetherington AJ, Dolan L. The evolution of lycopsid rooting structures: conservatism and disparity. New Phytol. 2017; 215(2): 538–44. https://doi.org/10.1111/nph.14324 PMID: 27901273

45. Keeley JE. *Isoetes howellii*: A submerged aquatic CAM plant? Am. J. Bot. 1981; 68(3): 420–4.

46. Skog JE, Hill CR. The Mesozoic herbaceous lycopsids. Ann. Mo. Bot. Gard. 1992; 1:648–75.

47. Retallack GJ. Earliest Triassic origin of Isoetes and quillwort evolutionary radiation. J. Paleontol. 1997; 71(3): 500–21.

48. Pigg KB. Isoetalean lycopsid evolution: From the Devonian to the present. Am. Fern J. 2001; 91(3): 99–114.

49. Ash SR, Pigg KB. A new Jurassic *Isoetites* (Isoetales) from the Wallowa terrane in Hells Canyon, Oregon and Idaho. Am. J. Bot. 1991; 78(12): 1636–42.

50. Pfeiffer NE. Monograph of the Isoetaceae. Ann. Mo. Bot. Gard. 1922; 9(2): 79–233.

51. Karrfalt E. Further observations on *Nathorstiana* (Isoetaceae). American journal of Botany. 1984; 71 (8): 1023–30.

52. Freund FD, Freyman WA, Rothfels CJ. Inferring the evolutionary reduction of corm lobation in *Isoëtes* using Bayesian model-averaged ancestral state reconstruction. Am. J. Bot. 2018; 105(2): 275–86. https://doi.org/10.1002/ajb2.1024 PMID: 29573405

53. Troia A, Pereira JB, Kim C, Taylor WC. The genus *Isoetes* (Isoetaceae): a provisional checklist of the accepted and unresolved taxa. Phytotaxa 2016; 277(2): 101–45.

54. PPG I. A community-derived classification for extant lycophytes and ferns. J. Syst. Evol. 2016; 54(6): 563–603.

55. Rydin C, Wikström N. Phylogeny of *Isoëtes* (Lycopsida): Resolving basal relationships using rbcL sequences. Taxon 2002; 51(1): 83–9.

56. Hoot SB, Taylor WC, Napier NS. Phylogeny and biogeography of *Isoëtes* (Isoëtaceae) based on nuclear and chloroplast DNA sequence data. Syst. Bot. 2006; 31(3): 449–60.

57. Kim C, Choi HK. Biogeography of North Pacific *Isoëtes* (Isoëtaceae) inferred from nuclear and chloroplast DNA sequence data. J. Plant Biol. 2016; 59(4): 386–96.

58. Pereira JBS, Labiak PH, Stützel T, Schulz C. Origin and biogeography of the ancient genus *Isoëtes* with focus on the Neotropics. Bot. J. Linn. Soc. 2017; 185(2): 253–71.

59. Banks JA. Selaginella and 400 million years of separation. Annu. Rev. Plant Biol. 2009; 60: 223–38. https://doi.org/10.1146/annurev.arplant.59.032607.092851 PMID: 19575581

60. Patil S, Rajput KS. The genus Isoetes from India: an overview. Plant Science Today. 2017; 4(4): 213–26.

61. Linder HP, Hardy CR, Rutschmann F. Taxon sampling effects in molecular clock dating: an example from the African Restionaceae. Mol. Phylogenet. Evol. 2005; 35(3): 569–82. https://doi.org/10.1016/j.ympev.2004.12.006 PMID: 15878126

62. Soares AE, Schrago CG. The influence of taxon sampling and tree shape on molecular dating: an empirical example from Mammalian mitochondrial genomes. Bioifnorm. Biol. Insights. 2012; 6: 129–43.

63. Hug LA, Roger AJ. The impact of fossils and taxon sampling on ancient molecular dating analyses. Mol. Biol. Evol. 2007; 24(8): 1889–97. https://doi.org/10.1093/molbev/msm115 PMID: 17556757

64. Xiang QY, Thomas DT, Xiang QP. Resolving and dating the phylogeny of Cornales–effects of taxon sampling, data partitions, and fossil calibrations. Mol. Phylogenet. Evol. 2011; 59(1):123–38. https://doi.org/10.1016/j.ympev.2011.01.016 PMID: 21300164

65. Schuettpelz E, Hoot SB. Inferring the root of *Isoëtes*: exploring alternatives in the absence of an acceptable outgroup. Systematic Botany. 2006 Apr 1; 31(2): 258–70.

