A transcription factor module regulating stomatal formation is co-opted for the development of setae in the astomatous liverwort *Marchantia polymorpha*

Authors
Kenta C. Moriya[1], Makoto Shirakawa[2], Jeanne Loue-Manifel[3,4], Yoriko Matsuda[5], Kentaro Tamura[6], Yoshito Oka[1], Tomonao Matsushita[1], Ikuko Hara-Nishimura[7], Gwyneth Ingram[3], Ryuichi Nishihama[5,8], Justin Goodrich[4], Takayuki Kohchi[5], Tomoo Shimada[1,*]

Affiliations
[1]Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan
[2]Graduate School of Biological Sciences, Nara Institute of Science and Technology (NAIST), Ikoma 630-0192, Japan
[3]Laboratoire Reproduction et Développement des Plantes, University of Lyon, ENS de Lyon, UCB Lyon 1, CNRS, INRAE, Lyon 69342, France
[4]Institute of Molecular Plant Sciences, University of Edinburgh, Daniel Rutherford Building, Max Born Crescent, Edinburgh EH9 3BF, UK
[5]Graduate School of Biostudies, Kyoto University, Kyoto 606-8502, Japan
[6]School of Food and Nutritional Sciences, University of Shizuoka, Shizuoka 422-8526, Japan
[7]Faculty of Science and Engineering, Konan University, Kobe 658-8501, Japan
[8]Department of Applied Biological Science, Faculty of Science and Technology, Tokyo University of Science, Noda 278-8510, Japan

*Correspondence: tshimada@gr.bot.kyoto-u.ac.jp

Abstract
The evolution of special types of cells requires the acquisition of new gene regulatory networks controlled by transcription factors (TFs). In stomatous plants, the stomatal formation is regulated by a TF module formed by subfamilies Ia and IIIb basic helix-loop-helix TFs (Ia-IIIb bHLH); however, how this module evolved during land plant diversification remains obscure. Here, we show that, in the astomatous liverwort *Marchantia polymorpha*, a Ia-IIIb bHLH module regulates the development of a unique sporophyte tissue, the setae, which are found in mosses and liverworts. The sole Ia bHLH gene, Mp*SETA*, and a IIIb bHLH gene, Mp*ICE2*, regulate cell division and/or differentiation of seta lineage cells. MpSETA can partially replace the stomatal function of Ia bHLH TFs in *Arabidopsis thaliana*, suggesting the common regulatory mechanism underlying setal and stomatal formation. Our findings reveal the co-option of a Ia-IIIb bHLH TF module for regulating the cell fate determination and/or cell division of distinct types of cells during the evolution of land plant.

Land plants develop unique types of cells and tissues to adapt to the terrestrial environment during evolution[1,2]. The acquisition of novel cells or tissues leading to complex body plans is related to the diversification of transcription factors (TFs)[2–4]. Basic helix-loop-helix (bHLH) TFs are TF superfamilies that play key roles in cell fate determination and cell division during the development of eukaryotes. In land plants, the number of genes encoding bHLH TFs has increased compared with those of chlorophyte and charophyte algae, suggesting that bHLH TFs are involved in terrestrial adaptation[5]. For example, the acquisition of stomata, a special tissue for gas exchange on the epidermis, is an important adaptation of plants to terrestrial environments; previous studies have revealed that stomatal formation is regulated by bHLH TFs as master TFs[6–9]. In *Arabidopsis thaliana*, three TFs belonging to subfamily Ia (SPEECHLESS [SPCH], MUTE, and FAMA)[6–8] form heterodimers with subfamily IIIb TFs (ICE1, also known as SCREAM [SCRM], and ICE2, also known as SCRM2) to promote stomatal formation by regulating downstream gene expression[9]. The molecular mechanism of stomatal formation by Ia and IIIb bHLH TFs is conserved in the moss *Physcomitrium patens*, in which stomata play an important role in spore dispersal by promoting dehydration and dehiscence of the sporangium[10].

Recent studies revealed that a gene encoding Ia bHLH is present in the genome of the astomatous liverwort *Marchantia polymorpha*[4,11–13]. Despite the importance of Ia bHLH in stomatal development, its function in plants lacking stomata has not yet been explored. Here, we report that Ia bHLH protein, designated as MpSETA, is a master regulator of the setal formation, which is a diploid tissue involved in long-distance spore dispersal in *M. polymorpha*. Furthermore, we show that Ia and IIIb bHLH positively regulate setal formation by heterodimerization, similar to the stomatal formation in other land plants. This outcome not only advances our understanding of the mechanisms of the evolution of plant tissue formation but also provides new insights into the co-option of gene expression regulatory networks (GRNs).

Results

MpSETA is the sole Ia bHLH protein in *M. polymorpha.*

To identify Ia bHLH coding genes in *M. polymorpha*, we constructed a phylogenetic tree of Ia bHLH proteins from various plant species using the bHLH domain and a C-terminal conserved domain called SMF (also known as the ACT-like domain)[14,15] (Fig. 1a and Extended Data Fig. 1). Our phylogenetic analysis suggested that MpBHLH35 (Mp2g04160) is the sole Ia bHLH in *M. polymorpha* (Fig. 1a). We named this gene Mp*SETA* based on its specific expression in the seta tissue of the sporophyte (see below). Multiple alignments of bHLH proteins revealed that the amino acid residues predicted to be important for E-box (CANNTG) binding and bHLH dimerization are highly conserved in MpSETA, although the amino acid sequence of the bHLH domain of MpSETA is relatively diversified to other Ia bHLH (Extended Data Fig. 1b). Thus, we can conclude that a single Ia bHLH gene, Mp*SETA*, is conserved in the genome of *M. polymorpha*.

Since the amino acid sequence of MpSETA is diversified, it is unclear if MpSETA has similar properties as the other Ia bHLH proteins. Therefore, we investigated whether MpSETA can act as a stomatal regulator instead of AtSPCH, AtMUTE, or AtFAMA. In this context, Mp*SETA* was expressed under the native promoters of At*SPCH*, At*MUTE*, and At*FAMA* in *spch-3*, *mute-2*, and *fama-1*

backgrounds, respectively (Fig. 1b and Extended Data Fig. 2a,b). Even though *mute-2* results in arrested meristemoids (self-renewing stomatal precursors), a few stomata were formed in *mute-2* expressing Mp*SETA* (Fig. 1b) (2.67 ± 1.56 and 2.33 ± 1.72 per abaxial side of the cotyledon in the lines #8-4 and #10-11, respectively [mean ± s.d.; n = 12]). Notably, hydathode pores (a modified form of stomatal pores) were often found in these lines. This might be due to the high expression activity of the At*MUTE* promoter in the hydathode of the cotyledons[16]. Mp*SETA* also exhibited the potential to rescue the *fama-1*. *A. thaliana fama* mutant displays caterpillar-like stomatal-lineage cells called "*fama* tumors," where the terminal symmetric division occurs more than once[8]. In *fama-1* expressing Mp*SETA*, excess cell divisions in stomatal lineage were suppressed, although no mature stomata were formed (Extended Data Fig. 2b,c). Neither stomata nor stomatal-lineage cells were found in *spch-3* expressing Mp*SETA* (Extended Data Fig. 2a). These results showing that Mp*SETA* is partially functional in stomatal-cell division and differentiation in *A. thaliana* suggest that it can interact with AtICE1 and AtSCRM2. Therefore, we tested this ability in yeast two-hybrid assays and found that MpSETA physically interacted with AtICE1 and AtSCRM2 (Extended Data Fig. 2d). Thus, our findings indicate that MpSETA from the astomatous liverwort is a bona fide Ia bHLH TF, albeit its amino acid sequence is diversified.

Mp*SETA* is preferentially expressed in developing sporophyte.

To investigate the expression pattern of Mp*SETA* in *M. polymorpha*, we reanalyzed the public RNA-seq dataset from several organs[4,17–20] and found that Mp*SETA* was preferentially expressed in the diploid sporophyte, whereas its expression level was low in the haploid gametophyte (Fig. 2a).

We examined in detail the expression pattern of Mp*SETA* by characterizing the different stages of development in the *M. polymorpha* sporophyte (Fig. 2b,c and Extended Data Fig. 3a); in wild type, sporophytes are divided into three tissues: foot, seta, and sporangium (Fig. 2b). The foot plays a key role in nutrient transport between the gametophyte and the sporophyte, while the seta, which consists of files of elongated cells that form a stalk suspending the sporangium, plays a pivotal role in spore dispersal[21–23]. After late spore maturation in sporophyte development, the seta extends through elongation of its cells and thrusts the sporangium outwards, causing it to break through the surrounding calyptra and pseudoperianth, which are tissues derived from the archegonium to protect the sporophyte during development[21–23]. As a result, spores are dispersed after sporangium dehiscence by desiccation. Based on the unique developmental events as shown in previous studies[21,24], we divided the late development of the sporophyte into eight stages, as follows (Fig. 2c and Extended Data Fig. 3a): (I) early-globular stage with a distinct amphithecium (the outermost tissue that forms the capsule wall) and the endothecium (the inner archesporial tissue); (II) late-globular stage; (III) archesporial-tissue stage with the visible differentiated foot, seta, and archesporial tissues; (IV) sporogenous-cell stage with elaterocytes and sporogenous cells, which are the precursors of elaters and sporocytes (spore mother cells); (V) sporocyte stage; (VI) spore-tetrad stage (after meiosis); (VII) maturated stage; (VIII) seta-elongated stage. Note that seta and foot are established between stages III and VII. In stage VI, we anatomically observed proliferative (symmetric) cell divisions in the seta region (Extended Data Fig. 3b). Moreover, cell division activity was detected not in the foot but rather in the seta of sporophytes between stages V and VI using a G2-M

phase reporter line, $_{pro}$Mp*CYCB;1:Dbox-GUS*[25] (Extended Data Fig. 3c). Thus, we concluded that the cell file of seta is established by a few proliferative cell divisions of the putative "seta mother cell" in the later developmental stage.

We detected the promoter activity of Mp*SETA* in the sporophyte by generating transformants expressing the β-glucuronidase gene (*GUS*) under the control of the Mp*SETA* promoter ($_{pro}$Mp*SETA:GUS*). GUS activity was found in developing sporophytes between stages II and V, especially in seta, whereas no GUS activity was found in stages VI and VII (Fig. 2d). In gametophytic tissues, GUS signals were detected only in young antheridia (Extended Data Fig. 4). These findings suggest that Mp*SETA* is expressed early in seta development and may regulate seta cell division and/or differentiation rather than the elongation of seta cells.

Mp*seta*^{ko} mutants show defects in setal formation in the sporophyte.

We generated loss-of-function mutants of Mp*SETA* by homologous recombination-mediated gene targeting to reveal the function of Mp*SETA* *in vivo*[26]. Therefore, we obtained two independent Mp*SETA* knock-out lines (Mp*seta-1*^{ko} and Mp*seta-2*^{ko}) in which the genomic regions encoding the bHLH domain were replaced with a hygromycin-resistant gene cassette (Extended Data Fig. 5a-c). We confirmed the loss of the full-length transcripts of the Mp*SETA* gene by reverse-transcription polymerase chain reaction (RT-PCR) analysis of homozygous mutant sporophytes produced from crosses between Mp*seta-1*^{ko} or Mp*seta-2*^{ko} males and females (Extended Data Fig. 5d). Although Mp*SETA* reporter gene expression was found in the developing antheridia (Extended Data Fig. 4), no obvious phenotype was observed during sperm formation in Mp*seta*^{ko} lines (Extended Data Fig. 5e), and mutant males were fertile.

We crossed males and females of Mp*seta*^{ko} and compared the resulting sporophytes with those of the wild type to investigate phenotypes of Mp*seta*^{ko} mutants in the diploid generation. In longitudinal sections of mature sporophytes of Mp*seta*^{ko}, we found anatomical defects in the development of setal cells (Fig. 3a,b). We did not observe any elongated seta cells or cell files of seta cells in Mp*seta*^{ko} mutants. The SF/SP ratio (a ratio of the length from the foot to the proximal side of sporangium to the total sporophyte length) and the number of cells around the seta region were significantly reduced in Mp*seta*^{ko} compared with wild type (Fig. 3c,d). Detailed analysis of earlier stages of sporophyte development revealed that defects in Mp*seta-1*^{ko} setae could be found even at stage V, the earliest stage at which seta cells are unequivocally recognized (Extended Data Fig. 6a). We concluded that Mp*seta*^{ko} mutants have defects in the differentiation of seta mother cells, which prevents the induction of the subsequent proliferative cell divisions. At one-month postfertilization, the wild-type sporangia were pushed out of the calyptras. By contrast, Mp*seta*^{ko} sporangia remained buried inside the calyptras, presumably due to the defects in seta development, and hence not exposed to the outside (Fig. 3e). Since sporangia are completely wrapped with the calyptras and pseudoperianths, spore dispersal does not occur unless the sporangia are pushed out of the calyptras by elongation of seta cells[27]. In genetic complementation experiments, we generated a transgenic line having the genomic region of Mp*SETA* introduced into Mp*seta-1*^{ko} (*g*Mp*SETA* Mp*seta-1*^{ko}). The sporophytes obtained by crossing these complemented lines had normal setae, and their sporangia were pushed out of the calyptra as the wild type (Fig. 3). Despite the obvious loss of setae, spores and other sporophytic tissues were normally formed in Mp*seta-1*^{ko}

(Extended Data Fig. 6). Consequently, these results suggest that Mp*SETA* is essential for the setal formation and spore dispersal in *M. polymorpha*.

MpICE2 physically interacts with MpSETA and regulates setal development.
Since the interaction between Ia and IIIb bHLHs is evolutionarily conserved[9,10,28], we examined whether Ia bHLH and IIIb bHLH physically interact with each other in *M. polymorpha*. We performed phylogenetic analyses to find IIIb bHLH genes in *M. polymorpha* and identified two genes encoding IIIb bHLH proteins in the genome of *M. polymorpha*, both of which are orthologous to At*ICE1* and At*SCRM2* (Extended Data Fig. 7). We named these genes Mp*ICE1* (Mp4g04910) and Mp*ICE2* (Mp4g04920). The amino acid sequences of the bHLH and ACT-like domains of IIIb bHLHs were highly conserved among land plants (Extended Data Fig. 8).

Reanalyzing public RNA-seq data revealed high expression of Mp*ICE2* at 13 days postfertilization in young sporophytes (from stage I to stage II)[17], in which putative seta mother cells were dividing and differentiating, while the expression level of Mp*ICE1* was almost constant in all tissues (Fig. 4a). Thus, we assumed that MpICE2 may predominantly function in cooperation with MpSETA during seta cell formation rather than MpICE1. We expressed *Citrine* (a yellow fluorescent protein; YFP), *GUS*, and nuclear localization-signal (*NLS*) fusion gene (*Citrine-GUS-NLS*) under the control of the Mp*ICE2* promoter to confirm the tissue/cell-level expression pattern of Mp*ICE2* in the sporophyte. In this line, GUS signals were detected overall in stages II and III sporophytes, specifically in the seta and foot in stages IV-VI sporophytes, and only in the foot in mature sporophytes together with Citrine and GUS signals in gametophytic tissues (Extended Data Fig. 9).

We performed yeast two-hybrid assays using full-length MpSETA in pairwise combinations with MpICE2 to test the interaction between MpSETA and MpICE2; in this context, a physical interaction was observed (Fig. 4b). Additionally, bimolecular fluorescence complementation (BiFC) assay was used to test the interaction between these bHLHs. YFP signals were detected in the nuclei of *Nicotiana benthaminana* leaves coexpressing MpSETA-nYFP and MpICE2-cYFP (Fig. 4c). These results suggest that MpSETA has the potential to interact with MpICE2 in *M. polymorpha*.

