bioRxiv preprint doi: https://doi.org/10.1101/2019.12.16.878298; this version posted December 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

3 **Authors:** Alexander J. Hetherington[1], David M. Emms[1], Steven Kelly[1] & Liam Dolan[1*]

4 **Author Affiliations:**

5 [1]Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK

6 **Corresponding Author:**
7 liam.dolan@plants.ox.ac.uk
8

9 **Keywords:**

10 Evolution; palaeobotany; roots; *Isoetes*; *Selaginella*; arborescent lycopsids; rhizomorph

11

12 **Abstract:**

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

27

28 ## Introduction

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

37 Stigmarian rootlets were interpreted as true roots by the majority of authors until the mid
38 20th century[5,6,10–14]. However, a suite of fossil findings in the second half of the 20th century, including

bioRxiv preprint doi: https://doi.org/10.1101/2019.12.16.878298; this version posted December 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

39 fossil embryos, rhizomorph apices and the abscission of rootlets[3,4,15–19] led to the revival of the
40 modified shoot hypothesis first suggested in 1872, which interpreted rootlets as modified leaves[7]..
41 Given that all rhizomorphic lycopsids (sensu[20–23]) form a monophyletic group, and that extinct
42 stigmarian rootlets were interpreted as modified leaves this suggested that the rootlets of all
43 rhizomorphic lycopsids were modified leaves, including the rootlets of extant *Isoetes*[3]. The
44 interpretation that the rootlets of extant *Isoetes* species were modified leaves was strikingly at odds
45 with all previous descriptions of *Isoetes* rootlets that had always been interpreted as roots similar to
46 the roots of other extant lycopsids[11,24–32].

47 New evidence that is inconsistent with the modified shoot hypothesis has been reported
48 since the seminal paper by Rothwell and Erwin[3]. First, the modified shoot hypothesis posits that the
49 ancestral embryo condition in the rhizomorphic lycopsids lacked an embryonic root, but instead
50 developed a single shoot axis that divided to give a typical shoot and modified rooting shoot axis
51 that developed modified leaves (rootlets). However, embryo development in the early diverging
52 rhizomorphic lycopsid, *Oxroadia* developed an embryonic root[20]. Therefore, the embryo of *Oxroadia*
53 does not support the hypothesis that a branching event in the embryo produced a rooting shoot axis
54 (rhizomorph) that developed root-like leaves (rootlets). Second, while the leaves of all plants species
55 develop exogenously[33], in a process that includes the outer-most layers of the shoot, roots of extant
56 *Isoetes* originate endogenously[34]. Therefore, the endogenous development of rootlets is inconsistent
57 with their interpretation as modified leaves[34]. Third, the discovery of the development of root hairs
58 on rootlets of extinct rhizomorphic lycopsids that are identical to the root hairs that develop on
59 extant lycopsids suggest that rootlets are root-like[2]. Together these three studies present an
60 emerging body of evidence that is incompatible with the modified shoot hypothesis.

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

70

71 **Results**

72 Development of a protocol to propagate *Isoetes echinospora* in axenic culture

73 To define gene expression signatures in the organs of *I. echinospora,* a population of plants was
74 collected from the wild (Fig. 1A) and protocols to grow the plants in axenic culture were developed.
75 The collected plants were grown in the green house and male spores (microspores) and female
76 spores (megaspores) were produced and then isolated. Megaspores and microspores were surface
77 sterilised and germinated together in sterile liquid media to generate a population of sporophytes in
78 axenic culture (Fig. 1B). Sporophytes were transferred to solid media three months after
79 germination (Fig. 1C). A population of c. 50 *I. echinospora* plants were grown for approximately four
80 months to a stage where plants were large enough to extract RNA from the three major organs;
81 leaves, corm and rootlets (Fig. 2A).

82

bioRxiv preprint doi: https://doi.org/10.1101/2019.12.16.878298; this version posted December 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

83 Assembly of an *Isoetes echinospora* sporophyte transcriptome

84 A sporophyte transcriptome was generated for rootlets corms and leaves. RNA was isolated from
85 each organ and sequenced. It was difficult to extract sufficient RNA from these plants because of the
86 challenge in isolating viable spores, getting the spores to germinate, effecting fertilisation and
87 getting sporophytes to develop in axenic culture. However, we extracted 1 technical replicate of
88 rootlets and 2 technical replicates of corm and leaves. The raw reads for all samples were pooled,
89 quality checked and assembled into contiguous transcripts. The assembled transcriptome comprised
90 113,464 transcripts with a mean sequence length of 940 base pairs (bp). There were 35,564
91 sequences over 1 kilobases (Kb) in the assembly, with an N50 of 1313 bp. Proteins were successfully
92 predicted for c. 95% of the transcripts. To investigate the completeness our transcriptome we next
93 performed a BUSCO[35] analysis to investigate the number of conserved BUSCO[35] groups in our
94 transcriptome. BUSCO[35] groups are near-universal single-copy orthologs. Identifing the percentage
95 of BUSCO[35] groups present in our *de-novo* transcriptome therefore provides a metric for the
96 completeness of our transcriptome. Of the 430 total BUSCO[35] groups searched for in the
97 Viridiplantae dataset[35], 318 (74.0%) were found complete, 87 (20.2%) were found fragmented and
98 only 25 (5.8%) were missing. These metrics indicate that the transcriptome assembly was high
99 quality. We next mapped the reads extracted from each of the three different organs; leaves, corms,
100 and rootlets to calculate the abundance levels for each transcript in each of the three organs
101 (Supplementary Table S1).

102

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

104 If *I. echinospora* rootlets were modified leaves, as predicted by the modified shoot hypothesis, we
105 might expect gene expression signatures to be similar in the rootlets and leaves. To test this
106 hypothesis, we compared gene expression in rootlets, leaves and corms using a principal coordinate
107 analysis (PCoA). The two leaf replicates, two corm replicates and the single rootlet sample were
108 plotted on the first two PCoA axes (which together account for 98.6% of the variance in the sample
109 (Fig. 2B)). The three tissue types are clearly distinct and separated in gene expression space. The first
110 PCoA axis accounts for 75.4% of the variance in gene expression and it distinguishes leaves and
111 corms from rootlets (Fig. 2B). The second PCoA axis accounts for 23.2% of the variance in gene
112 expression and distinguishes all three tissues from each other (Fig. 2B). The PCoA indicated that gene
113 expression profiles of rootlets and leaves are distinct and does not support the hypothesis that *I.*
114 *echinospora* rootlets are modified leaves.

115

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

118 If the rootlets of *I. echinospora* are true roots we expected similarities in gene expression between
119 rootlets and true roots of other land plant species such as the lycophytes *Selaginella moellendorffii*
120 and the seed plant *Arabidopsis thaliana*. To compare gene expression between these species we first
121 defined orthologous relationships between the genes of the three species using the OrthoFinder
122 software[36,37]. This analysis identified 1,737 single copy orthologs in common between these species.
123 Using these 1,737 orthologs we compared gene expression between the different species. We
124 compared average gene expression between *I. echinospora* rootlets and leaves (this study) with the
125 published gene expression in roots and leaves of *S. moellendorffii* [38] and roots and "aerial parts" of
126 *Arabidopsis thaliana* (based on EMBL-EBI accession E-GEOD-53197). To compare gene expression

bioRxiv preprint doi: https://doi.org/10.1101/2019.12.16.878298; this version posted December 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

127 between these different species and organs we subjected the gene expression dataset to a PCoA.
128 The first three principal coordinates accounted for 95.7% of the variance in the dataset. Axis 1
129 accounted for 43.6% of the variance and separated the samples by species (Fig. 3A, B). Axis 2,
130 accounted for 35.9% of the variance and distinguished the two lycophyte transcriptomes (*I.*
131 *echinospora* and *S. moellendorffii*) from that of the seed plant *A. thaliana* (Fig. 3A, C). PCoA axes one
132 and two therefore indicate that the majority of the differences in gene expression is accounted for
133 by differences between species rather than between roots and leaves. PCoA axis 3 accounted for
134 16.2% of the variance and distinguished between leaves and roots in all species (Fig. 3B, C). Leaf
135 samples clustered in the positive values and root samples clustered in the negative values of PCoA
136 axis 3 (Fig. 3B, C). The clustering of the *I. echinospora* rootlet sample with both the roots of *S.*
137 *moellendorffii* and *A. thaliana* on axis 3 (Fig. 3B, C) indicates that the gene expression signature of
138 the rootlets of *I. echinospora* is similar to the the gene expression signature of both *S. moellendorffii*
139 and *A. thaliana*. These gene expression data are consistent with the hypothesis that rootlets of *I.*
140 *echinospora* are roots.