66. Besnard G, Bianconi ME, Hackel J, Manzi S, Vorontsova MS, Christin P-A. Herbarium genomics retrace the origins of C4-specific carbonic anhydrase in Andropogoneae (Poaceae). Bot. Lett. 2018: 165(3–4); 419–433.

67. Bieker VC, Martin MD. Implications and future prospects for evolutionary analyses of DNA in historical herbarium collections. Bot. Lett. 2018: 165(3–4); 409–18.

68. Bakker FT, Lei D, Yu J, Mohammadin S, Wei Z, van de Kerke S et al. Herbarium genomics: plastome sequence assembly from a range of herbarium specimens using an Iterative Organelle Genome Assembly pipeline. Biol. J. Linn. Soc. 2016; 117(1): 33–43.

69. Bousquet J, Strauss SH, Doerksen AH, Price RA. Extensive variation in evolutionary rate of *rbcL* gene sequences among seed plants. Proc. Natl. Acad. Sci. USA 1992; 89(16): 7844–8. https://doi.org/10.1073/pnas.89.16.7844 PMID: 1502205

70. Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG. From algae to angiosperms–inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. BMC Evol. Biol. 2014; 14(1): 23.

71. Karol KG, Arumuganathan K, Boore JL, Duffy AM, Everett KD, Hall JD et al. Complete plastome sequences of *Equisetum arvense* and *Isoetes flaccida*: implications for phylogeny and plastid genome evolution of early land plant lineages. BMC Evol. Biol. 2010; 10(1): 321.

72. Wickett, et al. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. Proceedings of the National Academy of Sciences, 1–21.

73. 1KP consortium (2019). One thousand plant transcriptomes and the phylogenomics of green plants. Nature, 1–20.

74. Olofsson JK, Bianconi M, Besnard G, Dunning LT, Lundgren MR, Holota H et al. Genome biogeography reveals the intraspecific spread of adaptive mutations for a complex trait. Mol. Ecol. 2016; 25(24): 6107–23. https://doi.org/10.1111/mec.13914 PMID: 27862505

75. Olofsson JK, Cantera I, Van de Paer C, Hong-Wa C, Zedane L, Dunning LT, et al. Phylogenomics using low-depth whole genome sequencing: A case study with the olive tribe. Mol. Ecol. Resour. 2019; 001:1–16.

76. Boston HL, Adams MS. Evidence of crassulacean acid metabolism in two North American isoetids. Aquat. Bot. 1983; 15(4):381–386.

77. Lundgren MR, Besnard G, Ripley BS, Lehmann CER, Chatelet DS, Kynast RG et al. Photosynthetic innovation broadens the niche within a single species. Ecol. Lett. 2015; 18(10): 021–29.

78. Patel RK, Jain M. NGS QC Toolkit: A toolkit for quality control of next generation sequencing data. PLoS One 2012; 7(2): e30619. https://doi.org/10.1371/journal.pone.0030619 PMID: 22312429

79. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 2017; 45(4): e18. https://doi.org/10.1093/nar/gkw955 PMID: 28204566

80. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002; 30(14): 3059–66. https://doi.org/10.1093/nar/gkf436 PMID: 12136088

81. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014; 30(9): 1312–3. https://doi.org/10.1093/bioinformatics/btu033 PMID: 24451623

82. Belshaw R, Katzourakis A. BlastAlign: a program that uses blast to align problematic nucleotide sequences. Bioinformatics 2005; 21(1): 122–123. https://doi.org/10.1093/bioinformatics/bth459 PMID: 15310559

83. Goremykin V.V., Hirsch-Ernst K.I., Wolfl S. and Hellwig F.H. Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm. Mol. Biol. Evol. 2003; 20(9): 1499–1505. https://doi.org/10.1093/molbev/msg159 PMID: 12832641

84. Nashima K, Terakami S, Nishitani C, Kunihisa M, Shoda M, Takeuchi M, et al. Complete chloroplast genome sequence of pineapple (*Ananas comosus*). Tree Genet. Genomes. 2015; 11(3): 60.

