

Origin and evolution of the nuclear auxin response system

Sumanth K. Mutte^{1,#}, Hirotaka Kato^{1,#}, Carl Rothfels², Michael Melkonian³,
Gane Ka-Shu Wong^{4,5,6}, and Dolf Weijers^{1,*}

¹ Laboratory of Biochemistry, Wageningen University, Stippeneng 4, 6708 WE Wageningen, The Netherlands

² Department of Integrative Biology, University of California, Berkeley, CA, United States of America

³ Botanical Institute, Cologne Biocenter, University of Cologne, D50674 Cologne, Germany

⁴ Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada

⁵ Department of Medicine, University of Alberta, Edmonton, Alberta, Canada

⁶ BGI-Shenzhen, Bei Shan Industrial Zone, Yantian District, Shenzhen, China

These authors contributed equally to this work

*For correspondence: dolf.weijers@wur.nl

1 **Abstract**

2

3 The small signaling molecule auxin controls numerous developmental
4 processes in land plants, acting mostly by regulating gene expression. Auxin
5 response proteins are represented by large families of diverse functions, but
6 neither their origin nor their evolution is understood. Here we use a deep
7 phylogenomics approach to reconstruct both the origin and the evolutionary
8 trajectory of all nuclear auxin response protein families. We found that,
9 while all subdomains are ancient, a complete auxin response mechanism is
10 limited to land plants. Functional phylogenomics predicts defined steps in
11 the evolution of response system properties, and comparative
12 transcriptomics across six ancient lineages revealed how these innovations
13 shaped a sophisticated response mechanism. Genetic analysis in a basal land
14 plant revealed unexpected contributions of ancient non-canonical proteins in
15 auxin response as well as auxin-unrelated function of core transcription
16 factors. Our study provides a functional evolutionary framework for
17 understanding diverse functions of the auxin signal.

18

19

20

21 **Introduction**

22 The plant signaling molecule auxin controls a multitude of growth and
23 developmental processes in land plants by influencing cell differentiation,
24 division, and expansion which is mainly mediated by changes in gene
25 expression via the nuclear auxin pathway (NAP; Kato et al., 2017b).
26 Perturbation of this gene regulatory pathway interferes with most, if not all,
27 developmental responses (Weijers and Wagner, 2016). The NAP encompasses
28 three dedicated protein families (Figure 1A, B). Auxin is perceived by a
29 co-receptor complex consisting of TRANSPORT INHIBITOR RESPONSE
30 1/AUXIN SIGNALING F-BOX (TIR1/AFB) and AUXIN/INDOLE-3-ACETIC
31 ACID (Aux/IAA) proteins (Dharmasiri et al., 2005; Kepinski and Leyser,
32 2005). Subsequent ubiquitination of the Aux/IAA proteins causes their
33 degradation in the 26S proteasome (Gray et al., 2001). When not degraded,
34 Aux/IAA proteins bind to and inhibit DNA-binding transcription factors, the
35 AUXIN RESPONSE FACTORS (ARF) (Kim et al., 1997). Thus, auxin
36 de-represses ARFs, allowing these to activate or repress their direct target
37 genes (Ulmasov et al., 1999).

38 A central question in plant biology is how this simple transcriptional
39 system with only three dedicated components can generate a multitude of
40 local auxin responses to support various developmental functions. In
41 flowering plants such as *Arabidopsis thaliana*, it is evident that the size of
42 TIR1/AFB (six members), Aux/IAA (29 members) and ARF (23 members)
43 gene families allows combinatorial assembly of distinct, local auxin response

44 pathways. Given that diversity in auxin responses follows from
45 diversification in its response proteins, is it is still unclear how NAP
46 complexity evolved from simpler ancestral states. Furthermore, while
47 intuitive, a key question is whether increased NAP complexity indeed
48 enabled more complex and diverse auxin responses during plant evolution. A
49 third important question is where, when, and from what precursors the NAP
50 originated.

51 Eukaryotic photosynthetic organisms diverged into three groups,
52 Glaucophyta, Rhodophyta (red algae), and Viridiplantae more than 1.5
53 billion years ago (Yoon et al., 2004). Viridiplantae are further classified into
54 chlorophyte algae and streptophytes, which include charophyte algae and
55 land plants.. Bryophytes represent the earliest diverging land plants and
56 consist of three groups: hornworts, liverworts and mosses. After the split
57 from bryophytes, ancestral vascular plants changed their life cycle from
58 haploid-dominant to diploid-dominant and established a vascular system
59 and root architecture, forming the group of lycophytes and euphylllophytes
60 (ferns, gymnosperms and angiosperms).

61 The presence of a functional NAP with reduced genetic redundancy
62 has been reported in model bryophytes (Flores-Sandoval et al., 2015; Kato et
63 al., 2015; Prigge et al., 2010; Rensing et al., 2008), whereas the presence of
64 endogenous auxin is also reported in wide range of algal species (Žižková et
65 al., 2017). Thus, a prediction is that the auxin response system may predate
66 land plants, and that complexity evolved after the divergence of ancestral

67 vascular plants from bryophytes. A key challenge is to identify the origin of
68 the NAP system, as well as to reconstruct the steps in the evolution of its
69 complexity. However, only little genome data are currently available from
70 non-flowering land plants (Rensing, 2017), which makes such inferences
71 extremely challenging. In addition, studies using only selected model species
72 bear the risk of generalizing observations from non-representative genomes,
73 due to species-specific gene-duplication, -loss, and -diversification. Therefore,
74 it is necessary to analyze multiple species to understand evolutionary trends.

75 Here we describe a deep phylogenomic analysis of NAP components
76 using a large transcriptome dataset with more than 1,000 plant species
77 including many algae. This extensive dataset enabled us to reconstruct the
78 ancestral states of auxin response gene families at key nodes in plant
79 evolution. We infer plausible origins and evolutionary patterns for each
80 auxin response gene family and predict auxin response properties at
81 evolutionary nodes. Using comparative RNA-seq of six species, we tested and
82 extended these predictions. Finally, we used a genetic strategy in a
83 bryophyte to demonstrate surprising (non)contributions of the ancient ARF
84 class as well as of deeply conserved non-canonical NAP components to auxin
85 signaling. Our work provides a deep view into early steps in the origin,
86 evolution and design principles of the multi-functional auxin response
87 system.

88 **Results**

89

90 **A phylogenomic strategy for reconstructing ancestral states**

91 To reconstruct origin and early diversification in auxin response gene
92 families, we designed a strategy (Figure 1—Figure supplement 1) that uses a
93 large transcriptome dataset (OneKP) including multiple species for each
94 major branch in plant species phylogeny (Matasci et al., 2014). The depth
95 and quality of each individual RNA-seq-derived transcriptome is limited and
96 a further caveat of transcriptome-based gene identifications is that the
97 number of genes may be underestimated if a gene is not expressed under the
98 sampling conditions or in the sampled tissue. However, the availability of
99 transcriptomes from multiple tissue samples of multiple related species,
100 should allow deduction of the ancestral state that defines the gene
101 complement at each evolutionary node. It should be stressed that this
102 number represents the ancestral state at a given node, and species-specific
103 gene duplications and gene losses will have modified the gene complement in
104 individual species. Given our focus on early events in auxin response
105 evolution, we have used all available transcriptomes of red algae, green
106 algae, bryophytes, lycophytes, ferns, and gymnosperms from the OneKP
107 dataset (Supplementary file 1). We also included all available angiosperm
108 species in the Chloranthales, Magnoliids and ANA grade, as well as several
109 species in both monocots and dicots (Supplementary file 1). For reference
110 and quality control purposes, we included genome-based sequences from well

111 annotated model species.

112

113 **Origin of nuclear auxin response components**

114 Each of the three auxin response protein types (ARFs, Aux/IAAs, and
115 TIR1/AFBs) are multi-domain proteins and we initially focused on the origin
116 of these proteins. Therefore, we asked where domains, or parts thereof, were
117 found, and at what node the multi-domain proteins first appear.

118 ARF proteins carry an N-terminal DNA-binding domain (DBD)
119 which consists of a composite dimerization domain (DD; made up of two
120 separate subdomains [DD1 and DD2] that fold into a single unit), a B3-type
121 DNA-interaction domain, and an ancillary domain (AD) of unknown function
122 (Figure 1C; Boer et al., 2014). The C-terminal Phox and Bem 1 (PB1) domain
123 is shared among ARF and Aux/IAA proteins and mediates homo- and
124 hetero-oligomerization (Korasick et al., 2014; Nanao et al., 2014). Finally,
125 ARFs contain a less well-defined Middle Region (MR) separating the PB1
126 and DBD (Figure 1C). In red algae, we found proteins containing an
127 N-terminal portion of DD1, DD2, and AD, lacking a B3 or PB1 domain, but
128 instead flanked by a C-terminal bromodomain (InterPro ID: IPR001487;
129 Figure 1C). The DD1 and DD2 motifs in red algae are spaced by 20–30
130 conserved amino acids, which is much shorter than the B3 domain (~120
131 amino acids; Supplementary file 2). In chlorophytes, we found a protein with
132 only AD, flanked by a DNA-binding AT-rich interaction domain (ARID;
133 InterPro ID: IPR001606; Figure 1C). Furthermore, we found separate

134 proteins that either represented a B3 or a PB1 domain (Figure 1C). Thus, all
135 ARF subdomains had been established before the split of the streptophytes,
136 but not combined in a single protein. In contrast, we discovered full-length
137 ARF-like proteins containing a DBD with a B3 domain inserted between DD
138 and AD in charophytes (Figure 1C and Figure 1—Figure supplement 2).
139 Land plant ARFs can be grouped into three classes, A, B and C (Finet et al.,
140 2013). Based on transactivation assays, class A and B ARFs are classified as
141 transcriptional activators and repressors, respectively (Kato et al., 2015;
142 Ulmasov et al., 1999). Phylogenetic analysis revealed that the ARF-like
143 proteins in charophytes fall in two sister clades and likely represent separate
144 precursors of class C-ARFs (proto-C-ARFs) and A/B-ARFs (proto-A/B-ARFs)
145 of land plants (Figure 2 and Figure 1—Figure supplement 2). Interestingly,
146 we found the PB1 domain only in proto-C-ARFs, which could however be due
147 to sparse sampling in some charophyte lineages (Figure 1—Figure
148 supplement 2).

149 To understand if the proto-ARFs share conserved, functionally
150 important residues, we generated homology models based on available DBD
151 crystal structures of *A. thaliana* ARF1 and ARF5 (Boer et al., 2014). As no
152 class C-ARF structure is known, we first modeled the *A. thaliana* ARF10
153 DBD to compare with proto-C-ARFs. Next, homology models for proto-ARFs
154 in *Spirogyra pratensis* (SpARF; proto-C-ARF) and *Mesotaenium caldariorum*
155 (McARF; proto-A/B-ARF) were generated. We also included all three ARFs of
156 the bryophyte *M. polymorpha* (MpARF1-3) representing each major class,

157 and compared all models to *A. thaliana* ARF structures. This analysis
158 revealed that all proto-ARFs likely share a conserved structural topology
159 (Figure 3A). Strikingly, all DNA-binding residues follow the spatial
160 restraints needed for DNA binding in all ARFs tested, suggesting a
161 conserved mode of DNA binding. On the other hand, dimerization residues
162 are conserved only in the (proto-)A/B-ARFs (McARF, MpARF1, and
163 MpARF2) but not in the (proto-)C-ARFs (SpARF, MpARF3, and ARF10).
164 These results clearly demonstrate that canonical ARF proteins were
165 established and differentiated into two classes in charophyte algae.

166 In addition to the proteins with canonical ARF-like structure, we
167 found a group of charophyte proteins consisting of an AP2 DNA binding
168 domain along with B3 and PB1 domains (Figure 1C). Phylogenetic analysis
169 showed that these proteins position along with RELATED TO ABI3 AND
170 VP1 (RAV) family in land plants (Figure 2B and Figure 1—Figure
171 supplement 2). Interestingly, land plant RAV proteins do not have a PB1
172 domain, and the B3 domain of RAV and ARF binds different DNA sequences
173 (Boer et al., 2014; Matias-Hernandez et al., 2014). Thus, we classify these
174 proteins as proto-RAV. In the charophyte green algae, the two classes of
175 proto-ARFs and proto-RAVs are found in various combinations in each
176 species (Figure 2A). While sequencing depth may be insufficient to detect all
177 proto-ARFs and proto-RAVs, there does not appear to be a conserved pattern
178 in the order of appearance and retention of these genes.

179 We next considered the origin the Aux/IAA proteins. These proteins

180 contain two functional small domains in addition to a C-terminal PB1
181 domain (Figure 1B, C). The N-terminal domain I recruits the TOPLESS
182 (TPL) transcriptional co-repressor (Szemenyei et al., 2008). Domain II
183 mediates the auxin-dependent interaction with TIR1/AFB and thus acts as a
184 degron (Dharmasiri et al., 2005; Gray et al., 2001; Kepinski and Leyser,
185 2005). Because domain I and II are too small for reliable BLAST searches,
186 we used the PB1 domain to identify potential family members. No
187 PB1-containing proteins were identified in red algae, while we found
188 proteins with a PB1 domain but no DBD in chlorophytes (Figure 1C).
189 Phylogenetic analysis based on the PB1 domain indicated their close
190 relatedness to the PB1 domain of proto-RAV family (Figure 2B and Figure
191 2—Figure supplement 1). PB1 domain-containing proteins that lack a DBD
192 were also found in many of the charophyte algae (Figures 1C, 3B and Figure
193 2—Figure supplement 1). Most of them were placed along with proto-RAV in
194 phylogenetic tree, but the sequences from *Coleochaetae irregularis* were
195 placed along with the Aux/IAA in land plants that is separate from the PB1
196 of both ARFs and proto-RAV proteins (Figures 2B, 3B and Figure 2—Figure
197 supplement 1). Even though the N-terminal part of the PB1 domain is not as
198 conserved as the C-terminal part, several critical residues were found to be
199 conserved in Aux/IAA-like sequences (Figure 3B). These results indicate that
200 the PB1 domain of land plant ARFs and Aux/IAAs had separate precursors
201 in charophytes. We could, however, not detect domain I or II in Aux/IAA-like
202 genes of charophyte algae, even when scrutinizing individual sequences. We

203 thus conclude that Aux/IAA proteins with all three functional domains are
204 limited to land plants.

205 Finally, we explored the origin of the TIR1/AFB auxin co-receptor
206 that consist of an N-terminal F-box domain that anchors the protein to the
207 other subunits in the SCF E3 ubiquitin ligase complex, and a C-terminal
208 leucine-rich repeat (LRR) domain that contains the auxin binding pocket.
209 Auxin acts as a molecular glue to stabilize the interaction between
210 TIR1/AFB and Aux/IAAs (Tan et al., 2007). The closest homolog of the
211 TIR1/AFB proteins in *A. thaliana* is CORONATINE INSENSITIVE 1 (COI1),
212 which functions as a receptor of the jasmonic acid (JA) phytohormone (Katsir
213 et al., 2008). In our homology search, we could not identify any proteins
214 showing homology to either TIR1/AFB or COI1 in red algae and chlorophytes
215 (Figures 1C and 2A). We did find many proteins showing homology to
216 TIR1/AFB and COI1 in the transcriptomes of charophyte algae (Figures 1C
217 and 2A). However, phylogenetic analysis indicated that these proteins form a
218 sister group to both TIR1/AFB and COI1 in land plants (Figure 4 and Figure
219 4—Figure supplement 2), suggesting that charophytes had an ancestor that
220 gave rise to both auxin and JA receptors. To infer whether the
221 TIR1/AFB/COI1-like proteins of charophytes function as receptors for auxin
222 or JA, we generated homology models of the TIR1/AFB/COI1-like proteins
223 from *K. nitens* and *S. pratensis*, as well as the bryophyte *M. polymorpha*
224 MpTIR1 and MpCOI1, using the *A. thaliana* TIR1 and COI1 crystal
225 structures (Sheard et al., 2010; Tan et al., 2007) as templates for modeling.

226 All of the charophyte proteins lack several LRR domains and C-terminal
227 caps (Supplementary file 2). Moreover, charophyte proteins lack the 3₁₀ helix
228 that is critical to form binding pockets for auxin or JA and
229 hormone-contacting residues (Figure 3C and Supplementary file 2). These
230 results suggest that the charophyte TIR1/AFB/COI1 precursor may not act
231 as an auxin or JA receptor, and we conclude that dedicated receptors for
232 auxin and JA were established only in land plants. Taken together, our
233 analyses suggest that the components of NAP were established in the
234 common ancestor of land plants by combining pre-existing components and
235 that the system evolved to regulate pre-existing transcription factors.

236
237 **Evolution of complexity in the nuclear auxin response system**

238 All three gene families have evolved to considerable size and diversity in
239 angiosperms, and this diversity is thought to underlie multifunctionality of
240 auxin as a hormone. We next aimed to reconstruct the evolutionary history of
241 auxin response components across all land plant lineages.

242 Consistent with previous descriptions (Finet et al., 2013), our
243 phylogenetic analysis showed that all land plant ARFs are divided into three
244 phylogenetic lineages (Figure 4 and Figure 1—Figure supplement 2). Within
245 the class C lineage, we did not find any duplications in the ancestors of
246 non-angiosperm species. The split that generated *A. thaliana* ARF10/16 and
247 ARF17 likely occurred early in angiosperm evolution, while the PB1 domain
248 was lost in the ARF17 group (Figure 4 and Figure 1—Figure supplement 2).

