scientific reports

Check for updates

OPEN

Gene expression data support the hypothesis that *Isoetes* rootlets are true roots and not modified leaves

Alexander J. Hetherington[1,2], David M. Emms[1], Steven Kelly[1] & Liam Dolan[1,3✉]

Rhizomorphic lycopsids are the land plant group that includes the first giant trees to grow on Earth and extant species in the genus *Isoetes*. Two mutually exclusive hypotheses account for the evolution of terminal rooting axes called rootlets among the rhizomorphic lycopsids. One hypothesis states that rootlets are true roots, like roots in other lycopsids. The other states that rootlets are modified leaves. Here we test predictions of each hypothesis by investigating gene expression in the leaves and rootlets of *Isoetes echinospora*. We assembled the de novo transcriptome of axenically cultured *I. echinospora*. Gene expression signatures of *I. echinospora* rootlets and leaves were different. Furthermore, gene expression signatures of *I. echinospora* rootlets were similar to gene expression signatures of true roots of *Selaginella moellendorffii* and *Arabidopsis thaliana*. RSL genes which positively regulate cell differentiation in roots were either exclusively or preferentially expressed in the *I. echinospora* rootlets, S. *moellendorffii* roots and *A. thaliana* roots compared to the leaves of each respective species. Taken together, gene expression data from the *de-novo* transcriptome of *I. echinospora* are consistent with the hypothesis that *Isoetes* rootlets are true roots and not modified leaves.

The first giant (> 50 m) trees to grow on Earth, the arborescent clubmosses, were tethered to the ground by rooting structures termed stigmarian systems whose homology has been debated for more than 150 years[1–9]. Stigmarian rooting systems consisted of two components, a central axis (rhizomorph) on which developed large numbers of fine axes (rootlets). There are two competing hypotheses to explain the origin of stigmarian rootlets which we designate, the lycopsid root hypothesis and the modified shoot hypothesis. The lycopsid root hypothesis posits that rootlets are homologous to roots of other lycopsids. The modified shoot hypothesis posits that rootlets are modified leaves (microphylls) and homologous to the leaves of other lycosids.

Stigmarian rootlets were interpreted as true roots by the majority of authors until the mid twentieth century[5,6,10–14]. However, a suite of fossil findings in the second half of the twentieth century, including fossil embryos, rhizomorph apices and the abscission of rootlets[3,4,15–19], led to the revival of the modified shoot hypothesis first suggested in 1872, which interpreted rootlets as modified leaves[7]. Given that all rhizomorphic lycopsids (sensu[20–23]) form a monophyletic group, and that extinct stigmarian rootlets were interpreted as modified leaves this suggested that the rootlets of all rhizomorphic lycopsids were modified leaves, including the rootlets of extant *Isoetes*[3]. The interpretation that the rootlets of extant *Isoetes* species were modified leaves was strikingly at odds with all previous descriptions of *Isoetes* rootlets that had always been interpreted as roots similar to the roots of other extant lycopsids[11,24–32].

New evidence that is inconsistent with the modified shoot hypothesis has been reported since the seminal paper by Rothwell and Erwin[3]. First, the modified shoot hypothesis posits that the ancestral embryo condition in the rhizomorphic lycopsids lacked an embryonic root, but instead developed a single shoot axis that divided to give a typical shoot and modified rooting shoot axis that developed modified leaves (rootlets). However, embryo development in the early diverging rhizomorphic lycopsid, *Oxroadia* developed an embryonic root[20]. Therefore, the embryo of *Oxroadia* does not support the hypothesis that a branching event in the embryo produced a rooting shoot axis (rhizomorph) that developed root-like leaves (rootlets). Second, while the leaves of

[1]Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK. [2]Present address: Institute of Molecular Plant Sciences, School of Biological Sciences, University of Edinburgh, Max Born Crescent, Edinburgh EH9 3BF, UK. [3]Present address: Gregor Mendel Institute of Molecular Plant Biology GmbH Dr. Bohr-Gasse 3, 1030 Vienna, Austria. ✉email: liam.dolan@gmi.oeaw.ac.at
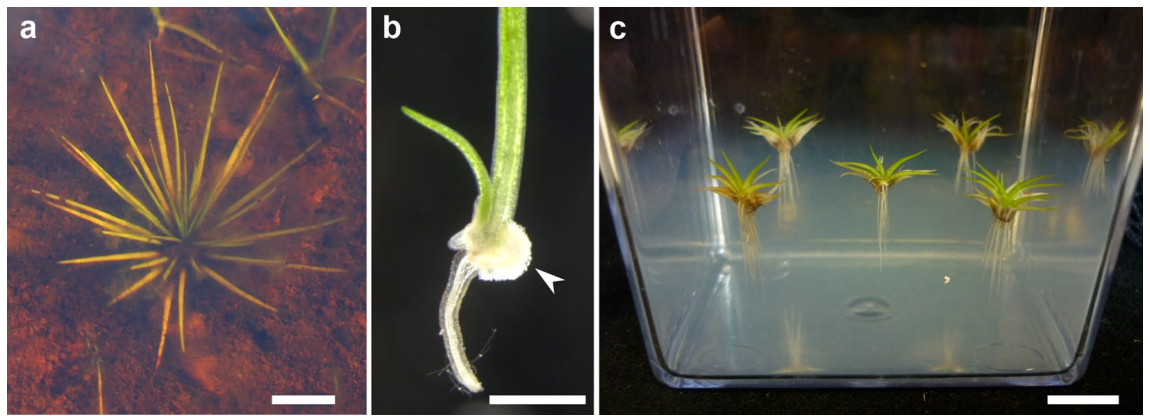
Figure 1. Growth of *I. echinospora* in axenic culture for RNA extraction. (**a**) *I. echinospora* growing submerged in its natural habitat, North West Sutherland (Scotland, UK). (**b**) Developing sporophyte emerging from the megaspore (highlighted with arrowhead) at the two leaf-, two rootlet-stage (**C**) *I. echinospora* sporophytes growing submerged on transparent solid media. Scale bars: (**a**) approximately 2 cm; (**b**) 1 mm; (**c**) 1 cm.
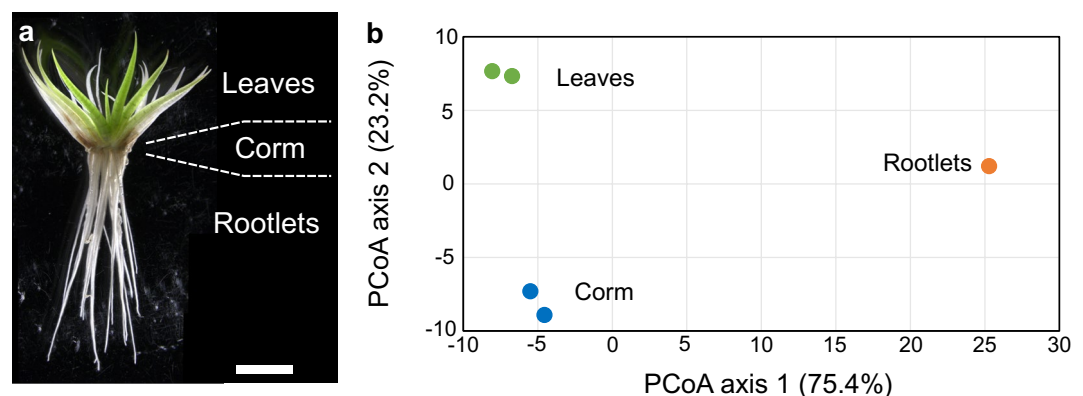
Figure 2. Gene expression profiles of leaves and rootlets are different in *I. echinospora*. (**a**) *I. echinospora* sporophyte at the stage when RNA was extracted from rootlets, corms and leaves. Scale bar 5 mm. (**b**) Comparison of gene expression profiles by principle coordinate analysis (PCoA) in the transcriptome of *I. echinospora*. Two technical replicates of gene expression profiles of the corm and leaves and single replicate of rootlets of *I. echinospora*. Two leaf replicates, green. Two corm replicates, blue. Single root replicate, orange. Values on PCoA axes are shown in thousands. Values in brackets on each axis describe the percentage of total variance accounted for by each axis.

