Title:

The *Azolla* fern symbiosis sexual reproduction requires far-red light and involves responsive CMADS1 homologue, miR319-controlled GAMYB, miR529 in the fern, and transporters in the symbiont

Authors:

Laura W. Dijkhuizen[1], Badraldin E. S. Tabatabaei[2], Paul Brouwer[1], Niels Rijken[1], Rushi S. Mehta[1], Erbil Güngör[1], Henriette Schluepmann[1]

Addresses:

1, Department of Biology, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

2, Department of Biotechnology, College of Agriculture, Isfahan University of Technology, Isfahan, 84156-83111, Iran

Author for correspondence:

H. Schluepmann, tel +31 30 2533289, e-mail h.schlupmann@uu.nl

Total word count: 6404

word counts for each section:

Introduction:	1041
M&M:	1291
Results:	2498
Discussion:	1574

Number of Figures: 7, all of which are in color.

Tables : 2

Supporting Information: Supplemental methods
Table S1-5
Fig. S1-8

Summary

- Regulation of the transition from vegetative to spore-forming phases in ferns is largely unknown yet a pre-requisite for their domestication. External and endogenous cues regulating the formation of sporocarps were researched in *Azolla* ferns, that are heterosporous and constitute a symbiosis with the cyanobacteria *Nostoc azollae*.
- Once external cues were identified, small RNA and mRNA from eukaryotic and prokaryotic partners in the symbiosis were analyzed for differential accumulation upon transition to sexual reproduction.
- *A. filiculoides* fern sporocarp formation required far-red light (FR) and was suppressed by medium nitrogen; FR-induced sporocarps were viable in crosses that verified species attribution of Dutch strains but not those from Iran's Anzali lagoon. FR-responsive transcripts encoded the *Azolla* MIKC^{C} homologue of CMADS1, miRNA319-controlled GAMYB in the fern, ycf2 in chloroplasts, and transporters in *N. azollae*. Loci of conserved microRNA in the fern lineage included miR172, but FR only induced mirR529 and miR535, and reduced miR159 and miR319.
- The clade of the FR-responsive MIKC^{C} radiated separately in ferns from the seed plant clade containing *AtSOC1* and floral homeotic genes A, C, D and E, yet the data supports that regulons controlling seed plant flowering originate from the diploid to haploid phase transition in the common ancestor of seed plants and ferns.

Keywords: *Azolla*, Dual RNA-sequencing, fern, GAMYB/MIKC^{C} transcription factors, *Nostoc azollae*, miRNA, Sexual reproduction, Symbiosis.

23 **INTRODUCTION (1041)**

24 Ferns from the genus *Azolla* thrive floating on freshwater. They require no nitrogen whilst growing
25 prolifically due to an unusual symbiosis with the cyanobacterium *Nostoc azollae* which fixates dinitrogen
26 with light energy (Brouwer *et al.*, 2017). Previously used as biofertilizer, they now are considered a
27 candidate crop for sustainable production of plant protein in subsiding delta regions but have never been
28 domesticated (Brouwer *et al.*, 2014, 2018). What controls their sexual reproduction is completely
29 unknown yet central to containment, propagation, breeding schemes or the establishment of biodiversity
30 and germplasm banks.

31 *Azolla* ferns are heterosporous with diploid sporophytes generating two types of sori mostly called
32 sporocarps (Nagalingum *et al.*, 2006; Coulter, 1910). Unlike homosporous ferns, the haploid phase of
33 *Azolla* ferns, the gametophytes, are not free-living but contained inside the sporocarps. The sexual
34 reproduction is therefore controlled by the sporophyte as in the case of seed plants. Unlike seed plants,
35 gametophytes do not develop inside the sporocarps when these are still on the sporophytes and therefore
36 fertilization is not guided by structures on the sporophyte such as flowers in angiosperms. *Azolla*
37 sporocarps typically detach when mature and sink to the sediment of ditches; they constitute resting
38 stages with nutrient reserves that survive drying or freezing and, thus, are key for dissemination practices
39 (Brouwer *et al.*, 2014).

40 The developmental transition to reproductive growth of the sporophyte (RD) occurs when shoot apices
41 of the sporophytes will form initials for sporocarps in addition to branches, leaves and roots. RD is not
42 well studied in ferns. It is analogous, or possibly homologous, to the transition in angiosperms of shoot
43 apical meristems (SAM) to inflorescence meristems: the meristem is switching to making branches with
44 leaves and sporocarps whereby the making of sporocarps is ancestral to the making of leaves (Vasco *et*
45 *al.*, 2016). Gene expression in lycophytes and ferns of class III HD-Zip transcription factors (C3HDZ) was
46 observed in both leaves and sporangia even though ferns and lycophyte fossil records have shown that
47 the lineages evolved leaves independently. Vasco *et al*. (2016), therefore, suggested that the common
48 C3HDZ role in sporangium development was co-opted for leaf development when leaf forms, the micro
49 and megaphylls, evolved in lycophytes and ferns, respectively.

50 In angiosperm SAM transition to RD, the LEAFY transcription factor (TF) induces floral homeotic genes and
51 a shoot meristem identity gene; LEAFY functions solely when bound to proteins in a complex (Sayou *et*
52 *al.,* 2016). LEAFY is conserved in plant lineages, but LEAFY does not promote formation of reproductive

53 structures in seed-free plants including ferns where, nonetheless, the protein was reported to maintain
54 apical stem cell activity (Plackett *et al.*, 2018). The pattern emerging, therefore, is that the LEAFY complex
55 is not among those that were already in place to link exogenous cues to the transition of SAM to RD in the
56 common ancestor of seed plants and ferns.

57 In angiosperms, MIKC^C-type TF are defined by the conserved MADS box, Intervening, Keratin and C-
58 terminal domains and form characteristic complexes that integrate endogenous and exogenous cues
59 turning SAM into RD (Theißen *et al.*, 2018). In *Arabidopsis thaliana (Arabidopsis)*, for example, gibberellic
60 acid (GA)-dependent and photoperiod/temperature signaling is integrated by MIKC^C from the FLC and SVP
61 clades that, unsurprisingly, are floral repressors. Consistent with separate transitions for inflorescence
62 and flower meristems in Arabidopsis, the MIKC^C floral integrators act on another, SOC1, that in turn
63 induces the expression of MIKC^C floral homeotic genes. Homeotic genes of particular interest specify
64 ovules, and stamens. These modules could be more ancient than those specifying carpels and sepals but
65 specification of ovules/stamens in angiosperms may not be homologous to that of mega- and
66 microsporocarps in *Azolla* ferns because heterospory evolved independently in ferns and angiosperms
67 (Sessa & Der, 2016).

68 In addition to MIKC^C TF, micro RNA (miR) are known to control phase transitions in angiosperms, including
69 RD. The miR156/172 age pathway controls flowering: miRNA156 decreases with the age of plants, and
70 thus also its repression of the targets *SQUAMOSA PROMOTER BINDING-LIKES* (*SPL*). In the annual
71 Arabidopsis, SPL promote expression of mir172 that targets another type of floral repressor, the AP2/TOE
72 TF (Hyun *et al.*, 2019). In the perennial *Arabis alpina*, the MIKC^C FLC homologue and miR156 regulate
73 expression of *SPL15* thus integrating temperature and age-dependent RD. Moreover, AP2/TOE and NIN
74 TF are known to mediate repression of *SOC1,* and thus RD, in Arabidopsis on high nitrate (Gras *et al.*, 2018;
75 Olas *et al.*, 2019). SOC1 is further known to control the steady state of miR319 that inhibits TCP protein
76 functions necessary for the transition to RD (Lucero *et al.*, 2017). Inextricably, therefore, transition to RD
77 is linked to conserved miRNA modules in angiosperms but we know little of miRNA loci in ferns (Berruezo
78 *et al.*, 2017; You *et al.*, 2017).

79 *Azolla* ferns and Arabidopsis were shown to share responses to nitrogen starvation, suggesting that the
80 ancestor of both had already evolved some of the responses to exogenous cues we know from
81 angiosperms (Brouwer *et al.*, 2017). GA applications to *Azolla* ferns, however, were reported to not be
82 effective to induce RD in *Azolla*, neither was stress treatment (Kar *et al.*, 2002). Systematic research of the
83 symbiosis responses to day-length and light quality is lacking. The coordinate development of the

84 cyanobacterial symbiont, *N. azollae,* and fern during the induction of sporocarps is documented:
85 hormogonia generated in the shoot apical colony are attracted to the developing sporocarps in the shoot
86 apex; attraction to either the leaf cavities or the sporocarps is likely mediated by secretions from
87 specialized trichomes (Perkins & Peters, 1993). How the development of symbionts and host are
88 coordinated upon transition to RD is unknown. Sequences of both genomes are now available, setting the
89 stage for gene-expression profiling of both the pro- and eukaryote RNA by dual RNA sequencing.

90 Here we identify external cues that induce or inhibit transition to RD in *A. filiculoides* in laboratory settings.
91 We test the viability of reproductive structures induced in different fern strains in crosses, herby defining
92 species attributions, and support the results obtained with phylogenetic analyses of the strains. To reveal
93 endogenous mechanisms mediating transition to RD, we profile transcript accumulation by dual RNA-
94 sequencing comparing sporophytes induced for RD by far-red light, at a time point when the reproductive
95 structures are not yet visible. We assign induced $MIKC^C$ and MYB TF to known clades using phylogenetic
96 analyses. In addition, we profile small RNA from the symbiosis, then identify miRNA and their targets in
97 the fern lineage. Finally, we discover regulatory modules of conserved miRNA and their targets responsive
98 to induction of RD in ferns.

99 **MATERIALS AND METHODS (1353)**

100 **Plant Materials, growth conditions and crosses**

101 Ferns were collected from different locations in the Netherlands, grown in liquid medium at a constant
102 temperature of 22ºC and a 16 h light period as described earlier (Brouwer *et al.*, 2017). Ferns from Spain
103 and Iran were processed to DNA on site. GPS coordinates from the collection sites of the strains used in
104 phylogenetic analyses are in Supporting Information (SI) Table S**1**. For sporocarp induction, ferns were
105 first raised under 16 h tube light (TL) periods, then transferred for induction under 16 h TL with far-red
106 (FR) Light Emitting Diodes - LED (Phillips GrowLED peaking at 735 nm, SI Fig. **S1a,b**). Photosynthetic Photon
107 Flux Density varied in the range 100-120 µmol m^{-2} s^{-1}. Sporophytes were generally kept at densities that
108 covered the surface. When testing sporocarp induction sporophytes were kept at set densities (with fresh
109 weight resets once per week, SI Fig. S**1c,d,e**) and set red to FR light ratios (SI Table S**2**). Phenotypic change
110 in mature sporophytes were photographed using a Nikon D300s with an AF Micro-Nikkor 60mm f/2.8D
111 objective or bellows on a rail to generate image reconstructions with Helicon focus software**.** Mega- and
112 microsporocarps were collected manually using tweezers. To initiate crosses, microsporocarps were
113 Dounce-homogenized to release massulae and then added to megasporocarps in distilled water (3-6 ml,

114 in 3 cm diameter petri dish), gently shaking to obtain clumps. The crosses were kept at room temperature
115 (about 18°C) under 16 h light conditions under TL at about 100 µmol m^{-2} s^{-1}. As sporelings emerged they
116 were transferred to liquid medium (Brouwer *et al.*, 2017).

117 **RNA extraction and quantitative RT-PCR** were as in (Brouwer *et al.*, 2017) and are detailed in Methods
118 **S1**.

119 **Phylogenetic analyses** are described in details in Methods **S1**. Chloroplastic marker regions were as in SI
120 Table S**3**. The $MIKC^C$ tree was subsampled from individual genome browsers with 1 KP as a guide (Leebens-
121 Mack *et al.*, 2019); the R2R3 MYB tree used data from respective genome repositories, the *Azolla* genome
122 (Li *et al.,* 2018) and Jiang & Rao, 2020.

123 **Reference genomes** used to align Dual RNA sequences onto in the *A. filiculoides* Galgenwaard strain are
124 detailed in Methods **S1**.

125 **Dual-RNA sequencing and data processing**

126 *A. filiculoides* sporophytes were from the sequenced accession Galgenwaard (Li *et al.*, 2018). The ferns
127 were maintained on liquid medium without nitrogen in long-day TL, then transferred to TL with and
128 without FR for 7 days, on fresh liquid medium without nitrogen; harvest was 2 h into the light period by
129 snap freezing. All samples were grown separately and harvested as triplicate biological replicates. Total
130 RNA was extracted using the Spectrum Plant Total RNA (Sigma-Aldrich) using protocol B, then DNAse
131 treated as described in (Brouwer *et al.*, 2017), and cleaned using the RNeasy MinElute Cleanup Kit
132 (Qiagen). rRNA depletions were carried out as per protocol using Ribo-Zero rRNA Removal Kit (Plant Leaf,
133 Illumina), and the RNA was then cleaned again using the RNeasy MinElute Cleanup Kit. Stranded libraries
134 were then synthesized using the Ovation Complete Prokaryotic RNA-Seq Library Systems kit (Nugen). The
135 libraries thus obtained were characterized and quantified before sequencing using a single lane Hiseq4000
136 and 50 cycles PE reading (Illumina).

137 Reads obtained were de-multiplexed then trimmed and quality filtered with trimmomatic (Bolger *et al.,*
138 2014), yielding 13-60 million quality PE reads per sample. Alignment of the reads to the genome assembly
139 and GFF vs1 of *A. filiculoides* (Azfi_vs1; Li *et al.*, 2018) was with STAR-aligner default settings (Dobin *et al.*,
140 2013) and was combined with the STAR read count feature (PE read counts corresponding to the HTseq
141 "Union" settings, (Anders *et al.*, 2015)). Read count tables were fed into DESeq2 (Love *et al.*, 2013) for
142 analysis of differentially accumulating transcripts.

