MPMI Vol. 30, No. 6, 2017, pp. 489–501. http://dx.doi.org/10.1094/MPMI-10-16-0218-R

e-Xtra*

Evolutionary History of Subtilases in Land Plants and Their Involvement in Symbiotic Interactions

Alexander Taylor and Yin-Long Qiu

University of Michigan, Department of Ecology and Evolutionary Biology, Ann Arbor, MI, U.S.A.

Accepted 22 March 2017.

Subtilases, a family of proteases involved in a variety of developmental processes in land plants, are also involved in both mutualistic symbiosis and host-pathogen interactions in different angiosperm lineages. We examined the evolutionary history of subtilase genes across land plants through a phylogenetic analysis integrating amino acid sequence data from full genomes, transcriptomes, and characterized subtilases of 341 species of diverse green algae and land plants along with subtilases from 12 species of other eukaryotes, archaea, and bacteria. Our analysis reconstructs the subtilase gene phylogeny and identifies 11 new gene lineages, six of which have no previously characterized members. Two large, previously unnamed, subtilase gene lineages that diverged before the origin of angiosperms accounted for the majority of subtilases shown to be associated with symbiotic interactions. These lineages expanded through both whole-genome and tandem duplication, with differential neofunctionalization and subfunctionalization creating paralogs associated with different symbioses, including nodulation with nitrogen-fixing bacteria, arbuscular mycorrhizae, and pathogenesis in different plant clades. This study demonstrates for the first time that a key gene family involved in plant-microbe interactions proliferated in size and functional diversity before the explosive radiation of angiosperms.

Subtilisin-like serine proteases (subtilases) of the S8A family are a large group of proteases with broad substrate specificity generally involved in protein turnover and organ development in land plants (Schaller et al. 2012; Siezen and Leunissen 1997). Subtilases have roles in developmental processes such as lateral root development, epidermal differentiation, cuticle formation, and xylem differentiation (Neuteboom et al. 1999; Tanaka et al. 2001; Zhao et al. 2000). In addition, subtilases have been shown to be involved in a variety of symbiotic interactions, including arbuscular mycorrhization (AM), nodulation, and pathogenesis (Table 1).

Despite the wide range of important functions of subtilases in plants, phylogenetic relationships in this gene family have not been updated since a phylogenetic study of subtilases in the single model organism *Arabidopsis thaliana* by Rautengarten et al. (2005). The availability of sequence data from a wide variety of fully sequenced genomes and transcriptomes across the tree of life now makes a much more comprehensive phylogenetic reconstruction possible. This not only allows for a more accurate analysis of relationships among gene clades but, also, permit assessment of evolutionary depth of each clade by comparison with the species tree. Having these two types of information can facilitate evolutionary interpretation of evidence on gene functions obtained from genetic and developmental studies. This study uses a phylogeny of 2,441 subtilase sequences from across 341 green plants, along with 19 subtilase sequences from 12 species of other eukaryotes, archaea, and bacteria, to examine the evolutionary history of this important gene family. Because the genetic basis for the evolutionary origin of nodulation is an important question with implications for agriculture as well as our understanding of convergence in complex symbiotic traits, our interpretation of the phylogeny focuses on the recruitment of subtilases to mediate AM and nodulation, two mutualistic symbiotic interactions plants form with microbes in their roots.

Corresponding author: A. Taylor; E-mail: abtaylor@umich.edu

*The *e*-Xtra logo stands for "electronic extra" and indicates that four supplementary tables are published online.

Subtilases in root mutualisms.

Expression of certain subtilases is induced by activation of the common symbiotic (*sym*) pathway that mediates the formation of two ecologically and economically important mutualisms that plants form with microbes in their roots, AM and nodulation with nitrogen-fixing bacteria (Gherbi et al. 2008; Hocher et al. 2011; Kistner and Parniske 2002; Kistner et al. 2005; Parniske 2008; Pawlowski et al. 2011). The *sym* pathway involves the plant perception of lipo-chito-oligosaccharide (LCO) signals convergently produced by rhizobial bacteria and AM fungi, engendering intracellular calcium spiking in the plant root epidermis (Maillet et al. 2011; Oldroyd 2013; Stracke et al. 2002). Subsequently, the plant forms a cytoplasmic bridge, called the "infection thread" for nodulation and "prepenetration apparatus" for AM, that guides the microbial symbiont to the site of symbiosome or arbuscule formation (Genre et al. 2005; Kistner et al. 2005; Parniske 2008; Szczyglowski et al. 1998). Subtilase genes in the *sym* pathway are expressed in infected root hair or epidermal cells forming the infection thread or prepenetration apparatus, with proteins secreted in the apoplastic space in the symbiotic membrane-membrane interface of the symbiosome or arbuscule, where they are thought to be involved in restructuring the plant membrane surface (Table 1) (Kistner et al. 2005; Laplaze et al. 2000; Svistoonoff et al. 2003; Takeda et al. 2009).

Several lines of evidence link subtilases to AM (Table 1). In early expression analyses using cDNA probe arrays, expression of *OsAM21* (renamed *OsSBTM1* by Takeda et al. [2011]) in *Oryza sativa* (Güimil et al. 2005) and AW584611 (*MtSBTM1*) in *Medicago truncatula* (Liu et al. 2003) was found to be induced in root tissues following inoculation with AM fungi (Table 1). Histochemical localization showed that the subtilases LjSBTM1, LjSBTM3, LjSBTM4, and LjSBTS are localized to

the apoplastic space in cells forming arbuscules in *Lotus japonicus*, following inoculation with the AM fungus *Glomus intraradices* (Takeda et al. 2009). Suppression of *LjSBTM1* and *LjSBTM3* expression by RNAi decreased arbuscule formation in *L. japonicus*, demonstrating that these subtilases play a role in AM (Table 1) (Takeda et al. 2009).

Spatial expression data in legumes also implicates subtilases in the formation of nodules with rhizobial bacteria (Table 1) (Kistner et al. 2005; Takeda et al. 2009). Expression of *LjSBTM4* and *LjSBTS* is induced by inoculation with the rhizobial bacterium *Mesorhizobium loti* (Takeda et al. 2009). Histochemical localization showed that *LjSBTS* expression is transiently induced in epidermal cells during early rhizobial infection, while *LjSBTM4* is induced in both epidermal and cortical cells around the infection thread and nodule primordia as well as in mature nodules (Takeda et al. 2009). A recent RNA-seq study conducted as part of the Expression Atlas project (Kapushesky et al. 2010) showed that exposure to LCO signals from *Sinorhizobium meliloti* induces expression of several subtilases in *M. truncatula* roots (van Zeijl et al. 2015).

Additionally, subtilase expression is induced during nodulation with *Frankia* bacteria in nodulating species in orders Fagales and Cucurbitales (Table 1) (Laplaze et al. 2000; Pawlowski et al. 2011; Ribeiro et al. 1995; Svistoonoff et al. 2003, 2004, 2014). Using β-glucuronidase (GUS) and green fluorescent protein (GFP) reporter gene constructs, Svistoonoff et al. (2003) showed that the subtilase gene *CG12* is expressed in root hair cells in *Casuarina glauca* during infection thread formation and in infected cortical nodule and prenodule cells during early nodulation but not during the period of nitrogen fixation, as predicted by bacterial *nifH* expression (Svistoonoff et al. 2003). Expression of a subtilase that is a close homolog to *CG12* is induced in *Datisca glomerata* during the initiation of

Table 1. Previously characterized subtilases included in this study[a]

Gene	Organism	Clade	Function	Evidence	Citation	Accession no.
SBT6.1	*Arabidopsis thaliana*	*SBT7*	Cell elongation, heat stress response	Genetic suppressor screen	Ghorbani et al. 2016; Liu et al. 2007	AAM97020
AtALE1	*A. thaliana*	*SBT2.4*	Cuticle formation and epidermal differentiation	Gene knockout	Tanaka et al. 2001	NP_564793.2
LlLIM9	*Lilium longiflorum*	*SBT2.7*	Microspore development	Immunocytochemical localization	Taylor et al. 1997	BAA04839.1
StSBT2.2a	*Solanum tuberosum*	*SBT2.2*	Pathogenesis with *Phytophthora infestans*	Expression	Gao et al. 2013	XP_006353035.1
StSBT2.2b	*S. tuberosum*	*SBT2.2*	Pathogenesis with *P. infestans*	Expression	Gao et al. 2013	XP_006363641.1
AtAIR3	*A. thaliana*	*SBT5.3*	Lateral root development, loosening cell walls	Spatial expression of GUS reporter gene fusion	Neuteboom et al. 1999	AAK74005
MtSBT5.3	*Medicago truncatula*	*SBT5.3*	Expressed in response to LCO signaling molecules from *Sinorhizobium meliloti*	Expression	van Zeijl et al. 2015	XP_003601486.1
AtSBT3.3	*A. thaliana*	*SBT3.3*	Immune priming in response to *Pseudomonas syringae* and *Hyaloperonospora arabidopsidis*	Expression, gene knockout, chromatin immunoprecipitation	Ramírez et al. 2013	NP_568255
MtSBTS	*M. truncatula*	*SBT4.16*	Expressed in response to LCO signaling molecules from *Sinorhizobium meliloti*	Expression	van Zeijl et al. 2015	XP_003606146.2
LjSBTS	*Lotus japonicus*	*SBT4.16*	AM with *Glomus intraradices* and nodulation with *Mesorhizobium loti*	Expression, expression reduced in *sym* pathway mutants	Kistner et al. 2005; Takeda et al. 2009	BAF95887.1
GmSLP1, *SSTP2*	*Glycine max*	*SBT4.19*	Seed coat development	Immunocytochemical localization	Beilinson et al. 2002; Rautengarten et al. 2008	AY033772.1, AAK53065
GmSLP2, *SSTP1*	*Glycine max*	*SBT4.19*	Cotyledon development	Immunocytochemical localization	Beilinson et al. 2002; Rautengarten et al. 2008	AF160513
AtXSP1	*A. thaliana*	*SBT4.6*	Xylem differentiation	Expression	Zhao et al. 2000	NP_568889
Cucumisin	*Cucumis melo*	*SBT4.21*	General protein degradation during fruit maturation	Immunocytochemical localization	Yamagata et al. 1994	BAA06905.1
AtSDD1	*A. thaliana*	*SBT1.2*	Regulates stomatal density	Gene knockout, RNA-blot localization	Berger and Altmann 2000	NP_563701
SlP69A	*S. lycopersicon*	*SBT1.10*	Pathogenesis with citrus exocortis viroid	Expression	Tornero et al. 1996	CAA64566.1
SlP69B	*S. lycopersicon*	*SBT1.10*	Pathogenesis with *Pseudomonas syringae*, *P. infestans*, citrus exocortis viroid	Expression, Coevolved pathogen inhibitor	Tornero et al. 1997; Jordá et al. 1999; Tian et al. 2004	CAA76725.1
SlP69C	*S. lycopersicon*	*SBT1.10*	Pathogenesis with *Pseudomonas syringae*	Expression	Jordá et al. 1999	Q9ZR46
SlP69D	*S. lycopersicum*	*SBT1.10*	Expressed in young leaves	Expression	Jordá et al. 1999	Q9SAN2

(*continued on next page*)

[a] GUS = β-glucuronidase; LCO = lipo-chito-oligosaccharide; AM = arbuscular mycorrhization; GFP = green fluorescent protein; RNAi = RNA interference.

nodulation with *Frankia* bacteria, based on expressed sequence tag data (Demina et al. 2013).

