Author names:

Laura W. Dijkhuizen[1], Badraldin Ebrahim Sayed Tabatabaei[2], Paul Brouwer[1], Niels Rijken[1], Valerie A. Buijs[1], Erbil Güngör[1], Henriette Schluepmann[1]

Affiliations:

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

Present address Valerie A. Buijs: Westerdijk Fungal Biodiversity Institute, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands.

Corresponding author e-mail address: h.schlupmann@uu.nl

Title: Control of the *Azolla* symbiosis sexual reproduction: ferns to shed light on the origin of floral regulation?

Short Title: The *Azolla* symbiosis sexual reproduction

Material distribution footnote:

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Henriette Schluepmann (h.schlupmann@uu.nl).

ABSTRACT (210)

Azolla ferns and the filamentous cyanobacteria *Nostoc azollae* constitute a model symbiosis that enabled colonization of the water surface with traits highly desirable for development of more sustainable crops: their floating mats capture CO_2 and fixate N_2 at high rates phototrophically. Their mode of sexual reproduction is heterosporous. Regulation of the transition from vegetative to spore-forming phases in ferns is largely unknown, yet a pre-requisite for *Azolla* domestication, and of particular interest since ferns represent the sister lineage of seed plants.

Far-red light (FR) induced sporocarp formation in *A. filiculoides.* Sporocarps obtained, when crossed, verified species attribution of Netherlands strains but not Iran's Anzali lagoon.

FR-responsive transcripts included CMADS1 MIKC^C-homologues and miRNA-controlled GAMYB transcription factors in the fern, transporters in *N.azollae,* and ycf2 in chloroplasts.

Loci of conserved miRNA in the fern lineage included miR172, yet FR only induced miR529 and miR535, and reduced miR319 and miR159.

Suppression of sexual reproduction in both gametophyte and sporophyte-dominated plant lineages by red light is likely a convergent ecological strategy in open fields as the active control networks in the different lineages differ. MIKC^C transcription factor control of flowering and flower organ specification, however, likely originated from the diploid to haploid phase transition in the homosporous common ancestor of ferns and seed plants.

22 **INTRODUCTION**

23 Ferns from the genus *Azolla* thrive floating on freshwater. They require no nitrogen whilst growing even
24 at the highest rates due to an unusual symbiosis with the cyanobacterium *Nostoc azollae,* which fixates
25 dinitrogen phototrophically (Brouwer *et al.*, 2017). They are a model plant symbiosis allowing to study
26 important traits: for example, highly efficient N_2 fixation imparted by a microbial consortium, building of
27 substrate mats with CO_2 draw-down (Speelman *et al.*, 2009), and adaptations to the aquatic floating
28 lifestyle. Representative of the Monilophyte lineage moreover, their study is of interest as a sister lineage
29 to seed plants to inform about the origin of seed plant traits. Previously used as biofertilizer, they now are
30 considered a candidate crop for sustainable high-yield production of plant protein in subsiding delta
31 regions but have never been domesticated (Brouwer *et al.*, 2014a, 2018). What controls their sexual
32 reproduction is completely unknown yet central to containment, propagation, breeding schemes or the
33 establishment of biodiversity and germplasm banks.

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

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

52 In the SAM transition to RD of the model angiosperm *Arabidopsis thaliana* (Arabidopsis), the LEAFY
53 transcription factor (TF) complex induces floral homeotic genes (Sayou *et al.,* 2016). LEAFY is conserved
54 in plant lineages, but LEAFY does not promote formation of reproductive structures in seed-free plants
55 including ferns where, nonetheless, the protein was reported to maintain apical stem cell activity (Plackett
56 *et al.*, 2018). The pattern emerging, therefore, is that the LEAFY complex is not among those that were
57 already in place to link exogenous cues to the transition of SAM to RD in the common ancestor of seed
58 plants and ferns.

59 In many angiosperms, MADS-box TF of MIKC^C-type form complexes that integrate endogenous and
60 exogenous cues turning SAM into RD (Theißen *et al.*, 2018). In Arabidopsis, for example, gibberellic acid
61 (GA)-dependent and photoperiod/temperature signaling is integrated by MIKC^C from the FLC and SVP
62 clades that, as predicted by Vasco *et al.* (2016), are floral repressors. Consistent with separate transitions
63 for inflorescence and flower meristems in Arabidopsis, the MIKC^C floral integrators act on another, SOC1,
64 that in turn induces the expression of MIKC^C floral homeotic genes. Homeotic genes of potential relevance
65 for ferns specify ovules, and stamens: these modules could be more ancient than those specifying carpels
66 and sepals. Specification of ovules/stamens in angiosperms, however, is unlikely homologous to that of
67 mega- and microsporocarps in *Azolla* ferns because heterospory evolved independently in ferns and
68 angiosperms (Sessa & Der, 2016).

69 In addition to MIKC^C TF, micro RNA (miR) are known to control phase transitions in angiosperms, including
70 RD. The miR156/172 age pathway controls flowering: miRNA156 decreases with the age of plants, and
71 thus also its repression of the targets *SQUAMOSA PROMOTER BINDING-LIKES* (*SPL*). In the annual
72 Arabidopsis, SPL promote expression of mir172 that targets another type of floral repressor, the AP2/TOE
73 TF (Hyun *et al.*, 2019). In the perennial *Arabis alpina*, the MIKC^C FLC homologue and miR156 regulate
74 expression of *SPL15* thus integrating temperature and age-dependent RD. Moreover, AP2/TOE and NIN
75 TF are known to mediate repression of *SOC1* in Arabidopsis on high nitrate (Gras *et al.*, 2018; Olas *et al.*,
76 2019). The MIKC^C SOC1 is further known to control levels of miR319 that inhibits TCP protein functions in
77 the FT-FD complex mediating photoperiod control of RD (Lucero *et al.*, 2017; Li et al., 2020). Inextricably,
78 therefore, transition to RD is linked to conserved miRNA modules in angiosperms but we know little of
79 miRNA loci in ferns (Berruezo *et al.*, 2017; You *et al.*, 2017).

80 *Azolla* ferns and Arabidopsis were shown to share responses to nitrogen starvation, suggesting that the
81 ancestor of both had already evolved some of the responses to exogenous cues we know from
82 angiosperms (Brouwer *et al.*, 2017). GA applications to *Azolla* ferns, however, did not induce RD in *Azolla*,

83 neither did stress treatment (Kar *et al.*, 2002). Systematic research of the symbiosis transition to RD is
84 lacking. Fern host and symbiont development are coordinated during the induction of sporocarps: mobile
85 filaments of *N. azollae* generated in the shoot apical colony move into the developing sporocarp initials in
86 the shoot apex (Figure 1). The host SAM developmental transition, therefore, may be influenced by
87 activities of the *N. azollae* apical colony residing at the SAM (Figure 1A). Alternatively, *N. azollae*
88 development may be controlled by secretions from specialized host trichomes in both the SAM, the
89 developing leaf and sporocarp initials (Cohen *et al.*, 2002; Perkins & Peters, 2006; Figure 1B and 1C).
90 Transition to RD will be best characterized by profiling expression of the fern host and the symbiont
91 simultaneously: we expect elaborate communication given that both genomes co-evolved and that *N.*
92 *azollae* is an obligate symbiont with an eroded genome (Ran et al., 2010; Dijkhuizen et al., 2018; Li et al.
93 2018).

94 Here we identify external cues that induce or inhibit transition to RD in *A. filiculoides* under laboratory
95 settings. We test the viability of reproductive structures induced in different fern accessions by crosses,
96 define species attributions and support the results obtained with phylogenetic analyses of the accessions.
97 To reveal endogenous mechanisms mediating transition to RD, we profile transcript accumulation by dual
98 RNA-sequencing comparing sporophytes grown with and without far-red light, at a time point when the
99 reproductive structures are not yet visible. We assign induced MIKC^{C} and MYB TF to known clades using
100 phylogenetic analyses. In addition, we profile small RNA from the symbiosis, then identify miRNA and their
101 targets in the fern lineage. Finally, we discover regulatory modules of conserved miRNA/targets involved
102 during induction of RD in ferns.

103 **RESULTS**

104 **Far-red light induces, while nitrogen in the medium represses, the transition to sexual reproduction in**
105 ***A. filiculoides***

106 A young fern growing at low density typically spreads over the water surface (Figure 2A). In contrast, a
107 fern that underwent transition to reproductive development in dense canopies formed a three
108 dimensional web characteristic of *Azolla* mats (Figure 2B). Unlike in the glasshouse, *A. filiculoides* ferns
109 never sporulated under tube light (TL, Supplemental Figure 1A); TL did not emit any photons in the far-
110 red range (FR, Supplemental Figure 1A). We, therefore, added far-red LED (FR, Supplemental Figure 1B)
111 during the entire 16 h light period. This initially resulted in the elongation of the stem internodes of the
112 fern (Figure 2C) and, after 2-3 weeks, in the formation of the sporocarps (Figure 2D). Sporocarps were

113 positioned as pairs on the branch side of branch points and were predominantly arranged as a mega- and
114 microsporocarp pair (Figure 2E).

115 To test the optimum light for sporocarp induction, decreasing ratios of red to far-red LED were applied to
116 ferns: sporophytes were first raised in TL light and then transferred to different red to far-red-light ratios
117 whilst they were also maintained at three differing densities (Supplemental Table 2, Supplemental Figures
118 1C, 1D and 1E). Far-red light was required for sporocarp formation regardless of the density of the fern
119 culture (Figure 2F). Increasing FR light whilst keeping a constant photosynthetic active radiation (PAR)
120 yielded more sporocarps per fern mass. The sporocarps per fern mass increased with time of exposure to
121 the FR up until five weeks and then decreased, presumably because ripe sporocarps detached. The
122 sporocarps per fern mass, furthermore, increased with the density of the fern culture. Reducing the PAR
123 at highest FR intensity (Figure 2F**,** FR4) revealed that PAR limits the production sporocarps at high FR light.

124 Adding 2 mM ammonium nitrate to the medium drastically reduced the spore count when inducing
125 sporocarp formation with FR (Figure 2G). Both, ammonium and nitrate added individually inhibited
126 sporocarp formation (data not shown), and inhibition explained why sporophytes without *N. azollae* kept
127 on various nitrogen sources never sporulated.