143 STAR was also used to align reads to the *N. azollae* and the chloroplast genomes, and read counts then
144 extracted using STAR default parameters or VERSE (Zhu *et al.*, 2016) with HTSeq setting "Union"
145 comparing single and paired end read counts on the polycistronic RNA. The three approaches yielded
146 similarly ranked genes but the Padj statistics of differential expression in count tables from the forced
147 single read counts yielded more dispersion in the data: only 20 *N. azollae* genes had P- adjusted values
148 smaller than 0.1 compared to up to 67 with the PE- approaches, we therefore present the PE-approach
149 using STAR read counts only.

150 STAR alignment to concatenated genomes of the *Azolla* symbiosis using default settings yielded a large
151 proportion of multi-mappers which interfered with PE read counting leading to losses of as much as 30%
152 of the reads. The strategy would reduce force mapping but is not necessary to extract differentially
153 accumulating transcripts. We, however, used alignments allowing no mismatches to concatenated
154 genomes when mapping the sRNA reads below, then recovering reads aligning to each genome using
155 Samtools fasta from the indexed .bam files.

156 **Small RNA data sequencing and data processing**

157 Ferns were maintained on liquid medium without nitrogen with 16 h TL, then transferred to TL with (F)
158 and without (T) FR for 7 days as described under Dual-RNA sequencing. In addition, sporophytes in F were
159 also grown on medium with 2 mM NH_4NO_3 (FN); ferns without cyanobacteria (C) were grown as in FN. All
160 sporophytes were harvested 7 d into the treatment, 2 h into the light period snap frozen. Samples were
161 harvested from triplicate growth experiments and thus represent triplicate biological replicates. Small
162 RNA was extracted as per protocol using the mirPremier microRNA Isolation Kit (Sigma-Aldrich), then
163 DNAse treated (Brouwer *et al.*, 2017), then cleaned using the RNeasy MinElute Cleanup Kit (Qiagen)
164 protocol for small RNA. Libraries of the sRNA were then generated using the SeqMatic TailorMix miRNA
165 Sample Preparation 12-reaction Kit (Version 2) as per protocol. Libraries were characterized using an
166 Agilent Technologies 2100 Bioanalyzer and quantified before sequencing using NextSeq500 and 1 x 75 b
167 read chemistry (Illumina).

168 Reads thus obtained were de-multiplexed, trimmed and quality filtered with trimmomatic, then aligned
169 with STAR (Dobin *et al.*, 2013) (set at maximal stringency and to output Bam files) to the concatenated
170 fasta files of the genomes Azfi_vs1, its chloroplast, and *N. azollae* 0708. The resulting bam files were then
171 split for each genome, and Samtools fasta (Li *et al.*, 2009) used to extract reads aligning to each genome.
172 fastx collapser (http://hannonlab.cshl.edu/fastx_toolkit) was then used to extract read counts for each

173 identical sequence, with the sequence defining the name in the lists obtained. A data matrix was
174 generated in bash, then exported to Excel or Rstudio for further analyses.

175 To determine differential expression, the lists obtained were restricted to sRNA with at least 10 reads per
176 sample joined in bash, and the tables thus obtained, first restricted to containing only reads of 20-22 nt
177 sRNA (to minimize gel excision bias during sequencing library preparation), then exported to R (Booth *et*
178 *al.*, 2018). DESeq2 was then used to display variation and dispersion in the data and identify differentially
179 accumulating sRNA using Padj values (Love *et al.*, 2013).

180 **miRNA discovery**

181 The miRNA loci in *A. filiculoides* were analyzed as in SI Fig. S**2**. Pre-miRNA were predicted with miRDEEP-
182 P2 (Kuang *et al.*, 2019); multi-mappers min. 100 and max. loop length 500) then extracted and subjected
183 to secondary structure prediction using RandFold default parameters (Bonnet *et al.*, 2004; Gruber *et al.*,
184 2008). miRNA candidates containing less than four mismatches within the hairpin structure were viewed
185 in IGV along with RNA expression data aligned to Azfi_vs1. Predicted miRNAs that were expressed in *A.*
186 *filiculoides* were then compared in miRbase (Kozomara *et al.*, 2019). Prediction of miRNA targets in
187 Azfi_vs1 was using the intersect of both psRNATarget (Dai *et al.*, 2018) and TargetFinder (Srivastava *et al.*,
188 2014) with a cutoff of 5.0. Predicted targets were linked to the closest homolog using the Mercator
189 annotation (Lohse *et al.*, 2014) of Azfi_vs1 predicted proteins.

190 **Data sharing**

191 The data that support the findings of this study are openly available in [repository name] at [URL],
192 reference number [reference number]. Data will be uploaded upon acceptance of the manuscript and will
193 include 1) read files for the dual RNA seq as well as small RNA seq on ENA 2) phylogeny files on Gitup.

194

195 **RESULTS (2498 words)**

196 **Far-red light induces, while presence of nitrogen in the medium represses, sexual reproduction in *A.***
197 ***filiculoides***

198 A young fern growing at low density typically spreads over the water surface (Fig. **1a**) in contrast to the
199 three dimensional web of a fern that underwent transition to reproductive development in dense
200 canopies (Fig. **1b**). Unlike in the glasshouse, *A. filiculoides* ferns never sporulated under tube light (TL, SI

201 Fig. S**1a**), TL did not emit any photons in the far-red range (FR, SI Fig. S**1a**). We, therefore, added far-red
202 LED (FR, SI Fig. S**1b**) during the entire 16 h light period. This initially resulted in the elongation of the stem
203 internodes of the fern (Fig. **1c**) and, after 2-3 weeks, in the formation of the sporocarps (Fig. **1d**).
204 Sporocarps were positioned as pairs on the branch side of branch points and were predominantly
205 arranged as a mega- and microsporocarp pair (Fig. **1e**).

206 To test the optimum light for sporocarp induction, decreasing ratios of red to far-red LED were applied to
207 ferns that had been raised in TL light and were then maintained at three differing densities (SI Table S**2**,
208 SI Fig. S**1c**,**d**,**e**). Far-red light was required for sporocarp formation regardless of the density of the fern
209 culture (Fig. **1f**). Increasing far-red whilst keeping a constant photosynthetic active radiation (PAR) yielded
210 more sporocarps per fern mass. The sporocarps per fern mass increased with time of exposure to the far-
211 red LED up until five weeks and then decreased, presumably because ripe sporocarps detached. The
212 sporocarps per fern mass, furthermore, increased with the density of the fern culture. Reducing the PAR
213 at highest far-red intensity revealed that PAR limits the production sporocarps at high far-red light (Fig.
214 **1f_**FR4).

215 Adding 2 mM ammonium nitrate to the medium drastically reduced the spore count when inducing
216 sporocarp formation with far-red LED (Fig. **1g**). Both, ammonium and nitrate added individually inhibited
217 sporocarp formation (data not shown), and inhibition explained why sporophytes without *N. azollae* kept
218 on various nitrogen sources never sporulated.

219 The ratio of mega to microsporocarp observed was approximately equal over nine weeks of induction (SI
220 Fig. S**1f**); and the marker *AzfiSOC1* (Brouwer *et al.*, 2014) was maximally induced when sporocarps became
221 visible four weeks after induction started.

222 **FR-induced sporocarps are viable which permits reciprocal crosses between strains to test phylogeny-**
223 **derived species attributions**

224 We next wondered whether the sporocarps induced with far-red LED were viable. We randomly tested
225 reciprocal crosses of the sequenced strain *A. filiculoides* Galgenwaard (Li *et al.*, 2018), six further strains
226 from the Netherlands, one from Gran Canaria (Spain) and another from the Anzali lagoon (Iran) (SI Table
227 S**1**). All strains could be induced to sporulate under far-red supplemented TL except Anzali. All sporulating
228 strains tested in crosses could be crossed suggesting that the ferns from the Netherlands and Spain were
229 *A. filiculoides* (Table **1**).

230 To research the taxonomy of the strains, intergenic regions of the choloroplast rRNA, used for
231 phylogenetic studies previously (Madeira *et al.,* 2016), were amplified and sequenced to compute
232 phylogenetic relations. The regions trnL-trnF (Fig. **2a**), and trnG-trnR (Fig. **2b**) yielded similar trees
233 confirming the close relationship of the *A. filiculoides* strains from the Netherlands and Spain. The trees
234 further revealed that the Anzali strain clusters along with two stains from South America (Southern Brazil
235 and Uruguay, SI Table S**1**). Use of the nuclear ITS region for phylogeny reconstruction proved futile when
236 amplifying the region on the *A. filiculoides* genome *in silico*: the ITS sequence is variable within the genome
237 (Fig. **2c**) and variations are very large when comparing to other species of *Azolla* (SI Fig. S**3a**).

238 The dorsal-side of the upper leaf lobes of the Anzali strain had two-celled, distinctly pointed, trichomes
239 unlike *A. filiculoides* that mostly, but bot exclusively, have single celled and rounded trichomes (SI Fig.
240 S**3b**). Consistent with its differing response to FR light, and its phylogenetic position based on the
241 chloroplast sequences, the Anzali strain, we conclude, is not *A. filiculoides*.

242 **Dual RNA sequencing by rRNA depletion profiles RNA from organelle, prokaryotic symbiont and fern**
243 **nucleus**

244 To research molecular processes altered during transition to reproductive development, ferns maintained
245 on TL were transferred to TL with and without FR for a week in triplicate cultures, then collected 2 h into
246 the 16 h light period snap frozen. Total RNA was then extracted, its plant rRNA depleted, and stranded
247 sequencing libraries generated which were depleted for rRNA from gram-negative bacteria. Libraries were
248 sequenced to obtain 12 to 60 million paired end (PE) 50 b reads, quality filtered and trimmed. To align the
249 reads obtained, nuclear and chloroplast genomes with annotations for the *A.filiculoides* Galgenwaard
250 strain were available (Li *et al.,* 2018) but not for its cyanobacterial symbiont.

251 Assembly of the *N. azollae* genome from metagenome short reads of the Galgenwaard strain yielded a
252 bin with 385 contigs that proved to be highly complete (95.8%) based on single copy marker genes. Contigs
253 aligning to the *N. azollae* 0708 reference genome (Ran *et al.*, 2010) were well over 99.5% identical (Fig.
254 **3a**). This degree of nucleotide identity was well above the maximum mismatch of STAR-aligner's default
255 setting; PE reads were thus aligned to the Galgenwaard strain nuclear and chloroplast genomes, and the
256 *N. azollae* 0708 genome, and read counts derived from their GFF annotation files. Alignments revealed a
257 high proportion of uniquely mapped reads to the fern nuclear and *N. azollae* genomes suggesting efficient
258 removal of rRNA (Fig. **3b**). The larger proportion of PE reads mapping to multiple loci in the chloroplast
259 indicated that fern chloroplast rRNA was not as efficiently removed (Fig. **3b**), yet read counts per gene

260 were still high because few genes are encoded in the chloroplast genome (84 without rRNA and tRNA
261 genes). Read counts in genome features further revealed that sense alignments dominated and thus that
262 the libraries were stranded (Fig. **3c**).

263 Dispersion of sense read count per gene was low which permitted identification of differentially
264 accumulating transcripts (defined as DEseq2 Padj <0.1; Love *et al.,* 2014) comparing ferns on TL and TL
265 with FR: 318 from the fern nucleus, 67 from *N. azollae* and 5 from the fern chloroplast (SI Table S**4**). We
266 next wondered whether the changes in transcript abundance reflected adaptations to light quality and
267 the transition to RD.

268 **FR triggers small transcriptional changes reflecting light-harvesting adaptations in chloroplasts and *N.***
269 ***azolla,* yet large changes in transporters and transposases of *N.azollae***

270 After a week exposure to FR, increased chloroplast transcripts encoded ycf2 (log2 of 2.9), energizing
271 transport of proteins into the chloroplast, and RNA polymerase subunit rpoA (log 2 of 1.3) (SI Table S**4**, *A.*
272 *filiculoides* chloroplast). Decreased chloroplast transcripts included psbD, previously reported as
273 responsive to light quality in plant lineages (Shimmura *et al.*, 2017).

274 Consistent with the high levels of N_2 fixation measured in *Azolla* ferns, the highest *N. azollae* read counts
275 were in transcripts of the *nif* operon, genes encoding photosystem I and II proteins and ATP synthase
276 subunits; in addition, read counts for *ntcA* and metabolic enzymes supporting high N_2 fixation were high
277 (SI Table S**5**). Read counts from the *N. azollae* transcripts thus reflected the activity of the symbiont and
278 we thus further analyzed the data for differential expression.

279 In *N. azollae*, FR led to accumulation of the PsbA transcript and a reduction of photosystem I reaction
280 center subunit XII, geranylgeranyl hydrogenase ChlP, and phycobilisome rod-core linker polypeptide
281 CpcG2 transcripts (SI Table S**4**, *N. azollae*). But these changes in light harvesting-related genes were small
282 compared to changes in hypothetical proteins of unknown function, transposases and highly expressed
283 transporters (TrKA, MFS and ABC transporters). We conclude that one week induction of reproductive
284 structures with FR triggered few transcriptional changes reflecting light-harvesting adaptations in *N.*
285 *azollae*. The larger differential accumulation of transporter transcripts reflect changes in metabolite
286 trafficking and communication with the host fern.

287 **FR alters transcripts from MIKC^C and R2R3MYB implicated in the reproductive transition of seed plants**

288 As in the case of *N. azollae* symbiont, the ferns largest changes in transcript accumulation under TL with
289 FR were in genes with unknown function, then in secondary- and lipid-metabolism (SI Table S**4**, *A.*
290 *filiculoides* nucleus). We noted the large induction of the SWEET-like transporter (Azfi_s0221.g058704
291 log2 7.5 fold), of transcripts related to cysteine-rich peptide signaling (Azfi_s0046.g030195 log2 7.1 fold;
292 Azfi_s0745.g084813 log2 4.0 fold) and of transcripts from jasmonate metabolism. Transcription factors
293 (TF) characterize the RD in seed plants and were, therefore, analyzed in more depth. Transcripts of MIKC^C
294 TF and of R2R3MYB TF were strikingly induced in sporophytes under TL with far-red LED (**Table 2**).