all plants species develop exogenously[33], in a process that includes the outer-most layers of the shoot, roots of extant *Isoetes* originate endogenously[34]. Therefore, the endogenous development of rootlets is inconsistent with their interpretation as modified leaves[34]. Third, the discovery of the development of root hairs on rootlets of extinct rhizomorphic lycopsids that are identical to the root hairs that develop on extant lycopsids suggest that rootlets are root-like[2]. Together these three studies present an emerging body of evidence that is incompatible with the modified shoot hypothesis.

To independently test the modified shoot hypothesis for the origin of lycopsid roots, we evaluated gene expression data of the extant rhizomorphic lycopsid, *Isoetes echinospora*. We generated, to our knowledge, the first organ specific transcriptome of an *Isoetes* species incorporating RNA from the three main organs of the sporophyte: rootlets, leaves and corms. If *I. echinospora* rootlets are modified leaves as predicted by the modified-shoot hypothesis we would expect gene expression profiles to be similar in rootlets and leaves. If, on the other hand, *I. echinospora* rootlets are true roots as predicted by the lycopsid root hypothesis we would expect that gene expression profiles would be different between leaves and rootlets, and gene expression profiles would be similar between *I. echinospora* rootlets and roots of *Selaginella* species.

Results

Assembly of an *Isoetes echinospora* sporophyte transcriptome. To define gene expression signatures in the organs of *I. echinospora,* a population of plants was collected from the wild (Fig. 1a) and protocols to grow the plants in axenic culture were developed (Fig. 1b,c). RNA was extracted from axincially grown plants from the three major organs; leaves, corm and rootlets and was sequenced (Fig. 2a). It was difficult to extract suf-

ficient RNA from these plants because of the challenge in isolating viable spores, getting the spores to germinate, effecting fertilisation and getting sporophytes to develop in axenic culture. However, we extracted 1 technical replicate of rootlets and 2 technical replicates of corm and leaves. A sporophyte transcriptome was generated for rootlets corms and leaves. The assembled transcriptome comprised 113,464 transcripts with a mean sequence length of 940 base pairs (bp). There were 35,564 sequences over 1 kilobases (Kb) in the assembly, with an N50 of 1313 bp. Proteins were successfully predicted for c. 95% of the transcripts. To investigate the completeness our transcriptome we next performed a BUSCO[35] analysis to investigate the number of conserved BUSCO[35] groups in our transcriptome. BUSCO[35] groups are near-universal single-copy orthologs. Identifing the percentage of BUSCO[35] groups present in our *de-novo* transcriptome therefore provides a metric for the completeness of our transcriptome. Of the 430 total BUSCO[35] groups searched for in the Viridiplantae dataset[35], 318 (74.0%) were found complete, 87 (20.2%) were found fragmented and only 25 (5.8%) were missing. These metrics indicate that the transcriptome assembly was high quality. We next mapped the reads extracted from each of the three different organs; leaves, corms, and rootlets to calculate the abundance levels for each transcript in each of the three organs (Supplementary Table S1).

Gene expression profiles are significantly different in *Isoetes echinospora* rootlets and leaves.

If *I. echinospora* rootlets were modified leaves, as predicted by the modified shoot hypothesis, we might expect gene expression signatures to be similar in the rootlets and leaves. To test this hypothesis, we compared gene expression in rootlets, leaves and corms using a principal coordinate analysis (PCoA), because similar approaches have proved successful for the investigation of lycophyte root transcriptomes[36,37]. The two leaf replicates, two corm replicates and the single rootlet sample were plotted on the first two PCoA axes [which together account for 98.6% of the variance in the sample (Fig. 2b)]. The three tissue types are clearly distinct and separated in gene expression space. The first PCoA axis accounts for 75.4% of the variance in gene expression and it distinguishes leaves and corms from rootlets (Fig. 2b). The second PCoA axis accounts for 23.2% of the variance in gene expression and distinguishes all three tissues from each other (Fig. 2b). The PCoA indicated that gene expression profiles of rootlets and leaves are distinct and does not support the hypothesis that *I. echinospora* rootlets are modified leaves.

Gene expression profiles of *Isoetes* rootlets clusters with gene expression of *Selaginella* and *Arabidopsis* roots.

