Published in final edited form as:
Evol Dev. 2011 ; 13(2): 182–192. doi:10.1111/j.1525-142X.2011.00468.x.

Sequence and function of bHLHs required for stomatal development in *Arabidopsis* are deeply conserved in land plants

Cora A. MacAlister[1,*] and **Dominique C. Bergmann**
Department of Biology, Stanford University, Stanford, California, 94305-5020, USA

Abstract

Stomata are a broadly conserved feature of land plants with a crucial role regulating transpiration and gas exchange between the plant and atmosphere. Stereotyped cell divisions within a specialized cell lineage of the epidermis generate stomata and define the pattern of their distribution. The behavior of the stomatal lineage varies in its detail among different plant groups, but general features include asymmetric cell divisions and an immediate precursor (the guard mother cell, GMC) that divides symmetrically to form the pair of cells that will differentiate into the guard cells. In *Arabidopsis*, the closely related basic helix-loop-helix subgroup Ia transcription factors SPEECHLESS, MUTE and FAMA promote asymmetric divisions, the acquisition of GMC identity and guard cell differentiation, respectively. Genome sequence data indicate that these key positive regulators of stomatal development are broadly conserved among land plants. While orthologies can be established among individual family members within the angiosperms, more distantly related groups contain subgroup Ia bHLHs of unclear affinity. We demonstrate group Ia members from the moss *Physcomitrella patens* can partially complement MUTE and FAMA and recapitulate gain of function phenotypes of group Ia genes in multiple steps in the stomatal lineage in *Arabidopsis*. Our data are consistent with a mechanism whereby a multifunctional transcription factor underwent duplication followed by specialization to provide the three (now non-overlapping) functions of the angiosperm stomatal bHLHs.

Keywords

stomata; bHLH transcription factor; Physcomitrella; Arabidopsis; plant evolution

Introduction

In the colonization of land, plants faced significant new challenges, particularly the threat of desiccation. The secretion of a waxy cuticle effectively prevents evaporation, but simultaneously limits acquisition of carbon dioxide from the atmosphere. This dilemma is solved in above-ground organs by the production of stomata, epidermal structures consisting of a pore flanked by a pair of guard cells. Changes in turgor pressure in the guard cells open and close the pore in response to changes in environmental and physiological conditions. Stomata were an ancient innovation of land plants dating back 400 million years in the fossil record, pre-dating even the evolution of leaves and roots (Ziegler 1987, Peterson et al. 2010). In extant plants, stomata are among the most conserved vegetative plant characters, though their patterned distribution on the surfaces of different organs is sufficiently varied that pattern commonly serves as a taxonomic trait (Edwards et al. 1998, Beerling 2005) (fig. 1).

[*]Corresponding author: Cora A. MacAlister, cmacalis@cshl.edu, (516)367-8330, fax (516)367-8369.
[1]Present address: Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 11724-2209, USA

The guard cells are produced in the epidermis by a series of stereotyped divisions varying in number and complexity. In the most primitive case (for example in the mosses) a young epidermal (protodermal) cell divides asymmetrically to form a small cell which can function directly as the guard mother cell (GMC) (Zeigler 1987, Payne 1979). The GMC divides symmetrically, sometimes incompletely, to form the two guard cells (Sack and Paolillo 1985). In other plants, the small cell created by protodermal division functions first as a meristemoid, undergoing additional rounds of asymmetric division (amplifying divisions), before developing into a GMC (fig. 1e). Amplifying divisions may be necessary to produce subsidiary cells adjacent to the stoma, or to establish spacing and pattern. Subsidiary cells are morphologically distinct from other epidermal cells and are specialized to assist the guard cells in opening and closing. In grasses, subsidiary cells have a different developmental origin; they are formed through highly polarized cell divisions in cells adjacent to the GMC (Hernandez et al. 1999).

Work in Arabidopsis has identified a trio of related basic helix-loop-helix (bHLH) type transcription factors, SPEECHLESS (SPCH), MUTE and FAMA, that are essential for stomatal production. Loss of function of any one of these proteins results in arrest of stomatal lineage cells, though at different stages. SPCH is required for asymmetric divisions, MUTE specifies GMC identity and FAMA is essential for the differentiation of the guard cells (fig. 1e) (MacAlister et al. 2007, Pillitteri et al. 2007 and Ohashi-Ito and Bergmann 2006). In addition, a pair of more distantly related bHLHs, INDUCER OF CBF EXPRESSION (ICE)/SCREAM (SCRM) and SCRM2 physically interact with SPCH, MUTE and FAMA in a plant bimolecular fluorescence complementation (BiFC) assay, and loss of SCRM/SCRM2 activity affects stomatal development (Kanaoka et al. 2008).

The bHLH family is one of the largest transcription factor families and contains at least 158 members in *Arabidopsis* (Pires and Dolan 2009). While all family members share the structural motif that gives the family its name, many of these proteins also bear other distinctive functional and conserved domains, allowing the family to be subdivided into 26 subgroups (Pires and Dolan 2009). SPCH, MUTE and FAMA form a single clade that, along with seven other proteins whose biological roles have not yet been defined, comprise *Arabidopsis* bHLH subgroup Ia (Ohashi-Ito 2007; Heim et al. 2003, Pires and Dolan 2009). Subgroup Ia members are united by a very similar bHLH domain with two additional regions of high sequence conservation. The first region is an extension immediately following the bHLH domain (dark blue in fig. 2b) whose degree of sequence conservation within this family is as high as that within the canonical bHLH domain. Since this extension after the bHLH domain is an invariant feature of the subgroup Ia bHLHs, we will treat the two as a single entity. The second conserved region of the subgroup Ia bHLHs is a shared domain at the C-terminus of the protein (Toledo-Ortiz et al. 2003 and Heim et al. 2003). This plant specific domain (referred to here as the SMF (SPCH, MUTE and FAMA) domain) is always found C-terminal to a bHLH domain. Most Arabidopsis Ia members with the exception of SPCH and MUTE also have a conserved region of 18 amino acids immediately N-terminal to the bHLH (fig 2b). Feller et al. (2006) also report the presence of a degenerate aspartate kinase-chorismate mutase-tyrA (prephenate dehydrogenase) or ACT-like domain in more than 30 Arabidopsis bHLH proteins from several subgroups including all subgroup Ia members except MUTE. In subgroup Ia the ACT-like domain appears in between the bHLH and SMF domains. The primary sequence conservation, however, is poor and a higher stringency motif-based search did not identify the ACT domain as a conserved feature of subgroup Ia (Pires and Dolan 2009).