295 Prediction of the most induced TF transcript encoding a SOC1-like MIKC^C (Azfi_s0028.g024032, log2 7.2
296 fold) missed the K-domain because this TF was not expressed in vegetative sporophytes. Manual
297 annotation allowed computation of a phylogenetic tree that placed both induced *Azolla* MIKC^C in the clade
298 of fern sequences radiating separately from the angiosperm and gymnosperm MIKC^C clade that gave rise
299 to AtSOC1, FLC, and to the A,C,D and E-function floral homeotic genes (SI Fig. S**4**). Induction of the
300 AzfiSOC1-paralogue was confirmed by RT-PCR (SI Fig. S**2g**). In Arabidopsis, AtSOC1 is known to control the
301 steady state of miR319 that inhibits TCP TF functions necessary for the transition to an inflorescence
302 meristem at the shoot apex (Lucero *et al.*, 2017); the link may also operate in *Azolla* ferns since a TCP TF
303 (Azfi_s0168.g054602) was induced in ferns on TL with far-red LED (SI Table S**4**, *A. filiculoides* nucleus).

304 Transcripts of R2R3MYB TF could be assigned to families using data from Jiang & Rao, 2020 (**Table 2**). The
305 highly induced class VIII-E TF (log2 6.1 fold) likely controls changes in secondary metabolism together with
306 the similarly changed bHLH TF (Güngör *et al.*, 2020), and therefore may not be linked to RD-directly. In
307 contrast, induced class V (log 2 2.1 fold) and VII (log 2 1.9 fold) TF are known to affect RD, with the GA
308 pathway associated class VII GAMYB from seed plants known to be regulated by miR319.

309 Prominence of transcription factors in Table **2** suspected to controlling RD as in seed plants begged the
310 question as to whether any RD pathways with conserved miRNA were conserved in the common ancestor
311 of seed plants and ferns. This required to first characterize miRNA loci in the *A.filiculoides* symbiosis.

312 **Small RNA profiles are characteristic for fern nucleus, chloroplast or *N. azollae***

313 Short RNA reads were obtained from sporophytes after a week growth under TL without and with FR from
314 the same cultures as those used for dual RNA sequencing. In addition, reads were obtained from
315 sporophytes grown under TL with FR on medium with 2 mM NH_4NO_3, and from sporophytes without *N.*
316 *azollae* on medium with 2 mM NH_4NO_3. All samples were collected two hours into the 16 h light period.
317 The sRNA sequencing library preparation chemistry did not distinguish pro- from eukaryote sRNA and

318 therefore was inherently a dual profiling method. Reads obtained were aligned to concatenated genomes
319 which yielded small RNA profiles characteristic for the fern nucleus, chloroplast and *N. azollae* (Fig. **4**; SI
320 Fig. S**5**).

321 The most abundant sRNA encoded in the fern nucleus were 21, 20 and 24 nt long, but the most diverse
322 were 24 nt long, consistent with abundant 21 nt miRNA discretely cut from a hair-pin precursor by dicer,
323 and the 24 nt siRNA generated by random processes (Fig. **4a**). The chloroplast most abundant sRNA was
324 also the most diverse at 28 nt, but the chloroplast have an unusual peak of sRNA of low diversity at 34 nt
325 (Fig. **4b**). The sRNA encoded in *N. azollae* are generally shorter than those of the chloroplast, yet, they
326 have a conspicuous peak of low diversity at 42 nt length, consistent in length with CRISPR RNA (Fig. **4c**).

327 **Conserved miRNA in *A. filiculoides* include miR172 and 396, but only miR529 and 535 increased in**
328 **response to FR**

329 Fern genome-encoded reads of 20-25 nt length were used for miRNA predictions allowing a hairpin of
330 maximally 500 bp. Loci predicted were then verified against miRNA criteria (Axtell & Meyers, 2018) as, for
331 example, in the case of the miRNA 172 locus Azfi_s0021g15800 (Fig. **5**): the locus was expressed as a pre-
332 miRNA (Fig. **5a**, RNAseq all conditions) and both the miRNA and miRNA* are found in replicate samples
333 and cut precisely with two base overhangs (Fig. **5b,c**, sRNAseq). The miR172 reads were derived from the
334 fern since they were also found in sporophytes without *N. azollae* (Fig. **5,** RNAseq -*N. azollae*). The locus
335 encoded a stable long hairpin which when folded released -65.76 kcal mol^{-1} at 37°C. MiR172 targets in
336 seed plants include the AP2/TOE transcription factors, and alignment of the target sites to all annotated
337 AP2/TOE transcripts from *A. filiculoides* revealed target sites for the AzfimiR172 (SI Fig. S**6**). The miR156a
338 locus was similarly verified (SI Fig. S**7**).

339 Mature miRNA were compared in miRbase (Kozomara *et al.*, 2019) which identified 11 of the conserved
340 miRNA families of 21 nt in the *A. filiculoides* genome and thus in the fern lineage (Fig. **6a**). The list included
341 miR396, as well as miR172, not yet confirmed in other seed-free plant lineages. MiRNA predictions further
342 uncovered a number of novel miRNA (miRNF).

343 Differential expression analysis of 20-22 nt reads exhibited more dispersion than the dualRNAseq of long
344 RNA, yet, it reliably identified sRNA with low Padj values when comparing ferns under TL with and without
345 FR (SI Fig. S**5d,e**). Reads from miR529c&d and miR535 increased (Padj < 0.1) whilst those of miR159,
346 miR319 and miRNF4 decreased (Padj <0.16) in sporophytes on TL with FR compared to TL light only (Fig.

347 **6b**). We conclude that the miRNA156/miR172 known to reciprocally control flowering are not altered in
348 the ferns undergoing transition to RD under TL with FR.

349 **miR319/GAMYB is responsive in sporophytes induced by FR**

350 miRNA targets were predicted combining psRNATarget (Dai *et al.*, 2018) and TargetFinder (Srivastava *et*
351 *al.*, 2014) then annotated with Mercator (Lohse *et al.*, 2014; Schwacke *et al.*, 2019). Predictions were
352 consistent with analyses in other land plant lineages (SI Table S**6**): miR172 targeted AP2/TOE-like
353 transcription factors and miR156, 529 and miR319 SPL-like proteins.

354 Transcripts of neither SPL nor AP2/TOE-like targets, however, were significantly altered in sporophytes
355 comparing TL with and without FR; this could be due to poor Azfi_vs1 annotation and incomplete
356 prediction of miR targets (SI Table S**6**, F vs T Padj). Nevertheless, consistent with an increased miR529c&d
357 the SPL from the Azfi_s0173.g055767 transcript is reduced (log2 0.74 fold) but with low significance (Padj
358 of 0.356). In contrast, transcripts of two GAMYB (Azfi_s0004.g008455 and Azfi_s0021.g015882) increased
359 up to 4 fold in sporophytes with decreased miR319 on TL with FR. The location of mir319 binding was
360 identified for the GAMYB encoded in Azfi_s0004.g008455 (SI Fig. S**8**).

361 Phylogenetic analyses revealed that the GAMYB targeted by miR319 were on the same leaf as the
362 GAMYB33 from *Arabidopsis thaliana* and were placed next to each other probably because they arose
363 from gene duplication (Fig. **7**). This is consistent with the whole genome duplication observed at the base
364 of *Azolla* evolution (Li *et al.*, 2018). We conclude, therefore, that miR319/GAMYB is responsive in
365 sporophytes undergoing transition to RD caused by FR.

366

367 **DISCUSSION (1574)**

368 **Far-red light induced stem elongation and reproductive development in ferns, an analogy to**
369 **angiosperms?**

370 The responses resemble a shade-avoidance syndrome but signal transduction pathways that mediate it in
371 ferns remain largely uncharacterized (Inoue *et al.*, 2017). Moreover, pathways causing shade response
372 components differ as they use alternative phytochromes and interacting factors (Possart *et al.*, 2014; Xie
373 *et al.*, 2020). Unlike *A. filiculoides*, Arabidopsis will transition to RD in the TL used for indoors cultivation
374 over the past decades. Most studies on the transition to RD, therefore, were done with red-light gated

375 plants, i.e. conditions leading to artefactual peaking of CONSTANS protein accumulation and florigen (FT)
376 expression (Song *et al.*, 2018). Arabidopsis as well as many other seed plants flower early when in ratios
377 of red to far-red light approaching one, the ratio encountered in day light, compared to TL. Far-red light-
378 dependent flowering appeared mediated by PHYB and D, eventually leading to SOC1 expression in the
379 meristem (Halliday *et al.*, 1994; Aukerman *et al.*, 1997; Lazaro *et al.*, 2015). ft mutant Arabidopsis,
380 however, flowered earlier under incandescent light (Martinez-Zapater & Somerville, 1990) or in TL with
381 far-red LED compared to in TL (Schluepmann *et al.,* unpublished) suggesting that FT is not solely involved.
382 Signaling was also reported via PHYB-regulated PIF4 protein complexes that bind the promoters of
383 miRNA156/172 directly with interference from PHYA- regulated TF (Sánchez-Retuerta *et al.*, 2018; Sun *et*
384 *al.*, 2018; Xie *et al.*, 2020). In spite of the apparent commonalities and the fern lineage representing the
385 sister lineage of seed plants, RD pathways from Arabidopsis may not operate in *Azolla*: the highly induced
386 $MIKC^C$ turned out a paralogue of SOC1, and miR156/172 steady states were unaltered by the far-red
387 triggered transition to RD in *A. filiculoides* (Fig. **6b**; SI Table S**6**).

388 PHYB in Arabidopsis has been shown to sense temperature by thermal reversion of the Pfr to fr state,
389 explaining why many responses to light quality can be tweaked by temperature regimes, in seed plants
390 (Legris *et al.*, 2017; Klose *et al.*, 2019). Phytochromes in ferns radiated separately from those of seed
391 plants, but have a similar domain structure (Li *et al.*, 2015). Temperature signaling by phytochromes has
392 not been researched in ferns. Initiation of fern, liverwort and bryophyte sporangia, however, is dependent
393 on temperature and photoperiod and differed with each species analyzed (Labouriau, 1958, Benson-
394 Evans, 1961; Nishihama *et al.*, 2015). Moreover, thermal reversion of phytochromes from some
395 cyanobacteria was described i*n vitro* as early as 1997 (Yeh *et al.*, 1997), but any effect on signal
396 transduction *in vivo* was not researched. Even though far-red LED did not induce sporocarps in the *Azolla*
397 strain Anzali, differing temperature regimes, therefore, may still work. Morphological/physiological
398 differences and phylogenetic analyses placed the Anzali strain closely to strains of a different species
399 collected in Uruguay/Southern Brazil which did not cluster with the representative *A. caroliniana* species.
400 Position of the Anzali strain was consistent with observations of distinct *Azolla* ferns in the Anzali lagoon
401 (Farahpour-Haghani *et al.*, 2017), and suggests that not only *A. filiculoides* may have been released as
402 nitrogen biofertilizer.

403 Whether analogous or homologous, repression of reproductive transition under dominating red light may
404 reflect an ecological strategy: when no neighbors are close, there is sufficient space to prolong vegetative
405 development and no need for dispersal by RD. Consistently, density of the *A. filiculoides* canopy had an

406 additional impact on the number of sporocarps observed per gram plant (Fig. **1f**) that may stem from
407 alternative cues such as volatile organic compounds released by neighbors (Vicherová *et al.*, 2020). Taken
408 together, results signify that for the first time RD of the *Azolla* fern symbiosis is amenable to experimental
409 enquiry.

410 **Far-red light responsive accumulation of transcripts and small RNA implicate miR529, key MIKC^C TF and**
411 **the miR319/GAMYB module in fern RD**

412 The GAMYB clade arose before land-plants evolved and before GA signaling (Aya *et al.*, 2011; Bowman *et*
413 *al.,* 2018, Jiang & Rao, 2020, Fig. **7**). *Azolla* GAMYB induced under FR, therefore, likely function in RD by
414 mediating cues perceived by the sporophyte into an ancestral gametophyte regulon. GAMYB from the
415 moss *Physcomitrella patens* were required for the formation of antheridiophores and spores but we do
416 not know whether the *Pp*GAMYB regulon plays a role in RD of the moss sporophytes. GAMYB in mosses
417 and lycophytes regulate expression of enzymes assigned to sporopollenin biosynthesis which are also
418 upregulated in *Azolla* sporocarps (Aya *et al.*, 2011; Brouwer *et al.,* 2014). *At*GAMYB are known to be
419 regulated by miR159 in Arabidopsis and by the related and more ancestral miR319 in seed-free plant
420 lineages (Achard *et al.*, 2004; Palatnik *et al.*, 2007). The *At*GAMYB/mir159 is part of the GA pathway
421 promoting flowering in Arabidopsis under short days (Millar *et al.*, 2019). Specifically, AtMYB33 was shown
422 to directly act on the promotors of both miR156 and AtSPL9 (Guo *et al.,* 2017). FR repressed miR159 and
423 miR319/increased *Azfi*GAMYB in *Azolla* were thus consistent with predicted Arabidopsis wiring. GA signal-
424 transduction evolved in lycophytes, however, and we have yet to find out how far-red light causes miR319
425 induction in TL-grown sporophytes. Or, rather, why red-light gated ferns have increased miR319.

426 Far-red induced RD in the haploid phase of the liverwort *Marchantia polymorpha* (Chiyoda *et al.*, 2008;
427 Kubota *et al.*, 2014; Yamaoka *et al.*, 2018; Tsuzuki *et al.*, 2019). The mechanism in these liverworts
428 involved the *Mp*miR529c/*Mp*SPL module which was necessary for development of reproductive branches.
429 *Mp*PHY and *Mp*PIF were required to mediate the response (Inoue *et al.*, 2019). Our analyses with *A.*
430 *filiculoides* reveal induction of AzfimiR529 under far-red light but only insignificant change in SPL targets.
431 The pathway may differ in the two plants because the specific MpmiR529c sequence involved in
432 *Marchantia* was not detected in our ferns.

