Title

Phylogenetic distribution and expression pattern analyses identified a divergent basal body assembly protein involved in land plant spermatogenesis

Shizuka Koshimizu[1,†], Naoki Minamino[2,†], Tomoaki Nishiyama[3], Emiko Yoro[4], Kazuo Ebine[2,5], Keiko Sakakibara[4], Takashi Ueda[2,5,*], and Kentaro Yano[1,*]

[†] These authors contributed equally

[*] Corresponding authors

[1] School of Agriculture, Meiji University, Kawasaki 214-8571, Japan

[2] Division of Cellular Dynamics, National Institute for Basic Biology, Okazaki 444-8585, Japan

[3] Research Center for Experimental Modeling of Human Disease, Kanazawa University, Kanazawa 920-0934, Japan

[4] Department of Life Science, Rikkyo University, Tokyo 171-8501, Japan

[5] Department of Basic Biology, SOKENDAI (The Graduate University for Advanced Studies), Okazaki 444-8585, Japan

E-mail addresses of corresponding authors

Kentaro Yano : kyano@meiji.ac.jp

Takashi Ueda : tueda@nibb.ac.jp

Abstract

Oogamy is a form of sexual reproduction and evolved independently in animals, fungi, and plants. In streptophyte plants, Charophyceae, Coleochaetophyceae, bryophytes, lycophytes, ferns (monilophytes), and some gymnosperms (Cycads and Ginkgo) utilize spermatozoids as the male gamete. Plant spermatozoids commonly possess characteristic structures such as the spline, which consists of a microtubule array, the multilayered structure (MLS) in which the uppermost layer is continuum of the spline, and multiple flagella. However, the molecular mechanisms underpinning plant spermatogenesis remain to be elucidated. To identify the genes involved in plant spermatogenesis, we performed computational analyses and successfully found deeply divergent *BLD10*s by combining multiple methods and omics-data. We then validated the functions of candidate genes in the liverwort *Marchantia polymorpha* and the moss *Physcomitrium patens* and found that Mp*BLD10* and Pp*BLD10* are required for normal basal body and flagella formation. Mp*bld10* mutants exhibited defects in remodeling of the cytoplasm and nucleus during spermatozoid formation, thus Mp*BLD10* should be involved in chromatin reorganization and elimination of the cytoplasm during spermiogenesis. Streptophyte BLD10s are orthologous to BLD10/CEP135 family proteins, which function in basal body assembly, but we found that BLD10s evolved especially fast in land plants and MpBLD10 might obtain additional functions in spermatozoid formation through the fast molecular evolution. This study provides a successful example of combinatorial study from evolutionary and molecular genetic perspectives that elucidated a function of the key protein of the basal body formation that fast evolved in land plants.

Introduction

Oogamy is a form of sexual reproduction using female and male gametes. The female gamete (egg cell) is non-motile and larger than the male gamete, whereas male gametes (sperm) are motile and smaller than female gametes. Oogamy evolved independently in animals, fungi, and plants (Spratt 1971; Simpson 2018), and it is a big question what genes drove evolution to oogamy, i.e., sperm production.

In streptophyte plants, sexual reproduction in Charophyceae, Coleochaetophyceae, and land plants are via oogamy. Among these organisms, Charophyceae, Coleochaetophyceae, bryophytes, lycophytes, ferns (monilophytes), and some gymnosperms (cycads and ginkgo) utilize spermatozoids as the male gamete. Oogamy in streptophyte plants is presumed to have originated from a single ancestor, then flagella of spermatozoids were lost independently in angiosperms, gymnosperms (in the common ancestor of cupressophytes, gnetophytes and Pinaceae), and Zygnematophyceae (Hodges et al. 2012). Plant spermatozoids commonly possess characteristic structures such as the spline, which consists of a microtubule array, the multilayered structure (MLS) in which the uppermost layer is a continuum of the spline and basal bodies are located on it, and multiple flagella. For decades, these structural features of spermatozoids have been investigated mainly by transmission electron

59 microscopy (Norstog 1967; Carothers and Kreitner 1968; Kreitner and Carothers 1976; Graham and
60 McBride 1979; Carothers and Duckett 1980; Renzaglia et al. 1985; Renzaglia and Duckett 1987). As
61 reviewed by Renzaglia and Garbary (2001), in the spermatozoids of Charophyceae, bryophytes, and
62 ferns, after the MLS develops, the nucleus becomes compacted and helically elongated along the
63 spline, during which a major part of the cytoplasm is eliminated. Unlike in these species, in
64 gymnosperms, the nucleus is neither condensed nor elongated, and the cytoplasm is not eliminated.
65 The number of flagella in a spermatozoid of bryophytes or lycophytes is two, but a spermatozoid of
66 ferns forms 20 - 50 or more flagella. Gymnosperms spermatozoids possess 1,000 - 50,000 flagella.
67 Although morphological studies have been well conducted, the molecular and genetic players in plant
68 spermatogenesis remain to be identified.
69 Currently, in addition to angiosperms, the genome sequences of a variety of streptophytes
70 have been determined by progress of sequencing technologies (Rensing et al. 2008; Banks et al. 2011;
71 Bowman et al. 2017; Li et al. 2018; Nishiyama et al. 2018; Zhao et al. 2019; Li et al. 2020; Wang et
72 al. 2020), and a vast amount of omics data such as transcriptome have been accumulating in an online
73 database, the Sequence Read Archive (SRA; Leinonen et al. 2010). Banks et al. (2011) reported that,
74 after gene clustering, 32 of 137 'angiospermLoss' groups (defined as present in at least two of the
75 following: Chlamydomonas [*Chlamydomonas reinhardtii*], Physcomitrella [*Physcomitrium patens*],
76 and Selaginella [*Selaginella moellendorffii*] but not in 15 angiosperms) harbored genes exhibiting
77 similarity to flagella or basal body-related genes, consistent with the presence of flagellated cells in
78 the three organisms. We envisioned that combining the phylogenetic distribution and expression data
79 would yield a more specific set that could test their function using molecular genetic methods. We
80 selected candidate genes specifically expressed in male reproductive tissues of Marchantia
81 (*Marchantia polymorpha*) and Physcomitrella but excluded genes apparently present in
82 Chlamydomonas to obtain a set that is worth investigating for the function in plant spermatogenesis.
83 The functions of the candidate genes were examined in the liverwort *Marchantia polymorpha* and the
84 moss *Physcomitrium* [*Physcomitrella*] *patens* (Rensing et al. 2020), for which molecular genetic
85 techniques have been established (Schaefer 1997; Nishiyama et al. 2000; Ishizaki et al. 2008; Kubota
86 et al. 2013). Loss-of-function mutants of the candidates Mp*BLD10* and Pp*BLD10* exhibited defects in
87 basal body and flagella formation during spermatogenesis, suggesting that these genes are required for
88 normal basal body and flagella formation. *BLD10*s were found to be putative orthologs of *BLD10* in
89 *Chlamydomonas reinhardtii* (Cr*BLD10*), the product is required for assembly of the basal body.
90 However, *BLD10*s evolved fast in the land plant lineage, and loss-of-function mutations in Mp*BLD10*
91 and Pp*BLD10* resulted in different phenotypes in the liverwort and moss, which were also distinct
92 from the phenotype in the chlorophyte Cr*bld10* mutant. Defects in reorganizing the cytoplasm and
93 nucleus during spermatozoid formation in Mpbld10 mutants suggested that MpBLD10 plays a role in
94 spermatozoid formation, in addition to the basal body formation. Thus, we present the results of a
95 successful combinatorial study encompassing phylogenetic distribution, gene expression, and

96 molecular genetics approaches, which unraveled that the function of the key component of the basal
97 body formation is diverged during plant evolution.
98
99 **Results**
100 **Four protein family groups were selected as candidates involved in spermatogenesis**
101 To identify genes involved in plant spermatogenesis, we performed *in silico* analyses combining
102 different methods as the first screening. The computational approach included the following steps
103 (Supplementary Fig. S1).
104 Step 1, selection of protein family groups specific to plant species that produce
105 spermatozoids. Because plant spermatozoids have structures distinct from animal sperm, such as a
106 MLS contiguous with the spline, a spiral-shaped nucleus elongated along the spline, and multiple
107 flagella, we hypothesized that the plant species producing spermatozoids would harbor specific genes
108 not present in animals that are needed to form these distinctive structures. Therefore, we classified
109 proteins of plants producing spermatozoids and animals into family groups using the OrthoFinder tool
110 (Emms and Kelly 2019) and then selected protein family groups present only in the plants producing
111 spermatozoids. In this step, we also used protein data forChlamydomonas, a flagellate green alga that
112 does not produce spermatozoids, so that we could remove known flagella proteins and unrelated
113 proteins for spermatogenesis from among the candidates by exclusion of protein families contained in
114 Chlamydomonas. Then, 938 protein family groups remained as primary candidates (Supplementary
115 Fig. S1, Step 1).
116 Step 2, extraction of protein family groups composed by genes highly expressed during
117 spermatogenesis. We expected that genes involved in spermatogenesis should be highly expressed
118 during spermatogenesis. Based on RNA-seq data for tissues in the vegetative and male reproductive
119 stages in Marchantia (Higo et al. 2016) and Physcomitrella (Koshimizu et al. 2018), we extracted
120 genes exhibiting higher expression levels during male reproductive stages compared to vegetative
121 stages. For families consisting of multiple genes, we selected as candidates those in which all
122 members are highly expressed at the male reproductive stages as candidates. In this step, 165 groups
123 were retained (Supplementary Fig. S1, Step 2).
124 Step 3, selection of protein family groups for which member proteins exhibit low BLAST
125 similarities with animal and Chlamydomonas proteins. In Step 1, we excluded protein family groups
126 shared between plants and animals or Chlamydomonas. In this step, we further eliminated protein
127 family groups that include proteins highly similar to those of animals or Chlamydomonas. Marchantia
128 proteins in the 165 protein family groups selected in Step 2 were used for the query, and we examined
129 sequence similarities against animals and Chlamydomonas proteins by BLASTP searching (Altschul
130 et al. 1997; Camacho et al. 2009). For exhaustive analysis, we used 'animals' and '*Chlamydomonas*
131 *reinhardtii*' taxa of NCBI (NCBI Resource Coordinators 2018) nr datasets for the BLASTP search.
132 When high sequence similarity to an animal or Chlamydomonas protein was detected (e-value < 0.001