If the rootlets of *I. echinospora* are true roots we expected similarities in gene expression between rootlets and true roots of other land plant species such as the lycophyte *Selaginella moellendorffii* and the seed plant *Arabidopsis thaliana*. To compare gene expression between these species we first defined orthologous relationships between the genes of the three species using the OrthoFinder software[38,39] (Supplementary Table S1). This analysis identified 1737 single copy orthologs in common between these species. We focussed our analysis on orthologs with a 1–1–1 orthology relationship between the three species as expression patterns of single copy orthologs are more likely to be conserved than those of dublicated genes[40]. Using these 1737 orthologs we compared gene expression between the different species. We compared average gene expression between *I. echinospora* rootlets and leaves (this study) with the published gene expression in roots and leaves of *S. moellendorffii*[36] and roots and "aerial parts" of *Arabidopsis thaliana* (based on EMBL-EBI accession E-GEOD-53197) (Supplementary Table S1). To compare gene expression between these different species and organs we subjected the gene expression dataset to a PCoA. The first three principal coordinates accounted for 95.7% of the variance in the dataset. Axis 1 accounted for 43.6% of the variance and separated the samples by species (Fig. 3a,b). Axis 2, accounted for 35.9% of the variance and distinguished the two lycophyte transcriptomes (*I. echinospora* and *S. moellendorffii*) from that of the seed plant *A. thaliana* (Fig. 3a,c). PCoA axes one and two therefore indicate that the majority of the differences in gene expression is accounted for by differences between species rather than between roots and leaves. PCoA axis 3 accounted for 16.2% of the variance and distinguished between leaves and roots in all species (Fig. 3b,c). Leaf samples clustered in the positive values and root samples clustered in the negative values of PCoA axis 3 (Fig. 3b,c). The clustering of the *I. echinospora* rootlet sample with both the roots of *S. moellendorffii* and *A. thaliana* on axis 3 (Fig. 3b,c) indicates that the gene expression signature of the rootlets of *I. echinospora* is similar to the the gene expression signature of both *S. moellendorffii* and *A. thaliana*. It is possible that the clustering of *I. echinospora* rootlets with root gene expression profiles of *S. moellendorffii* and *A. thaliana* resulted from the absence of photosynethetic-related genes in these rooting organs. To test if this accounted for the clustering that we observed, we carried out a further investigation of gene expression after removing genes encoding photosynethetic functions. Of the 1737 orthologs used in the analysis only 47 encoded for photosynthetic functions or were encoded in the chloroplast as assigned by MapMan[41] to the term "photosynthesis". We subjected the gene expression datasets for the remaining 1,690 orthologs to a PCoA. The results of the PCoA (Supplementary Fig. S1) were analogous to the result using all 1737 orthologs allowing us to rule out that the clustering of *I. echinospora* rootlets with the roots of *S. moellendorffii* and *A. thaliana* in gene expression space was due to genes encoding photosynthetic functions. These gene expression data are consistent with the hypothesis that rootlets of *I. echinospora* are roots.

The *RSL* root cell differentiation genes are expressed in *Isoetes echinospora* rootlets.

To verify our findings that gene expression of *I. echinospora* rootlets were similar to those of the true roots of *S. moellendorffii* and *A. thaliana* we next determined the expression of the root-specific *ROOT HAIR DEFECTIVE SIX-LIKE* (*RSL*) genes in *I. echinospora*. *ROOT HAIR DEFECTIVE SIX-LIKE* (*RSL*) genes positively regulate the development of root hairs in euphyllophytes including *A. thaliana*[42–45] and are expressed in *S. moellendorffii* roots[37,46]. *RSL* genes are markers for vascular plant roots because they are expressed at a much higher level in roots of *A. thaliana* (EMBL-EBI accession E-GEOD-53197) and *S. moellendorffii*[36] than in leaves and shoots
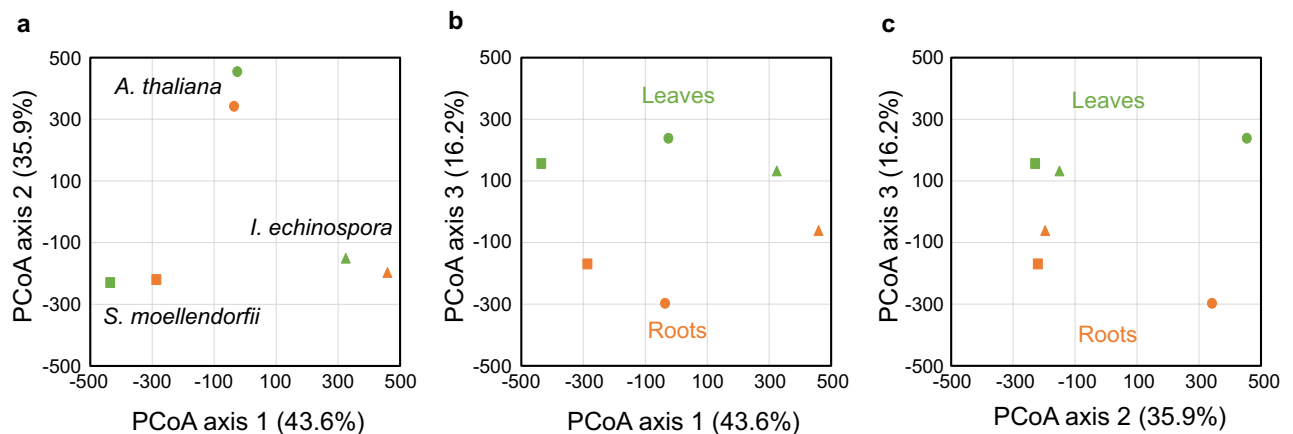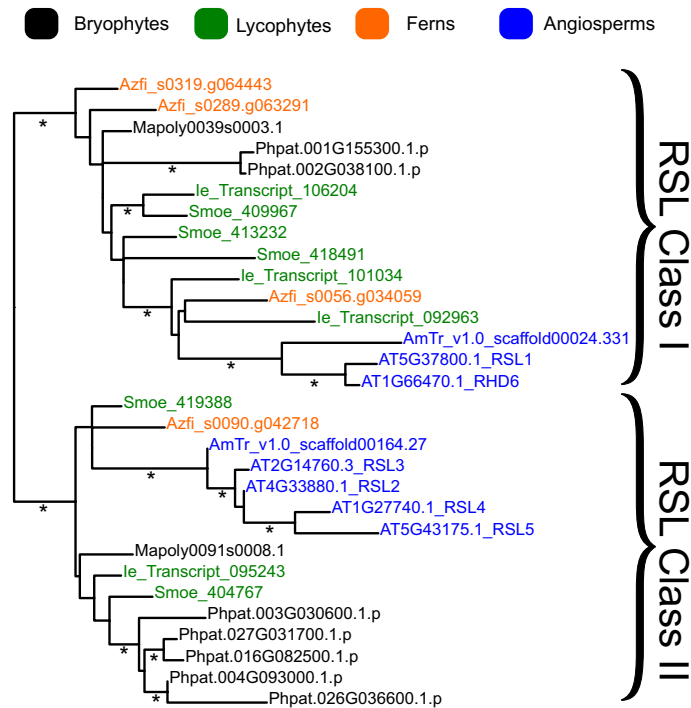
Figure 3. Comparison of root and leaf transcriptomes of *Arabidopsis thaliana*, *I. echinospora* and *Selaginella moellendorffii*. Comparison of gene expression profiles by PCoA in the transcriptomes of *A. thaliana*, circles, *S. moellendorffii*, squares and *I. echinospora*, triangles. Leaf samples coloured green, root samples orange. (**a**) comparison of principal coordinate axis 1 and 2. (**b**) comparison of principal coordinates axis 1 and 3. (**c**) comparison of principal coordinate axis 2 and 3. Values in brackets on each axis describe the percentage of total variance accounted for by each axis. Axis 1 separates gene expression in the three species. Axis 2 distinguishes gene expression between the two lycophytes transcriptomes, *I. echinospora* and *S. moellendorffii* from *A. thaliana*. Axis 3 distinguishes between the leaf samples and the root samples in each transcriptome.