249 The class A-ARF is represented by a single copy in bryophytes and
250 lycophytes. We found that a subset of genes lacking the DBD diverged from
251 class A-ARFs in early land plants, is missing in hornworts and has been
252 retained in liverworts, mosses and lycophytes (non-canonical ARF, ncARF;
253 Figures 3B, 4 and Figure 2—Figure supplement 1). A further gene
254 duplication event in the ancestor of euphyllophytes gave rise to two class A
255 sub-families corresponding to *A. thaliana* ARF5/7/19 and ARF6/8,
256 respectively. In the ancestor of seed plants a gene duplication caused
257 differentiation between the *A. thaliana* ARF5 and ARF7/19 subfamilies
258 (Figure 4 and Figure 1—Figure supplement 2). Finally, two gene duplication
259 events in the ancestral angiosperms led to ARF6 and ARF8 and to a
260 parologue of ARF7/19, which was lost in *A. thaliana* (Figure 4 and Figure
261 1—Figure supplement 2).

262 Class B-ARFs are represented by a single gene in the ancestor of
263 liverworts, mosses, lycophytes, and ferns. However, no hornwort species
264 appears to contain class B-ARFs (Figure 4 and Figure 1—Figure supplement
265 2). Gene duplications in the ancestral gymnosperms gave rise to three class
266 B-ARF copies, one representing *A. thaliana* ARF3/4, another leading to *A.*
267 *thaliana* ARF2 and the third generating the remainder of the class B-ARFs
268 in *A. thaliana* (Figure 4 and Figure 1—Figure supplement 2). Notably, the
269 reported lack of the PB1 domain in ARF3 (Finet et al., 2013) is an
270 independent loss in the common ancestor of monocots and eudicots (Figure
271 1—Figure supplement 2).

272 Our data indicated that an ancestral Aux/IAA gene lacking domain I
273 and II had been established during the evolution of charophytes, while “true”
274 Aux/IAAs with all functional domains are found only in land plants (Figure
275 1C). In addition to one copy of “true” Aux/IAA, we found another set of deeply
276 conserved non-canonical Aux/IAA-like sequences that lack the domain I and
277 II (ncIAA; Figures 2B, 3B, 4, Figure 2—Figure supplement 1, and Figure
278 4—Figure supplement 1). Strikingly, while the Aux/IAAs have diversified
279 through gene duplications, the ncIAA is found only in a single copy in all
280 evolutionary nodes examined here, and is represented by IAA33 in *A.
281 thaliana*. In the ancestor of euphylophytes, gene duplication events gave
282 rise to three Aux/IAAs, which were retained in the ancestral seed plants
283 (Figure 4 and Figure 4—Figure supplement 1). Common ancestor of
284 angiosperms have eleven Aux/IAA proteins, which is more than triple the
285 number found in gymnosperms (Figure 4 and Figure 4—Figure supplement
286 1). Finally, in addition to the ancient ncIAA generated in a first duplication
287 event, several independent later events generated non-canonical family
288 members lacking domains. For example, the lack of domain II in IAA20,
289 IAA30, IAA31, IAA32, and IAA34 of *A. thaliana* appears to be an
290 independent loss in their respective lineages in the core angiosperms (Figure
291 4—Figure supplement 1).

292 Our data indicated that ancestral charophyte green algae had one
293 common ancestor for both auxin (TIR1/AFB) and JA (COI1) F-box
294 co-receptors, and following duplication in the ancestor of all land plants,

295 developed into two independent receptors (Figure 4 and Figure 4—Figure
296 supplement 2). The common ancestor of bryophytes and lycophytes had a
297 single orthologue of *A. thaliana* TIR1/AFB. Gene duplication events in the
298 ancestor of euphyllophytes gave rise to three subgroups; one leading to
299 TIR1/AFB1-3, one leading to AFB4/5 and another which is widely present in
300 many species including the angiosperms, but has been lost in some monocots
301 and dicots including *A. thaliana* (Figure 4 and Figure 4—Figure supplement
302 2).

303 Thus, our analysis of the patterns of diversification in the ARF,
304 Aux/IAA and TIR1/AFB families identifies the auxin response complement at
305 each evolutionary node, and in addition reveals deeply conserved
306 non-canonical family members. Notably, many changes occurred in the
307 composition of NAP from the common ancestor of lycophytes to
308 euphyllophytes, which may have led to complex auxin response.
309

310 **Multi-species comparative transcriptome analysis reveals evolution of**
311 **response complexity**

312 The complements of auxin response components identified from
313 phylogenomic analysis allow for clear predictions of which species possess a
314 functional transcriptional auxin response system. Based on our predictions,
315 only land plants should be able to respond. In addition, it is intuitive that the
316 number of components in auxin response will relate to the complexity of
317 response, but as yet there is no experimental basis for such relationship. To

318 experimentally address the competence of species to respond to auxin, and to
319 explore the relationship between auxin response components and the
320 qualitative and quantitative aspects of auxin response, we performed
321 comparative transcriptome analysis. We selected six species that belong to
322 different ancient lineages and that each have a different complement of
323 auxin response components (Figure 5A). We used the charophyte algae
324 *Klebsormidium nitens* and *Spirogyra pratensis*, the hornwort *Anthoceros*
325 *agrestis*, the liverwort *Marchantia polymorpha*, the moss *Physcomitrella*
326 *patens*, and the fern *Ceratopteris richardii*. To detect only early
327 transcriptional responses, we treated plants with auxin for 1 h, and
328 performed RNA-seq followed by *de novo* transcriptome assembly and
329 differential gene expression analysis. To avoid inactivation of the natural
330 auxin IAA by conjugation or transport, we treated with 10 µM of the
331 synthetic auxin 2,4-dichlorophenoxyacetic acid (2,4-D). This compound was
332 shown to behave like IAA in the context of the NAP (Tan et al., 2007).
333 Importantly, 68–90% of the differentially expressed genes (DEG) from *de*
334 *novo* assemblies in *K. nitens*, *M. polymorpha* and *P. patens* matched with
335 genome-based differential gene expression performed in parallel (Figure
336 5—Figure supplement 1), thus validating our approach.

337 Transcriptome analysis after prolonged auxin treatment in *P. patens*
338 had identified a large set of auxin-responsive genes (Lavy et al., 2016).
339 Indeed, we found 105 and 1090 genes to be auxin-regulated in *M.*
340 *polymorpha* and *P. patens*, respectively (Figure 5A). Likewise, we found 159

341 and 413 genes to be auxin-regulated in *A. agrestis* and *C. richardii* (Figure
342 5A). Unexpectedly, despite lacking Aux/IAA and dedicated TIR1/AFB genes,
343 both charophyte algae species showed a strong transcriptional response to
344 2,4-D treatment. A total of 1094 and 1681 genes were differentially
345 expressed in *K. nitens* and *S. pratensis*, respectively (Figure 5A). Thus, there
346 is a clear transcriptional response to 1 hour of 2,4-D treatment in all species
347 analyzed, yet the number of genes is different, with an exceptionally large
348 number of responsive genes in charophytes. We next determined if the
349 number of DEG correlates with gene number in each transcriptome
350 assembly (Figure 5 – Figure Supplement 2), and found that differences in
351 DEG among species can not be explained by total gene number.

352 We next addressed whether there were differences in the
353 characteristics of regulation. Both charophyte species showed a high
354 percentage of gene repression. Only 37% and 33% of DEG were activated in
355 *K. nitens* and *S. pratensis*, respectively (Figure 5A). In contrast, the
356 distribution of fold change amplitude values differed between the two
357 charophytes where *S. pratensis* showed a general shift towards larger
358 amplitudes of regulation (Figure 5A). Even though the complement of auxin
359 response proteins are different, all three bryophytes showed a similar
360 pattern: 36–53% of DEG were activated, with very few genes showing an
361 amplitude over 2-fold up- or down-regulation (Figure 5A). In contrast, 82% of
362 DEG were activated in *C. richardii*. We also found that there was a notable
363 difference in the distribution of fold-change values, with a larger fraction of

364 genes being more strongly activated (maximum 28 fold; Figure 5A).

365 We found that the number of auxin-responsive genes is positively
366 correlated with the number of ARFs in land plants as seen in the expanded
367 number of ARFs and DEG in *P. patens* and *C. richardii*. A switch to gene
368 activation is not correlated with the number of ARFs, but rather with a
369 duplication in the class A-ARFs in the ancestor of euphylophytes and/or
370 increase of Aux/IAA and TIR1/AFB. The increase in amplitude of
371 auxin-dependent gene regulation in *C. richardii* could be a consequence of
372 higher activation upon treatment, increased repression in the absence of
373 auxin, or both. To determine its basis, we compared normalized expression
374 values for the 20 top-most auxin activated, and the 20 least auxin activated
375 genes in all species (Figure 5B). This revealed that the increased amplitude
376 of auxin regulation in *C. richardii* is not correlated with increased expression
377 in the presence of auxin, but rather caused by reduced expression in its
378 absence. This quantitative property of the auxin response system is
379 correlated with the increased numbers of Aux/IAA genes.

380

381 **Identification of a deeply conserved auxin-dependent gene set in land**
382 **plants**

383 Given that the mechanism of auxin response is ancient and conserved among
384 all land plants, a key question is whether responses in different species
385 involve regulation of a shared set of genes. Classical primary
386 auxin-responsive genes—the Aux/IAA, GH3 and SAUR families—were

387 shown to be auxin responsive in many angiosperm species (Abel and
388 Theologis, 1996). In our RNA-seq analysis, we found different bryophyte
389 species to show auxin-dependence in only some of these three gene families
390 (Fig 6A), yet no species showed regulation of all three gene families. In
391 contrast, *C. richardii* displayed auxin-dependence of members of all three
392 gene families (Figure 6A). Given that the Aux/IAA and GH3 proteins
393 themselves regulate auxin levels or response, this indicates that a robust
394 feedback mechanism evolved after the split of bryophytes and vascular
395 plants.

396 We further selected several known auxin-responsive genes in
397 angiosperms and determined the expression of their homologues in our
398 RNA-seq dataset. It is known that class II homeodomain-leucine zipper
399 (*C2HDZ*), *WIP*, and *YUCCA* (*YUC*) genes are regulated by auxin in *A.*
400 *thaliana* (Crawford et al., 2015; Sawa et al., 2002; Takato et al., 2017). We
401 identified *C2HDZ* and *WIP* to be commonly activated by auxin in all land
402 plants in our RNA-seq (note that no *WIP* gene was identified in the *A.*
403 *agrestis* assembly). Indeed, qPCR analysis confirmed auxin-activation of
404 *C2HDZ* (Figure 6B). We also identified the auxin biosynthesis gene *YUC* to
405 be commonly down-regulated among multiple land plant species (except *A.*
406 *agrestis*), and qPCR analysis demonstrated this to be true in *A. agrestis*, as
407 well (Figure 6B). While homologues of *C2HDZ* were detected in the
408 charophyte assemblies, none was regulated by auxin, which supports the
409 different nature of the auxin response system in these species. In summary,

410 land plants share a deeply conserved set of auxin up- and down-regulated
411 genes.

412

413 **Contributions of ancient components to auxin response**

414 Our phylogenomic analysis identified several components that are deeply
415 conserved, yet whose contributions to auxin response are unknown: two
416 deeply conserved non-canonical auxin signaling components lack important
417 domains (*ncIAA* and *ncARF*), while class C-ARFs diverged from all other
418 ARFs in green algae prior to establishment of the NAP. To investigate the
419 biological roles of these genes, we chose the liverwort *M. polymorpha*, the
420 only genetically tractable model plant encoding *ncIAA*, *ncARF* and C-ARF
421 genes. We first addressed *ncIAA* and *ncARF* function and performed
422 CRISPR/Cas9-mediated mutagenesis (Sugano et al., 2014) to obtain two
423 different alleles for each gene which presumably cause a loss-of-function by
424 frame shift mutation (*nciaa-6*, *nciaa-10*, *ncarf-2*, *ncarf-10*; Figure 7A, Figure
425 7—Figure supplement 1A, B,E). To investigate whether *ncIAA* and *ncARF*
426 are involved in auxin response, we grew mutants on auxin-containing
427 medium. Exogenously supplied auxin causes severe inhibition of thallus
428 growth and increased formation of rhizoids in wild-type (Figure 7B; Ishizaki
429 et al., 2012). *nciaa* mutants showed auxin response similar to wild-type,
430 while growth inhibition was strongly suppressed in *ncarf* mutants although
431 rhizoid formation was still promoted by auxin (Figure 7B). We next selected
432 two auxin-up-regulated genes (*EXP* and *WIP*) and one auxin-down-regulated

433 gene (*YUC2*, Eklund et al., 2015), and examined their expression in all
434 mutants by qPCR analysis (Figure 7C). In *nciaa* mutants, the expression of
435 auxin-up-regulated genes responded similarly to the wild-type, while the
436 expression of the auxin-repressed *YUC2* gene was significantly reduced in
437 the absence of auxin, but similarly repressed by auxin. In *ncarf* mutants, the
438 basal expression of auxin-upregulated genes was similar to WT, while the
439 expression after auxin treatment was significantly reduced in the mutants.
440 The expression of *YUC2* was reduced in mock condition and auxin treatment
441 did not change the expression. Thus, in *M. polymorpha*, ncIAA may have a
442 function in gene expression, but is not critical for auxin response itself. On
443 the other hand, ncARF represents a novel positive regulator of both
444 auxin-dependent gene activation and repression.

445 Finally, we focused on C-ARF function. While partial mutants have
446 been reported in *A. thaliana*, no plants completely lacking C-ARF have been
447 described. We used CRISPR/Cas9 gene editing to generate a series of
448 loss-of-function mutants in MpARF3, the single C-ARF of *M. polymorpha*
449 (*arf3^{ge1}-1*, *arf3^{ge1}-2*, *arf3^{ge2}-1*; Figure 7A and Figure 7—Figure supplement 1C,
450 D). All three *arf3* mutants showed dramatic defects in development, notably
451 in vegetative propagules (gemmae) which arrested before maturation,
452 consistent with ubiquitous ARF3 protein accumulation in these structures
453 (Figure 7D-F, and Figure 7—Figure supplement 1G). Gemmae development
454 was previously shown to depend on auxin response and the class A-ARF
455 (MpARF1; Kato et al., 2015; Kato et al., 2017a), and we hence tested if

456 transcriptional responses to auxin were altered in *arf3* mutants. Strikingly,
457 all auxin-responsive genes we tested showed similar responses in WT and
458 *arf3* mutants, while *arf1* mutants showed no auxin responses (Figure 7C).
459 This result suggests that, class C-ARF in *M. polymorpha* have different
460 target genes from A-ARF and may not be critical for auxin-dependent gene
461 regulation.

462 **Discussion**

463

464 *Deep origin of nuclear auxin response in the ancestor of land plants*

465 Phylogenetic analysis and domain structural analysis provided many
466 insights into the origin of NAP and its evolutionary trajectory. All
467 subdomains of dedicated auxin-response proteins were recovered in
468 transcriptomes from red algae and chlorophytes, but the multidomain
469 protein appears only in the charophyte and land plant lineage. These
470 findings show that proto-ARF transcription factor was established during
471 the evolution of ancestral charophytes by combining existing domains.
472 However, given that no defined Aux/IAA and TIR1/AFB could be identified in
473 charophytes, a complete nuclear auxin response system is limited to land
474 plants. Ancestors of TIR1/AFB and COI1 co-receptors could be identified in
475 charophytes, but detailed residue analysis suggested these to be neither
476 auxin- nor JA receptor. Thus, duplication of this gene, as well as multiple
477 mutations in the LRR domain, must have preceded the deployment of these
478 proteins as co-receptors. Auxin-dependence of ARFs is mediated by
479 auxin-triggered degradation of Aux/IAA proteins, bridging ARF and
480 TIR1/AFB proteins through two protein domains: the ARF-interacting PB1
481 domain and the TIR1/AFB-interacting domain II. We did find charophyte
482 PB1-containing proteins that form a sister clade of land plant Aux/IAA.
483 However, domain II was not detected in these proteins. Along with
484 innovations in the proto-TIR1/AFB/COI1 protein, gain of a minimal degron

485 motif in the Aux/IAA precursor likely completed the auxin response system
486 in the early ancestor of land plants.

487

488 Auxin responses in algal species

489 Despite the lack of defined Aux/IAA and TIR1/AFB auxin co-receptor, the
490 charophytes *K. nitens* and *S. pratensis* showed an extensive transcriptional
491 response to exogenously supplied 2,4-D within 1 hour. A recent independent
492 study showed IAA-dependent gene expression in *K. nitens* upon prolonged
493 treatment with higher concentrations (100 µM for 10 h to 7 days; Ohtaka et
494 al., 2017). While *S. pratensis* has a proto-C-ARF, *K. nitens* does not appear to
495 have proto-ARFs. Thus, by definition this response system must be different
496 from the land plant auxin response system. Indeed, the charophyte
497 orthologue of core land plant auxin responsive genes (C2HDZ) did not
498 respond to 2,4-D and IAA. There was little, if any, overlap between
499 auxin-responsive transcripts in the two charophytes, and in qPCR
500 experiments on individual genes we noticed a high variability between
501 experiments (not shown). Thus, it appears that charophytes do respond to
502 auxin-like molecules, but this response may not be robust, or it may strongly
503 depend on growth conditions. Auxin resembles indole and tryptophan, and it
504 is possible that the response to auxin observed is in fact a metabolic response
505 to nutrient availability. Presence of endogenous IAA is observed in a wide
506 range of algal species including charophytes, chlorophytes, rhodophytes,
507 chromista, and cyanobacteria (Žižková et al., 2017). Moreover,

508 non-photosynthetic bacteria and fungi produce IAA and use it for
509 communication with plants and algae (Amin et al., 2015; Fu et al., 2015), and
510 thus it is likely that a response mechanism independent of the NAP exists in
511 these species.