128 The ratio of mega- to microsporocarp observed was approximately equal over nine weeks of induction
129 with FR (Supplemental Figure 1F); and the marker *AzfiSOC1* (Brouwer *et al.*, 2014) was maximally induced
130 when sporocarps became visible four weeks after FR induction started (Supplemental Figure 1G).

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

133 We next wondered whether the sporocarps induced with FR were viable. We randomly tested reciprocal
134 crosses of the sequenced accession *A. filiculoides* Galgenwaard (Li *et al.*, 2018), six further accessions from
135 the Netherlands, one from Gran Canaria (Spain) and another from the Anzali lagoon (Iran) (Supplemental
136 Table 1). All accessions could be induced to sporulate under FR, except Anzali. All sporulating accessions
137 tested in crosses could be crossed suggesting that the ferns from the Netherlands and Spain were *A.*
138 *filiculoides* (Table 1).

139 To research the taxonomy of the accessions, intergenic regions of the chloroplast rRNA, used for
140 phylogenetic studies previously (Madeira *et al.,* 2016), were amplified and sequenced to compute
141 phylogenetic relations. The regions trnL-trnF (Figure 3A), and trnG-trnR (Figure 3B) yielded similar trees

142 confirming the close relationship of the *A. filiculoides* accessions from the Netherlands and Spain. The
143 trees further revealed that the Anzali accession clustered along with two accessions from South America
144 that do not cluster within the *A. caroliniana* group (Southern Brazil and Uruguay, Supplemental Table 1).
145 Use of the nuclear ITS region for phylogeny reconstruction proved futile when amplifying the region on
146 the *A. filiculoides* genome *in silico*: the *A. filiculoides* genome assembly contained over 500 copies of the
147 ITS sequence, and the copies of ITS were variable within the same genome (Figure 3C). Variations within
148 anyone genome were furthermore large when comparing variations between genomes of *Azolla* species
149 (Supplemental Figure 3A).

150 The dorsal side of the upper leaf lobes of the Anzali accession had two-celled, distinctly pointed, papillae
151 unlike *A. filiculoides* that mostly, but not exclusively, have single-celled and rounded papillae
152 (Supplemental Figure 3B). Consistent with its differing response to far-red light, and its phylogenetic
153 position based on the chloroplast sequences, the Anzali accession, we conclude, is not *A. filiculoides*.

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

156 We nest wondered which molecular processes are altered during the symbiosis transition to RD. Ferns
157 maintained on TL were transferred to TL with and without FR for a week in triplicate cultures, then
158 collected 2 h into the 16 h light period snap frozen. Total RNA was then extracted, its plant rRNA depleted,
159 and stranded sequencing libraries generated which were depleted for rRNA from gram-negative bacteria.
160 Libraries were sequenced to obtain 12 to 60 million paired-end (PE) 50 b reads, quality filtered and
161 trimmed. To align the reads obtained, nuclear and chloroplast genomes with annotations for the
162 *A.filiculoides* Galgenwaard accession were available (Li *et al.,* 2018) but not for its cyanobacterial
163 symbiont.

164 The *N. azollae* Galgenwaard genome was assembled both from plants kept in the lab (*A. filiculoides* lab; Li
165 et al., 2018) and plants taken from the original sampling location (*A. filiculoides* wild; Dijkhuizen et al.
166 2018). Both assemblies were from short-reads of the entire symbiosis metagenome and were fragmented
167 but highly complete: over 95% based on single copy marker genes. Contigs aligning to the *N. azollae* 0708
168 reference (Ran *et al.*, 2010) were well over 99.5% identical to each other (Figure 4A). This degree of
169 nucleotide identity was well above the maximum mismatch of STAR-aligner's default setting; PE reads
170 were thus aligned to the Galgenwaard accession nuclear and chloroplast genomes, and the *N. azollae*
171 0708 genome, and read counts derived from their GFF annotation files. Alignments revealed a high

172 proportion of uniquely mapped reads to the fern nuclear and *N. azollae* genomes suggesting efficient
173 removal of rRNA (Figure 4B). The larger proportion of PE reads mapping to multiple loci in the chloroplast
174 indicated that fern chloroplast rRNA was not as efficiently removed (Figure 4B), yet read counts per gene
175 were still high because few genes are encoded in the chloroplast genome (84 without rRNA and tRNA
176 genes). Read counts in genome features further revealed that sense alignments dominated and thus that
177 the libraries were stranded (Figure 4C).

178 Dispersion of sense read count per gene was low which permitted identification of differentially
179 accumulating transcripts (defined as DEseq2 Padj <0.1; Love *et al.,* 2014) comparing ferns on TL and TL
180 with FR: 318 from the fern nucleus, 67 from *N. azollae* and five from the fern chloroplast (Supplemental
181 Table 4). We next wondered whether the changes in transcript abundance reflect adaptations to light
182 quality and the transition to reproductive development.

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

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

189 Consistent with the high levels of N_2 fixation measured in *Azolla* ferns, the highest *N. azollae* read counts
190 were in transcripts of the *nif* operon, genes encoding photosystem I and II proteins and ATP synthase
191 subunits. In addition, read counts for *ntcA* and metabolic enzymes supporting high N_2 fixation were high
192 (Supplemental Table 5). Read counts from the *N. azollae* transcripts reflected the activity of the symbiont
193 and we thus further analyzed the data for differential expression.

194 In *N. azollae*, far-red light led to accumulation of the PsbA transcript, but reduction of photosystem I
195 reaction center subunit XII, geranylgeranyl hydrogenase ChlP, and phycobilisome rod-core linker
196 polypeptide CpcG2 transcripts (Supplemental Table 4, *N. azollae*). These changes in light harvesting-
197 related genes were small, however, compared to changes in hypothetical proteins of unknown function,
198 transposases and highly expressed transporters (TrKA, MFS and ABC transporters). We conclude that a
199 one-week induction of reproductive structures with far-red LED triggered few transcriptional changes
200 reflecting light-harvesting adaptations in *N. azollae*. The larger differential accumulation of transporter
201 transcripts reflects changes in metabolite trafficking and communication with the host fern.

202 **FR alters transcripts from MIKC^C and R2R3MYB TF related to those of the RD of seed plants**

203 As in the case of *N. azollae* symbiont, the ferns largest changes in transcript accumulation under FR were
204 in genes with unknown function, then in secondary- and lipid-metabolism (Supplemental Table 4, *A.*
205 *filiculoides* nucleus). We noted the large induction of the SWEET-like transporter (Azfi_s0221.g058704
206 log2 7.5 fold), of transcripts related to cysteine-rich peptide signaling (Azfi_s0046.g030195 log2 7.1 fold;
207 Azfi_s0745.g084813 log2 4.0 fold) and of transcripts from jasmonate metabolism. To examine the link
208 between RD in seed plants and ferns, however, TF were analyzed in more depth. Transcripts of MIKC^C TF
209 and of R2R3MYB TF were strikingly induced in sporophytes under FR (Table 2).

210 Prediction of the most induced TF transcript annotated as SOC1-like MIKC^C (Azfi_s0028.g024032, log2 7.2
211 fold) missed the K-domain because this TF was not expressed in vegetative sporophytes. Manual
212 annotation allowed computation of a phylogenetic tree (Supplemental Figure 4): both induced *Azolla*
213 MIKC^C were in the well-supported clade (95.5% bootstrap confidence) of fern sequences radiating
214 separately from the angiosperm and gymnosperm MIKC^C clade that gave rise to AtSOC1, FLC, and the
215 A,C,D and E-function floral homeotic genes (Supplemental Figure 4). Induction of the fern SOC1-paralogue
216 was confirmed by RT-PCR (Supplemental Figure 2G). In Arabidopsis, AtSOC1 is known to control the steady
217 state of miR319 that inhibits TCP TF functions necessary for the transition to an inflorescence meristem
218 at the shoot apex (Lucero *et al.*, 2017); the link may also operate in *Azolla* ferns since a TCP TF
219 (Azfi_s0168.g054602) was induced in ferns on TL with far-red LED (Supplemental Table 4, *A. filiculoides*
220 nucleus).

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

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

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

230 Short RNA reads were obtained from sporophytes after a week growth under TL without and with FR from
231 the same cultures as those used for dual RNA sequencing. In addition, reads were obtained from
232 sporophytes grown under TL with far-red LED on medium with 2 mM NH_4NO_3, and from sporophytes
233 without *N. azollae* on medium with 2 mM NH_4NO_3. All samples were collected two hours into the 16 h
234 light period. The sRNA sequencing library preparation chemistry did not distinguish pro- from eukaryote
235 sRNA and therefore was inherently a dual profiling method. Reads obtained were aligned to concatenated
236 genomes which yielded small RNA profiles characteristic for the fern nucleus, chloroplast and *N. azollae*
237 (Figure 5; Supplemental Figure 5).

238 The most abundant sRNA encoded in the fern nucleus were 21, 20 and 24 nt long, but the most diverse
239 were 24 nt long, consistent with abundant 21 nt miRNA discretely cut from a hair-pin precursor by dicer,
240 and the 24 nt siRNA generated by random processes (Figure 5A). The chloroplast most abundant sRNA of
241 28 nt was also the most diverse, but the chloroplast sRNA have an unusual peak of sRNA of low diversity
242 at 34 nt (Figure 5B). The sRNA encoded in *N. azollae* are generally shorter than those of the chloroplast,
243 yet, they have a conspicuous peak of low diversity at 42 nt length, consistent in length with CRISPR RNA
244 (Figure 5C).

245 **Conserved miRNA in *A. filiculoides* include miR172 and 396, but only miR319 and 159 were decreased**
246 **in response to FR, whilst miR529 and 535 were increased**

247 Fern genome-encoded sRNA reads of 20-25 nt length were used for miRNA predictions allowing a hairpin
248 of maximally 500 bp. Loci predicted were then verified against miRNA criteria (Axtell & Meyers, 2018) as,
249 for example, in the case of the miRNA 172 locus Azfi_s0021g15800 (Figure 6): the locus was expressed as
250 a pre-miRNA (Figure 6A, RNAseq all conditions) and both the miRNA and miRNA* are found in replicate
251 samples and cut precisely with two base overhangs (Figure 6B and 6C, sRNAseq). The miR172 reads were
252 derived from the fern since they were also found in sporophytes without *N. azollae* (Figure 6**,** RNAseq -*N.*
253 *azollae*). The locus encoded a stable long hairpin which when folded released -65.76 kcal mol^{-1} at 37°C.
254 MiR172 targets in seed plants include the AP2/TOE transcription factors, and alignment of the target sites
255 to all annotated AP2/TOE transcripts from *A. filiculoides* revealed target sites for the AzfimiR172
256 (Supplemental Figure 6). The miR156a locus was similarly verified (Supplemental Figure 7).