85. Wickett NJ, Zhang., Hansen SK, Roper JM, Kuehl JV, Plock SA, et al B. Functional gene losses occur with minimal size reduction in the plastid genome of the parasitic liverwort *Aneura mirabilis*. Mol. Biol. Evol. 2008; 25(2): 393–401. https://doi.org/10.1093/molbev/msm267 PMID: 18056074

86. Moore MJ, Bell CD, Soltis PS and Soltis DE. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proc. Natl. Acad. Sci. U.S.A. 2007; 104(49): 19363–8. https://doi.org/10.1073/pnas.0708072104 PMID: 18048334

87. Kim HT, Chung MG and Kim KJ. Chloroplast geome evolution in early diverged Leptosporangiate ferns. Mol. Cells. 2014; 37(5): 372. https://doi.org/10.14348/molcells.2014.2296 PMID: 24823358

88. Li X, Li Q, Lin X, Hu Z and Chen S. High-throughput multiplex sequencing reveals the prospect of chloroplast genomes as a plant super-barcode.

89. Wolf PG, Karol KG, Mandoli DF, Kuehl J, Arumuganathan K, Ellis et al. The first complete chloroplast genome sequence of a lycophyte, *Huperzia lucidula* (Lycopodiaceae). Gene. 2005; 350(2): 117–28. https://doi.org/10.1016/j.gene.2005.01.018 PMID: 15788152

90. Shimada H and Sugiura M. Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes. Nucleic Acids Res. 1991; 19 (5): 983–95. https://doi.org/10.1093/nar/19.5.983 PMID: 1708498

91. 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(1): 174.

92. Goremykin VV, Hirsch-Ernst KI, Wölfl S, Hellwig FH. The chloroplast genome of *Nymphaea alba*: whole-genome analyses and the problem of identifying the most basal angiosperm. Mol. Biol. Evol. 2004; 21(7): 1445–54. https://doi.org/10.1093/molbev/msh147 PMID: 15084683

93. Sugiura C, Kobayashi Y, Aoki S, Sugita C, Sugita M. Complete chloroplast DNA sequence of the moss *Physcomitrella patens*: evidence for the loss and relocation of rpoA from the chloroplast to the nucleus. Nucleic Acids Research. 2003; 31(18): 5324–31. https://doi.org/10.1093/nar/gkg726 PMID: 12954768

94. Tsudzuki J, Ito S, Tsudzuki T, Wakasugi T, Sugiura M. A new gene encoding tRNA Pro (GGG) is present in the chloroplast genome of black pine: a compilation of 32 tRNA genes from black pine chloroplasts. Curr. Genet. 1994; 26(2): 153–8. https://doi.org/10.1007/BF00313804 PMID: 8001170

95. Zhu A, Guo W, Gupta S, Fan W, Mower JP. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. New Phytol. 2016; 209(4): 1747–56. https://doi.org/10.1111/nph.13743 PMID: 26574731

96. Der JP. Genomic perspectives on evolution in bracken fern. Dissertation. 2010. Utah State University, Department of Biology. 2010.

97. Banks JA, Nishiyama T, Hasebe M, Bowman JL, Gribskov M, de Pamphilis C et al. The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. Science. 201; 332 (6032): 960–3. https://doi.org/10.1126/science.1203810 PMID: 21551031

98. Tsuji S, Ueda K, Nishiyama T, Hasebe M, Yoshikawa S, Konagaya A et al. The chloroplast genome from a lycophyte (microphyllophyte), *Selaginella uncinata*, has a unique inversion, transpositions and many gene losses. J. Plant Res. 2007; 120(2): 281–90. https://doi.org/10.1007/s10265-006-0055-y PMID: 17297557

99. Yi DK, Kim KJ. Complete chloroplast genome sequences of important oilseed crop *Sesamum indicum* L. PloS one. 2012; 7(5): e35872. https://doi.org/10.1371/journal.pone.0035872 PMID: 22606240

100. Oliver MJ, Murdock AG, Mishler BD, Kuehl JV, Boore JL, Mandoli DF, et al. Chloroplast genome sequence of the moss *Tortula ruralis*: gene content, polymorphism, and structural arrangement relative to other green plant chloroplast genomes. BMC genomics. 2010; 11(1): 143.