141

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

143 To verify our findings that gene expression of *I. echinospora* rootlets were similar to those of the
144 true roots of *S. moellendorffii* and *A. thaliana* we next determined the expression of the root-specific
145 *ROOT HAIR DEFECTIVE SIX-LIKE* (*RSL*) genes in *I. echinospora. ROOT HAIR DEFECTIVE SIX-LIKE* (*RSL*)
146 genes positively regulate the development of root hairs in euphyllophytes including *A. thaliana*[39–42]
147 and are expressed in *S. moellendorffii* roots[43,44]. *RSL* genes are markers for vascular plant roots
148 because they are expressed at a much higher level in roots of *A. thaliana* (EMBL-EBI accession E-
149 GEOD-53197) and *S. moellendorffii*[45] than in leaves and shoots (Supplementary Table S1). We
150 searched the *I. echinospora* transcriptome for RSL genes using the BLAST algorithm with RSL-specific
151 queries. RSL sequences were identified in the *I. echinospora* transcriptome. A gene tree was
152 generated and defined four *I. echinospora* RSL genes in two monophyletic groups (Fig. 4). There
153 were three transcripts in the RSL Class I clade (106204; 101034; 092963) and a single transcript in
154 the RSL Class II clade (095243). Average expression of the four RSL genes in rootlets was 4.24
155 transcripts per million (TPM) (Fig. 4). The average root expression was 5.78 TPM for the six RSL genes
156 of *A. thaliana* (EMBL-EBI accession E-GEOD-53197), demonstrating similarities in expression of RSL
157 genes between in *I. echinospora* and *A. thaliana*. In *I. echinospora*, expression of each RSL Class I
158 transcript was higher in rootlets than in leaves (Fig. 4). Furthermore, the single *I. echinospora* Class II
159 RSL gene transcript (095243) was expressed in rootlets and no expression was detected in the corm
160 or leaves. These data indicate that RSL genes are preferentially expressed in the *I. echinospora*
161 rootlets and not in leaf tissue, as in they are in *S. moellendorffii* of *A. thaliana* (Supplementary Table
162 S1). These data are consistent with the hypothesis that *I. echinospora* rootlets are roots.

163 To verify that RSL genes are markers of vascular plant roots we investigated the *RSL* genes in
164 *Azolla filiculoides,* a fern that develops roots with root hairs[30,46], and *Salvinia cucullata* a fern that has
165 secondarily lost roots with root hairs and instead modified leaves perform rooting functions[30,47,48].
166 We searched the *S. cucullata* and *A. filiculoides* genomes and proteomes[49] for RSL genes using the
167 BLAST algorithm with RSL-specific queries. A gene tree was constructed with the retrieved
168 sequences and allowed us to identify 3 RSL Class I genes and a single RSL Class II gene in the *A.*
169 *filiculoides* genome (Fig. 4, Supplementary Fig. S1). Consistant with their role in root development in
170 *A. filiculoides* the RSL genes were expressed in the roots[46]. However, there were no RSL genes in the
171 *S. cucullata* genome. *S. cucullate* sequences were identified in closely related basic-helix-loop-helix
172 transcription factor subfamily XI[50,51] but none were identified in RSL clade (Fig. 4, Supplementary Fig.

bioRxiv preprint doi: https://doi.org/10.1101/2019.12.16.878298; this version posted December 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

173 S1). We conclude that the loss of RSL genes accompanied the evolutionary loss of roots with root
174 hairs in *Salvinia cucullate*, which is consistent with RSL genes being markers of vascular plant roots.
175 Furthermore, if the rootlets of *I. echinospora* were modified leaves, similar to the root-like modified
176 leaves of *S. cucullate,* we might have expected that RSL genes would have also been lost from the *I.*
177 *echinospora* genome. Instead, RSL genes are preferentially expressed in *I. echinospora* rootlets just
178 as they are in *S. moellendorffii* roots. These data are consistent with the hypothesis that the *I.*
179 *echinospora* rootlet is a root and not a modified leaf.

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

185

186 **Discussion**

187 The homology of the rootlets of both extinct and extant rhizomorphic lycopsids have been
188 contentious for the past 150 years, with two competing hypotheses. The first, interprets the rootlets
189 as true roots similar to the roots of other lycopsids. The second, interprets rootlets as modified
190 leaves. Despite the second hypothesis that posits that rootlets are modified leaves being widely
191 accepted over the past 30 years[1,3,52] there is a growing body of evidence[2,34] that suggests that
192 rootlets should be interpreted as true roots. Here we report the *de novo* transcriptome of *I.*
193 *echinospora* that we used to test predictions of the two competing hypotheses. We discovered that
194 expression profiles in *I. echinospora* rootlets and leaves were different. We showed that gene
195 expression profiles of *I. echinospora* rootlets and *S. moellendorffii* and *A. thaliana* roots were similar.
196 Finally, RSL genes involved in root cell differentiation are preferentially expressed in *I. echinospora*
197 rootlets as they are in *S. moellendorffii* roots and the roots of euphyllophytes (A. thaliana, *Oryza*
198 *sativa* and *Brachypodium distachyon*[39–42]). Taking these three pieces of evidence together, we
199 conclude that *Isoetes* rootlets are true roots, like those of extinct and extant lycopsids and not
200 modified leaves.

201 The new evidence presented here adds to the growing and extensive list of similarities between
202 the rootlets of rhizomorphic lycopsids – *Isoetes* species and extinct taxa such as *Stigmaria* – and the
203 roots of other lycopsids[2,20,34]. This growing body of evidence supports the hypothesis that rootlets
204 are roots and not modified leaves. The rootlets of the rhizomorphic lycopsids and roots of all extant
205 lycopsids are indeterminate radially symmetric axes that branch by isotomous dichotomy, develop
206 endogenously within specialised structures, develop a root meristem with root cap and produce root
207 hairs[31,32,34,53]. If the modified shoot hypothesis were correct it would have required the direct
208 modification of a determinate leaf that did not branch, developed exogenously and was
209 characterised by a ligule, stomata and dorsiventral symmetry into a rootlet. Each of these leaf
210 characters would have had to be lost and all of the rootlet characters, which are shared among the
211 lycopsids, would have had to evolve independently. By contrast, if the lycopsid root hypothesis is
212 correct and rootlets are roots then there is no requirement for this large suite of character state
213 changes. Instead, the only character transitions required to account for rootlet character states were
214 the collateral positioning of the phloem, the regular rhizotaxy and rootlet abscission[4]. Although
215 these three characters (collateral positioning of the phloem, the regular rhizotaxy and rootlet
216 abscission) are predominately leaf characters, they are not exclusive to leaves; each has been
217 described in the roots of other species of land plants. The collateral position of the phloem is found

bioRxiv preprint doi: https://doi.org/10.1101/2019.12.16.878298; this version posted December 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

218 in Lycopodium roots including, *Lycopodium lucidulum*, *Lycopodium clavalum*, *Lycopodium obscurum*
219 and *Lycopodium complanalum*[54], regular rhizotaxy develops in *Ceratoptertis thalictroides*, *Cucurbita*
220 *maxima* and *Pontederia cordata*[55–57] and roots abscise in *Oxalis esculenta*, *Abies balsamea*, *Pinus*
221 *strobus*, *Tsuga canadensis* and *Azolla* species[58–63]. Based on character transitions alone we suggest
222 that the hypothesis that rootlets of the rhizomorphic lycopsids are roots, similar to other lycopsid
223 roots, is a more parsimonious hypothesis than interpreting rootlets as modified leaves.