(Supplementary Table S1). We searched the *I. echinospora* transcriptome for RSL genes using the BLAST algorithm with RSL-specific queries. RSL sequences were identified in the *I. echinospora* transcriptome. A gene tree was generated and defined four *I. echinospora* RSL genes in two monophyletic groups (Fig. 4). There were three transcripts in the RSL Class I clade (106204; 101034; 092963) and a single transcript in the RSL Class II clade (095243). We next investigated the expression of the *I. echinospora* RSL genes in the transcriptome, generated with single biological rootlet replicate. Average expression of the four RSL genes in rootlets was 4.24 transcripts per million (TPM) (Fig. 4). The average root expression was 5.78 TPM for the six RSL genes of *A. thaliana* (EMBL-EBI accession E-GEOD-53197), demonstrating similarities in expression of RSL genes between *I. echinospora* and *A. thaliana*. In *I. echinospora*, expression of each RSL Class I transcript was higher in rootlets than in leaves (Fig. 4). Furthermore, the single *I. echinospora* Class II RSL gene transcript (095243) was expressed in rootlets and no expression was detected in the corm or leaves. The analysis of our transcriptome demonstates that RSL genes are preferentially expressed in *I. echinospora* rootlets and not in leaf tissue; their expression is the same in *I. echinospora* as in *S. moellendorffii* of *A. thaliana* (Supplementary Table S1). The analysis of RSL gene expression is based on a single biological replicate because of the technical limitations of working with *I. echinospora* in axenic culture; we would have preferred to present data from replicated data of root transcriptomes. However, these results are presented here because they are consistent with the other data presented above that support the hypothesis that *I. echinospora* rootlets are roots.

To verify that RSL genes are markers of vascular plant roots we investigated the *RSL* genes in *Azolla filiculoides*, a fern that develops roots with root hairs[30,48], and *Salvinia cucullata* a fern that has secondarily lost roots with root hairs and instead modified leaves perform rooting functions[30,49,50]. We searched the *S. cucullata* and *A. filiculoides* genomes and proteomes[51] for RSL genes using the BLAST algorithm with RSL-specific queries. A gene tree was constructed with the retrieved sequences and allowed us to identify 3 RSL Class I genes and a single RSL Class II gene in the *A. filiculoides* genome (Fig. 4, Supplementary Fig. S2). Consistant with their role in root development in *A. filiculoides* the RSL genes were expressed in the roots[48]. However, there were no RSL genes in the *S. cucullata* genome. *S. cucullata* sequences were identified in closely related basic-helix-loop-helix transcription factor subfamily XI[52,53] but none were identified in the RSL clade (Fig. 4, Supplementary Fig. S2). We conclude that the loss of RSL genes accompanied the evolutionary loss of roots with root hairs in *Salvinia cucullata*, which is consistent with RSL genes being markers of vascular plant roots. Furthermore, if the rootlets of *I. echinospora* were modified leaves, similar to the root-like modified leaves of *S. cucullata*, we might have expected that RSL genes would have also been lost from the *I. echinospora* genome. Instead, RSL genes are preferentially expressed in *I. echinospora* rootlets just as they are in *S. moellendorffii* roots. These data are consistent with the hypothesis that the *I. echinospora* rootlet is a root and not a modified leaf.

Taken together these data—the distinct gene expression profiles of the rootlets and leaves of *I. echinospora*, the similarity in expression profiles of orthologous gene preferentially expressed in rootlets of *I. echinospora* and roots of *S. moellendorffii* and *A. thaliana*, and the expression of the *RSL* genes in the rootlets of *I. echinospora* and roots of *S. moellendorffii* and *A. thaliana* – support the lycopsid root hypothesis which posits that *Isoetes* rootlets are roots and not modified leaves.

		Expression (TPM)		
	Transcript number	Leaf	Corm	Rootlet
Class I	Ie_Transcript_106204	1.13	3.46	9.12
	Ie_Transcript_101034	0.51	0.72	3.21
	Ie_Transcript_092963	1.24	0.98	2.01
Class II	Ie_Transcript_095243	0	0	2.61

Figure 4. Gene tree analysis and expression analysis of *I. echinospora* RSL genes. Top, maximum likelihood gene tree of RSL genes generated in PhyML 3.0[47]. Gene names: black, bryophytes; green, lycopsids; orange, ferns; blue, angiosperms. The RSL genes are grouped into two monophyletic classes; Class I and Class II. Bottom, gene expression in the three RSL Class I genes and single RSL Class II gene in the *I. echinospora* transcriptome. Expression in transcripts per million (TPM) given for the single replicate of rootlets and as an average of the two technical replicates of leaf and corm. Species name abbreviations: Phpat, *Physcomitrella patens*; Mapoly, *Marchantia polymorpha*; Some, *Selaginella moellendorffii*; Sk, *Selaginella kraussiana*; Ie_Transcript, *Isoetes echinospora*; Sacu, *Salvinia cucullata*; Azfi, *Azolla filiculoides*; AmTr, *Amborella trichopoda*; AT *Arabidopsis thaliana*. *Indicate branches with over 0.85 aLRT SH-like support.

Discussion

The homology of the rootlets of both extinct and extant rhizomorphic lycopsids had been contentious for the past 150 years, with two competing hypotheses. The first, interprets the rootlets as true roots similar to the roots of other lycopsids. The second, interprets rootlets as modified leaves. Despite the second hypothesis that posits that rootlets are modified leaves being widely accepted over the past 30 years[1,3,54], there is a growing body of evidence[2,34] that suggests that rootlets should be interpreted as true roots. Here we report the de novo transcriptome of *I. echinospora* that we used to test predictions of the two competing hypotheses. We discovered that expression profiles in *I. echinospora* rootlets and leaves were different. We showed that gene expression profiles of *I. echinospora* rootlets and *S. moellendorffii* and *A. thaliana* roots were similar. Finally, RSL genes involved in root cell differentiation are preferentially expressed in *I. echinospora* rootlets as they are in *S. moellendorffii* roots and the roots of euphyllophytes (*A. thaliana*, *Oryza sativa* and *Brachypodium distachyon*[42–45]). Taking these three pieces of evidence together, we conclude that *Isoetes* rootlets are true roots, like those of extinct and extant lycopsids and not modified leaves.