133 or coverage > 10%), we excluded the protein family group. After this step, 31 groups remained
134 (Supplementary Fig. S1, Step 3; and Supplementary Table S1).
135 Step 4, human check of the expression level data obtained in Step 2. From the Marchantia
136 and Physcomitrella expression data used in Step 2, we selected seven genes exhibiting lower
137 expression levels in the vegetative growth stage and substantial differences in expression levels
138 between the vegetative and male reproductive stages. We then selected as candidates four protein
139 family groups (Group-75, Group-89, Group-230, and Group-339) composed of seven proteins. It
140 should be noted that although we could have initially selected genes exhibiting above-mentioned
141 expression patterns, we preferred to use a strategy that narrows down the number of candidates after
142 studying the functions of a broad range of proteins that could be involved in spermatogenesis
143 (Supplementary Fig. S1, Step 4).
144
145 **Two protein family groups were selected as final candidates for functional analysis**
146 Regarding Group-339, the function of a highly similar protein in Arabidopsis, AUG7 (AT5G17620.1),
147 in delocalization of γ-tubulin in the mitotic spindle and phragmoplast was reported (Hotta et al. 2012).
148 In spermatogenesis of bryophytes, centrioles serve as the microtubule organizing centers in the
149 spindle for the final mitosis (Vaughn and Renzaglia 1998), in which γ-tubulin localizes to the
150 centriolar centrosomes (Shimamura et al. 2004). These observations suggested that the proteins of
151 Group-339 should play a role in γ-tubulin localization during spermatogenesis; thus, we reserved
152 further functional analysis for Group-339. Group-75 included four proteins each of Marchantia and
153 Physcomitrella. Because the deletion of all genes for four members would be technically difficult
154 even in Marchantia or Physcomitrella, we also reserved this group for future analysis. The
155 Arabidopsis protein highly similar to Group-89 is DUO3 (AT1G64570.1), which regulates male
156 germline development, and is essential for sperm cell specification and fertilization (Brownfield et al.
157 2009). It would be interesting to investigate the function of this protein in spermatogenesis in
158 bryophytes. Group-230 proteins exhibited no similarity to Arabidopsis proteins, but it did exhibit a
159 weak similarity to Chlamydomonas BLD10 (Cre10.g418250.t1.2, CrBLD10), a cartwheel protein
160 essential for assembly of the basal body that functions as the origin of flagella (Matsuura et al. 2004;
161 Hiraki et al. 2007). In Phytozome ver. 5, which the analysis by Banks et al. (2011) was based on, the
162 member of Group-230, Selaginella SELMODRAFT_427424 and PHYPADRAFT_69693 (v1.1, older
163 model; same locus as Pp3c9_9040V3.2, but not the same exon-intron structure prediction), were
164 placed in the same group with CrBLD10. However, the similarity was so subtle that the relation was
165 not detected in Step 3. Current Phytozome ver. 12 "Gene Ancestry" for the viridiplantae places the
166 Physcomitrella and Marchantia to different groups containing only mosses and Marchantia,
167 respectively. No description on the encoding gene in Marchantia has been made, and the
168 corresponding gene is annotated as a 'structural maintenance of chromosomes smc family member',
169 with no publications reporting results of functional analyses in Physcomitrella. Thus, we decided to

170 conduct further functional analyses for Group-89 and Group-230 consisting of one member each in
171 Marchantia, particularly focusing our interest on spermatogenesis (Supplementary Fig. S1, Step 5).
172
173 **Mp*bld10* mutants are defective in spermatozoid formation**
174 To analyze the roles of these two genes, we generated knock-out lines by genome editing using the
175 CRISPR/Cas9 system (Ran et al. 2013; Sugano et al. 2018). No mutants were obtained for the
176 *Mapoly0029s0108* gene (Group-89). For the *Mapoly0001s0460* gene (Group-230), the guide-RNA
177 sequence was designed to target the first exon (Fig. 1A), and two independent lines harboring
178 frameshift mutations were obtained (Fig. 1A and Supplementary Fig. S2A). Hereinafter, we refer to
179 Group-230 family proteins as BLD10 because the phenotype and sequence analyses suggested an
180 orthologous relationship to CrBLD10. The mutants were designated Mp*bld10-1* and Mp*bld10-2*.
181 These mutations did not markedly affect vegetative growth of the thalli (Supplementary Fig. S2B) and
182 formation of antheridiophores (Supplementary Fig. S2C-S2E). However, moving spermatozoids of the
183 mutants were rarely observed for the mutants (Supplementary Movie 1-3). Intriguingly, cytoplasm
184 elimination, nuclear elongation, and flagella formation were incomplete in the mutant spermatozoids
185 compared with wild-type spermatozoids (Fig. 1B-1E). Immunostaining of centrin and acetylated
186 tubulin (ac-tubulin) in spermatids was performed to observe basal bodies and the axoneme in the
187 flagella (Fig. 1F-1I, Higo et al. 2018). Filaments of ac-tubulin were detected in a subpopulation of
188 Mp*bld10*-spermatids, but spermatids with no detectable ac-tubulin were also observed. In spermatids
189 positive for the ac-tubulin signal, short or coiled filaments in the cell bodies were frequently noted
190 (Fig. 1G and 1H). The puncta signals of centrin exhibiting an abnormal size were sometimes observed
191 in the mutant (Fig. 1I). These results suggested that Mp*BLD10* plays a crucial role in spermatozoid
192 formation in Marchantia.
193 To examine the effect of the mutation in Mp*BLD10* at an ultrastructural level, we conducted
194 a transmission electron microscopy (TEM) analysis of spermatids and spermatozoids of the Mp*bld10-*
195 *1 mutant*. The flagella of wild-type spermatids contained an axoneme comprising two central
196 microtubules and surrounding nine doublet microtubules (Fig. 2A). In the Mp*bld10-1 mutant*,
197 however, a major population of spermatids did not harbor flagella, and the flagella formed in a
198 subpopulation of spermatids exhibited a disordered axoneme structure (Fig. 2E). No structural
199 abnormalities were detected in the Mp*bld10-1* mutant MLS, a structure unique to plants that is
200 attached to the anterior mitochondrion in spermatids and consists of the spline, which is the
201 uppermost stratum containing arrayed microtubules, and a lower strata with high electron densities,
202 namely, the lamellar strip (Fig. 2B and 2F). We also observed wild-type basal bodies, which contain
203 nine triplet microtubules and are attached to the spline of the MLS (Fig 2C and 2D). In the Mp*bld10-1*
204 *mutant*, the wild type-like basal bodies were only occasionally observed, and amorphous electron-
205 dense regions were frequently observed instead of basal bodies (Fig. 2G and 2H). These results
206 strongly suggested that Mp*BLD10* is required for correct assembly of the basal body and flagella

207 during spermatogenesis. Furthermore, we observed that the Mp*bld10-1* spermatozoids also exhibited a
208 defect in chromatin compaction in the nucleus (Fig. 2I and 2J). Thus, Mp*BLD10* might also be
209 involved in chromatin organization during spermiogenesis in Marchantia.
210
211 **PpBLD10 is required for formation of basal bodies and flagella**
212 To further examine the functions of the Group-230 proteins (BLD10s) in bryophytes, we analyzed the
213 ortholog of Mp*BLD10* in Physcomitrella, the model of moss readily amenable to gene targeting and
214 genome editing. The Group-230 in Physcomitrella consists of only *Pp3c9_9040V3.2* (Pp*BLD10*). We
215 generated knock-out mutants of this gene using the CRISPR-Cas9 system. Pp*bld10-22*, which has an
216 11-bp deletion in the first exon, and Pp*bld10-30*, which has an approximately 10-kbp deletion
217 between exon 1 and exon 29 of the transcript XM_024529611.1, were used in further analyses
218 (Supplementary Fig. S3). Although these mutants did not exhibit any marked defects in protonemata,
219 gametophores, and gametangia (Supplementary Fig. S4), spermatozoids of these mutants were not
220 motile; no moving mutant spermatozoids were observed (Supplementary Movie 4-6). Spermatids
221 lacking the signal or with coiled filaments were observed in the mutants immunostained for ac-tubulin
222 (Supplementary Fig. S5C-S5F), indicating that the mutants are defective in flagella formation, similar
223 to Marchantia. No marked defects in elimination of the cytoplasm and nuclear elongation of the
224 mutants were observed (Supplementary Fig. S6A-S6C).
225 In the TEM analysis of spermatids and spermatozoids, no axoneme structure was observed
226 in most of the spermatids in Pp*bld10-30* (Supplementary Fig. S7A and S7C), and amorphous electron-
227 dense regions were observed instead of basal bodies (Supplementary Fig. S7B and S7D). With regard
228 to the MLS, no marked defects were observed in the mutant, similar to the Mp*bld10-1* mutant
229 (Supplementary Fig. S7B and S7D). In contrast to the Mp*bld10-1*, no defect in chromatin compaction
230 in the nucleus was observed in Pp*bld10-30* (Supplementary Fig. S7E and S7F). These results suggest
231 that PpBLD10 functions in basal body- and flagella assembly, but the requirement for PpBLD10 in
232 chromatin compaction and nuclear formation differs from the case in Marchantia.
233
234 **MpBLD10 and PpBLD10 localize in basal bodies during flagella formation**
235 We next examined the subcellular localization of MpBLD10. We generated a transgenic line of
236 Marchantia expressing mCitrine-fused MpBLD10 driven by its own promoter in Mp*bld10-1*.
237 Expression of mCitrine-MpBLD10 restored the defects in spermatozoid formation and motility in the
238 mutant, indicating that this chimeric protein retains the authentic function (Supplementary Movie 7).
239 In this transgenic line, we traced spermiogenesis according to developmental stage as previously
240 defined (Minamino et al. 2021). In stage 0 spermatids, mCitrine-MpBLD10 was observed as two
241 closely aligned rod-like structures (Fig. 3A). In stage 1 spermatids, mCitrine-positive structures were
242 located at the base of the flagella (Fig. 3B). In stage 2, the mCitrine-MpBLD10 signal became weak,
243 then ultimately disappeared in subsequent stages (Fig. 3C-3F). To verify the nature of these structures,

244 we performed co-immunostaining with centrin and ac-tubulin. As shown in Fig. 3G, mCitrine-
245 MpBLD10 was localized in close association with centrin at the proximal side of the flagella (Fig.
246 3G). These results suggest that MpBLD10 localize in the basal body during flagella formation and is
247 then degraded after flagella formation.
248 The localization of PpBLD10 in Physcomitrella was examined using Citrine knock-in lines
249 (Supplementary Fig. S3). Spermiogenesis was observed according to the developmental stages
250 defined for Marchantia (Minamino et al. 2021) (Supplementary Fig. S8A-S8E), but in Physcomitrella,
251 flagella formation was slower than in Marchantia and began when the nuclear shape was spindle-like;
252 thus, stages 0 and 1 as defined for Marchantia are indistinguishable in Physcomitrella. PpBLD10-
253 Citrine signals were observed in stage 0-1 and stage 2 but disappeared in subsequent stages, as in
254 Marchantia (Supplementary Fig. S8A-S8E). In stage 0-1, the PpBLD10-Citrine was observed as
255 puncta (Supplementary Fig. S8A); in stage 2, the signal was detected in a wider region of the basal
256 part of the flagella as compared with stage 0-1 (Supplementary Fig. S8B and S8F). The positional
257 relationship between PpBLD10-Citrine and centrin in Physcomitrella could not be observed because
258 the anti-centrin antibody did not immunostain Physcomitrella spermatids (data not shown).
259
260 **BLD10s exist in streptophytes with flagella and evolved fast in land plants**
261 CrBLD10, to which MpBLD10 and PpBLD10 exhibit weak similarity, is an ortholog of *Homo*
262 *sapiens* CEP135 (HsCEP135) (Carvalho-Santos et al. 2010), which plays a role in early basal
263 body/centriole biogenesis (Kleylein-Sohn et al. 2007). To obtain information on BLD10 proteins in
264 the streptophyte lineage and to assess their orthology to the BLD10/CEP135 family, we searched the
265 genome, transcript, and protein sequences of species ranging from streptophyte algae to angiosperms
266 (angiosperms: *Arabidopsis thaliana* and *Oryza sativa*; gnetophytes: *Gnetum montanum*; Pinaceae:
267 *Picea abies* and *Pinus taeda*; Ginkgo: *Ginkgo biloba*; monilophytes: *Salvinia cucullata* and *Azolla*
268 *filiculoides*; lycophytes: *Selaginella moellendorffii*; bryophytes: *Anthoceros punctatus*;
269 Zygnematophyceae: *Penium margaritaceum*, *Mesotaenium endlicherianum*, and *Spirogloea*
270 *muscicola*; Coleochaetophyceae: *Coleochaete orbicularis*; Charophyceae: *Chara braunii*;
271 Klebsormidiophyceae: *Klebsormidium nitens*; Mesostigmatophyceae: *Mesostigma viride*;
272 Chorokybophyceae: *Chlorokybus atmophyticus*). In Arabidopsis, Oryza, Gnetum, Picea, Pinus,
273 Penium, Mesotaenium, and Spirogloea, no hit sequences were obtained. Some of the hit sequences
274 were reconstructed based on RNA-seq data and similarity to preliminary alignments. Finally, Ginkgo
275 (GbBLD10a and GbBLD10b), Salvinia (Sacu_v1.1_s0007.g003760), Azolla (AfBLD), Selaginella
276 (SmBLD10), Anthoceros (Apun_evm.model.utg000038l.487.1), Coleochaete (GBSL01053926.1),
277 Chara (CbBLD10), Klebsormidium (kfl00353_0080_v1.1), Mesostigma (MvBLD10), and
278 Chlorokybus (Chrsp112S01623) were subjected to the alignment and phylogenetic analyses together
279 with MpBLD10, PpBLD10, and CrBLD10 and non-plant BLD10/CEP135 family sequences from *H.*
280 *sapiens* (HsCEP135), *Drosophila melanogaster* (DmBLD10), and *Tetrahymena thermophila*