433 Control over RD as it switched from the gametophyte, in gametophyte dominated seed-free plants, to the
434 diploid sporophyte in lycophytes, ferns and seed plants, also resulted from the evolution of MIKC^C
435 regulons specific for each plant lineage (Leebens-Mack *et al.*, 2019, SI Fig. S**4**). The MIKC^C clade of TF that

436 specify both the transition to RD in sporophytes, such as *At*SOC1 and *At*FLC, and the A,C, D and E homeotic
437 functions in the inflorescence development of angiosperms are peculiar in that the TF radiated in each
438 megaphyll plant lineage separately. The TF work as hetero-tetramers and may have evolved from an
439 ancient tandem gene duplication through subsequent polyploidy events (Zhao *et al.*, 2017). Sepals are
440 organs formed only in angiosperms. Consistently, TF with A functions specifying sepals occurred solely in
441 angiosperms. MIKC^C specifying organs that contain the reproductive structures in angiosperms (C,D and E
442 homeotic genes), had direct homologues in gymnosperms, but not in ferns. Instead, ferns radiated a sister
443 clade to that containing AtSOC1, and A, C, D and E MIKC^C, in which the *Azolla* MIKC^C responsive to far-red
444 light were placed. This fern clade comprises *Ceratopteris thalictoides* CMADS1 reported highly expressed
445 in the sporangial initial (Hasebe *et al.*,1998). At the base of the radiating clade of modern MIKC^C that
446 specify modern organs of the sporophytes controlling the gametophyte development within, is the LAMB1
447 protein expressed specifically in sporogenic structures of lycophytes (Svensson *et al.*, 2000). Combined
448 phylogenetic and RNAseq analyses, therefore, suggest that *Azolla* fern sporocarps are likely specified by
449 two MIKC^C TF from the fern sister-lineage to AtSOC1. The result further suggests that SOC1 control over
450 flowering has its roots in the common ancestor of seed plants and ferns when a MIKC^C regulon controlled
451 the diploid to haploid phase transition.

452 **Dual RNA sequencing to study coordinate development in assemblages of pro- and eukaryotic**
453 **organisms**

454 The dual RNAseq method could be applied more generally to study bacteria/plant assemblages common
455 in the seed-free plant lineages, representing an important base of the "tangled tree" of lineage evolution
456 (Quammen D 2018). For *Azolla* where co-evolution of *N. azollae* and fern host genomes demonstrated
457 that fitness selection occurs at the level of the metagenome (Li *et al.*, 2018), it delivered foundational data
458 and opens the way to studying the coordinate development of host and obligate symbiont. Efficient
459 removal of the rRNA was key to the success of reading long RNA. Fern chloroplast rRNA removal could be
460 improved still, using specific DNA probes integrated during synthesis of the stranded sequencing libraries.
461 Transcripts from the fern's mitochondrial genome are likely in the data but we have yet to assemble the
462 *A. filiculoides* mitochondrial genome to confirm this.

463 *N. azollae* extracts did not contain photosynthesis pigments absorbing/fluorescing in the far-red spectrum
464 and the *N. azollae* genome did not contain psbA4/Chlf-S like proteins (data not shown), yet it contained
465 phytochromes and cyanobacteriochromes capable of sensing far-red light as other filamentous
466 cyanobacteria do (Wiltbank & Kehoe, 2019). The RNA profiling did not reveal whether the *N. azollae*

467 trigger fern RD since roles of the far-red responsive transcripts in *N. azollae*, particularly those encoding
468 the transporters, have yet to be studied. An approach could be to study these in free living filamentous
469 cyanobacteria amenable to genetic manipulation. Manipulation of the symbiont would, however, be
470 preferred. Tri-parental mating protocols may be effective on *N. azollae in situ* if combined with high-
471 frequency RNA-guided integration with Cas12d enzyme complexes (Strecker *et al.,* 2020). *N. azollae* Cas
472 genes were found to be mostly pseudogenes and CRISPR arrays appeared missing (Cai *et al.,* 2013),
473 suggesting that incoming cargo plasmid may not be destroyed by Cas complexes. Nevertheless, sRNA
474 profiles from Fig. **5** show discrete accumulation of 31 and 42 b sRNA of very low complexity and a
475 functional Cas6 splicing CRISPR precursor RNA was induced on TL with FR (SI Table S**4**, *N. azollae*). CRISPR
476 arrays may not have been recognized, we conclude, if they contained few repeats due to the sheltered
477 life-style inside the fern. Rendering *A. filiculoides* and *N. azollae* amenable to genetic manipulation will be
478 crucial to provide evidence for hypotheses generated by transcript profiling in the future.

479 **Acknowledgements**

480 NWO-ALW EPS grant (ALWGS.2016.020) supported LWD, NWO-TTW grant (Project 16294) supported EG.
481 We are grateful to Bruno Huettel for advice on dual RNA sequencing library preparation methods. We
482 thank Bas Dutilh and Berend Snel for access to the Utrecht University bioinformatic computing resources.

483 **Author Contributions**

484 Fern induction experiments were done by BST and PB, fern crosses and comparisons were done by EG,
485 BST and unnamed graduate students, phylogenetic analyses and miR discovery were carried out by LWD
486 and NR, sRNA and dual RNA profiling was done by HS. HS conceived the manuscript with all authors
487 contributing to editing it.

488

489 **Reference list**

490 **Achard P, Herr A, Baulcombe DC, Harberd NP**. **2004**. Modulation of floral development by a gibberellin-
491 regulatedmicroRNA. *Development* **131**: 3357–3365.

492 **Anders S, Huber W**. **2010**. Differential expression analysis for sequence count data. *Genome Biology* **11**:
493 R106.

494 **Anders S, Pyl PT, Huber W**. **2015**. HTSeq-a Python framework to work with high-throughput sequencing
495 data. *Bioinformatics* **31**: 166–169.

496 **Aukerman MJ, Hirschfeld M, Wester L, Weaver M, Clack T, Amasino RM, Sharrock RA**. **1997**. A deletion
497 in the PHYD gene of the Arabidopsis Wassilewskija ecotype defines a role for phytochrome D in red/far-
498 red light sensing. *The Plant Cell* **9**: 1317–1326.

499 **Axtell MJ, Meyers BC**. **2018**. Revisiting Criteria for Plant MicroRNA Annotation in the Era of Big Data. *The*
500 *Plant Cell* **30**: 272–284.

501 **Aya K, Hiwatashi Y, Kojima M, Sakakibara H, Ueguchi-Tanaka M, Hasebe M, Matsuoka M**. **2011**. The
502 Gibberellin perception system evolved to regulate a pre-existing GAMYB-mediated system during land
503 plant evolution. *Nature Communications* **2**: 544.

504 **Benson-Evans K**. **1961**. Environmental Factors and Bryophytes. *Nature* **191**: 255–260.

505 **Berruezo F, de Souza FSJ, Picca PI, Nemirovsky SI, Martínez Tosar L, Rivero M, Mentaberry AN, Zelada**
506 **AM**. **2017**. Sequencing of small RNAs of the fern Pleopeltis minima (Polypodiaceae) offers insight into
507 the evolution of the microrna repertoire in land plants (T Unver, Ed.). *PLOS ONE* **12**: e0177573.

508 **Bonnet E, Wuyts J, Rouze P, Van de Peer Y**. **2004**. Evidence that microRNA precursors, unlike other non-
509 coding RNAs, have lower folding free energies than random sequences. *Bioinformatics* **20**: 2911–2917.

510 **Bowman JL, Kohchi T, Yamato KT, Jenkins J, Shu S, Ishizaki K, Yamaoka S, Nishihama R, Nakamura Y,**
511 **Berger F, *et al.* 2017**. Insights into Land Plant Evolution Garnered from the Marchantia polymorpha
512 Genome. *Cell* **171**: 287-304.e15.

513 **Brouwer P, Bräutigam A, Buijs VA, Tazelaar AOE, van der Werf A, Schlüter U, Reichart G-J, Bolger A,**
514 **Usadel B, Weber APM, *et al.* 2017**. Metabolic Adaptation, a Specialized Leaf Organ Structure and
515 Vascular Responses to Diurnal N2 Fixation by Nostoc azollae Sustain the Astonishing Productivity of
516 Azolla Ferns without Nitrogen Fertilizer. *Frontiers in Plant Science* **8**.

517 **Brouwer P, Bräutigam A, Külahoglu C, Tazelaar AOE, Kurz S, Nierop KGJ, van der Werf A, Weber APM,**
518 **Schluepmann H**. **2014**. Azolla domestication towards a biobased economy? *New Phytologist* **202**: 1069–
519 1082.

520 **Brouwer P, Schluepmann H, Nierop KGJ, Elderson J, Bijl PK, van der Meer I, de Visser W, Reichart G-J,**
521 **Smeekens S, van der Werf A**. **2018**. Growing Azolla to produce sustainable protein feed: the effect of
522 differing species and CO2 concentrations on biomass productivity and chemical composition. *Journal of*
523 *the Science of Food and Agriculture* **98**: 4759–4768.

524 **Chiyoda S, Ishizaki K, Kataoka H, Yamato KT, Kohchi T**. **2008**. Direct transformation of the liverwort
525 Marchantia polymorpha L. by particle bombardment using immature thalli developing from spores.
526 *Plant Cell Reports* **27**: 1467–1473.

527 **Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009.** trimAl: a tool for automated alignment
528 trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972-3.

529 **Coulter JM, Barnes CR, Cowles HC**. **1910**. *A Textbook of Botany for Colleges and Universities*. London:
530 American Book Company.

531 **Dai X, Zhuang Z, Zhao PX**. **2018**. psRNATarget: a plant small RNA target analysis server (2017 release).
532 *Nucleic Acids Research* **46**: W49–W54.

533 **Dijkhuizen LW, Brouwer P, Bolhuis H, Reichart G-J, Koppers N, Huettel B, Bolger AM, Li F-W, Cheng S,**
534 **Liu X, *et al.* 2018**. Is there foul play in the leaf pocket? The metagenome of floating fern Azolla reveals
535 endophytes that do not fix N 2 but may denitrify. *New Phytologist* **217**: 453–466.

536 **Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR**. **2013**.
537 STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.

538 **Farahpour-Haghani A, Hassanpour M, Alinia F, Nouri-Ganbalani G, Razmjou J, Agassiz D**. **2017**. Water
539 ferns Azolla spp. (Azollaceae) as new host plants for the small China-mark moth, Cataclysta lemnata
540 (Linnaeus, 1758) (Lepidoptera, Crambidae, Acentropinae). *Nota Lepidopterologica* **40**: 1–13.

541 **Gras DE, Vidal EA, Undurraga SF, Riveras E, Moreno S, Dominguez-Figueroa J, Alabadi D, Blázquez MA,**
542 **Medina J, Gutiérrez RA**. **2018**. SMZ/SNZ and gibberellin signaling are required for nitrate-elicited delay
543 of flowering time in Arabidopsis thaliana. *Journal of Experimental Botany* **69**: 619–631.

544 **Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL**. **2008**. The Vienna RNA Websuite. *Nucleic*
545 *Acids Research* **36**: W70–W74.

546 **Güngör E, Brouwer P, Dijkhuizen LW, Shaffar D, Nierop K, de Vos R, Toraño J, van der Meer I,**
547 **Schluepmann H. 2020.** *Azolla* ferns testify: seed plants and ferns share a common ancestor for
548 leucoanthocyanidin reductase enzymes. *New Phytologist in press.*

549 **Guo C, Xu Y, Shi M, Lai Y, Wu X, Wang H, Zhu Z, Poethig RS, Wu G**. **2017**. Repression of miR156 by
550 miR159 Regulates the Timing of the Juvenile-to-Adult Transition in Arabidopsis. *The Plant Cell* **29**: 1293–
551 1304.

552 **Halliday KJ, Koornneef M, Whitelam GC**. **1994**. Phytochrome B and at Least One Other Phytochrome
553 Mediate the Accelerated Flowering Response of Arabidopsis thaliana L. to Low Red/Far-Red Ratio. *Plant*
554 *Physiology* **104**: 1311–1315.

555 **Hasebe M, Wen CK, Kato M, Banks JA. 1998.** Characterization of MADS homeotic genes in the fern
556 Ceratopteris richardii. *Proceedings of the National Academy of Sciences* **95**:6222-6227.

557 **Hisanaga T, Okahashi K, Yamaoka S, Kajiwara T, Nishihama R, Shimamura M, Yamato KT, Bowman JL,**
558 **Kohchi T, Nakajima K**. **2019**. A cis -acting bidirectional transcription switch controls sexual dimorphism
559 in the liverwort. *The EMBO Journal* **38**: e100240.

560 **Hou S, Brenes-Álvarez M, Reimann V, Alkhnbashi OS, Backofen R, Muro-Pastor AM, Hess WR**. **2019**.
561 CRISPR-Cas systems in multicellular cyanobacteria. *RNA Biology* **16**: 518–529.

562 **Hyun Y, Vincent C, Tilmes V, Bergonzi S, Kiefer C, Richter R, Martinez-Gallegos R, Severing E, Coupland**
563 **G**. **2019**. A regulatory circuit conferring varied flowering response to cold in annual and perennial plants.
564 *Science* **363**: 409–412.

565 **Inoue K, Nishihama R, Araki T, Kohchi T**. **2019**. Reproductive Induction is a Far-Red High Irradiance
566 Response that is Mediated by Phytochrome and PHYTOCHROME INTERACTING FACTOR in Marchantia
567 polymorpha. *Plant and Cell Physiology* **60**: 1136–1145.

568 **Jiang C-K, Rao G-Y**. **2020**. Insights into the Diversification and Evolution of R2R3-MYB Transcription
569 Factors in Plants. *Plant Physiology* **183**: 637–655.