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

230 The gene expression data from the *de novo I. echinospora* transcriptome are consistent with the
231 hypothesis that the rootlets of the rhizomorphic lycopsids are roots and not modified leaves. We
232 therefore interpret the rootlets of the rhizomophic lycopsids as roots developing from a unique root
233 bearing organ; the rhizomorph[21,53,64]. This conclusion suggests that the dichotomously branching
234 rooting axis is conserved among all lycopsids and a distinguishing character of the group. The
235 dichotomous branching of these rooting axes has been conserved for over 400 million years and our
236 comparative transcriptomic analysis suggest that the RSL genes function during root development in
237 *Selaginella* and *Isoetes* has been conserved since these species shared a common ancestor at least
238 375 million years ago[65]. Our comparative analysis of the transcriptomes of extant lycophytes support
239 the hypothesis that the rooting systems of extant *Isoetes* species and their extinct giant ancestors
240 are homologous. The data also suggest that the development of the large rooting systems of the
241 lycopsid trees that were an important component of the Palaeozoic flora and played a key role in
242 changing the Earth's Carbon Cycle were controlled by the same genes that regulate root
243 development in their extant herbaceous descendants.

244

245 **Materials and Methods**

246 Plant collection and growth

247 Mature *I. echinospora* plants were collected from Loch Aisir and Loch Dubhaird Mor in September
248 2013 and 2014 from North West Sutherland (Scotland, UK) with the permission of the John Muir
249 Trust and the Scourie Estate. *I. echinospora* plants were identified on the basis of their echinate
250 megaspore ornamentation[66]. Mature *I. echinospora* plants were grown submerged in aquaria in
251 Levington M2 compost topped with coarse gravel in a glasshouse at Oxford University at 18°C under
252 a 16 h light : 8 h dark photoperiod.

253

254 Growth of *I. echinospora* in axenic culture

255 RNA was extracted from plants grown in axenic culture to ensure that there was no RNA
256 contamination from other organisms. A procedure was developed to surface sterilise *I. echinospora*
257 spores and germinate a population of axenically grown plants, based on previously developed
258 procedures[67–69]. Sporophylls were removed from the mature plant population growing in aquaria in
259 September (2013 and 2014) when sporangia were mature[67]. Using forceps (under a Leica M165 FC

bioRxiv preprint doi: https://doi.org/10.1101/2019.12.16.878298; this version posted December 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

260 stereo microscope) mega- and micro-sporangia were isolated from sporophylls. Intact sporangia
261 were washed in 1% (v/v) sodium dichloroisocyanurate (NaDCC) for 5 min. Sporangia were broken
262 and loose spores were washed in 0.1% NaDCC for a further 5 min. Following the NaDCC washes,
263 loose spores were rinsed for 5 min three times in ddH2O. Microspores were centrifuged for 5 min at
264 5000 rpm between washes). Once sterilised, mega and micro-spores were mixed together in ddH_2O
265 in a Petri dish. Petri dishes were sealed with parafilm, and incubated in darkness at 4^oC for 2 wk.
266 After 2 wk, Petri dishes were moved to a 16 h light : 8 h dark photoperiod at 18^oC. Approximately
267 30% of surface sterilised megasporangia contained megaspores that germinated, and within these
268 megasporangia c. 25% of the total megaspore population germinated. It was possible to identify
269 germinating megaspores because cracking of the megaspore wall was visible and the presence of
270 archegonia on the megagametophyte. Once fertilisation occurred, developing sporophytes were
271 identified by the presence of the first leaf. Sporophytes were left to continue to grow in ddH_2O
272 water until the two leaf two rootlet stage when they were moved to magenta boxes containing; ½
273 Gamborg's medium[70], supplemented with 1% phytogel (Sigma). Plants were embedded in Gamborg
274 media and submerged in liquid Bold's Basal Medium (Sigma, UK).

275

276 RNA extraction and sequencing

277 Total RNA was extracted from root, corm and leaf tissues from c. 50 *I. echinospora* plants. Total RNA
278 from leaves (two independent replicates), corm (two independent replicates) and rootlets (one
279 replicate) was extracted with the RNeasy plant mini kit (Qiagen). On-column DNase I treatment was
280 performed with RNase-free DNase I (Qiagen), according to the manufacturer's instructions. cDNA
281 was synthesised with ProtoScript II reverse transcriptase (New England Biolabs) according to the
282 manufacturer's instructions, using oligo(dT) primer. Total cDNA samples were quantified with a
283 Nanodrop ND-1000 spectrophotometer. RNA purity and quality were checked with an Agilent 2100
284 Bioanalyzer. cDNA was sequenced by the High-Throughput Genomics Group at the Wellcome Trust
285 Centre for Human Genetics, University of Oxford. Sequencing resulted in 195,072,304 paired end
286 reads separated into five samples: 2 leaves samples (35,718,157; 35,555,048 paired end reads), 2
287 corm samples (38,728,989; 44,379,751 paired end reads) and one rootlet sample (40,690,359 paired
288 end reads). The raw read libraries have been deposited under SRP135936 on the NCBI Sequence
289 Read Archive.

290

291 *De novo* transcriptome assembly, protein predictions and expression analysis

292 Raw reads were quality trimmed using Trimmomatic-0.32[71], to remove remaining Illumina adaptors
293 and low quality tails. Ribosomal RNA was filtered out using Sortmerna-1.9[72] and error corrected
294 using BayesHammer (SPAdes-16 3.5.0)[73] (with setting --only-error-correction) and Allpaths-LG-4832[74]
295 (with setting PAIRED_SEP=-20 and ploidy = 2). Reads were normalised using Khmer-0.7.1 with a
296 khmer size of 21. Before assembly, paired end reads were stitched together using Allpaths-LG-
297 4832[74]. A *de novo* transcriptome assembly was made with the cleaned, stitched reads using SGA[75],
298 SSPACE-v3[76], and CAP3[77]. Finally assembled scaffolds were corrected using Pilon-1.6[78]. The
299 Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the
300 accession GGKY00000000. The version described in this paper is the first version, GGKY01000000.
301 Proteins were predicted from the *de novo* transcriptome assembly using GeneMarkS-T[79], Prodigal[80]

bioRxiv preprint doi: https://doi.org/10.1101/2019.12.16.878298; this version posted December 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

302 and Transdecoder (part of the Trinity assembly program[81]), proteins were deposited on Zenodo
303 (http://doi.org/10.5281/zenodo.3574570). A BUSCO analysis[35], using BUSCO 3.1.0 and the
304 *viridiplantae_odb10* database.

305

306 Comparison of gene expression between *I. echinospora* organs

307 Using the sporophyte transcriptome assembly we next mapped the reads from the three organ
308 libraries – leaves, corms, and rootlets – to the transcriptome to measure the expression levels of
309 each transcript in the three tissues using Salmon[82]. To investigate the similarities between gene
310 expression in the different organs we carried out a PCoA on the three organ types. Euclidean
311 distances were derived from the expression of all transcripts (TPM) in each organ and were
312 subjected to PCoA in PAST[83] using a transformation exponent of 2.