281 (TtBLD10) as outgroup sequences (Fig. 4A and Supplementary Fig. S9). Additionally, we included
282 the additional sequences *Adiantum capillus-veneris* MBC9850943.1, *Chlamydomonas eustigma*
283 CEUSTIGMA_g448.t1, and *Bombus impatiens* XP_012244165.1 to stabilize the alignment and
284 phylogenetic tree by dividing long branches. Here, among multiple potential isoforms, ref_seq protein
285 XP_024385379.1 was used for the phylogenetic analysis because this isoform of PpBLD10 was well
286 supported by RNA-seq data and fit the alignment better.
287 In the phylogenetic analysis, the green plant, streptophyte, and land plant genes formed a
288 clade (Fig. 4A). Among land plants, the setaphytes (mosses + liverworts), and euphyllophytes formed
289 a clade with low and high bootstrap supports, respectively. Outside of land plants,
290 Coleochaetophyceae, Charophyceae, Klebsormidiophyceae, and Chlorokybophyceae branched in that
291 order, with low bootstrap support. The branches of the land plants were longer than those of
292 streptophyte algae (Fig. 4A). The molecular clock hypothesis was rejected between land plants and
293 streptophyte algae, with Chlamydomonas as out-group by Tajima's relative rate test ($p < 0.05$; Tajima
294 1993), indicating an increase in the evolutionary rate of BLD10s in the land plant lineage. Upon
295 application of a local clock model in PAML (Yang 2007) to the dataset for green plants, such that the
296 branches of land plants after divergence from Coleochaetophyceae has different rates than all other
297 branches, land plants had an approximately 2-fold higher evolutionary rate than green algae.
298
299 **Green plant BLD10s have a novel conserved domain close to the C-terminus**
300 **Functional regions were reported for** CrBLD10 and HsCEP135: a probably essential region for the
301 function of CrBLD10 (residues 850 - 1,050; Hiraki et al. 2007), and three binding regions in
302 HsCEP135 (microtubule-binding region, residues 1 - 190; CPAP-binding region, residues 50 - 460;
303 hSAS-6-binding region, residues 416 - 1,140; Lin et al. 2013). Among streptophyte BLD10s, three
304 conserved regions were identified (Fig. 4B and Supplementary Fig. S9). Region 1 does not seem to
305 exist in Selaginella, because the N-terminal region is annotated as a separate protein in the ref_seq
306 annotation (XP_024518287.1). The N-terminus of region 1 was not detected in either Ginkgo or
307 Adiantum. As such, the level of conservation shown in Fig. 4B and Supplementary Fig. S9 is low, but
308 region 1 is conserved in other streptophytes. The microtubule-binding region and the N-terminal half
309 of the CPAP-binding region correspond to region 1 in streptophytes, but a large proportion of the
310 BLD10/CEP135 family is so divergent that the conserved residues are rare (Supplementary Fig. S9).
311 The C-terminal half of the CPAP-binding region was found to be mostly conserved among green algae
312 including Coleochaete, but it was shortened in land plants. Region 2 includes the probably essential
313 region of CrBLD10 and overlaps with the hSAS-6-binding region. Region 2 is a long and conserved,
314 but a large deletion of 49 residues in the CrBLD10 essential region was found in MpBLD10
315 (Supplementary Fig. S9). A review of RNA-seq data mapped to the region confirmed that the loss in
316 MpBLD10 was not due to an annotation error skipping an exon. The 27 N-terminal residues are
317 highly conserved in green plants. In streptophytes, region 3 is a 65-residues long highly conserved

region close to the C-terminus (Supplementary Fig. S9). Although some residues aligned, the distances to human, fly, and Tetrahymena were so large that the homology is obscure, and the alignment was unstable to additional insect sequences. Similarity to outside of land plants cannot be detected through PSI-BLAST (Altschul et al. 1997) with land plant sequences in NCBI, and no conserved domain was found in the conserved domain database (CDD; Lu et al. 2020). An examination of the alignment revealed that, CrBLD10 has region 3 starting with the signature (KR)XX(ED)LE and extending to (LVM)(LV)X(LI)(LM)(SA)(KR)(VL)(DE)X(DE)(RK) except Q rich block, thus it was judged to be homologous. Region 3 constitutes a novel conserved domain in green plants. A 3-amino acid (aa) deletion and lower conservation in the N-terminus of the region was noted in Klebsormidium. The intron-exon structure is fully supported by RNA-seq data.

Discussion

In this study, we established a pipeline that allows for efficient and rapid searches of protein family groups involved in plant spermatogenesis through a computational analysis of large omics datasets stored in publicly available online databases. Our *in silico* approach involves two primary steps: (i) selection of protein families specific to streptophytes that produce spermatozoid, and (ii) selection of protein families encoded by genes highly and predominantly expressed during spermatogenesis. Using this pipeline, we extracted seven genes from approximately 19,000 Marchantia genes belonging to four protein family groups (Group-75, Group-89, Group-230, and Group-339). Among these protein families, we conducted functional analyses of Group-89 and Group-230 in Marchantia. No Group-89 mutants were obtained, but we successfully generated loss-of-function mutants for Group-230 (BLD10s). We then found that the BLD10s play a crucial role in spermatogenesis in Marchantia and Physcomitrella, thus showing the effectiveness and accuracy of our computational selection approach. This method is applicable to the identification of genes involved in a variety of biological processes other than spermatogenesis using large omics datasets.

Most spermatozoids of the Mpbld10 mutants did not have flagella. A few spermatozoids possessed short or coiled filaments within the cytoplasm (fig. 1G-1I), likely corresponding to incomplete flagella observed in a TEM image (fig. 2E). Similar results were observed in Pp*bld10* mutants (Supplementary Fig. S5C-S5F and Supplementary Fig. S7C). Generally, flagella are observed outside the cell body in the early stages of spermatogenesis (Minamino et al. 2021), but the immature flagella-like structures remained inside the cell in the Mp*bld10* and Pp*bld10* mutants. In addition to this shared characteristic of Marchantia and Physcomitrella, the Mp*bld10* mutants also exhibited defective cytoplasm elimination, chromatin compaction, and nuclear elongation (Fig. 1C-1E and Fig. 2J), which were not observed in the Pp*bld10* mutants (Supplementary Fig. S6B and S6C and Supplementary Fig. S7F). MpBLD10 and PpBLD10 were localized at the base of forming flagella in the early stages of spermiogenesis and disappeared after flagella formation was completed (Fig. 3A-3F and Supplementary Fig. S8A-S8E). This localization, together with the close association between

355 MpBLD10 and centrin at the base of the flagella (Fig. 3G), suggests that MpBLD10 and PpBLD10
356 are basal body proteins that function in basal body assembly. MpBLD10 is likely also involved in
357 chromatin reorganization during spermiogenesis. The additional effect of the Mp*bld10* mutation on
358 phenotypes shared between Mp*bld10 a*nd Pp*bld10* suggests that MpBLD10 plays an additional role
359 during spermiogenesis and that the functions of MpBLD10 and PpBLD10 partially diverged during
360 bryophyte evolution.

361 In the phylogenetic analysis, the branching order among streptophyte algae was congruent
362 with an organism tree based on phylotranscriptomic analyses (Wickett et al. 2014; Puttick et al. 2018;
363 Leebens-Mack et al. 2019) with low bootstrap support (Fig. 4A). Although the monophyly of land
364 plant genes was supported by a high bootstrap value, among land plants, the tree was congruent with
365 the phylogeny of organisms with low bootstrap supports, except for the position of the hornwort gene.
366 The placement of hornworts sister to all other land plants has been observed at a low frequency in low
367 copy number gene phylogenies of land plants (Li et al. 2020). Based on the congruence to the
368 phylogeny, these streptophyte BLD10s are putative orthologs of CrBLD10 (fig. 4A). Flagellate
369 species in streptophytes have BLD10; conversely, species without flagella (i.e., Zygnematophyceae,
370 conifers/gnetophytes, and angiosperms) lost BLD10 (Fig. 5). BLD10/CEP135 family proteins
371 function in assembly of the basal body/centriole, which is involved in cell division and serves as the
372 basis of flagella/cilia in human (Kleylein-Sohn et al. 2007; Lin et al. 2013), fly (Mottier-Pavie and
373 Megraw 2009; Carvalho-Santos et al. 2012), Tetrahymena (Bayless et al. 2012), and Chlamydomonas
374 (Matsuura et al. 2004; Hiraki et al. 2007). MpBLD10 and PpBLD10 also play a role in basal body
375 assembly, similar to other BLD10/CEP135 family proteins. In accordance with the differences in
376 BLD10 sequences between land plants and Chlamydomonas, the mutant phenotypes differed from the
377 mutants of other BLD10/CEP135 family proteins; basal bodies were completely lacking and flagella
378 never observed in Cr*bld10*, but incomplete basal bodies and flagella were formed in the Mp*bld10* and
379 Pp*bld10* mutants. Although DmBLD10 remains after the development of sperms in fly (Mottier-Pavie
380 and Megraw 2009), MpBLD10 and PpBLD10 disappeared, suggesting that MpBLD10 and PpBLD10
381 are necessary only during flagella formation, and no longer needed after the formation of flagella. In
382 addition, Mp*bld10* mutants exhibited defects in cytoplasmic reduction and nuclear elongation during
383 spermiogenesis, and this phenotype was observed specifically in Marchantia but not in
384 Physcomitrella. Changes in the role of basal body/centriole in the life cycle or environment of
385 fertilization (free water, limited water, or pollination droplet) in land plants might have affected the
386 evolutionary rate of BLD10s in land plants through positive selection or a relaxation of purifying
387 selection, which cannot be discerned from the current data. Note that centrioles are not observed
388 during cell division in land plants (Buschmann and Zachgo 2016) (Fig. 5). Despite *Klebsormidium*
389 *nitens* NIES-2285 do not show flagellated cells under laboratory conditions, BLD10 expression is
390 detected in RNA-seq data, implying their role in cell division rather than flagella formation. BLD10
391 probably played a dual role in the flagella formation and cell division in the ancestral green algae and