We generated two independent genome-edited lines using the CRISPR/Cas9 system (Mp*ice2-2*^{ge} and Mp*ice2-6*^{ge}) to assess the function of Mp*ICE2*[29,30]. The first one retained two amino acid substitutions (L503H and M504L) that were predicted to be important for the DNA-binding activity of the bHLH domain[31], while the second had a frame-shift mutation that caused deletion of the C-terminal half of the bHLH domain and the whole ACT-like domain (Extended Data Fig. 10). In these Mp*ice2* mutants, the setae of sporophytes were not formed (Fig. 4d,e) and the number of cells between foot and sporangium was significantly reduced (Fig. 4f,g). Additionally, we found that the number of cells in the seta region of Mp*ice2-2*^{ge} was higher than that of Mp*ice2-6*^{ge} (Fig. 4g). This is probably due to the translational product of Mp*ICE2* with 2-amino acid substitution in the bHLH domain, which might be partially functional in sporophytes of Mp*ice2-2*^{ge} compared with the null allele Mp*ice2-6*^{ge}. In both Mp*ice2*^{ge} mutants, the sporophytes did not emerge outside of the protective organs derived from the archegonia, similar to the case in Mp*seta*^{ko} mutants (Fig. 4h). Since the single mutants of Mp*ice2*^{ge} showed almost the same phenotype as that of the Mp*seta*^{ko} mutants, Mp*ICE1* may not be functionally redundant with Mp*ICE2* at least in the setal formation.

The Mp*ice2*^{ge} phenotype in the setal region was completely suppressed by introducing the genomic region of Mp*ICE2* into Mp*ice2-6*^{ge} (Fig. 4d-h). Therefore, we can conclude that the MpSETA-MpICE2 heterodimer plays an important role in the setal development of *M. polymorpha*.

Discussion

In this article, we showed that two transcription factors, MpSETA (Ia bHLH) and MpICE2 (IIIb bHLH), play a pivotal role in controlling the formation of the diploid tissue seta in the sporophyte of *M. polymorpha*, which is an astomatous liverwort (Figs. 3 and 4). Similarly, in land plants, a module formed by the subfamilies Ia and IIIb bHLH TFs regulates GRNs in stomatal development[2]. Mp*SETA* could partially complement the defects of *A. thaliana mute* and *fama*, suggesting similar properties of Ia bHLH TFs from *M. polymorpha* and *A. thaliana*. However, Mp*SETA* was unable to complement the *spch* mutant (Fig. 1b and Extended Data Fig. 2). These results are consistent with the previous hypothesis stating that the ancestral Ia bHLH proteins had a MUTE- and FAMA-like function[32,33]. Although the nature of Mp*SETA* expressing cells is still unknown, Mp*SETA* may function as a regulator of cell differentiation and asymmetric cell division during setal formation, similar to the stomatal formation (Fig. 5a).

Although stomata and setae are completely different morphologically, they both play a common role in promoting sporangial dehiscence and spore dispersal. In mosses and hornworts, stomata are present in the epidermis of the sporangia and function in the desiccation of the sporangia by a gas exchange to promote its dehiscence and spore dispersal[23,34]. In *A. thaliana*, At*ICE1* controls stomatal development on anther epidermis and can regulate dehydration and dehiscence of anther[35]. However, the seta is the tissue that supports the sporangia and has a role through cell elongation in thrusting the sporangia outside the surrounding maternal tissues to permit long-distance dispersal of spores[22,23]. In this study, we found that the lack of setae in Mp*seta*^{ko} and Mp*ice2*^{ge} mutants prevented sporangia dehiscence and spore dispersal due to the inability of the sporangia to break through the protective organs around it (Figs. 3 and 4). Thus, the Ia-IIIb bHLH module has a common role in the development of the tissues involved in spore or pollen dispersal.

Previous research suggested that land plant evolution occurred through the reusing and/or modifying of the preexisting GRNs[18,36–41]. Here, we hypothesize that the Ia-IIIb bHLH module was primarily used in the stomatal formation and secondarily co-opted to setal formation in the common ancestor of “Setaphyta” (Fig. 5b). Since only mosses and liverworts have setae, a mosses-liverworts clade is called Setaphyta[1,42,43]. However, the process of setal development differs between mosses and liverworts; while a transient intercalary meristem, called the seta meristem, produces seta in mosses, the body plan, including seta, foot, and sporangium, is established by formative cell division at an early stage in liverworts[23]. Therefore, whether the Ia-IIIb bHLH module is involved in the formation of seta in mosses, such as *P. patens*[44,45], should be confirmed.

In *A. thaliana*, the AtFAMA-AtICE1 heterodimer regulates not only the development of stomata but also of myrosin idioblasts, which are adjacent to vascular tissues and contribute to defense against herbivores[39,46,47]. Since myrosin cells are present only in the Brassicales, the AtFAMA-AtICE1 module for stomatal formation was co-opted to myrosin cells during the evolution of Brassicales (Fig. 5b). The

AtFAMA-AtICE1 module is thought to regulate the expression of different genes in stomatal and myrosin cell lineages[39,46,47]; however, the detailed mechanism is not yet known. Liverworts have already lost several stomatal-related genes[4,48], such as the leucine-rich-repeat receptor-like gene *TOO MANY MOUTHS* (*TMM*) and the secreted peptide gene *EPIDERMAL PATTERNING FACTOR 1/2* (*EPF1/2*). The mechanisms underlying the development of setal-cell lineage might be different from the ones related to stomatal-cell lineages. RNA-seq and ChIP-seq analyses could identify genes that function downstream of the MpSETA-MpICE2 module.

Are *SMF* genes and/or Mp*SETA* orthologs conserved in liverworts other than *M. polymorpha*? A BLAST search of 1,000 plant transcriptomes (OneKP)[49], using At*FAMA* as query, revealed the absence of Ia bHLH TFs in liverworts except for Mp*SETA*. Since the liverwort transcriptome samples used in OneKP often do not contain sporophytes, it is difficult to detect genes that are specifically or transiently expressed in developing sporophytes, such as Ia bHLH. Therefore, genome analyses of various plant species in the future could be useful to understand the relationship between Ia bHLH sequence diversification and stomata/seta formation.

Methods

Phylogenetic analysis

The classifications of bHLH are according to Pires and Dolan[5]. For phylogenetic analyses of subfamilies Ia and IIIb bHLH, amino acid sequence information was retrieved from MarpolBase (https://marchantia.info), Phytozome v.13 (https://phytozome-next.jgi.doe.gov/), OneKP (https://db.cngb.org/onekp/), and TAIR (http://www.arabidopsis.org/). The bHLH domain and C-terminal ACT-like domain were aligned using MAFFT[50] v.6.864 (https://www.genome.jp/tools-bin/mafft) with the default parameters, and the alignment gaps were removed manually. Amino acid sequences were visualized with Jalview[51] v.2.11.2.1. BLAST searches of NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi) were used to predict the amino acids important for nucleotide binding and dimer formation in the bHLH domains. Phylogenetic tree constructions were performed using the maximum-likelihood algorithm on MEGA 7[52] with the JTT+G+I substitution model for Ia and LG+G+I substitution model for IIIb bHLH. Bootstrap analyses with 1000 replicates were performed in each analysis to assess the statistical support for the topology. Subfamilies III(a+c) and III(d+e) bHLHs were chosen as the outgroup of the phylogeny for Ia and IIIb bHLH, respectively.

Plant materials and growth condition

Male and female accessions of *M. polymorpha* L., Takaragaike-1 (Tak-1), and Tak-2, respectively, were maintained asexually. A female progeny backcrossed to Tak-1 for three backcross generations (BC3-38) and male Tak-1 were used as wild type. Gemmae and thalli were cultured on half-strength Gamborg's B5 media containing 1% (w/v) agar and 1% (w/v) sucrose under 50-60 μmol photons m^{-2} s^{-1} continuous white light at 22□. To induce gametangiophore development, 10 to 12-day-old thalli were transferred to 16-h-light/8-h-dark conditions with 50-60 μmol photons m^{-2} s^{-1} white light and 50-60 μmol photons m^{-2} s^{-1} far-red light emitted from diodes (IR LED STICK 18W, Namoto) at 18□ and incubated for 1 month.

The *Arabidopsis thaliana* Columbia-0 (Col-0) accession was used as wild type except for *mute-2* where Wassilewskija-4 (Ws-4) was used. Seeds were surface

sterilized with 70% ethanol and then sown onto half-strength MS media containing 0.5% (w/v) gellan gum and 1% (w/v) sucrose. The seeds were incubated at 22°C under 50-60 μmol photons m^{-2} s^{-1} continuous white light.

Complementation tests of *A. thaliana* stomatal defective mutants

T-DNA insertion mutants *spch-3* (SAIL_36_B06) and *fama-1* (SALK_100073) were obtained from the Arabidopsis Biological Resource Center (ABRC); *mute-2* (FLAG_225D03) from the French National Institute for Agricultural Research (INRA). To construct $_{pro}$At*SPCH:*Mp*SETA*, firstly At*SPCH* promoter (2,572 bp upstream of the translational initiation site) was amplified from Col-0 gDNA, and MpSETA CDS was amplified from cDNA derived from Tak-1 antheridiophores. $_{pro}$At*SPCH* and MpSETA CDS fragments were fused by PCR using PrimeSTAR GXL polymerase (Takara Bio), and the resultant PCR fragment was subcloned into pENTR1A entry clones (Invitrogen) at the SalI and EcoRV restriction sites using In-Fusion HD Cloning Kit (Takara Bio) and transferred into the destination vector pGWB501[53] using Gateway LR Clonase II Enzyme mix (Thermo Fisher Scientific). $_{pro}$At*MUTE:*Mp*SETA* and $_{pro}$At*FAMA:*Mp*SETA* constructs were generated by LR recombination of the R4pGWB501[54] or R4pGWB601[54] with a pENTR1A containing Mp*SETA* CDS at the EcoRI sites and either pENTR5'/TOPO_proAtMUTE harboring $_{pro}$At*MUTE*, 1,930 bp upstream of the translational initiation site, or pENTR5'/TOPO_proAtFAMA harboring $_{pro}$At*FAMA*, 3,105 bp upstream of the translational initiation site (R4pGWB501_proAtMUTE:MpSETA and R4pGWB601_proAtFAMA:MpSETA). The resultant plasmids were introduced into *spch-3/+*, *mute-2/+*, or *fama-1/+* heterozygous plants by the previously described method[55] using *Agrobacterium tumefaciens* strain GV3101. We confirmed that all transformants had a single insertion event by segregation analyses. We used T_3 or T_4 homozygous plants. We stained the cotyledons by FM4-64 and observed them using an LSM780 laser scanning microscope (Carl Zeiss). Images were processed with Fiji (NIH). The sequences of primers used in this study are shown in Supplementary Tables 1 and 2.

Gene expression analysis

Publicly available transcriptome data were downloaded from Sequence Read Archive (SRA) repository. Accession numbers include: sporelings (SRR4450254, SRR4450255, SRR4450256)[4], male thalli (DRR118949, DRR118950, DRR118951)[18], female thalli (DRR118943, DRR118944, DRR118945)[18], antheridiophores (DRR050346, DRR050347, DRR050348)[19], archegoniophores (DRR050351, DRR050352, DRR050353)[19], antheridia (DRR050349, DRR050350)[19], archegonia (DRR209029, DRR209030, DRR209031, DRR209031)[20], 13 DPF sporophytes (SRR1553297, SRR1553298, SRR1553299)[17], mature sporophytes (SRR896223)[4]. RNA-seq data were preprocessed to filter out low-quality sequences using fastp[56] v.0.20.0. The sequence reads were mapped to the *M. polymorpha* genome v.6.1 (https://marchantia.info) by STAR[57] v.2.7.8a with default parameters. The post-processing of SAM/BAM files was performed using SAMtools[58] v.1.11. The read counts for each gene were used to calculate transcript per million (TPM) by using RSEM[59] v.1.3.0 with default parameters. Plots were created using the Rstudio v.1.4.1106 (https://www.rstudio.com/).

Histology

Plant samples were fixed with 2% (w/v) paraformaldehyde and 2% (v/v) glutaraldehyde in 0.05 M cacodylate buffer (pH 7.4) for 2 hours at room temperature, post-fixed with 2% (v/v) osmium tetroxide in 0.1 M cacodylate buffer for 2 hours at room temperature, dehydrated in an ethanol series, substituted with acetone, and then embedded in Spurr's resin (Polysciences). The Spurr's blocks were cut into semi-thin sections (0.75-2 μm) with glass knives on an ultramicrotome Leica Ultracut UCT (Leica Microsystems) and stained with a solution containing 1% (w/v) sodium tetra-borate and 1% (w/v) toluidine blue O. Sections were mounted on MAS-coated glass slides (Matsunami Glass). Images were obtained using a VB-7010 (KEYENCE)/AxioCam HRc (Zeiss) and were processed with Fiji or Adobe Photoshop Elements 9 (Adobe Systems).

Histochemical GUS staining

To construct $_{pro}$Mp*SETA:GUS*, we amplified the genomic fragment of 4,194 bp upstream region of the translational initiation site from Tak-1 gDNA using PrimeSTAR Max DNA polymerase (Takara Bio), subcloned into pENTR1A at the EcoRI restriction sites using In-Fusion HD Cloning Kit and then transferred into destination vector pMpGWB104[60] using Gateway LR Clonase II Enzyme mix (Thermo Fisher Scientific). To construct $_{pro}$Mp*ICE2:Citrine-GUS-NLS*, firstly we amplified genomic fragment of 3,060 bp upstream region of the translational initiation site from Tak-1 gDNA, *Citrine* ORF from pMpGWB107[60], and *GUS-NLS* ORF from pPZP211_35S-NG-GUS-NLS-nosT[61]. These fragments were fused by PCR, subcloned into pENTR1A at the SalI and EcoRV sites using In-Fusion HD Cloning Kit, and transferred into pMpGWB101[60]. The resultant plasmids were introduced into Tak-1 accession by the previously described method[62] using *A. tumefaciens* strain GV2260.

The tissues of $_{pro}$Mp*SETA:GUS* or $_{pro}$Mp*ICE2:Citrine-GUS-NLS* plants were vacuum-infiltrated and incubated at 37□ overnight in the GUS staining solution containing 10 mM EDTA (pH 8.0), 100 mM NaH_2PO_4 (pH 7.0), 0.1% (v/v) Triton X-100, 0.5 g L^{-1} 5-bromo-4-chloro-3-indolyl-β-D-glucuronic acid (X-Gluc), 5 mM potassium-ferrocyanide, and 5 mM potassium-ferricyanide. Samples were washed in 70% (v/v) ethanol and cleared with chloral-hydrate / glycerol solution.

Generation of Mp*seta*^{ko} mutants

Tak-1 genomic sequences of 3,125 bp upstream and 3,101 bp downstream of MpSETA bHLH domain coding region were amplified by PCR with a PrimeSTAR Max DNA polymerase and inserted into PacI and AscI sites of pJHY-TMp1[26], respectively. The vector was introduced into germinating F_1 spores from Tak-1 and Tak-2 cross via *A. tumefaciens* strain GV2260 as previously described[63]. The transformed T_1 plants carrying the targeted insertion were selected by PCR using GoTaq DNA polymerase (Promega). As T_1 plants of Mp*seta-1*^{ko} and Mp*seta-2*^{ko} were both females, male mutants were obtained from F_1 sporelings by crossing female mutants with Tak-1.