The broader functioning bHLHs, ICE1/SCRM and SCRM2, are members of the four member group IIIb clade. Group IIIb genes in *Arabidopsis* have a clade-specific domain C-terminal to the bHLH domain (Heim et al. 2003 and Pires and Dolan, 2009). Another

member of this group, bHLH093, has also been shown to physically interact with FAMA (Ohashi-Ito and Bergmann 2007), but the fourth (bHLH061) is uncharacterized.

Stomata are an excellent natural laboratory for understanding developmental evolution: they are likely monophyletic, having evolved a single time in the ancestor of all extant stomata bearing plants (Raven 2002); they require only a few stereotyped divisions to form; they are essential, and they are extremely well conserved across a broad evolutionary distance. The *Arabidopsis* group Ia stomatal bHLHs are a paradox; stomata are essential, yet without any one of the three (SPCH, MUTE and FAMA) stomata are not produced. How then could SPCH, MUTE and FAMA have evolved? Furthermore, if the *Arabidopsis* group Ia genes function as dimers with group IIIb genes, when did this relationship develop? Understanding the evolutionary history and diversity of these groups of bHLHs will be necessary to answer such questions. We have made a detailed comparative analysis of the SPCH, MUTE and FAMA sequences from a number of different plant species, with specific focus on the sequences from the distantly related moss *Physcomitrella patens*, and extended this with a functional analysis of the moss genes to generate a model for the evolution of this family.

Materials and Methods

Identification of SPCH, MUTE and FAMA homologues

Putative homologs of the stomatal bHLHs were identified from plants with a sequenced genome, or for which considerable other sequence data is available. The SPCH, MUTE or FAMA protein sequences were used as the query in a series of BLASTp or tBLASTn searches against protein or translated nucleotide databases, respectively. The *Ricinus communis* genome was searched from the TIGR portal for that species. Other angiosperm genomes were searched through NCBI and Phytozome v5.0 (http://www.phytozome.net) by similarity search as well as by genome synteny with the *Arabidopsis* genes using the Plant Genome Duplication Database (Tang et al. 2008). The rice and maize SPCH, MUTE and FAMA sequences have been previously reported (Liu et al. 2009).

The *Selaginella* and *Physcomitrella* sequences were identified by protein BLAST searches of the filtered, non-redundant protein models at JGI using all members of *Arabidopsis* group Ia as query. The *Selaginella* database was also searched with the Ia members from *Physcomitrella* and vice versa. In all cases, the same subgroup Ia members were identified indicating that all Ia members from these genome databases have been identified. Significant hits from bHLHs outside of subgroup Ia were discounted based on the absence of the SMF domain. These hits always exhibited weaker similarity in the bHLH domain and were always less significant than group Ia members. A search of the *Arabidopsis* genome by the *Physcomitrella* or *Selaginella* sequences identifies FAMA as the most similar to all of the *Physcomitrella* and *Selaginella* Ia proteins. Similar searches were made within the genomes of the algae *Chlamydomonas reinhardtii* and *Ostreococcus sp*. from which no subgroup Ia bHLHs were identified.

The predicted gene models for PpSMF2 and SmSMF3 encoded proteins missing the N-terminus and a significant portion of the bHLH domain, respectively. Manual re-examination of the genomic sequence found in-frame start codons preceding the predicted start sites. Use of the earlier start codons restored the missing portions of the proteins giving PpSMF2 a nearly identical N-terminus to PpSMF1 and giving SmSMF3 a complete bHLH domain as well as the N-terminal bHLH extension characteristic of most of the *Arabidopsis* Ia bHLHs. The predicted transcript for PtSPCH2 was found to omit portions of the MPKT domain and the N-terminus due to differences in the prediction of intron boundaries. GENSCAN (Burge and Karlin 1997) was used with the genomic sequences to predict start

sites and intron boundaries. In all cases the modified gene models were used in subsequent analysis.

Sequence alignments and tree construction

Multiple sequence alignments were generated using Clustal W (Larkin et al. 2007). Percent similarity calculations between two proteins were made using EMBOSS Pairwise Alignment Algorithms using a global alignment (needle) with a blossom 62 similarity matrix. Following sequence alignment, maximum parsimony phylogenetic trees were calculated with 1000 replicate bootstrap values by PHYLIP protpars via the Mobyle portal (Bertrand et al. 2009). At3g56980, a member of *Arabidopsis* bHLH subgroup Ib was used as an outgroup.

Growth conditions

Phycomitrella patens was grown on minimal media using standard culture techniques (Ashton et al. 1979). Production of sporophytes was induced by short photoperiods (8hr light, 16hr dark) and cool temperatures (18°C) in gametophores grown on solid media. For cross-species complementation and overexpression assays, the constructs described below were transformed into *Arabidopsis* ecotype Columbia by agrobacterium mediated transformation (Clough and Bent 1998). For rescue experiments, constructs were transformed into a background segregating the appropriate mutation (*spch-3*/+ (MacAlister et al. 2007) or *fama-1*/+, (Ohashi-Ito and Bergmann 2006)) or were crossed in (*mute-1*/+, Pilliterri et al, 2007). At least 6 T2 lines (20-24 plants/line) were screened for segregation of the mutant phenotype and rescue was determined based on statistically significant deviation from the expected segregation ratio of the mutant phenotype (1:3) based on χ^2 test in 10dpg seedlings. Plants that produced any stomata were scored as complemented. Rescue was further quantified by calculating the mean number of mature guard cells produced per cotyledon surface (20 individual sides) and total hypocotyls per line at 7 dpg. Statistical significance between the transgenic lines and the mutant background was determined by Student's T-test at a 95% confidence level. Estrogen induction of appropriate plant lines was achieved by germination on MS agar media containing 5μM β-estradiol and effects for scored at 12dpg for cotyledons and second leaves and ~22 dpg for leaf 6. For scoring phenotypes in table 2, the number of cells exhibiting phenotypes like those found in SPCH, MUTE or FAMA overexpression lines were quantified from 0.32 mm^2 regions taken from the basal quadrant in the region between the midvein and leaf margin. Lines in which at least half of the individuals scored had greater than 10 examples of the overexpression like phenotypes per region were marked as "+++". Those with mostly 5-10 examples per region were marked "++" and those with less than five per region as "+". or from