257 Mature miRNA were compared in miRbase (Kozomara *et al.*, 2019) which identified 11 of the conserved
258 miRNA families of 21 nt in the *A. filiculoides* genome and thus in the fern lineage (Figure 7A). The list

259 included miR396, as well as miR172, not yet confirmed in other seed-free plant lineages. MiRNA
260 predictions further uncovered a number of novel miRNA (miRNF).

261 Differential expression analysis of 20-22 nt reads exhibited more dispersion than the dualRNAseq of long
262 RNA, yet, it reliably identified sRNA with low Padj values when comparing ferns under TL with and without
263 FR (Supplemental Figures 5D and 5E). Reads from miR529c&d and miR535 increased (Padj < 0.1) whilst
264 those of miR159, miR319 and miRNF4 decreased (Padj <0.16) in sporophytes on TL with FR compared to
265 TL light only (Figure 7B). We conclude that the miRNA156/miR172 known to reciprocally control flowering
266 are not altered in the ferns undergoing transition to RD under FR.

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

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

272 Transcripts of neither SPL nor AP2/TOE-like targets, however, were significantly (Padj < 0.1) altered in
273 sporophytes comparing TL with and without FR; this could be due to poor Azfi_vs1 annotation and
274 incomplete prediction of miR targets (Supplemental Table 6, F vs T Padj). Nevertheless, consistent with an
275 increased miR529c&d the SPL from the Azfi_s0173.g055767 transcript is reduced (log2 0.74 fold) but with
276 low significance (Padj of 0.356). In contrast, transcripts of two GAMYB (Azfi_s0004.g008455 and
277 Azfi_s0021.g015882) increased up to 4 fold in sporophytes with decreased miR319 on TL with FR. The
278 location of mir319 binding is shown for the GAMYB encoded in Azfi_s0004.g008455 (Supplemental Figure
279 8).

280 Phylogenetic analyses revealed that the GAMYB targeted by miR319 were on the same leaf as the
281 GAMYB33 from Arabidopsis and were placed next to each other, probably because they arose from gene
282 duplication (Figure 8). This is consistent with the whole genome duplication observed at the base of *Azolla*
283 evolution (Li *et al.*, 2018). We conclude, therefore, that miR319/GAMYB is responsive in sporophytes
284 undergoing transition to RD caused by FR.

285

286 **DISCUSSION**

287 **Red-dominated light suppresses formation of dissemination stages in both gametophyte and**
288 **sporophyte-dominated lineages of plants and is, therefore, a convergent ecological strategy.**

289 Far-red light responses in *Azolla* ferns look alike a shade-avoidance syndrome but signal transduction
290 pathways that mediate them in ferns remain largely uncharacterized (Inoue *et al.*, 2017). Pathways
291 causing shade response components are known to use alternative phytochromes and interacting factors
292 (Possart *et al.*, 2013; Xie *et al.*, 2020).

293 Phytochromes in ferns radiated separately from those of seed plants (Li *et al.*, 2015), hence, their function
294 cannot be predicted using orthology with seed plants. Nevertheless, fern phytochromes have a similar
295 structure and, as in the case of PHYB in Arabidopsis, may sense temperature as well as red/far-red light
296 by thermal reversion of the Pfr to fr state (Legris *et al.*, 2017; Klose *et al.*, 2020): initiation of fern, liverwort
297 and bryophyte sporangia was dependent on temperature and photoperiod (Labouriau, 1958, Benson-
298 Evans, 1961; Nishihama *et al.*, 2015). Moreover, thermal reversion of phytochromes from some
299 cyanobacteria was described *in vitro* as early as 1997 (Yeh *et al.*, 1997). Therefore, if far-red light induces
300 RD in *A. filiculoides*, then differing temperature regimes combined with light quality changes may induce
301 RD in the Anzali ferns. Anzali ferns were not *A. filiculoides* but related to species from Uruguay and
302 Southern Brazil which did not cluster with, (Figure 2) and did not have the single-celled papillae of the
303 traditional *A. caroliniana* species (Pereira *et al*., 2001; Supplemental Figure 3). Position of the Anzali
304 accession was consistent with observations of distinct *Azolla* ferns in the Anzali lagoon (Farahpour-
305 Haghani *et al.*, 2017), and suggests that not only *A. filiculoides* may have been released as nitrogen
306 biofertilizer.

307 Unlike *A. filiculoides*, Arabidopsis will transition to RD in the TL generally used for indoors cultivation. Most
308 studies on the transition to RD, therefore, were done with red-light gated plants, i.e. conditions leading
309 to artefactual peaking of CONSTANS protein accumulation and florigen (FT) expression (Song *et al.*, 2018).
310 Nevertheless, Arabidopsis as well as many other seed plants flower early when in ratios of red to far-red
311 light approaching one, the ratio encountered in day light, compared to TL. Far-red light-dependent
312 flowering is mediated by PHYB and D, eventually leading to *SOC1* expression in the meristem (Halliday *et*
313 *al.*, 1994; Aukerman *et al.*, 1997; Lazaro *et al.*, 2015). ft mutant Arabidopsis flowered earlier under
314 incandescent light (Martinez-Zapater & Somerville, 1990) or in TL with far-red LED compared to in TL
315 (Schluepmann *et al.,* unpublished) suggesting that FT is not involved. Signaling was also reported via PHYB-
316 regulated PIF4 protein complexes that bind the promoters of miRNA156/172 directly with interference
317 from PHYA-regulated TF (Sánchez-Retuerta *et al.*, 2018; Sun *et al.*, 2018; Xie *et al.*, 2020). In the present

318 study, the highly induced MIKC^C turned out a paralogue of SOC1, and miR156/172 steady states were
319 unaltered by the far-red triggered transition to RD in *A. filiculoides* (Figure 7B; Supplemental Table 6)
320 suggesting that mechanisms in ferns and seed plant RD are not conserved.

321 Far-red induced sexual reproduction in the liverwort gametophytes *Marchantia polymorpha* (Chiyoda *et*
322 *al.*, 2008; Kubota *et al.*, 2014; Yamaoka *et al.*, 2018; Tsuzuki *et al.*, 2019). The mechanism in these
323 liverworts involved the MpmiR529c/MpSPL module, which was necessary for the development of
324 reproductive branches. MpPHY and MpPIF were required to mediate the response (Inoue *et al.*, 2019).
325 Our analyses with *A. filiculoides* reveal induction of AzfimiR529 under far-red light but no significant
326 change in SPL targets. Additionally, the specific MpmiR529c sequence involved in *M. polymorpha* was not
327 detected in our ferns. The pathways thus differ in the two seed-free plants.

328 Given that gametophyte and sporophyte dominated lineages require far-red light to initiate formation of
329 dissemination stages, we conclude that repression of RD under dominating red light in *Azolla* likely reflects
330 a convergent ecological strategy: in open surfaces, where red-light is abundant, there is sufficient space
331 to prolong vegetative development and no need for dispersal. Consistently, density of the *A. filiculoides*
332 canopy had an additional impact on the number of sporocarps observed per gram plant (Figure 2F) that
333 may stem from alternative cues such as volatile organic compounds released by neighbors (Vicherová *et*
334 *al.*, 2020). Taken together, results presented in this work signify that, for the first time, RD of the *Azolla*
335 fern symbiosis is amenable to experimental enquiry.

336 **Phylogenetic position of MIKC^C TF responsive during *Azolla* RD suggests that control of flowering in seed**
337 **plants originates from the diploid to haploid transition in the common ancestor of seed plants and ferns**

338 The GAMYB TF clade arose before land plants evolved and before GA signaling (Aya *et al.*, 2011; Bowman
339 *et al.,* 2017, Jiang & Rao, 2020, Figure 8). *Azolla* GAMYB induced under TL with far-red, therefore, likely
340 function in RD by mediating cues perceived by the sporophyte into an ancestral gametophyte regulon.
341 GAMYB from the moss *Physcomytrella patens* were required for the formation of antheridiophores by the
342 gametophyte, and of normal spores by the sporophyte (Aya *et al.*, 2011). The GAMYB regulon did not play
343 a role in the developmental transition of lycophyte sporophytes to formation of sporogenic structures.

344 AtGAMYB are known to be regulated by miR159 in Arabidopsis and by the related and more ancestral
345 miR319 in seed-free plant lineages (Achard *et al.*, 2004; Palatnik *et al.*, 2007). The AtGAMYB/miR159 is
346 part of the GA pathway promoting flowering in Arabidopsis under short days (Millar *et al.*, 2019).
347 Specifically, AtMYB33 was shown to directly act on the promotors of both miR156 and AtSPL9 (Guo *et al.,*

348 2017). Far-red light repressed miR159 and miR319, with increased AzfiGAMYB, in *Azolla* were thus
349 consistent with wiring in Arabidopsis; furthering their analysis in homosporous ferns with free
350 gametophytic stages will reveal whether they are important for the diploid to haploid transition.