313

314 OrthoFinder analysis and comparison of gene expression between *I. echinospora, S.*
315 *moellendorffii* and *Arabidopsis thaliana*

316 Orthologous relationships between *I. echinospora*, *S. moellendorffii* and *A. thaliana* proteins were
317 determined using OrthoFinder[36,37]. OrthoFinder was run with *I. echinospora* proteins and protein
318 datasets for 57 species from Phytozome (full list of species in Supplementary Table S2) including the
319 Rhodophyta *Porphyra umbilicalis*, seven species of chlorophytes, the bryophytes *Marchantia*
320 *polymorpha* and *Physcomitrella patens*, the lycophytes *Selaginella moellendorffii* and 46 angiosperm
321 species. This analysis resulted in the identification of 38,217 orthogroups, accounting for 82.6% of all
322 genes included in the analysis (The results of the OrthoFinder analysis were we deposited on
323 Zenodo, http://doi.org/10.5281/zenodo.3574570).

324 To compare gene expression between *I. echinospora*, *S. moellendorffii* and *A. thaliana* we
325 identified single copy orthologs between these species based on the OrthoFinder[36,37] analysis. In
326 total, 1,737 single copy orthologs were found between the three species. Using these 1,737
327 orthologs we contrasted gene expression between the different species. We investigated average
328 genes expression between *I. echinospora* rootlets and leaves (this study) with the published average
329 gene expression between roots and leaves of *S. moellendorffii*[38] and *A. thaliana*. *A. thaliana* gene
330 expression was based on average gene expression in "aerial part" and "root" of 17 different natural
331 accessions (EMBL-EBI accession E-GEOD-53197). To investigate similarities in gene expression
332 between these 1,737 orthologs we carried out a PCoA in PAST[83]. Euclidean distances were derived
333 from the Log10 transformed gene expression of the 1,737 orthologs (Supplementary Table S1).
334 Euclidean distances were subjected to a PCoA in PAST[83], using a transformation exponent of 4.

335

336 Phylogenetic Analyses

337 Phylogenetic analyses were carried out on the RSL genes. Blast queries were assembled based on
338 previously published gene trees of RSL genes[50]. Sequences were used to blast the protein databases
339 of the; *Marchantia polymorpha* "primary" (proteins) (version 3.1, November, 2015), *Physcomitrella*
340 *patens* "primary" (proteins) (version 3.0, January 12, 2014), *Selaginella moellendorffii* "primary"
341 (proteins) (version 1.0, January 12, 2014), *Amborella trichopoda* (proteins) (version 1.0, 2013) and
342 *Arabidopsis thaliana* "primary" (proteins) (TAIR10) on the http://marchantia.info/blast/ server. Two
343 fern protein databases were also searched; *Azolla filiculoides* protein v1.1 and *Salvinia cucullata*

bioRxiv preprint doi: https://doi.org/10.1101/2019.12.16.878298; this version posted December 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

344 proteins v1.2 [49] as well as the predicted proteins from the *I. echinospora* transcriptome generated in
345 this study. All proteins were aligned in MAFFT[84,85], manually edited in Bioedit[86]. Maximum likelihood
346 gene trees were generated in PhyML 3.0[87], using Jones, Taylor and Thorton (JTT) amino acid
347 substitution model. To verify the absence of *RSL* genes in the *S. cucullate* the genomes and
348 proteomes of *A. filiculoides* and *S. cucullate* were searched by Blast using the *A. thaliana* protein
349 sequence RSL1 (AT5G37800) using an E-value cut off of 1E-15. A gene tree was generated as
350 described above including the addition of *A. thaliana* protein sequence from subfamilies VIIIb and
351 XI[50,51] (Fasta alignments files for both gene trees we deposited on Zenodo,
352 http://doi.org/10.5281/zenodo.3574570).

353

354

355 **Acknowledgements**

356 This research was supported by a Biotechnology and Biological Sciences Research Council (Grant
357 BB/J014427/1) Doctoral Training Partnership Scholarship and the George Grosvenor Freeman
358 Fellowship by Examination in Sciences, Magdalen College (Oxford) to A.J.H. L.D. was funded by a
359 European Research Council Advanced Grant (EVO500, contract 250284), European Commission
360 Framework 7 Initial Training Network (PLANTORIGINS, project identifier 238640) and European
361 Research Council Grant (De NOVO-P, contract 787613). S.K. and D.E. were funded by a European
362 Union's Horizon 2020 research and innovation program under grant agreement number 637765 to
363 S.K. S.K. is a Royal Society University Research Fellow. We thank the High-Throughput Genomics
364 Group at the Wellcome Trust Centre for Human Genetics (funded by Wellcome Trust grant reference
365 090532/Z/09/Z) for the generation of sequencing data. We are grateful to the John Muir Trust and
366 the Scourie Estate for providing permission to collect *I. echinospora*, and to Dr Heather McHaffie
367 (Royal Botanic Garden Edinburgh) for assistance identifying *I. echinospora*.

368

369 **Competing interests**

370 The authors declare no competing interests.

371

372 **Data availability**

373 The raw read libraries have been deposited under SRP135936 on the NCBI Sequence Read Archive.
374 Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the
375 accession GGKY00000000. The orthofinder analysis, predicted protein sequences in the I.
376 echinopsora transcriptome, and fasta alignment files for gene trees were deposited at Zenodo,
377 http://doi.org/10.5281/zenodo.3574570.

378

379 **References**

380 1. Stewart, W. & Rothwell, G. W. *Paleobotany and the evolution of plants*. (Cambridge
381 University Press, 1993).

382 2. Hetherington, A. J., Berry, C. M. & Dolan, L. Networks of highly branched stigmarian rootlets
383 developed on the first giant trees. *Proc. Natl. Acad. Sci.* **113**, 6695–6700 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2019.12.16.878298; this version posted December 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

384 3. Rothwell, G. W. & Erwin, D. The rhizomorph apex of *Paurodendron*: implications for
385 homologies among the origins of Lycopsida. *Am. J. Bot.* **72**, 86–98 (1985).

386 4. Karrfalt, E. E. A further comparison of *Isoetes* roots and stigmarian appendages. *Can. J. Bot.*
387 **58**, 2318–2322 (1980).

388 5. Williamson, W. A monograph on the morphology and histology of *Stigmaria ficoides*.
389 *Palaeontogr. Soc.* **40**, 1–62 (1887).

390 6. Weiss, F. E. A re-examination of the stigmarian problem. *Proc. Linn. Soc. London* **144**, 151–
391 166 (1933).

392 7. Schimper, W. P. *Traité de paléontologie végétale ou la flore du monde primitive dans ses*
393 *rapports avec les formations géologiques et la flore du monde actuel*. (J. B. Baillière et Fils,
394 1872).

395 8. Solms-Laubach, H. G. *Fossil Botany being an introduction to palaeophytology from the*
396 *stanpoint of the botanist. Translated into English by Garnsey, H. E and Balfour, I. B*.
397 (Clarendon Press, Oxford (UK), 1891).

398 9. Stewart, W. A comparative study of stigmarian appendages and *Isoetes* roots. *Am. J. Bot.* **34**,
399 315–324 (1947).

400 10. Lang, W. H. On the apparently endogenous insertion of the roots of *Stigmaria*. *Mem. Proc.*
401 *Manchester lierary Philos. Soc.* **67**, 101–107 (1923).

402 11. Bower, F. O. *The origin of a land flora*. (Macmillan and Co., Limited, 1908).