The new evidence presented here adds to the growing and extensive list of similarities between the rootlets of rhizomorphic lycopsids—*Isoetes* species and extinct taxa such as *Stigmaria*—and the roots of other lycopsids[2,20,34]. This growing body of evidence supports the hypothesis that rootlets are roots and not modified leaves. The rootlets of the rhizomorphic lycopsids and roots of all extant lycopsids are indeterminate radially symmetric axes that branch by isotomous dichotomy, develop endogenously within specialised structures, develop a root

meristem with root cap and produce root hairs[31,32,34,55]. If the modified shoot hypothesis were correct it would have required the direct modification of a determinate leaf that did not branch, developed exogenously and was characterised by a ligule, stomata and dorsiventral symmetry into a rootlet. Each of these leaf characters would have had to be lost and all of the rootlet characters, which are shared among the lycopsids, would have had to evolve independently. By contrast, if the lycopsid root hypothesis is correct and rootlets are roots then there is no requirement for this large suite of character state changes. Instead, the only character transitions required to account for rootlet character states were the collateral positioning of the phloem, the regular rhizotaxy and rootlet abscission[4]. Although these three characters (collateral positioning of the phloem, the regular rhizotaxy and rootlet abscission) are predominately leaf characters, they are not exclusive to leaves; each has been described in the roots of other species of land plants. The collateral position of the phloem is found in Lycopodium roots including, *Lycopodium lucidulum*, *Lycopodium clavalum*, *Lycopodium obscurum* and *Lycopodium complanalum*[56], regular rhizotaxy develops in *Ceratoptertis thalictroides*, *Cucurbita maxima* and *Pontederia cordata*[57–59] and roots abscise in *Oxalis esculenta*, *Abies balsamea*, *Pinus strobus*, *Tsuga canadensis* and *Azolla* species[60–65]. Based on character transitions alone we suggest that the hypothesis that rootlets of the rhizomorphic lycopsids are roots, similar to other lycopsid roots, is a more parsimonious hypothesis than interpreting rootlets as modified leaves.

Our new evidence from the transcriptome of *I. echinospora* adds to the numerous traits that are common between the rootlets of rhizomorphic lycopsids and the roots of other lycopsids. It is not possible to rule out the hypothesis that all of these similarities in antomy, development and now gene expression may be the product of convergent evolution. However, we suggest that it is more parsimonious to interpret the rootlets of the rhizomophic lycopsids as true roots than modified leaves.

The gene expression data from the de novo *I. echinospora* transcriptome are consistent with the hypothesis that the rootlets of the rhizomorphic lycopsids are roots and not modified leaves. We therefore interpret the rootlets of the rhizomophic lycopsids as roots developing from a unique root bearing organ; the rhizomorph[21,55,66]. This conclusion suggests that the dichotomously branching rooting axis is conserved among all lycopsids and a distinguishing character of the group. The dichotomous branching of these rooting axes has been conserved for over 400 million years and our comparative transcriptomic analysis suggest that the RSL genes function during root development in *Selaginella* and *Isoetes* has been conserved since these species shared a common ancestor at least 375 million years ago[67]. Our comparative analysis of the transcriptomes of extant lycophytes supports the hypothesis that the rooting systems of extant *Isoetes* species and their extinct giant ancestors are homologous. These data also suggest that the development of the large rooting systems of the lycopsid trees that were an important component of the Palaeozoic flora and played a key role in changing the Earth's Carbon Cycle were controlled by the same genes that regulate root development in their extant herbaceous descendants.

Materials and methods

Plant collection and growth. Mature *I. echinospora* plants were collected from Loch Aisir and Loch Dubhaird Mor in September 2013 and 2014 from North West Sutherland (Scotland, UK) with the permission of the John Muir Trust and the Scourie Estate. *I. echinospora* plants were identified on the basis of their echinate megaspore ornamentation[68]. Mature *I. echinospora* plants were grown submerged in aquaria in Levington M2 compost topped with coarse gravel in a glasshouse at Oxford University at 18 °C under a 16 h light : 8 h dark photoperiod.

Growth of *I. echinospora* in axenic culture. RNA was extracted from plants grown in axenic culture to ensure that there was no RNA contamination from other organisms. A procedure was developed to surface sterilise *I. echinospora* spores and germinate a population of axenically grown plants, based on previously developed procedures[69–71]. Sporophylls were removed from the mature plant population growing in aquaria in September (2013 and 2014) when sporangia were mature[69]. Using forceps (under a Leica M165 FC stereo microscope) mega- and micro-sporangia were isolated from sporophylls. Intact sporangia were washed in 1% (v/v) sodium dichloroisocyanurate (NaDCC) for 5 min. Sporangia were broken and loose spores were washed in 0.1% NaDCC for a further 5 min. Following the NaDCC washes, loose spores were rinsed for 5 min three times in ddH$_2$O. Microspores were centrifuged for 5 min at 5000 rpm between washes with NaDCC. Once sterilised, mega and micro-spores were mixed together in ddH$_2$O in a Petri dish. Petri dishes were sealed with parafilm, and incubated in darkness at 4 °C for 2 week. After 2 weeks, Petri dishes were moved to a 16 h light:8 h dark photoperiod at 18 °C. Approximately 30% of surface sterilised megasporangia contained megaspores that germinated, and within these megasporangia c. 25% of the total megaspore population germinated. It was possible to identify germinating megaspores because cracking of the megaspore wall was visible and the presence of archegonia on the megagametophyte. Once fertilisation occurred, developing sporophytes were identified by the presence of the first leaf. Sporophytes were left to continue to grow in ddH$_2$O water until the two leaf two rootlet stage when they were moved to magenta boxes containing; ½ Gamborg's medium[72], supplemented with 1% phytogel (Sigma). Plants were embedded in Gamborg media and submerged in liquid Bold's Basal Medium (Sigma, UK).

RNA extraction and sequencing. Total RNA was extracted from root, corm and leaf tissues from c. 50 *I. echinospora* plants. Total RNA from leaves (two independent replicates), corm (two independent replicates) and rootlets (one replicate) was extracted with the RNeasy plant mini kit (Qiagen). On-column DNase I treatment was performed with RNase-free DNase I (Qiagen), according to the manufacturer's instructions. cDNA was synthesised with ProtoScript II reverse transcriptase (New England Biolabs) according to the manufacturer's instructions, using oligo(dT) primer. Total cDNA samples were quantified with a Nanodrop ND-1000 spectrophotometer. RNA purity and quality were checked with an Agilent 2100 Bioanalyzer. Library preparation and sequencing was carried out by the High-Throughput Genomics Group at the Wellcome Trust Centre

for Human Genetics, University of Oxford in 2015. mRNA was selected from the total RNA and converted to cDNA, followed by second strand synthesis to incorporate dUTP. cDNA was then end-repaired, A-tailed and adapter-ligated. Samples then underwent uridine digestion. The libraries were then size selected, mutiplexed and quality checked before paired end sequencing over one lane of a flow cell using Illumina HiSeq 2000. Sequencing resulted in 195,072,304 paired end reads separated into five samples: 2 leaves samples (35,718,157; 35,555,048 paired end reads), 2 corm samples (38,728,989; 44,379,751 paired end reads) and one rootlet sample (40,690,359 paired end reads). The raw read libraries have been deposited under SRP135936 on the NCBI Sequence Read Archive.