101. Choi KS, Chung MG, Park S. The complete chloroplast genome sequences of three Veroniceae species (Plantaginaceae): comparative analysis and highly divergent regions. Front. Plant Sci. 2016; 7: 355. https://doi.org/10.3389/fpls.2016.00355 PMID: 27047524

102. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. Bioinformatics 2004; 20(17): 3252–5. https://doi.org/10.1093/bioinformatics/bth352 PMID: 15180927

103. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J et al. De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. Nat. Protoc. 2013; 8(8): 1494–512. https://doi.org/10.1038/nprot.2013.084 PMID: 23845962

104. Notredame C, Higgins DG, Heringa J. T-coffee: a novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. 2000; 302(1): 205–17. https://doi.org/10.1006/jmbi.2000.4042 PMID: 10964570

105. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 2014; 30(22): 3276–8. https://doi.org/10.1093/bioinformatics/btu531 PMID: 25095880

106. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat. Methods 2012; 9(4): 357. https://doi.org/10.1038/nmeth.1923 PMID: 22388286

107. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N et al. The sequence alignment/map format and SAMtools. Bioinformatics 2009; 25(16): 2078–9. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943

108. Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E. EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. Genome Res. 2009; 19(2): 327–35. https://doi.org/10.1101/gr.073585.107 PMID: 19029536

109. Der JP, Barker MS, Wickett NJ, dePamphilis CW, Wolf PG. De novo characterization of the gametophyte transcriptome in bracken fern, *Pteridium aquilinum*. BMC Genomics 2011; 12(1): 99.

110. Canales J, Bautista R, Label P, Gómez-Maldonado J, Lesur I et al. De novo assembly of maritime pine transcriptome: implications for forest breeding and biotechnology. Plant Biotechnol. J. 2014; 12(3): 286–99. https://doi.org/10.1111/pbi.12136 PMID: 24256179

111. Aya K, Kobayashi M, Tanaka J, Ohyanagi H, Suzuki T, Yano K et al. De novo transcriptome assembly of a fern, *Lygodium japonicum*, and a web resource database, Ljtrans DB. Plant Cell Physiol. 2015; 56 (1): e5. https://doi.org/10.1093/pcp/pcu184 PMID: 25480117

112. Szövényi P, Perroud PF, Symeonidi A, Stevenson S, Quatrano RS, Rensing SA et al. De novo assembly and comparative analysis of the *Ceratodon purpureus* transcriptome. Mol. Ecol. Resour. 2014; 15 (1): 203–15. https://doi.org/10.1111/1755-0998.12284 PMID: 24862584

113. Ming R, Van Buren R, Wai CM, Tang H, Schatz MC, Bowers JE et al. The pineapple genome and the evolution of CAM photosynthesis. Nat. Genet. 2015; 47(12): 1435–42. https://doi.org/10.1038/ng.3435 PMID: 26523774

114. Yang T, Liu X. Comparative transcriptome analysis of *Isoetes sinensis* under terrestrial and submerged conditions. Plant Mol. Biol. Report. 2016; 34(1): 136–45.

115. http://www.medplantrnaseq.org/, last accessed 13/11/18

116. Fisher AE, Hasenstab KM, Bell HL, Blaine E, Ingram AL, Columbus JT. Evolutionary history of chloridoid grasses estimated from 122 nuclear loci. Mol. Phylogenet. Evol. 2016; 105: 1–14. https://doi.org/10.1016/j.ympev.2016.08.011 PMID: 27554759

117. Dunning LT, Lundgren MR, Moreno-Villena JJ, Namaganda M, Edwards EJ, Nosil P et al. Introgression and repeated co-option facilitated the recurrent emergence of C4 photosynthesis among close relatives. Evolution 2017; 71(6): 1541–55. https://doi.org/10.1111/evo.13250 PMID: 28395112

118. Abadi S, Azouri D, Pupko T, Mayrose I. Model selection may not be a mandatory step for phylogeny reconstruction. Nat. Commun. 2019; 10(1): 934. https://doi.org/10.1038/s41467-019-08822-w PMID: 30804347

119. Rolle F. Kryptogamen. In: Förster W, Kenngott A, Ladenburg D, Reichenow A, Shenk D, Schömilch D et al. (eds.). Encyklopaedie der Naturwissenschaften. 1885; 12. pp. 211–277.