DNA manipulations

DNA was extracted from moss protonemal tissue using a modified CTAB extraction procedure (http://biology4.wustl.edu/moss/methods.html). The coding regions of PpSMF1 and 2 were PCR amplified from genomic DNA using the primer pairs "caccatgaaccatttgaggccgaa" and "tcagaattgcagagagtgtagg" and "caccatgcacaatttggagcctaaacg" and "ctactgtagggagtgcggcg" respectively. The resulting fragments were cloned into pENTR/D-TOPO (Invitrogen). For overexpression, coding region entry clones were recombined into pMDC7 (estrogen inducible) and pMDC32 (35S) vectors (Curtis and Grossniklaus 2003). For rescue experiments, the SPCH, MUTE and FAMA promoters described previously (MacAlister et al. 2007, Ohashi-Ito and Bergmann 2006) were inserted into a NotI site preceding the coding regions of PpSMF1 and PpSMF2 and recombined into pMDC99 (Curtis and Grossniklaus 2003) for transformation into *Arabidopsis*. Promoter swaps constructs MUTEpro::FAMA and FAMApro::MUTE were created in the same manner, inserting the appropriate Not1 digested promoter in pENTR

vectors containing the FAMA and MUTE cDNAs described previously (MacAlister et al. 2007, Ohashi-Ito and Bergmann 2006). Expression of the MUTEpro::FAMA and FAMApro::MUTE transgenes were confirmed by RT-PCR using a reverse primer from the coding sequence and a forward primer from the promoter's 5′ UTR region. The primer sequences were "ttgaccttcattagacttaagc" for the FAMA 5′ UTR, "ctcagaagagaatcttggcg" for the MUTE 5′ UTR, "agttgttgtcgttgtcatcg" for the FAMA coding region and "cagagatgatctttacgagc" for the MUTE coding region.

Microscopy

Epidermal phenotypes were scored using DIC optics on plant material cleared in 70% ethanol and then overnight in Hoyer's solution (Liu and Meinke 1998) using protocols and equipment described in MacAlister et al. (2007). Confocal images were taken with a Leica SP5 with excitation/emission spectra of 561/591-635 for propidium iodide counterstaining and 561/652-691 for chloroplast autofluorescence. SEM images were taken of fresh material using an FEI Quanta 200 environmental SEM. Images were processed in ImageJ (NIH).

Results

Distribution of group Ia and IIIb among plant lineages

To determine the degree of sequence conservation and the evolutionary history of bHLHs involved in stomatal development, we used available expressed sequence tag (EST) and genome sequence to identify putative orthologues of SPCH, MUTE and FAMA from a variety of plant species including several angiosperms, the gymnosperm *Picea glauca*, the lycophyte *Selaginella moellendorffii* and the moss *Physcomitrella patens*. Our results are generally in agreement with a recent large-scale study of the plant bHLH family found that both Subgroup Ia and IIIb bHLHs can be identified across a broad variety of stomata-making plants but are absent from the unicellular algae Chlamydomonas reinhardtii, Ostreococcus lucimarinus and *O. tauri* (Pires and Dolan 2009).

SPCH, MUTE, and FAMA are conserved within the seed plants—Within the angiosperms, putative orthologues of SPCH, MUTE and FAMA can be clearly identified due to the high degree of sequence conservation and the unique domain architecture that distinguishes these three from other subgroup Ia members. The FAMA-like proteins have two highly conserved regions not found in SPCH and MUTE. The first is found in the N-terminus between the start of the protein and the bHLH domain and the second is immediately N-terminal to the SMF domain (red in fig. 2b). The MUTE-like sequences have a different region of unique conservation in this position (orange in fig. 2b). The MUTE sequences are also united by the absence of any residues preceding the bHLH domain, whereas the FAMA and SPCH-like sequences, as well as the other subgroup Ia members, have N-terminal sequences of varying lengths, often containing signatures of transcriptional activation domains. The primary uniting feature of the SPCH sequences is the presence of the mitogen activated protein (MAP) kinase target domain (MPKTD). This region has been shown to regulate SPCH activity in response to phosphorylation by MAP kinases (Lampard et al. 2008). Although the primary sequence conservation within this region is not high, all of the sequences contain four or five high stringency (P-X-S/T-P) MAP kinase phosphorylation target sites (fig. S3). The SPCH-like sequences also have a conserved stop position. The C-terminus of SPCH is known to be important for function; a mutation leading to the truncation of the last seven amino acids results in the same phenotype as a null allele (MacAlister et al. 2007). SPCH has the shortest C-terminal domain of any of the *Arabidopsis* Ia proteins. This stop site is a distinctive feature, as the coding regions of the SPCH homologues end at nearly the same residue (fig. S3). Although there are no apparent losses of SPCH, MUTE or FAMA in the angiosperm genomes surveyed here, SPCH has

undergone two independent duplications: once in the lineage leading to sorghum, rice and maize (Liu et al. 2009) and again in poplar (fig. 2a).

The gymnosperm and angiosperm lineages diverged approximately 300MYA (Wolf et al. 1989). We have found SPCH, MUTE and FAMA to be well conserved within the angiosperms. Despite much more limited sequence data available for gymnosperms, we have identified a SPCH-like cDNA sequence from *Picea glauca* (white spruce) from the Arborea white spruce gene catalogue. Although more distantly related, PgSPCH exhibits all the hallmarks of SPCH-like identity (fig S4). We also identified a loblolly pine (*Pinus taeda*) EST with high similarity to FAMA (98% identical within the bHLH domain). Although this EST (TIGR transcript assembly CO362181) is not the full coding sequence, it does include the N-terminal extensions to the bHLH found in all subgroup Ia members except SPCH and MUTE as well as the FAMA-unique conserved region preceding the SMF domain.

Subgroup Ia and IIIb bHLHs are found outside the seed plants—In more distantly related groups the primary sequence identity is more limited. Within the genome of the lycophyte *Selaginella moellendorffii* we have identified three members of bHLH subgroup Ia. One of these, SmSMF1, like MUTE, begins immediately at the bHLH domain and clusters with the angiosperm MUTEs in our phylogenetic analysis (fig. 2a). The other two *Selaginella* Ia members (SmSMF2 and 3) possess a more general Ia-like domain architecture including a long N-terminus (402 and 185 amino acids long, respectively) and the N-terminal extension of the bHLH domain that is seen in all *Arabidopsis* Ia bHLHs except MUTE and SPCH (fig. 2 and fig. S4). SmSMF2 and 3 cluster with the angiosperm FAMA sequences (fig. 2a). None of the putative *Selaginella* class Ia genes contain a MPKTD suggesting that SPCH is not conserved and likely evolved after the divergence of the lycophytes and seed plant lineages.

