Molecular Biology and Evolution

Dramatic Change in Function and Expression Pattern of a Gene Duplicated by Polyploidy Created a Paternal Effect Gene in the Brassicaceae

Shao-Lun Liu[1,2] and Keith L. Adams*[,1,2]

[1]UBC Botanical Garden and Centre for Plant Research, Vancouver, British Columbia, Canada

[2]Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada

***Corresponding author:** E-mail: keitha@interchange.ubc.ca.

Associate editor: Naoki Takebayashi

Abstract

New gene formation by polyploidy has been an ongoing process during the evolution of various eukaryotes that has contributed greatly to the large number of genes in their genomes. After duplication, some genes that are retained can acquire new functions or expression patterns, or subdivide their functions or expression patterns between duplicates. Here, we show that *SHORT SUSPENSOR* (*SSP*) and *Brassinosteroid Kinase 1* (*BSK1*) are paralogs duplicated by a polyploidy event that occurred in the Brassicaceae family about 23 Ma. *SSP* is involved in paternal control of zygote elongation in *Arabidopsis thaliana* by transcription in the sperm cells of pollen and then translation in the zygote, whereas *BSK1* is involved in brassinosteroid signal transduction. Comparative analysis of expression in 63 different organs and developmental stages revealed that *BSK1* and *SSP* have opposite expression patterns in pollen compared with all other parts of the plant. We determined that *BSK1* retains the ancestral expression pattern and function. Thus, *SSP* has diverged in function after duplication from a component of the brassinosteroid signaling pathway to a paternal regulator of the timing of zygote elongation. The ancestral function of *SSP* was lost by deletions in the kinase domain. Our sequence rate analysis revealed that *SSP* but not *BSK1* has experienced a greatly accelerated rate of amino acid sequence changes and relaxation of purifying selection. In addition, *SSP* has been duplicated to create a new gene (*SSP-like1*) with a completely different expression pattern, a shorter coding sequence that has lost a critical functional domain, and a greatly accelerated rate of amino acid sequence evolution along with evidence for positive selection, together indicative of neofunctionalization. This study illustrates two dramatic examples of neofunctionalization following gene duplication by complete changes in expression pattern and function. In addition, our findings indicate that paternal control of zygote elongation by *SSP* is an evolutionarily recent innovation in the Brassicaceae family.

Key words: gene duplication, polyploidy, neofunctionalization, paternal effects, gene expression, Brassicaceae.

Research article

Introduction

Whole genome duplication (WGD), or polyploidy, has been an ongoing process during eukaryotic evolution, with polyploidy events having occurred during the evolution of fish, frogs, yeasts, and flowering plants, among other groups (Otto and Whitton 2000; Wolfe 2001; Seoighe 2003; Jaillon et al. 2009; Van de Peer et al. 2009). Almost all angiosperms show evidence for at least one round of WGD sometime during their evolutionary history, with many plants having had multiple polyploidy events occur during the evolution of their lineage (Cui et al. 2006; Soltis et al. 2009). In addition to polyploidy, duplicated genes can be formed by segmental duplications of multiple genes along one chromosome, tandem duplication of individual genes, and duplicative retroposition. All the types of gene duplication have contributed greatly to the large number of genes in many eukaryotic genomes. After formation by duplication, the functions of duplicated genes can diverge by the acquisition of new function, neofunctionalization (Ohno 1970), or partitioning of ancestral function, subfunctionalization (Hughes 1994; Force et al. 1999). Expression patterns of duplicated genes can diverge by changes in gene regulation, including gain of a new expression pattern relative to the ancestral state or partitioning of an ancestral expression pattern between the duplicates, also referred to as neofunctionalization and subfunctionalization, respectively (Force et al. 1999). Functional and expression divergence are widely regarded as important mechanisms for the retention of duplicated genes.

In the genome of *Arabidopsis thaliana*, there have been at least three rounds of ancient WGD events during the evolution of its lineage, termed α-, β-, and γ-WGD events. Among them, α is specific to the Brassicaceae family, and its timing is currently estimated at about 23 Ma (Barker et al. 2009), β is specific to the Brassicales order after the divergence of the Brassicaceae and Caricaceae from a common ancestor, whereas γ is an older WGD event that is presumably eudicot specific (Blanc et al. 2003; Bowers et al. 2003; Jaillon et al. 2007; Ming et al. 2008). Expression patterns of genes duplicated by WGD and by other smaller scale mechanisms have been examined in a range of organ types, developmental stages, and stress conditions from published microarray data sets. Several studies have shown that there

has been considerable divergence in expression patterns across different organs and treatments, with over half of the duplicated pairs examined in each case showing significant divergence in expression patterns between duplicates (Blanc and Wolfe 2004; Haberer et al. 2004; Casneuf et al. 2006; Duarte et al. 2006; Ganko et al. 2007; Ha et al. 2007; Zou et al. 2009). Expression divergence plus accelerated and asymmetric sequence evolution (i.e., a much faster rate of sequence evolution in one duplicate compared with the other) have been interpreted as evidence for functional divergence (Blanc and Wolfe 2004). However, there are relatively few cases of experimentally demonstrated gain of a new function of duplicated genes during the evolution of the Brassicaceae family. Examples of neofunctionalization that have been reported include the nitrilase genes *NIT1* and *NIT4*, where *NIT1* has a new function and accelerated rate of sequence evolution, but expression patterns are similar (Blanc and Wolfe 2004), and the mercaptopyruvate sulfurtransferases *AtMST1* and *AtMST2* that have a different subcellular localization, to the mitochondria or cytoplasm (Nakamura et al. 2000), but similar expression patterns. Another example is the gene pair MEDEA and *SWINGER* that are differentially imprinted in the endosperm (*MEDEA* is paternally imprinted and *SWINGER* is not imprinted), have largely overlapping, but not identical, expression patterns (Spillane et al. 2007), and different but partially redundant functions (Wang et al. 2006). A case of neofunctionalization after duplicative retroposition is *CYP98A8*/*CYP98A9* compared with *CYP98A3*. Neofunctionalization of *CYP98A8*/*CYP98A9* led to a novel phenolic pathway in pollen of *A. thaliana*, and the genes' expression patterns are mostly limited to flowers, in contrast to *CYP98A3* which is expressed in most organs but not pollen (Matsuno et al. 2009). Less well documented are cases of neofunctionalization that show gain of a new function, elimination of the old function, gain of expression in a new organ types, and loss of expression in other organ types by one of the duplicates, yet such cases likely involve some of the most dramatic changes in function after gene duplication.

In this study, we identified that the *SHORT SUSPENSOR* (*SSP*) gene (Bayer et al. 2009) and the *Brassinosteroid Kinase 1* (*BSK1*) gene (Tang et al. 2008) are paralogs derived by the α-WGD at the base of the Brassicaceae family. We present analyses of gene expression and sequence evolution indicating that *SSP* has undergone neofunctionalization from being involved in brassinosteroid signal transduction to regulating the timing of zygote elongation by a unique paternal effect mechanism involving transcription in sperm cells of the pollen and translation in the zygote. In addition, we analyzed a duplicated copy of *SSP*, *SSP-like1*, which also has undergone neofunctionalization.

Materials and Methods

Microarray Data Analysis

The *Arabidopsis* ATH1 microarray data from 63 different developmental stages and organ types (#ME00319) (Schmid et al. 2005) and the Soybean Genome Array data (#GSE12286) (Haerizadeh et al. 2009) were obtained from the Gene Expression Omnibus at *National Center for Biotechnology Information* (NCBI) (http://www.ncbi.nlm.nih.gov/). The 63 different developmental stages and organ types used in *Arabidopsis* ATH1 microarray data are listed in supplementary table S1 (Supplementary Material online). Raw CEL files were processed and normalized using the MAS5.0 algorithm in Bioconductor (http://www.bioconductor.org/). To determine the absence or presence of expression, the "mas5calls" function in Bioconductor was implemented in the statistical package R (http://www.r-project.org/). This statistical procedure performs the Wilcoxon signed rank-based gene expression presence/absence detection algorithm and yields a detection call (i.e., a detection *P* value) to access if the detected transcript is significantly greater than background signal noise. A gene with a *P* value less than 0.05 is marked as "presence," whereas a gene with a *P* value equal to or greater than 0.05 is marked as "absence." Because each gene contains three replicates, only those assigned as absence of expression in at least 2 of 3 replicates were given an absence call.