Nodulation is restricted to the nitrogen-fixing clade (NFC) of rosids (orders Fabales, Fagales, Cucurbitales, and Rosales) but has evolved multiple times independently within that clade (Doyle 1994, 2011; Soltis et al. 1995; Swensen and Mullin 1997). The *sym* pathway, which mediates AM across land plants, was recruited during each evolutionary origin of nodulation for which the genetic basis has been examined (Demina et al. 2013; Gherbi et al. 2008; Hocher et al. 2011; Kistner and Parniske 2002; Kistner et al. 2005; Op den Camp et al. 2011; Pawlowski and Sprent 2008; Pawlowski et al. 2011; True and Carroll 2002). AM is the ancestral condition in land plants, and elements of the *sym* pathway are functionally conserved from legumes to liverworts (Wang and Qiu 2006; Wang et al. 2010) and, perhaps, even to charophytes (Delaux et al. 2015). The evolutionary origins of nodulation are, thus, an example of "deep homology" (Doyle 2011; Shubin et al. 1997, 2009), meaning that phylogenetically distinct nodulation symbioses originated by the repeated, independent recruitment of homologous genes from a plesiomorphic (ancestral) pathway for novel, homoplastic functions in different lineages. It is currently unknown how subtilases involved in nodulation and AM are related to each other, though one small gene phylogeny of 13 subtilases indicated that those mediating nodulation and AM in different plant lineages are of independent origins, in contrast to the pattern observed for *sym* pathway genes (Takeda et al. 2007).

Subtilases in pathogenesis.

In addition to their roles in nodulation and AM, subtilases are expressed during pathogenesis. The *Arabidopsis thaliana* subtilase gene *AtSBT3.3* is involved in immune priming in response to *Pseudomonas syringae* and *Hyaloperonospora arabidopsidis*, as demonstrated by weakened immune responses in *sbt3.3*

Table 1. (*continued from previous page*)

Gene	Organism	Clade	Function	Evidence	Citation	Accession no.
SlP69E	*S. lycopersicum*	*SBT1.10*	Expressed constitutively in roots	Expression	Jordá et al. 2000	Q9LWA4
SlP69F	*S. lycopersicum*	*SBT1.10*	Expressed in hydathodes	Expression	Jordá et al. 2000	CAB67120.1
LjSBTM1	*L. japonicus*	*SBT1.10*	AM with *Glomus intraradices*	Spatial expression of GUS reporter gene fusion, RNAi inhibition reduces arbuscule formation	Kistner et al. 2005; Takeda et al. 2009; van Zeijl et al. 2015	BAF95755.1
LjSBTM3	*L. japonicus*	*SBT1.10*	AM with *Glomus intraradices*	Expression, RNAi inhibition reduces arbuscule formation	Takeda et al. 2009	BAF95887.1
LjSBTM4	*L. japonicus*	*SBT1.10*	AM with *Glomus intraradices* and nodulation with *Mesorhizobium loti*	Spatial expression of GUS reporter gene fusion	Takeda et al. 2009; van Zeijl et al. 2015	BAF95753.1
MtSBTM1	*M. truncatula*	*SBT1.10*	AM with *Glomus versiforme*	Expression	Liu et al. 2003	AW584611, XP_003611196.1
OsSBTM1	*Oryza sativa*	*SBT1.13*	AM with *Glomus intraradices*	Expression	Güimil et al. 2005	NP_915664.1
AG12	*Alnus glutinosa*	*SBT1.13*	Nodulation with *Frankia* sp.	Expression	Ribeiro et al. 1995	CAA59964.1
CG12	*Casuarina glauca*	*SBT1.13*	Nodulation with *Frankia* sp. but not mycorrhization with ectomycorrhizal *Pisolithus alba* or AM *Glomus intraradices*	Spatial expression of GFP reporter gene fusion	Laplaze et al. 2000; Svistoonoff et al. 2003; Hocher et al. 2006	AAO62352.1
StCG12a	*S. tuberosum*	*SBT1.13*	Pathogenesis with *P. infestans*	Expression	Gao et al. 2013	XP_004249504.1
StCG12b	*S. tuberosum*	*SBT1.13*	Pathogenesis with *P. infestans*	Expression	Gao et al. 2013	XP_015089529.1
SlSBT3	*S. lycopersicum*	*SBT1.13*	Response to herbivory by *Manduca sexta*	Expression, gene knockout	Meyer et al. 2016	NP_001234774.1
MtCG12a	*M. truncatula*	*SBT1.13*	Expressed in response to LCO signaling molecules from *Sinorhizobium meliloti*	Expression	van Zeijl et al. 2015	XP_003624104.2
MtCG12b	*M. truncatula*	*SBT1.13*	Expressed in response to LCO signaling molecules from *Sinorhizobium meliloti*	Expression	van Zeijl et al. 2015	XP_003624105.1
AtSLP2	*A. thaliana*	*SBT1.6*	General protein turnover and metabolism	Expression	Golldack et al. 2003	AAL67071
TaSSP1	*Triticum aestivum*	*SBT1.4*	Peptide degradation in senescing leaves	Expression, proteolytic assay	Roberts et al. 2003	AGN03879.1
AtARA12, *AtSLP1*, *AtSCS1*	*A. thaliana*	*SBT1.7*	General protein degradation in intercellular space of stem; expression in leaves in young plants and leaves, roots, and stems in older plants	Expression	Hamilton et al. 2003; Golldack et al. 2003	NP_569048.1, AAK25995

mutants (Ramírez et al. 2013). *SlSBT3* is expressed in response to herbivory by the insect *Manduca sexta* in *Solanum lycopersicum* (Meyer et al. 2016). Expression of *SlP69* subtilase paralogs in *S. lycopersicum* is induced in leaf and stem tissues in response to a variety of pathogens, including *Pseudomonas syringae*, *Phytophthora infestans*, and the citrus exocortis viroid (Jordá et al. 1999, 2000; Tornero et al. 1996, 1997). More recently, as part of the Expression Atlas project (Kapushesky et al. 2010), transcriptome analysis showed differential regulation of subtilases during *P. infestans* infection in tuber tissue of *Solanum tuberosum* (Gao et al. 2013). Strengthening the case for its having a role in pathogenesis, the subtilase *SlP69B* is induced in *S. lycopersicum* during infection with *P. infestans* and the pathogen expresses a Kazal-like protease inhibitor that inhibits its activity, indicating pathogen coevolution in response to this subtilase (Tian et al. 2004, 2005).

Study aims.

This study aimed to determine the phylogenetic relationships of subtilases in the Viridiplantae kingdom, to enhance understanding of the origin and evolution of different lineages in this important gene family (Rautengarten et al. 2005; Schaller et al. 2012). Our analysis and interpretation of this phylogeny focused on subtilase lineages associated with symbiotic interactions in angiosperms. This study also aimed to elucidate the pattern of duplication that led to the multiple paralogous symbiosis-induced subtilases *LjSBTM1*, *LjSBTM3*, and *LjSBTM4* in *L. japonicus* and *SlP69* paralogs in *S. lycopersicum* and, particularly, to determine whether the duplication of symbiosis-induced subtilases occurred via tandem or whole-genome duplication (WGD). Tandem duplication and subsequent neofunctionalization is common in genes mediating pathogenesis (Michelmore and Meyers 1998), while WGD has been implicated in the origin of nodulation in legumes (Cannon et al. 2015; Vanneste et al. 2014).

RESULTS AND DISCUSSION

Relationships among plant subtilases.

Our phylogenetic analysis of 2,460 subtilase amino acid sequences for the first time produces a gene phylogeny that reconstructs evolutionary history of this important gene family across green plants (Fig. 1). We greatly expanded taxonomic sampling compared with the gene phylogeny of Rautengarten et al. (2005), incorporating subtilase homologs from 341 species of diverse Viridiplantae, seven species of nonviridiplantae eukaryotes, one species of archaea, and four species of bacteria (Supplementary Tables S1, S2, and S4). This expanded taxonomic coverage reveals the approximate age of each gene lineage through comparison of the gene phylogeny with the corresponding organismal phylogeny (Finet et al. 2010; Qiu et al. 2006, 2010; Soltis et al. 2011). There are two important caveats to this approach. One is the uneven sampling of genomic data—among non-angiosperm land plants, this study has only one bryophyte, *Physcomitrella patens*, and one lycophyte, *Selaginella moellendorffii*, and lacks any genomes from gymnosperms and monilophytes. The second is the fact that transcriptomic data likely contain only a small subset of the homologs in of a given species genome, those that happen to be expressed at the time of tissue sampling.

Despite these caveats, this expanded gene phylogeny allows for several insights into the evolution of subtilases in green plants. First, there is a general pattern of increased gene diversification in angiosperms, resulting in 21 of the 33 named gene lineages restricted to angiosperms (Fig. 1). Additionally, we identified 11 new gene lineages: *SBT2.7*, *SBT3.19*, *SBT4.16*, *SBT4.17*, *SBT4.18*, *SBT4.20*, *SBT4.21*, *SBT1.10*, *SBT1.11*, *SBT1.12*, and *SBT1.13*, six of which have no previously characterized members (Fig. 1). Some of these lineages were quite ancient, dating to the origin of seed plants (*SBT3.19*) or vascular plants (*SBT4.16*, *SBT4.17*), but were missing in the analysis by Rautengarten et al. (2005) because *A. thaliana* had no homologs in these lineages. This effect was especially notable in gene lineages containing subtilases mediating symbiotic interactions, such as *SBT4.16*, *SBT1.10*, *and SBT1.13*, all of which contain subtilase genes expressed during nodulation or AM interactions and which contain no homologs from *A. thaliana*. This is in keeping with a more general pattern of loss of AM-related genes in *A. thaliana*, a nonnodulating and nonmycorrhizal plant (Bravo et al. 2016; Delaux et al. 2015). Finally, while the topology of our gene phylogeny largely agreed with that of Rautengarten et al. (2005), there were major revisions to the gene clades *SBT5*and *SBT6* (discussed below).

We found strong bootstrap (BS) support for *SBT1* (96% BS), *SBT2* (100% BS), and *SBT3* (93% BS), three gene clades named by Rautengarten et al. (2005) (Fig. 1). However, *SBT4* was weakly supported (46% BS), though there was strong support for named lineages within this group (Fig. 1). Several clades at this high level would have to be recognized if *SBT4* was more narrowly defined to a clade with at least 70% BS support. While Rautengarten et al. (2005) found *SBT5* to be paraphyletic, we found the *AtSBT5* sequences to be polyphyletic, with *AtSBT5.1* and *AtSBT5.2* falling into the weakly supported *SBT4*, while the rest form a monophyletic group (99% BS). Finally, while Rautengarten et al. (2005) recognized *SBT6* as a monophyletic group with two sequences in *A. thaliana*, *AtSBT6.1* and *AtSBT6.2*, our analysis found that these two sequences fell into two large gene clades that, together with a small clade, formed a paraphyletic group at the base of the entire subtilase tree of Viridiplantae (Fig. 1), which are rooted with prokaryotic subtilase sequence (Fig. 1).

A series of two consecutive deep divergences of subtilase sequences right above the outgroup prokaryotic sequences, with moderate to strong BS support, indicates that this large gene family underwent two rounds of duplication during early stages of eukaryotic evolution, as the clades that define the two splits contain animal and green plant sequences (Fig. 1). The first divergence in eukaryotes is between a small ancient clade (100% BS) containing the subtilase homolog *PC9* from the animal *Mus musculus* and two homologs from the charophyte green alga *Klebsormidium flaccidum* and all other eukaryotic subtilase clades. The second divergence results in *SBT7* and an unnamed superclade that includes *SBT1*, *SBT2*, *SBT3*, *SBT4*, *SBT5*, and *SBT6* and a small ancient clade containing sequences from animals, fungi, and stramenopiles in addition to those from green plants. This result is similar to what Rautengarten et al. (2005) found, but the improved taxon sampling in this study significantly increased robustness of the conclusion.