Reverse transcription PCR

For gene expression analysis, 21 DPF sporophytes were collected from wild type, Mp*seta-1*^{ko}, and Mp*seta-2*^{ko}, and total RNA was extracted by RNeasy Plant Mini Kit (Qiagen) according to the manufacturer's protocol. The quality and quantity of total RNA were evaluated with NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific). First-strand cDNA was synthesized using ReverTra Ace qPCR RT Master

Mix with gDNA Remover (Toyobo), and semiquantitative RT-PCR was undertaken using Mp*EF1α* as a loading control[64].

Yeast two-hybrid assay

The coding sequences of MpSETA and MpICE2 were amplified from cDNA derived from mRNA of Tak-1 thalli by PCR using PrimeSTAR Max DNA polymerase or PrimeSTAR GXL polymerase (Takara Bio). The coding sequences of AtICE1 and AtSCRM2 were amplified from cDNA derived from Col-0 leaves using PrimeSTAR GXL polymerase. The resultant PCR fragments were subcloned into pENTR1A at the EcoRI sites or SalI and EcoRV sites using In-Fusion HD Cloning Kit. To generate a bait destination vector, pDEST-GBKT7-Amp^r, $_{pro}$*AmpR:AmpR* was amplified from pDEST-GADT7[65] by PCR using PrimeSTAR Max DNA polymerase, and the fragment was cloned into at SfoI site of pDEST-GBKT7[65]. The inserted fragments, MpSETA, MpICE2, AtICE1, and AtSCRM2 were transferred into pDEST-GADT7 and/or pDEST-GBKT7-Amp^r using Gateway LR Clonase II Enzyme mix. Bait and prey constructs were co-transformed into the yeast strain Y2HGold (Clontech) using the Frozen-EZ Yeast Transformation II Kit (Zymo Research) and the transformants were grown on the solid SD media lacking Leu and Trp (SD-LW). To examine the interaction between the bait and prey protein, transformants were grown on the solid SD media lacking Leu, Trp, His, and adenine with 40 mg L^{-1} X-α-gal and 200 μg L^{-1} Aureobasidin A (SD-LWHA/X/AbA) at 30□. pDEST-GBKT7-Amp^r and pDEST-GADT7 were used as negative controls (empty).

BiFC

The coding sequences of MpSETA and MpICE2 subcloned into pENTR1A described above were transferred into the pB4GWnY and/or pB4GWcY binary vector[66] using LR reaction to be fused with N-terminal or C-terminal half of YFP (Mp*SETA-nYFP* and Mp*ICE2-cYFP* driven by the *CaMV35S* promoter). Transformed *A. tumefaciens* strain GV3101 cells harboring expression vectors were cultured and resuspended in distilled water to a final optical density of OD_{600} = 1.0. Mixed *Agrobacterium* cultures were infiltrated into 4-week-old *Nicotiana benthamiana* leaves. Nuclei were stained with 4',6-diamidino-2-phenylindol (DAPI) at the 1 mg L^{-1} concentration for 30 min. Samples were observed 1.5 days post-inoculation (DPI) by an LSM780 laser scanning microscope. pB4GWnY and pB4GWcY were used as negative control (empty).

Generation of Mp*ice2^{ge}* mutants

To generate Mp*ice2^{ge}* mutants, Mp*ICE2* locus encoding the bHLH domain was edited using CRISPR/Cas9 based genome-editing system as previously described[29]. Two sgRNA were designed to generate Mp*ice2^{ge}* mutants. The oligonucleotides encoding sgRNA were cloned into pMpGE_En03[29] between BsaI sites and then introduced into pMpGE010[29]. Male and female mutants that do not harbor T-DNA containing *Cas9* were obtained from F_1 sporelings crossed with wild type and Mp*ice2-2^{ge}* or Mp*ice2-6^{ge}*.

Complementation tests of Mp*seta^{ko}* and Mp*ice2^{ge}*

To construct *g*Mp*SETA* for Mp*seta-1^{ko}* complementation, the genomic region containing 4,194 bp upstream region and coding sequences was amplified from Tak-1 genomic DNA. The fragment was cloned into pENTR1A between EcoRI sites and then

introduced into pMpGWB301[60]. The resultant plasmids were introduced into female Mp*seta-1*^{ko}. Male *g*Mp*SETA* Mp*seta-1*^{ko} was obtained from F_1 sporeling crossed with female *g*Mp*SETA* Mp*seta-1*^{ko} and male Mp*seta-1*^{ko}. To construct *g*Mp*ICE2* for Mp*ice2-6*^{ge}, the genomic region containing 3,060 bp upstream regions and coding sequences was amplified from Tak-1 genomic DNA. The fragment was cloned into pENTR1A between SalI and EcoRV sites and then introduced into pMpGWB101[60]. The resultant plasmids were introduced into female Mp*ice2-6*^{ge}. A male *g*Mp*ICE2* Mp*ice2-6*^{ge} was obtained from a F_1 sporeling derived from a cross between a female *g*Mp*ICE2* Mp*ice2-6*^{ge} and a male Mp*ice2-6*^{ge}.

References

1. Ligrone, R., Duckett, J. G. & Renzaglia, K. S. Major transitions in the evolution of early land plants: A bryological perspective. *Ann. Bot.* **109**, 851–871 (2012).
2. Romani, F. & Moreno, J. E. Molecular mechanisms involved in functional macroevolution of plant transcription factors. *New Phytol.* **230**, 1345–1353 (2021).
3. Catarino, B., Hetherington, A. J., Emms, D. M., Kelly, S. & Dolan, L. The stepwise increase in the number of transcription factor families in the precambrian predated the diversification of plants on land. *Mol. Biol. Evol.* **33**, 2815–2819 (2016).
4. Bowman, J. L. *et al.* Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell* **171**, 287–304 (2017).
5. Pires, N. & Dolan, L. Origin and diversification of basic-helix-loop-helix proteins in plants. *Mol. Biol. Evol.* **5**, 911–912 (2010).
6. MacAlister, C. A., Ohashi-Ito, K. & Bergmann, D. C. Transcription factor control of asymmetric cell divisions that establish the stomatal lineage. *Nature* **445**, 537–540 (2007).
7. Pillitteri, L. J., Sloan, D. B., Bogenschutz, N. L. & Torii, K. U. Termination of asymmetric cell division and differentiation of stomata. *Nature* **445**, 501–505 (2007).
8. Ohashi-Ito, K. & Bergmann, D. C. *Arabidopsis* FAMA controls the final proliferation/differentiation switch during stomatal development. *Plant Cell* **18**, 2493–2505 (2006).
9. Kanaoka, M. M. *et al.* SCREAM/ICE1 and SCREAM2 specify three cell-state transitional steps leading to *Arabidopsis* stomatal differentiation. *Plant Cell* **20**, 1775–1785 (2008).
10. Chater, C. C. *et al.* Origin and function of stomata in the moss *Physcomitrella patens*. *Nat. Plants* **2**, 16179 (2016).
11. Chater, C. C. C., Caine, R. S., Fleming, A. J. & Gray, J. E. Origins and evolution of stomatal development. *Plant Physiol.* **174**, 624–638 (2017).
12. Harris, B. J. *et al.* Phylogenomic Evidence for the Monophyly of Bryophytes and the Reductive Evolution of Stomata. *Curr. Biol.* **30**, 2001–2012.e2 (2020).
13. Flores-Sandoval, E., Romani, F. & Bowman, J. L. Co-expression and transcriptome analysis of *Marchantia polymorpha* transcription factors supports Class C ARFs as independent actors of an ancient auxin regulatory module. *Front. Plant Sci.* **9**, 1–21 (2018).

14. Feller, A., Hernandez, J. M. & Grotewold, E. An ACT-like domain participates in the dimerization of several plant basic-helix-loop-helix transcription factors. *J. Biol. Chem.* **281**, 28964–28974 (2006).
15. Seo, H., *et al.* Intragenic suppressors unravel the role of the SCREAM ACT-like domain for bHLH partner selectivity in stomatal development. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2117774119 (2022).
16. Pillitteri, L. J., Bogenschutz, N. L. & Torii, K. U. The bHLH protein, MUTE, Controls Differentiation of Stomata and the Hydathode Pore in *Arabidopsis*. *Plant Cell Physiol.* **49**, 934–943 (2008).
17. Frank, M. H. & Scanlon, M. J. Transcriptomic evidence for the evolution of shoot meristem function in sporophyte-dominant land plants through concerted selection of ancestral gametophytic and sporophytic genetic programs. *Mol. Biol. Evol.* **32**, 355–367 (2015).
18. Yamaoka, S. *et al.* Generative cell specification requires transcription factors evolutionarily conserved in land plants. *Curr. Biol.* **28**, 479–486.e5 (2018).
19. Higo, A. *et al.* Transcriptional framework of male gametogenesis in the liverwort *Marchantia polymorpha* L. *Plant Cell Physiol.* **57**, 325–338 (2016).
20. Hisanaga, T. *et al.* Deep evolutionary origin of gamete-directed zygote activation by KNOX/BELL transcription factors in green plants. *eLife* **10**, e57090 (2021).
21. Shimamura, M. *Marchantia polymorpha*: Taxonomy, phylogeny and morphology of a model system. *Plant Cell Physiol.* **57**, 230–256 (2016).
22. Haig, D. Filial mistletoes: The functional morphology of moss sporophytes. *Ann. Bot.* **111**, 337–345 (2013).
23. Ligrone, R., Duckett, J. G. & Renzaglia, K. S. The origin of the sporophyte shoot in land plants: A bryological perspective. *Ann. Bot.* **110**, 935–941 (2012).
24. Durand, E. J. The development of the sexual organs and sporogonium of *Marchantia polymorpha*. *Bull. Torrey Bot. Club* **7**, 321–335 (1908).
25. Hernández-García, J. *et al.* Coordination between growth and stress responses by DELLA in the liverwort *Marchantia polymorpha*. *Curr. Biol.* **31**, 3678–3686.e11 (2021).
26. Ishizaki, K., Johzuka-Hisatomi, Y., Ishida, S., Iida, S. & Kohchi, T. Homologous recombination-mediated gene targeting in the liverwort *Marchantia polymorpha* L. *Sci. Rep.* **3**, 1532 (2013).
27. Shaw, J. & Renzaglia, K. Phylogeny and diversification of bryophytes. *Am. J. Bot.* **91**, 1557–1581 (2004).
28. Wu, Z. *et al.* Multiple transcriptional factors control stomata development in rice. *New Phytol.* **223**, 220–232 (2019).
29. Sugano, S. S. *et al.* Efficient CRISPR/Cas9-based genome editing and its application to conditional genetic analysis in *Marchantia polymorpha*. *PLoS One* **13**, e0205117 (2018).
30. Sugano, S. S. *et al.* CRISPR/Cas9-mediated targeted mutagenesis in the liverwort *Marchantia polymorpha* L. *Plant Cell Physiol.* **55**, 475–481 (2014).
31. Chinnusamy, V. *et al.* ICE1: A regulator of cold-induced transcriptome and freezing tolerance in *Arabidopsis*. *Genes Dev.* **17**, 1043–1054 (2003).
32. Davies, K. A. & Bergmann, D. C. Functional specialization of stomatal bHLHs through modification of DNA-binding and phosphoregulation potential. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15585–15590 (2014).

33. MacAlister, C. A. & Bergmann, D. C. Sequence and function of basic helix-loop-helix proteins required for stomatal development in *Arabidopsis* are deeply conserved in land plants. *Evol. Dev.* **13**, 182–192 (2011).
34. Merced, A. & Renzaglia, K. S. Structure, function and evolution of stomata from a bryological perspective. *Bryophyt. Divers. Evol.* **39**, 7–20 (2017).
35. Wei, D. *et al.* INDUCER OF CBF EXPRESSION 1 is a male fertility regulator impacting anther dehydration in *Arabidopsis*. *PLoS Genet.* **14**, e1007695 (2018).
36. Higo, A. *et al.* Transcription factor DUO1 generated by neo-functionalization is associated with evolution of sperm differentiation in plants. *Nat. Commun.* **9**, 5283 (2018).
37. Menand, B. *et al.* An ancient mechanism controls the development of cells with a rooting function in land plants. *Science.* **316**, 1477–1480 (2007).
38. Koshimizu, S. *et al. Physcomitrella* MADS-box genes regulate water supply and sperm movement for fertilization. *Nat. Plants* **4**, 36–45 (2018).
39. Shirakawa, M. *et al.* FAMA is an essential component for the differentiation of two distinct cell types, myrosin cells and guard cells, in *Arabidopsis*. *Plant Cell* **26**, 4039–4052 (2014).
40. Yasui, Y. *et al.* GEMMA CUP-ASSOCIATED MYB1, an ortholog of axillary meristem regulators, is essential in vegetative reproduction in *Marchantia polymorpha*. *Curr. Biol.* **29**, 3987–3995.e5 (2019).
41. Proust, H. *et al.* RSL Class I genes controlled the development of epidermal structures in the common ancestor of land plants. *Curr. Biol.* **26**, 93–99 (2016).
42. Puttick, M. N. *et al.* The interrelationships of land plants and the nature of the ancestral embryophyte. *Curr. Biol.* **28**, 733–745.e2 (2018).
43. Renzaglia, K. S. & Garbary, D. J. Motile gametes of land plants: Diversity, development, and evolution. *CRC. Crit. Rev. Plant Sci.* **20**, 107–213 (2001).
44. Sakakibara, K., Nishiyama, T., Deguchi, H. & Hasebe, M. Class 1 KNOX genes are not involved in shoot development in the moss *Physcomitrella patens* but do function in sporophyte development. *Evol. Dev.* **10**, 555–566 (2008).
45. Coudert, Y., Novák, O. & Harrison, C. J. A KNOX-Cytokinin regulatory module predates the origin of indeterminate vascular plants. *Curr. Biol.* **29**, 2743–2750.e5 (2019).
46. Li, M. & Sack, F. D. Myrosin idioblast cell fate and development are regulated by the *Arabidopsis* transcription factor FAMA, the auxin pathway, and vesicular trafficking. *Plant Cell* **26**, 4053–4066 (2014).
47. Shirakawa, M., Tanida, M. & Ito, T. The cell differentiation of idioblast myrosin cells: Similarities with vascular and guard cells. *Front. Plant Sci.* **12**, 829541 (2022).
48. Li, F. W. *et al. Anthoceros* genomes illuminate the origin of land plants and the unique biology of hornworts. *Nat. Plants* **6**, 259–272 (2020).
49. Leebens-Mack, J. H. *et al.* One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**, 679–685 (2019).
50. Katoh, K. & Toh, H. Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.* **9**, 286–298 (2008).
51. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

52. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
53. Nakagawa, T. *et al.* Improved gateway binary vectors: High-performance vectors for creation of fusion constructs in transgenic analysis of plants. *Biosci. Biotechnol. Biochem*. **71**, 2095–2100 (2007).
54. Nakamura, S. *et al.* Gateway binary vectors with the bialaphos resistance gene, bar, as a selection marker for plant transformation. *Biosci. Biotechnol. Biochem.* **74**, 1315–1319 (2010).
55. Clough, S. J. & Bent, A. F. Floral dip: A simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).
56. Chen, S., Zhou, Y., Chen, Y. & Gu, J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
57. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
58. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
59. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
60. Ishizaki, K. *et al.* Development of gateway binary vector series with four different selection markers for the liverwort *Marchantia polymorpha*. *PLoS One* **10**, e0138876 (2015).
61. Matsushita, T., Mochizuki, N. & Nagatani, A. Dimers of the N-terminal domain of phytochrome B are functional in the nucleus. *Nature* **424**, 571–574 (2003).
62. Tsuboyama, S., Nona, S., Ezura, H. & Kodama, Y. Improved G-AgarTrap□: A highly efficient transformation method for intact gemmalings of the liverwort *Marchantia polymorpha*. *Sci. Rep.* **8**, 10800 (2018).
63. Ishizaki, K., Chiyoda, S., Yamato, K. T. & Kohchi, T. Agrobacterium-mediated transformation of the haploid liverwort *Marchantia polymorpha* L., an emerging model for plant biology. *Plant Cell Physiol.* **49**, 1084–1091 (2008).
64. Rövekamp, M., Bowman, J. L. & Grossniklaus, U. *Marchantia* Mp*RKD* regulates the gametophyte-sporophyte transition by keeping egg cells quiescent in the absence of fertilization. *Curr. Biol.* **26**, 1782–1789 (2016).
65. Rossignol, P., Collier, S., Bush, M., Shaw, P. & Doonan, J. H. *Arabidopsis* POT1A interacts with TERT-V(18), an N-terminal splicing variant of telomerase. *J. Cell Sci.* **120**, 3678–3687 (2007).
66. Hino, T. *et al.* Two Sec13p homologs, AtSec13A and AtSec13B, redundantly contribute to the formation of COPII transport vesicles in *Arabidopsis thaliana*. *Biosci. Biotechnol. Biochem.* **75**, 1848–1852 (2011).

