Plant Molecular Biology **56:** 929–945, 2004.

Genome-wide analysis of *S*-Locus F-box-like genes in *Arabidopsis thaliana*

Lei Wang[1,4,†], Li Dong[1,4,†], Yu'e Zhang[1,2], Yansheng Zhang[1], Weihua Wu[2], Xingwang Deng[3], and Yongbiao Xue[1,*]

[1]*Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100080, P.R. China (*author for correspondence; e-mail ybxue@genetics.ac.cn);* [2]*Department of Plant Sciences, College of Biological Sciences, State Key Laboratory of Plant Physiology and Biochemistry, China Agricultural University, Beijing 100094, P.R. China;* [3]*Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, Connecticut 06520-8104, USA;* [4]*The Graduate School of Chinese Academy of Sciences, Beijing, 100039, P.R. China;* [†]*these authors contributed equally to this work.*

Received 8 September 2004; accepted in revised form 14 November 2004

Key words: *Arabidopsis thaliana,* evolution, *S*-locus F-box-like protein, T-DNA insertion

Abstract

The *Antirrhinum S*-locus F-box gene, *AhSLF-S_2*, has been shown to determine the pollen function of S-RNase-mediated self-incompatibility (SI). Its initial identification led to the discovery of a large family of plant-specific F-box proteins, named the SLF (*S*-Locus F-box) family, including members from species with or without S-RNase SI system. To investigate the evolution and function of its family members in *Arabidopsis*, we first identified 92 *Arabidopsis* F-box proteins related to AhSLF-S_2, referred to as AtSFL (*S*-locus F-box-like) in this report. Phylogenetic analyses with family members from several plant species revealed that they could be classified into five subgroups, and the *SLF* genes appeared to have had a monophyletic origin. Yeast two-hybrid analyses showed that most AtSFL proteins could interact with one or more ASK (*Arabidopsis* Skp1-like) proteins, a component of the SCF (Skp1/Cullin or CDC53/F-box) complex, suggesting that AtSFLs may function in the process of ubiquitin/26S proteasome-mediated proteolysis. Transcript analysis found that most of *AtSFL* genes are expressed ubiquitously and only three of them (*AtSFL61*, *79* and *85*) displayed a tissue-specific pattern. In consistent, phenotypic observations for T-DNA insertion lines of 37 *AtSFL* genes revealed that most of them are functionally redundant, but inactivation of two *AtSFL* genes (*AtSFL 61* and *70*) appears to have caused developmental defects in embryo or female gametophyte. Our results show that a diversified expression and functional pattern are associated with *AtSFL* genes, indicating that they play important roles in various biological processes in *Arabidopsis*.

Introduction

Selective protein degradation is an important regulatory mechanism that allows cells to respond rapidly to intracellular signals and changing environmental conditions by adjusting the levels of key proteins. The ubiquitin/26S proteasome-mediated proteolytic pathway appears to be a dominant protein degradation system in plants (reviewed in Sullivan *et al.*, 2003; Smalle and Vierstra, 2004). In this pathway, the 76-amino acid protein ubiquitin (Ub) serves as marker attached to target protein. Ubiquitination of the target protein is carried out via a three-step conjugation cascade. First, the ubiquitin moiety is activated by the ubiquitin-activating enzyme (E1), then transferred to the ubiquitin conjugating enzyme (E2), and finally bound to a substrate

protein by the ubiquitin-protein ligase (E3). Polyubiquitinated proteins are recognized by the 26S proteasome for protein degradation. As the last step in the Ub-conjugation cascade, E3 acts as a ligase to recognize the substrate protein for ubiquitination (Patton *et al*., 1998). Currently, five types of E3 ligases have been identified according to their subunit organization and/or mechanism of Ub transfer and include N-end rule E3 or E3α, HECT (homologous to E6-APC-terminus) domain family, APC (anaphase promoting complex), RING (Real Interesting New Gene)/U-box and SCF (a complex of Skp1/Cullin or CDC53/F-box protein) (Hershko and Ciechanover, 1998; Deshaies, 1999). Except the N-end rule E3, the other four types of E3 have been described so far in plants (reviewed in Smalle and Vierstra, 2004). Among them, the SCF complex is a major type of the plant E3 ligases. F-box proteins contain a conserved F-box motif (approximately 40 amino acids) near their N-termini, which anchors the subunit to the rest of the SCF complex by interacting with Skp1. The C-termini of the F-box proteins usually contain one or a combination of several known protein–protein interaction motifs, e.g., leucine-rich repeat (LRR), WW (Trp-Trp), TPR (tetratricopeptide), Kelch repeat or WD40 repeat (beta-transducin related), for target identification. Many F-box proteins have been identified in numerous eukaryotes (reviewed in Winston *et al*., 1999; Kipreos and Pagano, 2000; Wang *et al*., 2002). In *Arabidopsis*, Gagne *et al*. (2002) have identified 694 potential F-box genes and classified them into 5 major families (A, B, C, D and E) with 20 subfamilies via a phylogenetic analysis. Yeast two-hybrid analyses have shown that representative members of the five major families of the putative F-box proteins could bind to one or more members of the ASK (*Arabidopsis* Skp1-like) family, providing support that many of these proteins are functional subunits of the SCF complexes. Meanwhile, Kuroda *et al*. (2002) have also identified at least 568 predicted F-box proteins encoded in the genome of *Arabidopsis*. Domain search analyses have identified several unique functional domains, allowing the classification of the potential *Arabidopsis* F-box proteins into 19 groups. So far, only 13 plant F-box proteins have been identified with a function in various aspects including hormonal responses, floral formation, photomorphogenesis, circadian rhythms, senescence and pathogen invasion (reviewed in Sullivan *et al*., 2003; Smalle and Vierstra, 2004).

The pollen-specific SLF (*S*-Locus F-box) protein was first identified in *Antirrhinum* by Lai *et al*. (2002). This gene, *AhSLF-S_2*, located 9-kbp downstream of *S_2-RNase* gene, encodes an F-box containing protein and homolog searches led to the discovery of a large family of F-box proteins in several plant species including *Arabidopsis*, referred to as the SLF (*S*-Locus F-box) family (Lai *et al*., 2002). Subsequently, similar pollen-specific *S*-locus F-box genes have been detected in several *Prunus* species (Entani *et al*., 2003; Ushijima *et al*., 2003; Yamane *et al*., 2003) and solanaceous species (Wang *et al*., 2003, 2004; Qiao *et al*., 2004a; Sijacic *et al*., 2004). Recently, Sijacic *et al*. (2004) have shown that PiSLF-S_2 of *Petunia inflata* controls the pollen function of self-incompatibility (SI). Qiao *et al*. (2004a) also have demonstrated that AhSLF-S_2 mediates the pollen function of S-RNase-based SI. Furthermore, Qiao *et al*. (2004b) have found that AhSLF-S_2 physically interacts with S-RNases likely through the formation of a $SCF^{AhSLF\text{-}S2}$ complex which possibly inhibits S-RNase activity during compatible pollination through the ubiquitin/26S proteasome-mediated protein degradation pathway.