SBT7, a newly named subfamily in this study, contains *AtSBT6.1*, which previously was placed in *SBT6* by Rautengarten et al. (2005). It clearly originated early in eukaryote evolution, before the divergence of Viridiplantae from Metazoa (Adl et al. 2012), as evidenced by the presence of the subtilase homolog *SIP* from *Homo sapiens* and a subtilase homolog from the phaeophyte *Ectocarpus siliculosus* in this clade. The clade containing *AtSBT6.1* also has subtilase homologs from across Viridiplantae, including the chlorophyte alga *Uronema belkae* (order Chaetophorales), the prasinophycean green alga *Ostreococcus tauri*, the charophyte algae *Klebsormidium flaccidum*, *Chlorokybus atmophyticus*, *Cylindrocystis cushleckae*, and *Coleochaete scutata* (Fig. 1) (Adl et al. 2012). The clade containing *AtSBT6.2* is restricted to streptophytes, containing a subtilase homolog from the charophyte alga *Klebsormidium flaccidum* as well as land plants.

AtSBT6.1 and *AtSBT6.2* are characterized by a stronger similarity to the mammalian kexins and pyrolysins than to plant subtilases, whereas all other *Arabidopsis thaliana SBT* subfamilies do not partition with any of the known human prohormone convertases (Fig. 1) (Rautengarten et al. 2005). In a phylogenetic analysis of all members of *A. thaliana SBT1*, *SBT2*, and *SBT6* subfamilies with yeast *Kex2p* and the human prohormone convertases (furin, *SK1*), *AtSBT6.1* and *AtSBT6.2* were embedded among the yeast and human sequences (Rautengarten et al. 2005). We found the clades containing *AtSBT6.1* and *AtSBT6.2*

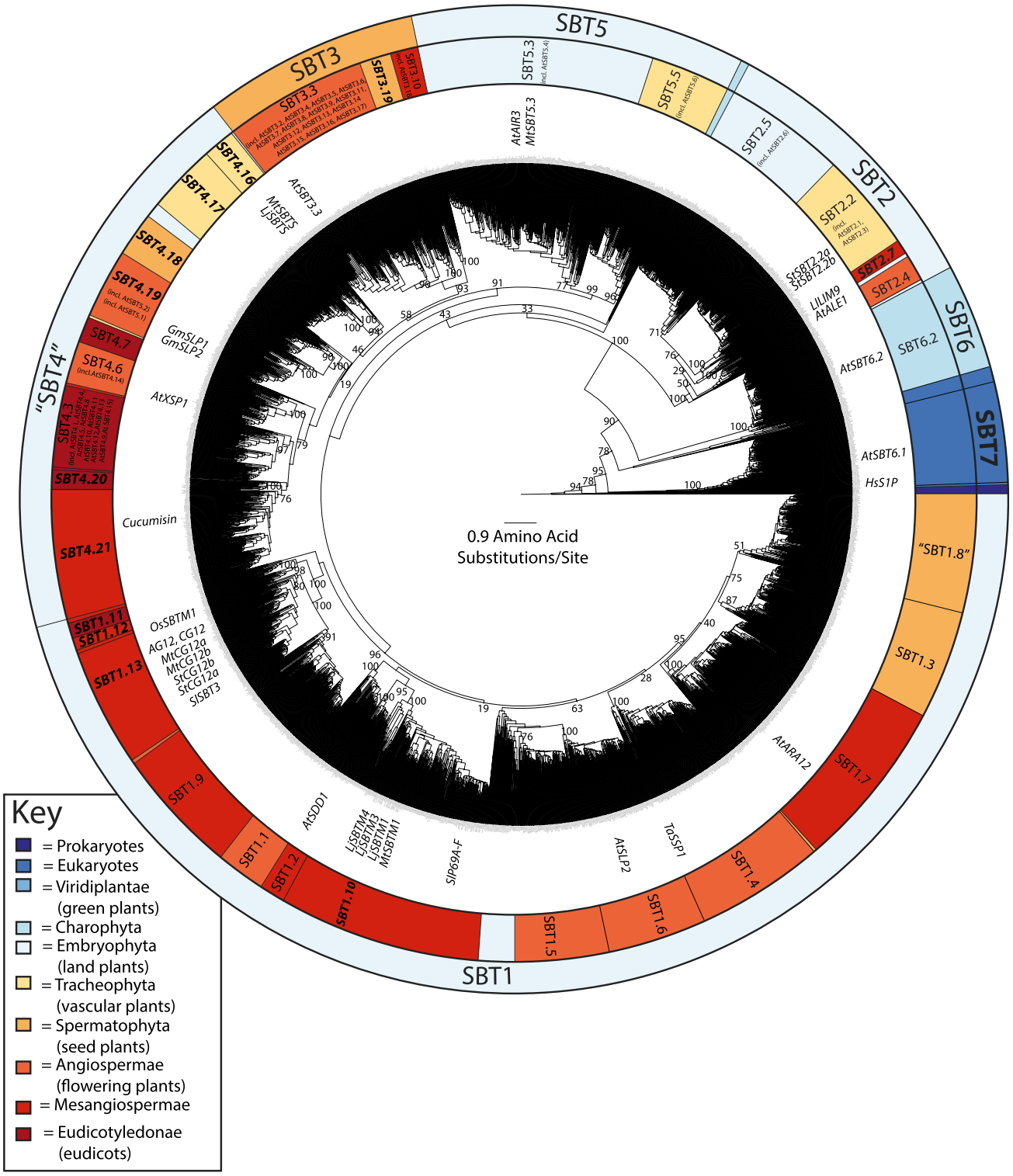

Fig. 1. Maximum-likelihood phylogenetic tree of 2,460 subtilases. including 2,441 subtilases in the Viridiplantae. Subtilases of interest are noted around the tree. Rings around the phylogeny delineate subtilase gene subfamilies and gene lineages, following the nomenclature of Rautengarten et al. (2005) when possible. Gene lineage names introduced in this paper are in bold italic font, names in quotation marks indicate less than 70% bootstrap support, and gene lineages containing many recently duplicated named *Arabidopsis thaliana* subtilase paralogs include those paralogs in parentheses. Phylogenetic depth of gene lineages is indicated by color in the key; criterion is the earliest-diverging plant lineage for which a subtilase homolog is present in that gene lineage. Magnification is approximately 2,000× for details of the tree.

to form a grade, with a clade between containing subtilase homologs from diverse eukaryotes, including animals (genera *Mus*, *Homo*), fungi (genera *Saccharomyces*, *Kluyveromyces*), diatom (genus *Fragilariopsis*), prasinophyte (genus *Nephroselmis)*, and a prymnesiophyte (genus *Emiliana*) (Fig. 1). Because the divergence of the clades containing *AtSBT6.1* and *AtSBT6.2* predates the Viridiplantae, it is no longer justifiable to place *AtSBT6.1* and *AtSBT6.2* in the same major group as Rautengarten et al. (2005) did. The clade containing *AtSBT6.1* is, thus, renamed *SBT7* in this study, with *SBT6* conserved for the clade containing *AtSBT6.2*. Both *SBT7* and *SBT6* seem to have low copy numbers throughout their phylogenetic distribution ranges, green plants and land plants, respectively (Fig. 1), and are well-represented in the 1KP dataset, suggesting that they are expressed in the young leaf and shoot tissue typically collected for the 1KP project.

The *SBT2* lineage originated early in the land plants, containing homologs in the liverworts *Pallavicinia lyelli* and *Frullania* sp., as well as the moss *Physcomitrella patens* (Fig. 1). The *SBT2* lineage likely underwent one round of duplication early in land plant evolution, as supported by the presence of bryophyte paralogs in the *SBT2.5* (*AtSLP3*) gene lineage and two parts at the base of the *SBT2* gene lineage (Fig. 1). The *SBT2.4* clade contains only angiosperm subtilases, including *AtALE1* (abnormal leaf shape 1), which is involved in cuticle formation and epidermal cell differentiation in embryos and juvenile *A. thaliana* plants (Tanaka et al. 2001). *LlLIM9*, which is induced during meiotic prophase in *Lilium longiflorum* (Kobayashi et al. 1994) during microspore development (Taylor et al. 1997), is monophyletic with subtilase sequences from several monocots and eudicots, suggesting that this gene lineage, now named *SBT2.7*, probably arose during the origin of mesangiosperms, a clade that includes all angiosperms, except two or three species-poor lineages at the base of angiosperm phylogeny, namely, Amborellales, Nymphaeales, and Austrobaileyales (Qiu et al. 2010; Soltis et al. 2011). Three characterized members of the *SBT2* lineage, *AtSBT2.5* (*AtSLP3*), *LlLIM9*, and *AtSBT2.4* (*AtALE1*), are all involved in tissue differentiation and organogenesis (Table 1), while *StSBT2.2a* and *StSBT2.2b* are found to be down- and up-regulated following infection with *P. infestans*, respectively (Table 1) (Gao et al. 2013). This lineage has retained relatively steady and low copy numbers through land plant evolution.

All other plant subtilases (*SBT1*, *SBT3*, *SBT4*, and *SBT5*) form a large, poorly supported clade (33% BS). If this lineage is real, it originated before the divergence of Embryophyta and Zygnematophyceae (Adl et al. 2012), as evidenced by the presence of subtilase homologs from three zygnematophycean green algae *Cylindrocystis cushleckae*, *Spirogyra* sp., and *Roya obtusa* in a basal position.

The *SBT5* gene clade contains two functionally characterized members. *AtAIR3*, involved in lateral root formation (Neuteboom et al. 1999), and *MtSBT5.3*, which is expressed in response to LCO signals from the nodulating bacteria *Sinorhizobium meliloti* (van Zeijl et al. 2015). *AtSBT5.3*, *AtSBT5.4*, *AtSBT5.5*, and *AtSBT5.6* form a well-supported monophyletic group (99% BS) that we call *SBT5*. This clade does not include *AtSBT5.1* and *AtSBT5.2*, which are in the *SBT4.19* clade. *SBT5* contains homologs mosses and liverworts, indicating an origin of this clade near the origin of land plants.

The *SBT3* clade is the only major clade of the subtilase gene family that has members exclusively in seed plants in our dataset, with the most basal taxon represented being the cycad *Encephalartos barteri* in the *SBT3.19* clade (Fig. 1). The only functionally characterized subtilase in this family is *AtSBT3.3*, which is involved in immune priming in *A. thaliana* (Ramírez et al. 2013). In *A. thaliana*, *SBT3* is the largest clade with 18 members (Rautengarten et al. 2005), but these genes seem to be derived from recent duplications that occurred within the *Arabidopsis* genus, the Brassicaceae family, or Brassicales order, as most of the 18 copies formed a moderately supported clade by themselves, not with any sequences from related species that are in our matrix, e.g., *Gossypium raimondii*, *Populus trichocarpa*, and *Ricinus communis*, which are in different orders (Fig. 1).

In the poorly-supported "*SBT4*" clade (46% BS), the earliest diverging subtilase lineage, *SBT4.16/SBT4.17*, appears to have originated in tracheophytes, with homologs in *Selaginella moellendorffii*, *Azolla caroliniana*, and an *Isoetes* sp. (Fig. 1). Several subtilases from the bryophyte grade are recovered as being nested in *SBT4*, but with low bootstrap support for specific placement, including homologs from the moss *Physcomitrella patens*, recovered as sister to *SBT4.18*, *SBT4.19*, *SBT4.7*, *SBT4.6*, *SBT4.3*, *SBT4.20*, and *SBT4.21* (19% BS), and a homolog from the hornwort *Nothoceros vincentianus* recovered to be sister to *SBT4.16* and *SBT4.17* (63% BS). *SBT4.16* includes *LjSBTS*, which is expressed during nodulation and AM in *L. japonicus* (Takeda et al. 2009), as well as a subtilase in *M. truncatula* that is upregulated in response to LCO signals from *Sinorhizobium meliloti* (van Zeijl et al. 2015), which we named *MtSBTS* to reflect its similarity to to *LjSBTS* in expression profile and phylogenetic position. The *SBT4* clade also contains *AtXSP1* (*AtSBT4.14*), involved in xylem differentiation (Zhao et al. 2000), the subtilase gene encoding cucumisin, which is involved in fruit ripening in *Cucumis melo* (Yamagata et al. 1994), and *GmSLP1* and *GmSLP2*, involved in seed coat development in *Glycine max* (Beilinson et al. 2002).