Phylogenetic Analysis of the *BSK* Gene Family

To identify the orthologous *BSK1*-like genes in outgroup species, we implemented a phylogenetic analysis of the *BSK* gene family with a focus on species with genome sequence data available, including 4 eudicot species (*A. thaliana*, *Carica papaya*, *Populus trichocarpa*, and *Vitis vinifera*), 2 monocot species (*Oryza sativa* and *Sorghum bicolor*), and 1 moss species (*Physcomitrella patens*). The sequences were retrieved from a Blast search with default settings in the Web site Plaza (http://bioinformatics.psb.ugent.be/plaza/) (Proost et al. 2009). In addition, we also included orthologous *BSK1*-like genes from *Brassica rapa* subsp. *pekinensis*, *Cleome spinosa*, *Gossypium hirsutum*, *Glycine max*, and *Helianthus annuus* in order to increase taxon sampling for the orthologous gene expression assay. Using *BSK1* in *A. thaliana* as query, the orthologous *BSK1*-like sequences from *Brassica*, *Cleome*, *Gossypium*, *Glycine*, and *Helianthus* were obtained from Blast searches (at least greater than 80% identity) of GenBank at NCBI (http://www.ncbi.nlm.nih.gov/). The *SSP* and *BSK1* genes from *A. lyrata* were obtained from Blast searches of Phytozome v4.0 at the Joint Genome Institute (http://www.phytozome.net/). Prior to phylogenetic analysis, sequences were aligned using TransAlign with the ClustalW program (Bininda-Emonds 2005) and manually checked using the program Bioedit (Hall 1999). A very divergent, short, sequence region at the 5′ end was removed and then the remainder of the gene sequence was used for further phylogenetic analysis (for details, see supplementary figs. S1 and S2, Supplementary Material online). Phylogenetic analysis was performed with a Bayesian method using MrBayes v3.1.2, described in supplementary fig. S4 (Supplementary Material online), and with a maximum likelihood (ML) method using Garli (Zwickl 2006). The nucleotide substitution model was automatically estimated from the empirical data. Statistical support for nodes was determined using bootstrapping with 100 ML replicates from Garli

and 50% majority rule of 250 best-fit Bayesian trees (standard deviation of split frequency <0.01) from MrBayes v3.1.2.

Plant Materials, Nucleic Acid Extraction, and RT-PCR

RNA was extracted from roots, stems, rosettes, leaves, flowers, ovules, or pollen from the following species: *Arabidopsis thaliana* (ecotype Columbia), *B. rapa* subsp. *pekinensis* (Chinese cabbage variety, MU525B, West Coast Seeds), *C. papaya* (cultivar Sun-Up), *H. annuus* (wild population from Utah), *G. hirsutum* (cultivar Maxxa), and *V. vinifera* (cultivar Pinot Noir). *Brassica rapa* plants were subjected to 4 °C for 2 months to stimulate bolting and flower production. Pollen from *Carica* and *Helianthus* was collected by tapping the flowers on a piece of paper and pollen from *Brassica* and *Vitis* was collected by using a vacuum cleaner method (Johnson-Brousseau and McCormick 2004). Collected pollen materials were examined for purity under light microscopy. Nucleic acid extraction and reverse transcription-polymerase chain reaction (RT-PCR) conditions followed those in Liu and Adams (2008). For RT-PCR, 25–35 reaction cycles were applied to assay gene expression level differences in different organ types. Gene-specific primers are listed in supplementary table S2 (Supplementary Material online). New sequences determined in this study were deposited in GenBank: *G. hirsutum* partial *BSK1.3* cds. and *C. papaya* partial *BSK11* cds. with accession numbers GU321198 and GU321199, respectively.

Selection Analysis

To test if there has been evidence of accelerated sequence evolution or positive selection acting on *SSP*, we used a phylogeny-based approach to examine if there has been any sequence rate acceleration or positive selection acting on *SSP* or *BSK1* using PAML (Yang 2007). Orthologous sequences from *P. trichocarpa* and *V. vinifera* were retrieved from CoGe (http://synteny.cnr.berkeley.edu/CoGe) (Lyons et al. 2008), and orthologous sequences from *C. papaya* were obtained from the Web site Plaza (http://bioinformatics.psb.ugent.be/plaza/) (Proost et al. 2009) based on collinear syntenic analyses. Pairwise ω (d_N/d_S)-ratio analysis, protein sequence rate acceleration, and positive selection were implemented using the program Codeml in PAML. For pairwise ω-ratio analysis, nonsynonymous (d_N) and synonymous (d_S) nucleotide substitution analysis was implemented using ML in Codeml. In the detection of sequence rate acceleration, we followed the analytical procedure described in Spillane et al. (2007). One ω-ratio model, two ω-ratio model, three ω-ratio model, and free ω-ratio model branch models were implemented. The first model assumes that only one ω-ratio leads to whole phylogeny branches; the second model assumes that one ω-ratio leads to pro-orthologs in the *BSK1* branch and another ω-ratio leads to the *SSP* branch; the third model assumes that three different ω-ratios lead to pro-ortholog branches, *BSK1* branch, and *SSP* branch, respectively; and the last model allows different ω-ratios for each branch of phylogeny. Then, twice the difference of their likelihood ratio between any two models (likelihood ratio test [LRT]) was compared against a chi-square distribution. The degree of freedom (df) was obtained based on the difference of parameters used in any two models. In the detection of positive selection, two branch-site models, model A test1 and model A test2, were implemented, and the LRT was conducted against a chi-square distribution with the 50:50 mixture of df = 0 and df = 1. Results from the branch-site model can allow us to evaluate which specific codons along sequence underwent positive selection ($\omega > 1$). In this study, we applied a branch-site model to detect positive selection on the *SSP*, *SSP*-like1, and *BSK1* genes. To correct multiple testing, a 5% false discovery rate control was used (Anisimova and Yang 2007).

Cis-Regulatory Element Analysis

Sequence up to 723 bases upstream of *SSP*, *SSP-like2*, and *BSK1* from *A. thaliana* were searched against the PlantCARE *cis*-element database (Lescot et al. 2002) to identify predicted *cis*-regulatory elements and compare among the three genes. To determine if there has been any insertion of transposable elements in the *cis*-element region, we performed Blast searches of CENSOR in the Repeat Masking of giri (http://www.girinst.org/censor/index.php) (Kohany et al. 2006).

Results

SSP and *BSK1* Are Whole Genome Duplicates with Completely Different Organ-Specific Expression Patterns

While studying genes duplicated by the most recent WGD during the evolutionary history of the Brassicaceae (Blanc et al. 2003; Bowers et al. 2003), also known as the α-WGD, we noticed that the *SSP* gene (locus At2g17090) and the *BSK1* gene (locus At4g35230) are paralogs. *SSP* is in a duplicated block on chromosome 2 and *BSK1* is in a duplicated block on chromosome 4, shown in figure 1 and supplementary figure S3 (Supplementary Material online). The two genes have very different functions: *SSP* regulates the timing of elongation of the embryo by activating the *YODA* signaling pathway, using a paternal control mechanism involving transcription in sperm cells of the pollen followed by translation only in the embryo (Bayer et al. 2009). In stark contrast, *BSK1* is part of the brassinosteroid signal transduction pathway (Tang et al. 2008). It is phosphorylated by the brassinosteroid receptor *BRI1* and it phosphorylates *BSU1* (Kim et al. 2009). Thus, there has been a change in function in one or both genes after gene duplication.