120. Freund FD. Characterizing quantitative variation in the glossopodia of three western North American *Isoëtes* species. Am. Fern J. 2016; 106(2): 87–115.

121. Hao S, Xue J, Wang Q, Liu Z. *Yuguangia ordinata* gen. et sp. nov., a new lycopsid from the Middle Devonian (late Givetian) of Yunnan, China, and its phylogenetic implications. Int. J. Plant Sci. 2007; 168(8): 1161–75.

122. Stein WE, Berry CM, Hernick LV, Mannolini F. Surprisingly complex community discovered in the mid-Devonian fossil forest at Gilboa. Nature 2012; 483(7387): 78–81. https://doi.org/10.1038/nature10819 PMID: 22382983

123. Berry CM, Marshall JEA. Lycopsid forests in the early Late Devonian paleoequatorial zone of Svalbard. Geology, 2015; 43(12): 1043–6.

124. Xu HH, Wang Y, Wang Q. A new homosporous, arborescent lycopsid from the Middle Devonian of Xinjiang, Northwest China. Palaeontology 2012; 55(5): 957–66.

125. Wang Y, Berry CM. A novel lycopsid from the Upper Devonian of Jiangsu, China. Palaeontology 2003; 46(6): 1297–311.

126. Cressler WL, Pfefferkorn HW. A Late Devonian isoetalean lycopsid, *Otzinachsonia beerboweri*, gen. et sp. nov., from north-central Pennsylvania, USA. Am. J. Bot. 2005; 92(7): 1131–40. https://doi.org/10.3732/ajb.92.7.1131 PMID: 21646135

127. Wellman CH. The invasion of the land by plants: when and where? New Phytol. 2010; 188(2), 306–9. https://doi.org/10.1111/j.1469-8137.2010.03471.x PMID: 20941845

128. Morris JL, Puttick MN, Clark JW, Edwards D, Kenrick P, Pressel S et al. The timescale of early land plant evolution. Proc. Natl. Acad. Sci. USA 2018; 115(1): E2274–83.

129. Rubinstein CV, Gerrienne P, de la Puente GS, Astini RA, Steemans P. Early Middle Ordovician evidence for land plants in Argentina (Eastern Gondwana). New Phytol. 2010; 188(2): 365–9. https://doi.org/10.1111/j.1469-8137.2010.03433.x PMID: 20731783

130. Wellman CH, Strother PK. The terrestrial biota prior to the origin of land plants (embryophytes): a review of the evidence. Palaeontology 2015; 58(4): 601–27.

131. Kenrick P. Palaeobotany: fishing for the first plants. Nature 2003; 425(6955): 248. https://doi.org/10.1038/425248a PMID: 13679900

132. Testo W, Sundue M. A 4000-species dataset provides new insight into the evolution of ferns. Mol. Phylogenet. Evol. 2016; 105: 200–11. https://doi.org/10.1016/j.ympev.2016.09.003 PMID: 27621129

133. Smith SA, Beaulieu JM, Donoghue MJ. An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. Proc. Natl. Acad. Sci. USA. 2010; 107(13): 5897–902. https://doi.org/10.1073/pnas.1001225107 PMID: 20304790

134. Rothfels CJ, Li FW, Sigel EM, Huiet L, Larsson A, Burge DO et al. The evolutionary history of ferns inferred from 25 low-copy nuclear genes. Am. J. Bot. 2015; 102(7): 1089–107. https://doi.org/10.3732/ajb.1500089 PMID: 26199366

135. Garratt MJ, Tims JD, Rickards RB, Chambers TC, Douglas JG. The appearance of *Baragwanathia* (Lycophytina) in the Silurian. Bot. J. Linn. Soc. 1984; 89(4): 355–8.

136. Kenrick P, Crane PR. The origin and early evolution of plants on land. Nature 1997; 389(6646): 33.

137. Magallón S, Hilu KW, Quandt D. Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. Am. J. Bot. 2013; 100(3): 556–73. https://doi.org/10.3732/ajb.1200416 PMID: 23445823

138. Edwards D, Feehan J. Records of *Cooksonia*-type sporangia from Late Wenlock Strata in Ireland. Nature. 1980; 287(5777). 41–2.