403 12. Leclercq, S. A monograph of *Stigmaria bacupensis*, Scott et Lang. *Ann. Bot.* **44**, 31–54 (1930).

404 13. Weiss, F. E. Morphology of *Stigmaria* and of its appendages in comparison with recent
405 lycopodiales. *Proc. Linn. Soc. London* **120**, 74–75 (1908).

406 14. Scott, D. H. *Studies in Fossil Botany*. (Adam and Charles Black, London, UK, 1900).

407 15. Rothwell, G. W. The apex of *Stigmaria* (Lycopsida), rooting organ of Lepidodendrales. *Am. J.*
408 *Bot.* **71**, 1031–1034 (1984).

409 16. Stubblefield, S. P. & Rothwell, G. W. Embryogeny and reproductive biology of
410 *Bothrodendrostrobus mundus* (Lycopsida). *Am. J. Bot.* **68**, 625–634 (1981).

411 17. Phillips, T. L. Reproduction of heterosporous arborescent lycopods in the Mississippian–
412 Pennsylvanian of Euramerica. *Rev. Palaeobot.* **27**, 239–289 (1979).

413 18. Frankenberg, J. M. & Eggert, D. A. Petrified Stigmaria from North America. I. *Stigmaria*
414 *ficoides*, the underground portions of *Lepidodendraceae*. *Palaeontogr. B,* **128**, 1–47 (1969).

415 19. Eggert, D. A. Petrified *Stigmaria* of Sigillarian origin from North America. *Rev. Palaeobot.*
416 *Palynol.* **14**, 85–99 (1972).

417 20. Bateman, R. M. Morphometric reconstruction, palaeobiology and phylogeny of *Oxroadia*
418 *gracilis* Alvin emend. and *O. conferta* sp. nov.: anatomically-preserved rhizomorphic lycopsids
419 from the Dinantian of Oxroad Bay, SE Scotland. *Palaeontogr. Abteilung B* **228**, 29–103 (1992).

420 21. Bateman, R. M., DiMichele, W. A. & Willard, D. A. Experimental cladistic analysis of
421 anatomically preserved arborescent Lycopsids from the Carboniferous or Euramerica: an
422 essay on paleobotanical phylogenetics. *Ann. Missouri Bot. Gard.* **79**, 500–559 (1992).

423 22. DiMichele, W. A. & Bateman, R. M. The rhizomorphic lycopsids: a case-study in

bioRxiv preprint doi: https://doi.org/10.1101/2019.12.16.878298; this version posted December 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

424 paleobotanical classification. *Syst. Bot.* **21**, 535–552 (1996).

425 23. Kenrick, P. & Crane, P. R. *The origin and early diversification of land plants: a cladistic study*.
426 (Smithsonian Series in Comparative Evolutionary Biology. Washington, DC, USA: Smithsonian
427 Institute Press, 1997).

428 24. Hofmeister, W. F. B. *On the Germination, Development, and Fructification of the Higher*
429 *Cryptogamia: And on the Fructification of the Coniferæ. Translated to English by Frederick*
430 *Currey*. (1862).

431 25. West, C. & Takeda, H. X. On *Isoetes japonica*, A. Br. *Trans. Linn. Soc. London. 2nd Ser. Bot.* **8**,
432 333–376 (1915).

433 26. Van Tieghem, P. & Douliot, H. Recherches comparatives sur l'origine des membres endogenes
434 dans les plantes vasculaires. *Ann. des Sci. Nat. Bot. Paris* **8**, 1–656 (1888).

435 27. Sachs, J. *Text-book of botany, morphological and physiological. Translated by A. W. Bennett*
436 *and W. T. Thiselton Dyer*. (Clarendon Press, Oxford (UK), 1875).

437 28. Foster, A. S. & Gifford, E. M. *Comparative morphology of vascular plants*. (W. H. Freeman and
438 Company, 1959).

439 29. Guttenberg, H. V. *Histogenese der Pteridophyten. Handbuch der Pflanzenanatomie vol. VII. 2.*
440 (Gebrüder Borntraeger, 1966).

441 30. Eames, A. J. *Morphology of vascular plants, lower groups*. (MCGraw-Hill Book Company, Inc,
442 1936).

443 31. Ogura, Y. *Comparative anatomy of vegetative organs of the pteridophytes. Handb*
444 *Pflanzenanat*. (Gebrüder Borntraeger, 1972).

445 32. Bierhorst, D. W. *Morphology of vascular plants*. (Macmillan, 1971).

446 33. Charlton, W. A. & Watson, J. Patterns of arrangement of lateral appendages on axes of
447 *Stigmaria ficoides* (Sternberg) Brongniart. *Bot. J. Linn. Soc.* **84**, 209–221 (1982).

448 34. Yi, S. & Kato, M. Basal meristem and root development in *Isoetes asiatica* and *Isoetes*
449 *japonica*. *Int. J. Plant Sci.* **162**, 1225–1235 (2001).

450 35. Waterhouse, R. M. *et al.* BUSCO Applications from Quality Assessments to Gene Prediction
451 and Phylogenomics. *Mol. Biol. Evol.* **35**, 543–548 (2018).

452 36. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome
453 comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157
454 (2015).

455 37. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative
456 genomics. *Genome Biol.* **20**, 238 (2019).

457 38. Mello, A., Efroni, I., Rahni, R. & Birnbaum, K. D. The *Selaginella* rhizophore has a unique
458 transcriptional identity compared with root and shoot meristems. *New Phytol.* **222**, 882–894
459 (2019).

460 39. Kim, C. M. & Dolan, L. ROOT HAIR DEFECTIVE SIX-LIKE Class I Genes Promote Root Hair
461 Development in the Grass *Brachypodium distachyon. PLOS Genet.* **12**, e1006211 (2016).

462 40. Masucci, J. D. & Schiefelbein, J. W. The rhd6 Mutation of *Arabidopsis thaliana* Alters Root-
463 Hair Initiation through an Auxin- and Ethylene-Associated Process. *Plant Physiol.* **106**, 1335–
464 1346 (1994).

bioRxiv preprint doi: https://doi.org/10.1101/2019.12.16.878298; this version posted December 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

465 41. Menand, B. *et al.* An ancient mechanism controls the development of cells with a rooting
466 function in land plants. *Science* **316**, 1477–1480 (2007).

467 42. Kim, C. M. & Dolan, L. Root hair development involves asymmetric cell division in
468 Brachypodium distachyon and symmetric division in *Oryza sativa*. *New Phytol.* **192**, 601–610
469 (2011).

470 43. Huang, L. & Schiefelbein, J. Conserved gene expression programs in developing roots from
471 diverse plants. *Plant Cell* **27**, 2119–2132 (2015).

472 44. Huang, L., Shi, X., Wang, W., Ryu, K. H. & Schiefelbein, J. Diversification of Root Hair
473 Development Genes in Vascular Plants. *Plant Physiol.* **174**, 1697–1712 (2017).

474 45. Mello, A., Efroni, I., Rahni, R. & Birnbaum, K. D. The *Selaginella* rhizophore has a unique
475 transcriptional identity compared with root and shoot meristems. *New Phytol.* **222**, 882–894
476 (2019).

477 46. de Vries, J. *et al.* Cytokinin-induced promotion of root meristem size in the fern *Azolla*
478 supports a shoot-like origin of euphyllophyte roots. *New Phytol.* **209**, 705–720 (2016).

479 47. Croxdale, J. G. Salvinia leaves. I. Origin and early differentiation of floating and submerged
480 leaves. *Can. J. Bot.* **56**, 1982–1991 (1978).

481 48. Croxdale, J. G. *Salvinia* leaves. III. Morphogenesis of the submerged leaf. *Can. J. Bot.* **59**,
482 2065–2072 (1981).