351 Control over sexual reproductive transition as it switched from the gametophyte, in gametophyte
352 dominated seed-free plants, to the diploid sporophyte in seed plants, resulted from the evolution of
353 inflorescence and floral meristem MIKC^C regulons. The MIKC^C clade of TF that specify both the transition
354 to RD in seed plant sporophytes, such as AtSOC1 and AtFLC, and the A,C, D and E homeotic functions are
355 peculiar in that the TF radiated in each megaphyll plant lineage separately (Leebens-Mack *et al.*, 2019;
356 Supplemental Figure 4). The TF work as hetero-tetramers and may have evolved from an ancient tandem
357 gene duplication through subsequent polyploidy events (Zhao *et al.*, 2017). Sepals are organs formed only
358 in angiosperms. Consistently, TF with A functions specifying sepals occurred solely in angiosperms. MIKC^C
359 specifying organs that contain the reproductive structures in angiosperms (C, D and E homeotic genes),
360 had direct homologues in gymnosperms, but not in ferns. Instead, ferns radiated a sister clade to the clade
361 containing AtSOC1, and A, C, D and E MIKC^C. This clade contained the *Azolla* MIKC^C responsive to far-red
362 light and the closely related *Ceratopteris* CMADS1. *In situ* hybridizations and northern blot analyses reveal
363 that *CMADS1* transcripts accumulate at high levels in the sporangia initials (Hasebe *et al.*, 1998). At the
364 base of the radiating clade of modern MIKC^C that specify modern organs of the sporophytes controlling
365 the gametophyte development within, is the LAMB1 protein expressed specifically in sporogenic
366 structures of lycophytes (Leebens-Mack *et al.*, 2019; Svensson *et al.*, 2000; and Supplemental Figure 4).
367 Combined phylogenetic and RNAseq analyses, therefore, suggest that the MIKC^C regulons controlling
368 flowering originate from the diploid to haploid phase transition of the common ancestor to ferns and seed
369 plants, not from regulons controlling sexual reproduction of the haploid phase.

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

372 The dual RNAseq method could be applied more generally to study bacteria/plant assemblages
373 particularly common in the seed-free plant lineages, representing an important base of the "tangled tree"
374 of plant lineage evolution (Quammen D. 2018). For *Azolla* where co-evolution of *N. azollae* and fern host
375 genomes demonstrated that fitness selection occurs at the level of the metagenome (Li *et al.*, 2018), it
376 delivered foundational data and opens the way to studying the coordinate development of host and
377 obligate symbiont. Efficient removal of the rRNA was key to the success of reading long RNA. Fern
378 chloroplast rRNA removal could be improved still, using specific DNA probes integrated during synthesis

379 of the stranded sequencing libraries. Transcripts from the fern's mitochondrial genome are likely in the
380 data but we have yet to assemble the *A. filiculoides* mitochondrial genome to confirm this.

381 *N. azollae* extracts did not contain photosynthesis pigments absorbing/fluorescing in the far-red spectrum
382 and the *N. azollae* genome did not contain psbA4/Chlf-S like proteins (data not shown). Yet genes
383 encoding phytochromes and cyanobacteriochromes capable of sensing far-red light as other filamentous
384 cyanobacteria are present (Wiltbank & Kehoe, 2019). The RNA profiling did not reveal whether the *N.*
385 *azollae* trigger fern RD since roles of the far-red responsive transcripts in *N. azollae*, particularly those
386 encoding the transporters, have yet to be studied. An approach could be to study these in free living
387 filamentous cyanobacteria amenable to genetic manipulation. Manipulation of the symbiont would,
388 however, be preferred. Tri-parental mating protocols may be effective on *N. azollae in situ* if combined
389 with high-frequency RNA-guided integration with Cas12d enzyme complexes (Strecker *et al.,* 2019). *N.*
390 *azollae* Cas genes were found to be mostly pseudogenes and CRISPR arrays appeared missing (Cai *et al.,*
391 2013), suggesting that incoming cargo plasmid may not be destroyed by Cas complexes. Nevertheless,
392 sRNA profiles from Figure 6 show discrete accumulation of 31 and 42 b sRNA of very low complexity and
393 a functional Cas6 splicing CRISPR precursor RNA was induced on TL with far-red LED (Supplemental Table
394 4, *N. azollae*). CRISPR arrays may not be recognized if they contained few repeats due to the sheltered
395 life-style inside the fern. Rendering *A. filiculoides* and *N. azollae* amenable to genetic manipulation will be
396 crucial to provide evidence for hypotheses generated by transcript profiling in the future.

397

398 **MATERIALS AND METHODS (1320)**

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

400 Ferns were collected from differing locations in the Netherlands, grown in liquid medium at a constant
401 temperature of 22°C under long days (16h light) as described earlier (Brouwer *et al.*, 2017). Ferns from
402 Spain and Iran were processed to DNA on site. GPS coordinates from the collection sites of the strains
403 used in phylogenetic analyses are in Supplemental Table S1. For sporocarp induction, ferns were first
404 raised under 16 h tube light (TL) periods, then transferred for induction under 16 h TL with far-red Light
405 Emitting Diodes - LED (Phillips GrowLED peaking at 735 nm, Supplemental Figures 1A and 1B).
406 Photosynthetic Photon Flux Density varied in the range 100-120 µmol m^{-2} s^{-1}. Sporophytes were generally
407 kept at densities that covered the surface as this inhibits growth of (blue) algae. When testing sporocarp
408 induction sporophytes were kept at set densities (with fresh weight resets once per week, Supplemental

409 Figures 1C, 1D and 1E) and set red to far-red light ratios (Supplemental Table **2**). Phenotypic change in
410 mature sporophytes were photographed using a Nikon D300s with an AF Micro-Nikkor 60mm f/2.8D
411 objective or bellows on a rail to generate image reconstructions with Helicon focus software**.**
412 Megasporocarps were collected manually using tweezers, so were microsporocarps. To initiate crosses
413 microsporocarps were Dounce-homogenized and massulae thus released added to megasporocarps in
414 distilled water (3-6 ml, in 3 cm diameter petri dish), shaking so as to obtain clumps. The crosses were kept
415 at room temperature (about 18°C) under 16 h light conditions under TL at about 100 µmol m^{-2} s^{-1}. As
416 sporelings emerged they were transferred to liquid medium (Brouwer *et al.*, 2017).

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

419 **Phylogenetic analyses** are described in details in Supplemental Methods. Chloroplastic marker regions
420 were as in Supplemental Table 3. The MIKC^{C} tree was subsampled from 1 KP (Leebens-Mack *et al.*, 2019);
421 the R2R3 MYB tree used data from respective genome repositories, the *Azolla* genome (Li *et al.,* 2018)
422 and Jiang & Rao, 2020.

423 **Reference genomes** used to align Dual RNA sequences onto the *A. filiculoides* Galgenwaard strain are
424 detailed in Supplemental Methods.

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

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

437 Reads obtained were de-multiplexed then trimmed and quality filtered with trimmomatic (Bolger *et al.,*
438 2014), yielding 13-60 million quality PE reads per sample. Alignment of the reads to the genome assembly
439 and GFF vs1 of *A. filiculoides* (Azfi_vs1; Li *et al.*, 2018) was with STAR-aligner default settings (Dobin *et al.*,
440 2013) and was combined with the STAR read count feature (PE read counts corresponding to the HTseq
441 "Union" settings, (Anders *et al.*, 2015)). Read count tables were fed into DESeq2 (Love *et al.*, 2013) for
442 analysis of differentially accumulating transcripts.

443 STAR was also used to align reads to the *N. azollae* and the chloroplast genomes, and read counts then
444 extracted using STAR default parameters or VERSE (Zhu *et al.*, 2016) with HTSeq setting "Union"
445 comparing single and paired End read counts on the polycistronic RNA. The three approaches yielded
446 similarly ranked genes but the Padj statistics of differential expression in count tables from the forced
447 single read counts were poor: only 20 *N. azollae* genes had P- adjusted values smaller than 0.1 compared
448 to up to 67 with the PE- approaches, we therefore present the PE-approach using STAR read counts only.

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

455 **Small RNA data sequencing and data processing**

456 Ferns were maintained on liquid medium without nitrogen with 16 h TL, then transferred to TL with and
457 without far-red LED for 7 days as described under Dual-RNA sequencing. In addition, sporophytes were
458 grown on medium with or without 2 mM NH_4NO_3; ferns without cyanobacteria were grown on medium
459 with 2 mM NH_4NO_3 in TL with far-red LED. All sporophytes were harvested 7 d into the treatment, 2 h into
460 the light period snap frozen. Samples were harvested from triplicate growth experiments and thus
461 represent triplicate biological replicates. Small RNA was extracted as per protocol using the mirPremier
462 microRNA Isolation Kit (Sigma-Aldrich), then DNAse treated (Brouwer *et al.*, 2017), then cleaned using the
463 RNeasy MinElute Cleanup Kit (Qiagen) protocol for small RNA. Libraries of the sRNA were then generated
464 using the SeqMatic TailorMix miRNA Sample Preparation 12-reaction Kit (Version 2) as per protocol.
465 Libraries were characterized using an Agilent Technologies 2100 Bioanalyzer and quantified before
466 sequencing using NextSeq500 and 1 x 75 b read chemistry (Illumina).

467 Reads thus obtained were de-multiplexed, trimmed and quality filtered with trimmomatic, then aligned
468 with STAR (Dobin *et al.*, 2013) (set at maximal stringency and to output Bam files) to the concatenated
469 fasta files of the genomes Azfi_vs1, its chloroplast, and *N. azollae* 0708. The resulting bam files were then
470 split for each genome, and Samtools fasta (Li *et al.*, 2009) used to extract reads aligning to each genome.
471 fastx collapser (http://hannonlab.cshl.edu/fastx_toolkit) was then used to extract read counts for each
472 identical sequence, with the sequence defining the name in the lists obtained. A data matrix was
473 generated in bash, then exported to Excel or Rstudio for further analyses.

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

479 **miRNA discovery**

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

489 **Data sharing**

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

493 **Supplemental Data**

494 **Supplemental Methods.** RNA extractions & quantitative RT-PCR, phylogenetic analyses, and metagenome

495 assemblies of *N.azollae* genomes from different *Azolla* species.

496 **Supplemental Table 1.** Location and coordinates of *Azolla* collection sites for strains used in the

497 phylogenetic analyses.

498 **Supplemental Table 2.** Red to far-red light intensity ratios and culture densities when testing induction
499 ofsporulation by far-red light.

500 **Supplemental Table 3.** Sequence accessions of the fern accessions from the phylogenetic trees in Figure
501 2A and 2B.

502 (separate Excell file with sheets for the Trn L-F and Trn G-R).

503 **Supplemental Table 4.** Differentially accumulating transcripts in sporophytes on tube light with and
504 without

505 far-red LED. (separate Excel file).

506 **Supplemental Table 5**. Read counts for transcripts of *N. azollae*. (separate Excell file).

507 **Supplemental Table 6.** miRNA and target loci transcript abundance in sporophytes in response to FR light.