392 then lost the role in cell division in land plants after other microtubule organization mechanisms for
393 chromosome separation during cell division were established; thus, the protein has a sperm-specific
394 function in bryophytes, lycophytes, and ferns. BLD10 gene was lost entirely in conifers/gnetophytes
395 and angiosperms, consistent with the loss of flagellated sperm cells.
396 We successfully identified genes that function in plant spermatogenesis using an *in silico*
397 analysis. Two identified proteins, MpBLD10 and PpBLD10, are basal body proteins that exhibit a
398 higher evolutionary rate despite the importance of their role in assembly of the basal body during
399 spermatogenesis. MpBLD10 possibly acquired a new role in spermatozoids formation—specifically,
400 in cytoplasm elimination and nuclear elongation—through plant evolution. Notably, the relationship
401 that streptophyte BLD10s are putative orthologs of BLD10 was not found in a simple run of
402 OrthoFinder, and identified only after a more detailed analysis. The importance of gene loss in the
403 evolution was recently documented in various fields/cases, and phylogenetic distribution of the gene
404 involved in a particular trait coincides with the distribution of the trait (Glastad et al. 2011; Zhang et
405 al. 2013; Lin et al. 2017; Griesmann et al. 2018; Sharma et al. 2018; Gluck-Thaler et al. 2020). In this
406 analysis, we demonstrated that the process can be reversed; that is, a gene responsible for the trait can
407 be identified by searching for a gene whose phylogenetic distribution coincides with the trait
408 combined with the expression at the site the trait is observed. The results of this study thus highlight
409 the power of combinatorial analyses with expression and multiple independent losses of traits and
410 underlying genes during evolution.
411
412 **Materials and Methods**
413 ***In silico* screening**
414 Ortholog groups were predicted using OrthoFinder (v2.3.3, Emms and Kelly 2019) with the protein
415 sequences of the species shown in Supplementary Table S2. For expression level quantification in
416 Marchantia and Physcomitrella, transcripts per million (TPM) values were calculated using RSEM
417 (v1.3.1, Li et al. 2011) with Bowtie2 (v2.3.5.1, Langmead et al. 2012) for RNA-seq data listed in
418 Supplementary Table S3 using the reference transcriptome datasets listed in Supplementary Table S2.
419
420 **Sequence searching and reconstruction of gene models**
421 BLD10 proteins in the streptophytes were searched against the genome, transcript, and protein
422 sequences using BLAST (v2.10.1+, Altschul et al. 1997; Camacho et al. 2009). In the protein datasets,
423 BLASTP searching was performed using MpBLD10 against the species shown in Supplementary
424 Table S2. Hit protein sequences are listed in Supplementary Table S4. The lengths of hit sequences in
425 Ginkgo and Chara were <600 aa (full length of MpBLD10 is 1,123 aa). Because the close gene ids
426 implied close positions in the genome, genomic locations were investigated. Gb_13822 is at the 279
427 Mb position on Chr12; Gb_03087-Gb03089, Gb_30854, and Gb_30855 are close to the 630 Mb
428 position on Chr12; Gb_39501 is close to the 277 Mb position on Chr7. Two gene models on Chr12

429 were reconstructed based on mapping of RNAseq data (*GbBLD10a* and *GbBLD10b*), and no reads
430 were found for Gb_39501. In Chara, we constructed a presumptive transcript (*CbBLD10*) from the
431 genome and RNA-seq data (Nishiyama et al. 2018). Azolla sequence Azfi_s0013.g013382 had a
432 deletion in the C-terminal region. The missed exons were supplied based on the RNA-seq data (Li et
433 al. 2018), and a gene model (Af*BLD10*) was constructed. In Selaginella, the hit sequence
434 SELMODRAFT_427424 had three deletions and one insertion relative to most of the green plant
435 BLD10s in the initial alignment. By investigating the genomic sequence with the preliminary
436 alignment as a guide, a gene model more similar to the conserved consensus was constructed
437 (*SmBLD10*), though the RNA-seq data of sperm-producing tissue in Selaginella were insufficient.
438 Although SELMODRAFT_427424 resides on scaffold_90, another copy (allele) is present on
439 scaffold_104, but contained an assembly gap (stretch of Ns) and was not annotated. Further,
440 transcriptome shotgun assemblies (TSA) were searched using TBLASTN at NCBI (NCBI Resource
441 Coordinators 2018) with CrBLD10. Hit transcript sequences for *Mesostigma viride* and *Coleochaete*
442 *orbicularis* were obtained (Supplementary Table S4). Mesostigma had split hits of the different
443 contigs with reasonable similarity, presumably constituting the corresponding protein. The three
444 contigs were mapped to scaffold_80 between 197,517 and 234,765. Thus, a gene model encoding
445 1744 residues (MvBLD10) was constructed for the region mostly based on RNA-seq mapping and a
446 little guess work. The *Coleochaete orbicularis* contig appeared to contain a complete coding
447 sequence. The reconstructed gene models are shown in Supplementary Data 1 and 2.
448
449 **Method for reconstructing gene models based on RNA-seq data or alignment similarity**
450 Ginkgo, Azolla, Selaginella, Chara, and Mesostigma BLD10 sequences were reconstructed according
451 to the following method. RNA-seq data were previously published or downloaded from SRA using
452 fastq-dump (Supplementary Table S3). The RNA-seq data were mapped to a corresponding reference
453 genome using hisat2 (v2.2.1; Kim et al. 2019). The mapped bam files were sorted according to the
454 coordinates using samtools (Danecek et al. 2021) and then indexed. The reference sequence and
455 annotations in gff files were loaded to JBrowse-1.15.4 (Buels et al. 2016) using prepare-reference and
456 flatfile-to-json, respectively. The indexed bam files were loaded using add-bam-track. After preparing
457 the data directory of JBrowse, the data were connected to Apollo-2.1.0 for manual editing. The gene
458 model or bam read alignment was chosen for the new gene models and merged or extended in the user
459 model editing pane. For the *Ginkgo biloba* reference, the bam file could not be indexed due to its large
460 reference size (>512 Mb). Thus, the reference was cut to 500 Mb, and Chr12, at >500 Mb, were
461 named as 'Chr12b' for subsequent processing. The records of bam files were screened for Chr7 at
462 <500 Mb and Chr12, and records for Chr12 >500 Mb were placed to Chr12b, and positions were
463 subtracted by 500 million. The reference genome gff was similarly edited. Thus, Chr7, Chr12 up to
464 500 Mb, and Chr12 over 500 Mb were loaded to JBrowse and apollo2.
465

466 **Phylogenetic analyses**

467 The protein sequences HsCEP135 (ENSP00000257287), DmBLD10 (FBpp0075391), and TtBLD10
468 (TTHERM_01164140) were obtained via Ensembl (Yates et al. 2020), FlyBase (Larkin et al. 2020),
469 and TGD (Stover et al. 2006), respectively. *Adiantum capillus-veneris* MBC9850943.1,
470 *Chlamydomonas eustigma* CEUSTIGMA_g448.t1, and *Bombus impatiens* XP_012244165.1 were
471 obtained from NCBI. Multiple sequence alignment was performed using MAFFT (Katoh and
472 Standley 2013) with the accurate option E-INS-i in Jalview software (v2.10.5, Waterhouse et al.
473 2009). The conservation level was calculated based on that used in the AMAS method for multiple
474 sequence alignment analysis (Livingstone and Barton 1993) in Jalview. A phylogenetic tree was
475 constructed based on the multiple alignment with complete deletion of gap sites using the maximum-
476 likelihood method of MEGA X software (Kumar et al. 2018), with 1,000 bootstrap replicates. Jones-
477 Taylor-Thornton (JTT) matrix-based model (Jones et al. 1992) with gamma-distribution among sites
478 and subtree-pruning-regrafting - extensive (SPR level 5) were used for amino acid substitution model
479 and heuristic methods. Protein regions used in the analyses are shown in Supplementary Fig. S9
480 (orange boxes). Differences in branch length between land plants and green algae were compared
481 using the local clock model in PAML4 (v4.9j; Yang 2007). Branches in land plants and the branch
482 leading to land plants after divergence with Coleochaetophyceae were assigned category #1 in the
483 comparison. JTT model was used for amino acid rate. Other parameters were the same as those of the
484 'aaml.ctl' file distributed with the software.

485

486 **Plant materials**

487 Male accession of Marchantia, Takaragaike-1 (Tak-1), was used for observing spermatogenesis in this
488 study. Plants were grown on 1/2× Gamborg's B5 medium containing 1.4% (w/v) agar at 22ºC in
489 continuous white light. Induction of sexual organ generation by far-red irradiation was performed as
490 described previously (Chiyoda et al. 2008).

491 Physcomitrella Gransden 'Cove-NIBB' line (Nishiyama et al. 2000) was used as the wild
492 type. Physcomitrella was cultured on BCDAT medium with 0.8% (w/v) agar at 25 °C under
493 continuous light conditions for protonemata (Nishiyama et al. 2000). Protonemata were transplanted
494 into sterile peat pots (Jiffy-7; Jiffy Products International AS) and cultured approximately one month
495 at 25 °C in continuous light for gametophores. To obtain gametangia and sporophytes, the peat pots
496 containing gametophores were incubated at 15 °C under short-day (8-h light and 16-h dark)
497 conditions (Sakakibara et al. 2008).

498

499 **Transformation of Marchantia**

500 To construct vectors for genome editing, target sequences were selected using CRISPR direct
501 (https://crispr.dbcls.jp/) (Naito et al. 2015), and double-stranded oligonucleotides of the target
502 sequences were inserted into the pMpGE_En03 vector (Sugano et al. 2018). The resultant gRNA

503 cassettes were introduced into the pMpGE010 vector using Gateway LR Clonase II Enzyme Mix
504 (Thermo Fisher Scientific). To express Citrine under regulation of the Mp*BLD10* promoter, 5 kb of
505 the 5' flanking region of Mp*BLD10* with a SmaI site was introduced into the pENTR/D-TOPO vector
506 (Invitrogen). The chimeric sequence was introduced into pMpGWB307 (Ishizaki et al. 2015) using
507 Gateway LR Clonase II Enzyme Mix. To construct the mCitrine-MpBLD10 vector, a genomic
508 fragment containing the coding region, intron, and 4 kb of the 3' flanking region of Mp*BLD10* was
509 inserted into the SmaI site of the pENTR/D-TOPO vector containing $_{pro}$Mp*BLD10* by using an In-
510 Fusion HD cloning kit (Clontech). A silent mutation was introduced into the PAM site by inverse
511 PCR. The CDS of monomeric Citrine was introduced into the SmaI site of the pENTR/D-TOPO
512 vector containing the Mp*BLD10* genomic fragment by using an In-Fusion HD cloning kit (Clontech).
513 The chimeric sequence was introduced into pMpGWB301 (Ishizaki et al. 2015) using Gateway LR
514 Clonase II Enzyme Mix. The primer sequences are listed in Supplementary Table S5.
515 Transformation of Marchantia was performed according to previously described methods
516 (Ishizaki et al. 2008; Kubota et al. 2013). Transgenic lines were selected with 10 mg l^{-1} hygromycin B
517 and 100 mg l^{-1} cefotaxime for pMpGE010, and 0.5 μM chlorsulfuron and 100 mg l^{-1} cefotaxime for
518 pMpGWB301.
519
520 **Transformation of Physcomitrella**
521 To perform CRISPR-mediated mutagenesis, we designed oligodeoxynucleotides used as single-guide
522 RNAs (sgRNAs) targeting Pp*BLD10* (Supplementary Fig. S3B) using CRISPRdirect
523 (https://crispr.dbcls.jp/; last visited April, 2021). The annealed oligodeoxynucleotides were cloned into
524 the BsaI site of pPpU6-sgRNA (LC494193) (Gu et al. 2020). To generate genome editing mutants,
525 protoplasts were co-transformed with a total of 6 μg of circular DNAs divided as follows: 2 μg of
526 pPpU6-Pp*BLD10*-sgRNA#1 or pPpU6-Pp*BLD10*-sgRNA#2, 2 μg of pAct-Cas9 (Collonnier et al.
527 2017), and 2 μg of pBHRF (Schaefer et al. 2010). After transient hygromycin-resistance selection,
528 transformants were recovered on a non-selective medium. To detect mutations, we amplified the
529 targeting region by PCR using three sets of primers (Supplementary Table S5) and sequenced the
530 fragments. To generate a Citrine knock-in line, the genomic sequence corresponding to intron 31-exon
531 35 and the 3' flanking region of Pp*BLD10* were cloned into pCTRN-NPTII-2 plasmid (AB697058).
532 The Citrine-fusion plasmid was amplified by PCR using the primers listed in Supplementary Table S5
533 and transformed into protoplasts. Transformation for the generation of both loss-of-function mutants
534 and Citrine knock-in lines were performed as previously described (Nishiyama et al. 2000). The
535 Citrine knock-in lines were screened using DNA gel blot analysis to confirm single integrations.
536 Genomic DNA (2 μg) was digested with EcoT22I, electrophoresed on 0.8% (w/v) SeaKem GTG
537 agarose (Lonza), and transferred onto a Hybond N+ nylon membrane (GE Healthcare). Probe
538 labelling, hybridization, and detection were performed using an AlkPhos direct labelling and detection
539 system with CDP-Star (GE Healthcare). A PCR-amplified fragment of the 3' untranslated region of