In *Arabidopsis*, only one sequence similar to AhSLF-S_2 called SON1 (Suppressor Of Nim-1) similar to AhSLF-S_2 has been studied, which negatively regulates a novel SAR (Systemic Acquired Resistance)- and NIM1 (Non-Inducible Immunity 1)-independent defense response (Kim and Delaney, 2002). Nevertheless, little is known about the presence and function of the SLF family members in other species without S-RNase based SI system, referred to as SFL (*S*-locus F-box-like), especially those in *Arabidopsis*. In this study, we identified a total of 155 predicted SLF and SFL polypeptides from several plant species including *Arabidopsis*. In particular, a genome-wide analysis revealed 92 *AtSFL* (*Arabidopsis thaliana S*-locus F-box-like) genes forming a large subfamily of the F-box proteins that could further classified into 4 classes based on their phylogenetic relationships. Interestingly, our phylogenetic analysis showed that the SLF proteins likely share a common origin. We also

provided evidence that several of the AtSFL proteins are capable of directly interacting with one or more ASK proteins, suggesting that they could be part of the SCF complexes. Furthermore, molecular analyses revealed that a diversified expression and functional pattern is associated with the *AtSFL* genes, indicating that they play important roles in various biological processes in *Arabidopsis*.

Materials and methods

Plant materials and growth conditions

Arabidopsis thaliana ecotype Columbia (Col-0) was used. Sixty-one T-DNA insertion lines (Table S1) were obtained from SALK or Garlic T-DNA insertion collection (http://signal.salk.edu and http://tmri.org/pages/collaborations). *Arabidopsis* plants were grown on MS medium containing 2% (W/V) sucrose and 0.8% (W/V) agar or grown in soil at 22 °C with a 16/8 h light/dark cycle.

Prediction of F-box proteins

AhSLF-S_2 (Accession number: CAC33022 and gi: 13161540) was used as query in BLAST search against EMBL-EBI database with an E-value cut-off level of 0.002. The proteins found by the EMBL-EBI blast were used for a domain search with Pfam (http://www.sanger.ac.uk/Software/Pfam/) and SMART (http://smart.embl-heidelberg.de/) databases. Accession numbers provided by EMBL-EBI were changed to AGI (Arabidopsis Genome Initiative) numbering system by BLAST of TAIR (The Arabidopsis Information Resource, http://www.arabidopsis.org/Blast/).

Amino acid sequence alignments, phylogenetic analysis and chromosome mapping

Multiple amino acid sequence alignments and phylogenetic tree construction were made using MegAlign program, a component of DNAstar package, followed by manual refinement. Protein structure prediction was performed using 3D-PSSM (http://www.sbg.bio.ic.ac.uk/~3dpssm/). Chromosomal maps were generated using the Chromosome Map Tool in TAIR (http://www.arabidopsis.org/jsp/ChromosomeMap/tool.jsp).

EST identification and search of microarray-based transcript data

Predicted mRNA sequences were used to conduct BLAST searches against GenBank *Arabidopsis* EST dataset. EST sequences with an 80–100% match were retrieved, checked for redundancy, and analyzed to verify the match. *Arabidopsis* microarray elements from the *Arabidopsis* Functional Genomics Consortium, Affymetrix 8K and 25K GeneChip®, and CATMA arrays were searched for expression profiles of *AtSFL* genes via the TAIR web site (http://www.arabidopsis.org/tools/bulk/microarray/index.jsp).

RNA preparation and reverse transcription-PCR

Total RNA was isolated from leaves, flower buds, siliques and roots of 4-week-old plants, using a RNA isolation kit (QIAGEN, CA, USA). Of the total RNA, 1 μg was digested with RNase-free DNase I (TaKaRa, Dalian, China) and used for reverse transcription with M-MLV reverse transcriptase (Invitrogen, CA, USA). After 1/10 dilution, 1 μl of the synthesized cDNA was used for reverse transcription (RT)-PCR (polymerase chain reaction). PCR conditions were 30 cycles of 94 °C for 1 min, 58 °C for 1 min, 72 °C for 1 min, and finally 72 °C for 5 min. PCR products were cloned and sequenced to check the identity.

Plasmid construction and yeast transformation assay

Coding sequences of 11 *AtSFL* genes (*AtSFL1*, *5*, *35*, *37*, *54*, *55*, *61*, *75*, *77*, *79* and *87*) representing members from each class of the AtSFL family were introduced to pGBKT7 bait plasmid (Clontech, CA, USA) to produce fusion proteins with the GAL4 DNA binding domain, and 17 *ASK* cDNA (*ASK1*, *2*, *3*, *4*, *6*, *7*, *8*, *9*, *10*, *11*, *12*, *13*, *14*, *16*, *17*, *18* and *19*) were inserted to pGADT7 prey plasmid containing the GAL4 activation domain. Yeast transformation, growth conditions, and assays for β-galactosidase activity were performed according to the manufacturer's instructions (Clontech, CA, USA).

Immunoblotting

Protein extracts for immunoblotting were prepared according to the method by Printen and Sprague (1994). Total proteins were separated by electrophoresis on 12% SDS-polyacrylamide gel and transferred onto nitrocellulose membranes (Micron Separations, Westborough, MA) using a Bio-Rad Transblot SD wet electroblotting apparatus (Hercules, CA, USA) in transfer buffer (10 mM Tris, 192 mM glycine, 20% methanol). After blotting, the membranes were blocked by incubation with 10% dried skimmed milk in TBS buffer (25 mM Tris–HCl, pH 7.5, 100 mM NaCl) for 1 h. Primary antibody of anti-Myc (Sigma-Aldrich) at a dilution of 1:10,000 was used to detect *BD*: *AtSFL* fusion proteins expression and anti-HA (Sigma-Aldrich, USA) to detect *AD*: *ASK* fusion proteins expression. Immune complexes were subsequently detected using alkaline phosphatase-conjugated secondary antibodies and nitroblue tetrazolumy 5-bromo-4-chloro-3-indolyl phosphate (Harlow and Lane, 1988).

Promoter analysis and Arabidopsis *transformation*

To construct an *AtSFL79* promoter-β-glucuronidase (GUS) fusion, a genomic DNA region corresponding to 2006 bp upstream predicted ATG codon of AtSFL79 ORF was amplified from *Arabidopsis* genomic DNA by PCR using two primers (5′-ATAAAGCTTTCACAGGGTCAAA GAAACGG-3′ and 5′-ATATCTAGATTCTGT TCTCCGGAAGAAAA-3′). Genomic DNA was isolated according to Liu *et al.* (1995). The PCR product was cloned into pBI101.2 binary vector between *Hin*dIII/*Xba*I sites (Clontech, CA, USA). Junction sequences of the resulting construct *AtSFL79::GUS* were determined. The construct was introduced into *Agrobacterium* strain GV3101 and subsequently transformed into wild type *Arabidopsis* (Col-0) plants according to Bechtold and Pelletier (1998). Kanamycin-resistant T_1 plants were used for GUS activity detection as described by Weigel and Galzebrook (2002) for both whole-mount and section observations.

PCR screening of T-DNA insertion lines

To identify plants with the T-DNA insertion, we performed PCR analyses of genomic DNA, using one gene-specific primer and one T-DNA-specific primer in the left border. Then gene-specific primer pairs flanking the insertion site were used to verify whether the plant is homozygous or not. Gene- and T-DNA-specific primers used are listed in Table S1.

T-DNA insertion copy number was determined by genomic DNA blot analysis as described previously (Xue *et al.*, 1996) using NPT II (neomycin phosphotransferase, for SALK insertion lines) or BAR (for Garlic insertion line) gene as probes. DNA (1 μg) was digested, separated on 0.8% agarose gel, and transferred onto Hybond N+ (Amersham, Buckinghamshire, UK) membrane. Prehybridization, hybridization, and washing of the blot were performed as recommended by the manufacturers. Probes were labeled with ^{32}P by random priming using the Prime-a-Gene labeling system (Promega, Madison, WI). To test linkage of T-DNA insertions to mutations, heterozygous insertion lines were crossed with wild type plant as male and progeny was tested by PCR analysis using T-DNA (LBb1: 5′ –GCGTGGACCGC TTGCTGCAACT-3′, for SALK insertion lines; G-LB1: 5′-GGATAAATAGCCTTGCTTCC-3′, for Garlic insertion lines) and gene-specific primers (Table S1).