To determine how the expression patterns of *SSP* and *BSK1* have evolved since gene duplication, we compared the expression patterns of *BSK1* and *SSP* using Affymetrix ATH1 microarray data from 63 different organ types and developmental stages in *A. thaliana* (Schmid et al. 2005) and RT-PCR expression assays in *B. rapa*. To analyze the microarray data, we normalized raw CEL files using the MAS5.0 algorithm and determined the absence or presence of

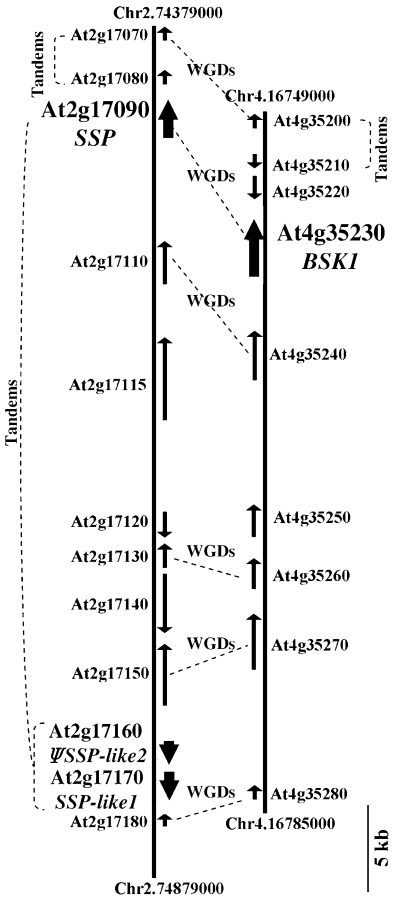

FIG. 1. Duplicated blocks on chromosomes 2 and 4 containing *BSK1*, *SSP*, *SSP-like1*, and *SSP-like2*. Scale bar indicates the nucleotide length. Abbreviations: Chr2, chromosome 2; Chr4, chromosom4; WGDs, genes duplicated by the most recent WGD; Tandems, genes formed by tandem duplication; Ψ, putative pseudogene. Information about *SSP-like1* and *SSP-like2* is presented in the last section of the Results.

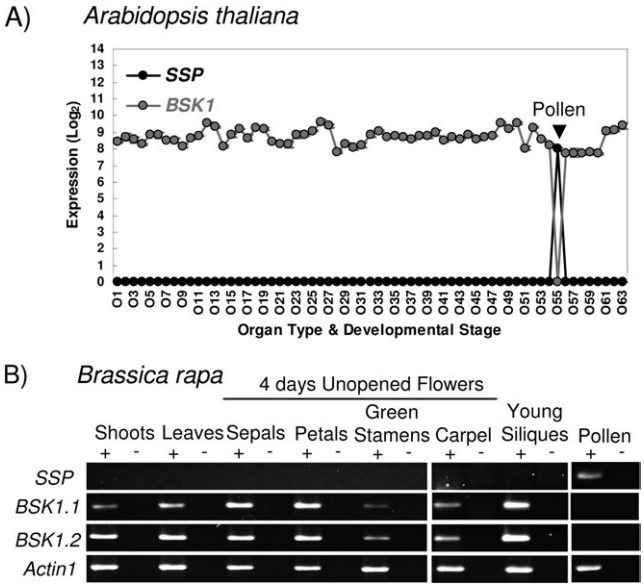

FIG. 2. *SSP* and *BSK1* show opposite organ-specific expression patterns. (*A*) Gene expression of *SSP* and *BSK1* in *Arabidopsis thaliana*. The MAS5-normalized microarray data were obtained from 63 different developmental stages and organ types. Absence and presence of expression above background was determined using mas5calls. Error bars show variance among three biological replicates. The different developmental stages and organ types are listed in supplementary table S1 (Supplementary Material online). (*B*) RT-PCR expression assays of *SSP*, *BSK1.1*, and *BSK1.2* in *Brassica rapa*. Plus signs indicate reactions containing reverse transcriptase and minus signs indicate reactions without reverse transcriptase.

expression by using a mas5calls function in Bioconductor (for details, see Materials and Methods). In *Arabidopsis*, *BSK1* was highly expressed in every organ type and developmental stage except for pollen where it is not expressed (fig. 2*A*). In complete contrast, *SSP* showed expression above background only in pollen (fig. 2*A*); the *SSP* results are consistent with data and analysis from Bayer et al. (2009). In *B. rapa*, one copy of *SSP* and two copies of *BSK1* were identified based on our phylogenetic analysis (fig. 3). Both copies of *BSK1* were highly expressed in every organ type that we examined except for pollen where neither is expressed, whereas *SSP* showed expression only in pollen (fig. 2*B*). Overall, the organ-specific expression patterns of *SSP* and *BSK1* are completely different and exactly opposite, and they are consistent between *Arabidopsis* and *Brassica*.

BSK1 Reflects the Ancestral Expression Pattern and Function

The dramatic difference in expression patterns between *SSP* and *BSK1* could be due to partitioning of the ancestral, preduplication, expression pattern between *BSK1* and *SSP*, if the ancestral state was expression in all organs, that would be an example of subfunctionalization of expression patterns. Alternatively, either *BSK1* or *SSP* could retain the ancestral expression pattern, with the other gene having undergone a complete change to gain a new expression pattern (i.e., neofunctionalization). To distinguish among these possibilities, we assayed expression of orthologous genes from outgroup species that diverged before the Brassicaceae-specific WGD. To identify the orthologs, we first reconstructed the phylogenetic relationships of the *BSK* gene family among the sequenced genomes from *A. thaliana*, *C. papaya* (papaya), *P. trichocarpa* (poplar), *V. vinifera* (grape), *O. sativa* (rice), and *S. bicolor* (sorghum) by using *P. patens* (a moss) as an outgroup (fig. 3). Sequences from additional eudicots were included for putative orthologs of *BSK1* and *BSK11*. The *BSK* gene phylogeny showed that there have been several rounds of gene duplication events during *BSK* gene family evolution in angiosperms. There are two major clades of genes related to *BSK1* that formed after the divergence of monocots and eudicots: the *BSK1* group and the *BSK11* group (fig. 3). The *BSK11* group is well supported as a clade to the exclusion of the *BSK1* sequences, but some relationships within the *BSK1* group are not well resolved. Nevertheless, within the *BSK1* clade, we identified orthologous genes from *C. spinosa*, *C. papaya*, *G. hirsutum*, *G. max*, *P. trichocarpa*, *V. vinifera*, and *H. annuus* (fig. 3).

After identifying *BSK1*-orthologous genes from outgroup species, we then assayed their expression pattern in