570 **Kaneko M, Inukai Y, Ueguchi-Tanaka M, Itoh H, Izawa T, Kobayashi Y, Hattori T, Miyao A, Hirochika H,**
571 **Ashikari M, *et al.* 2004**. Loss-of-Function Mutations of the Rice GAMYB Gene Impair α-Amylase
572 Expression in Aleurone and Flower Development. *The Plant Cell* **16**: 33–44.

573 **Katoh K, Rozewicki J, Yamada KD**. **2019.** MAFFT online service: multiple sequence alignment, interactive
574 sequence choice and visualization. *Briefings in bioinformatics* **20**: 1160-6.

575 **Klose C, Nagy F, Schäfer E**. **2020**. Thermal Reversion of Plant Phytochromes. *Molecular Plant*.

576 **Kozomara A, Birgaoanu M, Griffiths-Jones S**. **2019**. miRBase: from microRNA sequences to function.
577 *Nucleic Acids Research* **47**: D155–D162.

578 **Kuang Z, Wang Y, Li L, Yang X**. **2019**. miRDeep-P2: accurate and fast analysis of the microRNA
579 transcriptome in plants (I Birol, Ed.). *Bioinformatics* **35**: 2521–2522.

580 **Kubota A, Kita S, Ishizaki K, Nishihama R, Yamato KT, Kohchi T**. **2014**. Co-option of a photoperiodic
581 growth-phase transition system during land plant evolution. *Nature Communications* **5**: 3668.

582 **Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL**. **2004**. Versatile and
583 open software for comparing large genomes. *Genome biology* **5**: R12.

584 **Labouriau LG**. **1958**. Studies on the initiation of sporangia in ferns.

585 **Leebens-Mack JH, Barker MS, Carpenter EJ, Deyholos MK, Gitzendanner MA, Graham SW, Grosse I, Li**
586 **Z, Melkonian M, Mirarab S, *et al.* 2019**. One thousand plant transcriptomes and the phylogenomics of
587 green plants. *Nature* **574**: 679–685.

588 **Legris M, Nieto C, Sellaro R, Prat S, Casal JJ**. **2017**. Perception and signalling of light and temperature
589 cues in plants. *The Plant Journal* **90**: 683–697.

590 **Li F-W, Brouwer P, Carretero-Paulet L, Cheng S, de Vries J, Delaux P-M, Eily A, Koppers N, Kuo L-Y, Li Z,**
591 ***et al.* 2018**. Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nature Plants* **4**:
592 460–472.

593 **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R**. **2009**. The
594 Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

595 **Li F-W, Melkonian M, Rothfels CJ, Villarreal JC, Stevenson DW, Graham SW, Wong GK-S, Pryer KM,**
596 **Mathews S**. **2015**. Phytochrome diversity in green plants and the origin of canonical plant
597 phytochromes. *Nature Communications* **6**: 7852.

598 **Lohse M, Nagel A, Herter T, May P, Schroda M, Zrenner R, Tohge T, Fernie AR, Stitt M, Usadel B**. **2014**.
599 Mercator: a fast and simple web server for genome scale functional annotation of plant sequence data.
600 *Plant, Cell & Environment* **37**: 1250–1258.

601 **Love MI, Huber W, Anders S**. **2014.** Moderated estimation of fold change and dispersion for RNA-seq
602 data with DESeq2. *Genome biology* **15**: 550.

603 **Lucero LE, Manavella PA, Gras DE, Ariel FD, Gonzalez DH**. **2017**. Class I and Class II TCP Transcription
604 Factors Modulate SOC1-Dependent Flowering at Multiple Levels. *Molecular Plant* **10**: 1571–1574.

605 **Madeira PT, Hill MP, Dray FA, Coetzee JA, Paterson ID, Tipping PW**. **2016**. Molecular identification of

Azolla invasions in Africa: The Azolla specialist, Stenopelmus rufinasus proves to be an excellent taxonomist. *South African Journal of Botany* **105**: 299–305.

Martinez-Zapater JM, Somerville CR. **1990**. Effect of Light Quality and Vernalization on Late-Flowering Mutants of Arabidopsis thaliana. *Plant Physiology* **92**: 770–776.

Millar AA, Lohe A, Wong G. **2019**. Biology and Function of miR159 in Plants. *Plants* **8**: 255.

Nagalingum NS, Schneider H, Pryer KM. **2006**. Comparative Morphology of Reproductive Structures in Heterosporous Water Ferns and a Reevaluation of the Sporocarp. *International Journal of Plant Sciences* **167**: 805–815.

Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution* **32**: 268-74.

Nishihama R, Ishizaki K, Hosaka M, Matsuda Y, Kubota A, Kohchi T. **2015**. Phytochrome-mediated regulation of cell division and growth during regeneration and sporeling development in the liverwort Marchantia polymorpha. *Journal of Plant Research* **128**: 407–421.

Nobori T, Velásquez AC, Wu J, Kvitko BH, Kremer JM, Wang Y, He SY, Tsuda K. **2018**. Transcriptome landscape of a bacterial pathogen under plant immunity. *Proceedings of the National Academy of Sciences* **115**: E3055–E3064.

Olas JJ, Van Dingenen J, Abel C, Działo MA, Feil R, Krapp A, Schlereth A, Wahl V. **2019**. Nitrate acts at the Arabidopsis thaliana shoot apical meristem to regulate flowering time. *New Phytologist* **223**: 814–827.

Olm MR, Brown CT, Brooks B, Banfield JF. **2017**. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME Journal* **11**: 2864–2868.

Kar P, Mishra S, Singh D. **2002**. *Azolla* sporulation in response to application of some selected auxins and their combination with gibberellic acid. *Biology and Fertility of Soils* **35**: 314–319.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. *Genome biology* **5**:R12.

Palatnik JF, Wollmann H, Schommer C, Schwab R, Boisbouvier J, Rodriguez R, Warthmann N, Allen E, Dezulian T, Huson D, *et al.* 2007. Sequence and Expression Differences Underlie Functional Specialization of Arabidopsis MicroRNAs miR159 and miR319. *Developmental Cell* **13**: 115–125.

Perkins SK, Peters GA. **2006**. The Azolla-Anabaena symbiosis: endophyte continuity in the Azolla life-cycle is facilitated by epidermal trichomes †. *New Phytologist* **123**: 53–64.

Plackett ARG, Conway SJ, Hewett Hazelton KD, Rabbinowitsch EH, Langdale JA, Di Stilio VS. **2018**. LEAFY maintains apical stem cell activity during shoot development in the fern *Ceratopteris richardii*. *eLife* **7**.

Possart A, Hiltbrunner A. **2013**. An evolutionarily conserved signaling mechanism mediates far-red light responses in land plants. *The Plant Cell* **25**: 102–114.

Quammen D. 2018. The tangled tree: a radical new history of life. Simon & Schuster, New York.

643 **Ran L, Larsson J, Vigil-Stenman T, Nylander JAA, Ininbergs K, Zheng W-W, Lapidus A, Lowry S,**
644 **Haselkorn R, Bergman B**. **2010**. Genome erosion in a nitrogen-fixing vertically transmitted
645 endosymbiotic multicellular cyanobacterium (N Ahmed, Ed.). *PLoS ONE* **5**: e11486.

646 **Rolauffs S, Fackendahl P, Sahm J, Fiene G, Hoecker U**. **2012**. Arabidopsis COP1 and SPA genes are
647 essential for plant elongation but not for acceleration of flowering time in response to a low red light to
648 far-red light ratio. *Plant Physiology* **160**: 2015–2027.

649 **Sánchez-Retuerta C, Suaréz-López P, Henriques R**. **2018**. Under a New Light: regulation of light-
650 dependent pathways by non-coding RNAs. *Frontiers in Plant Science* **9**.

651 **Sayou C, Nanao MH, Jamin M, Posé D, Thévenon E, Grégoire L, Tichtinsky G, Denay G, Ott F, Peirats**
652 **Llobet M, *et al.* 2016**. A SAM oligomerization domain shapes the genomic binding landscape of the
653 LEAFY transcription factor. *Nature Communications* **7**: 11222.

654 **Schwacke R, Ponce-Soto GY, Krause K, Bolger AM, Arsova B, Hallab A, Gruden K, Stitt M, Bolger ME,**
655 **Usadel B**. **2019**. MapMan4: A Refined Protein Classification and Annotation Framework Applicable to
656 Multi-Omics Data Analysis. *Molecular Plant* **12**: 879–892.

657 **Sessa EB, Der JP**. **2016**. Evolutionary genomics of ferns and lycophytes. In: Advances in Botanical
658 Research. 215–254.

659 **Shimmura S, Nozoe M, Kitora S, Kin S, Matsutani S, Ishizaki Y, Nakahira Y, Shiina T**. **2017**. Comparative
660 Analysis of Chloroplast psbD Promoters in Terrestrial Plants. *Frontiers in Plant Science* **8**.

661 **Song YH, Kubota A, Kwon MS, Covington MF, Lee N, Taagen ER, Laboy Cintrón D, Hwang DY, Akiyama**
662 **R, Hodge SK, *et al.* 2018**. Molecular basis of flowering under natural long-day conditions in Arabidopsis.
663 *Nature Plants* **4**: 824–835.

664 **Srivastava PK, Moturu T, Pandey P, Baldwin IT, Pandey SP**. **2014**. A comparison of performance of plant
665 miRNA target prediction tools and the characterization of features for genome-wide target prediction.
666 *BMC Genomics* **15**: 348.

667 **Strecker J, Ladha A, Gardner Z, Schmid-Burgk JL, Makarova KS, Koonin E V., Zhang F**. **2019**. RNA-guided
668 DNA insertion with CRISPR-associated transposases. *Science* **365**: 48–53.

669 **Sun Z, Li M, Zhou Y, Guo T, Liu Y, Zhang H, Fang Y**. **2018**. Coordinated regulation of Arabidopsis
670 microRNA biogenesis and red light signaling through Dicer-like 1 and phytochrome-interacting factor 4
671 (L-J Qu, Ed.). *PLOS Genetics* **14**: e1007247.

672 **Svensson ME, Johannesson H, Engström P**. **2000**. The LAMB1 gene from the clubmoss, Lycopodium
673 annotinum, is a divergent MADS-box gene, expressed specifically in sporogenic structures. *Gene* **253**:
674 31–43.

675 **Theißen G, Rümpler F, Gramzow L**. **2018**. Array of MADS-Box Genes: Facilitator for Rapid Adaptation?
676 *Trends in Plant Science* **23**: 563–576.

677 **Tsuzuki M, Futagami K, Shimamura M, Inoue C, Kunimoto K, Oogami T, Tomita Y, Inoue K, Kohchi T,**
678 **Yamaoka S, *et al.* 2019**. An early arising role of the MicroRNA156/529-SPL module in reproductive
679 development revealed by the liverwort marchantia polymorpha. *Current Biology* **29**: 3307-3314.e5.

680 **Vasco A, Smalls TL, Graham SW, Cooper ED, Wong GK-S, Stevenson DW, Moran RC, Ambrose BA**. **2016**.
681 Challenging the paradigms of leaf evolution: Class III HD-Zips in ferns and lycophytes. *New Phytologist*
682 **212**: 745–758.

683 **Vicherová E, Glinwood R, Hájek T, Šmilauer P, Ninkovic V**. **2020**. Bryophytes can recognize their
684 neighbours through volatile organic compounds. *Scientific Reports* **10**: 7405.

685 **de Vries A, Ripley BD**. **2016**. Package 'ggdendro': Create Dendrograms and Tree Diagrams Using
686 'ggplot2' (R package). *R package version 0.1-20*.

687 **Wickham H**. **2011**. ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics* **3**: 180–185.

688 **Wiltbank LB, Kehoe DM**. **2019**. Diverse light responses of cyanobacteria mediated by phytochrome
689 superfamily photoreceptors. *Nature Reviews Microbiology* **17**: 37–50.

690 **Xie Y, Zhou Q, Zhao Y, Li Q, Liu Y, Ma M, Wang B, Shen R, Zheng Z, Wang H. 2020.** FHY3 and FAR1
691 Integrate Light Signals with the miR156-SPL Module-Mediated Aging Pathway to Regulate Arabidopsis
692 Flowering. *Molecular Plant* **13**:483-98.

693 **Yamaoka S, Nishihama R, Yoshitake Y, Ishida S, Inoue K, Saito M, Okahashi K, Bao H, Nishida H,**
694 **Yamaguchi K, *et al.* 2018**. Generative Cell Specification Requires Transcription Factors Evolutionarily
695 Conserved in Land Plants. *Current Biology* **28**: 479-486.e5.

696 **Yeh K**. **1997**. A Cyanobacterial Phytochrome Two-Component Light Sensory System. *Science* **277**: 1505–
697 1508.

698 **You C, Cui J, Wang H, Qi X, Kuo L-Y, Ma H, Gao L, Mo B, Chen X**. **2017**. Conservation and divergence of
699 small RNA pathways and microRNAs in land plants. *Genome Biology* **18**: 158.

700 **Zhao T, Holmer R, de Bruijn S, Angenent GC, van den Burg HA, Schranz ME**. **2017**. Phylogenomic
701 Synteny Network Analysis of MADS-Box Transcription Factor Genes Reveals Lineage-Specific
702 Transpositions, Ancient Tandem Duplications, and Deep Positional Conservation. *The Plant Cell* **29**:
703 1278–1292.

704 **Zhu Q, Fisher S, Shallcross J, Kim J**. **2016**. VERSE: a versatile and efficient RNA-Seq read counting tool.

705

706

Table 1 Sporeling counts obtained from random crosses of *A. filiculoides* strains collected in The Netherlands and Spain (Gran Canaria).

Megaspores:	Galgenwaard	Krommerijn	Hoogwoud	Gran Canaria	Den Bosch	Nijmegen	Nieuwebrug	Groningen
Massulae :								
Galgenwaard	>1000	17		1		13		2
Krommerijn	1				16		2	
Hoogwoud			20					2
Gran Canaria*	17				>40	1		
Den Bosch		9		1				
Nijmegen		12		33			2	
Nieuwebrug		6	3			22		
Groningen			1				5	

*crosses involving megaspores from Gran Canaria reproducibly germinated late, generally after 5 weeks instead of 2-3 weeks.