508 **Supplemental Figure 1.** Light quality and fern densities when testing induction of sporulation.

509 **Supplemental Figure 2.** Flow chart to discover conserved and novel miRNA in *A. filiculoides*.

510 **Supplemental Figure 3.** Taxonomy of the *Azolla* strains in this study**.**

511 **Supplemental Figure 4.** Phylogenetic analysis of the *A. filiculoides* MIKCC responsive to FR.

512 **Supplemental Figure 5** sRNA sequencing and mapping statistics.

513 **Supplemental Figure 6.** MiR172 target sites in AP2/TOE1 transcription factors (eaAP2 lineage) comparing
514 *A.filiculoides* and seed plants.

515 **Supplemental Figure 7.** The miRNA156α locus in A. filiculoides.

516 **Supplemental Figure 8.** The AzfiGAMYB locus of A.filiculoides targeted by miRNA319.

517 **ACKNOWLEDGEMENTS**

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

AUTHOR CONTRIBUTIONS

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

REFERENCES

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

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

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

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

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

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

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

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

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

Bowman JL, Kohchi T, Yamato KT, Jenkins J, Shu S, Ishizaki K, Yamaoka S, Nishihama R, Nakamura Y, Berger F, *et al.* 2017. Insights into Land Plant Evolution Garnered from the Marchantia polymorpha

550 Genome. *Cell* **171**: 287-304.e15.

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

555 **Brouwer P, Bräutigam A, Külahoglu C, Tazelaar AOE, Kurz S, Nierop KGJ, van der Werf A, Weber APM,**
556 **Schluepmann H**. **2014**. Azolla domestication towards a biobased economy? *New Phytologist* **202**: 1069–
557 1082.

558 **Brouwer P, Schluepmann H, Nierop KGJ, Elderson J, Bijl PK, van der Meer I, de Visser W, Reichart G-J,**
559 **Smeekens S, van der Werf A**. **2018**. Growing Azolla to produce sustainable protein feed: the effect of
560 differing species and CO_2 concentrations on biomass productivity and chemical composition. *Journal of*
561 *the Science of Food and Agriculture* **98**: 4759–4768.

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

565 **Cohen MF, Sakihama Y, Takagi YC, Ichiba T, Yamasaki H. 2002.** Synergistic effect of deoxyanthocyanins
566 from symbiotic fern *Azolla* spp. on hrmA gene induction in the cyanobacterium *Nostoc punctiforme*.
567 *Molecular plant-microbe interactions*. **15**: 875–82.

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

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

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

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

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

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

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

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

588 **Guo C, Xu Y, Shi M, Lai Y, Wu X, Wang H, Zhu Z, Poethig RS, Wu G**. **2017**. Repression of miR156 by
589 miR159 Regulates the Timing of the Juvenile-to-Adult Transition in Arabidopsis. *The Plant Cell* **29**: 1293–
590 1304.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

624 **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,**
625 ***et al.* 2018**. Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nature Plants* **4**:
626 460–472.

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

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

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

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

637 **Madeira PT, Hill MP, Dray FA, Coetzee JA, Paterson ID, Tipping PW**. **2016**. Molecular identification of
638 Azolla invasions in Africa: The Azolla specialist, Stenopelmus rufinasus proves to be an excellent
639 taxonomist. *South African Journal of Botany* **105**: 299–305.

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

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

643 **Nagalingum NS, Schneider H, Pryer KM**. **2006**. Comparative Morphology of Reproductive Structures in
644 Heterosporous Water Ferns and a Reevaluation of the Sporocarp. *International Journal of Plant Sciences*
645 **167**: 805–815.

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

649 **Olas JJ, Van Dingenen J, Abel C, Działo MA, Feil R, Krapp A, Schlereth A, Wahl V**. **2019**. Nitrate acts at
650 the Arabidopsis thaliana shoot apical meristem to regulate flowering time. *New Phytologist* **223**: 814–
651 827.

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

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

658 **Pereira AL, Teixeira G, Sevinate-Pinto I, Antunes T, Carrapiço F. 2001.** Taxonomic re-evaluation of the
659 *Azolla* genus in Portugal. *Plant Biosystems-An International Journal Dealing with all Aspects of Plant*
660 *Biology*. **135**:285-94.

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

663 **Plackett ARG, Conway SJ, Hewett Hazelton KD, Rabbinowitsch EH, Langdale JA, Di Stilio VS**. **2018**.
664 LEAFY maintains apical stem cell activity during shoot development in the fern *Ceratopteris richardii*.
665 *eLife* **7**:e39625.

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

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

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

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

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

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

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

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

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

687 **Speelman EN, van Kempen MM, Barke J, Brinkhuis H, Reichart GJ, Smolders AJ, Roelofs JG, Sangiorgi F,**
688 **de Leeuw JW, Lotter AF, Sinninghe Damste JS. 2009.** The Eocene Arctic *Azolla* bloom: environmental
689 conditions, productivity and carbon drawdown. *Geobiology* **7**:155-70.

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

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

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

698 **Svensson ME, Johannesson H, Engström P**. **2000**. The LAMB1 gene from the clubmoss, Lycopodium
699 annotinum, is a divergent MADS-box gene, expressed specifically in sporogenic structures. *Gene* **253**:
700 31–43.

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

703 **Truernit E, Bauby H, Dubreucq B, Grandjean O, Runions J, Barthélémy J, Palauqui JC. 2008.** High-
704 resolution whole-mount imaging of three-dimensional tissue organization and gene expression enables
705 the study of phloem development and structure in Arabidopsis. *The Plant Cell* **20**:1494-503.

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

709 **Vasco A, Smalls TL, Graham SW, Cooper ED, Wong GK-S, Stevenson DW, Moran RC, Ambrose BA**. **2016**.
710 Challenging the paradigms of leaf evolution: Class III HD-Zips in ferns and lycophytes. *New Phytologist*
711 **212**: 745–758.

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

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

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

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

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

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

725 **Yeh K**. **1997**. A Cyanobacterial Phytochrome Two-Component Light Sensory System. *Science* **277**: 1505–
726 1508.

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

729 **Zhao T, Holmer R, de Bruijn S, Angenent GC, van den Burg HA, Schranz ME**. **2017**. Phylogenomic
730 Synteny Network Analysis of MADS-Box Transcription Factor Genes Reveals Lineage-Specific
731 Transpositions, Ancient Tandem Duplications, and Deep Positional Conservation. *The Plant Cell* **29**:
732 1278–1292.

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

734

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

Megaspores: Massulae :	Galgenwaard	Krommerijn	Hoogwoud	Gran Canaria	Den Bosch	Nijmegen	Nieuwebrug	Groningen
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 the Gran Canaria strain reproducibly germinated late, generally after 5 weeks instead of 2-3 weeks.

743 **Table 2.** Transcription factors with largest changes in transcript abundance when comparing sporophytes
744 in TL with and without far-red LED.

745

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

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

747

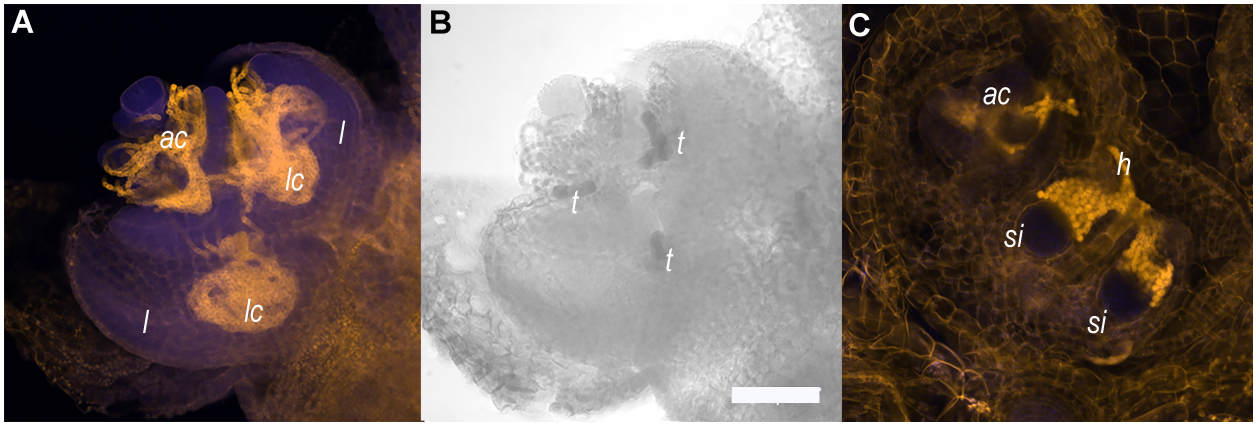

Figure 1. Shoot apices of *A. filiculoides* sporophytes contain the apical *N. azollae* colony, trichomes and sporocarp initials.

(A) Shoot apex with apical *N. azollae* colony (*ac*), outer leaves were removed for better visualization of developing leaf lobes (*l*) and bacteria.

(B) the same shoot apex as in (A) with trichomes (*t*) revealed by transmitting light absorption surrounded by the apical colony and in areas destined to form leaf cavities (*lc*); Bar: 50 μm.

(C) *N. azollae* hormogonia (*h*) invading the sporocarp initials (*si*). Sporophytes were stained with propidium idiode (Truernit *et al*., 2008), then fluorescence detected using 405 nm activation beam, and emission <560 nm filter for the yellow channel and 505-530 nm for the blue channel with the Confocal Laser Scanning microscope Zeiss LSM5 Pascal.