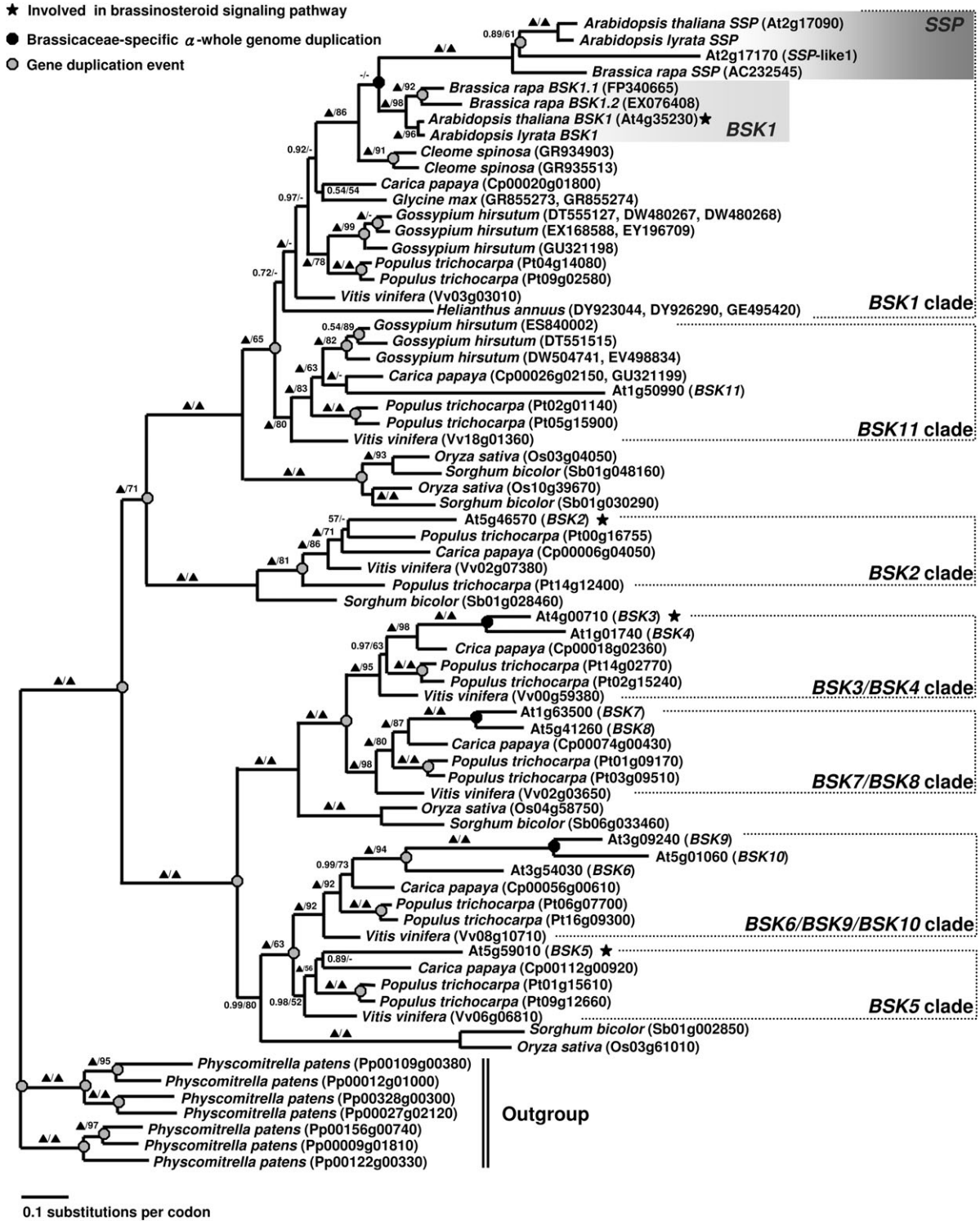

Fig. 3. Phylogenetic tree of the *BSK* gene family. Topology is inferred from a Bayesian method. Statistical support is indicated on branches. The first value is posterior probability from Bayesian analysis (▲ = 1) and the second value is 100 bootstrapping replicates using ML analysis (▲ = 100). Dash indicates either posterior probability <0.5 or bootstrap value <50. Stars indicate genes that have been shown to function in brassinosteroid signal transduction (Tang et al. 2008). Dark gray shading indicates *SSP* genes and *SSP-like1* gene, whereas light gray shading indicates *BSK1* genes.

pollen and in multiple organ types in *C. papaya* (papaya), *G. hirsutum* (cotton), *V. vinifera* (grape), and *H. annuus* (sunflower) by using RT-PCR. All the genes showed expression in various organ types but no expression in pollen (fig. 4A). In addition, we analyzed Affymetrix microarray data from *G. max* (soybean) (Haerizadeh et al. 2009). Expression of the *BSK1*-like gene in soybean was below background in pollen but at relatively high levels in all other

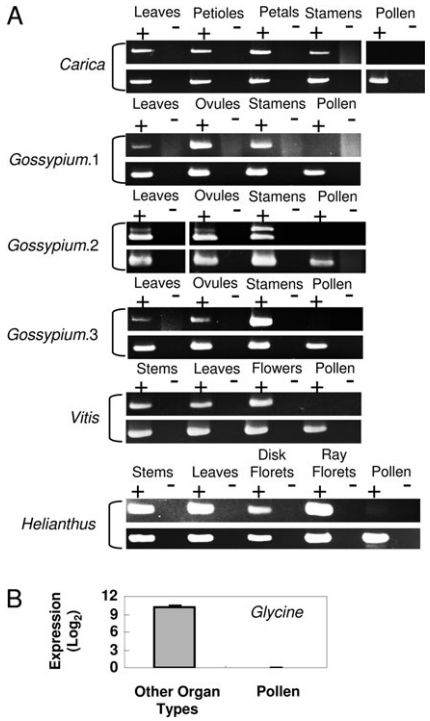

Fig. 4. Gene expression analyses of *BSK1* orthologs from outgroup species. (*A*) RT-PCR expression assays of *BSK1* orthologs in the upper panels of each pair, and actin genes in the lower panel of each pair were used as a control for DNA template concentration. See figure 3 and supplementary table S2 (Supplementary Material online) for gene accession numbers. Plus signs indicate reactions containing reverse transcriptase (RT) and minus signs indicate reactions without RT. Two bands are present in *Gossypium hirsutum BSK1.2* because of an intron retention alternatively spliced variant. (*B*) Microarray expression analysis of the *BSK1* ortholog in *Glycine max*. Error bars show variance among different biological replicates (two replicates in other organ types and three replicates in pollen).

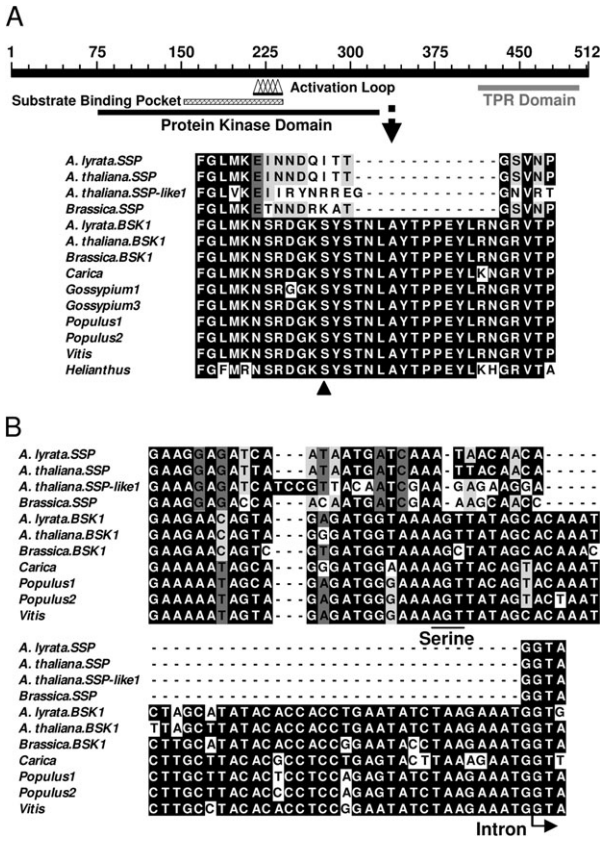

Fig. 5. Sequence analysis of functional domains in *SSP*, *SSP-like1*, and *BSK1*. Conserved functional domains were identified by using the Conserved Domain Database (CDD) with an interactive domain family analysis (Marchler-Bauer et al. 2007). (*A*) Diagram from the CDD showing two functional domains: the protein kinase domain and the TPR domain. The effects are illustrated of nucleotide substitutions and deletions in exon 4 on the activation loop of the protein kinase domain in *SSP* and *SSP-like1*. The location of serine-230, that is essential for *BSK1* function, is marked with a triangle. (*B*) Nucleotide alignment of exon 4 showing the locations of the nucleotide substitutions and deletions. The positions corresponding to serine-230 are indicated with a line and "serine." The start of intron 4 is indicated with an arrow. Shading: black, >45% shared identity; dark gray, 35–45% shared identity; light gray, 25–35% shared identity; and white, <25% shared identity.

organ types (fig. 4*B*). Thus, the *BSK1*/*SSP* orthologs in papaya, cotton, grape, soybean, and sunflower all show similar expression patterns to *BSK1* not to *SSP*. These observations indicate that the preduplication expression pattern of *SSP* and *BSK1* is no expression in pollen but expression in other organs. To infer the preduplication expression pattern of *SSP* and *BSK1* using another approach, we examined their most recent common ancestral (MRCA) expression state by implementing a ML method using the phylogeny of the *BSK* gene family in *A. thaliana* to reconstruct the MRCA expression state (supplementary fig. S4, Supplementary Material online). The results were consistent with the outgroup expression analysis.

The above results indicate that *BSK1* shows the ancestral expression state and potentially the ancestral function. Further support for the ancestral function comes from examining other members of the *BSK* gene family in a phylogenetic context. The *BSK* genes *BSK2*, *BSK3*, and *BSK5* have been functionally characterized as being involved in brassinosteroid signal transduction (Tang et al. 2008). All of those genes branch as an outgroup to the clade containing *BSK1* and *SSP* (fig. 3), strongly suggesting that an ancestral function of the *BSK* genes was involvement in brassinosteroid signal transduction, not in regulating embryo elongation and division.