Acknowledgements

We thank Drs. Tsuyoshi Nakagawa (Shimane University, Japan), Shoji Mano (National Institute for Basic Biology, Japan), and Shigeo S. Sugano (National Institute of Advanced Industrial Science and Technology, Japan) for sharing vectors. We also thank Dr. Keiji Nakajima (Nara Institute of Science and Technology, Japan) for sharing the figures of plants in Fig. 5b. This work was supported by MEXT/JSPS KAKENHI grants to M.S. (JP19K06722 and JP20H05416), to K.T. (JP26711017 and JP18K06283), to

T.M. (JP20H05905 and JP20H05906), to I.H.-N. (JP15H05776), to R.N. (JP20H04884) and to T.S. (JP18K06284); Grants-in-Aid JSPS Fellows to M.S. (JP12J05453) and to K.C.M. (JP21J14990); the Takeda Science Foundation, the Kato Memorial Bioscience Foundation, and the Ohsumi Frontier Science Foundation to M.S. J.L.-M. was supported by a Ph.D. studentship from the Darwin Trust of Edinburgh.

Author contributions

K.C.M. and T.S. conceived and designed the research in general; K.C.M. performed most of the experiments and analyzed the data; M.S. and Y.M. performed the experiments on Mp*SETA*; J.L.-M., G.I., and J.G. performed the experiments on Mp*ICE2*; K.T., Y.O., T.M., I.H.-N., R.N., J.G., and T.K. supervised the experiments; K.C.M. and T.S. wrote the manuscript; All authors read, edited, and approved the manuscript.

Competing interests

The authors declare no competing interests.

Figure legends

Fig. 1 | MpSETA is the only bHLH transcription factor that belongs to subfamily Ia in *M. polymorpha*. a, A maximum-likelihood bHLH phylogenetic tree of subfamilies Ia, Ib(1) (purple), II (gray), and III(a+c) (outgroup) is shown. Numbers at branches indicate bootstrap values calculated from 1,000 replicates. Ia bHLHs are divided into four groups: TEC1/6 clade (light blue), MYC70 clade (cream-yellow), SMF/FAMA clade (magenta), and SPCH/MUTE clade (green). Species are abbreviated as follows: Mp, *M. polymorpha* (liverwort); Pp, *P. patens* (moss); Cepur, *Ceratodon purpureus* (moss); Aagr, *Anthoceros agrestis* (hornwort); Sm, *Selaginella moellendorffii* (lycophyte); AmTr, *Amborella trichopoda* (basal angiosperm); Os, *Oryza sativa* (monocot); At, *Arabidopsis thaliana* (dicot). An arrow indicates MpSETA/MpBHLH35 (Mp2g04160), and arrowheads indicate bHLH TFs mentioned as FAMA-like bHLH TFs in a previous study[13]. Amino acid sequences from only *A. thaliana* and *M. polymorpha* were used for subfamilies Ib(1), II, and III(a+c). **b,** Confocal images of *A. thaliana* abaxial cotyledons of wild type (Ws-4), *mute-2*, and $_{pro}$At*MUTE:*Mp*SETA mute-2* at 9 days after stratification (DAS). The upper panels and lower panels show the middle area and tip area of the cotyledons, respectively (left image). Arrowheads and asterisks indicate stomata and hydathode pores, respectively. Bars, 100 μm (**b**, confocal images), and 1 mm (**b**, left).

Fig. 2 | Mp*SETA* is preferentially expressed in developing sporophyte. a, Box plot showing the expression profiles of Mp*SETA* across nine *M. polymorpha* tissues. Y-axis shows transcripts per million (TPM). Sporeling and thalli are vegetative gametophytic (*n*) organs. Antheridiophores and archegoniophores are haploid male and female reproductive receptacles, respectively. Antheridia and archegonia are the organs that produce sperms and egg cells, respectively. Sporophytes are diploid (2*n*) organs developed after fertilization. **b**, Tissue section of mature *M. polymorpha* sporophyte (stage VII). **c**, Developmental stages of *M. polymorpha* sporophytes after 7–40 days' postfertilization. (I) sporophyte differentiating amphithecium and endothecium, (II) later globular sporophyte, (III) sporophyte differentiating foot, seta, and archesporial tissues, (IV) sporophyte differentiating sporogenous cells and elaterocytes (or elater mother cell), (V) sporophyte differentiating sporocytes (or spore mother cells), (VI) sporophyte differentiating spore tetrads after meiosis, (VII) mature sporophyte before the seta elongation, (VIII) mature sporophyte after seta elongation. F, foot; s, seta; at, archesporial tissue; sp, sporangium. Arrows indicate the endothecium. Arrowheads indicate the cell wall of the first cell division. See also Extended Data Fig. 3 for the developmental process of seta and foot. **d**, Histochemical detection of β-glucuronidase (GUS) activity driven by Mp*SETA* promoter in the developing seta region from stage II to stage VII. f, foot; s, seta; at, archesporial tissue; sp, sporangium; ca, calyptra; p, pseudoperianth. Arrowheads indicate the cell wall of the first cell division. Bars, 50 μm (**b**, and **c**, top), and 100 μm (**c**, bottom, and **d**).

Fig. 3 | Mps*eta*^{ko} mutants show developmental defects in the seta. a,**b**, Longitudinal sections of sporophytes of indicated genotypes. Enlarged images of the square areas are shown in (**b**). Dashed and solid arrows indicate the length of the sporophyte (SP) and the length from the boundary between gametophyte and sporophyte to the proximal side of sporangium (SF), respectively. **c**,**d**, Quantitative data showing the SF/SP length ratio

and the number of cells in the seta region. Since the boundary between seta and foot is unclear, cells excluding transfer cells (the cells at the boundary between gametophyte and sporophyte) were used to count the number of cells. The different letters indicate significantly different mean values at $p < 0.01$ (Tukey's HSD test). ($n = 6$ independent lines). **e**, One month-post-fertilization archegoniophores that produce mature sporophytes. Arrowheads indicate sporangia exposed to the outside of calyptra due to elongation of seta cells. Bars, 100 μm (**a**,**b**), and 5 mm (**e**).

Fig. 4 | MpICE2 regulates setal development in cooperation with MpSETA. a. Box plot showing the expression profiles of Mp*ICE1* and Mp*ICE2* across nine *M. polymorpha* tissues. Y-axis shows transcripts per million (TPM). **b**, Y2H assays showing the interaction between MpSETA and MpICE2. MpSETA fused with GAL4 DNA-binding domain (DBD) was used as bait, while MpICE2 fused with GAL4 activation domain (AD) was used as prey. DBD alone and AD alone were used as the negative control. **c**, BiFC assays showing the interaction between MpSETA and MpICE2 in *N. benthamiana* leaf epidermal cells. MpSETA was fused to the N-terminal fragment of EYFP (nYFP), while MpICE2 was fused to the C-terminal fragment of EYFP (cYFP). nYFP alone and cYFP alone were used as the negative control. Nuclei were stained by 4',6-diamidino-2-phenylindole (DAPI). **d**,**e**, Longitudinal sections of sporophytes of indicated genotypes. Enlarged images of the square areas are shown in (**e**). Dashed and solid arrows indicate the length of the sporophyte (SP) and the length from the boundary between gametophyte and sporophyte to the proximal side of sporangium (SF), respectively. **f**,**g**, Quantitative data showing the SF/SP length ratio and the number of cells in the seta region. Since the boundary between seta and foot is unclear, cells excluding transfer cells (the cells at the boundary between gametophyte and sporophyte) were used to count the number of cells. The different letters indicate significantly different mean values at $p < 0.01$ (Tukey's HSD test) ($n = 6$ or 7 independent lines). **h**, One month-post-fertilization archegoniophores that produce mature sporophytes. Arrowheads indicate sporangia exposed to the outside of calyptra due to elongation of seta cells. Bars, 10 μm (**c**, YFP, DAPI, and Merge), 100 μm (**c**, left, **d**, and **e**), and 5 mm (**f**).

Fig. 5 | Function and co-option of the Ia-IIIb bHLH module during the evolution of land plants. a, Schematic model comparing the molecular functions of the Ia-IIIb bHLH TF modules in *A. thaliana* and *M. polymorpha* during cell fate determination. In *A. thaliana*, the heterodimer of Ia bHLHs (SPCH, MUTE, and FAMA) and IIIb bHLHs (ICE1 and SCRM2) control the development of stomata. In *M. polymorpha*, the heterodimer of Ia bHLH (MpSETA) and IIIb bHLH (MpICE2) regulates cell differentiation and cell division in the seta precursor transition and might directly or indirectly be involved in the symmetric division of a putative seta mother cell. **b**, An evolutionary model for the Ia-IIIb bHLH TF module. Co-option of the Ia-IIIb bHLH module might have occurred multiple times independently during the evolution of land plants. First, stomata and a transcriptional module consisting of Ia-IIIb bHLHs evolved in the common ancestor of land plants. In the ancestral plant, the Ia bHLHs may have had MUTE- and FAMA-like functions. Second, the Ia-IIIb bHLH TF module might have been co-opted to regulate setal development in the ancestor of the Setaphyta. Third, after mosses and liverworts diverged, the common ancestor of liverworts lost its stomata.

A co-option of the Ia-IIIb bHLH module occurred in the Brassicales plants for regulating myrosin idioblast development.

Extended Data

Extended Data Fig. 1 | Comparison of the domain architecture of Ia bHLHs from *M. polymorpha, P. patens,* and *A. thaliana*. a, A diagram of the domain architecture of MpSETA (*M. polymorpha*), PpSMF1, PpSMF2 (*P. patens*), AtSPCH, AtMUTE, and AtFAMA (*A. thaliana*). While no PEST domain was identified, MpSETA has a bHLH domain and SMF domain conserved at the C-terminus like other Ia bHLH proteins. SMF domain is structurally considered to be the ACT-like domain, which is a putative domain for protein-protein dimerization. **b**, Sequence alignment of the bHLH domain of Ia bHLH proteins. Ia bHLHs are surrounded by a black box, and others are Ib(1) bHLHs. Asterisks indicate amino acids that are assumed to be important for binding to the E-box (CANNTG), and the triangles indicate amino acids that are assumed to be important for dimerization of the bHLH domain. The yellow box indicates the LxCxE motif, which is a binding motif with Retinoblastoma-related (RBR). **c**, Sequence alignment of the C-terminal SMF domain of Ia bHLH proteins.

Extended Data Fig. 2 | Function of Mp*SETA* in *A. thaliana* Ia bHLH mutants. a, Confocal images of *A. thaliana* abaxial cotyledons of wild type (Col-0), *spch-3*, and $_{pro}$At*SPCH:*Mp*SETA spch-3* at 9 days after stratification (DAS). **b**, Confocal images of *A. thaliana* abaxial cotyledons of wild type (Col-0), *fama-1*, and $_{pro}$At*FAMA:*Mp*SETA fama-1* at 9 DAS. Brackets and arrows indicate *fama* tumors and stomatal-lineage cells, respectively. **c**, Quantitative data of the distribution of the number of cell divisions that occurred in the stomatal lineage in each genotype. (n > 320 cells per genotype, 9 DAS cotyledons). **d**, Y2H assays in which the MpSETA fused with GAL4 DNA-binding domain (DBD) was used as bait, and the AtICE1 and AtSCRM2 fused with GAL4 activation domain (AD) were used as prey. DBD alone and AD alone were used as the negative controls.

Extended Data Fig. 3 | Development of seta in *M. polymorpha*. a, Magnified images showing the foot and seta for each stage of wild-type sporophytes. Same images as used in Fig. 2c. **b**, The formation of the cell files of seta in the wild type stage VI sporophyte. Arrowheads indicate the plane of proliferative cell divisions. **c**, Histochemical detection of β-glucuronidase (GUS) activity in the sporophytes crossed with male wild type and female $_{pro}$Mp*CYCB;1:Dbox-GUS*. Bars, 100 μm.

Extended Data Fig. 4 | Expression analysis of Mp*SETA* in the gametophytic tissues. Histochemical detection of β-glucuronidase (GUS) activity driven by Mp*SETA* promoter in the developing antheridia. Bars, 1 mm.

Extended Data Fig. 5 | Generation and phenotypes of Mp*SETA* knock-out lines. a, Structure of the Mp*SETA* locus disrupted by homologous recombination. Knock-out lines have a deletion in the bHLH domain coding region. White boxes indicate the exons of the MpSETA coding sequence. DT-A, diphtheria toxin A fragment gene; Hgr^R, hygromycin resistant gene. **b**, Genotyping of the Mp$seta^{ko}$ lines used in this study to distinguish the sex. rbm27, a male-specific marker; rhf73, a female-specific marker. **c**, Genotyping of the Mp$seta^{ko}$ lines. The position of primers used for PCR is shown in (**a**).

M, Male; F, Female. **d**, RT-PCR to confirm the loss of the full-length Mp*SETA* transcript in Mp*seta*^{ko} lines in 21 DPF sporophytes. Mp*EF1α* was used as an internal control. **e**, Spermatogenesis process in WT and Mp*seta*^{ko} lines. All images are at the same scale. Bars, 10 μm (**e**).

Extended Data Fig. 6 | Comparative analyses of setal or spore development between wild type and Mp*seta-1*^{ko}. a, Longitudinal sections of stage V and stage VI Mp*seta-1*^{ko} sporophytes. **b**, Sporogenesis process in wild type and Mp*seta-1*^{ko}. Sporocytes are the cells before the pre-meiotic stage, and spore tetrads are the cells immediately after meiosis. All images are at the same scale. Bars, 100 μm (**a**) and 50 μm (**b**).

Extended Data Fig. 7 | Phylogenetic tree of IIIb bHLH TFs. A maximum-likelihood bHLH phylogenetic tree of subfamilies IIIb, III (a+c) (light blue), and III(d+e) (outgroup) is shown. Numbers at branches indicate bootstrap values calculated from 1,000 replicates. IIIb bHLHs are divided into 2 groups: ICE/SCRM clade (orange) and NFL clade (magenta). Species are abbreviated as follows: Mp, *M. polymorpha* (liverwort); Pp, *P. patens* (moss); Cepur, *Ceratodon purpureus* (moss); Aagr, *Anthoceros agrestis* (hornwort); Sm, *Selaginella moellendorffii* (lycophyte); AmTr, *Amborella trichopoda* (basal angiosperm); Os, *Oryza sativa* (monocot); At, *A. thaliana* (dicot). Arrows indicate MpICE1 (Mp4g04910) and MpICE2 (Mp4g04920). For the phylogenetic construction of subfamilies III(a+c) and III(d+e), we used the amino acid sequences from only *A. thaliana* and *M. polymorpha*.