483 49. Li, F.-W. *et al.* Fern genomes elucidate land plant evolution and cyanobacterial symbioses.
484 *Nat. Plants* **4**, 460–472 (2018).

485 50. Catarino, B., Hetherington, A. J., Emms, D. M., Kelly, S. & Dolan, L. The stepwise increase in
486 the number of transcription factor families in the Precambrian predated the diversification of
487 plants on land. *Mol. Biol. Evol.* **33**, 2815–2819 (2016).

488 51. Pires, N. & Dolan, L. Origin and diversification of basic-helix-loop-helix proteins in plants. *Mol.*
489 *Biol. Evol.* **27**, 862–74 (2010).

490 52. Taylor, E. L., Taylor, T. N. & Krings, M. *Paleobotany: the biology and evolution of fossil plants*.
491 (Academic Press, Burlington, MA, USA, 2009).

492 53. Hetherington, A. J. & Dolan, L. The evolution of lycopsid rooting structures: conservatism and
493 disparity. *New Phytol.* **215**, 538–544 (2017).

494 54. Pixley, E. Y. A study of the ontogeny of the primary xylem in the roots of *Lycopodium*. *Bot.*
495 *Gaz.* **129**, 156–160 (1968).

496 55. Mallory, T. E., Chiang, S. H., Cutter, E. G. & Gifford, E. M. Sequence and pattern of lateral root
497 formation in five selected species. *Amer. J. Bot.* **57**, 800–809 (1970).

498 56. Hou, G., Hill, J. P. & Blancaflor, E. B. Developmental anatomy and auxin response of lateral
499 root formation in *Ceratopteris richardii*. *J. Exp. Bot.* **55**, 685–693 (2004).

500 57. Charlton, W. A. Distribution of lateral roots and pattern of lateral initiation in *Pontederia*
501 *cordata* L. *Bot. Gaz.* **136**, 225–235 (1975).

502 58. Duncan, J. F. “Pull Roots” of *Oxalis Esculenta*. *Trans. Bot. Soc. Edinburgh* **29**, 192–196 (1927).

503 59. Tippett, J. T. Shedding of ephemeral roots in gymnosperms. *Can. J. Bot.* **60**, 2295–2302
504 (1982).

bioRxiv preprint doi: https://doi.org/10.1101/2019.12.16.878298; this version posted December 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

505 60. Konar, R. N. & Kapoor, R. K. Anatomical studies on *Azolla pinnata*. *Phytomorphology* **211**–**223**,
506 (1972).

507 61. Uheda, E., Nakamura, S. & Kitoh, S. Aspects of the very rapid abscission of *Azolla* branches:
508 anatomy and possible mechanism. *Int J Plant Sci* **156**, 756–763 (1995).

509 62. Fukuda, K., Yamada, Y., Miyamoto, K., Ueda, J. & Uheda, E. Separation of abscission zone cells
510 in detached *Azolla* roots depends on apoplastic pH. *J. Plant Physiol.* **170**, 18–24 (2013).

511 63. Yamada, Y., Koibuchi, M., Miyamoto, K., Ueda, J. & Uheda, E. Breakdown of middle lamella
512 pectin by OH during rapid abscission in Azolla. *Plant, Cell Environ.* **38**, 1555–1564 (2015).

513 64. Gensel, P. G. & Berry, C. M. Early lycophyte evolution. *Am. Fern J.* **91**, 74–98 (2001).

514 65. Wikström, N. & Kenrick, P. Evolution of *Lycopodiaceae* (Lycopsida): Estimating Divergence
515 Times from rbcL Gene Sequences by Use of Nonparametric Rate Smoothing. *Mol. Phylogenet.*
516 *Evol.* **19**, 177–186 (2001).

517 66. Taylor, W. A. Megaspore wall ultrastructure in *Isoetes*. *Am. J. Bot.* **80**, 165–171 (1993).

518 67. Kott, L. S. & Britton, D. M. A comparative study of spore germination of some *Isoetes* species
519 of northeastern North America. *Can. J. Bot.* **30**, 1679–1687 (1982).

520 68. Oh, M. J. *et al.* High frequency sporophytes regeneration from the spore culture of the
521 endangered aquatic fern *Isoetes coreana*. *Am. J. Plant Sci.* **04**, 14–20 (2013).

522 69. Caldeira, C. F. *et al.* Sporeling regeneration and ex situ growth of *Isoëtes cangae* (*Isoetaceae*):
523 Initial steps towards the conservation of a rare Amazonian quillwort. *Aquat. Bot.* **152**, 51–58
524 (2019).

525 70. Gamborg, O. L., Miller, R. A. & Ojima, K. Nutrient requirements of suspension cultures of
526 soybean root cells. *Exp. Cell Res.* **50**, 151–158 (1968).

527 71. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence
528 data. *Bioinformatics* **30**, 2114–2120 (2014).

529 72. Kopylova, E., Noe, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in
530 metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).

531 73. Nikolenko, S. I., Korobeynikov, A. I. & Alekseyev, M. A. BayesHammer: Bayesian clustering for
532 error correction in single-cell sequencing. *BMC Genomics* **14**, S7 (2013).

533 74. Butler, J. *et al.* ALLPATHS: de novo assembly of whole-genome shotgun microreads. *Genome*
534 *Res.* **18**, 810–20 (2008).

535 75. Simpson, J. T. & Durbin, R. Efficient de novo assembly of large genomes using compressed
536 data structures. *Genome Res.* **22**, 549–556 (2012).

537 76. Boetzer, M., Henkel, C. V, Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled
538 contigs using SSPACE. *Bioinformatics* **27**, 578–579 (2011).

539 77. Huang, X. & Madan, A. CAP3: A DNA sequence assembly program. *Genome Res.* **9**, 868–77
540 (1999).

541 78. Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection
542 and genome assembly improvement. *PLoS One* **9**, e112963 (2014).

543 79. Tang, S., Lomsadze, A. & Borodovsky, M. Identification of protein coding regions in RNA
544 transcripts. *Nucleic Acids Res.* **43**, 1–10 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/2019.12.16.878298; this version posted December 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

545 80. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site
546 identification. *BMC Bioinformatics* **11**, 119 (2010).

547 81. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a
548 reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).

549 82. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-
550 aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).

551 83. Hammer, Ø., Harper, D. A. T. & Ryan, P. D. Paleontological Statistics Software Package for
552 Education and Data Analysis. Palaeontologia Electronica. *Palaeontol. Electron.* **4**, 1–9 (2001).

553 84. Katoh, K. & Frith, M. C. Adding unaligned sequences into an existing alignment using MAFFT
554 and LAST. *Bioinformatics* **28**, 3144–3146 (2012).

555 85. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7:
556 Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

557 86. Hall, T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for
558 Windows 95/98/NT. *Nucleic Acids Symposium Series* vol. 41 95–98 (1999).

559 87. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies:
560 Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).