Loss of the Original Function of *SSP* by Mutations in the Kinase Activation Domain

SSP and *BSK1* have two major functional domains: a protein kinase domain and a tetratricopeptide repeat (TPR) domain (fig. 5). The protein kinase domain is responsible for the catalytic activity, and the TPR domain is involved in protein–protein interactions. Mutations in these two regions have been shown to be detrimental for the function of *SSP* (Bayer et al. 2009). When comparing with other orthologs, there is a deletion in the activation loop of the protein kinase domain in *SSP*, which resides in the substrate-binding pocket (fig. 5). In addition, *SSP* has a one nucleotide deletion in a codon corresponding to a serine

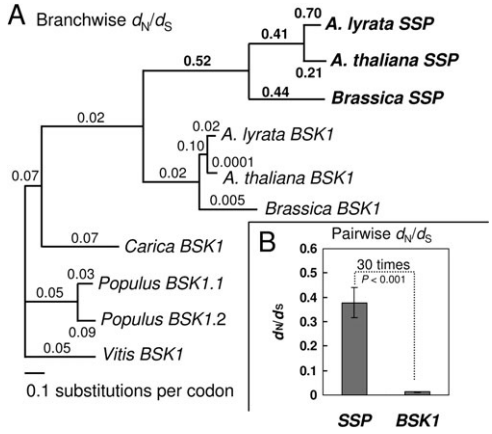

Fig. 6. Sequence rate and selection analysis of *SSP* and *BSK1*. (*A*) Phylogenetic tree of *SSP* genes from *Arabidopsis thaliana*, *A. lyrata*, and *Brassica rapa*, and *BSK1* genes from *A. lyrata*, *A. thaliana*, and *B. rapa*, *Carica papaya*, *Populus trichocarpa*, and *Vitis vinifera*. Tree branch length and branchwise d_N/d_S ratios above the branches were estimated with codeml using a free-ratio model. A higher evolutionary rate is found in *SSP*. The tree is unrooted. (*B*) Pairwise d_N/d_S ratios, d_N, and d_S analysis among *A. lyrata*, *A. thaliana*, *and B. rapa* showing that *SSP* has a d_N/d_S ratio 30 times higher than *BSK1*.

in *BSK1* that is critical for its function (fig. 5). Serine-230 in *BSK1* is the major site of phosphorylation by the brassinosteroid receptor kinase *BRI1* and mutations result in an 82% reduction in phosphorylation (Tang et al. 2008). *BRI1* phosphorylation of *BSK1* at Ser-230 promotes *BSK1* binding to the *BSU1* phosphatase and continuation of the signal transduction cascade (Kim et al. 2009). Mutation of the serine abolished binding of *BSK1* to *BSU1* (Kim et al. 2009). Those analyses indicate that deletion mutations in *SSP* have resulted in loss of the original *BSK1* function in brassinosteroid signal transduction. The complete change in expression pattern of *SSP* also probably contributed to loss of the original function because *SSP* is not expressed in most of the organs where the brassinosteroid receptor *BRI1* that starts the signal transduction cascade is expressed (*BRI1* does not appear to be expressed in pollen, from microarray data determined by Schmid et al. [2005]).

Rapid Amino Acid Sequence Evolution in *SSP*

Considering that *SSP* has changed in function after its formation by gene duplication, we then asked if there is any acceleration in sequence evolution or positive selection acting on the protein sequence of *SSP*. To answer this question, we implemented sequence rate analysis by using PAML. The *BSK1*-like genes from *Carica*, *Populus*, and *Vitis* were regarded as pro-orthologs based on their chromosomal syntenic collinear relationship with *SSP* and *BSK1* duplication block (Lyons et al. 2008; Proost et al. 2009). The ratio ($\omega = d_N/d_S$) of nonsynonymous (d_N) to synonymous (d_S) nucleotide substitution for the lineage ancestral to *SSP* is significantly higher than the one to *BSK1* (fig. 6*A*). A two ω-ratio model that allows one ω-ratio leading to pro-orthologs and *BSK1* branch (reflecting functional constraint) and another ω-ratio leading to the *SSP* branch (allowing functional diversification) fits better than the one ω-ratio model that only applies one ω-ratio for the whole phylogeny based on a LRT ($P < 0.0001$; supplementary table S3, Supplementary Material online). In addition, the three ω-ratio model that assumes three different ω-ratios leading to the pro-ortholog branch, the *BSK1* branch, and the *SSP* branch, respectively, fits better than two different ω-ratios, suggesting that the *BSK1* branch evolves much slower than the pro-ortholog branch and the *SSP* branch (supplementary table S3, Supplementary Material online). However, a free ω-ratio model marginally fits better than the three ω-ratio model, suggesting that the majority of changes in the protein sequence occurred after divergence of *SSP* and *BSK1* ($P < 0.05$; supplementary table S3, Supplementary Material online).

We found that the sequence rate acceleration in *SSP* is still an ongoing process. From pairwise d_N/d_S ratio comparisons among two *Arabidopsis* species and from *B. rapa*, *SSP* showed a 30 times significantly higher d_N/d_S ratio than *BSK1* (fig. 6*B*), suggesting that the sequence rate acceleration was an ongoing process after the speciation between the *Arabidopsis* and *Brassica* lineages and also between *A. thaliana* and *A. lyrata*. The higher d_N/d_S ratio in *SSP* is largely caused by the elevation of nonsynonymous nucleotide substitutions (supplementary table S4, Supplementary Material online). We next performed a branch-site model test to examine if there has been any positive selection acting on specific sites in *SSP*, but we did not detect any evidence of positive selection acting on *SSP* (supplementary table S5, Supplementary Material online). Both the free-ratio model and the branch-site model failed to detect evidence of positive selection, suggesting that the accelerated sequence evolution of *SSP* is due to the relaxation of purifying selection after gene duplication (Kondrashov et al. 2002; Jordan et al. 2004).

Neofunctionalization of *SSP-Like1* After Duplication of *SSP*

In addition to the WGD event that created *SSP*, we also observed two other duplications that created two loci closely related to *SSP*: At2g17170—here referred to as *SSP-like1*, and At2g17160—here referred to as *SSP-like2*. The two genes are close to *SSP* on chromosome 2 and they are tandem duplicates (fig. 1). Phylogenetic analysis shows that *SSP-like1* branches with *SSP* (fig. 3) and *SSP-like2* branches with *SSP-like 1* (data not shown). The finding that *SSP* and *SSP-like1* are paralogs is consistent with the results of a large-scale analysis of receptor-like kinase genes in *A. thaliana* and *O. sativa* where At2g17090 and At2g17170 branched together (Shiu et al. 2004). Evidence from their chromosome locations and our phylogenetic trees indicate that *SSP-like1*, *SSP-like2*, and *SSP* are derived by two tandem duplication events: *SSP* duplicated in tandem to create the ancestral *SSP-like1*/*SSP-like2* sequence, that gene was transposed a few genes downstream on the chromosome (fig. 1), and then there was either a partial duplication to create *SSP-like2* or a complete duplication followed by a deletion at the 5′ end of *SSP-like2* (fig. 7*A*).

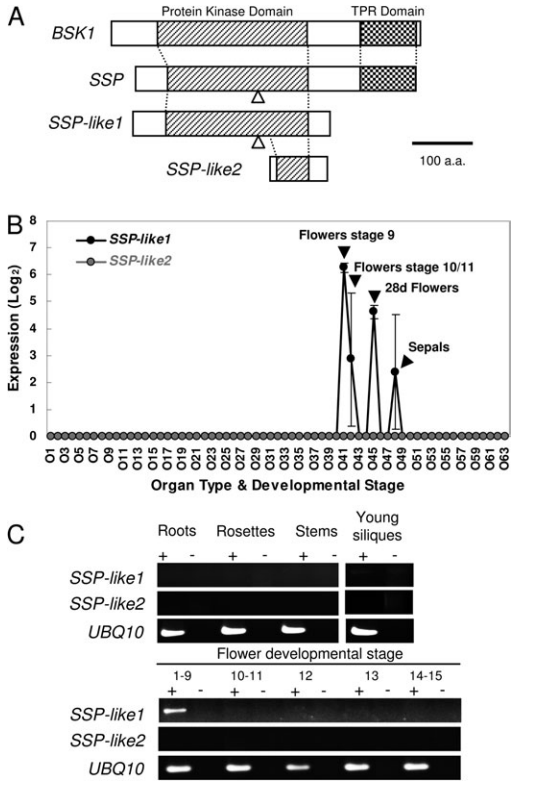

Fig. 7. Gene structure and expression of *SSP-like1* and *SSP-like2*. (*A*) Gene structure of *SSP-like1* and *SSP-like2* in comparison with *BSK1* and *SSP*. Arrowheads indicate the deletion in the protein kinase domain. (*B*) The MAS5-normalized microarray data from 63 different developmental stages and organ types. Expression values were background corrected. Error bars indicate variance among replicates. The developmental stages and organ types in the microarray data are listed in supplementary table S1 (Supplementary Material online). (*C*) Expression assay by RT-PCR verifying that *SSP-like1* is expressed in the early stage of unopened flower but not in other organs examined, and *SSP-like2* is not expressed across different organ types. *UBQ10* was used as a control for cDNA template concentration.