Whole-mount preparation

Siliques of different developmental stages were collected, dissected, cleared by Herr's fluid (Herr *et al.*, 1971) and observed using DIC optics (Olympus, Japan). Images were recorded with Olympus BX51 microscope.

Results

Arabidopsis *contains a large gene family similar to the S-locus F-box genes*

To identify *Arabidopsis* genes related to the *S*-locus F-box genes, we searched the entire *Arabidopsis* genome sequence from EMBL-EBI database using the WU-BLAST program, queried by the peptide sequence of AhSLF-S_2 of *Antirrhinum*. This revealed that 96 predicted *Arabidopsis* polypeptides are homologous to AhSLF-S_2. To further define them, we examined whether they contained a potential F-box domain and subse-

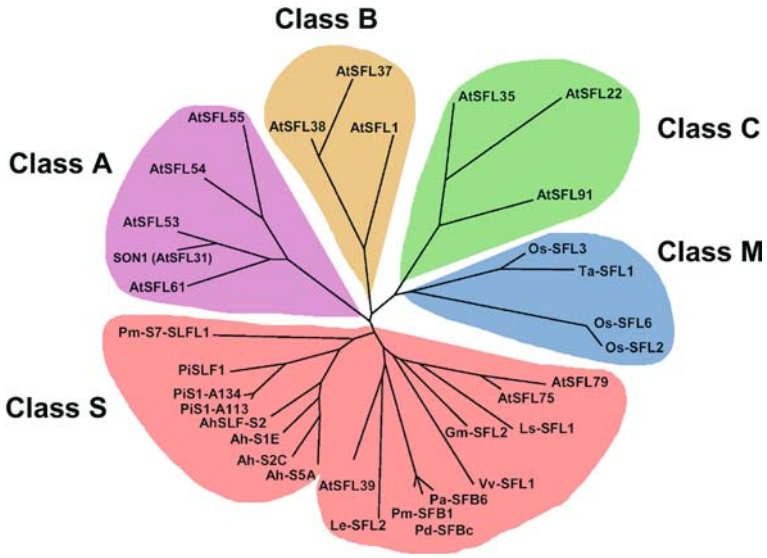

Figure 1. A representative phylogenetic tree of predicted SLF and SFL polypeptides in plants. The complete tree is shown in Figure S1. All of the sequences could be grouped into 5 major classes, namely S, A, B, C and M. The sequences of the Class S, A, B, C are from dicotyledonous plants [*Arabidopsis thaliana* (At), *Prunus dulcis* (Pd), *Prunus mume* (Pm), *Petunia inflata* (Pi), medick (Mt), tomato (Le), soybean (Gm), grape (Vv), lettuce (Ls) and potato (St)], while the sequences of the Class M come from monocotyledonous plants [rice (Os) and wheat (Ta)].

quently found that 92 of them are predicted to encode proteins with an F-box motif in their N terminal regions and were referred to as *AtSFL* (*Arabidopsis thaliana S*-locus F-box-like) genes (Table S2).

Phylogenetic analysis of predicted SFL and SLF polypeptides in plants

To examine the evolutionary relationships of the predicted SFL and SLF polypeptides in plants, an unrooted phylogenetic tree of the entire collection of these sequences was created with the Clustal W method (Figure S1). The phylogenetic tree also included SFL proteins identified from EST or genomic sequences from several other plant species including grape (*Vitis vinifera*), lettuce (*Lactuca sativa*), medick (*Medicago truncatula*), potato (*Solanum tuberosum*), rice (*Oryza sativa*), soybean (*Glycine max*), tomato (*Lycopersicon esculentum*) and wheat (*Triticum aestivum*). Most of the sequences were derived from dicot species except two monocot species (rice and wheat) (Table S3). A representative tree is shown in Figure 1. Manual inspection of the phylogenetic tree revealed that they could be organized into five groups (Class A, B, C, M and S). AtSFL39, 40, 66, 75, 79, 81and 87 are grouped together with the *S*-locus F-box proteins from *Antirrhinum*, *Prunus* and *Petunia*, forming the Class S. The rest AtSFLs are further divided into three groups, which we termed Class A, B, and C, based on their evolutionary distances to the Class S proteins. Class A, the nearest clade to the Class S, has 43 members, Class B 10, and Class C 32. The sequences from the monocotyledons form an independent group, referred to Class M, containing 14 predicted polypeptides. This analysis indicated that SFL proteins probably had evolved from a single ancestral protein and the SFL sequences were diversified before the origin of the dicotyledons and monocotyledons.

Besides *Arabidopsis*, all of the available SLF-like proteins from the dicotyledons were grouped together with the Class S (Figure 1). Interestingly, the class S is divided into two subgroups. One group encompasses all known *SFB* (*S*-haplotype-specifc F-box) genes of *Prunus*, which possibly control the pollen function of self-incompatibility (Ushijima *et al*., 2003, 2004; Yamane *et al*., 2003), whereas the other group includes all of pollen *S* genes found in *Antirrhinum* and *Petunia* (Qiao *et al*., 2004a, b; Wang *et al*., 2003, 2004; Sijacic *et al*., 2004), such as AhSLF-S_2 and Pi-SLF_2. This

result suggests that the pollen *S* genes may have adopted two separate routes during the evolution of self-incompatibility.

Analysis of conserved domains among the predicted SLF and SFL polypeptides

Many F-box proteins are known to contain protein-interacting domains at their C-terminal regions (Bai *et al*., 1996). Recent genome-wide analyses of the *Arabidopsis* F-box proteins (Gagne *et al*., 2002; Kuroda *et al*., 2002) have revealed that F-box proteins contain several known protein–protein interaction domains, such as LRR, WD40, Kelch repeats, ARM (armadillo repeats), PAS/PAC, actin-like and RING finger domains. Other domains were also found, such as zinc finger, HSF (heat shock factor), Myb and HLH (helix-loop-helix) PPR (pentatricopeptide repeat), RVT (reverse-transcriptase-like), Tub, FBD (domain in F-Box and BRCT domain containing plant proteins) domain and SEL1-like (suppressor and/or enhancer of *lin-12*) repeats. CRFA/FBD, CRFB, CRFC are the newly defined C-terminal domains, which are also predicted to be involved in the protein–protein interactions (Kuroda *et al*., 2002). To identify any possible known motifs in the C-terminal regions of predicted AtSFL polypeptides, we did NCBI web site search and revealed that, apart from AtSFL75 and AtSFL79 (both have a LRR domain), no known protein-interacting domains were present in the C-terminal regions of the predicted AtSFL proteins (data not shown). Nevertheless, alignment of the Class S proteins showed that there are four conserved regions down stream of the F-box motif, consistent with the previous reports (Lai *et al*., 2002; Zhou *et al*., 2003). They were referred to as C1, C2, C3 and C4 domains, respectively (Figure 2 and Figure S2A). Some amino acids are highly conserved like NGL/I and NPXXXXXXXLP (X denotes any amino acid) in C1, YKV/IV/I and V/I/LXXXXXXSWR/K in C2, GXXXW and VLX XXXXXXXFXXXXXP in C3 and IWXXXXX XXXXXW in C4, indicating that these regions are likely important for their structure and function.