De novo transcriptome assembly, protein predictions and expression analysis.   Raw reads were quality trimmed using Trimmomatic-0.32[73], to remove remaining Illumina adaptors and low quality tails. Ribosomal RNA was filtered out using Sortmerna-1.9[74] and error corrected using BayesHammer (SPAdes-16 3.5.0)[75] (with setting -only-error-correction) and Allpaths-LG-4832[76] (with setting PAIRED_SEP = – 20 and ploidy = 2). Reads were normalised using Khmer-0.7.1 with a khmer size of 21. Before assembly, paired end reads were stitched together using Allpaths-LG-4832[76]. A de novo transcriptome assembly was made with the cleaned, stitched reads using SGA[77], SSPACE-v3[78], and CAP3[79]. Finally assembled scaffolds were corrected using Pilon-1.6[80]. The Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GGKY00000000. The version described in this paper is the first version, GGKY01000000. Proteins were predicted from the de novo transcriptome assembly using GeneMarkS-T[81], Prodigal[82] and Transdecoder (part of the Trinity assembly program[83]), proteins were deposited on Zenodo (http://doi.org/10.5281/zenodo.3574570). A BUSCO analysis[35], using BUSCO 3.1.0 and the *viridiplantae_odb10* database.

Comparison of gene expression between *I. echinospora* organs.   Using the sporophyte transcriptome assembly we next mapped the reads from the three organ libraries—leaves, corms, and rootlets—to the transcriptome to measure the expression levels of each transcript in the three tissues using Salmon[84]. To investigate the similarities between gene expression in the different organs we carried out a PCoA on the three organ types. Euclidean distances were derived from the expression of all transcripts (TPM) in each organ and were subjected to PCoA in PAST[85] using a transformation exponent of 2.

OrthoFinder analysis and comparison of gene expression between *I. echinospora, S. moellendorffii* and *Arabidopsis thaliana*.   Orthologous relationships between *I. echinospora, S. moellendorffii* and *A. thaliana* proteins were determined using OrthoFinder[38,39]. OrthoFinder was run with *I. echinospora* proteins and protein datasets for 57 species from Phytozome (full list of species in Supplementary Table S2) including the Rhodophyta *Porphyra umbilicalis*, seven species of chlorophytes, the bryophytes *Marchantia polymorpha* and *Physcomitrella patens*, the lycophytes *Selaginella moellendorffii* and 46 angiosperm species. This analysis resulted in the identification of 38,217 orthogroups, accounting for 82.6% of all genes included in the analysis (the results of the OrthoFinder analysis were we deposited on Zenodo, http://doi.org/10.5281/zenodo.3574570).

To compare gene expression between *I. echinospora*, *S. moellendorffii* and *A. thaliana* we identified single copy orthologs between these species based on the OrthoFinder[38,39] analysis. In total, 1737 single copy orthologs were found between the three species. Using these 1737 orthologs we contrasted gene expression between the different species. We investigated average genes expression between *I. echinospora* rootlets and leaves (this study) with the published average gene expression between roots and leaves of *S. moellendorffii*[36] and *A. thaliana*. *A. thaliana* gene expression was based on average gene expression in "aerial part" and "root" of 17 different natural accessions (EMBL-EBI accession E-GEOD-53197). To investigate similarities in gene expression between these 1737 orthologs we carried out a PCoA in PAST[85]. Euclidean distances were derived from the Log10 transformed gene expression of the 1737 orthologs (Supplementary Table S1). Euclidean distances were subjected to a PCoA in PAST[85], using a transformation exponent of 4. To rule out that the results were not influenced by genes encoding photosynthetic fucntions, we investigated how many of the 1737 genes were assigned to the MapMan[41] term "photosynthesis". Of the 1737 only 47 were photosynthesis related or encoded in the chloroplast (Supplementary Table S1). We ran a PCoA on these 1,690 using the rsame methods described above.

Phylogenetic analyses.   Phylogenetic analyses were carried out on the RSL genes. BLAST queries were assembled based on previously published gene trees of RSL genes[52]. Sequences were used to BLAST the protein databases of the; *Marchantia polymorpha* "primary" (proteins) (version 3.1, November, 2015), *Physcomitrella patens* "primary" (proteins) (version 3.0, January 12, 2014), *Selaginella moellendorffii* "primary" (proteins) (version 1.0, January 12, 2014), *Amborella trichopoda* (proteins) (version 1.0, 2013) and *Arabidopsis thaliana* "primary" (proteins) (TAIR10) on the http://marchantia.info/blast/ server. Two fern protein databases were also searched; *Azolla filiculoides* protein v1.1 and *Salvinia cucullata* proteins v1.2[51] as well as the predicted proteins from the *I. echinospora* transcriptome generated in this study. All proteins were aligned in MAFFT[86,87], manually edited in Bioedit[88]. Maximum likelihood gene trees were generated in PhyML 3.0[47], using Jones, Taylor and Thorton (JTT) amino acid substitution model. To verify the absence of RSL genes in the *S. cucullata* genome and proteomes the genomes of *A. filiculoides* and *S. cucullata* were searched by BLAST using the *A. thaliana* protein sequence RSL1 (AT5G37800) using an E-value cut off of 1E-15. A gene tree was generated as described above including the addition of *A. thaliana* protein sequences from subfamilies VIIIb and XI[52,53] (Fasta alignments files for both gene trees we deposited on Zenodo, http://doi.org/10.5281/zenodo.3574570).