The *SBT1* clade, according to our analysis, is by far the largest of the subtilase subfamilies and clearly dates back to the beginning of land plant evolution. Two clades containing subtilases associated with symbiosis, *SBT1.10* and *SBT1.13*, account for almost one third of that expansion, which likely happened in the common ancestor of mesangiosperms (Fig. 1). Using only homolog counts from whole genomes, there are, on average, 21.69 *SBT1* homologs per species across land plants (compared with 11.78 *SBT4* homologs, the next largest clade), of which an average of 9.17 are in either the *SBT1.10* or *SBT1.13* lineages. The majority of previously characterized subtilases involved in symbiotic interactions are in the *SBT1* clade of the S8A subtilases, in the *SBT1.10* or *SBT1.13* clade, which are involved in symbiotic interactions (Table 1). Many of these subtilases exist as paralogous genes clustered in the genome, indicating tandem duplication (discussed below).

In addition to subtilases mediating biotic interactions, the *SBT1* clade contains several other characterized subtilases, mostly involved in developmental processes (Schaller et al. 2012). *SlP69D-F* genes are expressed during development in various tissues (Table 1) (Jordá et al. 1999, 2000). *AtSBT1.5* and *AtSBT1.6* are expressed constitutively in all cell types and are thought to be involved in nonspecific protein degradation and turnover (Rautengarten et al. 2005). In the *SBT1.4* lineage, *TaSSP1* is involved in protein degradation in senescing leaves of *Triticum aestivum* (Roberts et al. 2003). In the *SBT1.7* lineage, *AtARA12 (AtSLP1*, *AtSCS1)* is found in the intercellular spaces in *A. thaliana* stems and degrades proteins with little specificity (Hamilton et al. 2003), while in *Glycine max*, the ortholog of this gene is involved in seed coat development (Rautengarten et al. 2008). *SBT1.1* processes preproAtPSK4, a precursor for phytosulfokines, involved in mitogenic activity (Srivastava et al. 2008). *AtSBT1.2 (SDD1)* mediates stomatal density and distribution in *A. thaliana* (Berger and Altmann 2000).

Relationship among subtilases associated with symbiosis.

The majority of subtilase genes associated with symbiotic interactions fall into two newly identified distinct gene clades, *SBT1.10* (100% BS) and *SBT1.13* (100% BS) (Fig. 1; Table 1).

Both of these clades contain subtilase genes associated with both nodulation and AM as well as pathogenesis. While both clades have members from across the mesangiosperms, *SBT1.13* is embedded in a large, strongly supported clade with *SBT1.11*, *SBT1.12*, and *SBT1.9* (100% BS), containing a gene from the basal angiosperm *Nuphar advena* (Fig. 1), while *SBT1.10* is embedded in a clade with *SBT1.1* and *SBT1.2*, containing a gene from the basal angiosperm *Amborella trichopoda* (Fig. 1), suggesting a divergence between *SBT1.10* and *SBT1.13* early in angiosperm evolution. The two major symbiotic lineages are both in the *SBT1* subfamily, and their large size is likely due to many rounds of whole-genome and tandem duplication (discussed below). Neither of these symbiotic gene lineages was present or named in the gene phylogeny of Rautengarten et al. (2005), due to their absence in *A. thaliana*, a nonmycorrhizal and nonnodulating species.

The *SBT1.13* clade includes subtilases expressed during pathogenesis in *Solanum* spp., AM in *Oryza sativa*, nodulation in *Alnus glutinosa* and *Casuarina glauca*, and subtilases upregulated in response to rhizobial LCO signals in *M. truncatula*, as well as a subtilase involved in wound response to insect herbivory in *S. lycopersicum* (Fig. 1; Table 1). The *SBT1.10* clade includes subtilases expressed during pathogenesis in *Solanum* spp., AM in *L. japonicus*, and nodulation in the legumes *L. japonicus* and *M. truncatula* (Fig. 2; Table 1). Due to the absence of genomic and transcriptomic sequence data, it is unclear whether nodulating species in order Fagales have subtilase genes in the *SBT1.10* clade or whether these genes are expressed during nodulation.

With the ages of gene clades assessed from the corresponding plant clades, it is clear that appearance of both major symbiotic gene lineages significantly predates the origin of the NFC and, by extension, nodulation. This line of evidence, in turn, suggests that the genes recruited for nodulation were involved in different functions before plants acquired nodulating capability. Consistent with this interpretation, both of the major gene lineages containing nodulation-induced subtilases also include genes induced during AM development. *OsSBTM1*, which was weakly supported as being in the *SBT1.13* clade (34% BS), is expressed during AM formation in *Oryza sativa*, when the root is inoculated by *Glomus intraradices* (Table 1) (Güimil et al. 2005). In the *SBT1.10* clade, *LjSBTM4* is expressed in both AM and root nodule symbioses, suggesting an expansion of symbiont range to include nodulating bacteria (Kistner et al. 2005; Takeda et al. 2007, 2009, 2011). In both of these gene clades, subtilases expanded their expression to be induced by new symbionts; the function of these subtilases in restructuring plant cell walls may make these gene clades functionally labile in symbiotic interactions (Schaller et al. 2012).

Six subtilase genes associated with symbiosis did not fall into these two clades or into the *SBT1* clade and, instead, were scattered in four major clades that were in less-derived positions than *SBT1* (Fig. 1). *LjSBTS*, expressed during both nodulation and AM (Table 1) (Takeda et al. 2009), and *MtSBTS*, upregulated in response to LCO signals (Table 1) (van Zeijl et al. 2015), are members of a newly identified gene lineage, *SBT4.16* (100% BS). This gene clade likely arose during the origin of vascular plants, as evidenced by their presence in the lycophyte *Selaginella moellendorffii* (Fig. 1). Two *S. tuberosum* subtilases, *StSBT2.2a* and *StSBT2.2b*, found to be down- and up-regulated following infection with *P. infestans*, respectively (Table 1) (Gao et al. 2013), are in the *SBT2.2* clade. *MtSBT5.3*, a subtilase from *M. truncatula* that is upregulated in response to *Rhizobium* LCO signals (van Zeijl et al. 2015), is in the *SBT5.3* clade. *AtSBT3.3* in *A. thaliana* is involved in immune priming in response to *Pseudomonas syringae* and *Hyaloperonospora arabidopsidis* (Ramírez et al. 2013).

The above relationships of subtilase genes associated with symbiosis show a clear nonrandom distribution of these genes in the subtilase gene tree of green plants despite limited numbers of genes and organisms that have been characterized, since more than half the gene clades in the entire tree have no characterized members (Fig. 1). Subtilases have evolved roles in symbiotic interactions many times independently, but symbiosis-induced subtilases are concentrated in one major clade, *SBT1*, and, even in this clade, they are found only in two subclades, *SBT1.10* and *SBT1.13*, both of which extend to the beginning of mesangiosperm evolution and which diverged from one another even earlier. A series of tandem and whole-genome duplication events may help explain how such a distribution pattern arose.

Tandem and whole-genome duplication in the SBT1.10 clade.

New paralogs in the *SBT1.10* gene clade have been acquired independently in different plant lineages, primarily via tandem gene duplication but, also, through whole-genome duplication (Figs. 2 and 3). Subsequently, these paralogs have been recruited for new but related symbiotic interaction functions in different lineages, likely through neo- and subfunctionalization (He and Zhang 2005; Lynch and Force 2000; Ohno 1970), resulting in a gene lineage with paralogs involved in a variety of symbioses, including pathogenesis, AM, and nodulation (Fig. 1).

SBT1.10 is represented by multiple, lineage-specific paralogs in subfamily Papilionoideae (*SBTM1* and *SBTM3*) (Fig. 2B), family Rosaceae (Fig. 2C), order Malphigiales (Fig. 2D), and family Solanaceae (*P69A* to *P69F*) (Fig. 2F), whereas copy numbers of subtilase homologs in the *SBTM4* clade have remained relatively low during eudicot evolution (Figs. 2A and 3). The *SBTM1/M3* and *P69* paralogs arose from *SBTM4* early in eudicot evolution, before the divergence of asterids and rosids, as evidenced by the respective monophyly (Fig. 2E and C) of *SBTM1/M3* and *P69* paralogs and their synteny (Fig. 3A). In subfamily Papilionoideae, *SBTM1/SBTM3* duplicated in the ancestor of the subfamily at least once to produce *SBTM1* and *SBTM3*, as in the case of *Phaseolus vulgaris*, and many times in the case of the paralogs of *M. truncatula* (Fig. 2B).

To assess the mode of these duplications, we performed a synteny analysis of *SBT1.10* paralogs in selected taxa with well-annotated genomes (Fig. 3). In the genomes of *Fragaria vesca* (order Rosales) and *Populus trichocarpa* (Malphigiales), *SBT1.10* homologs underwent independent tandem duplication, resulting in multiple *SBT1.10* paralogs that are monophyletic in each order (Fig. 2, nodes C and D), and tandem with *SBTM4* (Fig. 3). Synteny is preserved between the regions containing *SBT1.10* and *SBTM4* paralogs in *F. vesca* and *Populus trichocarpa* (Fig. 3B).

In all sampled species in the Papilionoideae, *SBTM1* and *SBTM3* paralogs are found on a separate chromosome from *SBTM4*, as compared with their tandem arrangement in *Fragaria vesca* and *Populus trichocarpa* (Fig. 3). This evidence supports a scenario in which *SBTM1/M3* arose by tandem duplication from *SBTM4* before the divergence of orders Rosales and Fabales and *SBTM4* then ended up on a separate chromosome from *SBTM1* and *SBTM3* during a larger segmental or whole-genome duplication and subsequent pseudogenization. This derived condition, shared by all sampled papilionoids, likely occurred during the WGD event near the origin of subfamily Papilionoideae (Cannon et al. 2015; Vanneste et al. 2014), as is supported by the synteny of large chromosomal regions containing *SBTM4* with the regions containing *SBTM1* and *SBTM3* in subfamily Papilionoideae (Fig. 3C). Unlike other *sym* pathway genes that were retained and neofunctionalized or subfunctionalized after this whole-genome duplication (Vanneste et al. 2014), the ancestral tandem copy of *SBTM1/M3*