Extended Data Fig. 8 | Comparison of the domain architecture of IIIb bHLHs from *M. polymorpha, P. patens,* and *A. thaliana*. a, A diagram of the domain architecture of MpICE1, MpICE2 (*M. polymorpha*), PpSCRM1 (*P. patens*), AtICE1, and AtSCRM2 (*A. thaliana*). MpICE1 and MpICE2 have a bHLH domain and ACT-like domain conserved at the C-terminus like other IIIb bHLH proteins. **b**, Sequence alignment of the bHLH domain of IIIb bHLH proteins. IIIb bHLHs are surrounded by a black box, and others are outgroup. Asterisks indicate amino acids that are assumed to be important for binding to the E-box (CANNTG). **c**, Sequence alignment of the C-terminal ACT-like domain of IIIb bHLH proteins.

Extended Data Fig. 9 | The expression analysis of Mp*ICE2*. a, Histochemical detection of β-glucuronidase (GUS) activity driven by Mp*ICE2* promoter in the vegetative thallus. **b**, Confocal images of the dorsal epidermis of $_{pro}$Mp*ICE2:Citrine-GUS-NLS* line. Upper and lower panels indicate the epidermis around the apical notch and the epidermis around the midrib, respectively. Arrows indicate the air pores. **c**,**d**, Histochemical detection of GUS activity driven by Mp*ICE2* promoter in the gametophytic reproductive organs. An antheridiophore (**c**) and an archegoniophore (**d**) are shown. **e**, Expression pattern of Mp*ICE2* in the developing sporophytes. f, foot; s, seta; at, archesporial tissue; sp, sporangium; ca, calyptra; p, pseudoperianth (*n*). Arrowheads indicate the cell wall of the first cell division. Bars, 5 mm (**c** and **d**), 100 μm (**b** and **e**).

Extended Data Fig. 10 | Generation of Mp*ice2* mutants by CRISPR/Cas9. a,

Schematic representation of the Mp*ICE2* gene and the resulting mutations in the obtained CRISPR/Cas9-generated alleles. Gray, white, and blue boxes indicate the coding sequences (CDS), the untranslated regions (UTR), and the bHLH domain coding region, respectively. **b**, Sequence alignment of putative translational products of wild type and Mp*ice2*^{ge} mutants. Asterisks indicate the amino acids that are assumed to be important for binding to E-box.

Supplementary Information

Supplementary Tables 1 and 2

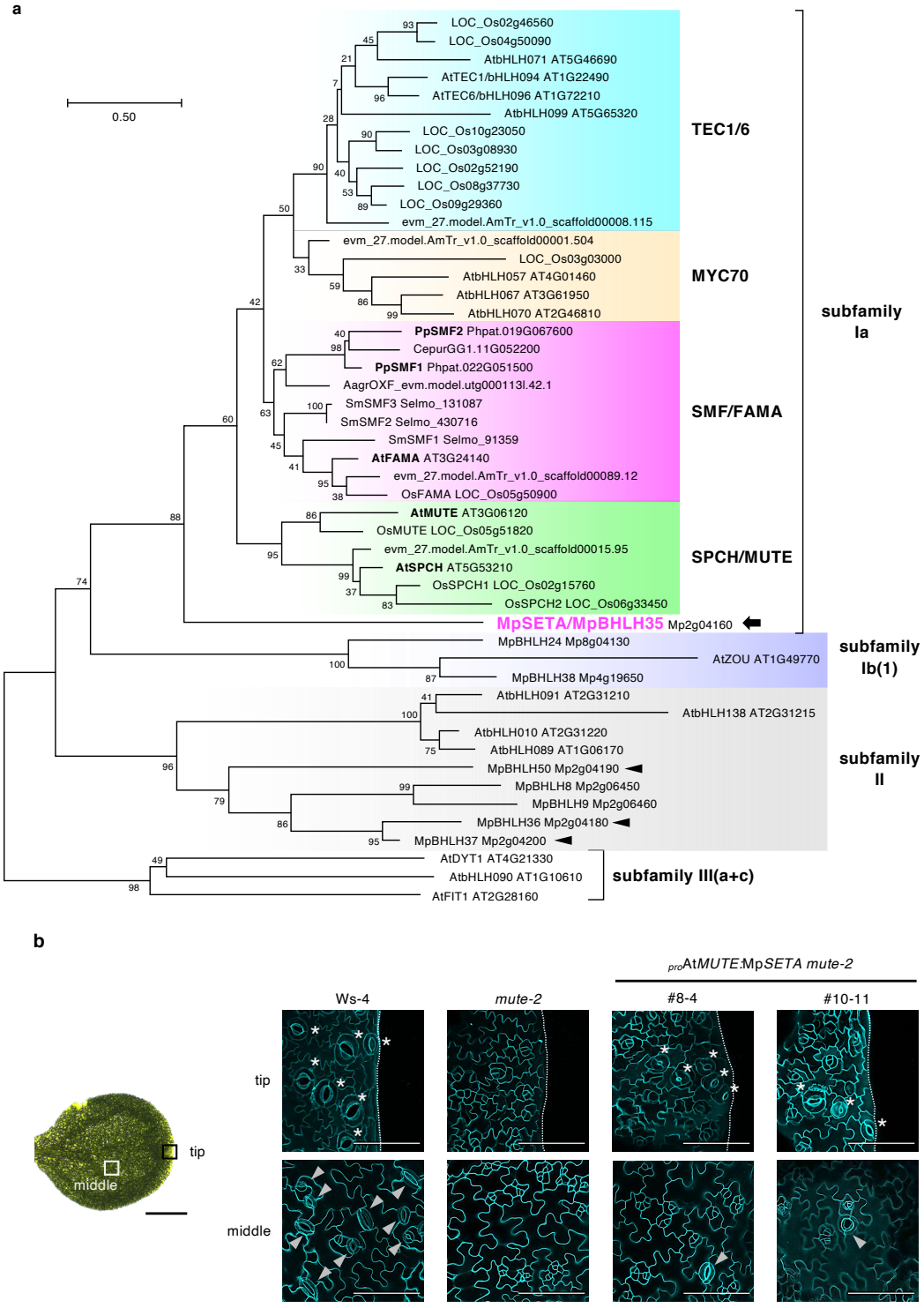


Fig. 1 l Mp*SETA* is the only bHLH transcription factor that belongs to subfamily Ia in *M. polymorpha*. a, A maximum-likelihood bHLH phylogenetic tree of subfamilies Ia, Ib(1) (purple), II (gray), and III(a+c) (outgroup) is shown. Numbers at branches indicate bootstrap values calculated from 1,000 replicates. Ia bHLHs are divided into four groups: TEC1/6 clade (light blue), MYC70 clade (cream-yellow), SMF/FAMA clade (magenta), and SPCH/MUTE clade (green). Species are abbreviated as follows: Mp, *M. polymorpha* (liverwort); Pp, *P. patens* (moss); Cepur, *Ceratodon purpureus* (moss); Aagr, *Anthoceros agrestis* (hornwort); Sm, *Selaginella moellendorffii* (lycophyte); AmTr, *Amborella trichopoda* (basal angiosperm); Os, *Oryza sativa* (monocot); At, *Arabidopsis thaliana* (dicot). An arrow indicates MpSETA/MpBHLH35 (Mp2g04160), and arrowheads indicate bHLH TFs mentioned as FAMA-like bHLH TFs in a previous study[49]. Amino acid sequences from only *A. thaliana* and *M. polymorpha* were used for subfamilies Ib(1), II, and III(a+c). **b,** Confocal images of *A. thaliana* abaxial cotyledons of wild type (Ws-4), *mute-2*, and $_{pro}$At*MUTE*:Mp*SETA mute-2* at 9 days after stratification (DAS). The upper panels and lower panels show the middle area and tip area of the cotyledons, respectively (left image). Arrowheads and asterisks indicate stomata and hydathode pores, respectively. Bars, 100 µm (**b**, confocal images), and 1 mm (**b**, left).

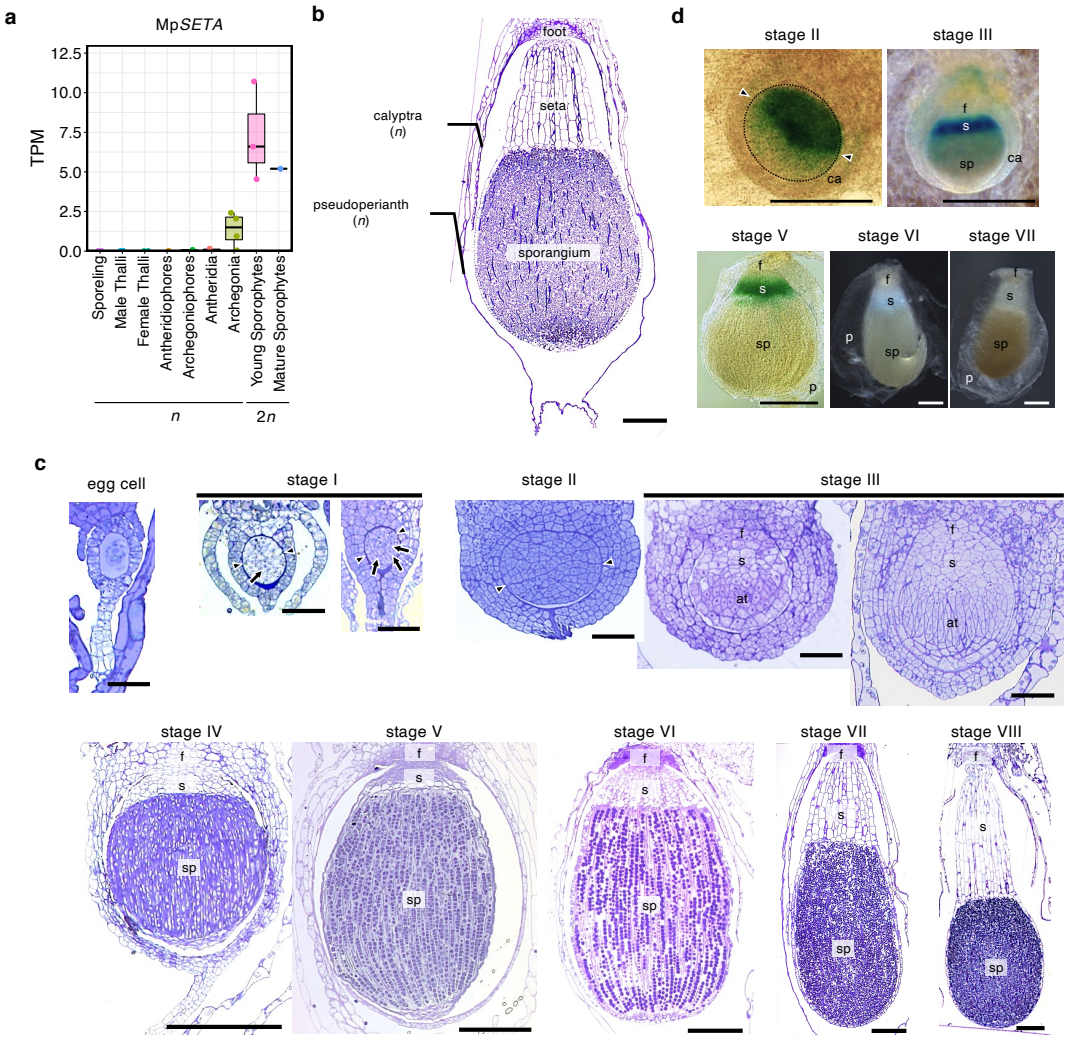


Fig. 2 | Mp*SETA* is preferentially expressed in developing sporophyte. a, Box plot showing the expression profiles of Mp*SETA* across nine *M. polymorpha* tissues. Y-axis shows transcripts per million (TPM). Sporeling and thalli are vegetative gametophytic (*n*) organs. Antheridiophores and archegoniophores are haploid male and female reproductive receptacles, respectively. Antheridia and archegonia are the organs that produce sperms and egg cells, respectively. Sporophytes are diploid (2*n*) organs developed after fertilization. **b**, Tissue section of mature *M. polymorpha* sporophyte (stage VII). **c**, Developmental stages of *M. polymorpha* sporophytes after 7–40 days' post-fertilization. (I) sporophyte differentiating amphithecium and endothecium, (II) later globular sporophyte, (III) sporophyte differentiating foot, seta, and archesporial tissues, (IV) sporophyte differentiating sporogenous cells and elaterocytes (or elater mother cell), (V) sporophyte differentiating sporocytes (or spore mother cells), (VI) sporophyte differentiating spore tetrads after meiosis, (VII) mature sporophyte before the seta elongation, (VIII) mature sporophyte after seta elongation. F, foot; s, seta; at, archesporial tissue; sp, sporangium. Arrows indicate the endothecium. Arrowheads indicate the cell wall of the first cell division. See also Extended Data Fig. 3 for the developmental process of seta and foot. **d**, Histochemical detection of β-glucuronidase (GUS) activity driven by Mp*SETA* promoter in the developing seta region from stage II to stage VII. f, foot; s, seta; at, archesporial tissue; sp, sporangium; ca, calyptra; p, pseudoperianth. Arrowheads indicate the cell wall of the first cell division. Bars, 50 µm (**b,** and **c**, top), and 100 µm (**c**, bottom, and **d**).