540 Pp*BLD10* was used as a DNA probe.
541
542 **Microscopy**
543 To observe Marchantia spermatids, antheridia were fixed for 60 min with 4% (w/v) paraformaldehyde
544 (PFA) in PME buffer (50 mM PIPES-KOH, 5 mM EGTA, and 1 mM $MgSO_4$ [pH 6.8]), and treated
545 for 30 min with cell wall digestion buffer (1% [w/v] cellulase, 0.25% [w/v] pectolyase Y-23, 1% [w/v]
546 BSA, 0.1% [w/v] NP-40, 1% glucose, and 1× cOmplete^{TM} EDTA-free protease inhibitor cocktail
547 [Roche Applied Science] in PME buffer). The samples were placed on a glass slide and then covered
548 with a cover slip in PBS containing 0.1% (v/v) Hoechst33342 (Dojindo). For immunostaining of
549 Marchantia spermatids, antheridia were fixed for 90 min with 4% (w/v) PFA in PME buffer and
550 treated for 30 min with cell wall digestion buffer. Cells were then treated with permeabilization buffer
551 (0.01% [v/v] Triton X-100 and 1% [w/v] BSA in PME buffer) for 10 min. After washing with PME
552 buffer three times, cells were placed on a MAS-coated glass slide (Matsunami) and incubated for 30
553 min at room temperature with blocking solution (1% [w/v] BSA in PBS buffer). After removal of
554 blocking solution, the cells were incubated with primary antibody in PBS buffer at 4°C overnight.
555 After washing with PBS buffer three times, the samples were incubated for 60 min at 37°C with the
556 secondary antibody and 0.1% (v/v) Hoechst33342 in PBS buffer. After washing with PBS buffer three
557 times, the slides were mounted using ProLong Diamond Antifade reagent (Thermo Fisher Scientific).
558 Samples were observed under a confocal microscope (LSM780, Carl Zeiss) with an oil immersion
559 lens (×63). For immunostaining of Physcomitrella spermatids, the same method was used with
560 modifications. The duration of cell wall digesting buffer treatment was changed to 40 min. Samples
561 were observed under a confocal microscope (LSM880, Carl Zeiss) with an oil immersion lens (×63).
562 For observation of Marchantia spermatozoids, freshly prepared spermatozoids in distilled
563 water were observed under a dark-field microscope (Olympus) equipped with an ORCA-Flash4.0 V2
564 camera (Hamamatsu Photonics). Physcomitrella spermatozoids were extracted by pressing the
565 antheridia between a glass slide and a cover slip with a 0.03-mm spacer (Koshimizu et al. 2018) and
566 observed under a dark-field microscope (BS-2040T, BioTools) equipped with a Michrome 5Pro
567 camera (BioTools).
568 The obtained images were processed using ImageJ (National Institutes of Health) and
569 Photoshop (Adobe Systems) software.
570
571 **Antibodies**
572 The monoclonal antibody against acetylated tubulin was purchased from Sigma-Aldrich (T7451) and
573 used at 1/10000 dilution for immunostaining. The polyclonal antibody against centrin was described
574 previously (Higo et al. 2018) and used at 1/5000 dilution for immunostaining. Alexa Fluor 488 plus
575 goat anti-mouse IgG, Alexa Fluor 594 plus goat anti-rabbit IgG, and Alexa Fluor 680 goat anti-mouse
576 IgG were purchased from Thermo Fisher Scientific and used at 1/1000 dilution for immunostaining.

577

578 **Transmission electron microscopy**

579 To observe spermatids, antheridia of Marchantia Tak-1, the Mp*bld10-1* mutant, Physcomitrella wild
580 type, and Pp*bld10-30* were collected and fixed with 2% PFA and 2% glutaraldehyde (GA) in 0.05 M
581 cacodylate buffer (pH 7.4) at 4ºC overnight. The fixed samples were washed three times with 0.05 M
582 cacodylate buffer for 30 min each and then post-fixed with 2% osmium tetroxide in 0.05 M
583 cacodylate buffer at 4ºC for 3 h. The samples were dehydrated in graded ethanol solutions (50% and
584 70% ethanol for 30 min each at 4ºC, 90% for 30 min at room temperature, four times with 100%
585 ethanol for 30 min at room temperature, and 100% ethanol overnight at room temperature). The
586 samples were infiltrated with propylene oxide (PO) twice for 30 min each and placed into a 50:50
587 mixture of PO and resin (Quetol-651; Nisshin EM Co.) for 3 h. The samples were transferred to 100%
588 resin and polymerized at 60ºC for 48 h. To observe mature Marchantia spermatozoids, spermatozoids
589 of Tak-1 and the Mp*bld10-1* mutant were collected in water, centrifuged at 5000×*g* for 3 min, and
590 fixed with 2% PFA and 2% GA in 0.05M cacodylate buffer (pH 7.4) at 4ºC overnight. The fixed
591 samples were washed three times with 0.05 M cacodylate buffer for 30 min each and post-fixed with
592 2% osmium tetroxide in 0.05 M cacodylate buffer at 4ºC for 2 h. The samples were dehydrated in
593 graded ethanol solutions (50% and 70% ethanol for 20 min each at 4ºC, 90% for 20 min at room
594 temperature, and four times with 100% ethanol for 20 min at room temperature). The samples were
595 infiltrated with PO twice for 30 min each and placed into a 70:30 mixture of PO and resin (Quetol-
596 651; Nisshin EM Co.) for 1 h, then the tube cap was opened, and PO was allowed to volatilize
597 overnight. The samples were transferred to 100% resin and polymerized at 60ºC for 48 h. The
598 polymerized resins were ultra-thin-sectioned at 70 nm with a diamond knife using an ultramicrotome
599 (Ultracut UCT; Leica) and the sections were mounted on copper grids. The sections were stained with
600 2% uranyl acetate for 15 min at room temperature, washed with distilled water, and secondary-stained
601 with lead stain solution (Sigma-Aldrich) for 3 min at room temperature. The grids were observed
602 under a transmission electron microscope (JEM-1400Plus; JEOL Ltd.) at an acceleration voltage of
603 100 kV. Digital images (3296×2472 pixels) were acquired using a CCD camera (EM-14830RUBY2;
604 JEOL Ltd.).

605

606

607 **Acknowledgements**

608 We thank N. Kawakami (Meiji Univ.), K. Yoshimoto (Meiji Univ.), and H. Kaku (Meiji Univ.) for
609 support in environment of experiments, F. Nogué (INRA), Y. Horiuchi (NIBB), and M. Hasebe
610 (NIBB) for providing plasmids for genome editing of Physcomitrella. We also thank T. Kimura
611 (Hokkaido Univ.) and M. Shimamura (Hiroshima Univ.) for anti-centrin antibody and H. Mano
612 (NIBB) for lending electron microscopy instrumentation and for general advice. Computations were
613 partially performed on the NIG supercomputer at the ROIS National Institute of Genetics and the Data

614 Integration and Analysis Facility at the National Institute for Basic Biology. Electron microscopy
615 analyses were partly supported by the EM facility of the National Institute for Physiological Sciences.
616 Rooms for cultivating Marchantia were provided by the Model Plant Research Facility, NIBB
617 Bioresource Center. This work was supported by JSPS KAKENHI grants to K.Y. (19H04870), K.E.
618 (19H04872), T.U. (19H05675 and 21H02515), N.M. (20K15824), T.N (15H04413 and 19K22448)
619 and K.S. (18K06367).
620

621 **Author contributions**

622 S.K. and K.Y. performed the computational analyses. N.M., K.E., and T.U. conducted the functional
623 analyses in Marchantia. S.K., E.Y., and K.S. conducted the functional analyses in Physcomitrella. S.K.
624 and T.N. performed sequence and phylogenetic analyses. All authors analyzed the data and
625 participated in writing the manuscript.
626

627 **Figure legends**

628 Fig. 1. Mp*bld10* mutants exhibit severe defects in spermatozoid formation. (A) Schematic structure of
629 the Mp*BLD10* (Mapoly0001s0460) gene. Nucleotide and amino acid sequences around mutation sites
630 in wild-type (WT) and Mp*bld10-1* are aligned. The target and PAM sequences are indicated by
631 underlining and bold font, respectively. (B-E) Maximum-intensity projection images of spermatozoids
632 of wild type (B), Mp*bld10-1* (C and D), and Mp*bld10-2* (E) stained with Hoechst33342. Scale bars =
633 10 μm (F-I) Maximum-intensity projection images of spermatids immunostained with anti-centrin and
634 anti-acetylated tubulin antibodies. Nuclei were visualized using Hoechst33342. Blue, green, and
635 magenta pseudo colors indicate Hoechst33342, Alexa 488, and Alexa 594, respectively. Scale bars = 5
636 μm.
637

638 Fig. 2. Transmission electron microscopy (TEM) of spermatids and spermatozoids in wild type and
639 Mp*bld10-1*. (A-H) TEM images in spermatids of wild type (A-D) and Mp*bld10-1* (E-H). Axonemes in
640 flagella (A and E), multilayered structures (B and F), and basal bodies (C, D, G, and H) are shown. (I
641 and J) TEM images of nuclei in spermatozoids of wild type (I) and Mp*bld10-1* (J). Scale bars = 200
642 nm.
643

644 Fig. 3. Subcellular localization of the MpBLD10 protein. (A-F) Differential interference contrast
645 microscopy (DIC) and maximum-intensity projection images of spermatids and spermatozoids
646 expressing mCitrine-MpBLD10 (green) driven by its own promoter at stage 0 (A), stage 1 (B), stage 2
647 (C), stage 3 (D), stage 4 (E), and stage 5 (F); nuclei were stained with Hoechst33342 (blue).
648 Developmental stages were classified according to Minamino et al. (2021). (G) Maximum-intensity
649 projection images of a spermatid expressing mCitrine-MpBLD10 (green) immunostained with anti-
650 centrin (magenta) and anti-acetylated tubulin (yellow) antibodies. The nucleus was stained with