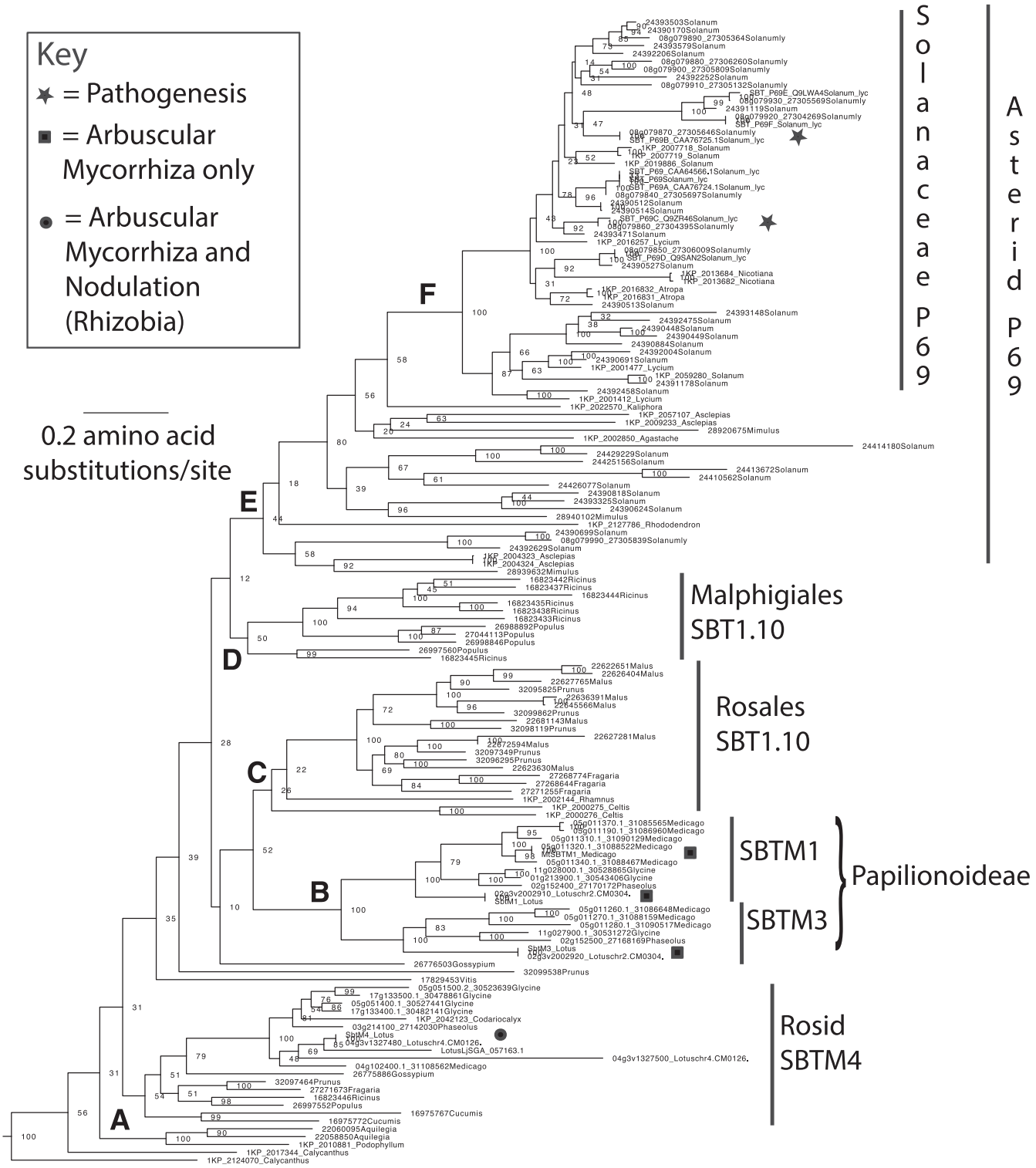

Fig. 2. A portion of the maximum-likelihood phylogenetic tree showing details of the SBT1.10 gene clade. Bootstrap values of 100 replicates are found at each node. Notable function characterizations are provided with symbols in the key. **A,** Orthologous gene lineage containing *LjSBTM4*. **B,** Gene lineage containing *LjSBTM1* and *LjSBTM3*, duplication leading to *LjSBTM1* and *LjSBTM3* occurring before origin Papilionoideae and restricted to this clade. **C,** Orthologous gene lineage of *SBT1.10* genes, specific to order Rosales. **D,** Orthologous gene lineage of SBT1.10 genes, containing multiple paralogs restricted to order Malphigiales. **E,** Gene lineage containing *P69* paralogs specific to asterids. **F,** Gene lineage containing all described P69 paralogs, which are restricted to the family Solanaceae. Gene names starting with letters were directly downloaded from the National Center for Biotechnology Information. Gene names starting with "1KP" are from the 1KP project, with 1KP sequence identification number and genus provided. All other genes are from full proteome data, contain Phytozome PAC or Kazusa (for *Lotus japonicus*) identification numbers, and begin with the *Arabidopsis* Genome Initiative (AGI) code if available in annotation. Solanum_tub = *Solanum tuberosum* and Solanum_lyc = *Solanum lycopersicum*.

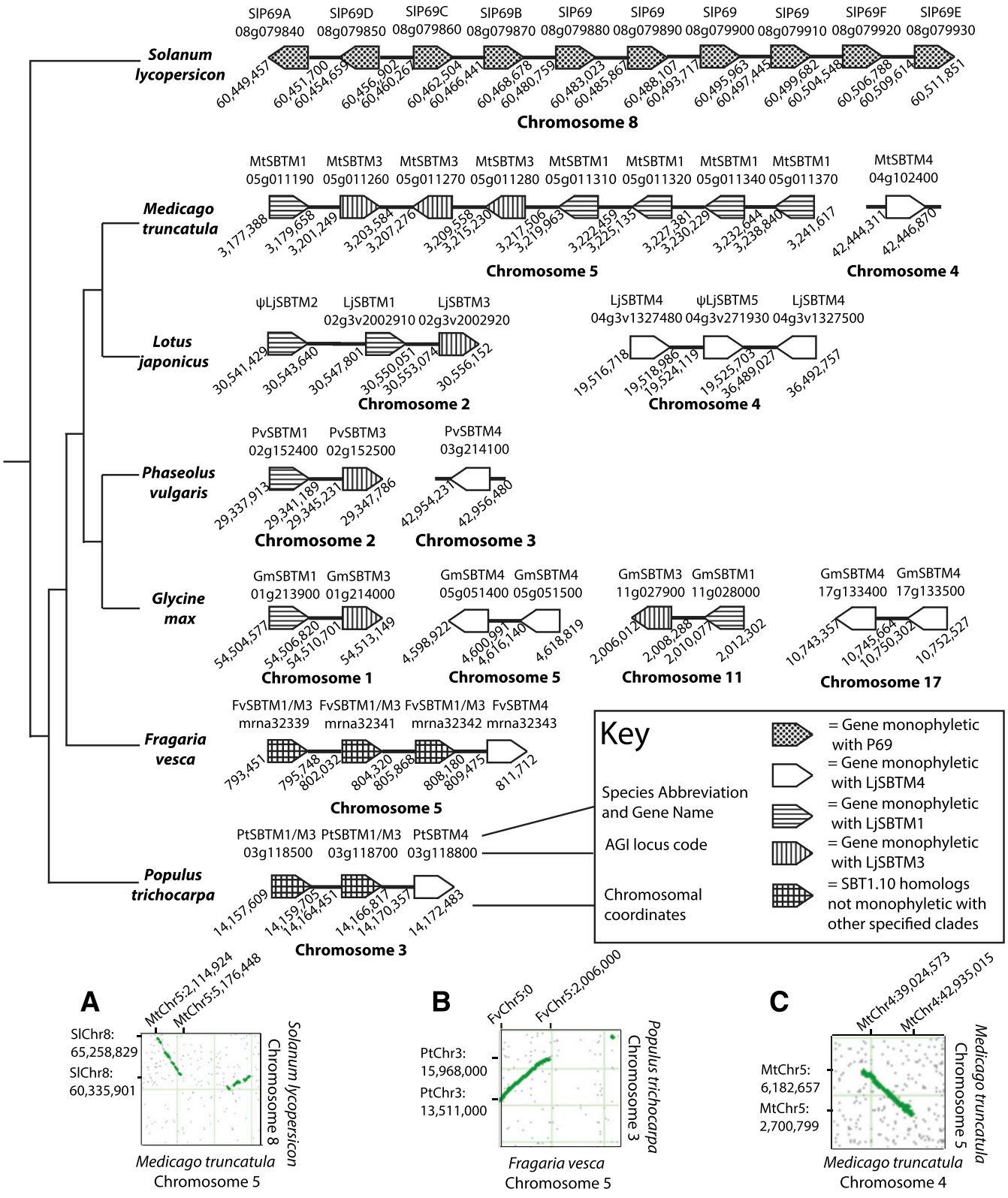

Fig. 3. Schematic representation (not to scale) showing syntenic relationships of SBT1.10 genes in eudicots, with different paralogs, coordinates, and strandedness, retrieved from Phytozome annotations. **A, B,** and **C,** Panels, retrieved from our synteny analyses in CoGe (comparative genomics) SynMap tool, showing retained synteny. **A,** Conserved synteny between *Solanum lycopersicum* chromosome 8 and *Medicago truncatula* chromosome 5 in the regions containing SBT1.10 paralogs. **B,** Conserved synteny between *Populus trichocarpa* chromosome 3 and *Fragaria vesca* chromosome 5, supporting a conserved state of tandem arrangement between *SBTM4* and *SBTM1*/*M3* between malvids and the nitrogen-fixing clade. **C,** Conserved synteny between *M. truncatula* chromosome 5 and chromosome 4, showing that the chromosomal region containing the *SBTM4* homolog shares ancestry with the chromosomal region containing *SBTM1*/*M3*.

near *SBTM4* in legumes was likely lost due to functional redundancy and vice-versa. This is supported by the presence of the pseudogenes ψ*SBTM2* tandem to *SBTM1* and *SBTM3* and ψ*SBTM5* tandem to *SBTM4* in *L. japonicus* (Takeda et al. 2009) (Fig. 3).

An exception to both the low copy number of *SBTM4* and the solely tandem duplication of *SBTM1* and *SBTM3* is found in *Glycine max*, in which tandem *SBTM1* and *SBTM3* paralogs are found in syntenic regions of chromosomes 1 and 11 (Fig. 3). Likewise, the four *SBTM4* paralogs are arranged as two tandem copies on two syntenic regions of chromosomes 5 and 17 (Fig. 3), suggesting one tandem duplication and then a subsequent duplication during the whole-genome duplication in the *Glycine* genus (Schmutz et al. 2010; Shoemaker et al. 2006). This is in keeping with the high copy number of subtilases in *Glycine max* generally, which has 104 subtilase paralogs against an average of 64 in the rosids, likely due to recent polyploidization in this genus (Schmutz et al. 2010; Shoemaker et al. 2006). *L. japonicus* has three copies of *SBTM4*, two of which are arranged tandemly and one quite distant on the same chromosome (Fig. 3). Because genera *Lotus* and *Glycine* belong to two separate major clades of the subfamily Papilionoideae (The Legume Phylogeny Working Group 2013) and species in basal lineages of both clades that have been sequenced, *P. vulgaris* and *M. truncatula*, have a single copy of *SBTM4*, the high copy number of the gene in *Glycine max* and *L. japonicus* is clearly the result of two recent, independent duplication events.

The retention of multiple *P69* paralogs across lamiids suggests that they have an adaptive function, perhaps coevolving with specific symbionts (Michelmore and Meyers 1998). This interpretation is supported by studies showing the coevolution of those pathogens with *P69* subtilases (Table 1) (Tian et al. 2004). Genes mediating pathogenesis and mutualism often show different evolutionary patterns, with those mediating mutualism remaining static over time, while those involved in pathogenesis showing accelerated rates of evolution, in an "arms race," though we did not recover that pattern here (Bravo et al. 2016; Kimbrel et al. 2013).

The pattern of widespread tandem duplication and neofunctionalization seen in these subtilases reflects a general pattern of evolution for host genes involved in coevolution with symbionts, particularly in organisms with innate (rather than adaptive) immune systems, and is found in the leucine-rich repeat class of resistance genes (Jones and Dangl 2006; Michelmore and Meyers 1998). This pattern has been proposed to create multiple paralogs that can each coevolve with specific symbionts, free of constraints of retaining functionality with ancestral symbionts, in a divergent selection regime (Michelmore and Meyers 1998). Other paralogs in the nodulation pathway, such as LysM-domain receptor kinases mediating Nod-factor reception in *L. japonicus* (*LjNFR1a*, *LjNFR1b*, and *LjNFR1c*), arose by tandem duplication (De Mita et al. 2014; Limpens et al. 2003). This legume-specific gene duplication has been proposed to account for derived traits of symbiont specificity in legumes, by allowing these paralogous receptors to expand host range and coevolve with different symbionts without constraint (Nakagawa et al. 2011; Radutoiu et al. 2007).