However, the conservation varies among the other classes of AtSFL proteins. Besides the Class M, which possesses only the C1, C3 and C4 domains, all of the AtSFL proteins contain the C1, C2 and C4 domains. In addition to the C1, C2 and C4 domains, the predicted proteins of the Class A, B and C each contain an extra domain, respectively. They are CA (conserved domain in Class A), CB (conserved domain in Class B) and CC (conserved domain in Class C). CA domain corresponds to SXXGNXYW and FDFXXXXXXX XLP, CB to NGSLXW, DLHTE and LCV, and CC to GXLXYXA, IVXFXXXXXE and YXGKL (Figure S2B). All of them are deduced to derive from the C3 domain.

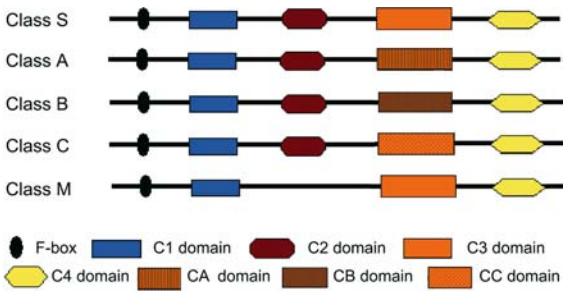

Figure 2. Conserved domains in the C-terminal regions of predicted SFL polypeptides. Besides the F-box motif, the Class S sequences contain the C1, C2, C3 and C4 domains, the Class A sequences C1, C2, CA and C4 domains, the Class B sequences C1, C2, CB and C4 domains, the Class C sequences C1, C2, CC and C4 domains, the Class M sequences C1, C3 and C4 domains. CA, CB and CC are diversified the C3 domain.

Nevertheless, C1, C2 and C4 were not conserved completely but with some variations. The C1 and C2 domains in the Class A proteins correspond to the conserved region 1 and 2 of CRFB domain (Kuroda *et al*., 2002). Kuroda *et al*. (2002) also classified 88.4% (38/43) of the predicted Class A proteins as the CRFB-containing proteins, whereas the C2 and CC domains in the Class C correspond to the region 1 and region 2, 3 and 4 of CRFC domain, respectively. Of the 32 predicted Class C peptides, 29(90.6%) were originally classified as the CRFC-containing F-box proteins (Table S2). In the classification system of Gagne *et al*. (2002), the distribution of AtSFL peptides is more systematic and have classified all of the 694 F-box genes in *Arabidopsis* into five families (A, B, C, D and E) with 20 subfamilies (A1–A6, B1–B7, C1–C5, D and E). The predicted Class A AtSFL polypeptides, containing A Type-I domain, were grouped into the A1 and A2 subfamilies. The Class B containing A Type-II domain and four peptides of the Class S (AtSFL39, 66, 81 and 87) were sorted into A3, and the Class C into A4 and A5, respectively. AtSFL75 and 79, the

LRR-containing proteins, belong to B1 subfamilies, and AtSFL40 (a polypeptide of the Class S) to the C4 family (Table S2). The A Type-I and -II domains, corresponding to the C2, C3 and C4 domains in AtSFL peptides, are related 200-aa hydrophobic regions with positionally conserved tryptophans and can be distinguished from each other by additional conserved regions (Gagne *et al.*, 2002). If the predicted SFL proteins could function as ubiquitin ligases (E3) as AhSLF-S_2, their variable C-terminal domains may indicate a diversity of the substrates that they could recognize. The highly conserved C1, C2 and C4 domains suggest that they may play a role in protein stability and function.

To further examine protein-interacting domains, we predicted possible secondary structures of the entire sequences of several AtSFL proteins representing the five classes against the 3D-PSSM server (http://www.sbg.bio.ic.ac.uk/~3dpssm/). The results showed that all of the conserved domains consist of 2–6 β-sheet structures, except the F-box motif, which forms a α-helix structure. The serial β-sheet structures are hydrophobic, which may be structurally important for AtSFLs (Figure S2).

Chromosomal location and genomic structure of AtSFL *genes*

To examine the genomic organization of the *AtSFL* genes, we searched the locations of the *AtSFL* loci on *Arabidopsis* chromosomes. Figure S3 shows the distribution of 92 predicted *AtSFL* genes on all five chromosomes. Overall, 36 *AtSFL* genes are arranged in tandem repeats of two to seven copies. Chromosome III has a region containing 26 *AtSFL* genes. Additionally, 29 of the 92 genes were within five genes of a closely related gene in the phylogenetic tree (see Figure S1) and 33 of them belong to adjacent genes. Similar genomic organizations also were observed in *Antirrhinum*, *Prunus* and *Petunia* (Ushijima *et al.*, 2003; Zhou *et al.*, 2003; Wang *et al.*, 2004). S_1, S_2, S_4 and S_5 haplotypes of *Antirrhinum* each have two or three highly homologous F-box genes (Zhou *et al.*, 2003). Similarly, S^c and S^d haplotypes of *Prunus dulcis* also contain two to three paralogous copies of *SFB*/*SLF* genes (Ushijima *et al.*, 2003) and S_1 haplotype of *Prunus mume* has 3 F-box genes and S_7 has 4 F-box genes (Entani *et al.*, 2003). In *Petunia inflata*, at least two paralogous copies of *PiSLF* have been detected (Wang *et al.*, 2004). These findings suggest that tandem duplication of chromosomal regions might have played a role in creating the *SFL* loci.

To examine the genomic structures of *AtSFL* genes, we did gene model analysis which revealed that 85.9% (79) of *AtSFL*s are predicted to be intron-free, a much higher percentage than that predicted for overall *Arabidopsis* genes (20.7%) (Arabidopsis Genome Initiative, 2000), and 12 of them (13.0%) have one intron and only 1 (*AtSFL24*) has two introns (1.1%) (Table S4). In *Antirrhinum*, *Prunus* and *Petunia*, the *SLF* also are intron-less (Lai *et al.*, 2002; Entani *et al.*, 2003; Ushijima *et al.*, 2003; Zhou *et al.*, 2003; Wang *et al.*, 2004).

By now, all identified *SLF* genes are closely linked to *S-RNase* genes (Lai *et al.*, 2002; Entani *et al.*, 2003; Ushijima *et al.*, 2003; Zhou *et al.*, 2003; Wang *et al.*, 2004). Although SI in Brassicaceae, to which *Arabidopsis* belongs, is sporophytically regulated, *S-like RNases* genes still exist in *Arabidopsis* (Taylor *et al.*, 1993; Bariola *et al.*, 1994). To examine whether *AtSFL* genes are closely linked to *S-like RNases* (*RNS1, RNS2, RNS3, RNS4 and RNS5*) in *Arabidopsis*, we calculated the physical distances between the four *RNS*s and corresponding nearest *AtSFLs*. The distances between *RNS1* and *AtSFL26* and *7* and *RNS5* as well as *AtSFL36* and *RNS2* are 389,697, 450,850 and 348,248 bp, respectively (Figure 3, also see Figure S3). However, we found several other types of F-box genes surround the four *RNSs* and one of them is only 31.4 kbp from *RNS4* (Figure 3). We noticed that some of them contain Kelch repeat protein interaction domain. Except the F-box gene, many genes participating in the protein degradation pathway are found in the *RNSs* surrounding regions such as *Cul-3, UBC2, ASKs* and *UBC1* (Figure 3). Moreover, like *Antirrhinum* and *P. inflata* (Lai *et al.*, 2002; Wang *et al.*, 2004), abundant retrotransposons are found in the *RNS1* surrounding regions (Figure 3).