651 Hoechst33342 (blue). Scale bars = 5 μm.
652
653 Fig. 4. Comparisons of BLD10/CEP135 family protein sequences. (A) Phylogenetic tree of
654 BLD10/CEP135 family proteins. Out-groups are Tetrahymena, human, and insects. Salmon, yellow,
655 green, blue, purple, and gray background colors indicate gymnosperms, monilophytes, lycophytes,
656 bryophytes, streptophyte algae, and chlorophytes, respectively. Branch lengths are proportional to the
657 estimated number of amino acid substitutions/site (scale upper left). (B) Architecture of the
658 MpBLD10 sequence (gaps removed) with conservation level among BLD10 proteins in the
659 streptophytes in (A). Three conserved regions are indicated. Values of conservation levels were
660 obtained using Jalview software.
661
662 Fig. 5. Existence of BLD10 proteins, flagella (basal body), and centriole during cell division in the
663 plant species. Orange dots show presumed flagella loss. Groups indicated by gray color were not
664 investigated due to insufficient sequence data. For phylogenetic relationships and presence of
665 centrioles, refer to Puttick et al. (2018) and Buschmann and Zachgo (2016), respectively.
666
667 **References**
668 Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped
669 BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*
670 25:3389–3402.
671 Banks JA, Nishiyama T, Hasebe M, Bowman JL, Gribskov M, DePamphilis C, Albert VA, Aono N,
672 Aoyama T, Ambrose BA, et al. 2011. The Selaginella genome identifies genetic changes associated
673 with the evolution of vascular plants. *Science* 332:960–963.
674 Bayless BA, Giddings TH, Winey M, Pearson CG. 2012. Bld10/Cep135 stabilizes basal bodies to
675 resist cilia-generated forces. *Mol. Biol. Cell* 23:4820–4832.
676 Bowman JL, Kohchi T, Yamato KT, Jenkins J, Shu S, Ishizaki K, Yamaoka S, Nishihama R,
677 Nakamura Y, Berger F, et al. 2017. Insights into land plant evolution garnered from the *Marchantia*
678 *polymorpha* genome. *Cell* 171:287-304.e15.
679 Brownfield L, Hafidh S, Durbarry A, Khatab H, Sidorova A, Doerner P, Twell D. 2009. *Arabidopsis*
680 DUO POLLEN3 is a key regulator of male germline development and embryogenesis. *Plant Cell*
681 21:1940–1956.
682 Buels R, Yao E, Diesh CM, Hayes RD, Munoz-Torres M, Helt G, Goodstein DM, Elsik CG, Lewis
683 SE, Stein L, et al. 2016. JBrowse: a dynamic web platform for genome visualization and analysis.
684 *Genome Biol.* 17:66.
685 Buschmann H, Zachgo S. 2016. The evolution of cell division: from streptophyte algae to land plants.
686 *Trends Plant Sci.* 21:872–883.
687 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+:

688 architecture and applications. *BMC Bioinformatics* 10:421.
689 Carothers ZB, Duckett JG. 1980. The bryophyte spermatozoid: a source of new phylogenetic
690 information. *Bull. Torrey Bot. Club* 107:281.
691 Carothers ZB, Kreitner GL. 1968. Studies of spermatogenesis in the Hepaticae. II. Blepharoplast
692 structure in the spermatid of Marchantia. *J. Cell Biol.* 36:603–616.
693 Carvalho-Santos Z, Machado P, Branco P, Tavares-cadete F, Rodrigues-martins A, Pereira-leal JB,
694 Bettencourt-dias M. 2010. Stepwise evolution of the centriole-assembly pathway. *J. Cell Sci.*
695 123:1414–1426.
696 Chiyoda S, Ishizaki K, Kataoka H, Yamato KT, Kohchi T. 2008. Direct transformation of the liverwort
697 *Marchantia polymorpha* L. by particle bombardment using immature thalli developing from spores.
698 *Plant Cell Rep.* 27:1467–1473.
699 Collonnier C, Guyon-Debast A, Maclot F, Mara K, Charlot F, Nogué F. 2017. Towards mastering
700 CRISPR-induced gene knock-in in plants: survey of key features and focus on the model
701 *Physcomitrella patens*. *Methods* 121–122:103–117.
702 Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy
703 SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* 10:giab008.
704 Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics.
705 *Genome Biol.* 20:238.
706 Glastad KM, Hunt BG, Yi S V., Goodisman MAD. 2011. DNA methylation in insects: on the brink of
707 the epigenomic era. *Insect Mol. Biol.* 20:553–565.
708 Gluck-Thaler E, Cerutti A, Perez-Quintero AL, Butchacas J, Roman-Reyna V, Madhavan VN,
709 Shantharaj D, Merfa M V., Pesce C, Jauneau A, et al. 2020. Repeated gain and loss of a single gene
710 modulates the evolution of vascular plant pathogen lifestyles. *Sci. Adv.* 6:4516–4529.
711 Graham LE, McBride GE. 1979. The occurrence and phylogenetic significance of a multilayered
712 structure in coleochaete spermatozoids. *Am. J. Bot.* 66:887-894.
713 Griesmann M, Chang Y, Liu X, Song Y, Haberer G, Crook MB, Billault-Penneteau B, Lauressergues
714 D, Keller J, Imanishi L, et al. 2018. Phylogenomics reveals multiple losses of nitrogen-fixing root
715 nodule symbiosis. *Science* 361:eaat1743.
716 Gu N, Tamada Y, Imai A, Palfalvi G, Kabeya Y, Shigenobu S, Ishikawa M, Angelis KJ, Chen C,
717 Hasebe M. 2020. DNA damage triggers reprogramming of differentiated cells into stem cells in
718 Physcomitrella. *Nat. Plants* 6:1098–1105.
719 Higo A, Kawashima T, Borg M, Zhao M, López-vidriero I, Sakayama H, Montgomery SA, Sekimoto
720 H, Hackenberg D, Shimamura M, et al. 2018. Transcription factor DUO1 generated by neo-
721 functionalization is associated with evolution of sperm differentiation in plants. *Nat. Commun.* 9:1–
722 13.
723 Higo A, Niwa M, Yamato KT, Yamada L, Sawada H, Sakamoto T, Kurata T, Shirakawa M, Endo M,
724 Shigenobu S, et al. 2016. Transcriptional framework of male gametogenesis in the liverwort

Marchantia polymorpha L. *Plant Cell Physiol.* 57:325–338.
Hiraki M, Nakazawa Y, Kamiya R. 2007. Report Bld10p constitutes the cartwheel-spoke tip and stabilizes the 9-fold symmetry of the centriole. *Curr. Biol.* 17:1778–1783.
Hodges ME, Wickstead B, Gull K, Langdale JA. 2012. The evolution of land plant cilia. *New Phytol.* 195:526–540.
Hotta T, Kong Z, Ho CMK, Zeng CJT, Horio T, Fong S, Vuong T, Lee YRJ, Liu B. 2012. Characterization of the *Arabidopsis* augmin complex uncovers its critical function in the assembly of the acentrosomal spindle and phragmoplast microtubule arrays. *Plant Cell* 24:1494–1509.
Ishizaki K, Chiyoda S, Yamato KT, Kohchi T. 2008. Agrobacterium-mediated transformation of the haploid liverwort *Marchantia polymorpha* L., an emerging model for plant biology. *Plant Cell Physiol.* 49:1084–1091.
Ishizaki K, Nishihama R, Ueda M, Inoue K, Ishida S, Nishimura Y, Shikanai T, Kohchi T. 2015. Development of gateway binary vector series with four different selection markers for the liverwort *Marchantia polymorpha*. *PLoS One* 10:e0138876.
Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* 8:275–282.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30:772–780.
Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37:907–915.
Kleylein-Sohn J, Westendorf J, Clech M Le, Habedanck R, Stierhof Y, Nigg EA. 2007. Plk4-Induced centriole biogenesis in human cells. *Dev. Cell* 13:190–202.
Koshimizu S, Kofuji R, Sasaki-Sekimoto Y, Kikkawa M, Shimojima M, Ohta H, Shigenobu S, Kabeya Y, Hiwatashi Y, Tamada Y, et al. 2018. *Physcomitrella* MADS-box genes regulate water supply and sperm movement for fertilization. *Nat. Plants* 4:36–45.
Kreitner GL, Carothers ZB. 1976. Studies of spermatogenesis in the hepaticae V. Blepharoplast development in *Marchantia polymorpha*. *Am. J. Bot.* 63:545.
Kubota A, Ishizaki K, Hosaka M, Kohchi T. 2013. Efficient Agrobacterium-mediated transformation of the liverwort *Marchantia polymorpha* using regenerating thalli. *Biosci. Biotechnol. Biochem.* 77:167–172.
Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35:1547–1549.
Larkin A, Marygold SJ, Antonazzo G, Attrill H, dos Santos G, Garapati P V., Goodman JL, Sian Gramates L, Millburn G, Strelets VB, et al. 2021. FlyBase: updates to the *Drosophila melanogaster* knowledge base. *Nucleic Acids Res.* 49:D899–D907.
Leebens-Mack JH, Barker MS, Carpenter EJ, Deyholos MK, Gitzendanner MA, Graham SW, Grosse I, Li Z, Melkonian M, Mirarab S, et al. 2019. One thousand plant transcriptomes and the

762 phylogenomics of green plants. *Nature* 574:679–685.
763 Leinonen R, Sugawara H, Shumway M. 2011. The Sequence Read Archive. Nucleic Acids Res.
764 39:D19–D21.
765 Li FW, Brouwer P, Carretero-Paulet L, Cheng S, De Vries J, Delaux PM, Eily A, Koppers N, Kuo LY,
766 Li Z, et al. 2018. Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nat.*
767 *Plants* 4:460–472.
768 Li FW, Nishiyama T, Waller M, Frangedakis E, Keller J, Li Z, Fernandez-Pozo N, Barker MS,
769 Bennett T, Blázquez MA, et al. 2020. *Anthoceros* genomes illuminate the origin of land plants and the
770 unique biology of hornworts. *Nat. Plants* 6:259–272.
771 Lin JJ, Wang FY, Li WH, Wang TY. 2017. The rises and falls of opsin genes in 59 ray-finned fish
772 genomes and their implications for environmental adaptation. *Sci. Rep.* 7:1–13.
773 Lin YC, Chang C, Hsu W, Tang CC, Lin Yi-nan, Chou E, Wu C, Tang TK. 2013. Human
774 microcephaly protein CEP135 binds to hSAS-6 and CPAP, and is required for centriole assembly.
775 *EMBO J.* 32:1141–1154.
776 Livingstone CD, Barton GJ. 1993. Protein sequence alignments: a strategy for the hierarchical
777 analysis of residue conservation. *Bioinformatics* 9:745–756.
778 Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler
779 GH, Song JS, et al. 2020. CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids
780 Res. 48:D265–D268.
781 Matsuura K, Lefebvre PA, Kamiya R, Hirono M. 2004. Bld10p, a novel protein essential for basal
782 body assembly in *Chlamydomonas*: localization to the cartwheel, the first ninefold symmetrical
783 structure appearing during assembly. *J. Cell Biol.* 165:663–671.
784 Minamino N, Norizuki T, Mano S, Ebine K, Ueda T. 2021. Remodeling of organelles and
785 microtubules during spermiogenesis in the liverwort *Marchantia polymorpha*.
786 bioRxiv:2021.07.10.451882
787 Mottier-Pavie V, Megraw TL. 2009. *Drosophila* Bld10 is a centriolar protein that regulates centriole,
788 basal body, and motile cilium assembly. *Mol. Biol. Cell* 20:2605–2614.
789 Naito Y, Hino K, Bono H, Ui-Tei K. 2015. CRISPRdirect: software for designing CRISPR/Cas guide
790 RNA with reduced off-target sites. *Bioinformatics* 31:1120–1123.
791 NCBI Resource Coordinators. 2018. Database resources of the National Center for Biotechnology
792 Information. *Nucleic Acids Res.* 46:D8–D13.
793 Nishiyama T, Hiwatashi Y, Sakakibara K, Kato M, Hasebe M. 2000. Tagged mutagenesis and gene-
794 trap in the moss, *Physcomitrella patens* by shuttle mutagenesis. *DNA Res.* 7:9–18.
795 Nishiyama T, Sakayama H, de Vries J, Buschmann H, Saint-Marcoux D, Ullrich KK, Haas FB,
796 Vanderstraeten L, Becker D, Lang D, et al. 2018. The *Chara* genome: secondary complexity and
797 implications for plant terrestrialization. *Cell* 174:448-464.e24.
798 Norstog K. 1967. Fine structure of the spermatozoid of Zamia with special reference to the flagellar