512

513 **Function of the ancestral ARFs**

514 Our data clearly indicate that ARF transcription factors were established in
515 common ancestor of charophyte green algae and land plants. Structural
516 homology models suggest that all the important residues for DNA binding
517 are conserved in proto-ARFs, suggesting that these should bind the same
518 target DNA sequences. Given that there is a core set of auxin-regulated
519 genes shared in all land plants, an intriguing possibility is that proto-ARFs
520 already regulated this core set of genes that only became auxin-dependent
521 upon establishment of TIR1/AFB and Aux/IAA proteins. Identification of the
522 transcriptional targets of these proto-ARFs should help address this
523 question. In any event, proto-ARFs—as well as critical residues for DNA
524 binding—have been retained in many algal genomes for hundreds of millions
525 of years, which suggests that they perform a biologically relevant function.
526 Whether this function is related to the processes that auxin controls in land
527 plants is an open question.

528 Interestingly, our phylogenetic analysis indicated that the split
529 between class C- and class A/B-ARFs occurred in charophytes before the
530 establishment of Aux/IAA-TIR1/AFB co-receptor, and by extension likely

531 before proto-ARFs were auxin-dependent. This suggests that class C-ARFs
532 are fundamentally different from class A/B-ARFs. Indeed, genetic analysis in
533 *M. polymorpha* revealed that its C-ARF likely does not act in
534 auxin-dependent gene regulation. Several studies in *A. thaliana* showed that
535 C-ARFs are involved in auxin response but the proposed role was different
536 between studies (Ding and Friml, 2010; Liu et al., 2010; Mallory et al., 2005;
537 Wang et al., 2005; Yang et al., 2013). In addition, C-ARFs of *A. thaliana*
538 generally have weak affinity to Aux/IAA proteins (Piya et al., 2014). To
539 clarify the function of this ancient ARF subfamily, auxin-responsiveness of
540 C-ARF proteins and relationship with A- or B-ARFs should be investigated
541 in different species.

542

543 Novel components in auxin response

544 A surprising outcome of the phylogenomic analysis was the discovery of two
545 deeply conserved non-canonical proteins: ncIAA and ncARF. Charophytes
546 have an Aux/IAA-like protein containing a PB1 domain, but lacking domain
547 II, which is critical for auxin perception. This protein could regulate the
548 function of proto-ARF (or proto-RAV), but not in an auxin-dependent manner.
549 While the canonical Aux/IAA gave rise to a large gene family, the ncIAA clade
550 represented by a single member in every evolutionary node. The retention of
551 a single ncIAA gene across plants suggests a fundamental function.
552 Unfortunately, our mutant analysis in *M. polymorpha* could not reveal the
553 function of ncIAA in auxin response and development in vegetative phase.

554 ncIAA might have a function only in other developmental stages, or under
555 specific stress conditions or environmental signals. No mutant in the
556 *Arabidopsis* IAA33 gene has yet been reported, and perhaps such a mutant
557 will help understand the ancient function of this protein.

558 This work revealed that a class-A ARF-derived ncARF subfamily
559 lacking a DBD is evolutionarily conserved among liverworts, mosses, and
560 lycophytes. Mutant analysis using *M. polymorpha* clearly showed that
561 ncARF functions as positive regulator in transcriptional auxin responses.
562 There are two hypothetical models for ncARF function. 1) ncARF protects
563 canonical ARFs from AUX/IAA-mediated inactivation through the
564 interaction of PB1 domain. 2) ncARF interacts with target gene loci by
565 interaction with canonical ARFs and help activate expression by recruiting
566 co-factors. Irrespective of the mechanism of ncIAA and ncARF function,
567 future models of auxin response will need to incorporate these conserved
568 components.

569

570 *Functional impact of increased complexity in NAP components*

571 Through comparative transcriptomics we infer that the number of
572 DNA-binding ARF transcription factors scales with the number of
573 auxin-regulated genes. Both *P. patens* and *C. richardii* have an expanded set
574 of ARFs and display substantially more auxin-responsive genes than *A.*
575 *agrestis* and *M. polymorpha*. It is likely that later duplications in the ARF
576 family in the seed plants led to the thousands of auxin-responsive genes in

577 these species (Paponov et al., 2008).

578 Another key evolutionary change is the transition from mostly gene
579 repression to gene activation. We infer that this transition occurred in a
580 common ancestor of euphylophytes, and transcriptome analysis in *A.*
581 *thaliana* and *O. sativa* shows this pattern persists in angiosperms (Jain and
582 Khurana, 2009; Paponov et al., 2008). There is a defining difference between
583 bryophyte and euphylophyte ARF families—a persisting duplication in the
584 class A-ARFs. We hypothesize that the euphylophytes duplication created
585 an ARF copy that is more potent, or perhaps even specialized for gene
586 activation. The comparative transcriptomics also adds an interesting twist to
587 our understanding of the functional distinction among ARF classes. Class
588 A-ARFs are considered activators, and class B-ARFs repressors, perhaps
589 through competing with class A-ARFs (Lavy et al., 2016; Ulmasov et al.,
590 1999). Despite a complete lack of class B-ARFs, the hornwort *A. agrestis*
591 showed comparable auxin-dependent gene repression to the other
592 bryophytes, suggesting that auxin-dependent gene repression may not be
593 mediated by class B-ARFs. Based on these findings, the role of class B-ARFs
594 in auxin response may need to be reconsidered.

595 A remarkable difference between bryophyte and
596 euphylophyte auxin-dependent transcriptomes is the appearance of genes
597 with a large amplitude of regulation in the latter. Many auxin-responsive
598 genes that were first identified in angiosperms such as *A. thaliana* have very
599 high amplitudes (Lee et al., 2009), but this appears to be a later innovation

600 in the response system. The high amplitude is caused by more effective
601 repression of gene activity in the no-auxin state, a property that is likely
602 mediated by Aux/IAA proteins. Indeed, ferns have a much larger set of
603 Aux/IAA proteins, as do all seed plants, and we propose that expansion of the
604 Aux/IAA family enabled plants to articulate a clear distinction between on
605 and off states in auxin response. In summary, this analysis reveals several
606 design principles of the auxin response system.

607 **Materials and Methods**

608 Plant materials and culture condition

609 Male *M. polymorpha* strain Takaragaike-1 (Tak-1) was used as wild type and
610 cultured as described previously (Kato et al., 2015). *K. nitens* (NIES-2285), *P.*
611 *patens* (Gransden), and *A. agrestis* (Oxford) were cultured on BCD medium
612 (Cove et al., 2009) solidified with 1% agar under the same condition with *M.*
613 *polymorpha*. *S. pratensis* (UTEX928) was cultured on Guillard's Woods Hole
614 medium (Nichols, 1973), pH7.9 containing 1% agar under white light with a
615 16-h light/8-h dark cycle at 22°C. *C. richardii* (Hn-n) was cultured on C-fern
616 medium (Plackett et al., 2015) under continuous white light at 28°C.

617

618 Data used

619 Data access to 1000 plant transcriptomes was provided by the OneKP
620 consortium (www.onekp.com; Matasci et al., 2014). All the transcriptome
621 assemblies of the species from red algae, green algae, bryophytes, lycophytes,
622 monilophytes, gymnosperms and basal angiosperms that were safely
623 identified as non-contaminated has been used for this analysis
624 (Supplementary file 1). CDS and protein sequences encoding all the
625 orthologous genes in the three (ARF, Aux/IAA and TIR1/AFB) gene families
626 from *M. polymorpha*, *P. patens*, *Amborella trichopoda*, *Oryza sativa*, *Zea*
627 *mays*, *Solanum lycopersicum* and *A. thaliana* were obtained from Phytozome
628 ver11 (phytozome.jgi.doe.gov/pz/portal.html). Aux/IAA genes from *Picea*
629 *abies* were obtained from Spruce Genome Project (www.congenie.org). *K.*

630 *nitens* genome information was accessed from *Klebsormidium nitens*
631 NIES-2285 genome project (Hori et al., 2014).

632

633 Phylogeny construction

634 BLAST database for all the selected species were generated using
635 'makeblastdb' module in BLAST+ v2.2.28 (<https://blast.ncbi.nlm.nih.gov>).
636 Protein sequences from *A. thaliana*, *M. polymorpha* and *P. patens* were used
637 to query each database independently for each gene family using tBLASTn.
638 All the scaffolds with the BLAST hits were extracted from the respective
639 transcriptomes and further translated using TransDecoder (ver2.0.1;
640 <http://transdecoder.github.io>). This provided the CDS and protein sequences
641 of all the scaffolds of the BLAST hits to any of the query sequences. The
642 protein sequences were run through the InterProScan database
643 (ver5.19-58.0; <http://www.ebi.ac.uk/interpro>) to look for conserved domains.
644 MAFFT (ver7.123b; Katoh and Standley, 2013) iterative refinement
645 algorithm (E-INS-i) was used to align the CDS sequences. Alignment
646 positions with more than 50% gaps were removed using the Phyutility
647 program (ver2.2.6; <http://blackrim.org/programs/phyutility/>) before the
648 phylogeny construction. PartitionFinder (ver1.1.1; Lanfear et al., 2012) was
649 used to identify the most suitable evolutionary model for all the three gene
650 families using the complete trimmed alignments on all the domains.
651 Maximum likelihood algorithm implemented in RAxML (ver8.1.20;
652 Stamatakis, 2014) with General Time Reversible (GTR) model of evolution

653 under GAMMA rate distribution with bootstrapping criterion (up to a
654 maximum of 1000 bootstraps) was used for the phylogenetic analysis.
655 Obtained trees were visualized using the iTOL (ver3; <http://itol.embl.de/>)
656 phylogeny visualization program. Phylogenetic trees were cleaned up
657 manually for misplaced sequences as well as for clades with long branch
658 attraction.

659

660 Auxin treatment

661 *M. polymorpha* gemmae or thallus explant without meristem and *A. agrestis*
662 small thalli were planted on the medium covered with nylon mesh (100 µm
663 pore) and grown for 10 days. *P. patens* protonematal tissues were grown on
664 the medium covered with cellophane for 10 days. Sterilised spores of *C.*
665 *richardii* were grown for 2 weeks after which fertilization was performed by
666 adding 5 ml of water on the plate. 7 days after fertilization, prothalli
667 carrying sporophytic leaves were transferred on the medium covered with
668 nylon mesh and grown for a further 7 days, after which sporophytes
669 contained 3-4 leaves. After growing, plants with mesh or cellophane were
670 submerged into liquid medium and cultured for 1 day. After pre-cultivation,
671 2,4-D was added to a final concentration of 10 µM and plants were incubated
672 for 1h. Excess liquid medium were removed with paper towels and plants
673 were frozen in liquid nitrogen. *K. nitens* and *S. pratensis* were streaked on
674 solid medium and grown for 2 weeks. Algal cells were collected into 40 ml of
675 liquid medium and cultured for 1 day with shaking at ~120 rpm. Then 2,4-D

676 was added so that final concentration became 10 μM, followed by incubation
677 for 1 h with shaking. After auxin treatment, algal cells were collected using
678 filter paper and frozen in liquid nitrogen.

679

680 **RNA extraction and sequencing**

681 Frozen plant sample were grinded into fine powder with mortar and pestle.
682 RNA from *K. nitens*, *S. pratensis*, *M. polymorpha*, and *P. patens* were
683 extracted using Trizol Reagent (Thermo Fisher Scientific) and RNeasy Plant
684 Mini Kit (QIAGEN). RNA from *A. agrestis* and *C. richardii* were extracted
685 using Spectrum™ Plant Total RNA Kit (Sigma-Aldrich). Total RNA was
686 treated with RNase-free DNase I set (QIAGEN) and purified with RNeasy
687 MinElute Clean Up Kit (QIAGEN). RNA-seq library construction with
688 TruSeq kit (Illumina) and 100 bp paired-end sequencing with Hiseq4000
689 (Illumina) were performed by BGI TECH SOLUTIONS (HONGKONG).

690

691 **Quantitative RT-PCR**

692 cDNA was synthesized with iScript cDNA Synthesis Kit (Bio-Rad).
693 Quantitative PCR was performed using iQ™ SYBR® Green Supermix
694 (Bio-Rad) and CFX384 Touch™ Real-Time PCR Detection System. A
695 two-step cycle consisting of denaturation at 95°C for 10 seconds followed by
696 hybridization/elongation at 60°C for 30 seconds, was repeated 40 times and
697 then followed by a dissociation step. Three technical and biological replicates

698 were performed for each condition. PCR efficiencies were calculated using
699 CFX Manager (Bio-Rad) software in accordance with the manufacturer's
700 instructions. For *Marchantia polymorpha*, relative expression values were
701 normalized by the expression of *EF1α* (Saint-Marcoux et al., 2015). All
702 primers used for the analysis are listed in Supplementary file 3.

703

704 **RNA-seq data analysis**

705 Obtained raw fastq reads were checked for quality control using FastQC
706 (www.bioinformatics.babraham.ac.uk/projects/fastqc). *De novo*
707 transcriptome assemblies for all 6 species were generated using Trinity
708 (<http://trinityrnaseq.github.io>) with default settings. To avoid any possible
709 contamination from sequencing method and to improve the data quality, raw
710 reads from land plants were mapped against charophyte *de novo* assemblies
711 using Bowtie2 (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>) in
712 default settings and all the perfectly mapped pairs were removed, after
713 which new assemblies were generated from pure raw read data for each
714 species. In a similar way, contamination was removed in charophytes by
715 mapping them against land plant transcriptome assemblies. Once the pure
716 *de novo* transcriptome assemblies were generated, again Bowtie2 was used
717 to map individual sample to the respective transcriptome assemblies using
718 default parameters. Further, to improve the read count estimation and
719 reduce the redundancy in Trinity transcripts, Corset (Davidson and Oshlack,
720 2014) was implemented to estimate raw read counts using the Bowtie2

721 mapped alignment data. The obtained raw read counts were normalized and
722 differentially expressed genes ($P_{adj} < 0.01$) were identified using DEseq2
723 (Love et al., 2014) implemented in R Bioconductor package. All the RNAseq
724 raw reads were deposited in NCBI Short Read Archive (SRA) under the
725 BioProjectID: PRJNA397394 (www.ncbi.nlm.nih.gov/bioproject/397394).

726

727 Alignments and homology modelling

728 All other protein alignments mentioned in the manuscript were generated
729 using ClustalOmega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>). Visualization
730 of the alignments were generated using Esprift (esprift.ibcp.fr). Homology
731 models were generated using Modeller v9.17 (<https://salilab.org/modeller/>).
732 Modelled 3d structures were visualized using PyMol v1.7.4 (The PyMOL
733 Molecular Graphics System, Schrödinger, LLC).

734

735 Core auxin gene set

736 To identify if the auxin regulated genes (*GH3*, *SAUR*, *WIP*, *YUCCA*, and
737 C2HDZ) were conserved in the charophytes and basal land plants, protein
738 sequences from *A. thaliana*, *M. polymorpha* and *P. patens* were used as
739 queries to find the homologous sequences. Obtained BLAST hits were
740 checked in InterProScan to confirm if they belong to the respective gene
741 families. Finally they were checked for differential expression in the above
742 obtained data.

743

744 Mutant generation for *M. polymorpha*

745 To generate the entry clones carrying sgRNA cassette, pairs of oligo DNAs
746 (HK001/HK002 or HK003/HK004 for ARF3, HK162/HK163 for ncARF,
747 HK168/HK169 for ncIAA) were annealed and cloned into pMpGE_En03
748 (Addgene) using *Bsa*I site. The sequence of oligo DNAs are listed in
749 Supplementary file 3. Resultant sgRNA cassette were transferred into
750 pMpGE_010 (Addgene) by LR reaction using Gateway® LR Clonase® II
751 Enzyme Mix (Thermo Fisher Scientific). Transformation into Tak-1 was
752 performed as described previously (Kubota et al., 2013) using Agrobacterium
753 strain GV3101:pMp90. For genotyping, genomic DNA was extracted by
754 simplified CTAB (cetyltrimethylammonium bromide) method
755 (<http://moss.nibb.ac.jp/protocol.html>). Genomic region including target site of
756 sgRNA was amplified with PCR using the primer set HK079/HK131 (ARF3),
757 HK172/HK173 (ncARF) and HK174/HK175 (ncIAA), and sequenced. All
758 primers used in this study are listed in Supplementary file 3.

759

760 Expression analysis of MpARF3 protein

761 MpARF3 promoter fragment including 5' UTR and 3 kb up stream region
762 was amplified with PCR using the primer set HK111/HK026 and cloned into
763 pMpGWB307 (Ishizaki et al., 2015) using *Xba*I site (pJL002). Genomic CDS
764 of MpARF3 without stop codon was amplified with PCR using the primer set
765 HK027/028 and subcloned into pENTR/D-TOPO vector (Thermo Fisher
766 Scientific). Mutation which confers resistant to sgRNA was introduced by

767 PCR using primer set HK137/138. Then mutated CDS fragment was
768 transferred into pJL002 by LR reaction and fused with promoter and
769 Citrine tag (pHKDW103). All primers used in this study are listed in
770 Supplementary file 3. Resultant vector was transformed into *arf3^{ge2-1}*
771 mutant thallus as described previously. Citrine signal and bright field
772 images were captured using a Leica SP5-II confocal laser scanning
773 microscope system, with excitation at 514 nm and detection at 520–600 nm.