Whether duplications in genes recruited for nodulation occurred at a whole-genome scale or local scale determines the extent to which a genetic "predisposition" for nodulation can be claimed, for which derived states (e.g., crack entry versus root hair deformation), and at which nodes of the plant phylogeny. Some paralogs recruited for nodulation, such as the ethylene-responsive transcription factors *MtERN1* and *MtERN2* (Middleton et al. 2007), arose through whole-genome duplication in the papilionoid legumes (Vanneste et al. 2014). This wholesale duplication of all genes has been suggested as a possible genetic mechanism for an evolutionary predisposition for nodulation by supplying selectively unconstrained genetic material for neofunctionalization (Li et al. 2013; Soltis et al. 1995; Swensen and Mullin 1997; Werner et al. 2014), though recent work has cast doubt on this hypothesis (Cannon et al. 2015).

Conclusions.

Subtilases play a role in a wide variety of developmental processes in land plants through the processing and degradation of proteins in the apoplastic space between cells (Schaller et al. 2012). By incorporating genomic and transcriptomic data into a large dataset spanning land plants, our analysis shows that the subtilase gene family underwent multiple rounds of duplication and diversification, resulting in many subtilase clades with different functions. This diversification was particularly prominent in the angiosperms, to which 21 of the 33 named subtilase lineages are restricted (Fig. 1).

The ability of subtilases to restructure cell walls may be adaptive for a variety of symbiotic interactions with nodulating bacteria, AM fungi, and pathogens, as subtilases were recruited to mediate these interactions multiple times across land plant evolution (Fig. 1; Table 1). Here, we show that the majority of subtilases shown to mediate symbiotic interactions fall into two gene lineages, the *SBT1.10* lineage and the *SBT1.13* lineage. However, in both of these lineages, homologous subtilases mediate at least three different symbiotic interactions in different plant species.

Further, in the *SBT1.10* clade, patterns of duplication as well as neo- and subfunctionalization were specific to each nodulating lineage (Fig. 2B, C, and D). Most genes in the *sym* pathway are single-copy orthologs, but in those represented by more than one copy, the specific pattern of gene duplication and recruitment has been shown to have functional consequences for nodulation in different lineages (Op den Camp et al. 2011). Whole-genome duplications in the rosids have been proposed as a mechanism for the genetic predisposition to nodulation in the NFC (Li et al. 2013; Swensen and Mullin 1997; Vanneste et al. 2014). Tandem duplication and neofunctionalization is a common pattern in genes mediating biotic interactions (Michelmore and Meyers 1998) and has been proposed to account for synapomorphies (clade-specific derived traits) in the evolution of nodulation (De Mita et al. 2014; Radutoiu et al. 2007; Zhang et al. 2007).

Nodulation evolved via the recruitment of the pre-existing *sym* pathway, which mediates AM formation across land plants (Wang et al. 2010) and which has elements that originated before the divergence of charophytes and land plants (Delaux et al. 2015). Here, we show that the evolution of the subtilase gene family, which contributed to the evolutionary origin of nodulation, involved multiple rounds of duplication and changes in symbiont specificity across the gene phylogeny.

MATERIALS AND METHODS

Full proteome data were retrieved for 23 selected taxa across the land plant phylogeny from Phytozome v10.2 (Goodstein et al. 2012), and the *L. japonicus* proteome was retrieved from the Kazusa DNA Research Institute. These 24 full proteome sequences were assembled into a local BLAST+ database (Altschul et al. 1990), and an additional BLAST search was performed of the fully sequenced *Klebsormidium flaccidum* genome, to increase taxonomic sampling to 25 fully sequenced genomes across the Viridiplantae.

Amino acid sequences of 54 of the 56 subtilases from the *A. thaliana* subtilase gene phylogeny by Rautengarten et al. (2005) were retrieved from GenBank (National Center for Biotechnology Information [NCBI]); two sequences found to be pseudogenes in alignment, *AtSBT3.1* and *AtSBT4.2*, were excluded. Eight of the 54 *A. thaliana* subtilases have been further functionally characterized in other studies. A literature review of plant subtilases was performed to catalog characterized plant subtilases in species other than *A. thaliana*, of which 23 were

used in BLAST searches, for a total of 77 subtilases from the literature to capture the phylogenetic breadth of the subtilase. Additional expression data on subtilases was retrieved from the Expression Atlas project (Kapushesky et al. 2010). The 77 sequences from the literature were used for a local BLASTP query of the 24 full proteomes with an e-value cutoff of 0, to discover the full complement of subtilase homologs in each proteome. These were supplemented by a BLASTP query of transcriptomes from 316 taxa in the 1KP database (Matasci et al. 2014), for a total of 341 taxa sampled. An additional 19 subtilase sequences from 12 species outside of the Viridiplantae (one archaea, four bacteria, and seven eukaryotes) were added in order to determine the phylogenetic depth of ancient subtilase clades *SBT6* and *SBT7*.

After removal of duplicates and sequences under 300 amino acids, these searches yielded, in total, 2,460 subtilase amino acid sequences: 77 subtilases retrieved from the literature, 1,159 subtilases from 316 taxa in the 1KP database, 19 subtilases from outside the Viridiplantae kingdom retrieved from NCBI, and 1,205 subtilases retrieved from the 25 selected proteomes downloaded from Phytozome v10.2 and the Kazusa DNA Research Institute. A multiple sequence alignment of amino acid sequences was performed using CLUSTALO (Thompson et al. 1997).

Sequence alignments were uploaded onto the CIPRES Science Gateway (Miller et al. 2010) and a maximum likelihood phylogeny was constructed using RAxML (Stamatakis 2006), using a Dayhoff substitution model and 100 BS replicates (Dayhoff et al. 1978). The tree was rooted using subtilase sequences from *Bacillus amyloliquefaciens* as an outgroup. The generalized time reversible (GTR) substitution model was also tested, and topologies relevant to the conclusions presented here remained stable through multiple phylogenetic analyses with different substitution models (Lanave et al. 1984). The resulting gene trees were visually inspected and custom python scripts counting taxon names in tip names were used to record and count genes in orthologous gene lineages (orthogroups), in order to describe patterns of specific gene lineage expansions in different land plant clades.

The *A. thaliana* subtilase gene nomenclature proposed by Rautengarten et al. (2005) was used as a foundation for our nomenclature system for subtilases in kingdom Viridiplantae. In cases when a large, monophyletic, well-supported (BS value >70%) gene lineage did not have a representative in *A. thaliana*, we assigned names based on the number sequence used by Rautengarten et al. (2005). For example, the highest number in Rautengarten's *AtSBT4* clade was *AtSBT4.15*, so we named the first unnamed lineage in *SBT4* "*SBT4.16*." Some small or paraphyletic groups were left unnamed, for example, the small lineage sister to *SBT6*. For names of major clades in green plants, we followed Cavalier-Smith (1981) for Viridiplantae, Lewis and McCourt (2004) for Charophyta, Mishler and Qiu (in press) for Embryophyta, and Cantino et al. (2007) for Tracheophyta, Spermatophyta, Angiospermae, Mesangiospermae, and Eudicotyledoneae.

Synteny analysis of a large clade containing the majority of characterized symbiosis-induced subtilases was conducted to investigate patterns of duplication of these paralogs in a phylogenetic framework. The genomic position and strandedness of sequences in these orthogroups were identified in the most recent genome assemblies available from Phytozome v10.2 or the Kazusa DNA Research Institute (Supplementary Table S3). Syntenic relationships of these genes were investigated using CoGe (comparative genomics) tools SynMap (for chromosome-scale synteny) and GEvo (for fine-scale synteny).

LITERATURE CITED

Adl, S. M., Simpson, A. G., Lane, C. E., Lukeš, J., Bass, D., Bowser, S. S., Brown, M. W., Burki, F., Dunthorn, M., Hampl, V., Heiss, A., Hoppenrath, M., Lara, E., Le Gall, L., Lynn, D. H., McManus, H., Mitchell, E. A., Mozley-Stanridge, S. E., Parfrey, L. W., Pawlowski, J., Rueckert, S., Shadwick, L., Schoch, C. L., Smirnov, A., and Spiegel, F. W. 2012. The revised classification of eukaryotes. J. Eukaryot. Microbiol. 59:429-514.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403-410.

Beilinson, V., Moskalenko, O. V., Livingstone, D. S., Reverdatto, S. V., Jung, R., and Nielsen, N. C. 2002. Two subtilisin-like proteases from soybean. Physiol. Plant. 115:585-597.

Berger, D., and Altmann, T. 2000. A subtilisin-like serine protease involved in the regulation of stomatal density and distribution in *Arabidopsis thaliana*. Genes Dev. 14:1119-1131.

Bravo, A., York, T., Pumplin, N., Mueller, L. A., and Harrison M. J. 2016. Genes conserved for arbuscular mycorrhizal symbiosis identified through phylogenomics. Nature Plants 2:15208.

Cannon, S. B., McKain, M. R., Harkess, A., Nelson, M. N., Dash, S., Deyholos, M. K., Peng, Y., Joyce, B., Stewart, C. N., Jr., Rolf, M., Kutchan, T., Tan, X., Chen, C., Zhang, Y., Carpenter, E., Wong, G. K., Doyle, J. J., and Leebens-Mack, J. 2015. Multiple polyploidy events in the early radiation of nodulating and nonnodulating legumes. Mol. Biol. Evol. 32:193-210.

Cantino, P. D., Doyle, J. A., Graham, S. W., Judd, W. S., Olmstead, R. G., Soltis, D. E., Soltis, P. S., and Donoghue, D. J. 2007. Towards a phylogenetic nomenclature of Tracheophyta. Taxon 56:822-846.

Cavalier-Smith, T. 1981. Eukaryote kingdoms: Seven or nine? Biosystems 14:461-481.

Dayhoff, M., Schwartz, R., and Orcutt, B. 1978. A model of evolutionary change in protein. Atlas Protein Seq. Struct 5:345-352.

De Mita, S., Streng, A., Bisseling, T., and Geurts, R. 2014. Evolution of a symbiotic receptor through gene duplications in the legume-rhizobium mutualism. New Phytol. 201:961-972.

Delaux, P. M., Radhakrishnan, G. V., Jayaraman, D., Cheema, J., Malbreil, M., Volkening, J. D., Sekimoto, H., Nishiyama, T., Melkonian, M., Pokorny, L., Rothfels, C. J., Sederoff, H. W., Stevenson, D. W., Surek, B., Zhang, Y., Sussman, M. R., Dunand, C., Morris, R. J., Roux, C., Wong, G. K., Oldroyd, G. E., and Ané, J. M. 2015. Algal ancestor of land plants was preadapted for symbiosis. Proc. Natl. Acad. Sci. U.S.A. 112: 13390-13395.

Demina, I. V., Persson, T., Santos, P., Plaszczyca, M., and Pawlowski, K. 2013. Comparison of the nodule vs. root transcriptome of the actinorhizal plant *Datisca glomerata*: Actinorhizal nodules contain a specific class of defensins. PLoS One 8:e72442.

Doyle, J. J. 1994. Phylogeny of the legume family: An approach to understanding the origins of nodulation. Annu. Rev. Ecol. Syst. 25: 325-349.

Doyle, J. J. 2011. Phylogenetic perspectives on the origins of nodulation. Mol. Plant-Microbe Interact. 24:1289-1295.