799 apparatus. *Am. J. Bot.* 54:831.
800 Puttick MN, Morris JL, Williams TA, Cox CJ, Edwards D, Kenrick P, Pressel S, Wellman CH,
801 Schneider H, Pisani D, et al. 2018. The interrelationships of land plants and the nature of the ancestral
802 embryophyte. *Curr. Biol.* 28:733-745.e2.
803 Rensing SA, Goffinet B, Meyberg R, Wu SZ, Bezanilla M. 2020. The moss *Physcomitrium*
804 (*Physcomitrella*) *patens*: a model organism for non-seed plants. *Plant Cell* 32:1361–1376.
805 Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud P-F,
806 Lindquist EA, Kamisugi Y, et al. 2008. The *Physcomitrella* genome reveals evolutionary insights into
807 the conquest of land by plants. *Science* 319:64–69.
808 Renzaglia KS, Carothers ZB, Duckett JG. 1985. Comparative ultrastructural studies of
809 spermatogenesis in the Metzgeriales (Hepaticae). I. The blepharoplast of Pallavicinia lyellii. Am. J.
810 Bot. 72:588–595.
811 Renzaglia KS, Duckett JG. 1987. Spermatogenesis in *Blasia pusilla* : from young antheridium through
812 mature spermatozoid. *Bryologist* 90:419–449.
813 Renzaglia KS, Garbary DJ. 2001. Motile gametes of land plants: diversity, development, and
814 evolution. *CRC. Crit. Rev. Plant Sci.* 20:107–213.
815 Sakakibara K, Nishiyama T, Deguchi H, Hasebe M. 2008. Class 1 KNOX genes are not involved in
816 shoot development in the moss *Physcomitrella patens* but do function in sporophyte development.
817 *Evol. Dev.* 10:555–566.
818 Schaefer DG, Delacote F, Charlot F, Vrielynck N, Guyon-Debast A, Le Guin S, Neuhaus JM,
819 Doutriaux MP, Nogué F. 2010. RAD51 loss of function abolishes gene targeting and de-represses
820 illegitimate integration in the moss *Physcomitrella patens*. *DNA Repair (Amst).* 9:526–533.
821 Schaefer DG, Zrÿd JP. 1997. Efficient gene targeting in the moss *Physcomitrella patens*. *Plant J.*
822 11:1195–1206.
823 Sharma V, Hecker N, Roscito JG, Foerster L, Langer BE, Hiller M. 2018. A genomics approach
824 reveals insights into the importance of gene losses for mammalian adaptations. *Nat. Commun.* 9:1–9.
825 Shimamura M, Brown RC, Lemmon BE, Akashi T, Mizuno K, Nishihara N, Tomizawa KI, Yoshimoto
826 K, Deguchi H, Hosoya H, et al. 2004. γ-tubulin in basal land plants: characterization, localization, and
827 implication in the evolution of acentriolar microtubule organizing centers. *Plant Cell* 16:45–59.
828 Simpson MG. 2018. Evolution and diversity of green and land plants. In: Plant systematics (3rd
829 edition). Amsterdam: Elsevier Science. pp. 55–74.
830 Spratt NT. 1971. Developmental biology. Belmont: Wadsworth Publishing Company
831 Stover NA, Krieger CJ, Binkley G, Dong Q, Fisk DG, Nash R, Sethuraman A, Weng S, Cherry JM.
832 2006. Tetrahymena Genome Database (TGD): a new genomic resource for *Tetrahymena thermophila*
833 research. *Nucleic Acids Res.* 34:D500–D503.
834 Sugano SS, Nishihama R, Shirakawa M, Takagi J, Matsuda Y, Ishida S, Shimada T, Hara-Nishimura I,
835 Osakabe K, Kohchi T. 2018. Efficient CRISPR/Cas9-based genome editing and its application to

conditional genetic analysis in *Marchantia polymorpha*. *PLoS One* 13:e0205117.

Tajima F. 1993. Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* 135:599–607.

Vaughn KC, Renzaglia KS. 1998. Origin of bicentrioles in Anthocerote spermatogenous cells. In: Bryology for the Twenty-first Century. London: Routledge. pp. 189–203.

Wang S, Li L, Li H, Sahu SK, Wang H, Xu Y, Xian W, Song B, Liang H, Cheng S, et al. 2020. Genomes of early-diverging streptophyte algae shed light on plant terrestrialization. *Nat. Plants* 6:95–106.

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25:1189–1191.

Wickett NJ, Mirarab S, Nguyen N, Warnow T, Carpenter E, Matasci N, Ayyampalayam S, Barker MS, Burleigh JG, Gitzendanner MA, et al. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc. Natl. Acad. Sci. USA.* 111:E4859–E4868.

Wingfield J, Lechtreck K-F. 2018. *Chlamydomonas* basal bodies as flagella organizing centers. *Cells* 7:79.

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24:1586–1591.

Yates AD, Achuthan P, Akanni W, Allen James, Allen Jamie, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R, et al. 2020. Ensembl 2020. *Nucleic Acids Res.* 48:D682–D688.

Zhang R, Guo C, Zhang W, Wang P, Li L, Duan X, Du Q, Zhao L, Shan H, Hodges SA, et al. 2013. Disruption of the petal identity gene APETALA3-3 is highly correlated with loss of petals within the buttercup family (Ranunculaceae). *Proc. Natl. Acad. Sci. USA.* 110:5074–5079.

Zhao YP, Fan G, Yin PP, Sun S, Li N, Hong X, Hu G, Zhang H, Zhang FM, Han JD, et al. 2019. Resequencing 545 ginkgo genomes across the world reveals the evolutionary history of the living fossil. *Nat. Commun.* 10:1–10.

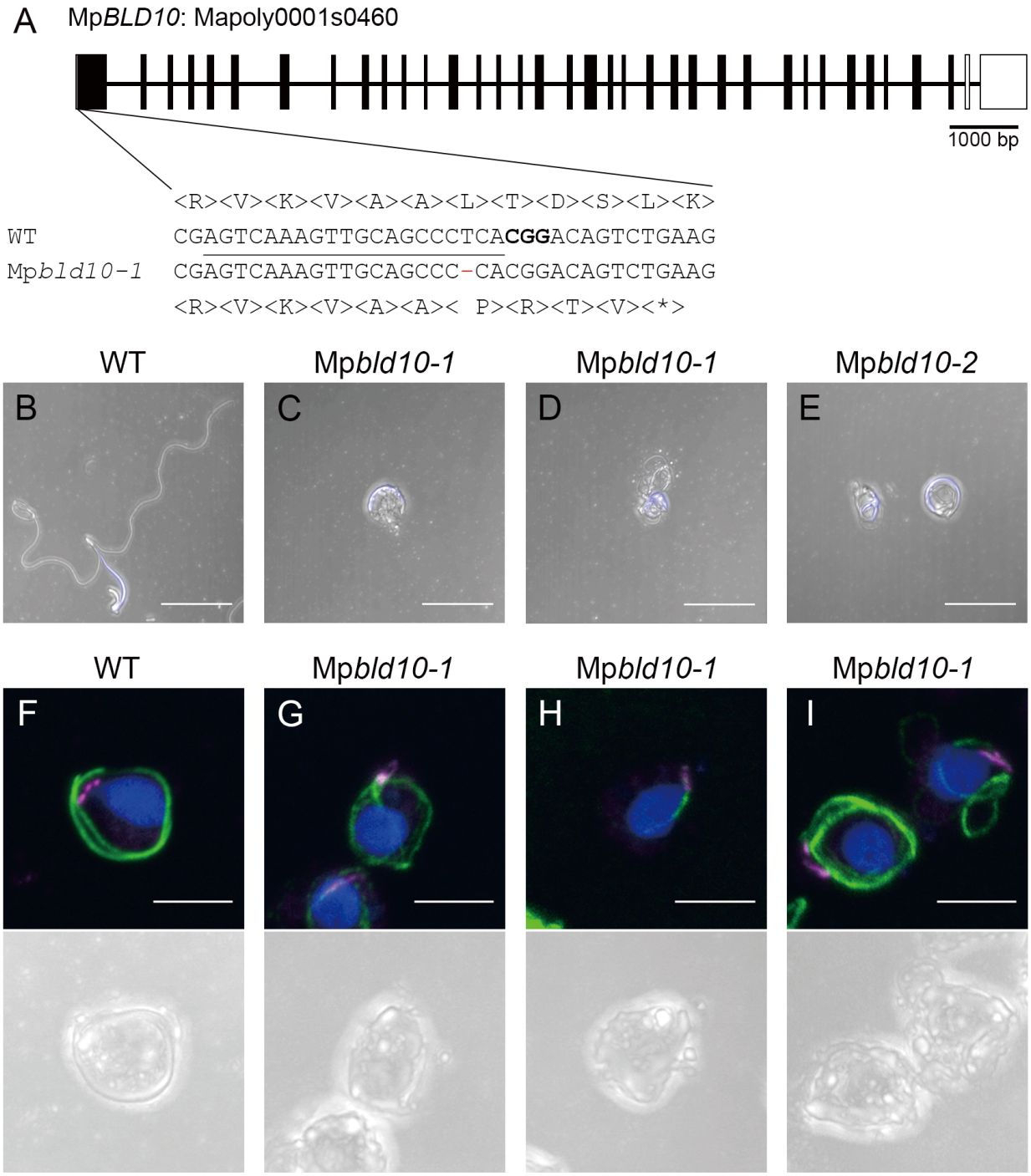


Fig. 1. Mp*bld10* mutants exhibit severe defects in spermatozoid formation.

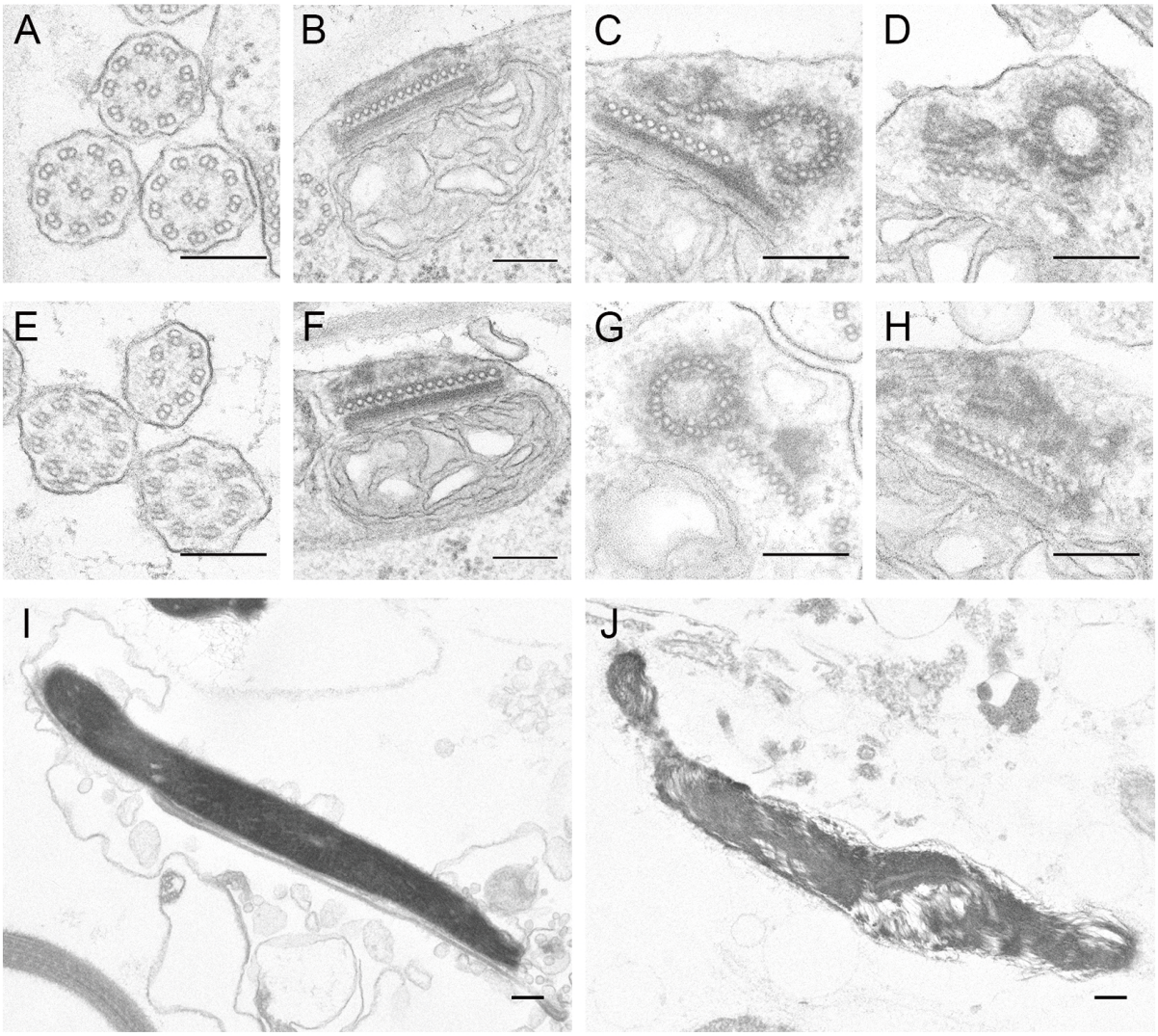


Fig. 2. Transmission electron microscopy (TEM) of spermatids and spermatozoids in wild type and Mp*bld10-1*.