774

775 **Data availability**

776 Raw read data of RNA-seq can be accessed in NCBI Short Read Archive (ID:
777 PRJNA397394).

778

779 **Acknowledgments**

780 We thank Jane A. Langdale for distributing plant materials of *A. agrestis*
781 and *C. richardii*, and Jasper Lamers and Lisa Olijslager for contributing to
782 MpARF3 analysis. We are grateful to all contributors of the OneKP project
783 for generating a comprehensive transcriptome database, and Eric Carpenter
784 for providing access. We thank Kuan-Ju Lu and Nicole van 't Wout Hofland
785 for helpful comments on the manuscript. This study was supported by an
786 EMBO Long-Term Postdoctoral Fellowship (ALTF 415-2016) to H.K. and a
787 VICI grant from the Netherlands Organization for Scientific Research
788 (NWO; 865.14.001) to D.W.

789

790 **Competing interests**

791 The authors declare that no competing interests exist.

792 **References**

793

- 794 Abel, S., and Theologis, A. (1996). Early genes and auxin action. *Plant
795 Physiol.* **111**, 9-17.
- 796 Amin, S.A., Hmelo, L.R., van Tol, H.M., Durham, B.P., Carlson, L.T., Heal,
797 K.R., Morales, R.L., Berthiaume, C.T., Parker, M.S., Djunaedi, B., *et al.*
798 (2015). Interaction and signalling between a cosmopolitan
799 phytoplankton and associated bacteria. *Nature* **522**, 98-101.
- 800 Boer, D.R., Freire-Rios, A., van den Berg, W.A., Saaki, T., Manfield, I.W.,
801 Kepinski, S., López-Vidrio, I., Franco-Zorrilla, J.M., de Vries, S.C.,
802 Solano, R., *et al.* (2014). Structural basis for DNA binding specificity by
803 the auxin-dependent ARF transcription factors. *Cell* **156**, 577-589.
- 804 Cove, D.J., Perroud, P.F., Charron, A.J., McDaniel, S.F., Khandelwal, A., and
805 Quatrano, R.S. (2009). Culturing the moss *Physcomitrella patens*. *Cold
806 Spring Harb. Protoc.* **2009**, pdb prot5136.
- 807 Crawford, B.C., Sewell, J., Golembeski, G., Roshan, C., Long, J.A., and
808 Yanofsky, M.F. (2015). Genetic control of distal stem cell fate within root
809 and embryonic meristems. *Science* **347**, 655-659.
- 810 Davidson, N.M., and Oshlack, A. (2014). Corset: enabling differential gene
811 expression analysis for de novo assembled transcriptomes. *Genome Biol.*
812 **15**, 410.
- 813 Dharmasiri, N., Dharmasiri, S., and Estelle, M. (2005). The F-box protein
814 TIR1 is an auxin receptor. *Nature* **435**, 441-445.

- 815 Ding, Z., and Friml, J. (2010). Auxin regulates distal stem cell differentiation
816 in *Arabidopsis* roots. Proc. Natl. Acad. Sci. USA *107*, 12046-12051.
- 817 Eklund, D.M., Ishizaki, K., Flores-Sandoval, E., Kikuchi, S., Takebayashi, Y.,
818 Tsukamoto, S., Hirakawa, Y., Nonomura, M., Kato, H., Kouno, M., *et al.*
819 (2015). Auxin Produced by the Indole-3-Pyruvic Acid Pathway Regulates
820 Development and Gemmae Dormancy in the Liverwort *Marchantia*
821 *polymorpha*. Plant Cell *27*, 1650-1669.
- 822 Finet, C., Berne-Dedieu, A., Scutt, C.P., and Marlétaz, F. (2013). Evolution of
823 the ARF gene family in land plants: old domains, new tricks. Mol. Biol.
824 Evol. *30*, 45-56.
- 825 Flores-Sandoval, E., Eklund, D.M., and Bowman, J.L. (2015). A Simple
826 Auxin Transcriptional Response System Regulates Multiple
827 Morphogenetic Processes in the Liverwort *Marchantia polymorpha*.
828 PLoS Genet. *11*, e1005207.
- 829 Fu, S.F., Wei, J.Y., Chen, H.W., Liu, Y.Y., Lu, H.Y., and Chou, J.Y. (2015).
830 Indole-3-acetic acid: A widespread physiological code in interactions of
831 fungi with other organisms. Plant Signal. Behav. *10*, e1048052.
- 832 Gray, W.G., Kepinski, S., Rouse, D., Leyser, O., and Estelle, M. (2001). Auxin
833 regulates SCFTIR1-dependent degradation of AUX/IAA proteins. Nature
834 *414*, 271-276.
- 835 Hayashi, K., Neve, J., Hirose, M., Kuboki, A., Shimada, Y., Kepinski, S., and
836 Nozaki, H. (2012). Rational design of an auxin antagonist of the
837 SCF(TIR1) auxin receptor complex. ACS Chem. Biol. *7*, 590-598

- 838 Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M.,
839 Sato, S., Yamada, T., Mori, H., Tajima, N., *et al.* (2014). *Klebsormidium*
840 *flaccidum* genome reveals primary factors for plant terrestrial
841 adaptation. *Nat. Commun.* *5*, 3978.
- 842 Ishizaki, K., Nishihama, R., Ueda, M., Inoue, K., Ishida, S., Nishimura, Y.,
843 Shikanai, T., and Kohchi, T. (2015). Development of Gateway Binary
844 Vector Series with Four Different Selection Markers for the Liverwort
845 *Marchantia polymorpha*. *PloS one* *10*, e0138876.
- 846 Ishizaki, K., Nonomura, M., Kato, H., Yamato, K.T., and Kohchi, T. (2012).
847 Visualization of auxin-mediated transcriptional activation using a
848 common auxin-responsive reporter system in the liverwort *Marchantia*
849 *polymorpha*. *J. Plant Res.* *125*, 643-651.
- 850 Jain, M., and Khurana, J.P. (2009). Transcript profiling reveals diverse roles
851 of auxin-responsive genes during reproductive development and abiotic
852 stress in rice. *FEBS J.* *276*, 3148-3162.
- 853 Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L.,
854 Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., *et al.* (2011).
855 Ancestral polyploidy in seed plants and angiosperms. *Nature* *473*,
856 97-100.
- 857 Kato, H., Kouno, M., Takeda, M., Suzuki, H., Ishizaki, K., Nishihama, R.,
858 and Kohchi, T. (2017a). The Roles of the Sole Activator-Type Auxin
859 Response Factor in Pattern Formation of *Marchantia polymorpha*. *Plant*
860 *Cell Physiol.* *58*, 1642-1651.

- 861 Kato, H., Ishizaki, K., Kouno, M., Shirakawa, M., Bowman, J.L., Nishihama,
862 R., and Kohchi, T. (2015). Auxin-Mediated Transcriptional System with
863 a Minimal Set of Components Is Critical for Morphogenesis through the
864 Life Cycle in *Marchantia polymorpha*. PLoS Genet. 11, e1005084.
- 865 Kato, H., Nishihama, R., Weijers, D., and Kohchi, T. (2017b). Evolution of
866 nuclear auxin signaling: lessons from genetic studies with basal
867 land plants. J. Exp. Bot. doi: 10.1093/jxb/erx267.
- 868 Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment
869 software version 7: improvements in performance and usability. Mol.
870 Biol. Evol. 30, 772-780.
- 871 Katsir, L., Schilmiller, A.L., Staswick, P.E., He, S.Y., and Howe, G.A. (2008).
872 COI1 is a critical component of a receptor for jasmonate and the
873 bacterial virulence factor coronatine. Proc. Natl. Acad. Sci. US A 105,
874 7100-7105.
- 875 Kepinski, S., and Leyser, O. (2005). The Arabidopsis F-box protein TIR1 is an
876 auxin receptor. Nature 435, 446-451.
- 877 Kim, J., Harter, K., and Theologis, A. (1997). Protein-protein interactions
878 among the Aux/IAA proteins. Proc. Natl. Acad. Sci. USA 94,
879 11786-11791.
- 880 Korasick, D.A., Westfall, C.S., Lee, S.G., Nanao, M.H., Dumas, R., Hagen, G.,
881 Guilfoyle, T.J., Jez, J.M., and Strader, L.C. (2014). Molecular basis for
882 AUXIN RESPONSE FACTOR protein interaction and the control of
883 auxin response repression. Proc. Natl. Acad. Sci. USA 111, 5427-5432.

- 884 Kubota, A., Ishizaki, K., Hosaka, M., and Kohchi, T. (2013). Efficient
885 *Agrobacterium*-mediated transformation of the liverwort *Marchantia*
886 *polymorpha* using regenerating thalli. Biosci. Biotechnol. Biochem. 77,
887 167-172.
- 888 Kumar, S., Stecher, G., Suleski, M., and Hedges, S.B. (2017). TimeTree: A
889 Resource for Timelines, Timetrees, and Divergence Times. Mol. Biol.
890 Evol. 34, 1812-1819.
- 891 Lanfear, R., Calcott, B., Ho, S.Y., and Guindon, S. (2012). Partitionfinder:
892 combined selection of partitioning schemes and substitution models for
893 phylogenetic analyses. Mol. Biol. Evol. 29, 1695-1701.
- 894 Lavy, M., Prigge, M.J., Tao, S., Shain, S., Kuo, A., Kirchsteiger, K., and
895 Estelle, M. (2016). Constitutive auxin response in *Physcomitrella*
896 reveals complex interactions between Aux/IAA and ARF proteins. eLife
897 5.
- 898 Lee, D.J., Park, J.W., Lee, H.W., and Kim, J. (2009). Genome-wide analysis of
899 the auxin-responsive transcriptome downstream of iaa1 and its
900 expression analysis reveal the diversity and complexity of
901 auxin-regulated gene expression. J. Exp. Bot. 60, 3935-3957.
- 902 Liu, X., Huang, J., Wang, Y., Khanna, K., Xie, Z., Owen, H.A., and Zhao, D.
903 (2010). The role of floral organs in carpels, an *Arabidopsis*
904 loss-of-function mutation in MicroRNA160a, in organogenesis and the
905 mechanism regulating its expression. Plant J. 62, 416-428.
- 906 Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold

- 907 change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**,
908 550.
- 909 Matasci, N., Hung, L.H., Yan, Z., Carpenter, E.J., Wickett, N.J., Mirarab, S.,
910 Nguyen, N., Warnow, T., Ayyampalayam, S., Barker, M., *et al.* (2014).
911 Data access for the 1,000 Plants (1KP) project. *GigaScience* **3**, 17.
- 912 Matías-Hernández, L., Aguilar-Jaramillo, A.E., Marín-González, E.,
913 Suárez-López, P., and Pelaz, S. (2014). *RAV* genes: regulation of floral
914 induction and beyond. *Ann. Bot.* **114**, 1459-1470.
- 915 Mallory, A.C., Bartel, D.P., and Bartel, B. (2005). MicroRNA-directed
916 regulation of *Arabidopsis* AUXIN RESPONSE FACTOR17 is essential
917 for proper development and modulates expression of early auxin
918 response genes. *Plant Cell* **17**, 1360-1375.
- 919 Nanao, M.H., Vinos-Poyo, T., Brunoud, G., Thévenon, E., Mazzoleni, M.,
920 Mast, D., Lainé, S., Wang, S., Hagen, G., Li, H., *et al.* (2014). Structural
921 basis for oligomerization of auxin transcriptional regulators. *Nat.*
922 *Commun.* **5**, 3617.
- 923 Nichols, H.W. (1973). Growth media - freshwater. In *Handbook of*
924 *Phycological Methods*, J.R. Stein, ed. (London: Cambridge University
925 Press), pp. 7-24.
- 926 Ohtaka, K., Hori, K., Kanno, Y., Seo, M., and Ohta, H. (2017). Primitive
927 Auxin Response without TIR1 and Aux/IAA in the Charophyte Alga
928 *Klebsormidium nitens*. *Plant Physiol.* **174**, 1621-1632.
- 929 Paponov, I.A., Paponov, M., Teale, W., Menges, M., Chakrabortee, S., Murray,

- 930 J.A., and Palme, K. (2008). Comprehensive transcriptome analysis of
931 auxin responses in *Arabidopsis*. *Mol. Plant* *1*, 321-337.
- 932 Piya, S., Shrestha, S.K., Binder, B., Stewart, C.N., Jr., and Hewezi, T. (2014).
933 Protein-protein interaction and gene co-expression maps of ARFs and
934 Aux/IAAs in Arabidopsis. *Front. Plant Sci.* *5*, 744.
- 935 Plackett, A.R., Rabinowitsch, E.H., and Langdale, J.A. (2015). Protocol:
936 genetic transformation of the fern *Ceratopteris richardii* through
937 microparticle bombardment. *Plant Methods* *11*, 37.
- 938 Prigge, M.J., Lavy, M., Ashton, N.W., and Estelle, M. (2010). *Physcomitrella*
939 *patens* auxin-resistant mutants affect conserved elements of an
940 auxin-signaling pathway. *Curr. Biol.* *20*, 1907-1912.
- 941 Rensing, S.A. (2017). Why we need more non-seed plant models. *The New*
942 *Phytol.* doi: 10.1111/nph.14464.
- 943 Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H.,
944 Nishiyama, T., Perroud, P.F., Lindquist, E.A., Kamisugi, Y., *et al.* (2008).
945 The *Physcomitrella* genome reveals evolutionary insights into the
946 conquest of land by plants. *Science* *319*, 64-69.
- 947 Saint-Marcoux, D., Proust, H., Dolan, L., and Langdale, J.A. (2015).
948 Identification of reference genes for real-time quantitative PCR
949 experiments in the liverwort *Marchantia polymorpha*. *PloS one* *10*,
950 e0118678.
- 951 Sawa, S., Ohgishi, M., Goda, H., Higuchi, K., Shimada, Y., Yoshida, S., and
952 Koshiba, T. (2002). The *HAT2* gene, a member of the HD-Zip gene family,

- 953 isolated as an auxin inducible gene by DNA microarray screening,
954 affects auxin response in *Arabidopsis*. Plant J. 32, 1011-1022.
- 955 Sheard, L.B., Tan, X., Mao, H., Withers, J., Ben-Nissan, G., Hinds, T.R.,
956 Kobayashi, Y., Hsu, F.F., Sharon, M., Browse, J., *et al.* (2010). Jasmonate
957 perception by inositol-phosphate-potentiated COI1-JAZ co-receptor.
958 Nature 468, 400-405.
- 959 Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and
960 post-analysis of large phylogenies. Bioinformatics 30, 1312-1313.
- 961 Sugano, S.S., Shirakawa, M., Takagi, J., Matsuda, Y., Shimada, T.,
962 Hara-Nishimura, I., and Kohchi, T. (2014). CRISPR/Cas9-mediated
963 targeted mutagenesis in the liverwort *Marchantia polymorpha* L. Plant
964 Cell Physiol. 55, 475-481.
- 965 Szemenyei, H., Hannon, M., and Long, J.A. (2008). TOPLESS mediates
966 auxin-dependent transcriptional repression during *Arabidopsis*
967 embryogenesis. Science 319, 1384-1386.
- 968 Szövényi, P., Frangedakis, E., Ricca, M., Quandt, D., Wicke, S., and Langdale,
969 J.A. (2015). Establishment of *Anthoceros agrestis* as a model species for
970 studying the biology of hornworts. BMC Plant Biol. 15, 98.
- 971 Takato, S., Kakei, Y., Mitsui, M., Ishida, Y., Suzuki, M., Yamazaki, C.,
972 Hayashi, K.I., Ishii, T., Nakamura, A., Soeno, K., *et al.* (2017). Auxin
973 signaling through SCFTIR1/AFBs mediates feedback regulation of IAA
974 biosynthesis. Biosci. Biotechnol. Biochem. 81, 1320-1326.
- 975 Tan, X., Calderon-Villalobos, L.I., Sharon, M., Zheng, C., Robinson, C.V.,

- 976 Estelle, M., and Zheng, N. (2007). Mechanism of auxin perception by the
977 TIR1 ubiquitin ligase. *Nature* *446*, 640-645.
- 978 Ulmasov, T., Hagen, G., and Guilfoyle, T.J. (1999). Activation and repression
979 of transcription by auxin response factors. *Proc. Natl. Acad. Sci. USA* *96*,
980 5844-5849.
- 981 Wang, J.W., Wang, L.J., Mao, Y.B., Cai, W.J., Xue, H.W., and Chen, X.Y.
982 (2005). Control of root cap formation by MicroRNA-targeted auxin
983 response factors in *Arabidopsis*. *Plant Cell* *17*, 2204-2216.
- 984 Weijers, D., and Wagner, D. (2016). Transcriptional Responses to the Auxin
985 Hormone. *Annu. Rev. Plant Biol.* *67*, 539-574.
- 986 Yang, J., Tian, L., Sun, M.X., Huang, X.Y., Zhu, J., Guan, Y.F., Jia, Q.S., and
987 Yang, Z.N. (2013). AUXIN RESPONSE FACTOR17 is essential for pollen
988 wall pattern formation in *Arabidopsis*. *Plant Physiol.* *162*, 720-731.
- 989 Yoon, H.S., Hackett, J.D., Ciniglia, C., Pinto, G., and Bhattacharya, D. (2004).
990 A molecular timeline for the origin of photosynthetic eukaryotes.
991 *Molecular Biol. Evol.* *21*, 809-818.
- 992 Žižková, E., Kubeš, M., Dobrev, P.I., Přibyl, P., Šimura, J., Zahajská, L.,
993 Záveská Drábková, L., Novák, O., and Motyka, V. (2017). Control of
994 cytokinin and auxin homeostasis in cyanobacteria and algae. *Ann. Bot.*
995 *119*, 151-166.