The *SSP-like1* gene compared with *SSP* is missing two exons and part of a third exon at the 3′ end of the gene including the region corresponding to the TPR protein-binding domain (fig. 7*A*). The TPR domain is essential for the function of *SSP* (Bayer et al. 2009), and thus, its absence is highly suggestive that *SSP-like1* does not have the same function as *SSP*. The function of *SSP-like1* is currently unknown. *SSP-like2* only contains about one-fourth of the protein kinase domain, in addition to lacking the TPR domain, and it likely is a pseudogene fragment (fig. 7*A*).

Analysis of microarray data from 63 different organ types and developmental stages of *A. thaliana* (Schmid et al. 2005) indicated that *SSP-like1* shows expression above background only in unopened flowers (stage 9–11), 28-day whole flowers, and sepals (stage 15), whereas *SSP-like2* shows no expression above background across any of the developmental stages and organ types (fig. 7*B*). We verified some of the microarray results using RT-PCR with six different organ types. In *SSP-like1*, expression was seen in unopened flowers of stage 9 and earlier among the organs types examined (fig. 7*C*). In contrast to *SSP-like1*, no expression of *SSP-like2* was observed (fig. 7*C*), further suggesting that *SSP-like2* is a pseudogene fragment. The expression pattern of *SSP-like1* contrasts greatly to *SSP*, and thus, the expression pattern of *SSP-like1* has considerably changed after gene duplication.

To test if there has been adaptive evolution acting on *SSP-like1*, we performed sequence rate and positive selection analysis using PAML. Compared with *SSP* and *BSK1*, *SSP-like1* has experienced rate acceleration after its formation, especially at nonsynonymous sites (supplementary figs. S5 and S6, Supplementary Material online). The rate acceleration is comparable in scale with the sequence evolution of *SSP* (supplementary fig. S5, Supplementary Material online). Although the free ratio did not show any evidence of positive selection (i.e., $d_N/d_S > 1$) on the branch leading to *SSP-like1* (supplementary fig. S5, Supplementary Material online), the branch-site model suggests that *SSP-like1* shows evidence for positive selection at many sites (supplementary fig. S6; supplementary table S6, Supplementary Material online), suggesting that *SSP-like1* might have undergone adaptive evolution after gene duplication.

Discussion

Neofunctionalization of *SSP* and *SSP-Like1* by Complete Changes in Expression Pattern, Function, Amino Acid Changes, and Deletions

SSP shows several hallmarks of a gene that has undergone neofunctionalization after duplication, and it is uncommon to find all of these features in a single neofunctionalized gene: 1) The function of *SSP* has changed from being a component of the brassinosteroid signal transduction pathway to regulating elongation of the embryo by an intriguing paternal effect mechanism. The dramatic functional change is surprising, although neofunctionalization of a duplicated gene sometimes produces a paralog with a very different function. 2) Functional divergence was caused in part by deletions in the kinase activation domain that abolished kinase activity and binding of the *SSP* predecessor to its interaction partner *BSU1*. Not only has *SSP* gained a new function but the gene also has lost its original function, unlike some duplicated and neofunctionalized genes that are still partially redundant. 3) An accelerated rate of amino acid changes in *SSP*, relative to *BSK1*, also probably was involved in functional divergence of *SSP*. 4) The organ-specific expression pattern of *SSP* has changed and it is completely opposite from its duplicated partner *BSK1* in pollen compared with 62 other organs and developmental stages. Such a drastic change in expression pattern has been found rarely, if at all, in duplicated genes, although expression data sets of comparable sizes are available only in a very small number of multicellular eukaryotes.

We hypothesize that the expression pattern of *SSP* changed before the functional change, from expression in all organs except pollen to expression only in pollen. That might have occurred by gain of expression in pollen followed by loss of expression in all other organs, or perhaps

expression in all organs (except pollen) was lost followed by gain of expression in pollen before the gene could suffer a pseudogenization mutation. If instead the function of *SSP* changed first to its current function before the expression pattern changed, it would hyperactivate the *YODA* pathway in green tissues and result in developmental defects, including lack of stomata, as inferred by data from seedlings expressing *SSP* from a strong, broadly active promoter (Bayer et al. 2009). Alternatively, the deletions in the kinase catalytic domain of *SSP* that abolished the original function could have occurred first, followed by changes in expression pattern and gain of the new function. After becoming expressed in pollen *SSP* was free to evolve rapidly in amino acid sequence.

Why would paternal control of zygote elongation and division evolve from a duplicated gene whose original function immediately upon WGD was involvement in brassinosteroid signal transduction? One possibility is that both the *SSP* and *BSK1* proteins are plasma membrane bound and contain a TPR domain that is important for mediating protein–protein interactions (Tang et al. 2008; Bayer et al. 2009). It is hypothesized that *SSP* may exert its function in regulating the timing of zygote elongation by recruiting an unidentified pathway activator and thus the importance for protein–protein interactions (Bayer et al. 2009). Alternatively, *SSP* may have evolved into a regulator of zygote elongation by chance co-option of a duplicated gene that had undergone accelerated amino acid sequence evolution and deletions.

In addition to neofunctionalization of *SSP*, we found evidence for neofunctionalization of *SSP-like1* after forming by duplication of *SSP*. *SSP-like1* has a different organ-specific expression pattern from *SSP*, most notably that it does not appear to be expressed in mature pollen where *SSP* is exclusively expressed (except for the *SSP* transcripts provided by the sperm to the zygote). Thus, *SSP-like1* functions in different organ types from *SSP*, and it probably has a different function. *SSP-like1* has a greatly accelerated rate of amino acid substitutions, even more so than *SSP*, and it shows evidence for positive selection at specific sites. However, it is not known if any of those sites are amino acids critical for function. *SSP-like1* has lost the TPR domain at the C-terminus. The TPR domain is essential for the function of *SSP* (Bayer et al. 2009) and its loss in *SSP-like1*, in combination with the accelerated amino sequence evolution and positive selection, further indicate that *SSP* and *SSP-like1* have diverged in function. *SSP-like1* probably has lost its original function and gained a new function as has *SSP*. The sequence of events that created *SSP-like1* and *SSP* by gene duplication and the events involved in neofunctionalization are summarized in figure 8.

We hypothesize that *SSP-like1* was created by duplication of *SSP* after the divergence of the *Arabidopsis* and *Brassica* lineages from a common ancestor, based on phylogenetic evidence. Our phylogenetic analysis shows that *SSP-like1* in *A. thaliana* branches with *SSP* in *A. thaliana* and *A. lyrata* instead of basal to *SSP* in *Brassica* (fig. 3). However, the statistical support level for the branch separating *SSP* in *Brassica* from *SSP* and *SSP-like1* in *Arabidopsis* is relatively low, so our inference of the timing of the duplication should be regarded as tentative. Once the *B. rapa* genome is mostly or fully sequenced, it should be possible to determine if there is or is not a homolog of *SSP-like1* in *Brassica*.