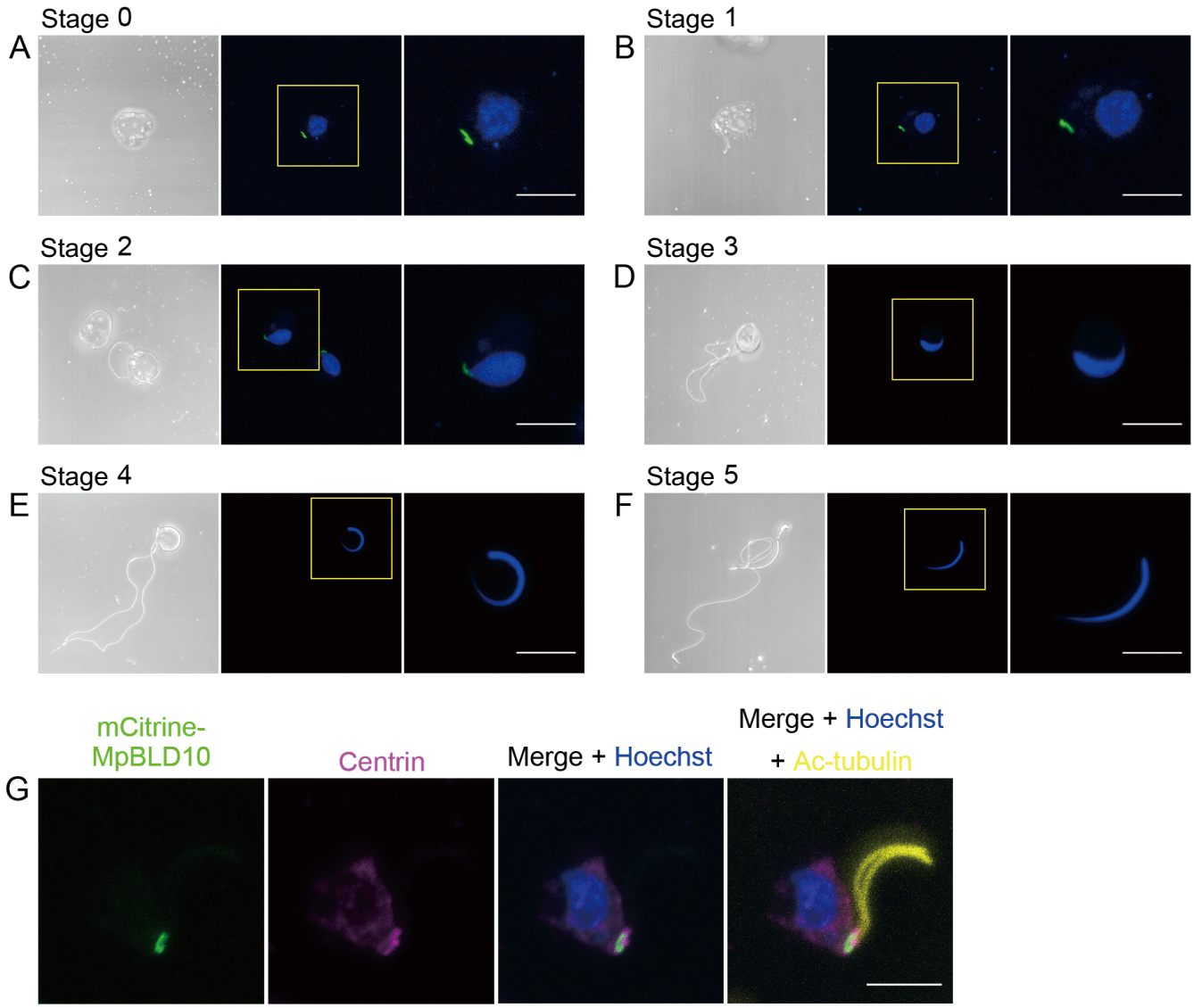


Fig 3. Subcellular localization of the MpBLD10 protein.

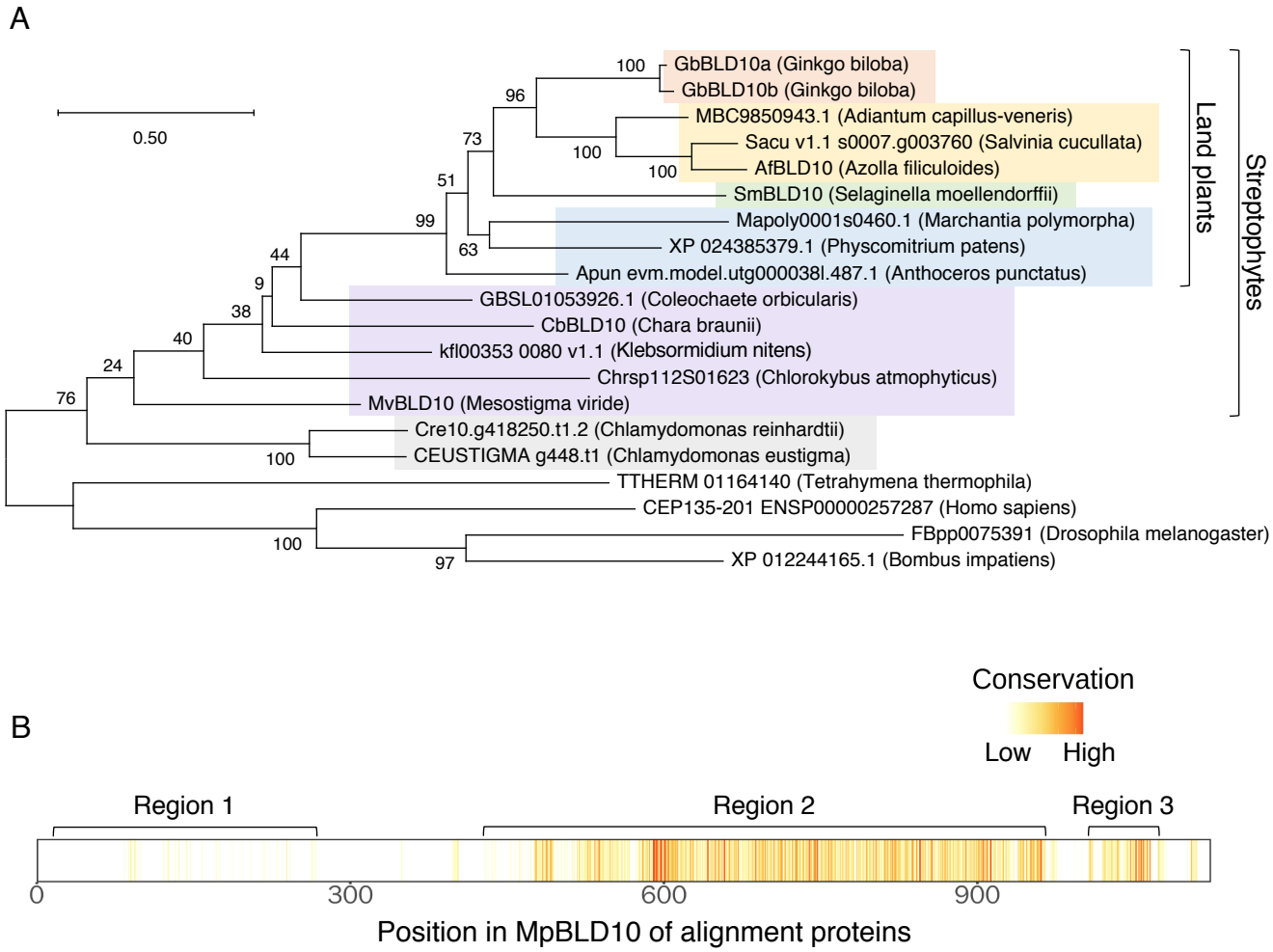
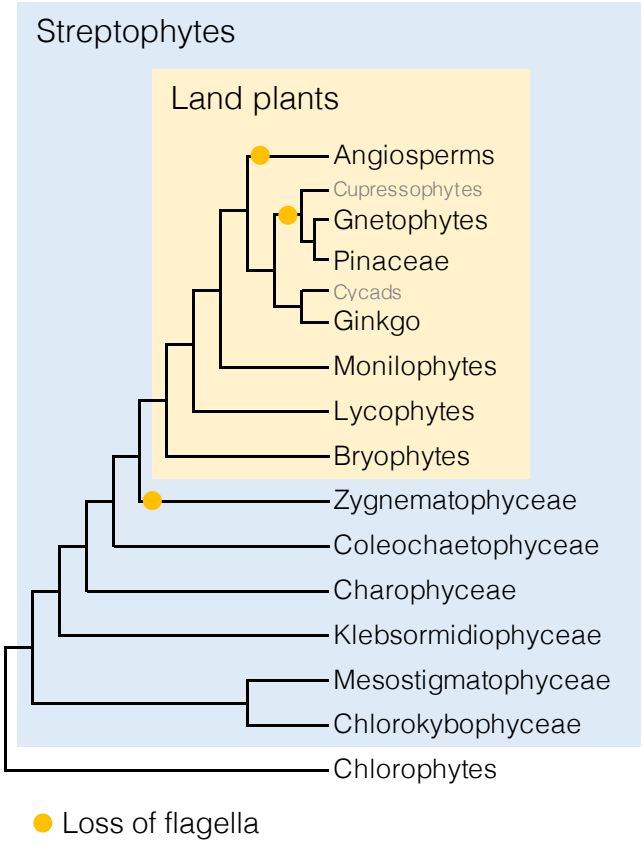


Fig. 4. Sequence comparisons of BLD10/CEP135 family proteins.



BLD10	Flagella (basal body)	Centriole
Absent	Absent	Absent
Absent	Absent	Absent
Absent	Absent	Absent
Present	Spermatozoid	Absent
Present	Spermatozoid	Absent
Present	Spermatozoid	Absent
Present	Spermatozoid	Absent
Absent	Absent	Absent
Present	Spermatozoid	Present
Present	Spermatozoid	Absent
Present	Zoospore	Present
Present	Motile cell	Present
Present	Zoospore	Present
Present	Motile cell	Present

Fig. 5. Existence of BLD10 proteins, flagella, and basal body/centriole in the plant species.