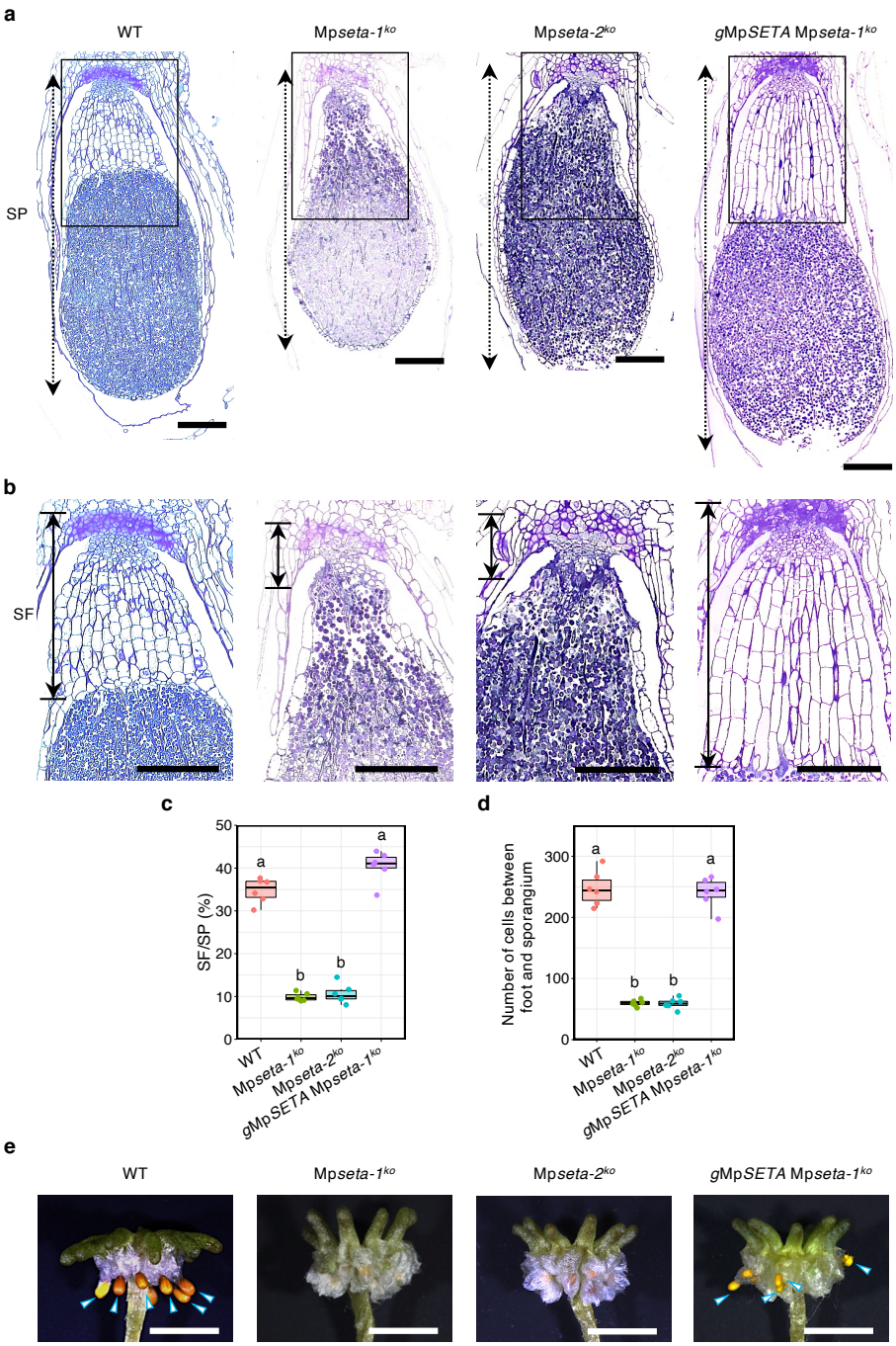
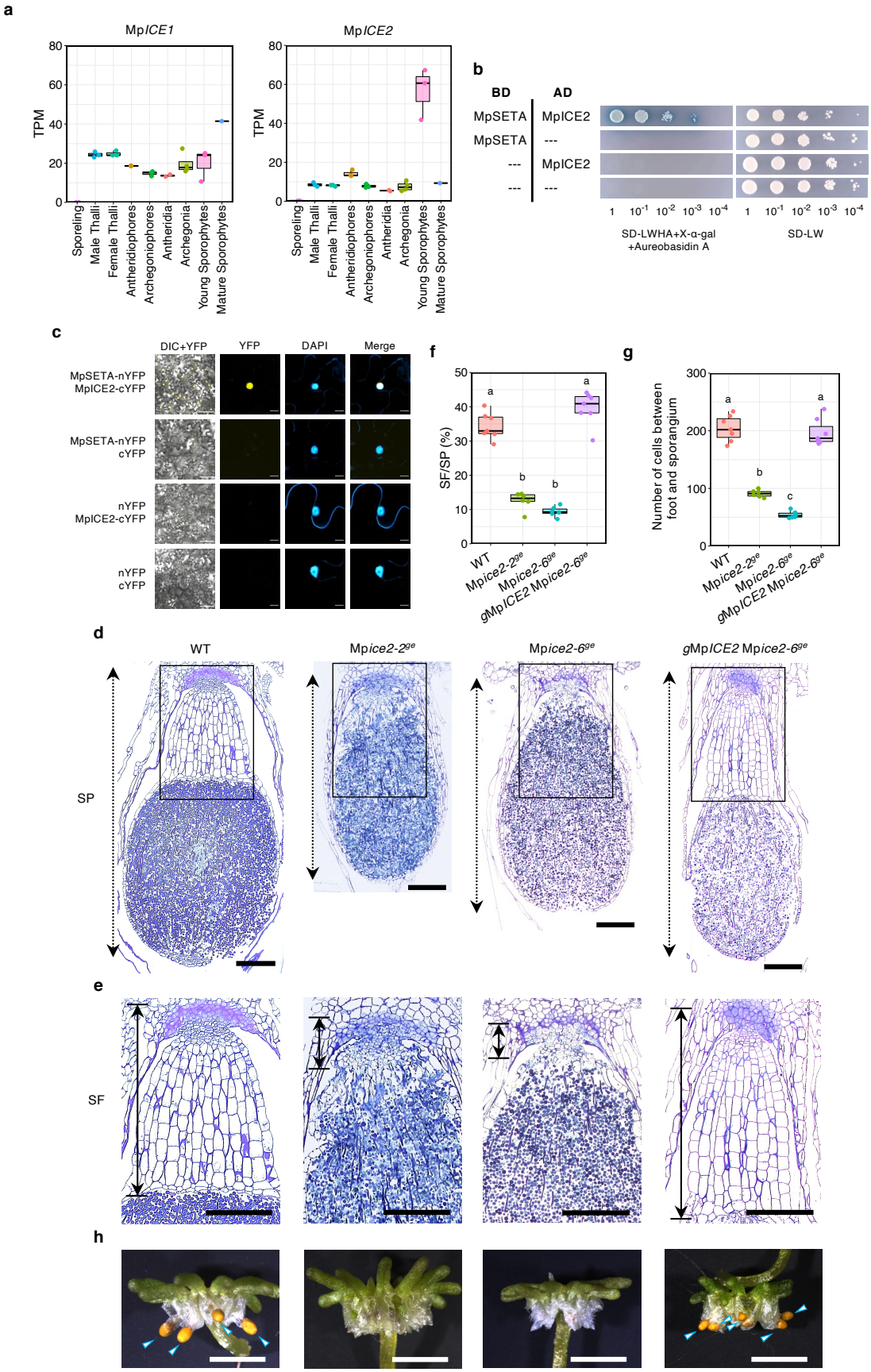


Fig. 3 | Mp*seta*^{ko} mutants show developmental defects in the seta. a,**b**, Longitudinal sections of sporophytes of indicated genotypes. Enlarged images of the square areas are shown in (**b**). Dashed and solid arrows indicate the length of the sporophyte (SP) and the length from the boundary between gametophyte and sporophyte to the proximal side of sporangium (SF), respectively. **c**,**d**, Quantitative data showing the SF/SP length ratio and the number of cells in the seta region. Since the boundary between seta and foot is unclear, cells excluding transfer cells (the cells at the boundary between gametophyte and sporophyte) were used to count the number of cells. The different letters indicate significantly different mean values at $p < 0.01$ (Tukey's HSD test). (n = 6 independent lines). **e**, One month-post-fertilization archegoniophores that produce mature sporophytes. Arrowheads indicate sporangia exposed to the outside of calyptra due to elongation of seta cells. Bars, 100 µm (**a**,**b**), and 5 mm (**e**).


a
MpICE1
MpICE2
TPM
Sporeling
Male Thalli
Female Thalli
Antheridiophores
Archegoniophores
Antheridia
Archegonia
Young Sporophytes
Mature Sporophytes
b
BD
AD
MpSETA
MpICE2

SD-LWHA+X-α-gal
+Aureobasidin A
SD-LW
c
DIC+YFP
YFP
DAPI
Merge
MpSETA-nYFP
MpICE2-cYFP
cYFP
nYFP
f
SF/SP (%)
WT
Mpice2-2ge
Mpice2-6ge
gMpICE2 Mpice2-6ge
g
Number of cells between foot and sporangium
d
SP
e
SF
h

Fig. 4 | MpICE2 regulates setal development in cooperation with MpSETA. a. Box plot showing the expression profiles of Mp*ICE1* and Mp*ICE2* across nine *M. polymorpha* tissues. Y-axis shows transcripts per million (TPM). **b**, Y2H assays showing the interaction between MpSETA and MpICE2. MpSETA fused with GAL4 DNA-binding domain (DBD) was used as bait, while MpICE2 fused with GAL4 activation domain (AD) was used as prey. DBD alone and AD alone were used as the negative control. **c**, BiFC assays showing the interaction between MpSETA and MpICE2 in *N. benthamiana* leaf epidermal cells. MpSETA was fused to the N-terminal fragment of EYFP (nYFP), while MpICE2 was fused to the C-terminal fragment of EYFP (cYFP). nYFP alone and cYFP alone were used as the negative control. Nuclei were stained by 4',6-diamidino-2-phenylindole (DAPI). **d**,**e**, Longitudinal sections of sporophytes of indicated genotypes. Enlarged images of the square areas are shown in (**e**). Dashed and solid arrows indicate the length of the sporophyte (SP) and the length from the boundary between gametophyte and sporophyte to the proximal side of sporangium (SF), respectively. **f**,**g**, Quantitative data showing the SF/SP length ratio and the number of cells in the seta region. Since the boundary between seta and foot is unclear, cells excluding transfer cells (the cells at the boundary between gametophyte and sporophyte) were used to count the number of cells. The different letters indicate significantly different mean values at $p < 0.01$ (Tukey's HSD test) ($n = 6$ or 7 independent lines). **h**, One month-post-fertilization archegoniophores that produce mature sporophytes. Arrowheads indicate sporangia exposed to the outside of calyptra due to elongation of seta cells. Bars, 10 µm (**c**, YFP, DAPI, and Merge), 100 µm (**c**, left, **d**, and **e**), and 5 mm (**f**).

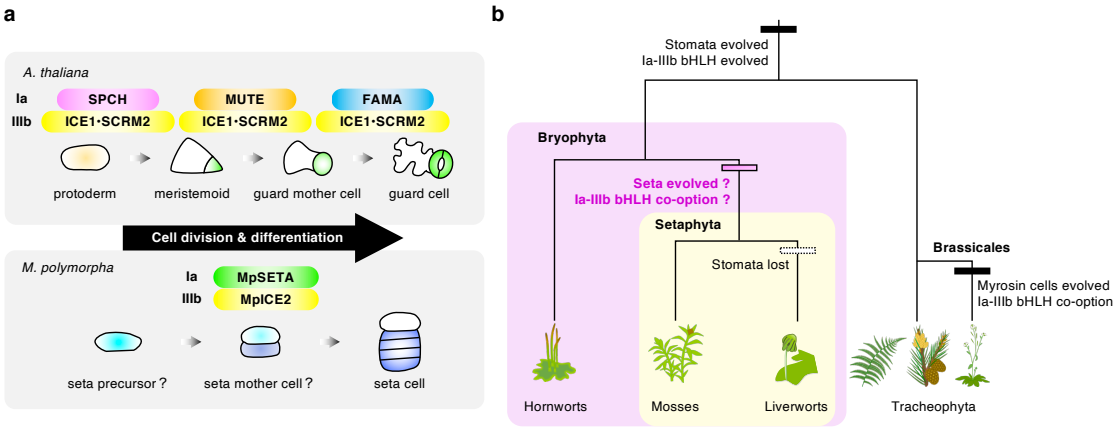
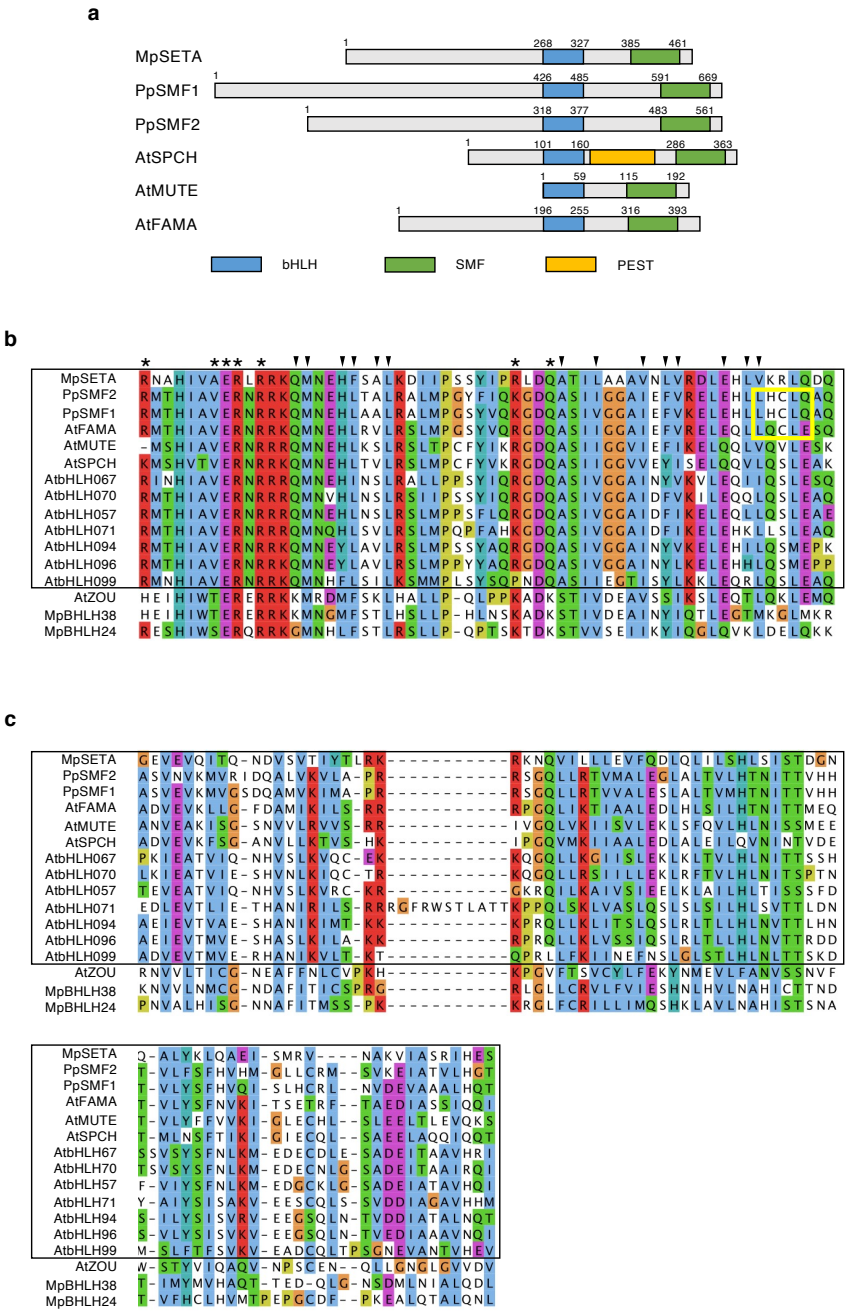
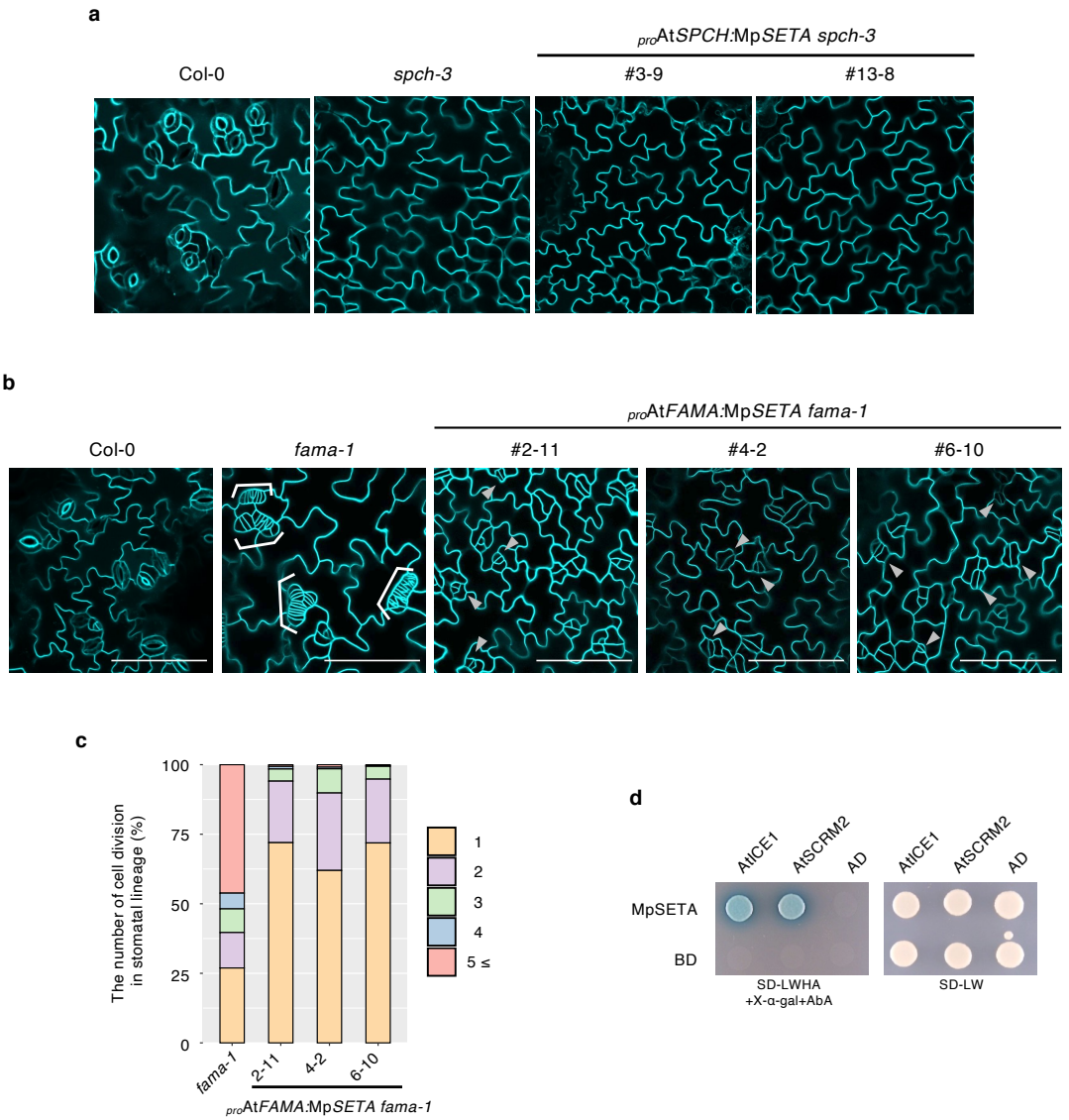