Data availability

References

1. Stewart, W. & Rothwell, G. W. *Paleobotany and the Evolution of Plants* (Cambridge University Press, Cambridge, 1993).
2. Hetherington, A. J., Berry, C. M. & Dolan, L. Networks of highly branched stigmarian rootlets developed on the first giant trees. *Proc. Natl. Acad. Sci.* **113**, 6695–6700 (2016).
3. Rothwell, G. W. & Erwin, D. The rhizomorph apex of *Paurodendron*: implications for homologies among the origins of Lycopsida. *Am. J. Bot.* **72**, 86–98 (1985).
4. Karrfalt, E. E. A further comparison of *Isoetes* roots and stigmarian appendages. *Can. J. Bot.* **58**, 2318–2322 (1980).
5. Williamson, W. A monograph on the morphology and histology of *Stigmaria ficoides*. *Palaeontogr. Soc.* **40**, 1–62 (1887).
6. Weiss, F. E. A re-examination of the stigmarian problem. *Proc. Linn. Soc. Lond.* **144**, 151–166 (1933).
7. Schimper, W. P. *Traité de Paléontologie Végétale ou la Flore du Monde Primitive Dans ses Rapports Avec les Formations Géologiques et la Flore du Monde Actuel* (J. B. Baillière et Fils, London, 1872).
8. Solms-Laubach, H. G. *Fossil Botany being an introduction to palaeophytology from the stanpoint of the botanist. Translated into English by Garnsey, H. E and Balfour, I. B.* (Clarendon Press, Oxford, 1891).
9. Stewart, W. A comparative study of stigmarian appendages and *Isoetes* roots. *Am. J. Bot.* **34**, 315–324 (1947).
10. Lang, W. H. On the apparently endogenous insertion of the roots of *Stigmaria*. *Mem. Proc. Manchester Lierary Philos. Soc.* **67**, 101–107 (1923).
11. Bower, F. O. *The Origin of a Land Flora* (Macmillan and Co., London, 1908).
12. Leclercq, S. A monograph of *Stigmaria bacupensis*, Scott et Lang. *Ann. Bot.* **44**, 31–54 (1930).
13. Weiss, F. E. Morphology of *Stigmaria* and of its appendages in comparison with recent lycopodiales. *Proc. Linn. Soc. Lond.* **120**, 74–75 (1908).
14. Scott, D. H. *Studies in Fossil Botany* (Adam and Charles Black, London, 1900).
15. Rothwell, G. W. The apex of *Stigmaria* (Lycopsida), rooting organ of Lepidodendrales. *Am. J. Bot.* **71**, 1031–1034 (1984).
16. Stubblefield, S. P. & Rothwell, G. W. Embryogeny and reproductive biology of *Bothrodendrostrobus mundus* (Lycopsida). *Am. J. Bot.* **68**, 625–634 (1981).
17. Phillips, T. L. Reproduction of heterosporous arborescent lycopods in the Mississippian-Pennsylvanian of Euramerica. *Rev. Palaeobot.* **27**, 239–289 (1979).
18. Frankenberg, J. M. & Eggert, D. A. Petrified Stigmaria from North America. I. Stigmaria ficoides, the underground portions of Lepidodendraceae. *Palaeontogr. B* **128**, 1–47 (1969).
19. Eggert, D. A. Petrified *Stigmaria* of sigillarian origin from North America. *Rev. Palaeobot. Palynol.* **14**, 85–99 (1972).
20. Bateman, R. M. Morphometric reconstruction, palaeobiology and phylogeny of *Oxroadia gracilis* Alvin emend. and *O. conferta* sp. nov.: anatomically-preserved rhizomorphic lycopsids from the Dinantian of Oxroad Bay, SE Scotland. *Palaeontogr. Abteilung B* **228**, 29–103 (1992).
21. Bateman, R. M., DiMichele, W. A. & Willard, D. A. Experimental cladistic analysis of anatomically preserved arborescent Lycopsids from the Carboniferous or Euramerica: an essay on paleobotanical phylogenetics. *Ann. Missouri Bot. Gard.* **79**, 500–559 (1992).
22. DiMichele, W. A. & Bateman, R. M. The rhizomorphic lycopsids: a case-study in paleobotanical classification. *Syst. Bot.* **21**, 535–552 (1996).
23. Kenrick, P. & Crane, P. R. *The Origin and Early Diversification of Land Plants: A Cladistic Study* (Smithsonian Institute Press, Washington, DC, 1997).
24. Hofmeister, W. F. B. *On the Germination, Development, and Fructification of the Higher Cryptogamia: And on the Fructification of the Coniferæ. Translated to English by Frederick Currey*. (1862).
25. West, C. & Takeda, H. X. On *Isoetes japonica*, A. Br. *Trans. Linn. Soc. Lond. Bot.* **8**, 333–376 (1915).
26. Van Tieghem, P. & Douliot, H. Recherches comparatives sur l'origine des membres endogenes dans les plantes vasculaires. *Ann. Sci. Nat. Bot. Paris* **8**, 1–656 (1888).
27. Sachs, J. *Text-Book of Botany, Morphological and Physiological. Translated by A. W. Bennett and W. T. Thiselton Dyer* (Clarendon Press, Oxford, 1875).
28. Foster, A. S. & Gifford, E. M. *Comparative morphology of vascular plants* (W. H. Freeman and Company, London, 1959).
29. Guttenberg, H. V. *Histogenese der Pteridophyten. Handbuch der Pflanzenanatomie vol. VII. 2.* (Gebrüder Borntraeger, 1966).
30. Eames, A. J. *Morphology of Vascular Plants, Lower Groups* (MCGraw-Hill Book Company Inc, New York, 1936).
31. Ogura, Y. *Comparative anatomy of vegetative organs of the pteridophytes. Handb Pflanzenanat*. (Gebrüder Borntraeger, 1972).
32. Bierhorst, D. W. *Morphology of Vascular Plants* (Macmillan, London, 1971).
33. Charlton, W. A. & Watson, J. Patterns of arrangement of lateral appendages on axes of *Stigmaria ficoides* (Sternberg) Brongniart. *Bot. J. Linn. Soc.* **84**, 209–221 (1982).
34. Yi, S. & Kato, M. Basal meristem and root development in *Isoetes asiatica* and *Isoetes japonica*. *Int. J. Plant Sci.* **162**, 1225–1235 (2001).
35. Waterhouse, R. M. *et al.* BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* **35**, 543–548 (2018).
36. Mello, A., Efroni, I., Rahni, R. & Birnbaum, K. D. The *Selaginella* rhizophore has a unique transcriptional identity compared with root and shoot meristems. *New Phytol.* **222**, 882–894 (2019).
37. Huang, L. & Schiefelbein, J. Conserved gene expression programs in developing roots from diverse plants. *Plant Cell* **27**, 2119–2132 (2015).
38. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
39. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
40. Huerta-Cepas, J., Dopazo, J., Huynen, M. A. & Gabaldon, T. Evidence for short-time divergence and long-time conservation of tissue-specific expression after gene duplication. *Brief. Bioinform.* **12**, 442–448 (2011).
41. Thimm, O. *et al.* MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* **37**, 914–939 (2004).
42. Kim, C. M. & Dolan, L. ROOT HAIR DEFECTIVE SIX-LIKE Class I genes promote root hair development in the grass *Brachypodium distachyon*. *PLOS Genet.* **12**, e1006211 (2016).