Expression profiling of AtSFL *genes*

Initial searches against the *Arabidopsis* EST database revealed 95 ESTs for 31 of *AtSFL* loci, representing 33.7% of *AtSFL* genes (Table S4). Overall, the number of *AtSFL* genes with ESTs

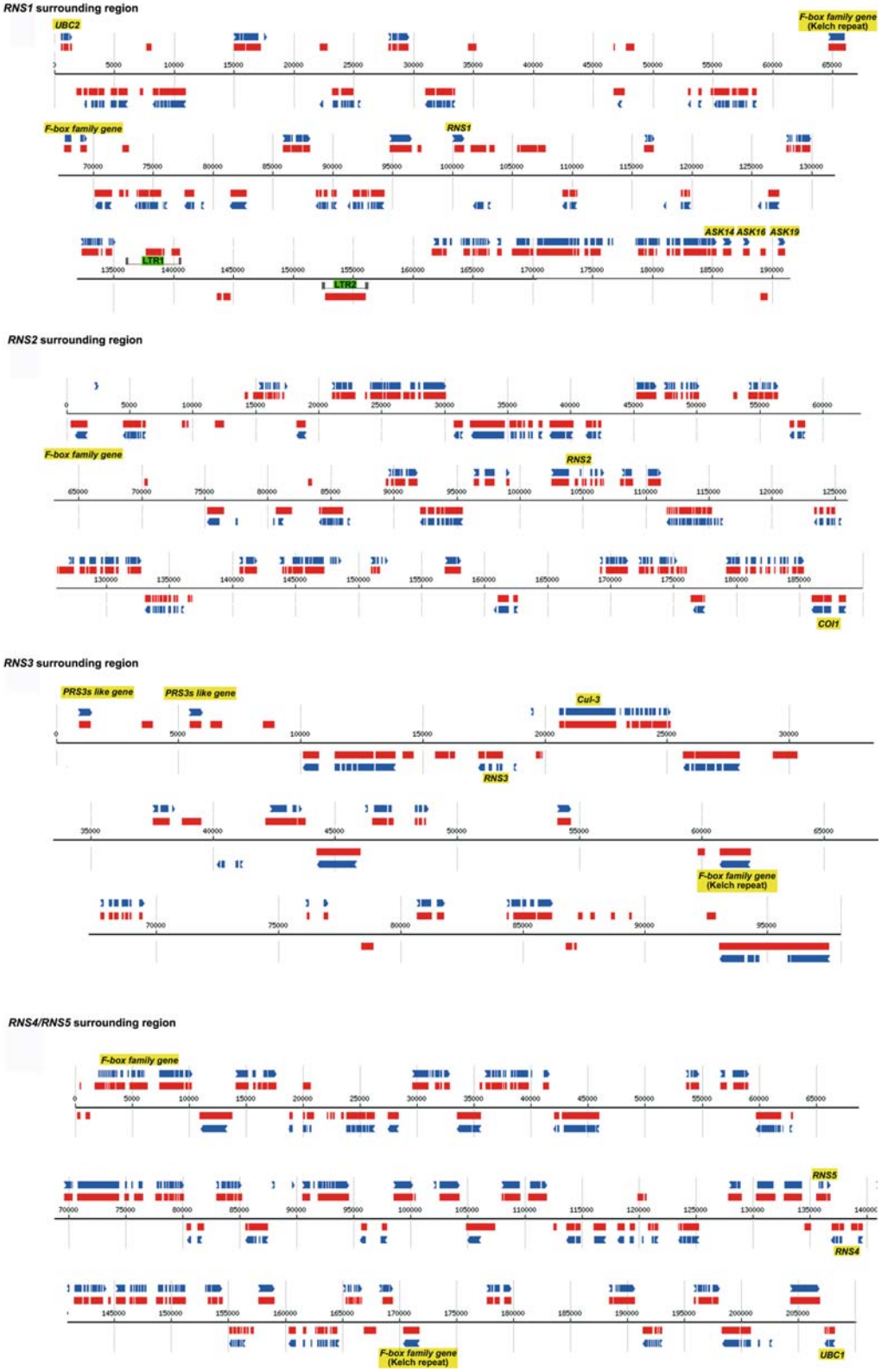
RNS1 surrounding region
UBC2
F-box family gene (Kelch repeat)
F-box family gene
RNS1
ASK14 ASK16 ASK19
LTR1
LTR2
RNS2 surrounding region
F-box family gene
RNS2
COI1
RNS3 surrounding region
PRS3s like gene
PRS3s like gene
Cul-3
RNS3
F-box family gene (Kelch repeat)
RNS4/RNS5 surrounding region
F-box family gene
RNS5
RNS4
F-box family gene (Kelch repeat)
UBC1

Figure 3. Genomic structures of *AtRNS*-containing regions. The nucleotide sequences of the four *AtRNS*-specific regions were characterized by Fgenesh, BLASTX, and AutoPredLTR search of RiceGAAS (Sakata *et al*., 2002). Putative *RNSs*, F-box and other ubiquitin/26S proteasome pathway-related genes are represented with yellow boxes. Putative LTRs are represented with green boxes. Blue, red, and gray boxes represent the results of Fgenesh, BLAST, and AutoPredLTR searches, respectively. Each pair of LTRs is linked by lines.

was much lower than that estimated for whole genome (60.3%) (Arabidopsis Genome Initiative, 2000), suggesting that many are expressed at low levels or restricted to specific cells or tissues or developmental stages. Nevertheless, some genes have more matched EST than others, for example, *AtSFL5* (*At1g12860*) has 9 ESTs, but *AtSFL 77* (*At4g19940*) has only one, suggesting that they are expressed differentially.

To further examine the expression patterns of *AtSFL* genes, we performed RT-PCR for all of the 92 genes in 4 major tissues including leaves, siliques, flower buds and roots (Figure 4). PCR products from cDNA templates were sequenced to ensure that the amplified sequences match the known sequences of the corresponding genes. RT-PCR for the *Arabidopsis tubulin* gene was also included as a control for RNA extraction and PCR. In total, the expression of 24 *AtSFL* genes was detected including 6 genes (*AtSFL35, 44, 61, 70, 73* and *78*) that have no previously reported ESTs (Table S4). Most of the tested *AtSFL* genes (*AtSFL1, 5, 12, 35, 37, 39, 40, 44, 46, 53, 54, 55, 66, 70, 73, 75, 77, 78, 81, 86* and *87*) were expressed constitutively in all tissues tested, suggesting that they may have basic or broad functions. Although some genes accumulated in all tissues, they appeared to be expressed differentially in the four tested tissues. For example, *AtSFL5*, *39*, *81*, *87* and *73* appeared to have a lower expression in one tissue (*AtSFL5* and *81* in roots, *AtSFL39* in leaves, *AtSFL87* in flower buds and *AtSFL73* in siliques), indicating that they are less active in these tissues. *AtSFL40*, *75*, *78* and *86* are highly expressed in one tissue (*AtSFL40* in roots, *AtSFL75* and *78* in flower buds and *AtSFL86* in leaves), and they may function specifically for the tissue development and other events that the tissue participates. *AtSFL75* was expressed highly in reproductive tissues: flower buds and siliques, suggesting that it may function during the reproduction development. Only 3 of them appeared to be tissue-specific. *AtSFL79* and *61* have a similar expression pattern and their transcripts are only detected in siliques and inflorescences, indicating that it may function during the reproduction development. *AtSFL85* was only expressed in siliques without any detectable expression in other tissues and may have a more restricted function(s).