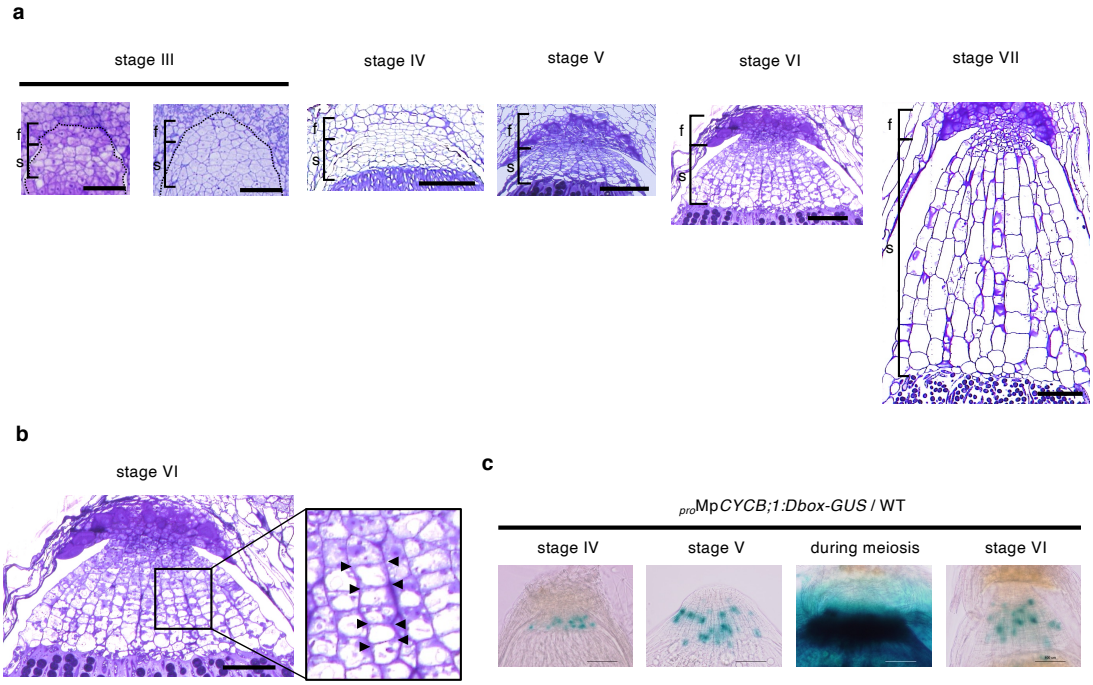
Fig. 5 | Function and co-option of the Ia-IIIb bHLH module during the evolution of land plants. a, Schematic model comparing the molecular functions of the Ia-IIIb bHLH TF modules in *A. thaliana* and *M. polymorpha* during cell fate determination. In *A. thaliana*, the heterodimer of Ia bHLHs (SPCH, MUTE, and FAMA) and IIIb bHLHs (ICE1 and SCRM2) control the development of stomata. In *M. polymorpha*, the heterodimer of Ia bHLH (MpSETA) and IIIb bHLH (MpICE2) regulates cell differentiation and cell division in the seta precursor transition and might directly or indirectly be involved in the symmetric division of a putative seta mother cell. **b**, An evolutionary model for the Ia-IIIb bHLH TF module. Co-option of the Ia-IIIb bHLH module might have occurred multiple times independently during the evolution of land plants. First, stomata and a transcriptional module consisting of Ia-IIIb bHLHs evolved in the common ancestor of land plants. In the ancestral plant, the Ia bHLHs may have had MUTE- and FAMA-like functions. Second, the Ia-IIIb bHLH TF module might have been co-opted to regulate setal development in the ancestor of the Setaphyta. Third, after mosses and liverworts diverged, the common ancestor of liverworts lost its stomata. A co-option of the Ia-IIIb bHLH module occurred in the Brassicales plants for regulating myrosin idioblast development.



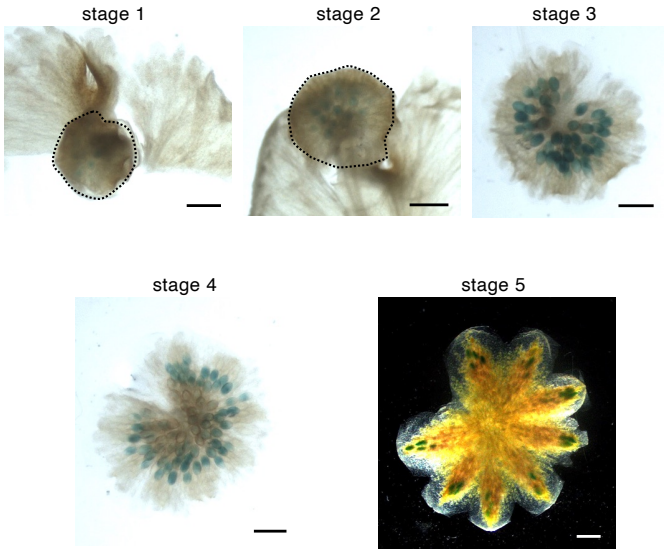
Extended Data Fig. 1 | Comparison of the domain architecture of Ia bHLHs from *M. polymorpha*, *P. patens*, and *A. thaliana*. a, A diagram of the domain architecture of MpSETA (*M. polymorpha*), PpSMF1, PpSMF2 (*P. patens*), AtSPCH, AtMUTE, and AtFAMA (*A. thaliana*). While no PEST domain was identified, MpSETA has a bHLH domain and SMF domain conserved at the C-terminus like other Ia bHLH proteins. SMF domain is structurally considered to be the ACT-like domain, which is a putative domain for protein-protein dimerization. **b**, Sequence alignment of the bHLH domain of Ia bHLH proteins. Ia bHLHs are surrounded by a black box, and others are Ib(1) bHLHs . Asterisks indicate amino acids that are assumed to be important for binding to the E-box (CANNTG), and the triangles indicate amino acids that are assumed to be important for dimerization of the bHLH domain. The yellow box indicates the LxCxE motif, which is a binding motif with Retinoblastoma-related (RBR). **c**, Sequence alignment of the C-terminal SMF domain of Ia bHLH proteins.



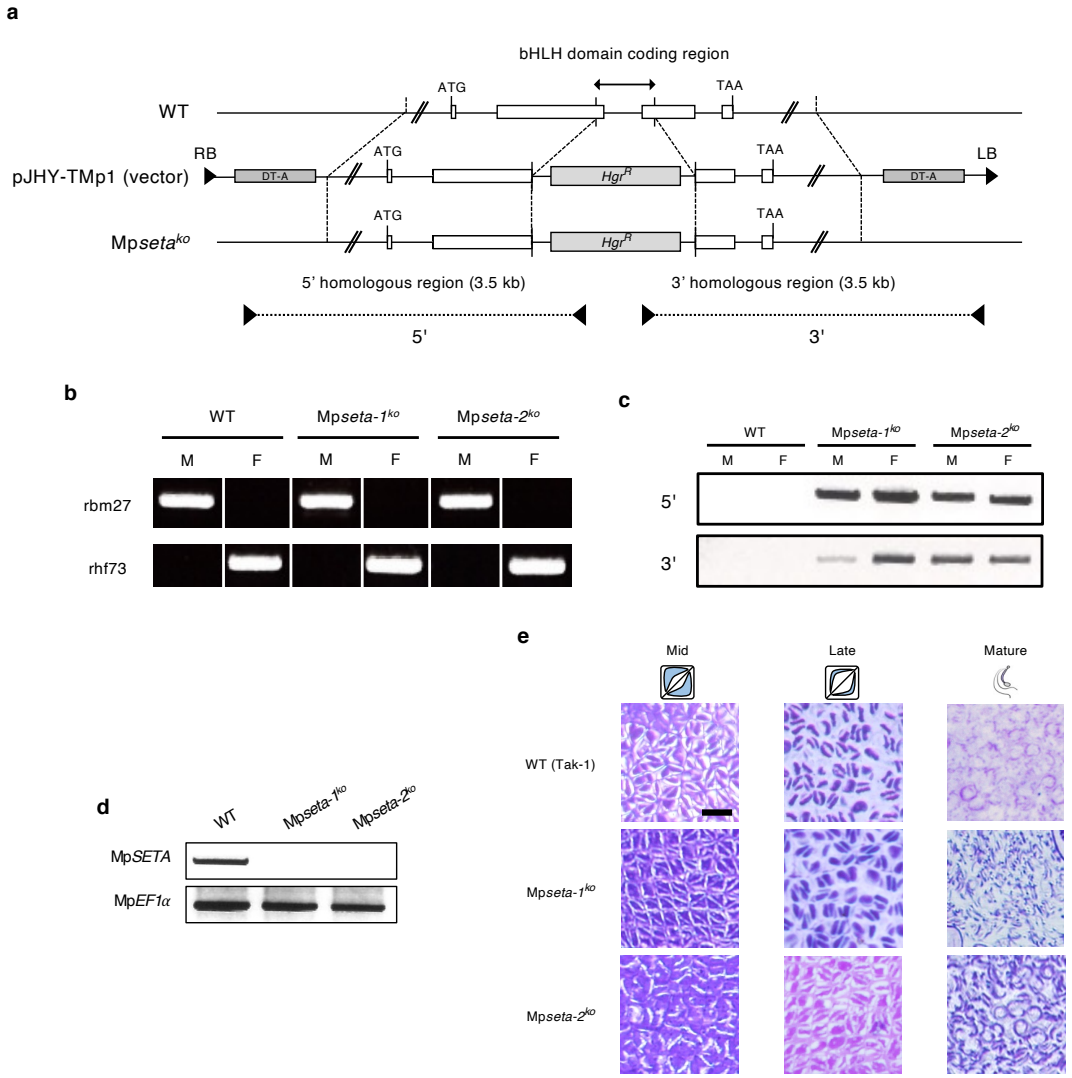
Extended Data Fig. 2 | Function of Mp*SETA* in *A. thaliana* Ia bHLH mutants. a, Confocal images of *A. thaliana* abaxial cotyledons of wild type (Col-0), *spch-3*, and $_{pro}$At*SPCH*:Mp*SETA spch-3* at 9 days after stratification (DAS). **b**, Confocal images of *A. thaliana* abaxial cotyledons of wild type (Col-0), *fama-1*, and $_{pro}$At*FAMA*:Mp*SETA fama-1* at 9 DAS. Brackets and arrows indicate *fama* tumors and stomatal-lineage cells, respectively. **c**, Quantitative data of the distribution of the number of cell divisions that occurred in the stomatal lineage in each genotype. (n > 320 cells per genotype, 9 DAS cotyledons). **d**, Y2H assays in which the MpSETA fused with GAL4 DNA-binding domain (DBD) was used as bait, and the AtICE1 and AtSCRM2 fused with GAL4 activation domain (AD) were used as prey. DBD alone and AD alone were used as the negative controls.



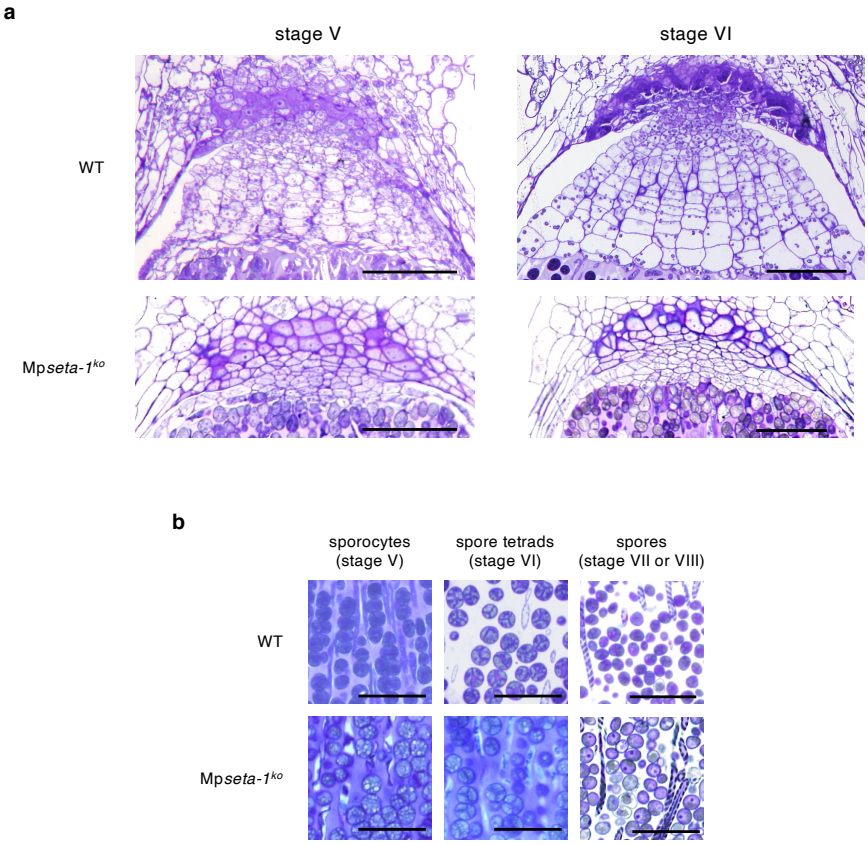
Extended Data Fig. 3 I Development of seta in *M. polymorpha*. a, Magnified images showing the foot and seta for each stage of wild-type sporophytes. Same images as used in Fig. 2c. **b**, The formation of the cell files of seta in the wild type stage VI sporophyte. Arrowheads indicate the plane of proliferative cell divisions. **c**, Histochemical detection of β-glucuronidase (GUS) activity in the sporophytes crossed with male wild type and female $_{pro}$Mp*CYCB;1:Dbox-GUS*. Bars, 100 µm.