43. Masucci, J. D. & Schiefelbein, J. W. The *rhd6* mutation of *Arabidopsis thaliana* alters root-hair initiation through an auxin- and ethylene-associated process. *Plant Physiol.* **106**, 1335–1346 (1994).
44. Menand, B. *et al.* An ancient mechanism controls the development of cells with a rooting function in land plants. *Science* **316**, 1477–1480 (2007).
45. Kim, C. M. & Dolan, L. Root hair development involves asymmetric cell division in *Brachypodium distachyon* and symmetric division in *Oryza sativa*. *New Phytol.* **192**, 601–610 (2011).
46. Huang, L., Shi, X., Wang, W., Ryu, K. H. & Schiefelbein, J. Diversification of root hair development genes in vascular plants. *Plant Physiol.* **174**, 1697–1712 (2017).
47. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 30. *Syst. Biol.* **59**, 307–321 (2010).
48. de Vries, J. *et al.* Cytokinin-induced promotion of root meristem size in the fern *Azolla* supports a shoot-like origin of euphyllophyte roots. *New Phytol.* **209**, 705–720 (2016).
49. Croxdale, J. G. *Salvinia* leaves. I. Origin and early differentiation of floating and submerged leaves. *Can. J. Bot.* **56**, 1982–1991 (1978).
50. Croxdale, J. G. *Salvinia* leaves. III. Morphogenesis of the submerged leaf. *Can. J. Bot.* **59**, 2065–2072 (1981).
51. Li, F.-W. *et al.* Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nat. Plants* **4**, 460–472 (2018).
52. Catarino, B., Hetherington, A. J., Emms, D. M., Kelly, S. & Dolan, L. The stepwise increase in the number of transcription factor families in the Precambrian predated the diversification of plants on land. *Mol. Biol. Evol.* **33**, 2815–2819 (2016).
53. Pires, N. & Dolan, L. Origin and diversification of basic-helix-loop-helix proteins in plants. *Mol. Biol. Evol.* **27**, 862–874 (2010).
54. Taylor, E. L., Taylor, T. N. & Krings, M. *Paleobotany: The Biology and Evolution of Fossil Plants* (Academic Press, Burlington, 2009).
55. Hetherington, A. J. & Dolan, L. The evolution of lycopsid rooting structures: conservatism and disparity. *New Phytol.* **215**, 538–544 (2017).
56. Pixley, E. Y. A study of the ontogeny of the primary xylem in the roots of *Lycopodium*. *Bot. Gaz.* **129**, 156–160 (1968).
57. Mallory, T. E., Chiang, S. H., Cutter, E. G. & Gifford, E. M. Sequence and pattern of lateral root formation in five selected species. *Am. J. Bot.* **57**, 800–809 (1970).
58. Hou, G., Hill, J. P. & Blancaflor, E. B. Developmental anatomy and auxin response of lateral root formation in *Ceratopteris richardii*. *J. Exp. Bot.* **55**, 685–693 (2004).
59. Charlton, W. A. Distribution of lateral roots and pattern of lateral initiation in *Pontederia cordata* L. *Bot. Gaz.* **136**, 225–235 (1975).
60. Duncan, J. F. "Pull Roots" of *Oxalis esculenta*. *Trans. Bot. Soc. Edinb.* **29**, 192–196 (1927).
61. Tippett, J. T. Shedding of ephemeral roots in gymnosperms. *Can. J. Bot.* **60**, 2295–2302 (1982).
62. Konar, R. N. & Kapoor, R. K. Anatomical studies on *Azolla pinnata*. *Phytomorphology* **1**, 211–223 (1972).
63. Uheda, E., Nakamura, S. & Kitoh, S. Aspects of the very rapid abscission of *Azolla* branches: anatomy and possible mechanism. *Int. J. Plant Sci.* **156**, 756–763 (1995).
64. Fukuda, K., Yamada, Y., Miyamoto, K., Ueda, J. & Uheda, E. Separation of abscission zone cells in detached *Azolla* roots depends on apoplastic pH. *J. Plant Physiol.* **170**, 18–24 (2013).
65. Yamada, Y., Koibuchi, M., Miyamoto, K., Ueda, J. & Uheda, E. Breakdown of middle lamella pectin by ·OH during rapid abscission in *Azolla*. *Plant Cell Environ.* **38**, 1555–1564 (2015).
66. Gensel, P. G. & Berry, C. M. Early lycophyte evolution. *Am. Fern J.* **91**, 74–98 (2001).
67. Wikström, N. & Kenrick, P. Evolution of Lycopodiaceae (Lycopsida): estimating divergence times from *rbcL* gene sequences by use of nonparametric rate smoothing. *Mol. Phylogenet. Evol.* **19**, 177–186 (2001).
68. Taylor, W. A. Megaspore wall ultrastructure in *Isoetes*. *Am. J. Bot.* **80**, 165–171 (1993).
69. Kott, L. S. & Britton, D. M. A comparative study of spore germination of some *Isoetes* species of northeastern North America. *Can. J. Bot.* **30**, 1679–1687 (1982).
70. Oh, M. J. *et al.* High frequency sporophytes regeneration from the spore culture of the endangered aquatic fern *Isoetes coreana*. *Am. J. Plant Sci.* **04**, 14–20 (2013).
71. Caldeira, C. F. *et al.* Sporeling regeneration and ex situ growth of *Isoëtes cangae* (Isoetaceae): initial steps towards the conservation of a rare Amazonian quillwort. *Aquat. Bot.* **152**, 51–58 (2019).
72. Gamborg, O. L., Miller, R. A. & Ojima, K. Nutrient requirements of suspension cultures of soybean root cells. *Exp. Cell Res.* **50**, 151–158 (1968).
73. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
74. Kopylova, E., Noe, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).
75. Nikolenko, S. I., Korobeynikov, A. I. & Alekseyev, M. A. BayesHammer: Bayesian clustering for error correction in single-cell sequencing. *BMC Genomics* **14**, S7 (2013).
76. Butler, J. *et al.* ALLPATHS: de novo assembly of whole-genome shotgun microreads. *Genome Res.* **18**, 810–820 (2008).
77. Simpson, J. T. & Durbin, R. Efficient de novo assembly of large genomes using compressed data structures. *Genome Res.* **22**, 549–556 (2012).
78. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578–579 (2011).
79. Huang, X. & Madan, A. CAP3: a DNA sequence assembly program. *Genome Res.* **9**, 868–877 (1999).
80. Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).
81. Tang, S., Lomsadze, A. & Borodovsky, M. Identification of protein coding regions in RNA transcripts. *Nucleic Acids Res.* **43**, 1–10 (2015).
82. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **11**, 119 (2010).
83. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
84. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
85. Hammer, Ø., Harper, D. A. T. & Ryan, P. D. Paleontological statistics software package for education and data analysis. *Palaeontol. Electron.* **4**, 1–9 (2001).
86. Katoh, K. & Frith, M. C. Adding unaligned sequences into an existing alignment using MAFFT and LAST. *Bioinformatics* **28**, 3144–3146 (2012).
87. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
88. Hall, T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**, 95–98 (1999).

Acknowledgements

Author contributions

A.J.H. and L.D. designed the project. A.J.H. carried out the analyses with guidance from S.K. and D.E. A.J.H. and L.D. wrote the paper with constructive comments from S.K. and D.E. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-78171-y.

Correspondence and requests for materials should be addressed to L.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.