996 **Figure legends**

997

998 **Figure 1.** Proteins in nuclear auxin pathway; mechanism and origin of the
999 domains. **(A, B)** Scheme of NAP in land plants. In the absence of auxin,
1000 Aux/IAA inhibit ARF via their PB1 domains, and by recruiting the TPL
1001 co-repressor. Auxin stabilizes the interaction between Aux/IAA and
1002 TIR1/AFB, followed by proteasome-mediated degradation of Aux/IAA. **(C)**
1003 Domain structure of NAP components in land plants and presence of each
1004 domain in algae, as recovered in transcriptomes. (Sub-)domains are
1005 indicated by colors, that match those in **(B)**. *: difficult to assign to ARF or
1006 Aux/IAA family; #: forming basal clade to both TIR1/AFB and COI1 in land
1007 plants.

1008 **Figure supplement 1.** The work flow of phylogenetic tree construction.

1009 **Figure supplement 2.** Phylogenetic tree of ARF and RAV proteins

1010

1011 **Figure 2.** Distribution of auxin signaling proteins precursors in algal
1012 lineages. **(A)** Occurrence of NAP components in red algae, chlorophytes, and
1013 charophytes. Empty circles and filled circles indicate the absence and
1014 presence of that particular component, respectively. #: PB1 domain is present
1015 in proto-RAV but not in land plant RAV. *: charophytes have TIR1/AFB/COI1
1016 precursor but land plants have separate genes encoding TIR1 and COI1
1017 proteins. **(B)** Overview of the phylogenetic arrangement of RAV1, Aux/IAA
1018 and ARFs based on the DBD tree and PB1 tree.

1019 **Figure supplement 1.** Phylogenetic tree based on PB1 domain

1020

1021

1022 **Figure 3.** Homology models of ancestral ARF, Aux/IAA and TIR1/AFB
1023 proteins. **(A)** Homology models for ARF DBDs. The crystal structure of
1024 *Arabidopsis thaliana* ARF1-DBD is shown on the left with important
1025 residues for DNA binding (top) and dimerization (bottom). Homology models
1026 for (proto-)ARFs are overlaid on *A. thaliana* ARF1 in right panels (brown).
1027 **(B)** Alignment of PB1 domain of (proto-)ARF, Aux/IAA and proto-RAV
1028 proteins. Numbering is based on the ARF5 protein of *A. thaliana*. Arrows
1029 and helices indicate β-sheets and α-helices in ARF5 and IAA17 of *A. thaliana*,
1030 respectively. Blue and red triangles indicate positive (+) and negative (-)
1031 faces, respectively. Golden asterisks represent the residues of polar
1032 interactions. **(C)** Homology models for TIR1/AFB and COI1 proteins. Left
1033 panel shows crystal structure of *A. thaliana* TIR1 from top view. Auxin
1034 binding pocket of TIR1/AFB and jasmonate binding pocket of COI1 are
1035 shown in right panels. The *Spirogyra pratensis* protein lacks secondary
1036 structures for hormone binding. Ci: *Coloechaete irregularis*, Da: *Desmidium*
1037 *aptogonum*, Kn: *Klebsormidium nitens*, Mc: *Mesotaenium caldarium*, Mp:
1038 *Marchantia polymorpha*, Pk: *Parachlorella kessleri*, Ps: *Pisum sativum*, Sp:
1039 *Spirogyra pratensis*.

1040

1041 **Figure 4.** Reconstruction of ancestral state of NAP components in plant

1042 evolution. Phylogeny of taxonomic classes are shown in left. Time point of the
1043 lineage diversification was calculated using TimeTree database (Kumar et al.,
1044 2017). Black stars indicate whole genome duplication events (Jiao et al.,
1045 2011). Right: phylogenetic trees show the copy number and phylogenetic
1046 relationship of each protein family in the common ancestors. Each circle is
1047 colored according to protein type as indicated in the box. In the top row,
1048 numbers indicate which genes of *Arabidopsis thaliana* belong to each
1049 subfamily and red circles indicates missing subfamilies in *A. thaliana*.

1050 **Figure supplement 1.** Phylogenetic tree of Aux/IAA

1051 **Figure supplement 2.** Phylogenetic tree of TIR1/AFB

1052 See also Figure 1—Figure supplement 2

1053

1054 **Figure 5.** Auxin-dependent gene regulation across basal plant species. (A)
1055 Histograms represent the distribution of \log_2 fold change among
1056 differentially expressed genes ($P_{adj} < 0.01$). Pie charts indicate the total
1057 number of up- and down-regulated genes in each species. Circles in the top
1058 left of each graph indicate the number of NAP components. (B) Violin plots of
1059 \log_2 normalized expression values (y-axis) of 20 least auxin activated (Low)
1060 and 20 top-most auxin upregulated (High) genes in each six species. White
1061 dot indicated the median expression value.

1062 **Figure supplement 1.** Number of DEG in de novo assembly- or genome-based
1063 transcriptome analysis.

1064 **Figure supplement 2.** . Summary statistics of comparative RNA-seq analysis

1065

1066 **Figure 6.** Identification of deeply conserved auxin-responsive genes. **(A)**
1067 Auxin-dependence of six well-known angiosperm auxin-responsive gene
1068 families (top) surveyed from *de novo* assembly-based transcriptomes in 6
1069 species. Each circle indicates a gene copy of each gene family. Red, blue and
1070 grey circle indicate up-, down- and non-regulated genes in response to auxin.
1071 **(B)** qPCR analysis of conserved auxin-responsive genes. Relative expression
1072 values are normalized by the expression of *EF1 α* in *Marchantia polymorpha*
1073 or the amount of total RNA in *Anthoceros agrestis* and *Ceratopteris richardii*.
1074 Each bar indicates average of expression with SD (biological replicates ≥ 3).
1075 *: p < 0.01 (t-test).

1076

1077 **Figure 7.** Genetic analysis of ancient components in *Marchantia polymorpha*.
1078 **(A)** Diagrams of gene structure and CRISPR/Cas9-mediated mutation in
1079 *ncIAA*, *ncARF* and ARF3 loci. Arrowheads indicate sgRNAs target sites.
1080 Gray and black boxes indicate UTR and CDS, respectively. Red and blue bars
1081 indicate the region coding PB1 and DBD. **(B)** 10-day-old gemmalings grown
1082 without or with 3 μ M 2,4-D. Scale bars: 5 mm. **(C)** Expression analysis of
1083 auxin-responsive genes in WT, *nciaa*, *ncarf*, and *arf3* mutants by qPCR.
1084 10-day-old gemmalings (*nciaa* and *ncarf*) or regenerating thalli (*arf1* and
1085 *arf3*) were treated with 10 μ M 2,4-D for 1 h. Each bar indicates average \pm SD
1086 (biological replicates = 3). Asterisks indicate significant differences. *: p <
1087 0.01 (Tukey test), **: p < 0.05 (t-test). **(D, E)** Thallus tips grown for 2 weeks

1088 (D) and gemma cups (E) of WT and *arf3^{ge1}-1* mutant. *arf3^{ge1}-1* showed growth
1089 retardation and no mature gemmae, similar to the other alleles. (F)
1090 Expression analysis of *proARF3:ARF3-Citrine* in *arf3^{ge2}-1* background. Left
1091 and right panel show developing and mature gemmae, respectively.
1092 Scale bars: 5 mm in (B and D), 0.5 mm in (E), 50 μ m in (F).

1093 **Figure supplement 1.** CRISPR/Cas9-mediated mutagenesis in *M.*
1094 *polymorpha.*

1095 **Figure supplement legends**

1096

1097 **Figure 1—Figure supplement 1.** The work flow of phylogenetic tree
1098 construction.

1099

1100 **Figure 1—Figure supplement 2.** Phylogenetic tree of ARF and RAV proteins.

1101 Label color shows the taxonomic group of each protein as indicated in the box
1102 above. Numbers along with the branches indicate branch length. Orange
1103 circles indicate the bootstraps less than 50. Colored boxes connected with
1104 gray bar shows the domain structure of each protein. Red: B3, green:
1105 blue: PB1, gray: AP2. The complete tree can be found at
1106 <http://itol.embl.de/shared/dolfweijers> (interactive Tree of Life; iTOL)

1107

1108 **Figure 2—Figure supplement 1.** Phylogenetic tree based on PB1 domain.
1109 Colored branches indicate protein families. Orange: Chlorophytes, green:
1110 proto-RAV, blue: Aux/IAA, black: (proto-)ARF. Numbers along with the
1111 branches indicate bootstrap values. The complete tree can be found at
1112 <http://itol.embl.de/shared/dolfweijers> (interactive Tree of Life; iTOL)

1113

1114 **Figure 4—Figure supplement 1.** Phylogenetic tree of Aux/IAA. Label color
1115 shows the taxonomic group of each protein as indicated in top. Colored boxes
1116 connected with gray bar shows the domain structure of each protein.
1117 Magenta: domain I, yellow domain II, blue: PB1. Numbers along with the

1118 branches indicate branch length. Orange circles indicate bootstrap values
1119 less than 50. The complete tree can be found at
1120 <http://itol.embl.de/shared/dolfweijers> (interactive Tree of Life; iTOL)

1121

1122 **Figure 4—Figure supplement 2.** Phylogenetic tree of TIR1/AFB. Label color
1123 shows the taxonomic group of each protein as indicated in left. Numbers
1124 along with the branches indicate branch length. Orange circles indicate
1125 bootstrap values less than 50.

1126

1127 **Figure 5—Figure supplement 1.** Number of DEG in de novo assembly- or
1128 genome-based transcriptome analysis.

1129

1130 **Figure 5—Figure supplement 2.** . Summary statistics of comparative
1131 RNA-seq analysis

1132

1133 **Figure 7—Figure supplement 1.** CRISPR/Cas9-mediated mutagenesis in *M.*
1134 *polymorpha*. (A-D) Mutations detected by sequencing analysis. The amino
1135 acid (AA) sequences encoded in WT are shown at the top. WT sequence is
1136 shown with the PAM sequence highlighted in red and the target sequence of
1137 sgRNA in blue. Purple bases indicate mutation. *nciaa-6*: 6 bp deletion and 20
1138 bp insertion, *nciaa-10*: 776 bp deletion and 75 bp insertion, *ncarf-2*: 1 bp
1139 insertion, *ncarf-10*: 486 bp deletion and 6bp insertion, *arf3ge1-1*: 5 bp deletion,
1140 *arf3ge1-2*: 11 bp deletion and 72 bp insertion, *arf3ge2-1*: 9 bp deletion. (E)

1141 3-week-old gemmalings. Arrows indicate the thalli formed with
1142 up-side-down.

1143 (F) qPCR analysis on 10-day-old gemmalings with or without 10 µM 2,4-D
1144 treatment for 1 h. Relative expression values are normalized by the
1145 expression of *EF1α*. Each bar indicates average with SD (biological replicate
1146 = 3). Each asterisk indicates significant difference between WT and mutants
1147 in the same condition ($p < 0.01$, Tukey test). (G) Thallus tips of WT and *arf3*
1148 mutants grown for 2 weeks. Scale bars = 5 mm.

1149

1150 **Supplementary files**

1151

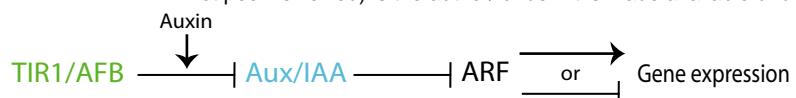
1152 **Supplementary file 1.** Species used in phylogenomic analysis.

Supplementary file 2. Multiple sequence alignments used in the study.

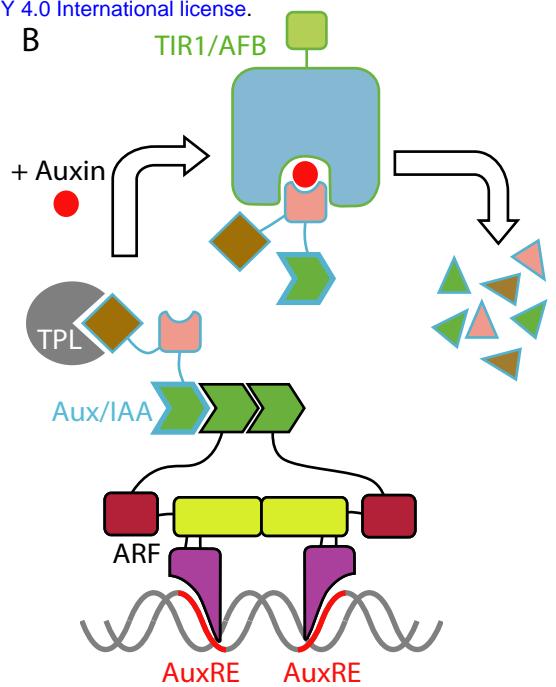
Supplementary file 3. Primers used in this study.

1153

A



B



C

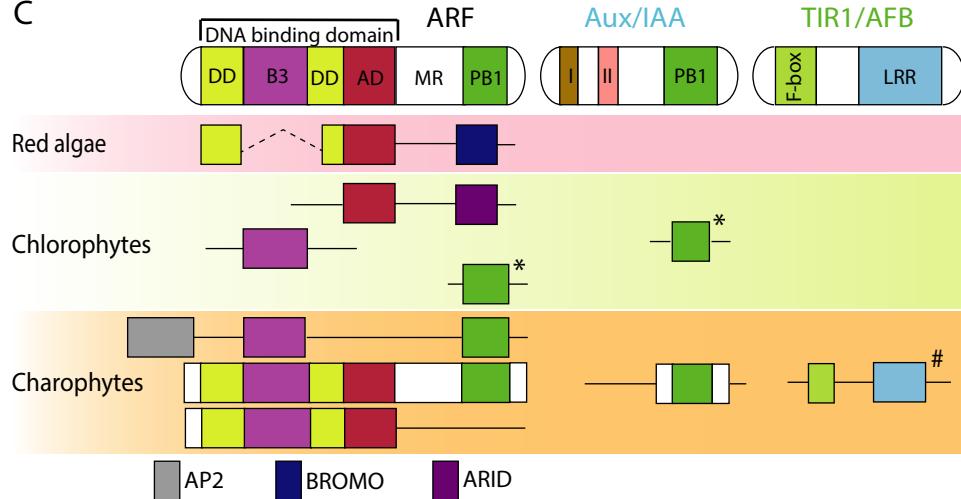
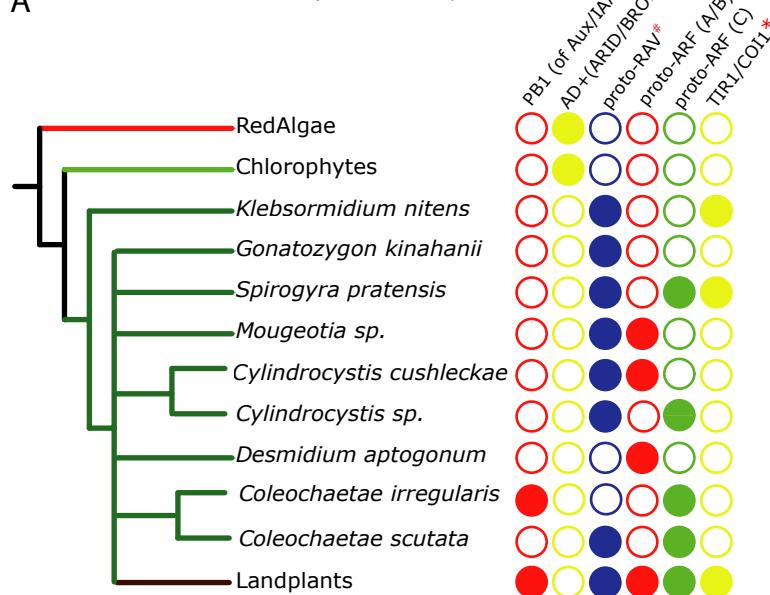


Figure 1. Proteins in nuclear auxin pathway; mechanism and origin of the domains. (A, B) Scheme of NAP in land plants. In the absence of auxin, Aux/IAA inhibit ARF via their PB1 domains, and by recruiting the TPL co-repressor. Auxin stabilizes the interaction between Aux/IAA and TIR1/AFB, followed by proteasome-mediated degradation of Aux/IAA. (C) Domain structure of NAP components in land plants and presence of each domain in algae. (Sub-)domains are indicated by colors, that match those in (B). *: difficult to assign to ARF or Aux/IAA family; #: forming basal clade to both TIR1/AFB and COI1 in land plants.