715 **Table 2** Transcription factors with highest changes of read counts in sporophytes after a week on TL with
716 versus without far-red LED.

717

A. filiculoides locus	baseMean	Log2Fold Change	DESeq2 padj	Mercator 4.0 annotation	Closest Arabidopsis Homolog
Azfi_s0028.g024032	**9**	**7.22**	**0.001**	**15.5.14 MADS/AGL**	**AGL20/SOC1, MIKC^{C}**
Azfi_s0113.g045874	155	6.41	0.002	15.5.2 R2R3MYB	AtMYB 32 (AT4G34990.1), R2R3MYB VIII-E*
Azfi_s0083.g038807	59	5.75	0	15.5.32 Basic Helix-Loop-Helix	EDA33, IND1, INDEHISCENT (AT4G00120.1)
Azfi_s0003.g007560	8	5.27	0.011	15.5.32 Basic Helix-Loop-Helix	FIT1 (regulates iron transport)
Azfi_s0003.g007559	34	3.49	0	15.5.32 Basic Helix-Loop-Helix	FIT1 (regulates iron transport)
Azfi_s0096.g043715	31	3.13	0.001	15.5.7 DREB subfamily A-2 of ERF/AP2	AT5G18450.1
Azfi_s0003.g007710	1778	2.82	0	**15.5.14 MADS/AGL**	**AGL6, MIKC^{C}**
Azfi_s0015.g013719	111	2.57	0	15.5.2 G2-like, GARP	HHO2, NIGT1.2
Azfi_s0112.g045798	66	2.2	0	15.5.51.1 NF-Y component NF-YA	AT5G12840.4
Azfi_s0014.g013539	105	2.07	0.005	**15.5.2 R2R3MYB**	**LOF2, R2R3MYB V***
Azfi_s0015.g014012	165	1.99	0.026	15.5.17 NAC domain	NAC025
Azfi_s0004.g008455	**103**	**1.93**	**0.003**	**15.5.2 R2R3MYB**	**miRNA319 controlled GAMYB 33, R2R3MYB VII***
Azfi_s0132.g049213	63	-2.35	0.004	15.5.32 Basic Helix-Loop-Helix	No hits
Azfi_s0182.g056462	10	-4.46	0.049	15.5.3 HD-ZIP I/II	HB16

718 *R2R3 MYB classification according to Jiang & Rao, 2020.

719

720 **Figure legends**

721 **Figure 1** Far-red light and nitrogen affect the reproductive phase transition in *A. filiculoides*. (a)
722 Vegetatively growing ferns. (b) Ferns in the reproductive phase with micro (mi) and megaspores (me). (c)
723 growth habit under TL and TL with far-red LED (FR). (d) A mature sporophyte and (e) its schematic
724 representation: a root is formed at each branch point, and sporocarp pairs are found close-by on the
725 branch side of the branch point but not at all branch points. (f) Sporulation frequencies at increasing plant
726 density (41.7, 128.8 and 208.3 g m^{-2} dry weight) and with increasing red to far-red light ratios (FR0-FR4).
727 The red to far-red ratios were 19.13 (FR0), 1.16 (FR1), 0.63 (FR2), 0.50 (FR3) and 0.31 (FR4); to obtain the
728 FR4 ratio, the TL intensity was halved. Each data point represents the average of three independent
729 continuous growth replicates. (g) Induction of sporulation without nitrogen (-N) or with 2 mM NH_4NO_3
730 (+N) in the medium, data are averages from three replicates with standard deviations.

731 **Figure 2** Chloroplast marker regions in strains from the Netherlands, Spain and Iran, and the rRNA
732 intergenic regions of the *A. filiculoides* genome. Phylogenetic trees of chloroplast regions trnL-trnF (a) and
733 trnG-trnR (b). The maximum likelihood trees were bootstrapped 1000x. All bootstrap values >0.70 are
734 displayed. In bold: sequences from accessions in this study labelled after their collection site and
735 sequences from the differing species extracted from genome shot-gun sequencing (Li *et al.,* 2018). In
736 regular type: sequences from the Madeira et al. (2016) study. (c), Sequence logo of all ITS1 gene regions
737 in the *A. filiculoides* genome (Li *et al.,* 2018). The gene regions were extracted and aligned (MEGA 7) then
738 used for *in-silico* PCR. Results were visualized using WebLogo (vs 2.8.2 Crooks *et al.,* 2004): at each position
739 of the ITS, the height of the stack indicates the sequence conservation, while the height of symbols within
740 the stack indicates the relative frequency of each base.

741 **Figure 3** Profiling eukaryotic and prokaryotic RNA simultaneously in *A. filiculoides* (dual RNAseq). (a)
742 Similarity of the *N. azollae* 0708 genome and Metagenome Assembled Genomes (MAGs) from *N. azollae*
743 strains in *Azolla* species. Average Nucleotide Identity (ANI) was calculated with the dRep implementation
744 of nucmer using the ANImf preset (Kurtz *et al.*, 2004; Olm *et al.*, 2017). Pairwise average ANI is shown in
745 an UPGMA dendrogram (de Vries & Ripley, 2016; Wickham, 2011). Aligned MAGS covered over 90% of
746 the reference *N. azollae* 0708. (b) Proportion (percent) of uniquely aligned PE reads (dark color) and PE
747 reads aligning at multiple loci (light color) on the genomes of fern (green), fern chloroplast (purple) and
748 *N. azollae 0708* (red). PE reads were from samples of sporophytes grown for a week on TL-light (replicates
749 T1, T2, T3) or TL light with far-red LED (replicates F1, F2, F3). STAR-aligner was used in PE mode. (c) Yields
750 of PE uniquely mapped compared with PE read counts on features and PE reads that STAR aligned

751 ambiguously to several features. Features were specified by current general features files (GFF) of each
752 of the fern (green), chloroplast (purple) and *N.azollae* 0708 (red) genomes.

753 **Figure 4** sRNA size distribution and diversity for each of the genomes in the *A. filiculoides* symbiosis. sRNA
754 reads were mapped with 0 mismatches, no overhang, to the concatenated genomes; bam2fasta was then
755 used to extract sRNA sequences aligning per genome, the fasta files obtained were then collapsed to
756 compute read counts for each sRNA, the lists obtained were then joined to a read count matrix imposing
757 at least 1 read per sRNA in each of the three replicate sample types: f (far-red), t (no far-red) and fn (far-
758 red and nitrogen in the medium). sRNA counts were then normalized to the total read counts per sample
759 for that genome. Fern nucleus. Chloroplast, included samples from sporophytes without cyanobacteria in
760 the analysis. *N. azollae*, read count files joined imposing at least 10 read per sRNA in every sample

761 **Figure 5** The miRNA 172α locus in *A.filiculoides*. Alignments at the Azfi_miRNA172α locus are visualized
762 using the Integrative Genomics Viewer (Thorvaldsdóttir *et al.,* 2013); S, read summary; i, individual reads
763 (a) Reads from pre-miRNA at the miR172α locus. RNAseq all conditions, all reads pooled from the diel
764 RNAseq experiment in (Brouwer *et al.,* 2017) . (b) Reads of the miR172 in sRNA libraries including those
765 of ferns without *N.azollae* (sRNAseq-*N.azollae*), ferns grown with and without far-red LED (sRNAseq+FR,
766 sRNAseq TL-FR). (c) Reads of the miR172*. (d) Azfi_miRNA172α hairpin folded using Vienna RNAfold
767 (Gruber *et al.,* 2008; predicts -104,07 kcal mol^{-1} at 22 ^{o}C upon folding); miRNA (purple) and miRNA* (pink)
768 stem arms are shown with 3'overhangs.

769 **Figure 6** Conserved and far-red light responsive miRNA in the fern *A. filiculoides*. (a) conserved miRNA in
770 the *A.filiculoides* fern compared to other land plant lineages. miRNA from the fern were predicted
771 computationally by miRDEEP-P2 using sRNA seq reads then curated manually, the remainder of the table
772 was from Axtell & Meyers, 2018. (b), miRNA with altered steady state in response to far-red light. Shown
773 are average read counts per million 20-22 nt reads in samples of sporophytes after one week on TL light
774 with far-red LED (f) or without (t); raw read counts were submitted to DESeq2 for statistical analyses
775 yielding P-adjusted values: * indicates Padj <0.1; ^ indicates Padj <0.16.

776 **Figure 7** R2R3MYB phylogenetic analyses and fold change comparing sporophytes with and without FR
777 LED. Sequences were extracted from the genome browsers of each species, aligned with
778 MAFFT linsi or einsi, then trimmed with trimAL. Phylogenetic inferences were computed with IQtree and
779 its internal model fitter; non-parametric bootstrap values are shown. *Chara autralis* (Ca), *Chara braunii*
780 (Cb), *Chara leptospora* (Cl), *Marchantia polymorpha* (Mp), *Selaginella moellendorfii* (Sm), *Azolla*

781 *filiculoides* (Azfi),), *Salvinia cuculata* (Sc), *Picea abies* (Pa), *Amborella trichopoda* (Atri), *Arabidopsis*
782 *thaliana* (At). Sequence names are color coded after the R2R3 MYB subfamilies defined in Jiang & Rao
783 (2020). Fold change in response to FR was calculated by DESeq2, green stars mark significant changes with
784 Padj <0.1.

Figure 1 Far-red light and nitrogen affect the reproductive phase transition in *A. filiculoides*. (a) Vegetatively growing ferns. (b) Ferns in the reproductive phase with micro (mi) and megaspores (me). (c) growth habit under TL and TL with far-red LED (FR). (d) A mature sporophyte and (e) its schematic representation: a root is formed at each branch point, and sporocarp pairs are found close-by on the branch side of the branch point but not at all branch points. (f) Sporulation frequencies at increasing plant density (41.7, 128.8 and 208.3 g m^{-2} dry weight) and with increasing red to far-red light ratios (FR0-FR4). The red to far-red ratios were 19.13 (FR0), 1.16 (FR1), 0.63 (FR2), 0.50 (FR3) and 0.31 (FR4); to obtain the FR4 ratio, the TL intensity was halved. Each data point represents the average of three independent continuous growth replicates. (g) Induction of sporulation without nitrogen (-N) or with 2 mM NH_4NO_3 (+N) in the medium, data are averages from three replicates with standard deviations.