The seven genes closely related to *AhSLF, PmSLF (PdSLF)* and *PiSLF* genes in the phylogenetic tree showed differential expression patterns. *AtSFL39*, *40*, *66*, *75*, *81* and *87* are all constitutively expressed in all tissues tested,

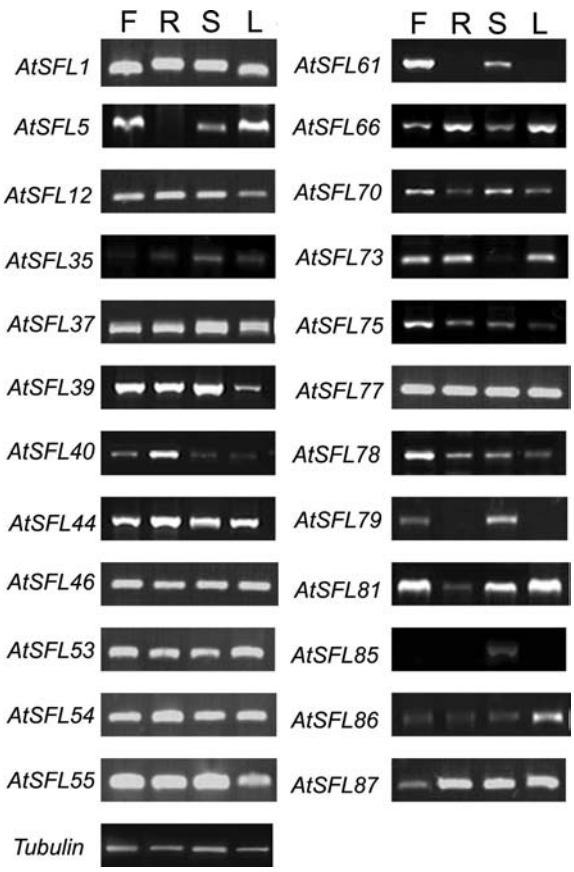

Figure 4. Expression of *AtSFL* genes in various tissues of 4-week-old *Arabidopsis*. Total RNAs treated with DNaseI from flower buds (F), siliques (S), leaves (L) and roots (R) were used for RT-PCR analysis. The *tubulin* gene was used as a positive control.

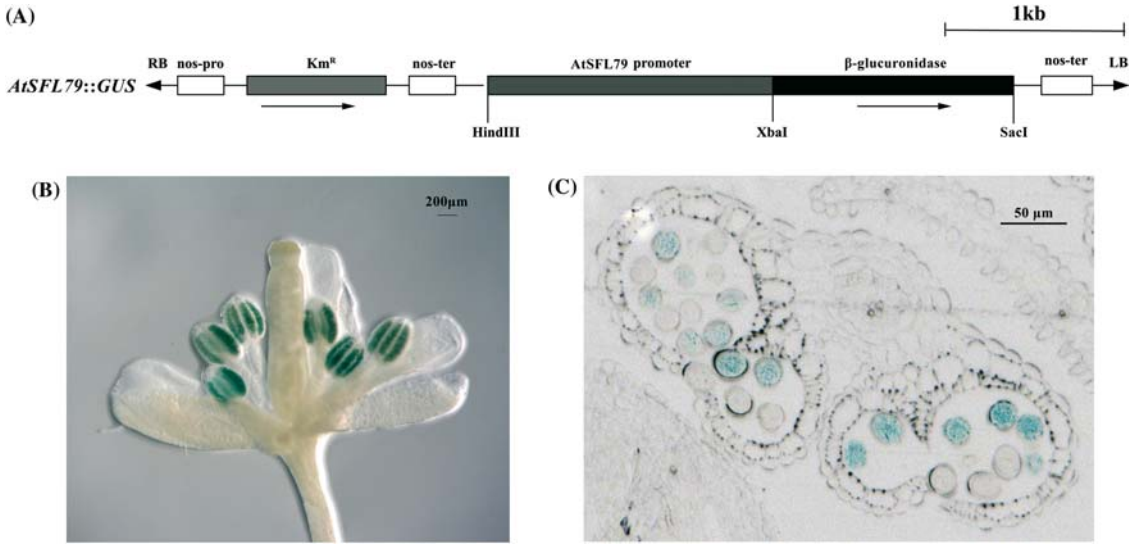

Figure 5. Expression of *AtSFL79* promoter and GUS fusion in transgenic *Arabidopsis*. (A) A schematic diagram of *AtSFL79*::*GUS*. RB and LB, right and left borders of the T-DNA; Km^r, kanamycin resistance (neomycin phosphortransferase) gene. Several restriction sites also are indicated, (B) Whole-mount examination of GUS activity of mature anther in transgenic plant and (C) A transverse mature anther section showing that GUS product was specifically detected in pollen. Scale bars are shown.

whereas *AtSFL79* is specific to the reproductive tissue. To further examine the expression of *AtSFL79*, we used its predicted promoter to drive *β*-glucuronidase (GUS) reporter gene in transgenic *Arabidopsis*. As shown in Figure 5, the reporter gene product was specifically detected in mature pollen, confirming that *AtSFL79* is specifically expressed in pollen. No GUS expression was detected in other tissues (data not shown). All of the *SLF* genes in *Antirrhinum*, *Petunia* and *Prunus* are pollen-specific (Ushijima *et al*., 2003; Zhou *et al*., 2003; Wang *et al*., 2004), suggesting that *AtSFL79* may possess function similar to that in *Antirrhinum*, *Petunia* and *Prunus*.

Expression patterns of AtSFLs are associated with diversified biological processes

To investigate possible biological processes in that AtSFL proteins are involved, we searched *Arabidopsis* microarray elements via the TAIR web site (http://www.arabidopsis.org/tools/bulk/microarray/index.jsp). The data for 13 *AtSFL* genes (*AtSFL36*, *37*, *39*, *40*, *53*, *59*, *61*, *65*, *75*, *79*, *81*, *87* and *88*) were collected. Transcriptional products of two genes, *AtSFL40* and *53* were detected in most of the microarray experiments, which makes it possible to perform a cluster analysis with other genes. To find out a useful cutoff for the correlation analysis, the sub-cluster of *tubulin* genes was extracted from the treeview result of AFGC microarray elements. Tubulin is a heterodimer of tubulin-*α* and -*β* that constitutes the monomer for microtubule assembly. The cluster includes 5 of 12 *tubulin* genes in *Arabidopsis* (Figure S4A). The correlation ratio between the *tubulin* *α*-2/-4 and the *tubulin* *β*-4 is 0.731, and 0.520 between the *tubulin* *α*-3/-5 and the *tubulin* *β*-5. In our analyses, 0.500 was thus selected as a cutoff value in 292 microarray experiments. Eighty-one genes were congregated in the *AtSFL40* cluster (Figure S4B), including some specific genes involved in the 26S proteasome pathway. They are *At5g42970* encoding for a subunit of COP9 complex (Serino *et al*., 1999), *At4g39910* for an ubiquitin-specific protease 3 (Chandler *et al*., 1997), *At5g67250* for SKIP2 (SKP1 interacting partner), *At2g21950* for SKIP6 and two F-box proteins (At3g48880 and At2g40920). In addition to these genes, we also found several putative protein kinase genes (*At2g03890*, *At4g10730* and *At5g58350*) and C3HC4-type RING finger genes (*At2g37150*, *At3g61180* and *At4g35410*) in this cluster. The *AtSFL53* cluster had only 5 genes (cutoff of 0.500) including an F-box gene (*At1g78100*) and a C4HC3-type zinc finger (PHD finger) gene (*At3g01460*). However, when the cutoff was reduced to 0.255, the genes that were

homologous to those in the *AtSFL40* cluster were also included (Figure S4C). They were genes encoding an F-box protein (At1g53320), putative protein kinases (At5g66880 and At3g56050), and a C3HC4-type zinc finger protein (At5g58780). In addition, genes encoding cytochrome P450s were found in the two clusters. These results suggest that these genes are co-regulated and likely function in the same pathway(s).