The *Physcomitrella patens* genome encodes only two subgroup Ia bHLHs. These two are 69% identical at the protein sequence level and share the same domain architecture including the standard Ia domains and the N-terminal bHLH extension that is absent from SPCH and MUTE (fig. 2b and fig. S4). The strong similarity throughout these two proteins suggests that the primordial group Ia bHLH was duplicated in the mosses as a secondary event. The *Selaginella* and *Physcomitrella* Ia genes share with SPCH, MUTE and FAMA a perfectly conserved intron position in the bHLH domain (fig. S4).

Subgroup IIIb is also represented in *Physcomitrella* and *Selaginella* genomes, however, unlike the subgroup Ia proteins which have expanded in number in the angiosperms (ten members in *Arabidopsis* compared to two in *Physcomitrella*) subgroup IIIb has maintained a relatively stable number of members (three *Physcomitrella* genes compared to four in *Arabidopsis*) (Pires and Dolan 2009).

Functional analysis of *Physcomitrella* group Ia genes

The *Physcomitrella* sequences were chosen for further study since *Physcomitrella* contains the fewest subgroup Ia bHLHs of the plants surveyed here and they are the most distantly related to the *Arabidopsis* sequences. In order to determine if the limited degree of protein sequence conservation between the *Physcomitrella* Ia bHLHs and the stomatal lineage Ia bHLHs is sufficient for function, we expressed the PpSMF1 and 2 coding sequences in *Arabidopsis*.

Overexpression of SPCH, MUTE or FAMA confers specific and distinct epidermal phenotypes reflective of each gene's endogenous functions. SPCH overexpression results in supernumerary divisions, MUTE overexpression produces aberrantly shaped stomata

resulting from epidermal cells taking on GMC identity in the absence of asymmetric divisions and FAMA overexpression bypasses the GMC stage and produces unpaired guard cells (Ohashi-Ito and Bergmann 2006, MacAlister et al. 2007, Pilliterri et al. 2007). We overexpressed PpSMF1 and 2 with the constitutive 35S promoter and with an estrogen inducible system that provides higher levels of expression (Curtis and Grossniklaus 2003 and Zuo et al. 2000). Overexpression of PpSMF1 recapitulated elements of the SPCH, MUTE and FAMA overexpression phenotypes, including supernumerary epidermal cell divisions (fig. 3d), misshapen stomata (arrow in fig 3b) and unpaired guard cells (* in fig. 3b). Overexpression of PpSMF2 was also able to produce these phenotypes, but to a much smaller degree (table 2). The estrogen inducible system is able to drive higher levels of expression than the 35S promoter (Zuo et al. 2001). Therefore, as expected we see stronger, but qualitatively similar effects in the estrogen induction and 35S lines. The subgroup Ia bHLHs are generally required for stomatal production in all stomata-producing organs including the embryonically derived cotyledons and the post-embryonic leaves. We examined the both adaxial and abaxial epidermal phenotypes in the second and sixth leaves in the estrogen inducible PpSMF1 and PpSMF2 lines and found that the ability of the Physcomitrella sequences to promote stomatal lineage identity is not limited to the cotyledons (table 2). The ability of a single protein to exhibit functionality of SPCH, MUTE and FAMA was somewhat surprising given that the *Arabidopsis* proteins have highly compartmentalized activity. To determine if the *Physcomitrella* Ia proteins are capable of driving stomatal fate when expressed in a context more like the endogenous state, PpSMF1 and 2 were expressed under the native SPCH, MUTE and FAMA promoters (MacAlister et al. 2007 and Ohashi-Ito and Bergmann 2006) in the *spch*, *mute* and *fama* mutant backgrounds, respectively, and scored for their ability to rescue the *Arabidopsis* mutant phenotypes.

Neither PpSMF1 nor PpSMF2 was able to rescue the *spch* phenotype (0/4 lines for PpSMF1 and 0/2 lines for PpSMF2 produced any mature guard cells in a total of 60 seedlings). However, PpSMF1 exhibited partial rescue of both *fama* (fig. 3i-k) and *mute* (fig. 3g and k). The rescue is partial in the sense that some, but not all stomatal lineage cells will proceed past the point of blockage, and there can be unpaired guard cells in the *fama;PpSMF1 lines* We define “rescued” mature guard cells as those indistinguishable in morphology from wild type *Arabidopsis* guard cells (* in fig. 3j). The ability of PpSMF1 to rescue both *mute* and *fama* mutants suggests that PpSMF1 is able to promote the GMC fate and promote guard cell differentiation, tasks which, in *Arabidopsis*, are apportioned between two distinct proteins.

These results are particularly interesting in light of our previous experiments that demonstrated that *Arabidopsis* SPCH and FAMA could not substitute for each other (MacAlister et al. 2007). To determine whether *Arabidopsis* MUTE and FAMA, or the biological events of GMC formation and guard cell differentiation are inherently similar, we generated promoter swap constructs MUTEpro::FAMA and FAMApro::MUTE and attempted to rescue *mute* and *fama* as we had with the *Physcomitrella* sequences. Neither MUTE nor FAMA was capable of rescuing the other (0/7 lines for *fama*; 0/4 lines for *mute*), suggesting that PpSMF1 possesses dual functionalities while MUTE and FAMA are each now specialized for a single step.

Discussion

bHLH transcription factors are common to all eukaryotes. They are especially numerous in multicellular organisms where they often play a prominent role in development (Degnan et al. 2009). Phylogenetic analysis indicates that independent expansions in the animal and plant lineages resulted in unique subfamilies in these two groups (Ledent and Vervoort

2001). Subgroup Ia bHLHs are a broadly conserved, plant specific group of transcription factors with an essential role in the production of stomata. The paralogs SPCH, MUTE and FAMA have well defined biological roles and characteristic domain architectures that make them an excellent system to understand how a multistep division of labor may evolve in the context of gene family evolution.