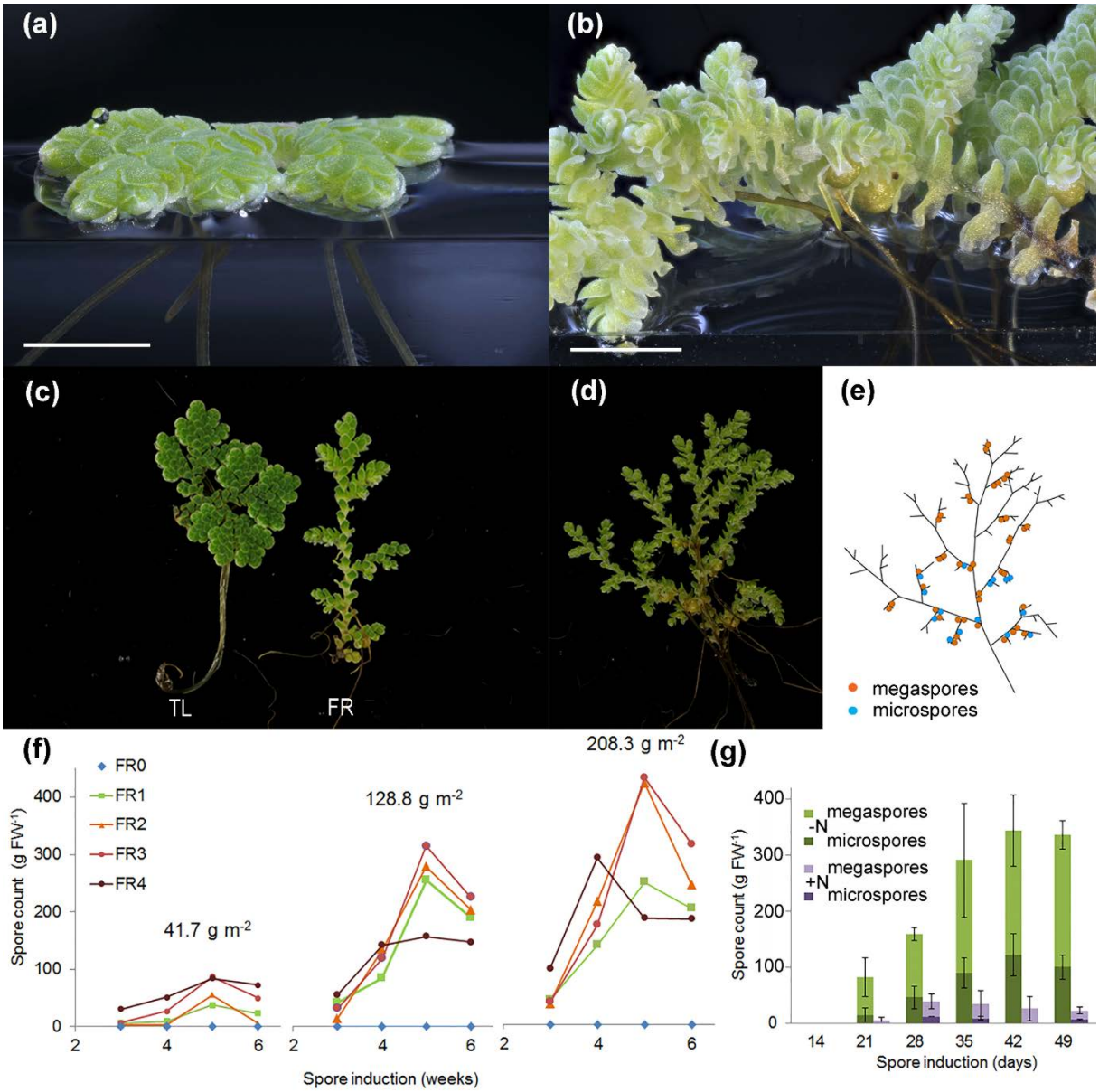
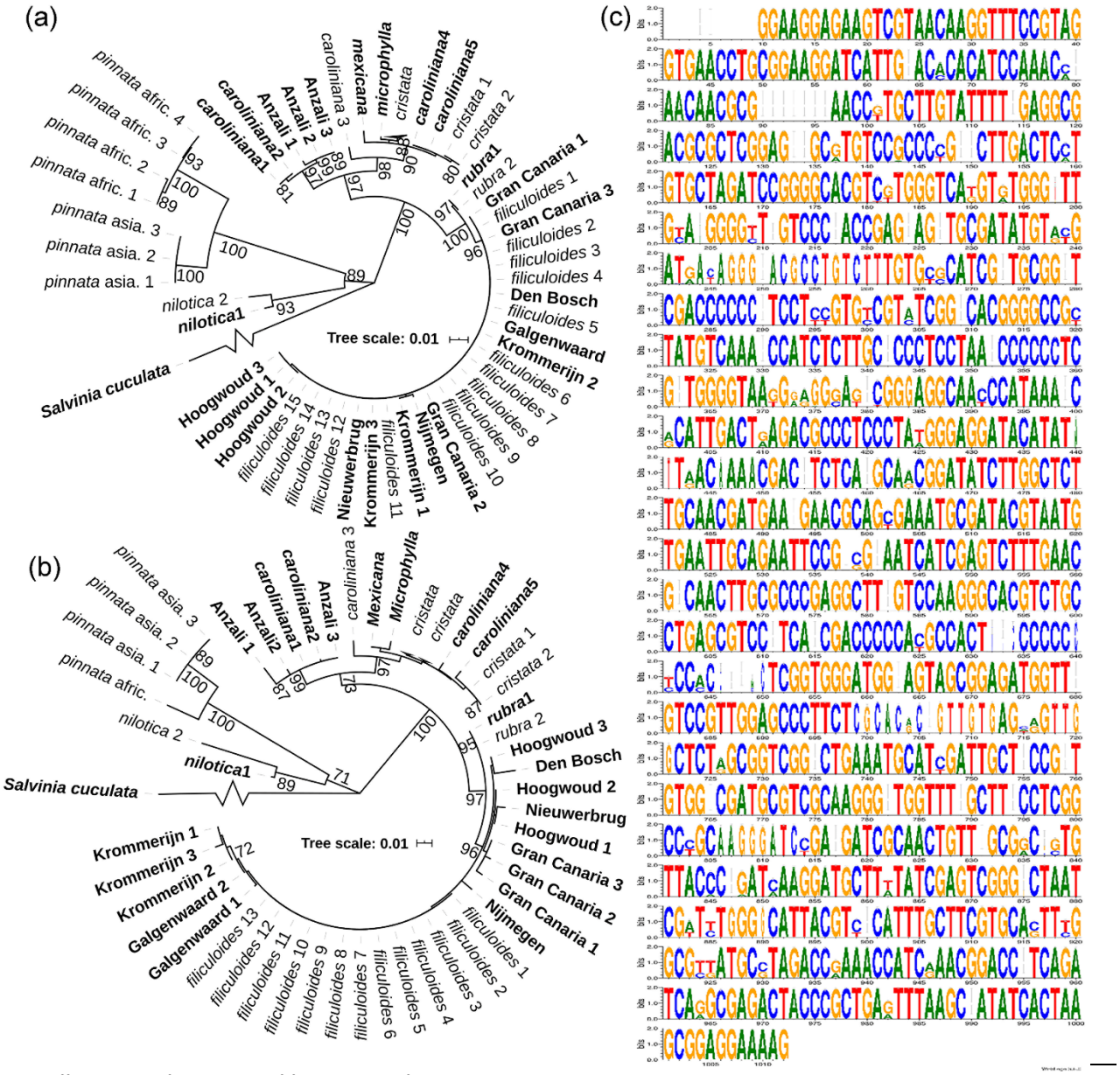
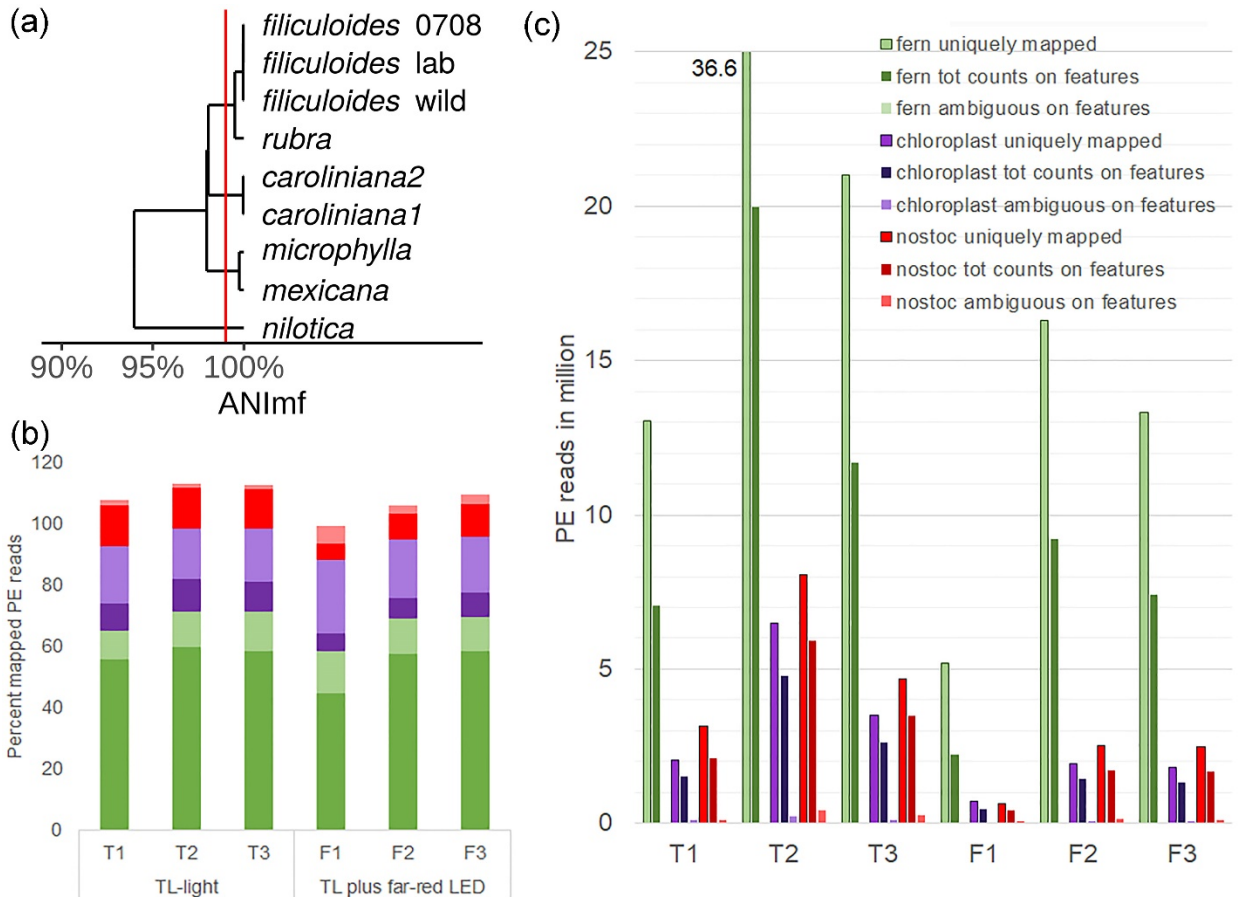


Figure 2 Chloroplast marker regions in strains from the Netherlands, Spain and Iran, and the rRNA intergenic regions of the *A. filiculoides* genome. Phylogenetic trees of chloroplast regions trnL-trnF (a) and trnG-trnR (b). The maximum likelihood trees were bootstrapped 1000x. All bootstrap values >0.70 are displayed. In bold: sequences from accessions in this study labelled after their collection site and sequences from the differing species extracted from genome shot-gun sequencing (Li *et al.,* 2018). In regular type: sequences from Madeira et al. (2016). (c) Sequence logo of all ITS1 gene regions in the *A. filiculoides* genome (Li *et al.,* 2018). The ITS1 sequences were extracted and aligned, then alignments visualized using WebLogo (vs 2.8.2 Crooks *et al.,* 2004): at each position of the ITS1, the height of the stack indicates the sequence conservation, while the height of symbols within the stack indicates the relative frequency of each base.



808 **Figure 3** Profiling eukaryotic and prokaryotic RNA simultaneously in *A. filiculoides* (dual RNAseq). (a)
809 Similarity of the *N. azollae* 0708 genome and Metagenome Assembled Genomes (MAGs) from *N. azollae*
810 strains in *Azolla* species. Average Nucleotide Identity (ANI) was calculated with the dRep implementation
811 of nucmer using the ANImf preset (Kurtz *et al.*, 2004; Olm *et al.*, 2017). Pairwise average ANI is shown in
812 an UPGMA dendrogram (de Vries & Ripley, 2016; Wickham, 2011). Aligned MAGS covered over 90% of
813 the reference *N. azollae* 0708. (b) Proportion (percent) of uniquely aligned PE reads (dark color) and PE
814 reads aligning at multiple loci (light color) on the genomes of fern (green), fern chloroplast (purple) and
815 *N. azollae 0708* (red). PE reads were from samples of sporophytes grown for a week on TL-light (replicates
816 T1, T2, T3) or TL light with far-red LED (replicates F1, F2, F3). STAR-aligner was used in PE mode. (c) Yields
817 of PE uniquely mapped compared with PE read counts on features and PE reads that STAR aligned
818 ambiguously to several features. Features were specified by current general features files (GFF) of each
819 of the fern (green), chloroplast (purple) and *N.azollae* 0708 (red) genomes.