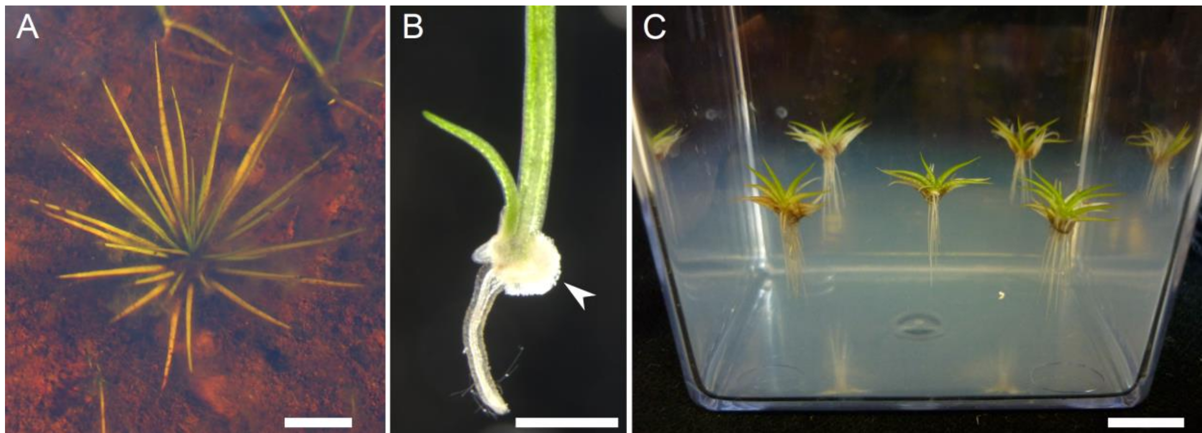
561

562

563

bioRxiv preprint doi: https://doi.org/10.1101/2019.12.16.878298; this version posted December 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

564 **Text Figures**

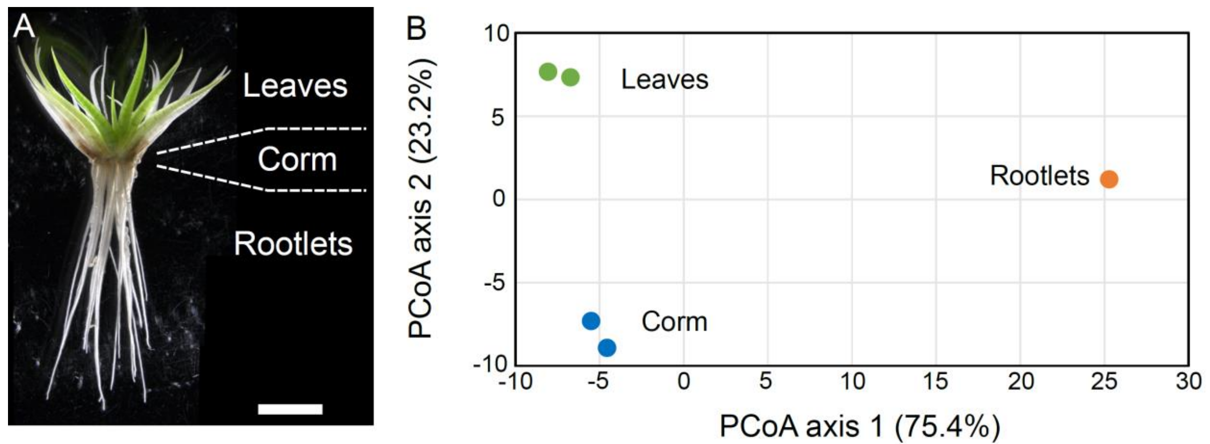


565

566 Fig. 1 Growth of *I. echinospora* in axenic culture for RNA extraction. (*A*) *I. echinospora* growing
567 submerged in its natural habitat, North West Sutherland (Scotland, UK). (*B*) Developing sporophyte
568 emerging from the megaspore (highlighted with arrowhead) at the two leaf-, two rootlet- stage (*C*) *I.*
569 *echinospora* sporophytes growing submerged on transparent solid media. Scale bars: A,
570 approximately 2 cm; B, 1 mm; C, 1 cm.

571

bioRxiv preprint doi: https://doi.org/10.1101/2019.12.16.878298; this version posted December 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



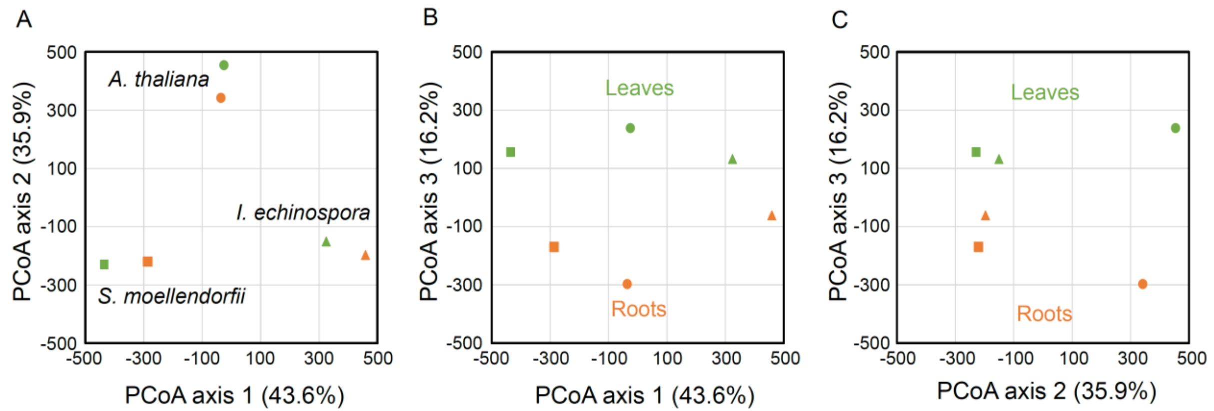
572

573 Fig. 2 Gene expression profiles of leaves and rootlets are different in *I. echinospora*. (A) *I.*
574 *echinospora* sporophyte at the stage when RNA was extracted from rootlets, corms and leaves. Scale
575 bar 5 mm. Comparison of gene expression profiles by principle coordinate analysis (PCoA) in the
576 transcriptome of *I. echinospora*. Two technical replicates of gene expression profiles of the corm and
577 leaves and single replicate of rootlets of *I. echinospora.* Two leaf replicates, green. Two corm
578 replicates, blue. Single root replicate, orange. Values on PCoA axes are shown in thousands. Values
579 in brackets on each axis describe the percentage of total variance accounted for by each axis.

580

bioRxiv preprint doi: https://doi.org/10.1101/2019.12.16.878298; this version posted December 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