Queried for the array elements with a fold change value of 2 at a replicate level, we found that *AtSFLs* may function in various biological processes (Table S5). *AtSFL36*, *37* and *40* may be associated with the regulation of circadian rhythm. We extracted 21 microarray experimental data corresponding to the circadian rhythm on 3 *AtSFLs* and *ZTL*, the known clock-associated gene (Somers *et al*., 2000). Their expression profiles are clustered in Figure S4D with *AtSFL53* as a negative control because of its nearly consistent expression under various circadian rhythms in Col wild type and Ler or Col mutants. The expression of *ZTL* is known to be independent of light/dark cycles and the circadian clock (Somers *et al*., 2000). In whole plants after 24-h period in continuous light, it is up regulated, whereas decreased from 28 h on relative to the standard. LHY and CCA1 are negative regulatory elements required for the central oscillator function (Alabadi *et al*., 2002). Figure S4D also shows that the *ZTL* is down regulated in *lhy cca1* null and *CCA1* overexpressing mutants no matter how long the light period is. *AtSFL36, 37* and *40* followed a similar pattern. In addition, the expression level of *AtSFL40* increased in the zinc transport experiment and decreased in the tissue-cultured shoots, indicating that it is also involved in other processes. Putative functions, such as epidermal cell differentiation and regulation of an overlapping set of genes involved in cytosine methylation that other *AtSFLs* may be associated are listed in Table S5. These results indicate that *AtSFLs* play roles in different biological processes and different *AtSFLs* or F-box genes may function in the same process.

Physical interaction of AtSFLs with ASKs

In general, F-box proteins directly bind to Skp1 through F-box domains (Patton *et al*., 1998). In the system of Gagne *et al*. (2002), except AtSFL75 and 79 in B1 and AtSFL40 in C4, most AtSFL proteins belong to A1–A5 subfamily (see Table S2). Their yeast two-hybrid results showed that representative proteins of A2 subfamily could interact with ASK11, A4 with ASK16, A5 with ASK1, 2, 11and 13, C4 with ASK1, 2, 4, 11, 13 and 18 and A1 with none of 10 ASKs they selected, while no data showed the interaction of B1subfamily with ASKs (Gagne *et al*., 2002). In the system of Kuroda *et al*. (2002), 38 AtSFL proteins belong to CRFB group and 29 AtSFL proteins belong to CRFC group (see Table S2). However, their results showed that representative protein sequences of these two families could not interact with ASK1.

Nevertheless, except AtSFL65 (CRFB), 80 (CRFC) and 22 (CRFC), which could not interact with any of 12 selected ASKs (Takahashi *et al*., 2004), no data directly showed the interaction of the other AtSFL proteins with ASKs (Gagne *et al*., 2002; Kuroda *et al*., 2002; Risseeuw *et al*., 2003). To test whether the F-box proteins from each of the four classes of AtSFLs could interact with 19 ASK proteins, we performed a yeast two-hybrid test (Figure 6). We amplified cDNA clones of 17 *ASK* genes but failed to amplify *ASK5* (*At3g60020*) and *15* (*At3g25650*) probably because of their low abundance. All of the 17 amplified ASK cDNAs were cloned into a Gal-4 activation domain vector pGADT7 and then transformed to yeast strain HF7C. We also selected 11 F-box proteins from the four classes, cloned their full-length cDNAs into a Gal-4 DNA-binding domain vector pGBKT7 and transformed to the same yeast stain.

As shown in Figure 6A and B, not all of the F-box proteins could interact with the ASK proteins. In the Class S, two of them (AtSFL75 and 79) share a similar interaction profile. They all interacted with ASK2, 9 and 14, but AtSFL75 also interacted with ASK18. AtSFL87 showed the physical interaction with 9 ASKs, more than any other F-box proteins tested. In the three selected proteins of the Class A, only AtSFL61 interacted with ASK1, 2, 9, 11, 14, 16, 18 and 19, the other two showed no interaction with any of the ASKs. In the Class C, AtSFL35 could only interact with ASK14 and AtSFL77 with ASK14 and 16. However, none of the three selected proteins of the Class B showed physical interaction with ASK. Interestingly, ASK14 interacted with all of the AtSFL proteins, of which the positive interactions

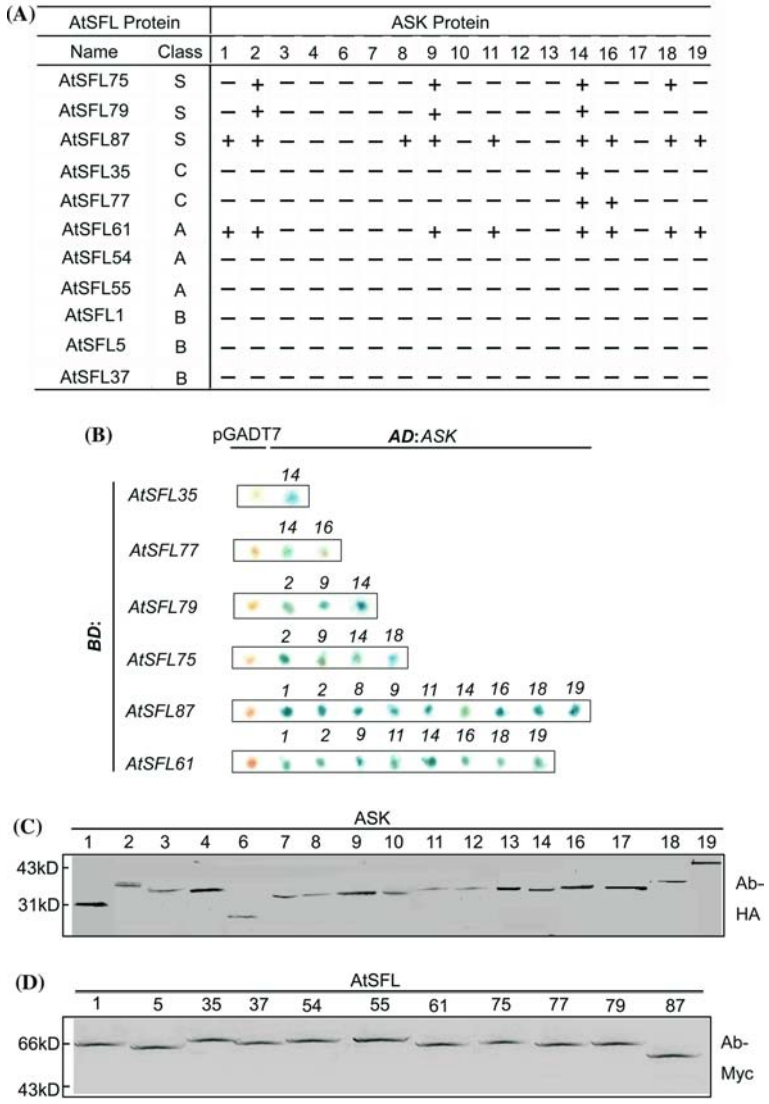

Figure 6. Physical interaction between AtSFL and ASK proteins detected by yeast two-hybrid screening. (A) A table summarizing the results of the interactions of ASK preys and AtSFL baits. Plus (+) and minus (−) indicate positive and negative interactions, respectively, (B) β-galactosidase activities in yeast, (C) Immunoblot analysis of ASK proteins fused to the Gal4 DNA-binding domain and (D) Immunoblot analysis of AtSFL proteins fused to the Gal4 DNA-activating domain.

were detected, but ASK3, 4, 6, 7, 12, 13 and 17 showed no interaction with any of the F-box proteins selected, suggesting that ASK14 has a broad interaction with these F-box proteins, whereas other ASKs are more specific for them. The expression of the 17 ASK fusion proteins were detected by Western blotting using HA monoclonal antibody (Figure 6C). Western blotting using c-Myc antibody showed that the AtSFL fusion proteins were expressed in yeast (Figure 6D). Taken together, the physical interactions of the members of AtSFL family with different ASKs suggest that they likely form diversified functional SCF complexes that may target various substrates for the ubiquitin/26S proteasome-mediated proteolysis.