The role of SPCH, MUTE and FAMA in Arabidopsis stomatal development has been well described (MacAlister et al. 2007, Pillitteri et al. 2007 and Ohashi-Ito and Bergmann 2006). Putative orthologues of all three are readily identifiable from the genomes of other angiosperms, suggesting conserved function. The available expression pattern data also support this possibility. Microarray data for the expression of the poplar SPCH and MUTE-like sequences show higher levels in developing leaves than in mature leaves (Wilkins et al. 2009), a pattern consistent with a role in stomatal development. More direct evidence is provided by Liu, Ohashi-Ito and Bergmann (2009) who showed that stomatal development in rice is dependent on OsFAMA function and is modulated by OsSPCH, a conservation of function that spans the more than 140 million years since the divergence of the monocots and dicots (Chaw et al. 2004). While this conservation of stomatal bHLH sequence and function within the angiosperms is consistent with the group Ia class having a conserved role in stomatal development, stomata are produced in a much wider group of plant taxa. If the stomatal bHLHs are generally required for stomatal development they should be generally conserved in the plants that produce them. Although SPCH, MUTE or FAMA orthologues cannot be unambiguously identified, the *Selaginella* and *Physcomitrella* genomes contain subgroup Ia bHLHs with the key Ia domains (fig. 2 and fig. S4). Subgroup IIIb is also conserved in *Physcomitrella* and *Selaginella*, suggesting that the heterodimerization between Ia and IIIb bHLHs could have been established early in plant evolution, even before the duplications that produced SPCH, MUTE and FAMA.

The mosses diverged from the ancestors of the angiosperms approximately 400MYA (Wolfe et al. 1989), yet a *Physcomitrella* Ia member (PpSMF1) is capable of promoting multiple steps of the stomatal lineage in *Arabidopsis*. This type of multi-functionality of a single protein suggests a possible mechanism for the evolution of the stomatal bHLHs (fig 4). The duplication of a multi-functional ancestor sequence with a domain architecture similar to that of the present day Physcomitrella sequences would allow for alteration of the domain architecture (including MUTE losing the N-terminus and SPCH acquiring the MPKTD) and specialization for discrete stomatal lineage stages and expression patterns. FAMA appears to be closest to this ancestral form. MUTE, with its truncated N-terminus may have arisen by disruption of the initial start codon forcing the use of a later start codon at the beginning of the bHLH domain, where there is a conserved methionine residue. SPCH's acquisition of the MPKTD is somewhat more difficult to trace since it is not a domain that is found in other Arabidopsis proteins. Insertion of a sequence with an abundance of serine and proline residues may have coincidentally created MAPK phosphorylation sites (P-X-S/T-P) that were able to capitalize on the MAPK pathway for SPCH regulation. Over time the functionally significant residues were maintained as the SPCH orthologues diverged, producing the islands of high conservation we now observe in this region (fig. S3). Functional data from Arabidopsis also supports this evolutionary scenario. For example, deletion of the N-terminal region of FAMA renders it capable of producing MUTE-like overexpression phenotypes (Ohashi-Ito and Bergmann, 2006), and deletion of the MPKTD of SPCH also generates a protein whose function more closely resembles that of MUTE (Lampard et al. 2008).

How might a single transcription factor have functioned to promote stomatal development in the ancestral case? A shorter stomatal lineage (for example, one without amplifying divisions as in the extant mosses) would simplify the problem. In the ancestral land plant, at

the minimum, FAMA and MUTE-like activity would be essential to produce paired guard cells. FAMA-like activity is needed to specify the guard cell identity and drive specific differentiation programs like stomatal pore formation, whereas MUTE-like activity ensures that two guard cells will coordinately form a stoma. This is achieved by MUTE promoting a specific precursor from which guard cells are derived (the GMC). The GMC has well-defined proliferative potential (limited to a single symmetric division). Only in this cell and its daughters is FAMA expressed, ensuring that only two guard cells are formed per stomatal unit, and that they are directly adjacent with a common wall in which the pore could form. The asymmetric divisions currently regulated by SPCH may be a relatively late addition to the stomatal lineage that allowed for stomatal patterning, flexibility and other fine control of epidermal development. In addition, the MAP kinase regulation of early stages of stomatal development via SPCH would also allow environmental stimuli to feed into control of stomatal density (Lampard et al. 2008). Throughout the land plants, guard cells morphologies are remarkably similar—two kidney shaped cells--except in the grasses, where guard cells are dumb-bell shaped. The ability of grass stomata to rapidly change pore size is dependent on subsidiary cells, themselves formed by asymmetric division (Franks and Farquhar, 2006). Interestingly, the unique and arguably more complex stomatal development in grasses correlates with the duplication of the SPCH gene in this lineage (Liu et al. 2009).

Our cross-species complementation results highlight that the similarity in guard cell development across different taxa may arise from an underlying molecular similarity: the ancient and widespread group Ia bHLHs. The subgroup Ia bHLHs in stomatal development may be an example of a gene class that has always been specifically tied to a particular cell type. From studies with *Arabidopsis* and maize, however, it is clear that many of the other genes required for guard cell specification and function are used in other processes. For example, the group IIIb bHLHs ICE1/SCRM and SCRM2 also have a well described role in cold response (Chinnusamy et al. 2003) whereas the MAPK signaling pathway that regulates SPCH activity also has various functions including establishment of polarity in the early embryo (Lukowitz et al. 2004) and stress and pathogen response (Asai et al. 2002 and Kovtun et al. 2000). Although it is impossible to know which function came first, the selective pressure for well functioning stomata could have driven the recruitment of these components for stomatal development and patterning as a secondary event.

Subgroup Ia bHLHs enjoy a unique function in the production of stomata. The evolution of this group of paralogs is informative not only for understanding how important plant structures came to be, but suggests that these genes are excellent targets for breeding or transgenic approaches to modifying stomatal production, and hence CO_2 acquisition, temperature and drought tolerance, across a wide range of economically important plant species.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding for this project was provided by US National Science Foundation grants IOS-0544895 and IOS-0844521 and a Terman fellowship (to DCB) CAM was supported in part by the US National Institutes of Health NHGRI institutional training grant in genome science 2 T32 HG0000044 to Stanford University. We are indebted to Ralph Quatrano (Washington University, St. Louis) for the gift of *Physcomitrella* gametophytes and for protocols and reagents, and to Julie Gray, Caspar Chater, David Beerling (University of Sheffield) and Thomas Münster (Max Planck Institute, Cologne) for discussion and communication of results before publication.

References

Asai T, Tena G, Plotnikova J, Willmann MR, Chiu W, Gomez-Gomez L, Boller T, Ausubel FM, Sheen J. MAP kinase signaling cascade in Arabidopsis innate immunity. Nature. 2002; 415:977–985. [PubMed: 11875555]

Ashton NW, Cove DJ, Featherstone DR. The isolation and physiological analysis of mutants of the moss Physcomitrella patens, which over-produce gametophores. Planta. 1979; 144:437–442.