A



B

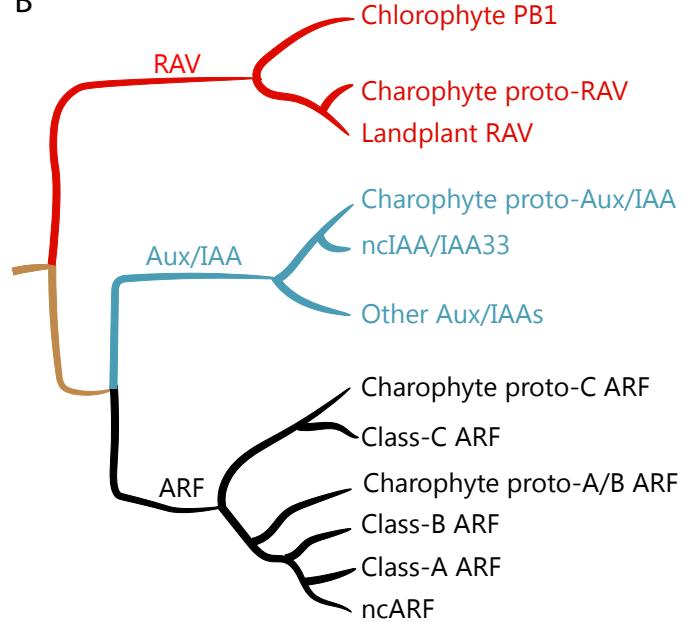


Figure 2. Distribution of auxin signaling proteins precursors in algal lineages. (A) Occurrence of NAP components in red algae, chlorophytes, and charophytes. Empty circles and filled circles indicate the absence and presence of that particular component, respectively. #: PB1 domain is present in proto-RAV but not in land plant RAV. *: charophytes have TIR1/AFB/COI1 precursor but land plants have separate genes encoding TIR1 and COI1 proteins. (B) Overview of the phylogenetic arrangement of RAV1, Aux/IAA and ARFs based on the DBD tree and PB1 tree.

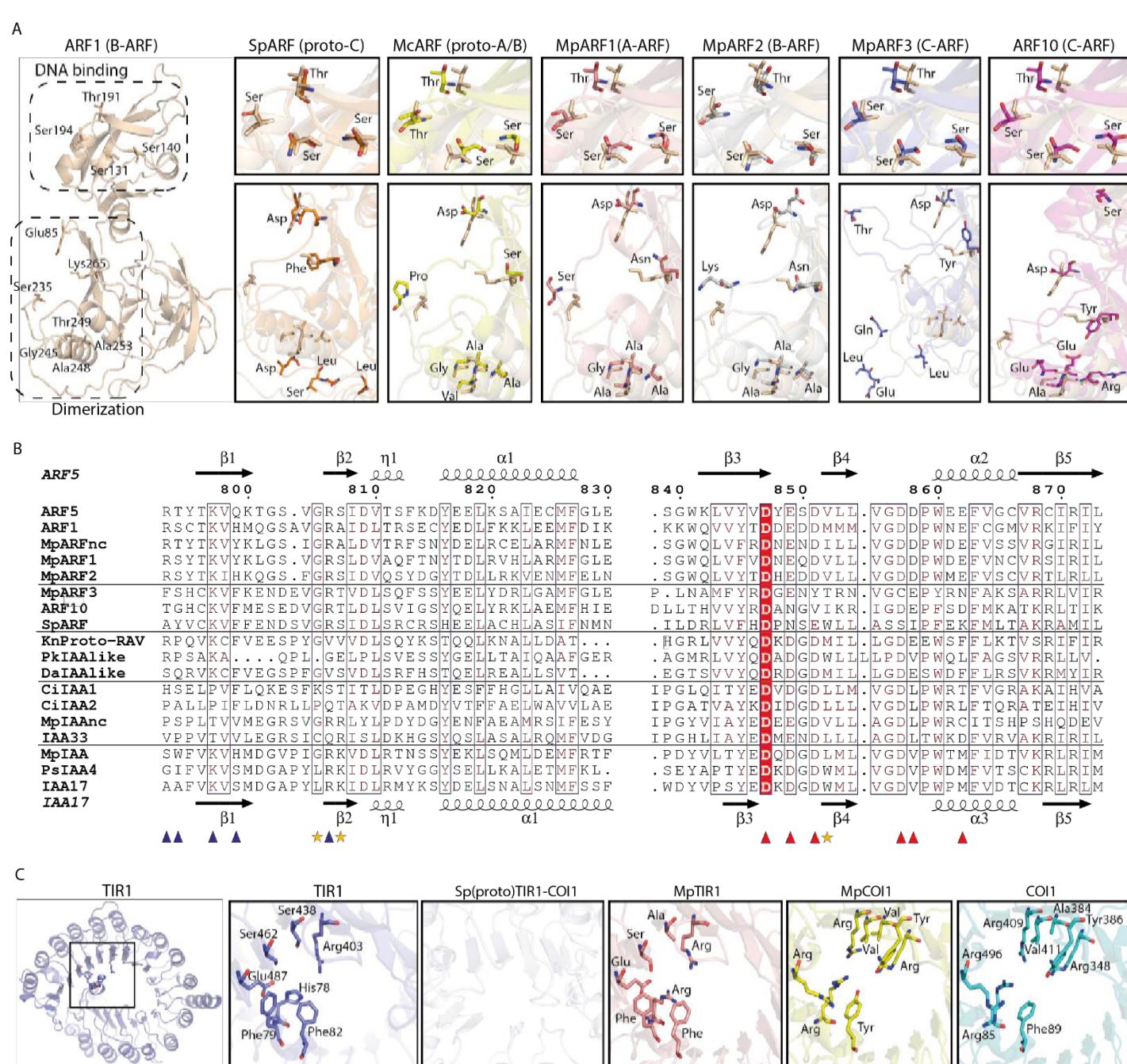


Figure 3. Homology models of ancestral ARF, Aux/IAA and TIR1/AFB proteins. **(A)** Homology models for ARF DBDs. The crystal structure of *Arabidopsis thaliana* ARF1-DBD is shown on the left with important residues for DNA binding (top) and dimerization (bottom). Homology models for (proto-)ARFs are overlaid on *A. thaliana* ARF1 in right panels (brown). **(B)** Alignment of PB1 domain of (proto-)ARF, Aux/IAA and proto-RAV proteins. Numbering is based on the ARF5 protein of *A. thaliana*. Arrows and helices indicate β-sheets and α-helices in ARF5 and IAA17 of *A. thaliana*, respectively. Blue and red triangles indicate positive (+) and negative (-) faces, respectively. Golden asterisks represent the residues of polar interactions. **(C)** Homology models for TIR1/AFB and COI1 proteins. Left panel shows crystal structure of *A. thaliana* TIR1 from top view. Auxin binding pocket of TIR1/AFB and jasmonate binding pocket of COI1 are shown in right panels. The *Spirogyra pratensis* protein lacks secondary structures for hormone binding. Ci: *Coloechaete irregularis*, Da: *Desmidium aptogonium*, Kn: *Klebsormidium nitens*, Mc: *Mesotaenium caldariorum*, Mp: *Marchantia polymorpha*, Pk: *Parachlorella kessleri*, Ps: *Pisum sativum*, Sp: *Spirogyra pratensis*.

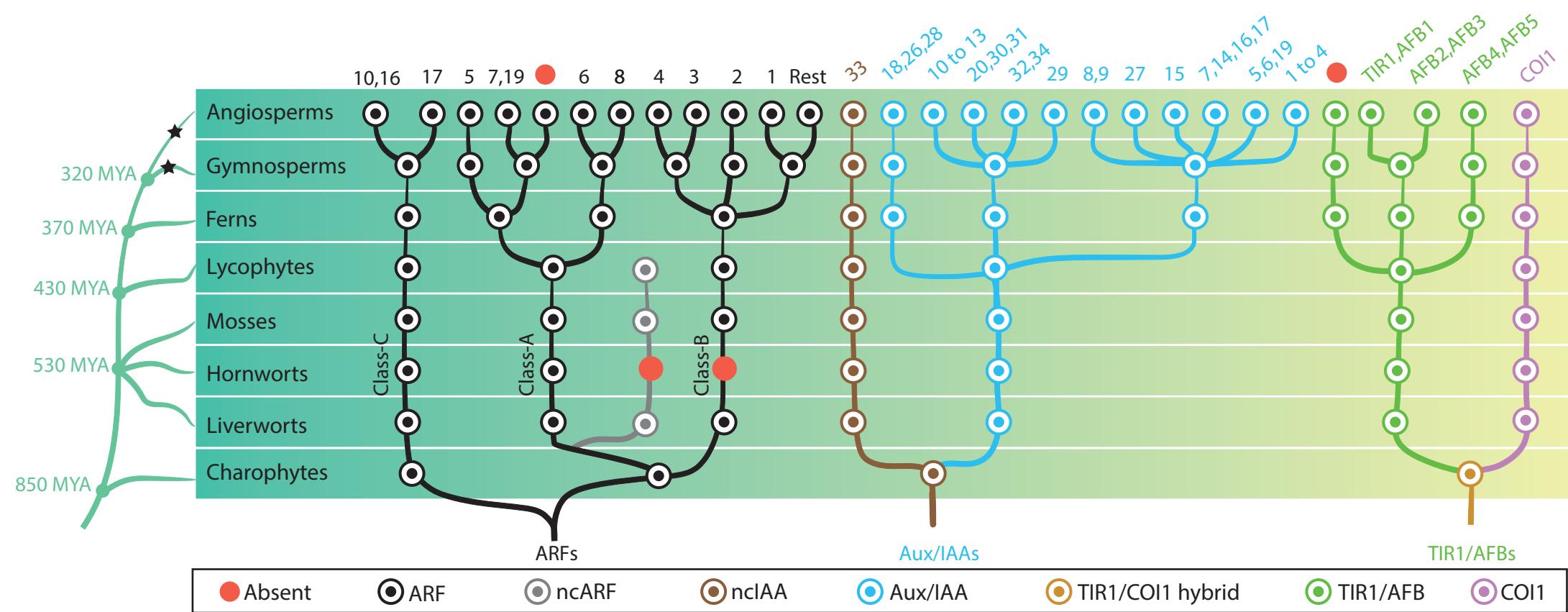


Figure 4. Reconstruction of ancestral state of NAP components in plant evolution. Phylogeny of taxonomic classes are shown in left. Time point of the lineage diversification was calculated using TimeTree database (Kumar et al., 2017). Black stars indicate whole genome duplication events (Jiao et al., 2011). Right: phylogenetic trees show the copy number and phylogenetic relationship of each protein family in the common ancestors. Each circle is colored according to protein type as indicated in the box. Top numbers indicate which genes of *Arabidopsis thaliana* belong to each subfamily.

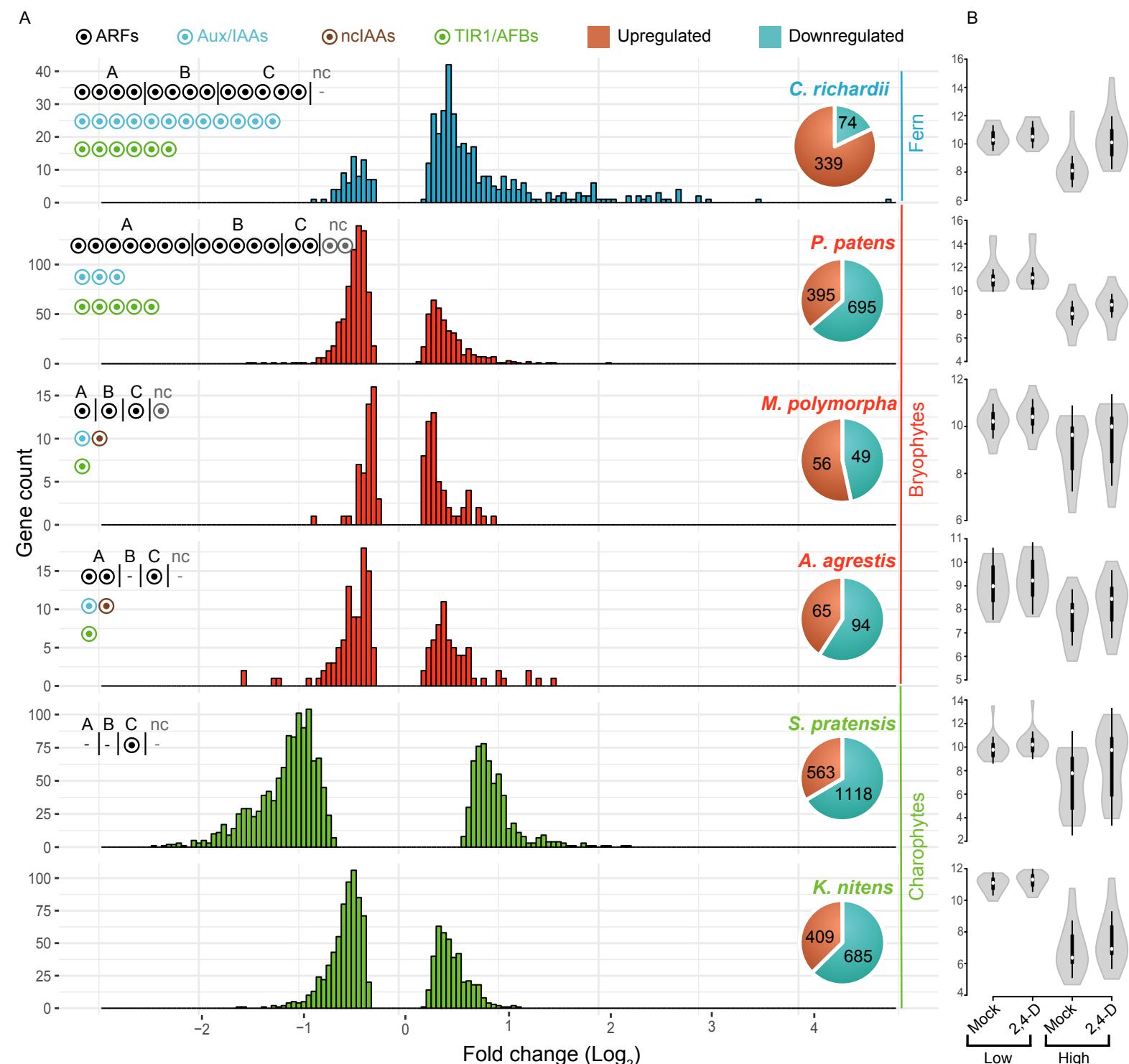


Figure 5. Auxin-dependent gene regulation across basal plant species. (A) Histograms represent the distribution of \log_2 fold change among differentially expressed genes ($\text{Padj} < 0.01$). Pie charts indicate the total number of up- and down-regulated genes in each species. Circles in the top left of each graph indicate the number of NAP components. (B) Violin plots of \log_2 normalized expression values (y-axis) of 20 least auxin activated (Low) and 20 top-most auxin activated (High) genes in each six species. White dot indicated the median expression value.

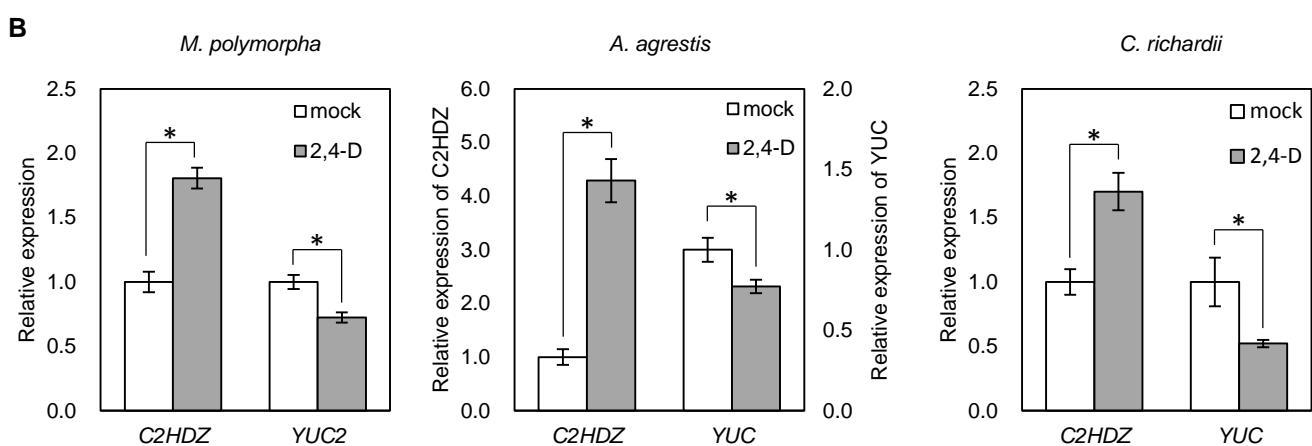
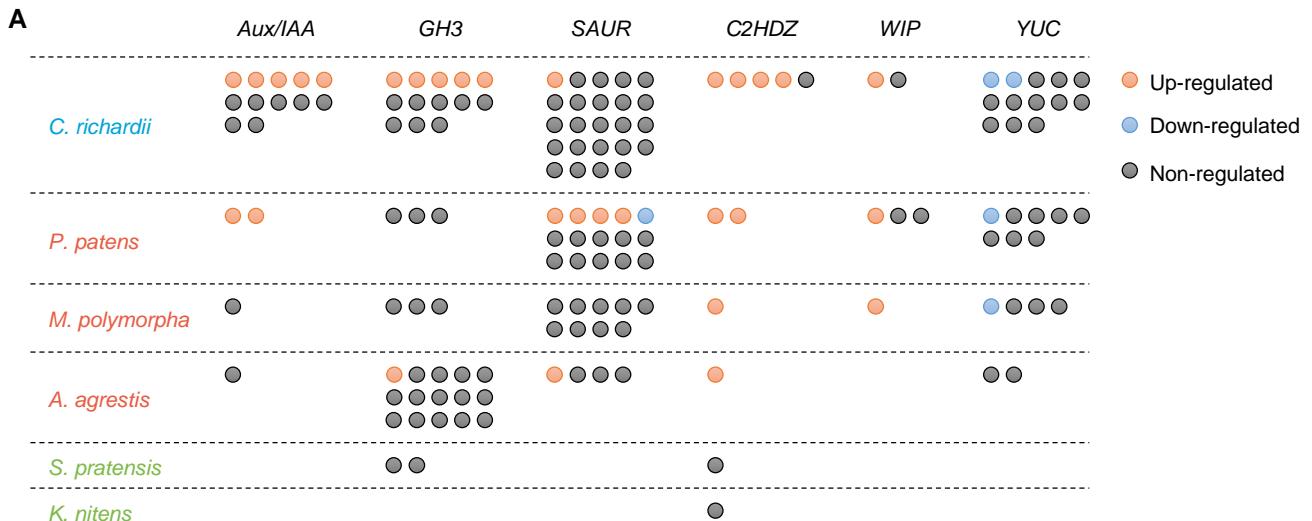


Figure 6. Identification of deeply conserved auxin-responsive genes. **(A)** Auxin-dependence of six well-known angiosperm auxin-responsive gene families (top) surveyed from *de novo* assembly-based transcriptomes in 6 species. Each circle indicates a gene copy of each gene family. Red, blue and grey circle indicate up-, down- and non-regulated genes in response to auxin. **(B)** qPCR analysis of conserved auxin-responsive genes. Relative expression values are normalized by the expression of *EF1 α* in *Marchantia polymorpha* or the amount of total RNA in *Anthoceros agrestis* and *Ceratopteris richardii*. Each bar indicates average of expression with SD (biological replicates ≥ 3). *: $p < 0.01$ (t-test).