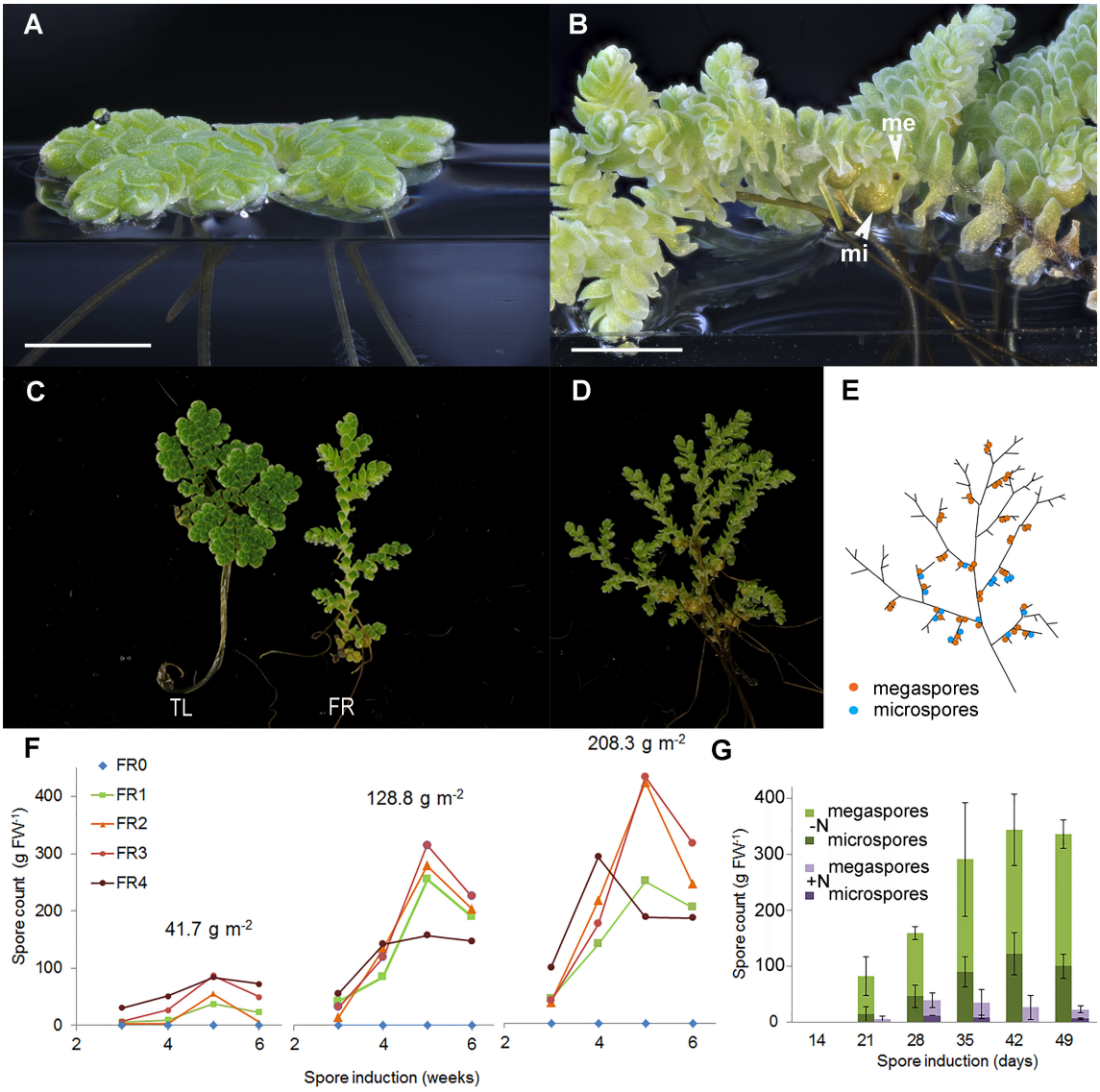

Figure 2. Far-red light and nitrogen affect the reproductive phase transition in *A. filiculoides*.

(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.

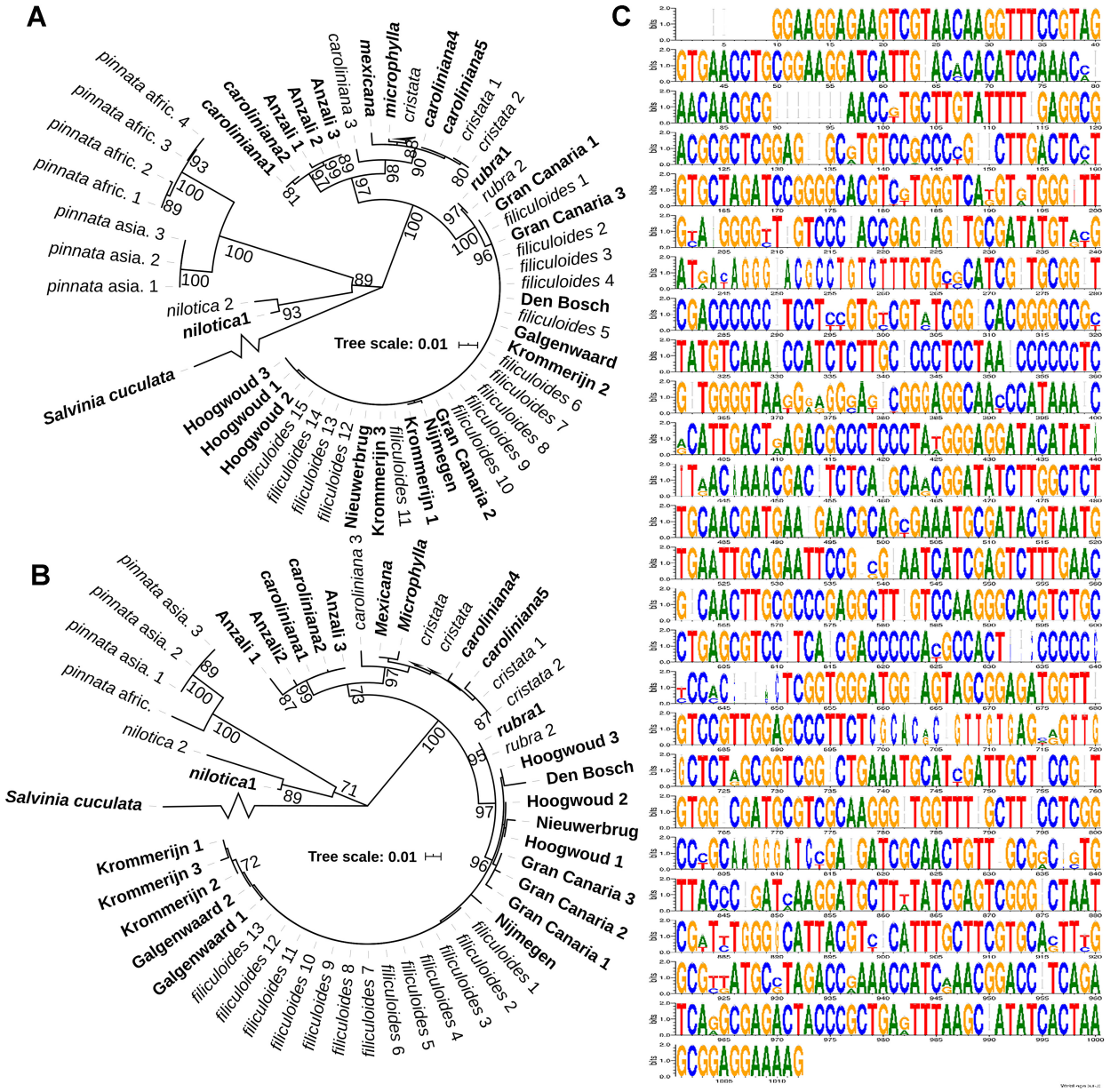

Figure 3. Chloroplast marker regions in strains from the Netherlands, Spain and Iran, and the rRNA intergenic regions of the *A. filiculoides* genome.

(A) Phylogenetic tree of chloroplast regions trnL-trnF. 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 the Madeira et al. (2016) study.

(B) Phylogenetic tree of chloroplast regions trnG-trnR.

(C) Sequence logo of all ITS1 gene regions in the *A. filiculoides* genome (Li *et al.,* 2018). The gene regions were extracted and aligned (MEGA 7) then used for *in-silico* PCR. Results were visualized using WebLogo (vs 2.8.2 Crooks *et al.,* 2004): at each position of the ITS, the height of the stack indicates the sequence conservation, while the height of symbols within the stack indicates the relative frequency of each base.

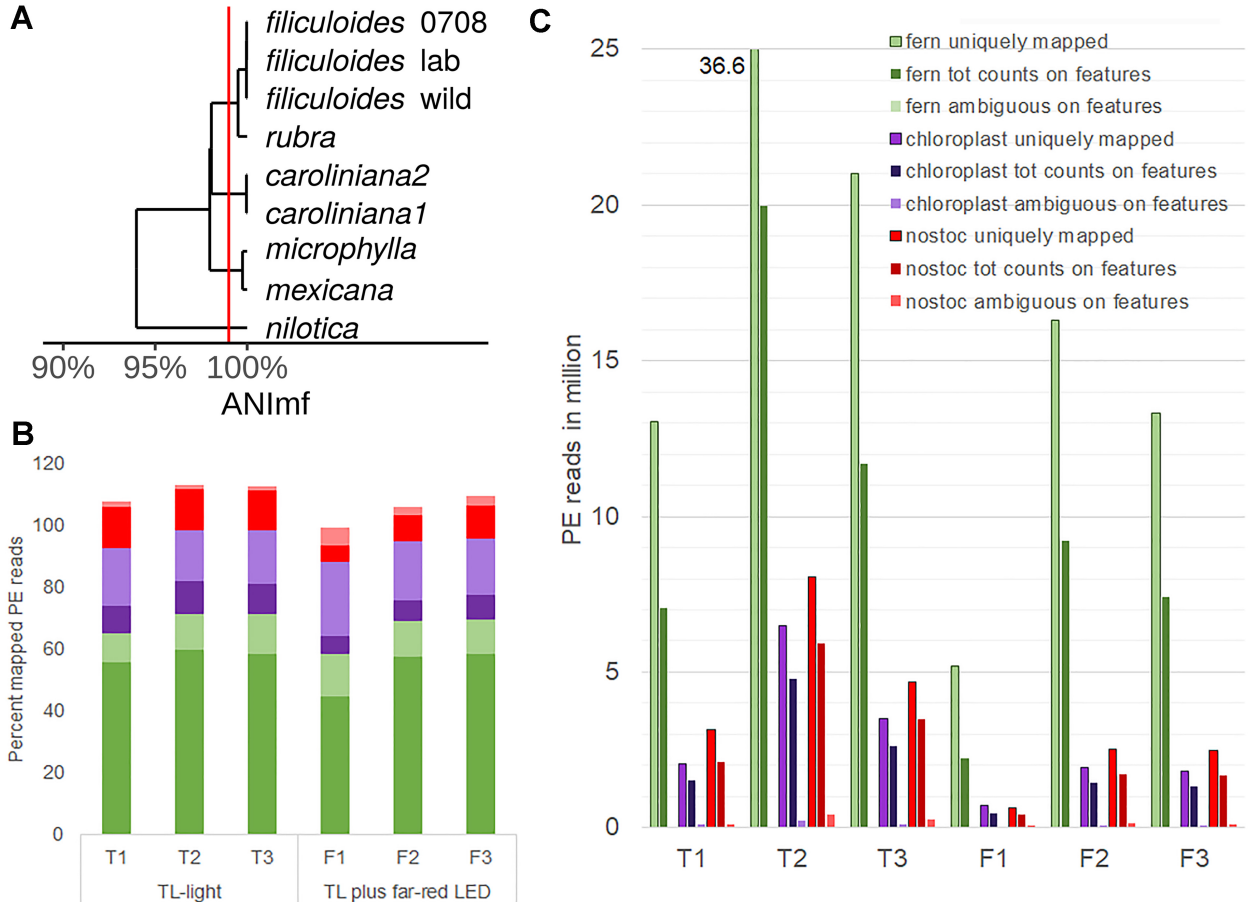

Figure 4. Profiling eukaryotic and prokaryotic RNA simultaneously in *A. filiculoides* (dual RNAseq).