Finet, C., Timme, R. E., Delwiche, C. F., and Marlétaz, F. 2010. Multigene phylogeny of the green lineage reveals the origin and diversification of land plants. Curr. Biol. 20:2217-2222.

Gao, L., Tu, Z. J., Millett, B. P., and Bradeen, J. M. 2013. Insights into organ-specific pathogen defense responses in plants: RNA-seq analysis of potato tuber-*Phytophthora infestans* interactions. BMC Genomics 14: 340.

Genre, A., Chabaud, M., Timmers, T., Bonfante, P., and Barker, D. G. 2005. Arbuscular mycorrhizal fungi elicit a novel intracellular apparatus in *Medicago truncatula* root epidermal cells before infection. Plant Cell 17: 3489-3499.

Gherbi, H., Markmann, K., Svistoonoff, S., Estevan, J., Autran, D., Giczey, G., Auguy, F., Péret, B., Laplaze, L., Franche, C., Parniske, M., and Bogusz, D. 2008. SymRK defines a common genetic basis for plant root endosymbioses with arbuscular mycorrhiza fungi, rhizobia, and *Frankia*bacteria. Proc. Natl. Acad. Sci. U.S.A. 105:4928-4932.

Ghorbani, S., Hoogewijs, K., Pečenková, T., Fernandez, A., Inzé, A., Eeckhout, D., Kawa, D., De Jaeger, G., Beeckman, T., Madder, A., Van Breusegem, F., and Hilson, P. 2016. The SBT6.1 subtilase processes the GOLVEN1 peptide controlling cell elongation. J. Exp. Bot. 67:4877-4887.

Golldack, D., Vera, P., and Dietz, K. -J. Expression of subtilisin-like serine proteases in *Arabidopsis thaliana* is cell- specific and responds to jasmonic acid and heavy metals with developmental differences. Physiol. Plant. 118: 64-73.

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., and Rokhsar, D. S. 2012. Phytozome: A comparative platform for green plant genomics. Nucleic Acids Res. 40:D1178-D1186.

Güimil, S., Chang, H. S., Zhu, T., Sesma, A., Osbourn, A., Roux, C., Ioannidis, V., Oakeley, E. J., Docquier, M., Descombes, P., Briggs, S. P., and Paszkowski, U. 2005. Comparative transcriptomics of rice reveals an

ancient pattern of response to microbial colonization. Proc. Natl. Acad. Sci. U.S.A. 102:8066-8070.
Hamilton, J. M., Simpson, D. J., Hyman, S. C., Ndimba, B. K., and Slabas, A. R. 2003. Ara12 subtilisin-like protease from *Arabidopsis thaliana*: Purification, substrate specificity and tissue localization. Biochem. J. 370:57-67.
He, X., and Zhang, J. 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. Genetics 169:1157-1164.
Hocher, V., Alloisio, N., Auguy, F., Founier, P., Doumas, P., Pujic, P., and Gherbi, H. 2011. Transcriptomics of actinorhizal symbioses reveals homologs of the whole common symbiotic signaling cascade. Plant Physiol. 156:700-711.
Hocher, V., Auguy, F., Argout, X., Laplaze, L., Franche, C., and Bogusz, D. 2006. Expressed sequence-tag analysis in *Casuarina glauca* actinorhizal nodule and root. New Phytol. 169:681-688.
Jones, J. D., and Dangl, J. L. 2006. The plant immune system. Nature 444: 323-329.
Jordá, L., Coego, A., Conejero, V., and Vera, P. 1999. A genomic cluster containing four differentially regulated subtilisin-like processing protease genes is in tomato plants. J. Biol. Chem. 274:2360-2365.
Jordá, L., Conejero, V., and Vera, P. 2000. Characterization of *P69E* and *P69F*, two differentially regulated genes encoding new members of the subtilisin-like proteinase family from tomato plants. Plant Physiol. 122: 67-74.
Kapushesky, M., Emam, I., Holloway, E., Kurnosov, P., Zorin, A., Malone, J., Rustici, G., Williams, E., Parkinson, H., and Brazma, A. 2010. Gene Expression Atlas at the European bioinformatics institute. Nucleic Acids Res. 38:D690-D698.
Kimbrel, J. A., Thomas, W. J., Jiang, Y., Creason, A. L., Thireault, C. A., Sachs, J. L., and Chang, J. H. 2013. Mutualistic co-evolution of type III effector genes in *Sinorhizobium fredii* and *Bradyrhizobium japonicum*. PLoS Pathog. 9:e1003204.
Kistner, C., and Parniske, M. 2002. Evolution of signal transduction in intracellular symbiosis. Trends Plant Sci. 7:511-518.
Kistner, C., Winzer, T., Pitzschke, A., Mulder, L., Sato, S., Kaneko, T., Tabata, S., Sandal, N., Stougaard, J., Webb, K. J., Szczyglowski, K., and Parniske, M. 2005. Seven *Lotus japonicus* genes required for transcriptional reprogramming of the root during fungal and bacterial symbiosis. Plant Cell 17:2217-2229.
Kobayashi, T., Kobayashi, E., Sato, S., Hotta, Y., Miyajima, N., Tanaka, A., and Tabata, S. 1994. Characterization of cDNAs induced in meiotic prophase in lily microsporocytes. DNA Res. 1:15-26.
Lanave, C., Preparata, G., Saccone, C., and Serio, G. 1984. A new method for calculating evolutionary substitution rates. J. Mol. Evol. 20:86-93.
Laplaze, L., Ribeiro, A., Franche, C., Duhoux, E., Auguy, F., Bogusz, D., and Pawlowski, K. 2000. Characterization of a *Casuarina glauca* nodule-specific subtilisin-like protease gene, a homolog of *Alnus glutinosa* ag12. Mol. Plant-Microbe Interact. 13:113-117.
Lewis, L. A., and McCourt, R. M. 2004. Green algae and the origin of land plants. Am. J. Bot. 91:1535-1556.
Li, Q. G., Zhang, L., Li, C., Dunwell, J. M., and Zhang, Y. M. 2013. Comparative genomics suggests that an ancestral polyploidy event leads to enhanced root nodule symbiosis in the Papilionoideae. Mol. Biol. Evol. 30:2602-2611.
Limpens, E., Franken, C., Smit, P., Willemse, J., Bisseling, T., and Geurts, R. 2003. LysM domain receptor kinases regulating rhizobial Nod factor-induced infection. Science 302:630-633.
Liu, J., Blaylock, L. A., Endre, G., Cho, J., Town, C. D., VandenBosch, K. A., and Harrison, M. J. 2003. Transcript profiling coupled with spatial expression analyses reveals genes involved in distinct developmental stages of an arbuscular mycorrhizal symbiosis. Plant Cell 15:2106-2123.
Liu, J.-X., Srivastava, R., Che, P., and Howell, S. H. 2007. Salt stress responses in Arabidopsis utilize a signal transduction pathway related to endoplasmic reticulum stress signaling. Plant J. 51:897-909.
Lynch, M., and Force, A. 2000. The probability of duplicate gene preservation by subfunctionalization. Genetics 154:459-473.
Maillet, F., Poinsot, V., André, O., Puech-Pagès, V., Haouy, A., Gueunier, M., Cromer, L., Giraudet, D., Formey, D., Niebel, A., Martinez, E. A., Driguez, H., Bécard, G., and Dénarié, J. 2011. Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. Nature 469:58-63.
Matasci, N., Hung, L. H., Yan, Z., Carpenter, E. J., Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., Ayyampalayam, S., Barker, M., Burleigh, J. G., Gitzendanner, M. A., Wafula, E., Der, J. P., dePamphilis, C. W., Roure, B., Philippe, H., Ruhfel, B. R., Miles, N. W., Graham, S. W., Mathews, S., Surek, B., Melkonian, M., Soltis, D. E., Soltis, P. S., Rothfels, C., Pokorny, L., Shaw, J. A., DeGironimo, L., Stevenson, D. W., Villarreal, J. C., Chen, T., Kutchan, T. M., Rolf, M., Baucom, R. S., Deyholos, M. K., Samudrala, R., Tian, Z., Wu, X., Sun, X., Zhang, Y., Wang, J., Leebens-Mack, J., and Wong, G. K. 2014. Data access for the 1,000 Plants (1KP) project. Gigascience 3:17.
Meyer, M., Huttenlocher, F., Cedzich, A., Procopio, S., Stroeder, J., Pau-Roblot, C., Lequart-Pillon, M., Pelloux, J., Stintzi, A., and Schaller, A. 2016. The subtilisin-like protease SBT3 contributes to insect resistance in tomato. J. Exp. Bot. 67:4325-4338.
Michelmore, R. W., and Meyers, B. C. 1998. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res. 8:1113-1130.
Middleton, P. H., Jakab, J., Penmetsa, R. V., Starker, C. G., Doll, J., Kaló, P., Prabhu, R., Marsh, J. F., Mitra, R. M., Kereszt, A., Dudas, B., VandenBosch, K., Long, S. R., Cook, D. R., Kiss, G. B., and Oldroyd, G. E. 2007. An ERF transcription factor in *Medicago truncatula* that is essential for Nod factor signal transduction. Plant Cell 19:1221-1234.
Miller, M. A., Pfeiffer, W., and Schwartz, T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Pages 1-8 in: Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA. IEEE, New York.
Mishler, B. D., and Qiu, Y.-L. Embryophyta. Phylonyms: A companion to the PhyloCode. K. de Queiroz, P. D. Cantino, and J. Gauthier, eds. University of California Press, Berkeley. In Press.
Nakagawa, T., Kaku, H., Shimoda, Y., Sugiyama, A., Shimamura, M., Takanashi, K., Yazaki, K., Aoki, T., Shibuya, N., and Kouchi, H. 2011. From defense to symbiosis: Limited alterations in the kinase domain of LysM receptor-like kinases are crucial for evolution of legume-*Rhizobium* symbiosis. Plant J. 65:169-180.
Neuteboom, L. W., Veth-Tello, L. M., Clijdesdale, O. R., Hooykaas, P. J., and van der Zaal, B. J. 1999. A novel subtilisin-like protease gene from *Arabidopsis thaliana* is expressed at sites of lateral root emergence. DNA Res. 6:13-19.
Ohno, S. 1970. Evolution by Gene Duplication. Springer-Verlag, New York.
Oldroyd, G. E. 2013. Speak, friend, and enter: Signalling systems that promote beneficial symbiotic associations in plants. Nat. Rev. Microbiol. 11:252-263.
Op den Camp, R., Streng, A., De Mita, S., Cao, Q., Polone, E., Liu, W., Ammiraju, J. S., Kudrna, D., Wing, R., Untergasser, A., Bisseling, T., and Geurts, R. 2011. LysM-type mycorrhizal receptor recruited for *rhizobium* symbiosis in nonlegume *Parasponia*. Science 331:909-912.
Parniske, M. 2008. Arbuscular mycorrhiza: The mother of plant root endosymbioses. Nat. Rev. Microbiol. 6:763-775.
Pawlowski, K., Bogusz, D., Ribeiro, A., and Berry, A. M. 2011. Progress on research on actinorhizal plants. Funct. Plant Biol. 38:633-638.
Pawlowski, K., and Sprent, J. I. 2008. Comparison between actinorhizal and legume symbiosis. Pages 261-288 in: Nitrogen-fixing actinorhizal symbioses. K. Pawlowski and W. E. Newton, eds. Springer Netherlands, Dordrecht, The Netherlands.
Qiu, Y.-L., Li, L., Wang, B., Chen, Z., Knoop, V., Groth-Malonek, M., Dombrovska, O., Lee, J., Kent, L., Rest, J., Estabrook, G. F., Hendry, T. A., Taylor, D. W., Testa, C. M., Ambros, M., Crandall-Stotler, B., Duff, R. J., Stech, M., Frey, W., Quandt, D., and Davis, C. C. 2006. The deepest divergences in land plants inferred from phylogenomic evidence. Proc. Natl. Acad. Sci. U.S.A. 103:15511-15516.
Qiu, Y.-L., Li, L., Wang, B., Xue, J. Y., Hendry, T. A., and Li, R. Q. 2010. Angiosperm phylogeny inferred from sequences of four mitochondrial genes. J. Syst. Evol. 48:391-425.
Radutoiu, S., Madsen, L. H., Madsen, E. B., Jurkiewicz, A., Fukai, E., Quistgaard, E. M. H., Albrektsen, A. S., James, E. K., Thirup, S., and Stougaard, J. 2007. LysM domains mediate lipochitin-oligosaccharide recognition and *Nfr* genes extend the symbiotic host range. EMBO J. 26: 3923-3935.
Ramírez, V., López, A., Mauch-Mani, B., Gil, M. J., and Vera, P. 2013. An extracellular subtilase switch for immune priming in *Arabidopsis*. PLoS Pathog. 9:e1003445.
Rautengarten, C., Steinhauser, D., Büssis, D., Stintzi, A., Schaller, A., Kopka, J., and Altmann, T. 2005. Inferring hypotheses on functional relationships of genes: Analysis of the *Arabidopsis thaliana* subtilase gene family. PLOS Comput. Biol. 1:e40.
Rautengarten, C., Usadel, B., Neumetzler, L., Hartmann, J., Büssis, D., and Altmann, T. 2008. A subtilisin-like serine protease essential for mucilage release from Arabidopsis seed coats. Plant J. 54:466-480.
Ribeiro, A., Akkermans, A. D. L., van Kammen, A., Bisseling, T., and Pawlowski, K. 1995. A nodule-specific gene encoding a subtilisin-like protease is expressed in early stages of actinorhizal nodule development. Plant Cell 7:785-794.
Roberts, I. N., Murray, P. F., Caputo, C. P., Passeron, S., and Barneix, A. J. 2003. Purification and characterization of a subtilisin-like serine protease induced during the senescence of wheat leaves. Physiol. Plant. 118:483-490.