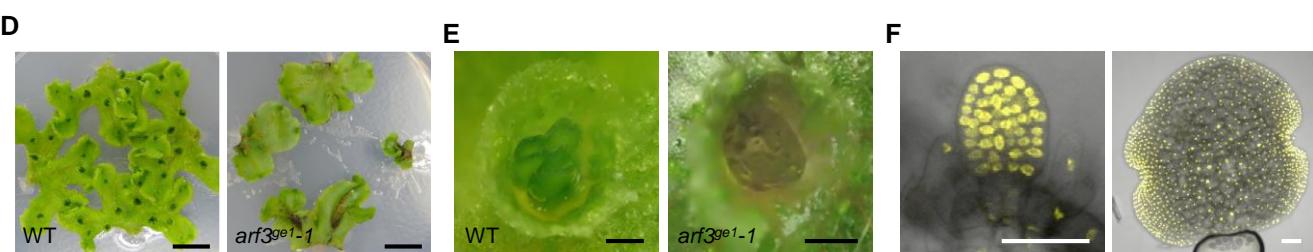
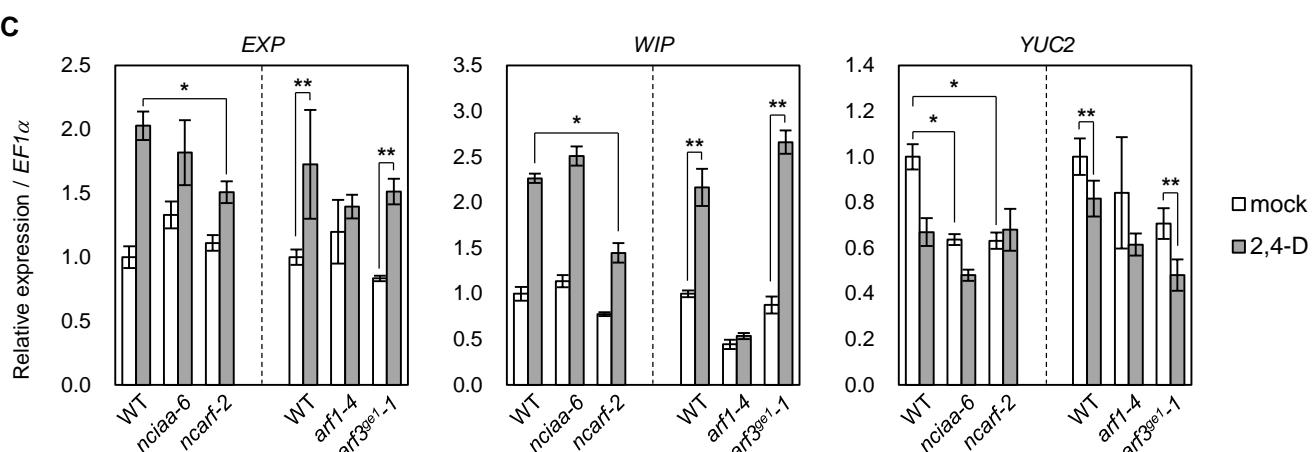
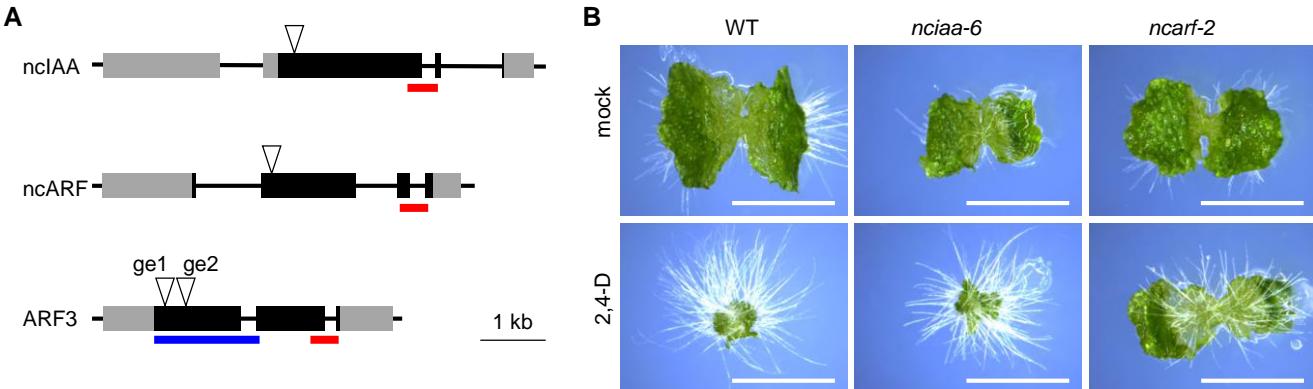


Figure 7. Genetic analysis of ancient components in *Marchantia polymorpha*. **(A)** Diagrams of gene structure and CRISPR/Cas9-mediated mutation in *nclAA*, *ncARF* and *ARF3* loci. Arrowheads indicate sgRNAs target sites. Gray and black boxes indicate UTR and CDS, respectively. Red and blue bars indicate the region coding PB1 and DBD. **(B)** 10-day-old gemmalings grown without or with 3 μ M 2,4-D. Scale bars: 5 mm. **(C)** Expression analysis of auxin-responsive genes in WT, *nciaa*, *ncarf*, and *arf3* mutants by qPCR. 10-day-old gemmalings (*nciaa* and *ncarf*) or regenerating thalli (*arf1* and *arf3*) were treated with 10 μ M 2,4-D for 1 h. Each bar indicates average \pm SD (biological replicates = 3). Asterisks indicate significant differences. *: $p < 0.01$ (Tukey test), **: $p < 0.05$ (t-test). **(D, E)** Thallus tips grown for 2 weeks (**D**) and gemma cups (**E**) of WT and *arf3ge1-1* mutant. *arf3ge1-1* showed growth retardation and no mature gemmae, similar to the other alleles. **(F)** Expression analysis of *proARF3:ARF3-Citrine* in *arf3ge2-1* background. Left and right panel show developing and mature gemmae, respectively.

Scale bars: 5 mm in **(B and D)**, 0.5 mm in **(E)**, 50 μ m in **(F)**.

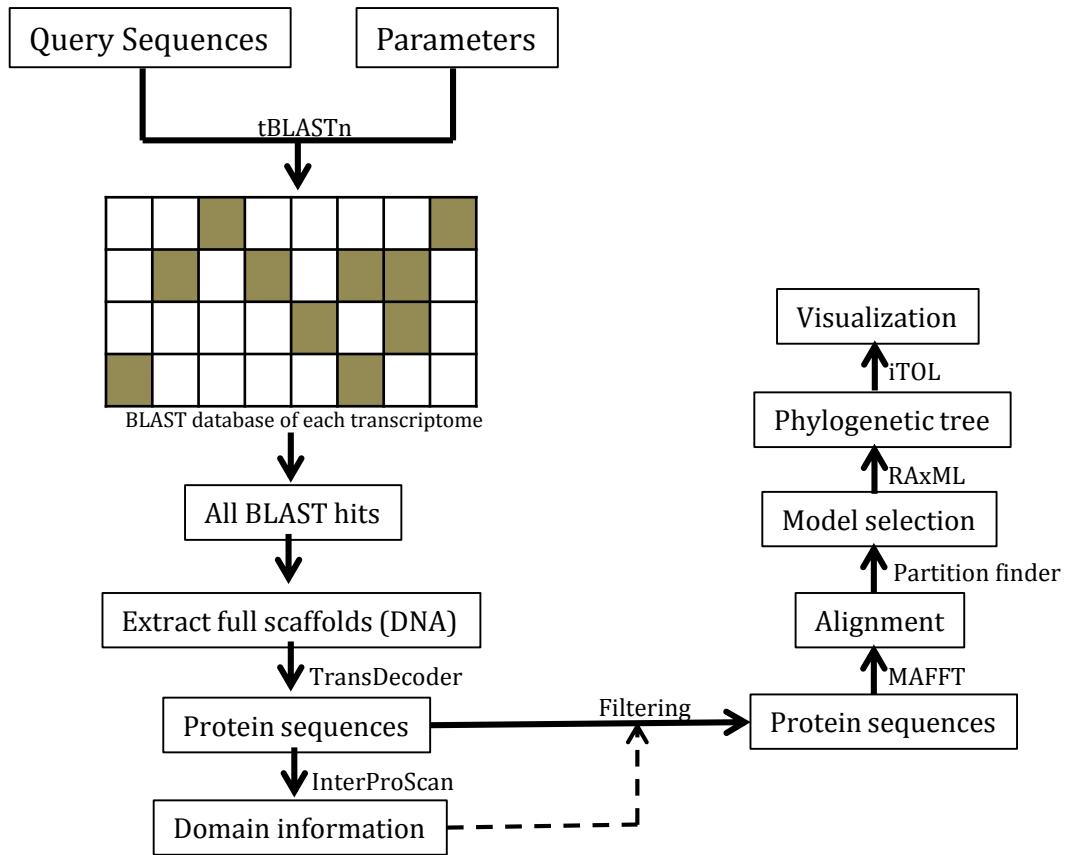
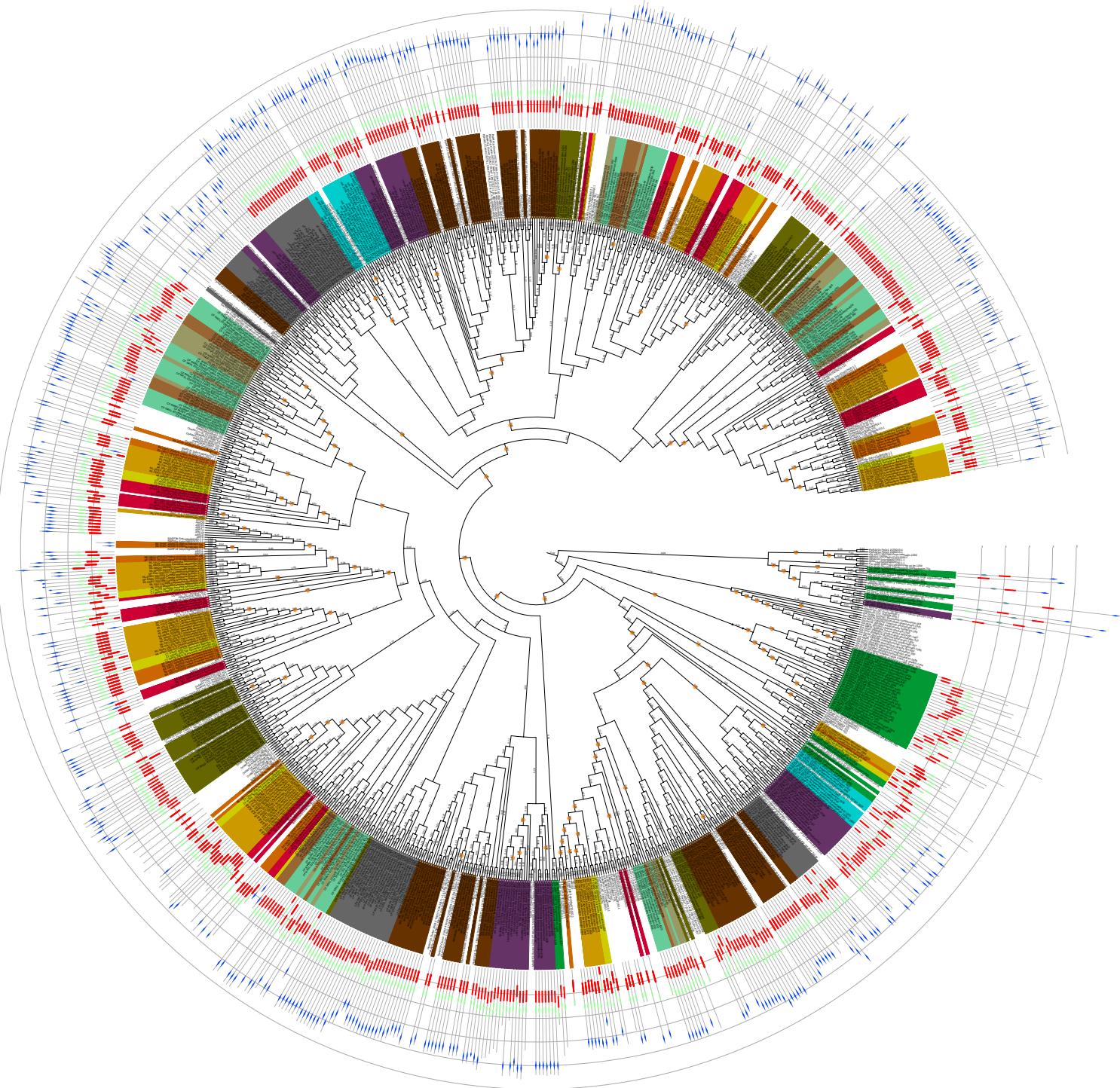


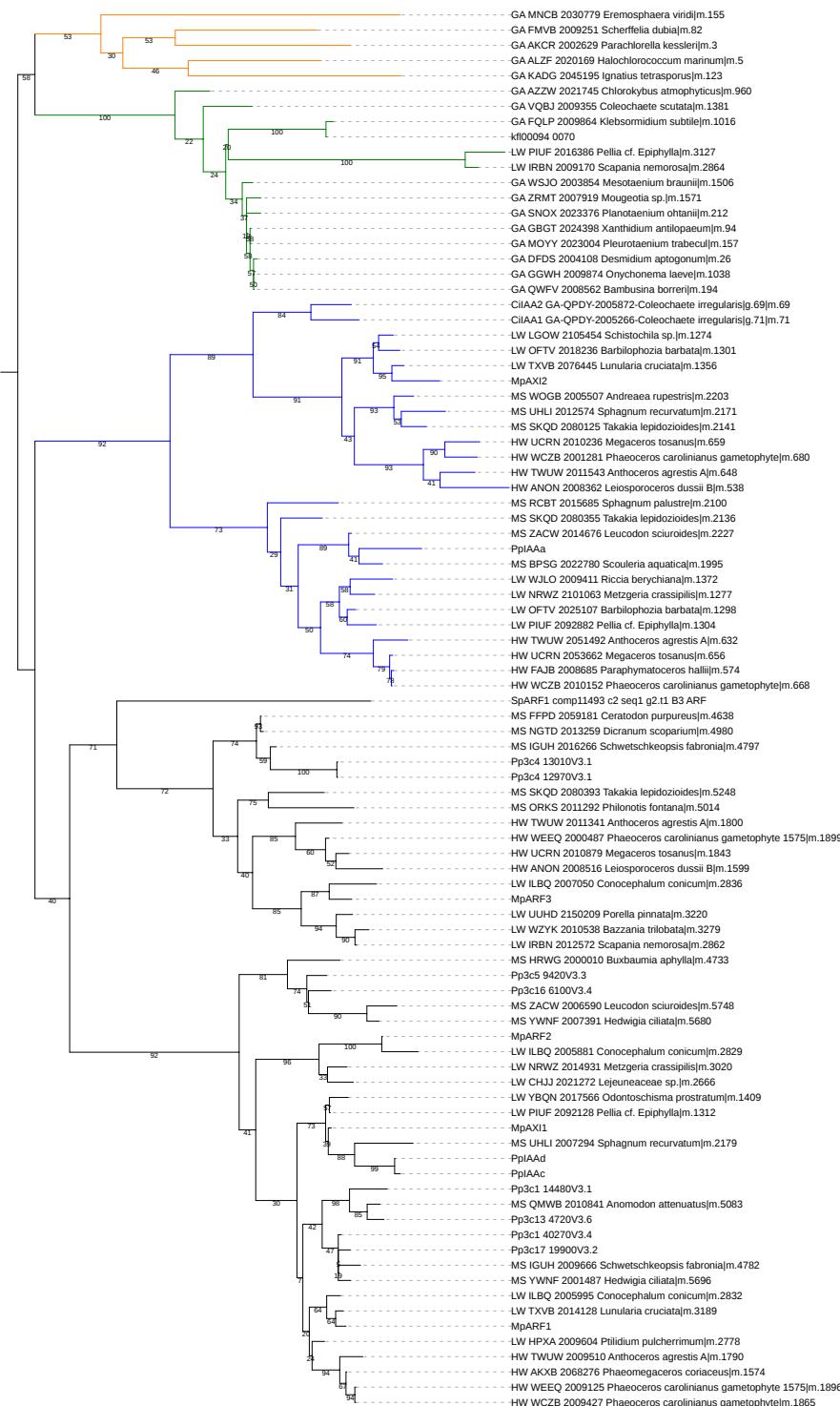
Figure1-Figure supplement 1. The work flow of phylogenetic tree construction.