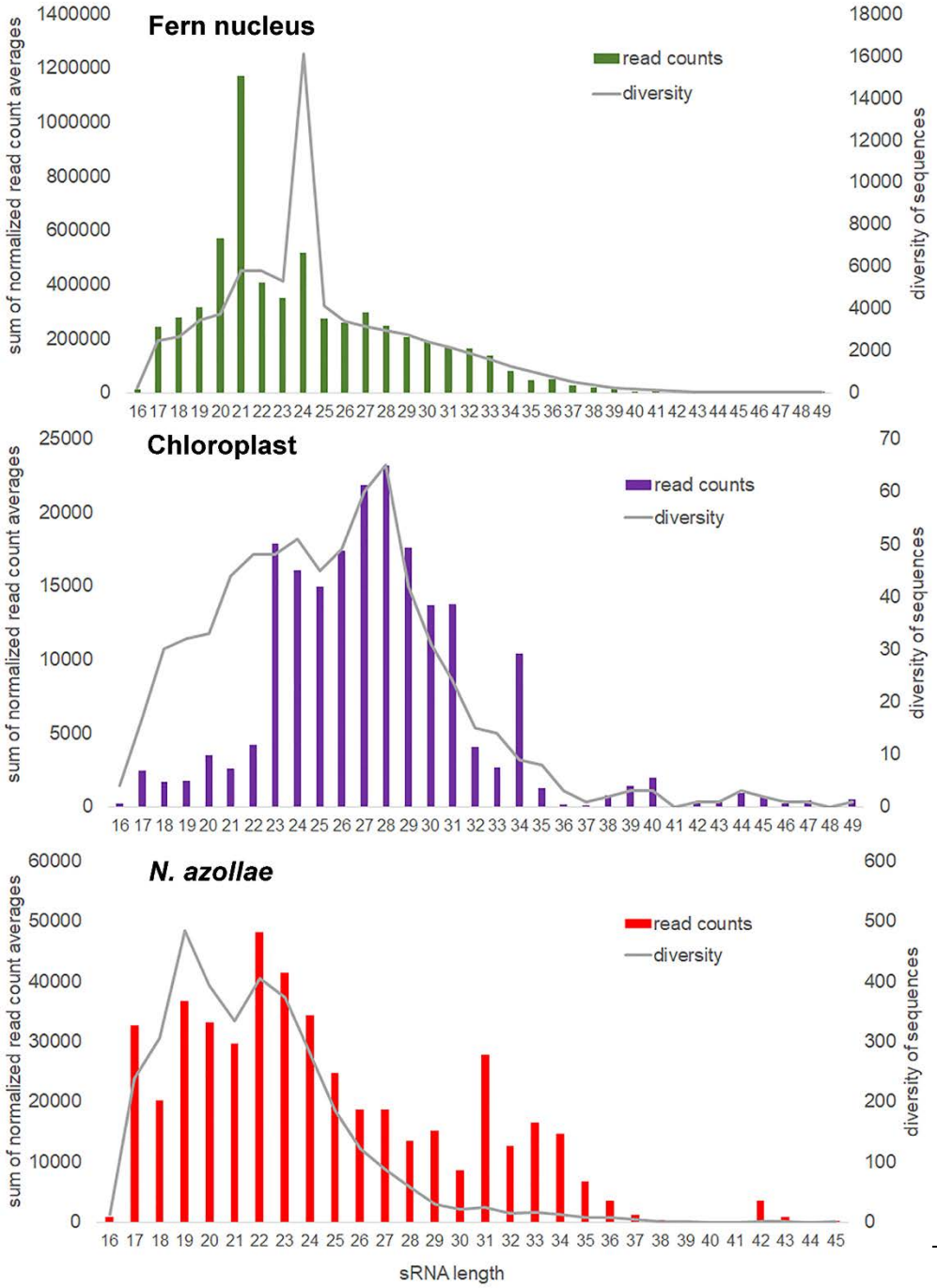
820



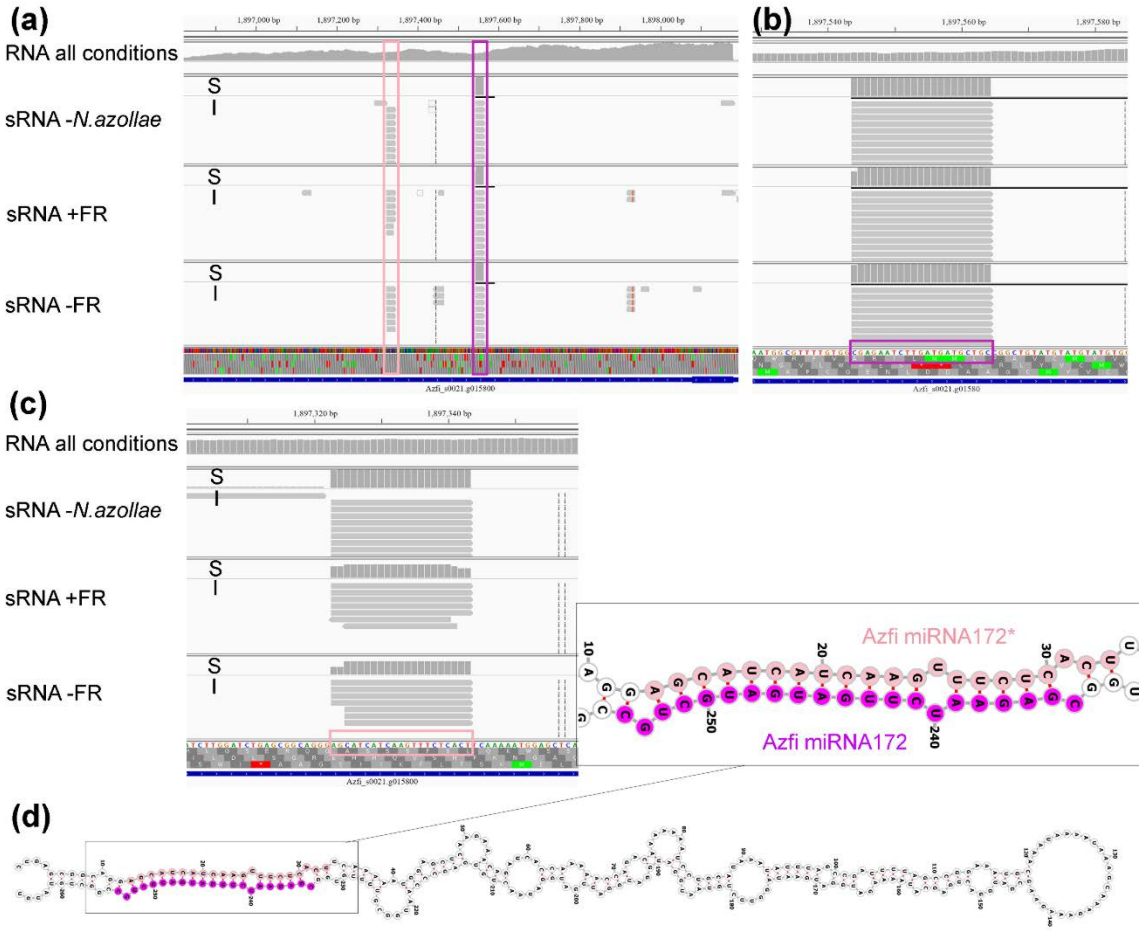
821 **Figure 4** sRNA size distribution and diversity for each of the genomes in the *A. filiculoides* symbiosis. sRNA
822 reads were mapped, imposing zero mismatches and no overhang, to the concatenated genomes;
823 bam2fasta was then used to extract sRNA sequences aligning per genome, the fasta files obtained were
824 then collapsed to compute read counts for each sRNA, the lists obtained were then joined to a read count
825 matrix imposing at least 1 read per sRNA in each of the three replicate sample types: f (far-red), t (no far-
826 red) and fn (far-red and nitrogen in the medium). sRNA counts were then normalized to the total read
827 counts per sample for that genome. Fern nucleus. Chloroplast, included samples from sporophytes
828 without cyanobacteria in the analysis. *N. azollae*, read count files joined imposing at least 10 read per
829 sRNA in e

830

831



832 **Figure 5** The miRNA 172α locus in *A.filiculoides*. Alignments at the Azfi_miRNA172α locus were visualized
833 using the Integrative Genomics Viewer (Thorvaldsdóttir *et al.,* 2013); S, read summary; i, individual reads
834 (a) Reads from pre-miRNA at the miR172α locus. RNAseq all conditions, all reads pooled from the diel
835 RNAseq experiment in (Brouwer *et al.,* 2017) . (b) Reads of the miR172 in sRNA libraries including those
836 of ferns without *N.azollae* (sRNAseq-*N.azollae*), ferns grown with and without far-red LED (sRNAseq+FR,
837 sRNAseq TL-FR). (c) Reads of the miR172*. (d) Azfi_miRNA172α hairpin folded using Vienna RNAfold
838 (Gruber *et al.,* 2008; predicts -104,07 kcal mol^{-1} at 22 °C upon folding); miRNA (purple) and miRNA* (pink)
839 stem arms are shown with 3'overhangs.



840

841

842 **Figure 6** Conserved and far-red light responsive miRNA in the fern *A. filiculoides*. (a) Conserved miRNA in
843 the *A.filiculoides* fern compared to other land plant lineages. miRNA from the fern were predicted
844 computationally by miRDEEP-P2 using sRNA seq reads then curated manually, the remainder of the table
845 was from Axtell & Meyers, 2018. (b) miRNA in *A. filiculoides* sporophytes responsive to far-red light. Shown
846 are average read counts per million 20-22 nt reads in samples of sporophytes after one week on TL light
847 with far-red LED (sRNA +FR) or without (sRNA -FR); raw read counts were submitted to DESeq2 for
848 statistical analyses yielding P-adjusted values: * indicates Padj <0.1; ^ indicates Padj <0.16.

849

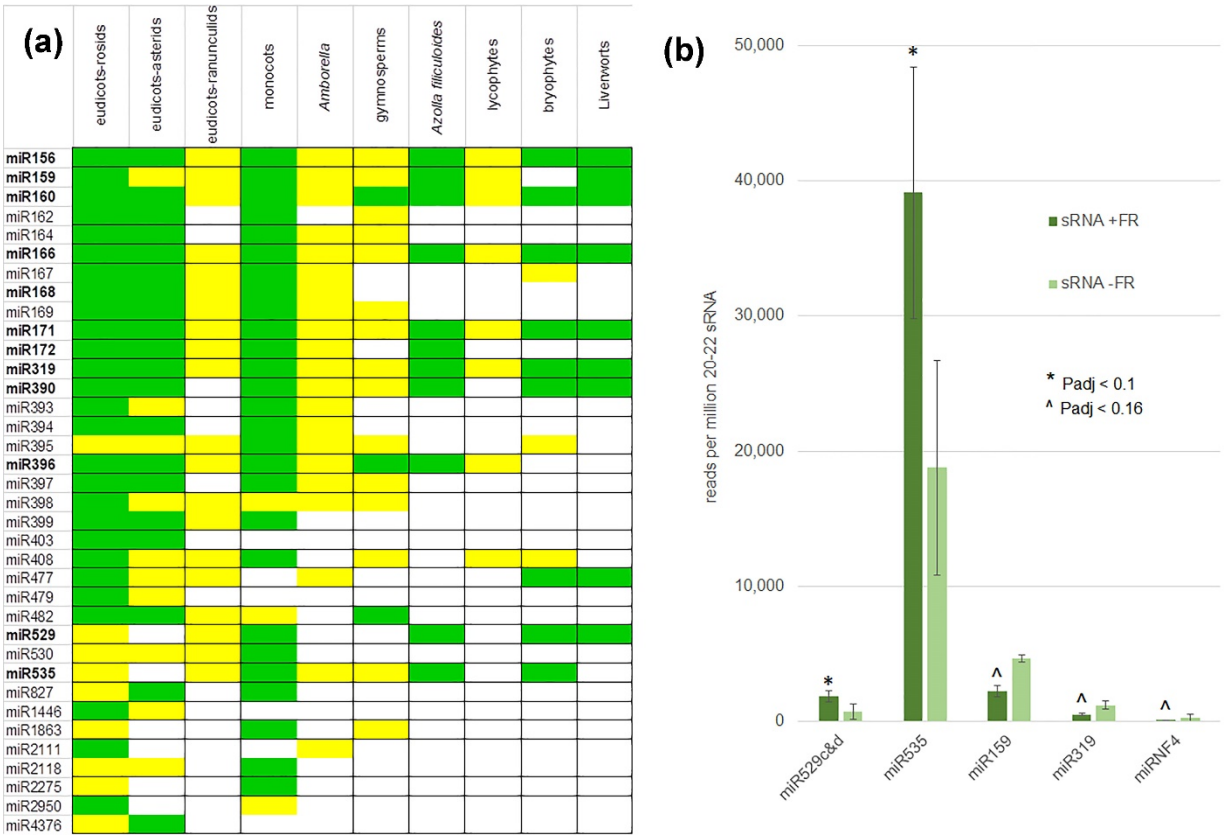
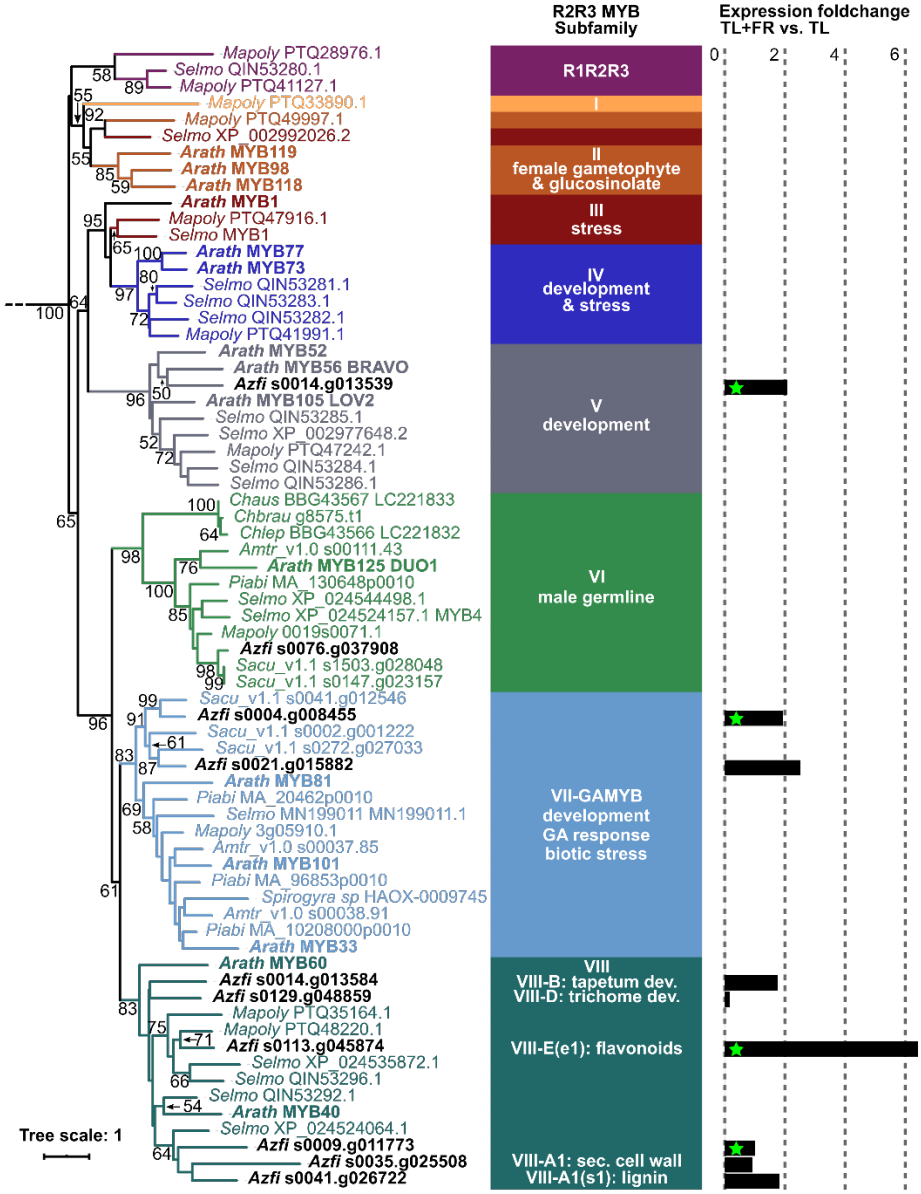



Figure 7 *Azolla* R2R3MYB phylogenetic analysis and response to far-red light (FR). Sequences extracted from the genome browsers of each species were aligned with MAFFT linsi or einsi (Katoh *et al.*, 2019), then trimmed with trimAL (Capella-Gutiérrez *et al.*, 2009). Phylogenetic inferences were computed with IQtree (Nguyen *et al.*, 2015) and its internal model fitter; non-parametric bootstrap values are shown. Zygnematophyceae: *Spirogyra sp*. Charophyceae: *Chara autralis* (*Chaus*), *Chara braunii* (*Chbrau*), *Chara leptospora* (*Chlep*). Bryophyte: *Marchantia polymorpha* (*Mapoly*). Lycophyte: *Selaginella moellendorfii* (*Selmo*). Pteridophytes: *Azolla filiculoides* (Azfi), *Salvinia cuculata* (Sacu). Gymnosperm: *Picea abies* (Piabi). Angiosperms: *Amborella trichopoda* (Amtr), *Arabidopsis thaliana* (Arath). Sequence names are color coded after the R2R3 MYB subfamilies defined in Jiang & Rao (2020). Foldchange in response to FR was calculated by DESeq2 (Love *et al*., 2014); green stars mark significant changes of Padj <0.1.