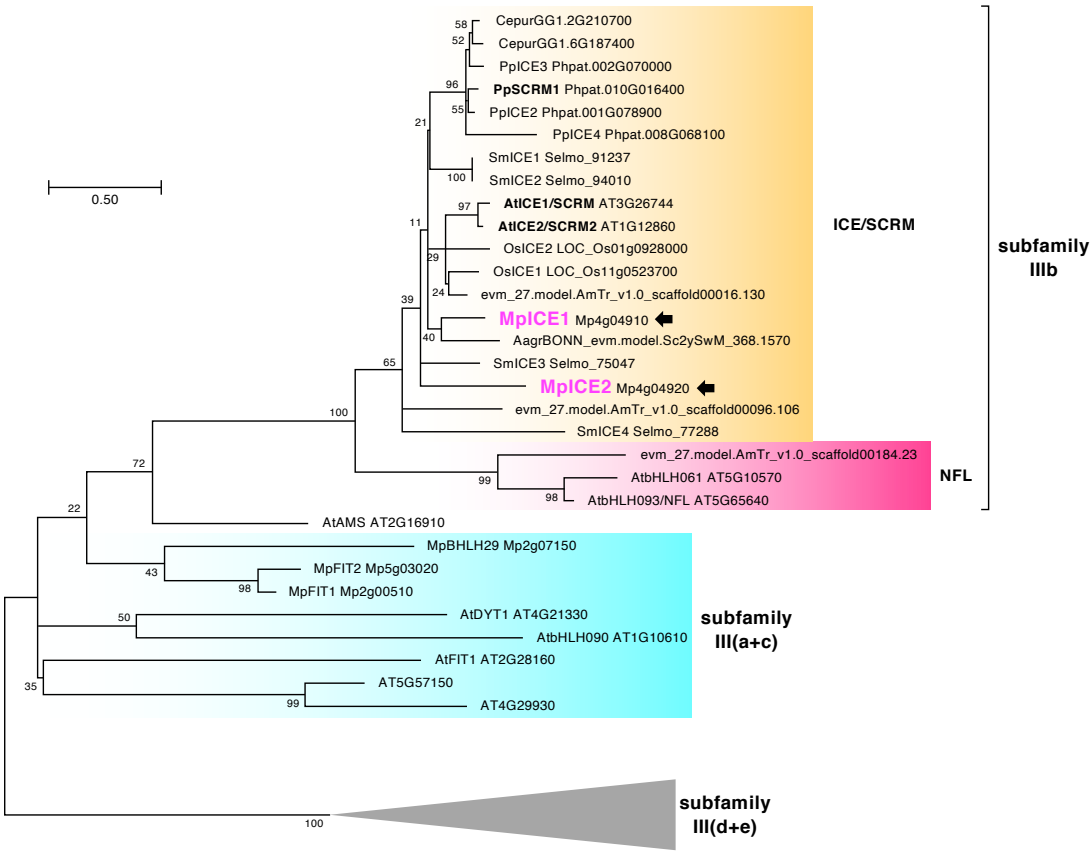
Extended Data Fig. 4 | Expression analysis of Mp*SETA* in the gametophytic tissues. Histochemical detection of β-glucuronidase (GUS) activity driven by Mp*SETA* promoter in the developing antheridia. Bars, 1 mm.



Extended Data Fig. 5 | Generation and phenotypes of Mp*SETA* knock-out lines. a, Structure of the Mp*SETA* locus disrupted by homologous recombination. Knock-out lines have a deletion in the bHLH domain coding region. White boxes indicate the exons of the MpSETA coding sequence. DT-A, diphtheria toxin A fragment gene; Hgr^R, hygromycin resistant gene. **b**, Genotyping of the Mp*seta*^{ko} lines used in this study to distinguish the sex. rbm27, a male-specific marker; rhf73, a female-specific marker. **c**, Genotyping of the Mp*seta*^{ko} lines. The position of primers used for PCR is shown in (**a**). M, Male; F, Female. **d**, RT-PCR to confirm the loss of the full-length Mp*SETA* transcript in Mp*seta*^{ko} lines in 21 DPF sporophytes. Mp*EF1α* was used as an internal control. **e**, Spermatogenesis process in WT and Mp*seta*^{ko} lines. All images are at the same scale. Bars, 10 µm (**e**).

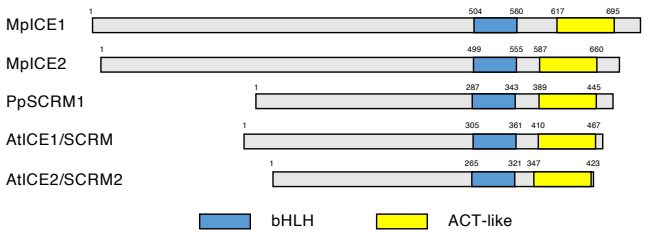


Extended Data Fig. 6 | Comparative analyses of setal or spore development between wild type and Mp*seta-1*^{ko}. a, Longitudinal sections of stage V and stage VI Mp*seta-1*^{ko} sporophytes. **b**, Sporogenesis process in wild type and Mp*seta-1*^{ko}. Sporocytes are the cells before the pre-meiotic stage, and spore tetrads are the cells immediately after meiosis. All images are at the same scale. Bars, 100 µm (**a**) and 50 µm (**b**).

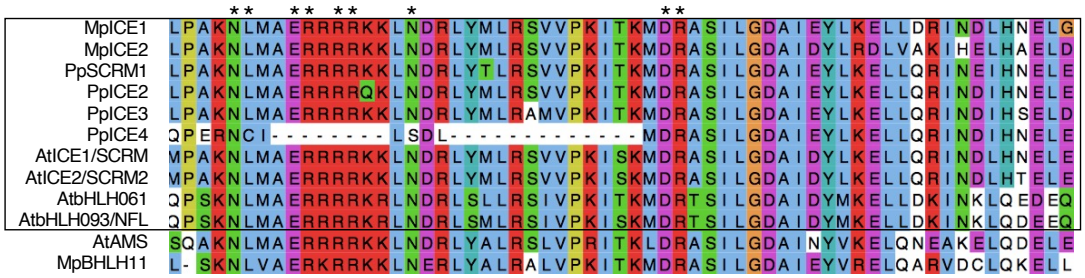


Extended Data Fig. 7 | Phylogenetic tree of IIIb bHLH TFs. A maximum-likelihood bHLH phylogenetic tree of subfamilies IIIb, III (a+c) (light blue), and III(d+e) (outgroup) is shown. Numbers at branches indicate bootstrap values calculated from 1,000 replicates. IIIb bHLHs are divided into 2 groups: ICE/SCRM clade (orange) and NFL clade (magenta). Species are abbreviated as follows: Mp, *M. polymorpha* (liverwort); Pp, *P. patens* (moss); Cepur, *Ceratodon purpureus* (moss); Aagr, *Anthoceros agrestis* (hornwort); Sm, *Selaginella moellendorffii* (lycophyte); AmTr, *Amborella trichopoda* (basal angiosperm); Os, *Oryza sativa* (monocot); At, *A. thaliana* (dicot). Arrows indicate MpICE1 (Mp4g04910) and MpICE2 (Mp4g04920). For the phylogenetic construction of subfamilies III(a+c) and III(d+e), we used the amino acid sequences from only *A. thaliana* and *M. polymorpha*.

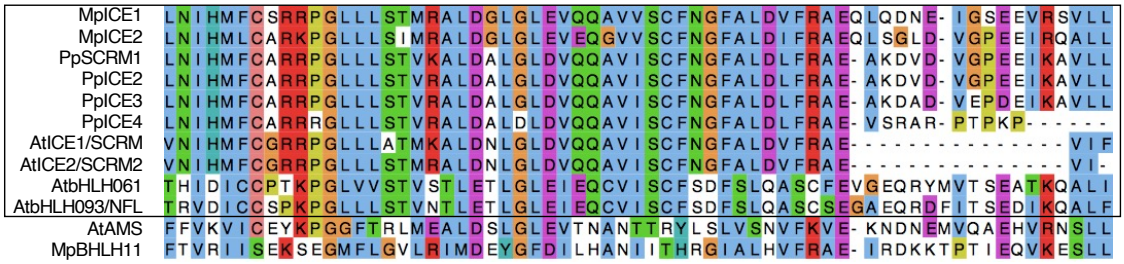
a



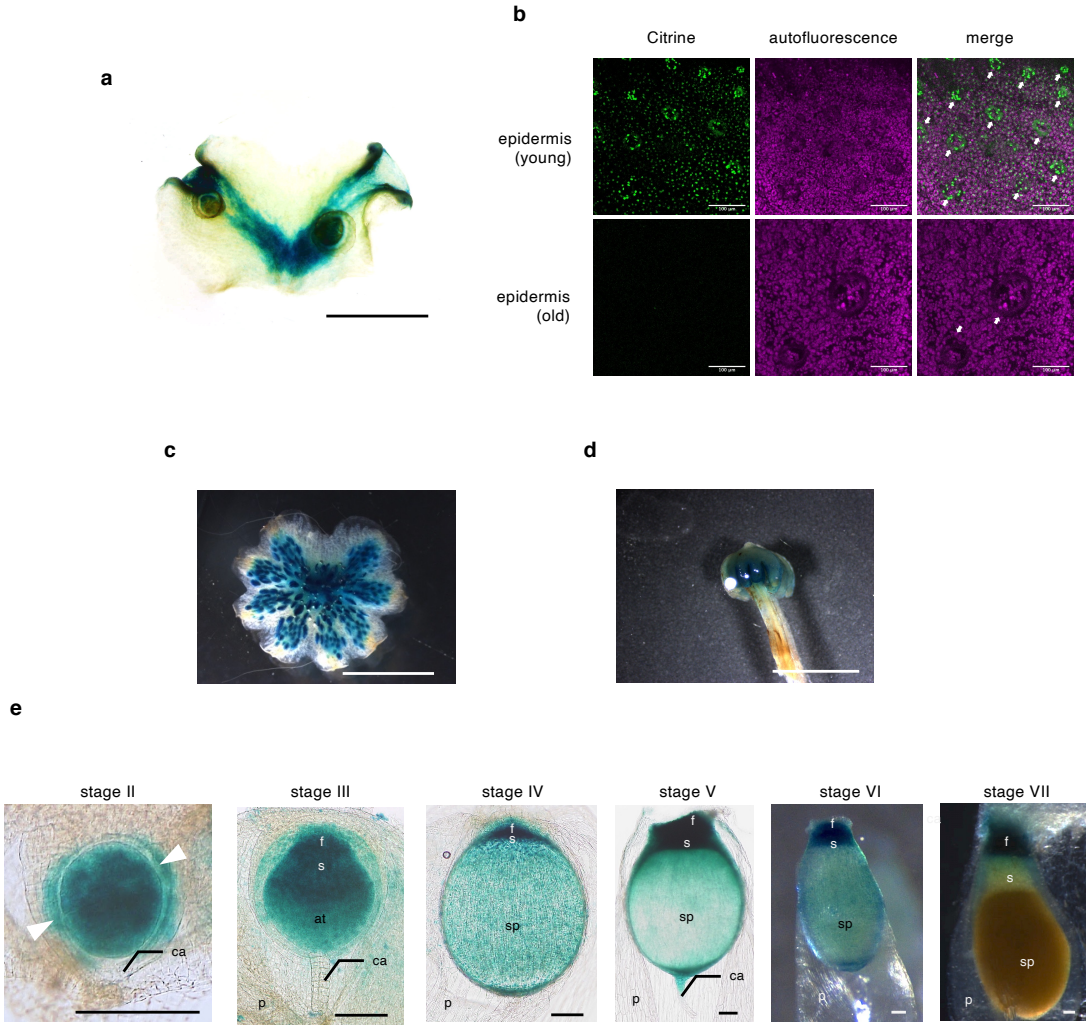
b



c

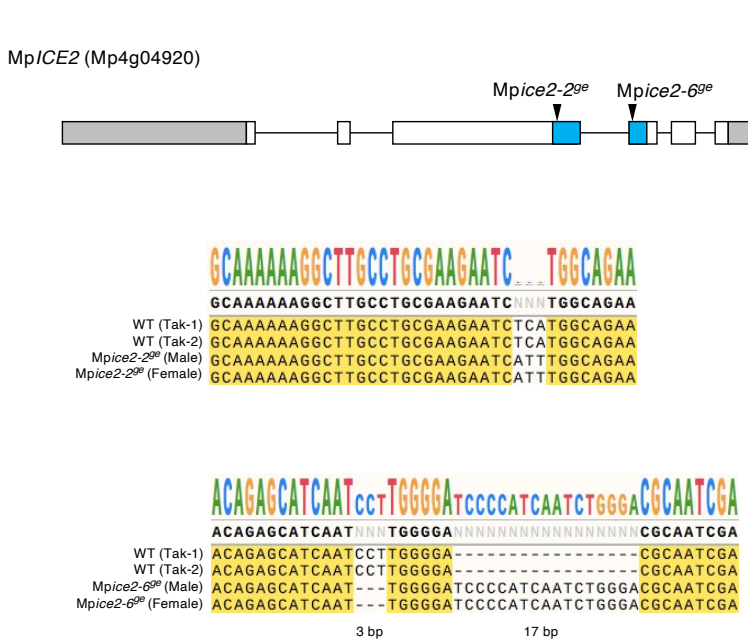


Extended Data Fig. 8 | Comparison of the domain architecture of IIIb bHLHs from *M. polymorpha*, *P. patens*, and *A. thaliana*. **a**, A diagram of the domain architecture of MpICE1, MpICE2 (*M. polymorpha*), PpSCRM1 (*P. patens*), AtICE1, and AtSCRM2 (*A. thaliana*). MpICE1 and MpICE2 have a bHLH domain and ACT-like domain conserved at the C-terminus like other IIIb bHLH proteins. **b**, Sequence alignment of the bHLH domain of IIIb bHLH proteins. IIIb bHLHs are surrounded by a black box, and others are outgroup. Asterisks indicate amino acids that are assumed to be important for binding to the E-box (CANNTG). **c**, Sequence alignment of the C-terminal ACT-like domain of IIIb bHLH proteins.



Extended Data Fig. 9 | The expression analysis of Mp*ICE2*. **a**, Histochemical detection of β-glucuronidase (GUS) activity driven by Mp*ICE2* promoter in the vegetative thallus. **b**, Confocal images of the dorsal epidermis of $_{pro}$Mp*ICE2:Citrine-GUS-NLS* line. Upper and lower panels indicate the epidermis around the apical notch and the epidermis around the midrib, respectively. Arrows indicate the air pores. **c**,**d**, Histochemical detection of GUS activity driven by Mp*ICE2* promoter in the gametophytic reproductive organs. An antheridiophore (**c**) and an archegoniophore (**d**) are shown. **e**, Expression pattern of Mp*ICE2* in the developing sporophytes. f, foot; s, seta; at, archesporial tissue; sp, sporangium; ca, calyptra; p, pseudoperianth (*n*). Arrowheads indicate the cell wall of the first cell division. Bars, 5 mm (**c** and **d**), 100 μm (**b** and **e**).

a



b

Extended Data Fig. 10 | Generation of Mp*ice2* mutants by CRISPR/Cas9. a, Schematic representation of the Mp*ICE2* gene and the resulting mutations in the obtained CRISPR/Cas9-generated alleles. Gray, white, and blue boxes indicate the coding sequences (CDS), the untranslated regions (UTR), and the bHLH domain coding region, respectively. **b**, Sequence alignment of putative translational products of wild type and Mp*ice2*^{ge} mutants. Asterisks indicate the amino acids that are assumed to be important for binding to E-box.