Colored clades

- Cycadales
- Conifers
- Ferns
- GreenAlgae
- Gnetales
- Hornworts
- Lycophtyes
- Liverworts
- Mosses
- BasalAngiosperms
- BasalEudicots
- Chloranthales
- Magnoliids



Tree scale: 0.1



Colored ranges

- Cycadales
- Conifers
- Ferns
- Gnetales
- Hornworts
- Lycophytes
- Liverworts
- Mosses
- BasalAngiosperms
- BasalEudicots
- Magnoliids
- Chloranthales

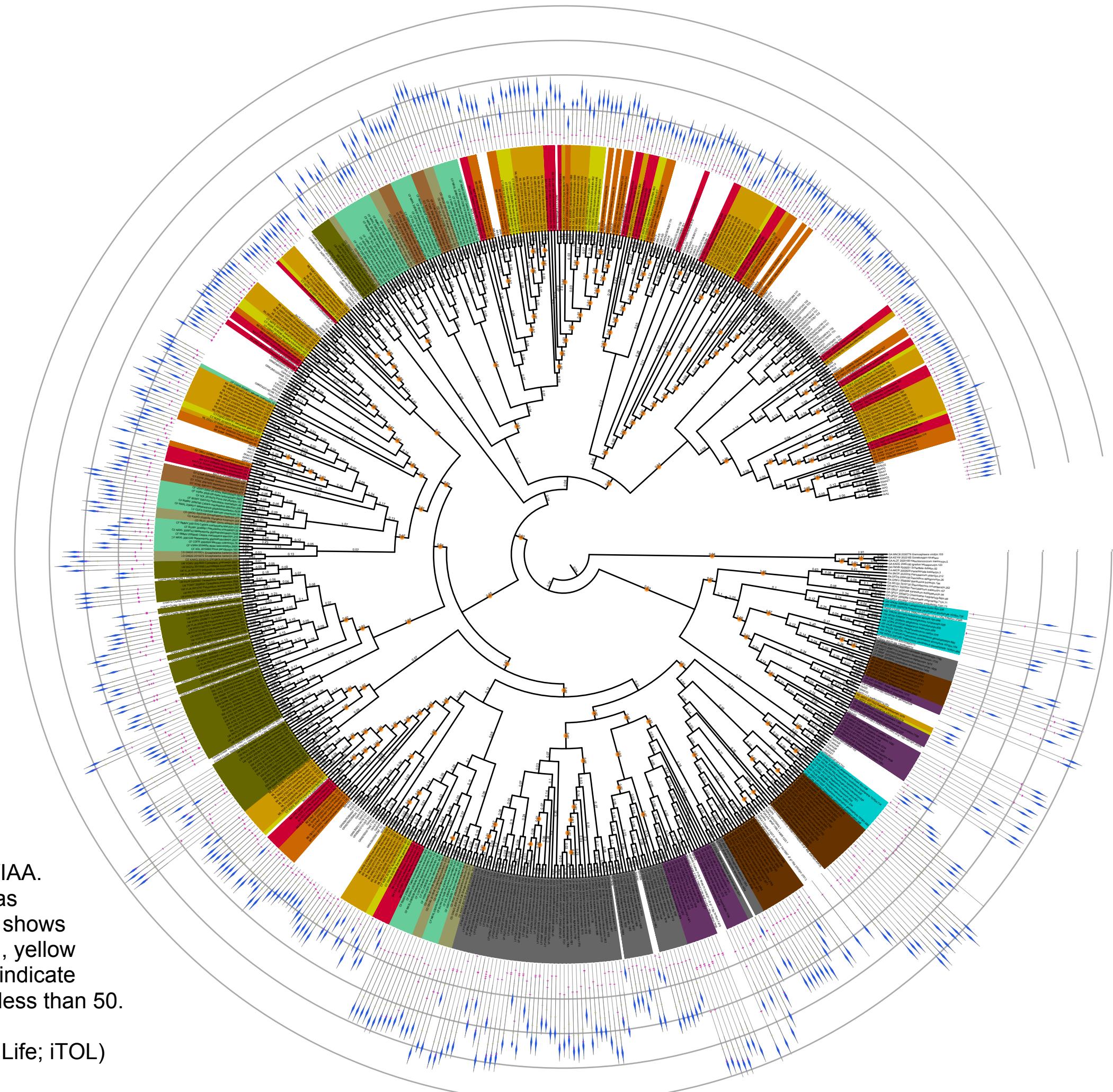


Figure 4-Figure supplement 1. Phylogenetic tree of Aux/IAA. Label color shows the taxonomic group of each protein as indicated in top. Colored boxes connected with gray bar shows the domain structure of each protein. Magenta: domain I, yellow domain II, blue: PB1. Numbers along with the branches indicate branch length. Orange circles indicate bootstrap values less than 50. The complete tree can be found at <http://itol.embl.de/shared/dolfweijers> (interactive Tree of Life; iTOL)

Colored clades

- Cycadales
- Conifers
- Ferns
- GreenAlgae
- Ginkgo
- Gnetales
- Hornworts
- Lycophtyes
- Liverworts
- Moss
- BasalAngiosperms
- BasalEudicots
- Chloranthales
- Magnoliids

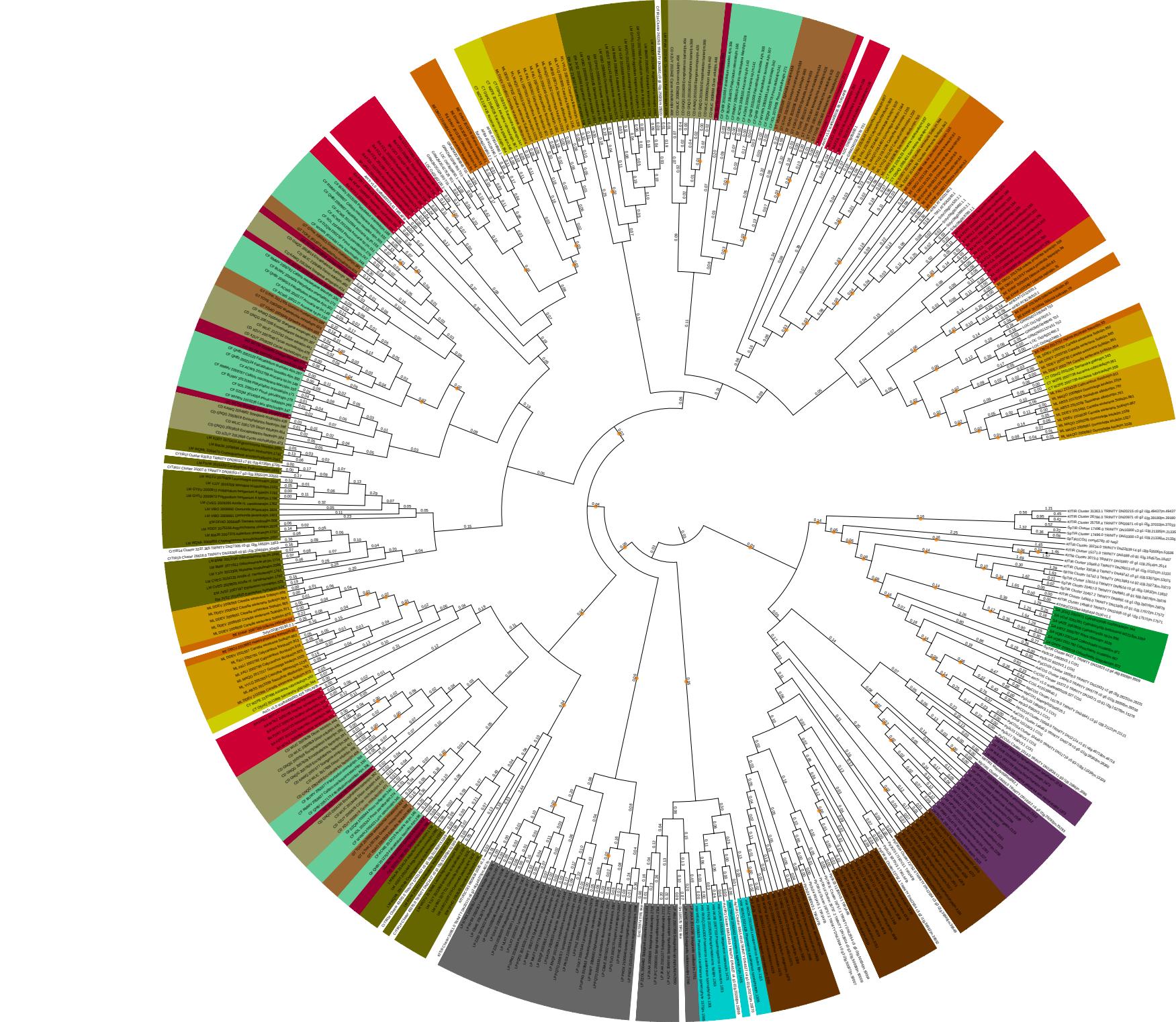


Figure 5—Figure supplement 1. Number of DEG in *de novo* assembly- or genome-based transcriptome analysis.

	De novo assembly-based		Genome-based		% Match (b/a)
	Number of DEG	Matches to genome data (a)	Number of DEG	Matches to de novo data (b)	
<i>K. nitens</i>	1094	970	1265	874	90.1
<i>M. polymorpha</i>	105	98	98	67	68.4
<i>P. patens</i>	1090	1035	1138	840	81.2

Figure 5—Figure supplement 2. Summary statistics of comparative RNA-seq analysis.

Species	Trinity Assembly Length (bp)	Corset Assembly Length (bp)	Trinity Transcripts	Corset Transcripts	Trinity Genes	Corset Genes (Clusters)	GC %	DEG (up)	DEG (down)	DEG (total)
<i>K. nitens</i>	103,789,047	96,017,668	90,079	62,718	62,978	36,209	55	409	685	1,094
<i>S. pratensis</i>	73,580,548	67,121,353	79,576	55,877	47,425	26,062	39	563	1,118	1,681
<i>A. agrestis</i>	305,335,425	302,084,291	155,454	143,697	43,514	36,512	51	65	94	159
<i>M. polymorpha</i>	127,811,269	121,162,080	87,527	63,666	52,502	30,749	48	56	49	105
<i>P. patens</i>	294,940,461	288,755,562	163,952	141,424	52,712	36,521	46	395	695	1,090
<i>C. richardii</i>	208,082,319	190,903,669	191,760	130,844	108,289	57,950	43	339	74	413

A

AA	L S M L C V F P I Q P D M D S C P G R S L G V A R K P F
WT	5' TTATCAATGCTCTGTGTGTTCCCATTCAACCTGATATGGACTCTTGGCCAGGCCGCTTTGGAGTTGCGCGAAAACCTTC 3'
nciaa-6	5' TTATCAATGCTCTGTGTGTTCCCATTCAACCTGATATGGACTCTTGGCCAGGCCGCTTTGGAGTTGCGCGAAAACCTTC 3'
nciaa-10	5' TTATCCGTATATGTATATCCATGCACAATACAGGATTGGTATATGGATATATGGATAACCGGCCATGTTAA----- 3'

AA	P P P	R L R S V K	E K P H Q F K
WT	5' CCA CCT CCG -	- AGGTT CGG TCC GT AAAG	GAAAAACCCACCAATTAAAG 3'
nciaa-6	5' CCACCTCCGTTGGAGTGGCTGAGTCA CGGTCCGTAAGG	GAAAAACCCACCAATTAAAG 3'	
nciaa-10	5' -----	CCCACCAATTAAAG 3'	

B

AA	W M R L Q Q Q K L E P A A P T T P L	S L P P L P
WT	5' TGGATGCGACTACAGCAGCAGAAATTAGAA CTC GCA - GCACCGACA ACT CCT CTC	AGCCTTCCCTCCCTTGCT 3'
ncarf-2	5' TGGATGCGACTACAGCAGCAGAAATTAGAACCTGCA CGACCGACA ACT CCT CTC	AGCCTTCCCTCCCTTGCT 3'
ncarf-10	5' TGGAT -----	TCATTAGCCT 3'

C

AA	Q L W H A C	A G G G M V
WT	5' CAGCTGTGGCACGCCCTG -----	CGCAGGA GGCA TGGTG 3'
arf3 ^{ge1} -1	5' CAGCTGTGGCACGCCCTG -----	GAGGCA TGGTG 3'
arf3 ^{ge1} -2	5' CAGCTGGCGTGCACAGCTCGCGATCAGGCCGGCATCAGGCCGGCATGCCTCCTGCGGATCAGGCCGGCAGCTC CGCAGGAGGCA TGGTG 3'	

D

AA	E K P A S F A K T L T Q S D A N
WT	5' GAGAAGCCGGCCTCGTTG CCA AGACGCTACGCAGAGCGACGCCAAC 3'
arf3 ^{ge2} -1	5' GAGAAGCCGGCCTCGT ----- CGCTCACGCAGAGCGACGCCAAC 3'

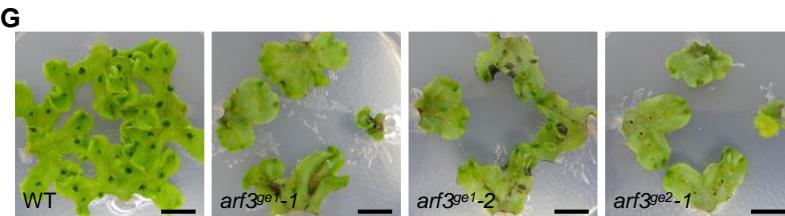
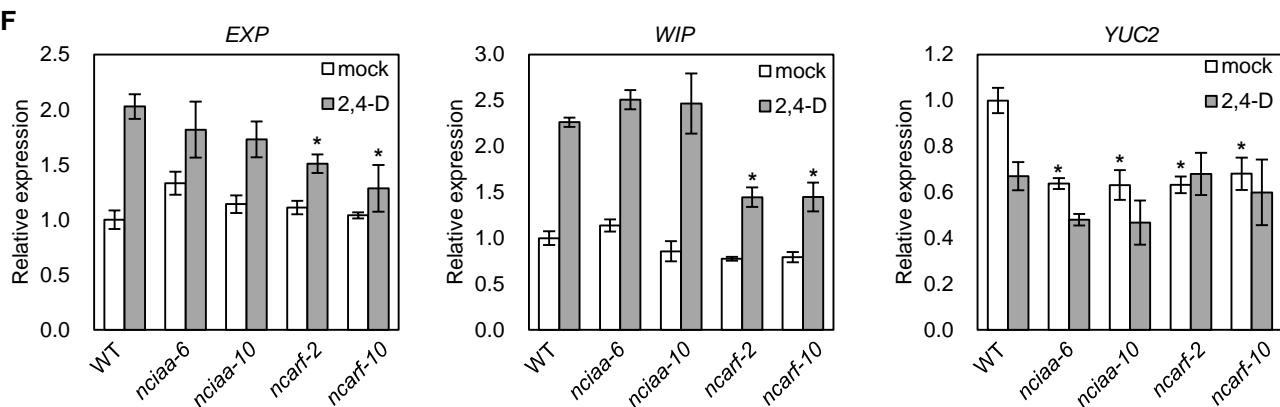
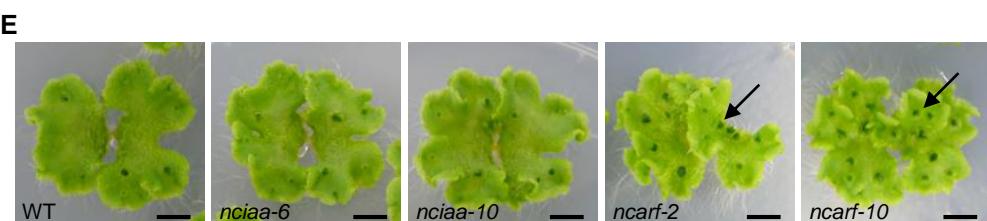


Figure 7—Figure supplement 1. CRISPR/Cas9-mediated mutagenesis in *M. polymorpha*

(A-D) Mutations detected by sequencing analysis. The amino acid (AA) sequences encoded in WT are shown at the top. WT sequence is shown with the PAM sequence highlighted in red and the target sequence of sgRNA in blue. Purple bases indicate mutation. *nciaa-6*: 6 bp deletion and 20 bp insertion, *nciaa-10*: 776 bp deletion and 75 bp insertion, *ncarf-2*: 1 bp insertion, *ncarf-10*: 486 bp deletion and 6bp insertion, *arf3^{ge1}-1*: 5 bp deletion, *arf3^{ge1}-2*: 11 bp deletion and 72 bp insertion, *arf3^{ge2}-1*: 9 bp deletion.

(E) 3-week-old gemmalings. Arrows indicate the thalli formed with up-side-down. Scale bars = 5 mm.

(F) qPCR analysis on 10-day-old gemmalings with or without 10 μ M 2,4-D treatment for 1 h.

(G) Thallus tips of WT and *arf3* mutants grown for 2 weeks. Scale bars = 5 mm.