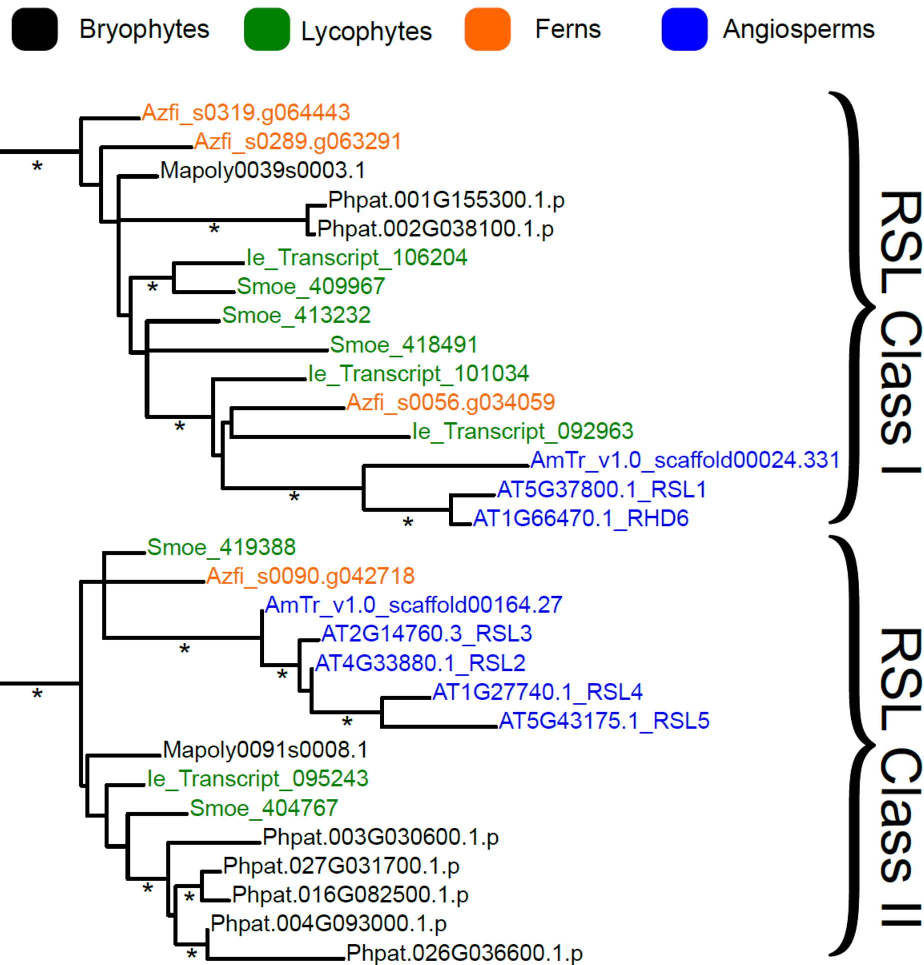
581



582 Fig. 3 Comparison of root and leaf transcriptomes of *Arabidopsis thaliana*, *I. echinospora* and
583 *Selaginella moellendorffii*. Comparison of gene expression profiles by PCoA in the transcriptomes of
584 *A. thaliana,* circles, *S. moellendorffii,* squares and *I. echinospora,* triangles. Leaf samples coloured
585 green, root samples orange. A, comparison of principal coordinate axis 1 and 2. B, comparison of
586 principal coordinates axis 1 and 3. C, comparison of principal coordinate axis 2 and 3. Values in
587 brackets on each axis describe the percentage of total variance accounted for by each axis. Axis 1
588 separates gene expression in the three species. Axis 2 distinguishes gene expression between the
589 two lycophytes transcriptomes, *I. echinospora* and *S. moellendorffii* from *A. thaliana.* Axis 3
590 distinguishes between the leaf samples and the root samples in each transcriptome.

591

bioRxiv preprint doi: https://doi.org/10.1101/2019.12.16.878298; this version posted December 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

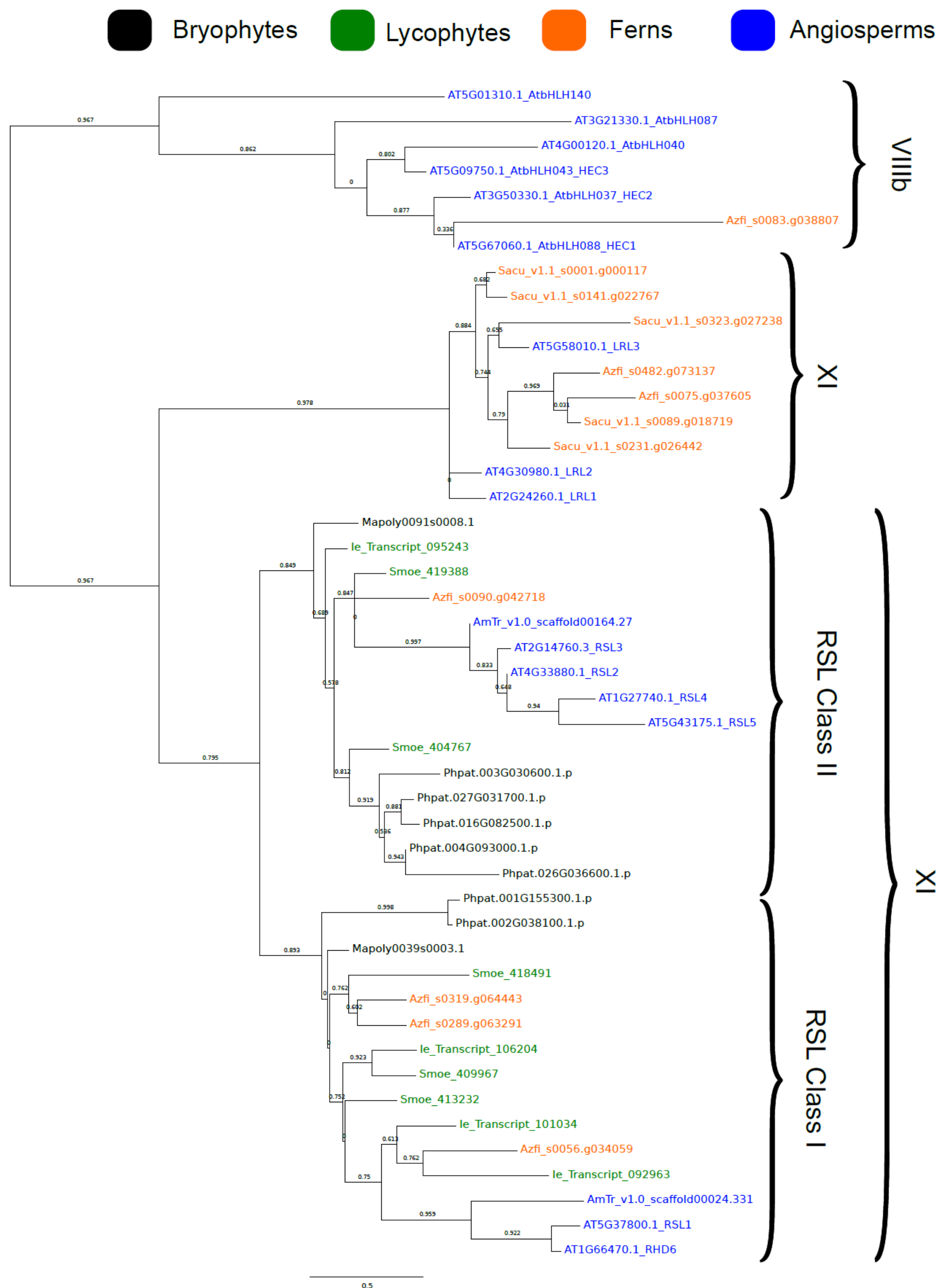


	Transcript number	Expression (TPM) Leaf	Corm	Rootlet
Class I	Ie_Transcript_106204	1.13	3.46	9.12
	Ie_Transcript_101034	0.51	0.72	3.21
	Ie_Transcript_092963	1.24	0.98	2.01
Class II	Ie_Transcript_095243	0	0	2.61

Fig. 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[87]. 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.

bioRxiv preprint doi: https://doi.org/10.1101/2019.12.16.878298; this version posted December 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

603 **Supplementary figure**



604

605 Supplementary Figure S1. There are no RSL genes in the *Salvinia cucullata* genome or proteome.
606 Gene tree analysis of RSL and related basic Helix-Loop-Helix (bHLH) transcription factors in a subset
607 of land plant species. There are not *S. cucullata* genes in the RSL clade instead closely related *S.*
608 *cucullate* genes are members of subfamily XI. Maximum likelihood gene tree of bHLH transcription
609 factors generated in PhyML 3.0. Gene names: black, bryophytes; green, lycopsids; orange, ferns;
610 blue, angiosperms. The RSL genes are grouped into two monophyletic classes; Class I and Class II.
611 Species name abbreviations: Phpat, *Physcomitrella patens*; Mapoly, *Marchantia polymorpha*; Some,
612 *Selaginella moellendorffii*; Sk, *Selaginella kraussiana*; Ie_Transcript, *Isoetes echinospora*; Sacu,

bioRxiv preprint doi: https://doi.org/10.1101/2019.12.16.878298; this version posted December 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

613 *Salvinia cucullata*; Azfi, *Azolla filiculoides*; AmTr, *Amborella trichopoda*; AT *Arabidopsis thaliana*.
614 Branch support as Shimodaira-Hasegawa-like approximate likelihood ratio tests.
615