Schaller, A., Stintzi, A., and Graff, L. 2012. Subtilases—Versatile tools for protein turnover, plant development, and interactions with the environment. Physiol. Plant. 145:52-66.

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D. L., Song, Q., Thelen, J. J., Cheng, J., Xu, D., Hellsten, U., May, G. D., Yu, Y., Sakurai, T., Umezawa, T., Bhattacharyya, M. K., Sandhu, D., Valliyodan, B., Lindquist, E., Peto, M., Grant, D., Shu, S., Goodstein, D., Barry, K., Futrell-Griggs, M., Abernathy, B., Du, J., Tian, Z., Zhu, L., Gill, N., Joshi, T., Libault, M., Sethuraman, A., Zhang, X. C., Shinozaki, K., Nguyen, H. T., Wing, R. A., Cregan, P., Specht, J., Grimwood, J., Rokhsar, D., Stacey, G., Shoemaker, R. C., and Jackson, S. A. 2010. Genome sequence of the palaeopolyploid soybean. Nature 463:178-183.

Shoemaker, R. C., Schlueter, J., and Doyle, J. J. 2006. Paleopolyploidy and gene duplication in soybean and other legumes. Curr. Opin. Plant Biol. 9: 104-109.

Shubin, N., Tabin, C., and Carroll, S. 1997. Fossils, genes and the evolution of animal limbs. Nature 388:639-648.

Shubin, N., Tabin, C., and Carroll, S. 2009. Deep homology and the origins of evolutionary novelty. Nature 457:818-823.

Siezen, R. J., and Leunissen, J. A. 1997. Subtilases: The superfamily of subtilisin-like serine proteases. Protein Sci. 6:501-523.

Soltis, D. E., Smith, S. A., Cellinese, N., Wurdack, K. J., Tank, D. C., Brockington, S. F., Refulio-Rodriguez, N. F., Walker, J. B., Moore, M. J., Carlsward, B. S., Bell, C. D., Latvis, M., Crawley, S., Black, C., Diouf, D., Xi, Z., Rushworth, C. A., Gitzendanner, M. A., Sytsma, K. J., Qiu, Y. L., Hilu, K. W., Davis, C. C., Sanderson, M. J., Beaman, R. S., Olmstead, R. G., Judd, W. S., Donoghue, M. J., and Soltis, P. S. 2011. Angiosperm phylogeny: 17 genes, 640 taxa. Am. J. Bot. 98:704-730.

Soltis, D. E., Soltis, P. S., Morgan, D. R., Swensen, S. M., Mullin, B. C., Dowd, J. M., and Martin, P. G. 1995. Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. Proc. Natl. Acad. Sci. U.S.A. 92:2647-2651.

Srivastava, R., Liu, J. X., and Howell, S. H. 2008. Proteolytic processing of a precursor protein for a growth-promoting peptide by a subtilisin serine protease in *Arabidopsis*. Plant J. 56:219-227.

Stamatakis, A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688-2690.

Stracke, S., Kistner, C., Yoshida, S., Mulder, L., Sato, S., Kaneko, T., Tabata, S., Sandal, N., Stougaard, J., Szczyglowski, K., and Parniske, M. 2002. A plant receptor-like kinase required for both bacterial and fungal symbiosis. Nature 417:959-962.

Svistoonoff, S., Hocher, V., and Gherbi, H. 2014. Actinorhizal root nodule symbioses: What is signalling telling on the origins of nodulation? Curr. Opin. Plant Biol. 20:11-18.

Svistoonoff, S., Laplaze, L., Auguy, F., Runions, J., Duponnois, R., Haseloff, J., Franche, C., and Bogusz, D. 2003. *cg12* expression is specifically linked to infection of root hairs and cortical cells during *Casuarina glauca* and *Allocasuarina verticillata* actinorhizal nodule development. Mol. Plant-Microbe Interact. 16:600-607.

Svistoonoff, S., Laplaze, L., Liang, J., Ribeiro, A., Gouveia, M. C., Auguy, F., Fevereiro, P., Franche, C., and Bogusz, D. 2004. Infection-related activation of the *cg12* promoter is conserved between actinorhizal and legume-rhizobia root nodule symbiosis. Plant Physiol. 136:3191-3197.

Swensen, S. M., and Mullin, B. C. 1997. The impact of molecular systematics on hypotheses for the evolution of root nodule symbioses and implications for expanding symbioses to new host plant genera. Plant Soil 194:185-192.

Szczyglowski, K., Shaw, R. S., Wopereis, J., Copeland, S., Hamburger, D., Kasiborski, B., and Dazzo, F. B. 1998. Nodule organogenesis and symbiotic mutants of the model legume *Lotus japonicus*. Mol. Plant Microbe Interact. 11:684-697.

Takeda, N., Haage, K., Sato, S., Tabata, S., and Parniske, M. 2011. Activation of a *Lotus japonicus* subtilase gene during arbuscular mycorrhiza is dependent on the common symbiosis genes and two cis-active promoter regions. Mol. Plant-Microbe Interact. 24:662-670.

Takeda, N., Kistner, C., Kosuta, S., Winzer, T., Pitzschke, A., Groth, M., Sato, S., Kaneko, T., Tabata, S., and Parniske, M. 2007. Proteases in plant root symbiosis. Phytochemistry 68:111-121.

Takeda, N., Sato, S., Asamizu, E., Tabata, S., and Parniske, M. 2009. Apoplastic plant subtilases support arbuscular mycorrhiza development in *Lotus japonicus*. Plant J. 58:766-777.

Tanaka, H., Onouchi, H., Kondo, M., Hara-Nishimura, I., Nishimura, M., Machida, C., and Machida, Y. 2001. A subtilisin-like serine protease is required for epidermal surface formation in *Arabidopsis* embryos and juvenile plants. Development 128:4681-4689.

Taylor, A. A., Horsch, A., Rzepczyk, A., Hasenkampf, C. A., and Riggs, C. D. 1997. Maturation and secretion of a serine proteinase is associated with events of late microsporogenesis. Plant J. 12:1261-1271.

The Legume Phylogeny Working Group. 2013. Legume phylogeny and classification in the 21st century: Progress, prospects and lessons for other species-rich clades. Taxon 62:217-248.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. 1997. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25:4876-4882.

Tian, M., Benedetti, B., and Kamoun, S. 2005. A second Kazal-like protease inhibitor from *Phytophthora infestans* inhibits and interacts with the apoplastic pathogenesis-related protease P69B of tomato. Plant Physiol. 138:1785-1793.

Tian, M., Huitema, E., Da Cunha, L., Torto-Alalibo, T., and Kamoun, S. 2004. A Kazal-like extracellular serine protease inhibitor from *Phytophthora infestans* targets the tomato pathogenesis-related protease P69B. J. Biol. Chem. 279:26370-26377.

Tornero, P., Conejero, V., and Vera, P. 1996. Primary structure and expression of a pathogen-induced protease (PR-P69) in tomato plants: Similarity of functional domains to subtilisin-like endoproteases. Proc. Natl. Acad. Sci. U.S.A. 93:6332-6337.

Tornero, P., Conejero, V., and Vera, P. 1997. Identification of a new pathogen-induced member of the subtilisin-like processing protease family from plants. J. Biol. Chem. 272:14412-14419.

True, J. R., and Carroll, S. B. 2002. Gene co-option in physiological and morphological evolution. Annu. Rev. Cell Dev. Biol. 18:53-80.

van Zeijl, A., Op den Camp, R. H., Deinum, E. E., Charnikhova, T., Franssen, H., Op den Camp, H. J., Bouwmeester, H., Kohlen, W., Bisseling, T., and Geurts, R. 2015. *Rhizobium* lipo-chitooligosaccharide signaling triggers accumulation of cytokinins in *Medicago truncatula* roots. Mol. Plant 8:1213-1226.

Vanneste, K., Maere, S., and Van de Peer, Y. 2014. Tangled up in two: A burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution. Philos. Trans. R. Soc. Lond. B Biol. Sci. 369:20130353.

Wang, B., and Qiu, Y. L. 2006. Phylogenetic distribution and evolution of mycorrhizas in land plants. Mycorrhiza 16:299-363.

Wang, B., Yeun, L. H., Xue, J. Y., Liu, Y., Ané, J. M., and Qiu, Y. L. 2010. Presence of three mycorrhizal genes in the common ancestor of land plants suggests a key role of mycorrhizas in the colonization of land by plants. New Phytol. 186:514-525.

Werner, G. D., Cornwell, W. K., Sprent, J. I., Kattge, J., and Kiers, E. T. 2014. A single evolutionary innovation drives the deep evolution of symbiotic N2-fixation in angiosperms. Nature Commun. 5:4087.

Yamagata, H., Masuzawa, T., Nagaoka, Y., Ohnishi, T., and Iwasaki, T. 1994. Cucumisin, a serine protease from melon fruits, shares structural homology with subtilisin and is generated from a large precursor. J. Biol. Chem. 269:32725-32731.

Zhang, X. C., Wu, X., Findley, S., Wan, J., Libault, M., Nguyen, H. T., Cannon, S. B., and Stacey, G. 2007. Molecular evolution of lysin motif-type receptor-like kinases in plants. Plant Physiol. 144:623-636.

Zhao, C., Johnson, B. J., Kositsup, B., and Beers, E. P. 2000. Exploiting secondary growth in Arabidopsis. Construction of xylem and bark cDNA libraries and cloning of three xylem endopeptidases. Plant Physiol. 123: 1185-1196.

AUTHOR-RECOMMENDED INTERNET RESOURCES

CoGe platform: https://genomevolution.org/coge
Expression Atlas project website: https://www.ebi.ac.uk/gxa