(A) Similarity of the *N. azollae* 0708 genome and Metagenome Assembled Genomes (MAGs) from *N. azollae* strains in *Azolla* species. Average Nucleotide Identity (ANI) was calculated with the dRep implementation of nucmer using the ANImf preset (Kurtz *et al.*, 2004; Olm *et al.*, 2017). Pairwise average ANI is shown in an UPGMA dendrogram (de Vries & Ripley, 2016; Wickham, 2011). Aligned MAGS covered over 90% of the reference *N. azollae* 0708.

(B) Proportion (percent) of uniquely aligned PE reads (dark color) and PE reads aligning at multiple loci (light color) on the genomes of fern (green), fern chloroplast (purple) and *N. azollae 0708* (red). PE reads were from samples of sporophytes grown for a week on TL-light (replicates T1, T2, T3) or TL light with far-red LED (replicates F1, F2, F3). STAR-aligner was used in PE mode.

(C) Yields of PE uniquely mapped compared with PE read counts on features and PE reads that STAR aligned ambiguously to several features. Features were specified by current general features files (GFF) of each of the fern (green), chloroplast (purple) and *N.azollae* 0708 (red) genomes.

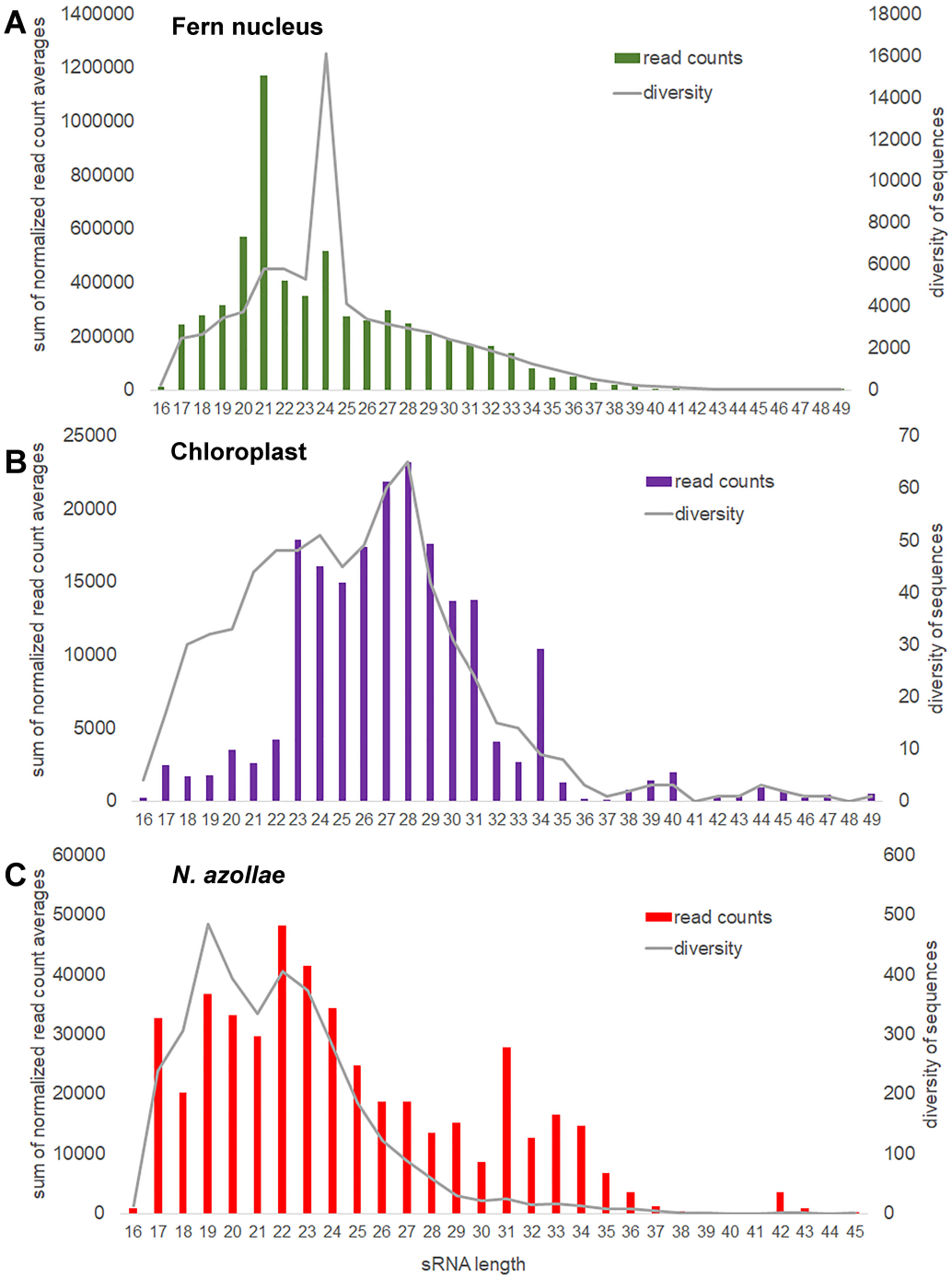

Figure 5. sRNA size distribution and diversity for each of the genomes in the *A. filiculoides* symbiosis. sRNA reads were mapped with 0 mismatches, no overhang, to the concatenated genomes; bam2fasta was then used to extract sRNA sequences aligning per genome, the fasta files obtained were then collapsed to compute read counts for each sRNA, the lists obtained were then joined to a read count matrix imposing at least 1 read per sRNA in each of the three replicate sample types: f (far-red), t (no far-red) and fn (far-red and nitrogen in the medium). sRNA counts were then normalized to the total read counts per sample for that genome.

(A) Fern nucleus.

(B) Chloroplast, included samples from sporophytes without cyanobacteria in the analysis.

(C) *N. azollae*, read count files joined imposing at least 10 reads per sRNA in every sample.

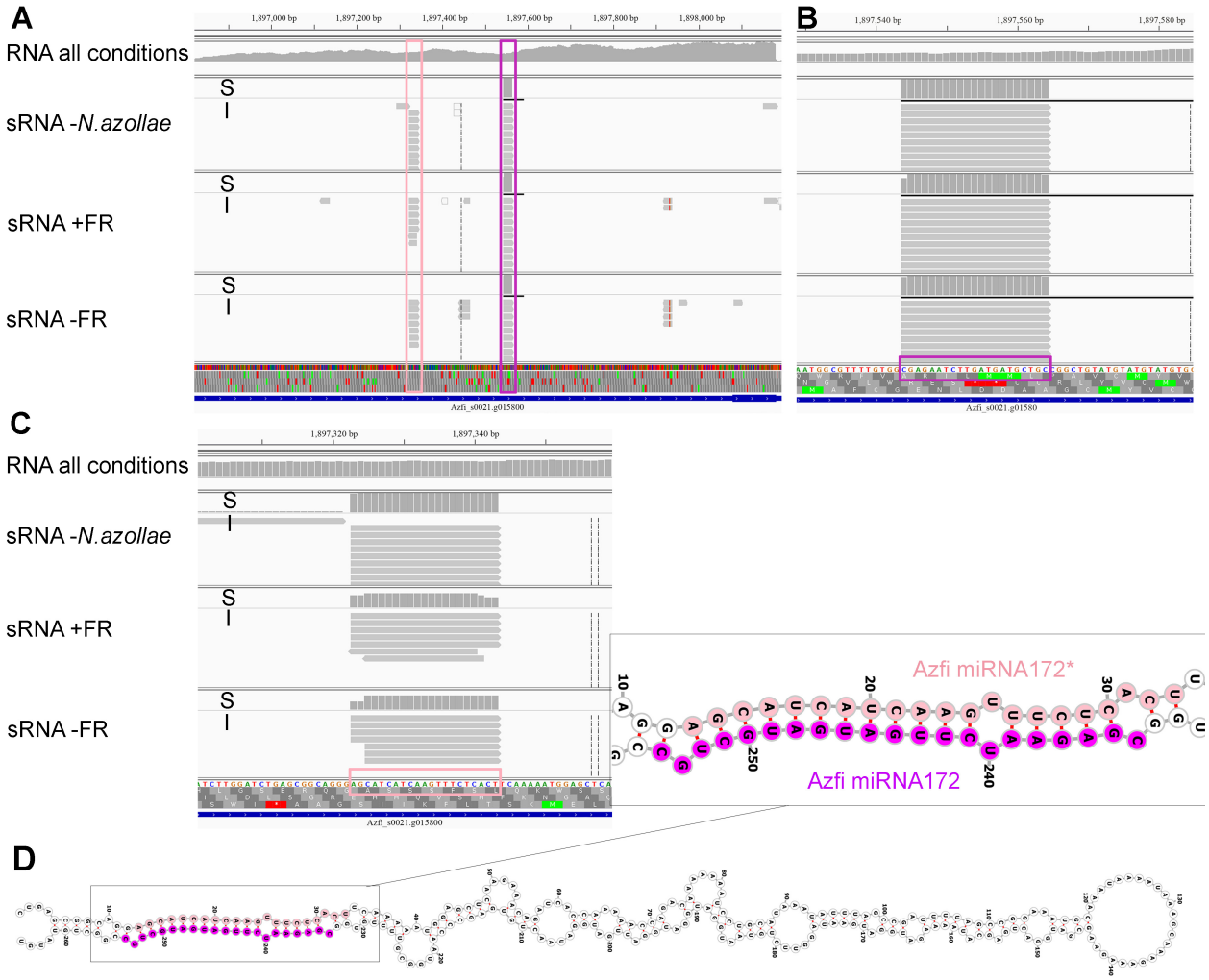

Figure 6. The miRNA 172α locus in *A.filiculoides*. Alignments at the Azfi_miRNA172α locus are visualized using the Integrative Genomics Viewer (Thorvaldsdóttir *et al.,* 2013); S, read summary; i, individual reads. **(A)** Reads from pre-miRNA at the miR172α locus. RNAseq all conditions, all reads pooled from the diel RNAseq experiment in (Brouwer *et al.,* 2017).