Beerling DJ. Leaf evolution: gases, genes and geochemistry. Ann Bot. 2005; 96:345–52. [PubMed: 15965270]

Bertrand N, Ménager H, Maufrais C, Joly N, Maupetit J, Letort S, Carrere S, Tuffery P, Letondal C. Mobyle: a new full web bioinformatics framework. Bioinformatics. 2009; 25:3005–3011. [PubMed: 19689959]

Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. 1997; 268:78–94. [PubMed: 9149143]

Chaw S, Chang C, Chen H, Li W. Dating the monocot–dicot divergence and the origin of core eudicots using whole chloroplast genomes. J. Mol. Evol. 2004; 58:424–441. [PubMed: 15114421]

Chinnusamy V, Ohta M, Kanrar S, Lee B, Hong X, Agarwal M, Zhu JK. ICE1: a regulator of cold-induced transcriptome and freezing tolerance in Arabidopsis. Genes & Dev. 2003; 17:1043–1054. [PubMed: 12672693]

Clough SJ, Bent AF. Floral dip: A simplified method for agrobacterium-mediated transformation of *Arabidopsis thaliana*. Plant J. 1998; 16:735–743. [PubMed: 10069079]

Curtis MD, Grossniklaus U. A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant Physiol. 2003; 133:462–469. [PubMed: 14555774]

Degnan BM, Vervoort M, Larroux C, Richards GS. Early evolution of metazoan transcription factors. Curr. Opin. Genet. Dev. 2009; 19:591–599. [PubMed: 19880309]

Edwards D, Kerp H, Hass H. Stomata in early land plants: an anatomical and ecophysiological approach. J. Exp. Bot. 1998; 49:255–278.

Feller A, Hernandez JM, Grotewold E. An ACT-like domain participates in the dimerization of several plant basic-helix-loop-helix transcription factors. J. Biol. Chem. 2006; 281:28964–28974. [PubMed: 16867983]

Heim MA, Jakoby M, Werber M, Martin C, Weisshaar B, Bailey PC. The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. Mol. Biol. Evol. 2003; 20:735–747. [PubMed: 12679534]

Hernandez ML, Passas HJ, Smith LG. Clonal analysis of epidermal patterning during maize leaf development. Dev. Biol. 1999; 216:646–658. [PubMed: 10642799]

Kanaoka MM, Pillitteri LJ, Fujii H, Yoshida Y, Bogenschutz NL, Takabayashi J, Zhu JK, Torii KU. SCREAM/ICE1 and SCREAM2 specify three cell-state transitional steps leading to Arabidopsis stomatal differentiation. Plant Cell. 2008; 20:1775–1785. [PubMed: 18641265]

Kovtun Y, Chiu W, Tena G, Sheen J. Functional analysis of oxidative stress-activated mitogen-activated protein kinase cascade in plants. Proc. Natl Acad. Sci. USA. 2000; 97:2940–2945. [PubMed: 10717008]

Lampard GR, MacAlister CA, Bergmann DC. Arabidopsis stomatal initiation is controlled by MAPK-mediated regulation of the bHLH SPEECHLESS. Science. 2008; 322:1113–1116. [PubMed: 19008449]

Larkin MA, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007; 23:2947–2948. [PubMed: 17846036]

Ledent V, Vervoort M. The basic helix-loop-helix protein family: comparative genomics and phylogenetic analysis. Genome Res. 2001; 11:754–770. [PubMed: 11337472]

Liu CM, Meinke DW. The *titan* mutants of Arabidopsis are disrupted in mitosis and cell cycle control during seed development. Plant J. 1998; 16:21–31. [PubMed: 9807824]

Liu T, Ohashi-Ito K, Bergmann DC. Orthologs of *Arabidopsis thaliana* stomatal bHLH genes and regulation of stomatal development in grasses. Development. 2009; 136:2265–2276. [PubMed: 19502487]

Lukowitz W, Roeder A, Parmenter D, Somerville C. A MAPKK kinase gene regulates extra-embryonic cell fate in Arabidopsis. Cell. 2004; 116:109–119. [PubMed: 14718171]

MacAlister CA, Ohashi-Ito K, Bergmann DC. Transcription factor control of asymmetric cell divisions that establish the stomatal lineage. Nature. 2007; 445:537–540. [PubMed: 17183265]

Ohashi-Ito K, Bergmann DC. Arabidopsis FAMA controls the final proliferation/differentiation switch during stomatal development. Plant Cell. 2006; 18:2493–2505. [PubMed: 17088607]

Peterson KM, Rychel AL, Torii KU. Out of the mouths of plants: the molecular basis of the evolution and diversity of stomatal development. Plant Cell. 2010; 22:296–306. [PubMed: 20179138]

Pires N, Dolan L. Origin and diversification of basic-Helix-Loop-Helix proteins in plants. Mol. Biol. Evol. 2010; 27:862–874. [PubMed: 19942615]

Pillitteri LJ, Sloan DB, Bogenschutz NL, Torii KU. Termination of asymmetric cell division and differentiation of stomata. Nature. 2007; 445:501–505. [PubMed: 17183267]

Raven JA. Selection pressures on stomatal evolution. New Phytol. 2002; 153:371–386.

Sack FD, Paolillo DJ Jr. Incomplete cytokinesis in Funaria stomata. American J. Bot. 1985; 72:1325–1333.

Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. Synteny and collinearity in plant genomes. Science. 2008; 320:486–488. [PubMed: 18436778]

Toledo-Ortiz G, Huq E, Quail PH. The Arabidopsis basic/helix-loop-helix transcription factor family. Plant Cell. 2003; 15:1749–1770. [PubMed: 12897250]

Wilkins O, Nahal H, Foong J, Provart NJ. Expansion and diversification of the Populus R2R3-MYB family of transcription factors. Plant Physiol. 2009; 149:981–993. [PubMed: 19091872]

Wolfe KH, Gouy M, Yang Y, Sharpt PM, Li WH. Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc. Natl Acad. Sci. USA. 1989; 86:6201–6205. [PubMed: 2762323]

Ziegler, H. The evolution of stomata. In: Zeiger, E.; Farquhar, GD.; Cowan, IR., editors. Stomatal Function. Stanford University Press; 1987.