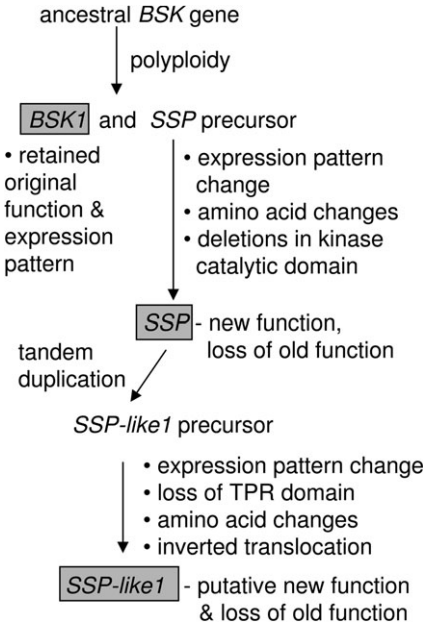

Fig. 8. Model for duplication and neofunctionalization of *SSP* and *SSP-like1*. See text for explanation of the steps.

The very different organ-specific expression patterns of *SSP*, *BSK1*, and *SSP-like1* suggest that changes have occurred in *cis*-regulatory elements of *SSP* and *SSP-like1*. We used the *cis*-regulatory element prediction program PlantCARE (Lescot et al. 2002) to predict and compare *cis*-regulatory elements among the three genes. Numerous putative *cis*-regulatory elements were detected, with eight being unique to *SSP* and five being unique to *SSP-like1* (supplementary fig. S7, Supplementary Material online). The unique *cis*-regulatory elements might contribute to their organ-specific expression organs. However, it is difficult to say how many of the predicted *cis*-regulatory elements are actually acting as regulatory elements and which ones are spurious matches to potential *cis*-regulatory elements. Further analysis with experimental constructs would be necessary to determine which regulatory elements have changed to give *SSP* and *SSP-like1* their unique expression patterns. In addition, we found a 769-bp helitron 1,660-bp upstream of the start codon of *SSP* in *A. thaliana*, but the helitron was not present in *SSP* in *B. rapa*. Considering that both *Brassica* and *Arabidopsis* show expression of *SSP* in pollen, the helitron does not appear to be involved in the pollen-specific expression of *SSP* in *A. thaliana*.

Neofunctionalization After Gene Duplication in Plants

SSP and *SSP-like1* add to the small number of cases of gain of a new function after gene duplication and loss of the old function during the evolution of a plant family. Studies of

neofunctionalization of genes duplicated by WGD, as well as other types of gene duplication, have revealed several types of neofunctionalization involving regulation and/or sequence and structural changes. Changes in protein function can occur by mutations in the amino acid sequence or by structural changes in the sequence including deletions and insertions, especially in functional domains. Some genes show either amino acid changes or structural changes, whereas other genes like *SSP* and *SSP-like1* show both. Duplicate genes that evolve new functions can either loose their old function, like *SSP*, or retain the old function with neofunctionalization having the effect of diversifying the gene's function. Gain of a new function by a duplicated gene can be accompanied by changes in expression patterns, as with *SSP* and *SSP-like1*, or instead expression patterns can remain largely the same. Likewise, regulatory neofunctionalization (new expression patterns) can occur with or without changes in the function of the protein coded by the gene. Regulatory neofunctionalization has been proposed to act in either a qualitative manner, with gain of a completely new expression pattern after duplication, or a quantitative manner, with changes in the expression level of one copy after gene duplication (Force et al. 1999; Duarte et al. 2006).

Polyploidy events provide a large number of new genes that could potentially undergo neofunctionalization. Neofunctionalization might occur relatively soon (within about 1 My) after polyploidy in plants that are still cytologically polyploids, or it may be a process that mostly happens several million years later, during or after cytological diploidization. The only currently known cases of neofunctionalization in an evolutionarily recent plant polyploid, to our knowledge, are 15 genes in *G. hirsutum* (tetraploid cotton) that show regulatory neofunctionalization (Chaudhary et al. 2009). In contrast, there are numerous potential cases of regulatory neofunctionalization (neofunctionalization of expression patterns) after the α-WGD event in the Brassicaceae, based on a combination of expression divergence and asymmetric sequence evolution between the duplicates (Blanc and Wolfe 2004; Ganko et al. 2007) or expression divergence and ancestral state inference (Duarte et al. 2006; Zou et al. 2009). However, changes in function have not been studied or shown for most of those cases of regulatory neofunctionalization. Neofunctionalization of expression patterns can be detected much more readily than the evolution of new functions because the latter requires experimentally determined functional information.

An alternative fate of duplicated genes is escape from adaptive conflict (EAC) and sometimes it may be mistaken for neofunctionalization (discussed in Des Marais and Rausher 2008). In the EAC model, a single (preduplication) gene undergoes selection to perform a new function in addition to its original function. However, the gene is constrained from improving either function because of detrimental effects on the other function. After duplication, one copy improves one function and the other copy improves the other function. In the case of *SSP* and *BSK1*, the ancestral function was the current function of *BSK1* and not the current function of *SSP*; that is inconsistent with EAC. In addition, the EAC model predicts that both duplicates will undergo adaptive change instead of showing purifying selection. *BSK1* is undergoing purifying selection, whereas *SSP* exhibits relaxation of selection but it does not show evidence for positive selection. We conclude that *SSP* and *BSK1* have not undergone EAC, and instead, *SSP* has experienced neofunctionalization.

Recent Evolutionary Origin of Paternal Control of Embryonic Patterning

In addition to being a dramatic example of neofunctionalization, our study of *SSP* also provides insights into the timing of the evolution of the gene's intriguing and novel paternal effect mechanism for control of zygote elongation after fertilization. *SSP* regulates the *YODA* pathway that activates elongation and asymmetric division of the zygote after fertilization to create the embryo precursor and the elongated suspensor cell, using a paternal control mechanism of translation of *SSP* transcripts that were provided by the sperm cells in the pollen instead of there being maternal *SSP* expression (Bayer et al. 2009). Our study shows that *SSP* originated about 23 Ma at the base of the Brassicaceae family, and thus the *SSP*-mediated paternal control of embryonic patterning is restricted to the Brassicaceae. Thus, other angiosperms must use a different mechanism to regulate elongation of the zygote after fertilization. One possibility would be transcripts from another gene provided by the pollen, using a similar paternal effect mechanism as does *SSP*. Another possibility would be expression of a gene in the zygote only after fertilization that regulates the *YODA* pathway. The mechanism involving *SSP* has replaced the ancestral mechanism, as the zygotes in *SSP* mutants do not undergo normal elongation (Bayer et al. 2009).

Yet a different mechanism for regulating the timing of zygote elongation may occur in apomictic plants in the genus *Boechera*, within the lineage encompassed by the α-WGD (Bailey et al. 2006), and that undergo embryogenesis without fertilization by pollen. Apomictic *Boechera* plants would lack paternally supplied *SSP* transcripts, and thus *Boechera* probably has an alternative genetic basis for controlling zygote elongation. One possibility would be expression of the maternal allele of *SSP* in the zygote at the proper time for elongation.

In addition to *SSP*, uniparental expression of genes involved in embryo and endosperm development also includes imprinted genes where only one allele is expressed and the other allele is epigenetically silenced in a parent-of-origin specific manner. Two of the paternally imprinted genes, *MEDEA* and *FWA*, arose from the α-WGD at the base of the Brassicaceae from a nonimprinted ancestral gene, and each gene has a nonimprinted paralog (Nakamura et al. 2006; Spillane et al. 2007). Interestingly, *SSP*, *MEDEA*, and *FWA* all originated from the same WGD by neofunctionalization and gain of uniparental expression, albeit with uniparental expression being accomplished with different mechanisms. Thus, the creation of imprinting and other parent-of-origin expression effects during seed development are ongoing evolutionary processes in plants.

Supplementary Material

Supplementary figures S1–S7 and tables S1–S6 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Huei-Jiun Su and Liang-Chi Wang for collecting papaya pollen and assaying gene expression and Jonathan Wendel's laboratory for providing pollen cDNA from *Gossypium*. Also thanks to Michael Barker and Quentin Cronk for helpful discussions and people in K. Adams' laboratory for comments on the manuscript. This study was funded by a grant from the Natural Science and Engineering Research Council of Canada. S.-L.L. was supported in part by funding (SAS972161CA02) from the Ministry of Education, Taiwan.

References

Anisimova M, Yang Z. 2007. Multiple hypothesis testing to detect lineages under positive selection that affects only a few sites. *Mol Biol Evol.* 24:1219–1228.