139. Edwards D, Feehan J, Smith DG. A late Wenlock flora from Co Tipperary, Ireland. Bot. J. Linn. Soc. 1983; 86(1–2), 19–36.

140. Libertín M, Kvaček J, Bek J, Žárský V, Štorch P. Sporophytes of polysporangiate land plants from the early Silurian period may have been photosynthetically autonomous. Nature plants. 2018; 4(5): 269 https://doi.org/10.1038/s41477-018-0140-y PMID: 29725100

141. Parham JF, Donoghue PC, Bell CJ, Calway TD, Head JJ, Holroyd P et al. Best practices for justifying fossil calibrations. Systematic Biology. 2011; 61(2): 346–59. https://doi.org/10.1093/sysbio/syr107 PMID: 22105867

142. Sanderson MJ. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 2003; 19(2): 301–2. https://doi.org/10.1093/bioinformatics/19.2.301 PMID: 12538260

143. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 2007; 7(1): 214.

144. Sanderson MJ. Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. Mol. Biol. Evol. 2002; 19(1): 101–9. https://doi.org/10.1093/oxfordjournals.molbev.a003974 PMID: 11752195

145. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. PLoS Biol. 2006; 4(5): e88. https://doi.org/10.1371/journal.pbio.0040088 PMID: 16683862

146. Felsenstein J. {PHYLIP} (Phylogeny Inference Package) version 3.6a3. 2002. Distributed by the author.

147. Desmet P, Brouillet L. Database of Vascular Plants of Canada (VASCAN): a community contributed taxonomic checklist of all vascular plants of Canada, Saint Pierre and Miquelon, and Greenland. PhytoKeys. 2013(25): 55. https://doi.org/10.3897/phytokeys.25.3100 PMID: 24198712

148. Britton DM, Brunton DF, Talbot SS. Isoetes in Alaska and the Aleutians. Am. Fern J. 1999: 133–41.

149. Flora of North America http://www.efloras.org/flora_page.aspx?flora_id=1, last accessed 20/03/20.

150. Ronquist F, Lartillot N, Phillips MJ. Closing the gap between rocks and clocks using total-evidence dating. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 2016; 371(1699): 20150136. https://doi.org/10.1098/rstb.2015.0136 PMID: 27325833

151. Schneider H, Schuettpelz E, Pryer KM, Cranfill R, Magallón S, Lupia R. Ferns diversified in the shadow of angiosperms. Nature. 2004; 428(6982):553–7. https://doi.org/10.1038/nature02361 PMID: 15058303

152. Testo W, Field A, Barrington D. Overcoming among-lineage rate heterogeneity to infer the divergence times and biogeography of the clubmoss family Lycopodiaceae. J Biogeogr. 2018; 45(8):1929–41.

153. Klaus KV, Schulz C, Bauer DS, Stützel T. Historical biogeography of the ancient lycophyte genus *Selaginella*: early adaptation to xeric habitats on Pangaea. Cladistics. 2017; 33(5): 469–80.

154. Olierook HK, Jourdan F, Merle RE, Timms NE, Kusznir N, Muhling JR. Bunbury Basalt: Gondwana breakup products or earliest vestiges of the Kerguelen mantle plume? Earth Planet Sci. Lett. 2016; 440: 20–32.

155. Brunton DF. Quillwort dispersal: which way is the wind blowing. Chinquapin. 2001; 9(20): 24–6.

156. Troia A. Dispersal and colonization in heterosporous lycophytes: palynological and biogeographical notes on the genus *Isoetes* in the Mediterranean region. Webbia. 2016 2; 71(2): 277–81.

157. Wood DP, Olofsson JK, McKenzie SW, Dunning LT. Contrasting phylogeographic structures between freshwater lycopods and angiosperms in the British Isles. Bot Lett. 2018; 165(3–4): 476–86.

158. Schafran PW. Molecular Systematics of *Isoëtes* (Isoëtaceae) in Eastern North America. PhD Thesis. Old Domion University. 2019. Available from: https://digitalcommons.odu.edu/biology_etds/110/ (last accessed 28/04/20)