Zuo J, Niu QW, Chau NH. An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in plants. Plant J. 2000; 24:265–273. [PubMed: 11069700]

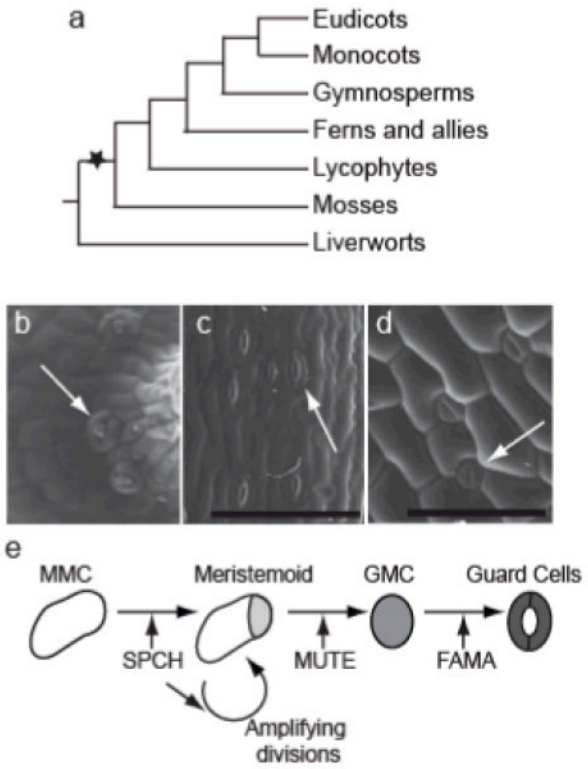

Fig. 1. Stomata are an ancient innovation in plants

(a) Land plant phylogeny including major groups described in this report; the first appearance of stomata is marked by a star. (b-d) Scanning electron micrographs of stomata (indicated by white arrows) in representative plant lineages: (b) *Physcomitrella patens* stomata on base of spore capsule, (c) Lycophyte *Selaginella moellendorffii* stomata on leaf, (d) Eudicot *Eschscholzia californica* (California poppy) stomata on leaf. (e) Diagram of stages in Arabidopsis (dicot) stomatal lineages with the point of action of the transcription factors, SPCH, MUTE and FAMA marked. A protodermal cell, the meristemoid mother cell (MMC) divides asymmetrically to form a small meristemoid (light grey). The meristemoid may self-renew by further asymmetric divisions or transition to a guard mother cell (GMC) (dark grey) which divides symmetrically to form the two guard cells. In Arabidopsis, asymmetric divisions require SPCH activity; GMC identity requires MUTE and the differentiation of the GCs requires FAMA function. Scale bars =100μm

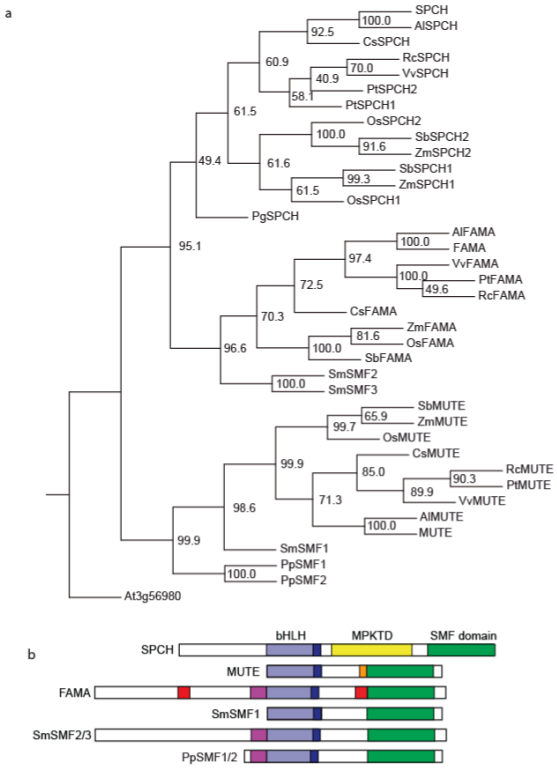

Fig. 2. Phylogenetic tree and domain architecture of SPCH, MUTE and FAMA

(a) Maximum parsimony phylogentic tree based on complete predicted protein coding sequences of putative orthologues of stomatal bHLHs. Bootstrap values for 1000 replicates are given in nodes as percents. At3g56980 (a member of subgroup Ib, Heim et al. 2003) is used as an outgroup. Zm=*Zea mays* (maize), Os=*Oryza sativa* (rice), Vv=*Vitis vinifera* (grape), Rc= *Ricinus communis* (castor bean), Pt=*Populus trichocarpa* (poplar), Cs= *Cucumis sativus* (cucumber), Pg= *Picea glauca* (white spruce), Sm= *Selaginella moellendorffii* and Pp= *Physcomitrella patens*. (b) Diagram of domain architecture of SPCH, MUTE and FAMA-like sequences from angiosperms and subgroup Ia members from *Selaginella* and *Physcomitrella*. Regions that are unique to FAMA, MUTE and SPCH are red, orange and yellow respectively. The bHLH is in light blue with the Ia characteristic N-terminal extension in purple and the C-terminal extension in dark blue. The conserved C-terminal SMF domain is green