Bailey CD, Koch MA, Mayer M, Mummenhoff K, O'Kane SL Jr, Warwick SI, Windham MD, Al-Shehbaz IA. 2006. Toward a global phylogeny of the Brassicaceae. *Mol Biol Evol.* 23:2142–2160.

Barker MS, Vogel H, Schranz ME. 2009. Paleopolyploidy in the Brassicales: analyses of the *Cleome* transcriptome elucidate the history of genome duplications in *Arabidopsis* and other Brassicales. *Genome Biol Evol.* 2009:391–399.

Bayer M, Nawy T, Giglione C, Galli M, Meinnel T, Lukowitz W. 2009. Paternal control of embryonic patterning in *Arabidopsis thaliana*. *Science* 323:1485–1488.

Bininda-Emonds OR. 2005. transAlign: using amino acids to facilitate the multiple alignment of protein-coding DNA sequences. *BMC Bioinformatics.* 6:156.

Blanc G, Hokamp K, Wolfe KH. 2003. A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. *Genome Res.* 13:137–144.

Blanc G, Wolfe KH. 2004. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell.* 16:1679–1691.

Bowers JE, Chapman BA, Rong J, Paterson AH. 2003. Unraveling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422:433–438.

Casneuf T, De Bodt S, Raes J, Maere S, Van de Peer Y. 2006. Nonrandom divergence of gene expression following gene and genome duplications in the flowering plant *Arabidopsis thaliana*. *Genome Biol.* 7:R13.

Chaudhary B, Flagel L, Stupar RM, Udall JA, Verma N, Springer NM, Wendel JF. 2009. Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (*Gossypium*). *Genetics* 182:503–517.

Cui L, Wall PK, Leebens-Mack JH, et al. (13 co-authors). 2006. Widespread genome duplications throughout the history of flowering plants. *Genome Res.* 16:738–749.

Des Marais DL, Rausher MD. 2008. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* 454:762–765.

Duarte JM, Cui L, Wall PK, Zhang Q, Zhang X, Leebens-Mack J, Ma H, Altman N, dePamphilis CW. 2006. Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of *Arabidopsis*. *Mol Biol Evol.* 23:469–478.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531–1545.

Ganko EW, Meyers BC, Vision TJ. 2007. Divergence in expression between duplicated genes in *Arabidopsis*. *Mol Biol Evol.* 24:2298–2309.

Ha M, Li W-H, Chen ZJ. 2007. External factors accelerate expression divergence between duplicate genes. *Trends Genet.* 23:162–166.

Haberer G, Hindemitt T, Meyers BC, Mayer KF. 2004. Transcriptional similarities, dissimilarities, and conservation of cis-elements in duplicated genes of *Arabidopsis*. *Plant Physiol.* 136:3009–3022.

Haerizadeh F, Wong CE, Bhalla PL, Gresshoff PM, Singh MB. 2009. Genomic expression profiling of mature soybean (*Glycine max*) pollen. *BMC Plant Biol.* 9:25.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser.* 41:95–98.

Hughes AL. 1994. The evolution of functionally novel proteins after gene duplication. *Proc Biol Sci.* 256:119–124.

Jaillon O, Aury JM, Noel B, et al. (56 co-authors). 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449:463–467.

Jaillon O, Aury JM, Wincker P. 2009. "Changing by doubling", the impact of whole genome duplications in the evolution of eukaryotes. *C R Biol.* 332:241–253.

Johnson-Brousseau SA, McCormick S. 2004. A compendium of methods useful for characterizing *Arabidopsis* pollen mutants and gametophytically-expressed genes. *Plant J.* 39:761–775.

Jordan IK, Wolf YI, Koonin EV. 2004. Duplicated genes evolve slower than singletons despite the initial rate increase. *BMC Evol Biol.* 4:22.

Kim TW, Guan S, Sun Y, Deng Z, Tang W, Shang JX, Sun Y, Burlingame AL, Wang ZY. 2009. Brassinosteroid signal transduction from cell-surface receptor kinases to nuclear transcription factors. *Nat Cell Biol.* 11:1254–1260.

Kohany O, Gentles AJ, Hankus L, Jurka J. 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics.* 7:474.

Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV. 2002. Selection in the evolution of gene duplications. *Genome Biol.* 3:research0008.

Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P, Rombauts S. 2002. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 30:325–327.

Liu S-L, Adams KL. 2008. Molecular adaptation and expression evolution following duplication of genes for organellar ribosomal protein S13 in rosids. *BMC Evol Biol.* 8:25.

Lyons E, Pedersen B, Kane J, et al. (11 co-authors). 2008. Finding and comparing syntenic regions among *Arabidopsis* and the outgroups papaya, poplar, and grape: CoGe with rosids. *Plant Physiol.* 148:1772–1781.

Marchler-Bauer A, Anderson JB, Derbyshire MK, et al. (28 co-authors). 2007. CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res.* 35:D237–D240.

Matsuno M, Compagnon V, Schoch GA, et al. 2009. Evolution of a novel phenolic pathway for pollen development. *Science* 325:1688–1692.

Ming R, Hou S, Feng Y, et al. (85 co-authors). 2008. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* 452:991–996.

Nakamura M, Katsumata H, Abe M, Yabe N, Komeda Y, Yamamoto KT, Takahashi T. 2006. Characterization of the class IV homeodomain-Leucine Zipper gene family in *Arabidopsis*. *Plant Physiol.* 141:1363–1375.

Nakamura T, Yamaguchi Y, Sano H. 2000. Plant mercaptopyruvate sulfurtransferases: molecular cloning, subcellular localization and enzymatic activities. *Eur J Biochem.* 267:5621–5630.

Ohno S. 1970. Evolution by gene duplication. New York: Springer-Verlag.

Otto SP, Whitton J. 2000. Polyploid incidence and evolution. *Annu Rev Genet.* 34:401–437.

Proost S, Van Bel M, Sterck L, Billiau K, Van Parys T, Van de Peer Y, Vandepoele K. 2009. PLAZA: a comparative genomics resource to study gene and genome evolution in plants. *Plant Cell.* 21:3718–3731.

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU. 2005. A gene expression map of *Arabidopsis* thaliana development. *Nat Genet.* 37:501–506.

Seoighe C. 2003. Turning the clock back on ancient genome duplication. *Curr Opin Genet Dev.* 13:636–643.

Shiu S-H, Karlowski WM, Pan R, Tzeng YH, Mayer KF, Li W-H. 2004. Comparative analysis of the receptor-like kinase family in Arabidopsis and rice. *Plant Cell.* 16:1220–1234.

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, dePamphilis CW, Wall PK, Soltis PS. 2009. Polyploidy and angiosperm diversification. *Am J Bot.* 96:336–348.

Spillane C, Schmid KJ, Laoueillé-Duprat S, Pien S, Escobar-Restrepo J-M, Baroux C, Gagliardini V, Page DR, Wolfe KH, Grossniklaus U. 2007. Positive darwinian selection at the imprinted *MEDEA* locus in plants. *Nature* 448:349–352.

Tang W, Kim TW, Oses-Prieto JA, Sun Y, Deng Z, Zhu S, Wang R, Burlingame AL, Wang ZY. 2008. *BSK*s mediate signal transduction from the receptor kinase *BRI1* in *Arabidopsis*. *Science* 321:557–560.

Van de Peer Y, Maere S, Meyer A. 2009. The evolutionary significance of ancient genome duplications. *Nat Rev Genet.* 10:725–732.

Wang D, Tyson MD, Jackson SS, Yadegari R. 2006. Partially redundant functions of two *SET*-domain polycomb-group proteins in controlling initiation of seed development in *Arabidopsis*. *Proc Natl Acad Sci U S A.* 103:13244–13249.

Wolfe KH. 2001. Yesterday's polyploids and the mystery of diploidization. *Nat Rev Genet.* 2:333–341.

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

Zou C, Lehti-Shiu MD, Thomashow M, Shiu S-H. 2009. Evolution of stress-regulated gene expression in duplicate genes of *Arabidopsis thaliana*. *PLoS Genet.* 5:e1000581.

Zwickl DJ. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Austin (TX): The University of Texas at Austin.