(B) Reads of the miR172 in sRNA libraries including those of ferns without *N.azollae* (sRNAseq-*N.azollae*), ferns grown with and without far-red LED (sRNAseq+FR, sRNAseq TL-FR).

(C) Reads of the miR172*.

(D) Azfi_miRNA172α hairpin folded using Vienna RNAfold (Gruber *et al.,* 2008; predicts -104,07 kcal mol^{-1} at 22 °C upon folding); miRNA (purple) and miRNA* (pink) stem arms are shown with 3'overhangs.

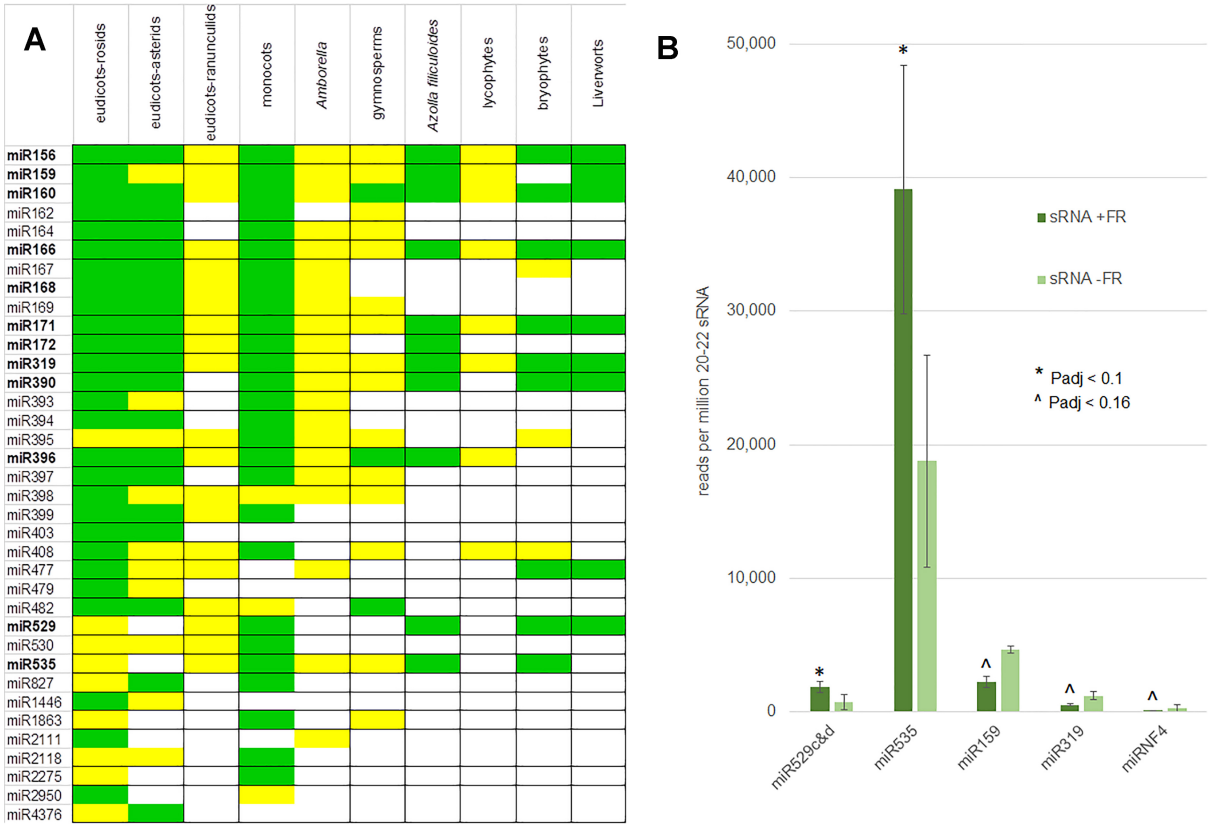

Figure 7. Conserved and far-red light responsive miRNA in the fern *A. filiculoides*.

(A) Conserved miRNA in the *A.filiculoides* fern compared to other land plant lineages. miRNA from the fern were predicted computationally by miRDEEP-P2 using sRNA seq reads then curated manually, the remainder of the table was from Axtell & Meyers, 2018. Green, high confidence miRNA; yellow low confidence miRNA.

(B) miRNA with altered steady state in response to far-red light. Shown are average read counts per million 20-22 nt reads in samples of sporophytes after one week on TL light with far-red LED (f) or without (t); raw read counts were submitted to DESeq2 for statistical analyses yielding P-adjusted values: * indicates Padj <0.1; ^ indicates Padj <0.16.

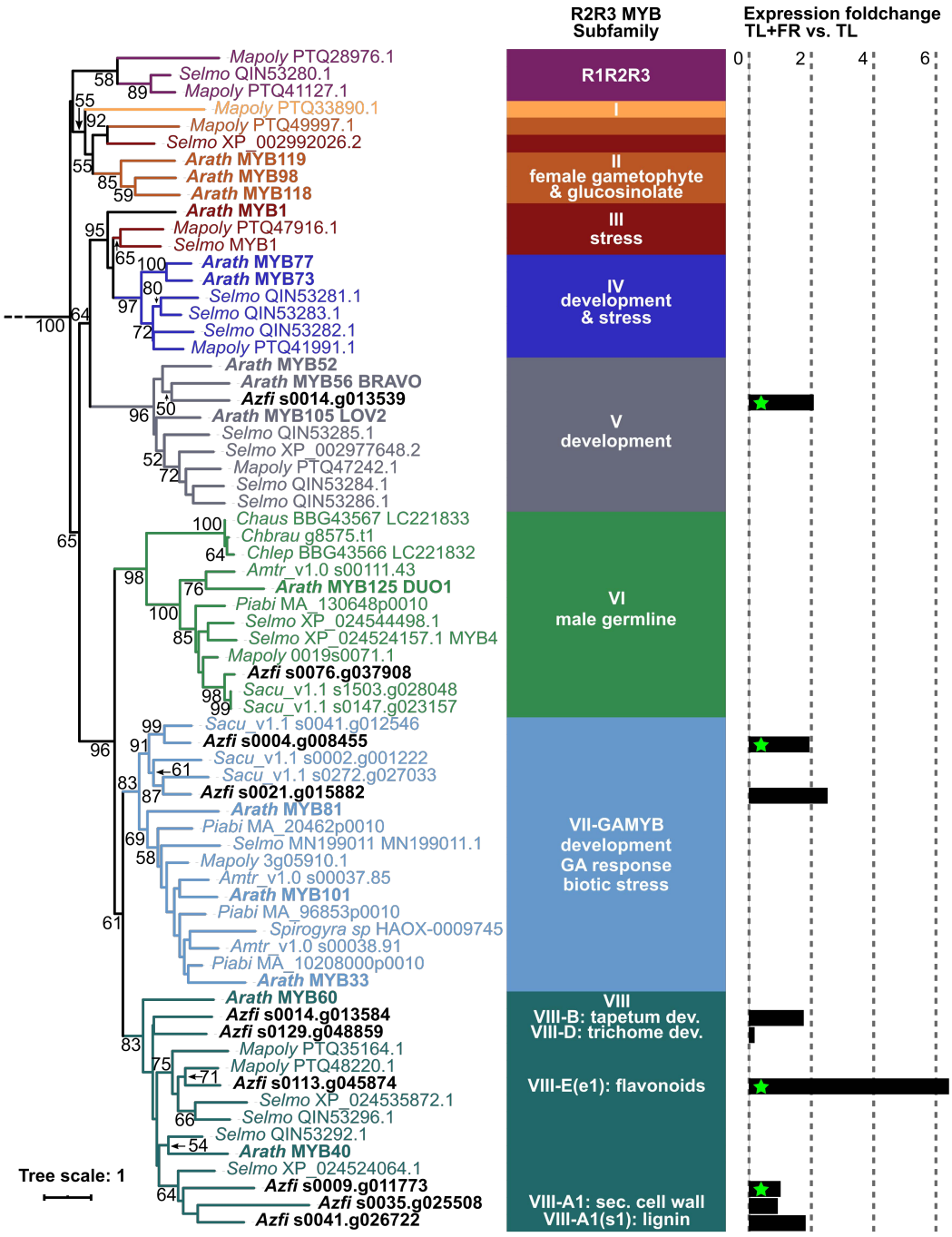

Figure 8. R2R3MYB phylogenetic analyses and fold change comparing sporophytes with and without FR LED. Sequences were extracted from the genome browsers of each species, aligned with MAFFT linsi or einsi, then trimmed with trimAL. Phylogenetic inferences were computed with IQtree and its internal model fitter; non-parametric bootstrap values are shown. *Chara autralis* (Ca), *Chara braunii* (Cb), *Chara leptospora* (Cl), *Marchantia polymorpha* (Mp), *Selaginella moellendorfii* (Sm), *Azolla filiculoides* (Azfi),), *Salvinia cuculata* (Sc), *Picea abies* (Pa), *Amborella trichopoda* (Atri), *Arabidopsis thaliana* (At). Sequence names are color coded after the R2R3 MYB subfamilies defined in Jiang & Rao (2020). Fold change in response to FR was calculated by DESeq2, green stars mark significant changes with Padj <0.1.