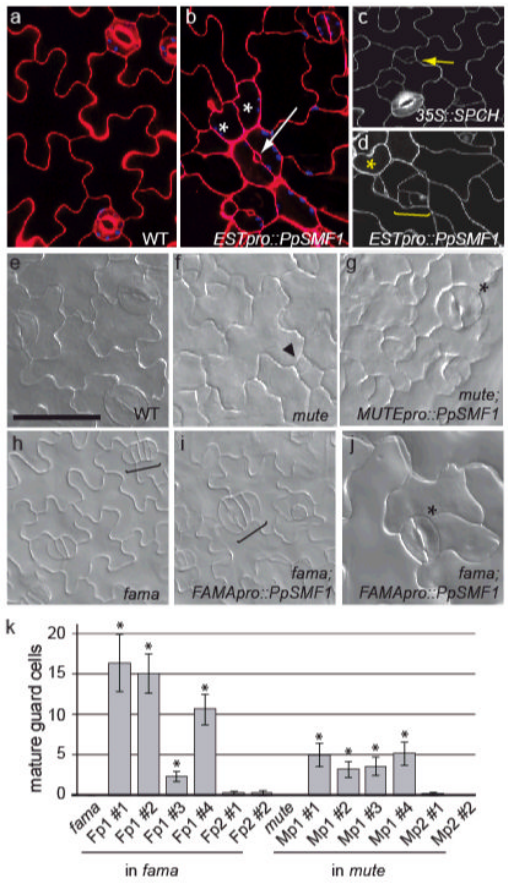

Fig. 3. Functional conservation of Physcomitrella and Arabidopsis group Ia bHLHs

(a-d) Confocal image of 10dpg cotyledons demonstrating that overexpression of PpSMF1 promotes SPCH, MUTE and FAMA-like overexpression phenotypes. (a) Wild type (b) induced ESTpro::PpSMF1 with unpaired guard cells marked with white *, and aberrantly shaped stomata by indicated by white arrow. Cell outlines are visualized by propidium iodide staining (red), chloroplast autofluorescence is blue. (c) 35S::SPCH, extra cell divisions noted by yellow arrow (d) induced ESTpro::PpSMF1 with extra SPCH-like divisions indicated in brackets and unpaired guard cells marked with yellow * (e-j) DIC images of 8 dpg cotyledons showing that expression of PpSMF1 can substitute for MUTE and FAMA function. (e) wild type (f) *mute* and (h) *fama* mutants; *mute* and *fama* do not produce mature guard cells but arrest at the meristemoid (black arrowhead) and GMC stage (black bracket) respectively. (g) expression of PpSMF1 under the MUTE promoter can partially rescue the *mute* phenotype; a normal stoma is marked with a black '*'. PpSMF1 rescue of *fama* may result in normal stomata (j, black *) or stomata within a *fama* GMC cluster, (i, black bracket). Scale bars are 50μm. a-d and e-j are at the same magnification. (k) quantification of rescue of the *mute* and *fama* phenotype as mean number of mature guard cells produced per seedling plus or minus standard error in independent transgenic lines (n=20 sides for each line) * mark statistically significant deviation from the mutant background. … do summary for the notation for the lines… for 4 lines of PpSMF1 and 2 of PpSMF2…

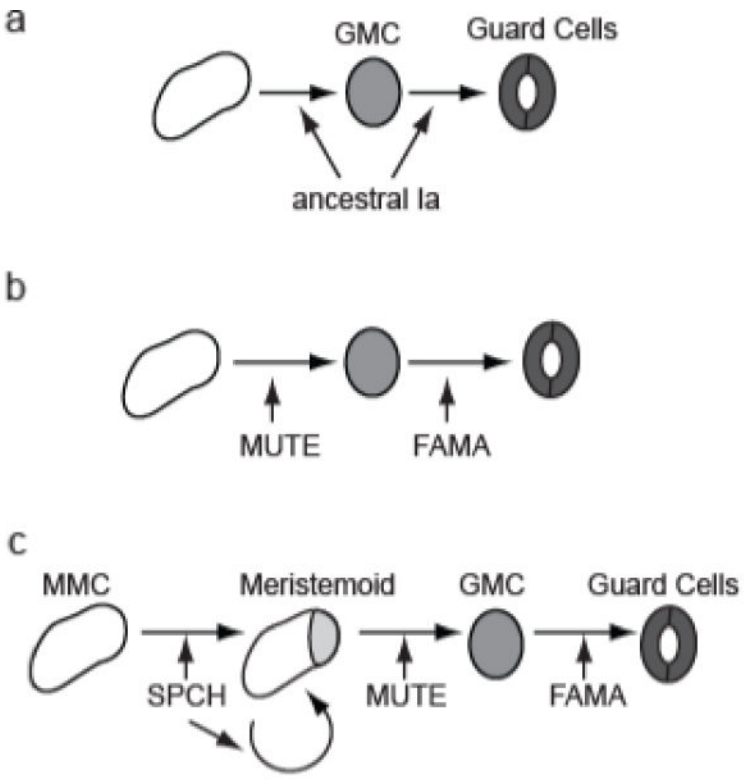

Fig.4. Model of the expansion of the developmental complexity of the stomatal lineage

(a) A single, multifunctional Ia member drives both the specification of GMC identity and guard cell differentiation. (b) Duplication of the ancestral Ia member allows for specialization for GMC specification (MUTE-like activity) and guard cell differentiation (FAMA-like activity). (c) Another duplication produces a third Ia member and allows for the evolution SPCH and the current stomatal lineage, including amplifying divisions.

Table 1
List of public database identification numbers for sequences identified

Physcomitrella patens	**JGI Protein ID**
PpSMF1	184221
PpSMF2	156270
Selaginella moellendorffii	**JGI Protein ID**
SmSMF1	91359
SmSMF2	430716
SmSMF3	131087
Populus trichocarpa	**Phytozome**
PtSPCH1	POPTR_0012s01370.1
PtSPCH2	POPTR_0015s06160
PtMUTE	POPTR_0008s20720
PtFAMA	POPTR_0001s32160.1
Vitis vinifera	**NCBI reference**
VvSPCH	XM_002267709.1
VvMUTE	XM_002277642.1
VvFAMA	XM_002266799.1
Ricinus communis	**NCBI reference**
RcSPCH	XM_002513543.1
RcMUTE	XM_002517114.1
RcFAMA	XM_002518062.1
Picea glauca	**Genbank**
PgSPCH	GQ04006P07
Sorghum bicolor	**Phytozome**
SbSPCH1	Sb10g020490
SbSPCH2	Sb04g009390
SbMUTE	Sb09g030930
SbFAMA	Sb09030060
Cucumis sativus	**Phytozome**
CsSPCH	Cucsa.012930.1
CsMUTE	Cucsa.254010.1
CsFAMA	Cucsa.256420.1
Arabidopsis lyrata	**JGI Protein ID**
AlSPCH	916961
AlMUTE	928234
AlFAMA	930468

Table 2

Phenotypic characterization of PpSMF1 and PpSMF2 transgenic lines across different tissues

Organ	Surface	Line	SPCH	MUTE	FAMA
Cotyledons	adaxial	est::PpSMF1 #1	+++	+++	+++
		est::PpSMF1 #2	+++	+++	+++
		est::PpSMF2 #1	+	+	++
		est::PpSMF2 #2	−	+	++
	abaxial	est::PpSMF1 #1	++	+	++
		est::PpSMF1 #2	++	++	++
Leaf 2	adaxial	est::PpSMF1 #1	+++	+++	++
		est::PpSMF1 #2	+++	+++	++
Leaf 6	adaxial	est::PpSMF1 #1	+	+	+
		est::PpSMF1 #2	+	+	+