Analysis of T-DNA insertion lines of AtSFL *genes*

To examine possible functions of the *AtSFL* genes, we obtained various T-DNA insertion lines of them. We searched various T-DNA insertion lines such as

SALK, Garlic, GABI-Kat and FLAG FST available from their public stock centers. In the 92 *AtSFL* genes, nearly each had several T-DNA insertion lines (see Table S4). So far, we obtained a total of 61 T-DNA insertion lines (32 SALK and 29 GARLIC lines) corresponding to 40 individual genes (see Table S4). To confirm the insertions in these lines, 20 plants from each line were planted at normal conditions and verified for their insertion using a gene-specific primer and a T-DNA specific primer. Among the 61 lines, 50 had the specific amplified insertions (data not shown) and then these PCR products were sequenced to verify the insertions. Figure S5 shows the T-DNA insertion sites of the 50 lines, corresponding to 37 individual genes.

To further examine phenotypes of the 50 verified insertion lines, their seed germination, rosette growth, inflorescence emergence, flower development, silique ripening were compared to the wild type during a complete life cycle. Surprisingly, no obvious phenotypic differences were found between most of them and wild type Col-0 plants, suggesting that an extensive functional redundancy might exist for these *AtSFL* genes. However, 4 lines that are related to 3 genes (*AtSFL24*, *AtSFL61* and *AtSFL70*) produced no homozygous progeny, suggesting that inactivation of these genes likely leads to embryonic lethality. To examine the linkage of T-DNA insertions to the mutations in these lines, the T-DNA insertion copy numbers were determined for them. One line (AtSFL24) showed a multiple copy insertion (data not shown) and was not used for further analysis. Two lines corresponding to *AtSFL61* and *AtSFL70* separately both had a single copy of T-DNA insertion (Figure S6A) and progeny analysis showed that the insertions were linked to the mutations (Figure S6B, S6C), indicating that the mutations were likely derived from T-DNA insertions.

To examine this possibility further, their siliques were dissected to observe the phenotype of ovules (Figure 7A). Normally, the developmental stages of embryo are synchronous and the size of the ovules is identical in the same silique of wild type. However, in the *atsfl70*/+ mutant, the mature siliques contained normal, green, fertilized seeds and aborted ovules at a ratio approximate to 1:1, suggesting a female gametophytic lethality. While in the *atsfl61*/+ mutant, the size of ovules is not even, indicating that development of partial embryos

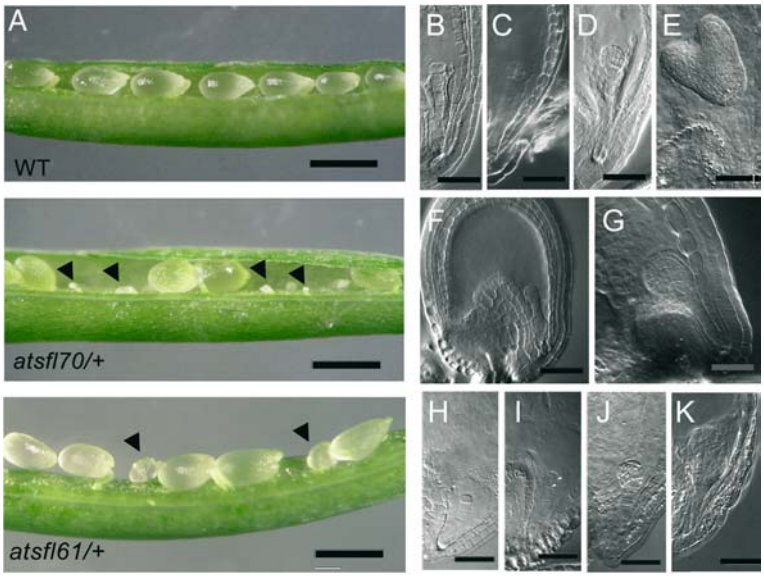

Figure 7. The embryonic development of two T-DNA insertion lines producing no homozygous progeny. (A) The open siliques derived from selfed progeny of wild type, *atsfl70*/+ and *atslf61*/+ plants. Black triangles indicate aborted or retarded ovules from the heterozygous plants. Scale bar: 500 μm. The embryonic developmental profile of wild type is shown in B–E (B, two-cell stage; C, octant stage; D, dermatogen stage; E, heart stage). In the *atsfl70*/+ line, the development of suspensor was arrested in a few ovules (F and G). In nearly 25% *atslf61*/+ ovules, the suspensor is short. The cell division of hypophysis, even the embryo, is irregular frequently (H–K). The white triangles denote the abnormal pattern of cell division. B, F and H, C and I, D and J, E, G and K depict the same stage. Scale bars: 50 μm.

may be arrested at different stages. To confirm the phenotypes of these lines, the differential interference contrast (DIC) microscope was applied to visualize the embryo development. In *Arabidopsis* embryogenesis, the pattern of cell division commences with an approximately 3-fold elongation of the zygote and followed by an asymmetric cell division, yielding a smaller apical and a larger basal cell (Mansfield and Briarty, 1991). Through two vertical (Figure 7B) and one horizontal division, the apical cell gives rise to a sphere composed of eight cells, the octant (Figure 7C). It comprises an upper and a lower tier, which will later give rise to the apical and much of the basal portion of the seedling, respectively. By contrast, the basal cell divides horizontally, and thereby produces a filamentous structure, the suspensor. In the octant stage embryo, a single round of tangential division separates an outer layer of eight epidermal precursor (or protoderm) cells from eight inner cells (Figure 7D). It was referred to dermatogen stage. Then the succeeding oriented cell division of the two tiers with apical-basal-polarity generates globular, heart (Figure 7E), torpedo then bent cotyledon embryos sequentially. Although nearly 50% ovules of *atsfl70*/+ could not survive beyond the zygotic division, a few of abnormal embryos, which exhibit retarded development of suspensor, were also found in the heterozygote (Figure 7F, G). In nearly 25% ovules of *atsfl61*/+, the suspensors were short. Frequently, the cell division of the hypophysis, even the embryo, was irregular (Figure 7H–K). These results indicated that two genes are likely involved in the development of embryo or female gametophyte. The product of *AtSFL61* was detected in the flower buds and siliques (see Figure 4), whereas the expression of *AtSFL70* was weakly detected in the tested tissues (see Figure 4), suggesting that its expression level was either low or spatial–temporally restricted.

Discussion

The AtSFL proteins may be involved in the ubiquitin/26S proteolysis pathway as a subunit of SCF complex

Previous studies have shown that most F-box proteins are associated with the ubiquitin/26S proteasome-mediated protein degradation, as a component of various SCF complexes (Hershko and Ciechanover, 1998). We found that 6 of *AtSFLs* are able to physically interact with ASK proteins by the yeast two-hybrid analysis, supporting that these proteins act as components of the SCF complexes. In particular, ASK2, 9 and 14, especially ASK14, interact with most AtSFL proteins examined. By contrast, most other known *Arabidopsis* F-box proteins interact mainly with ASK1 and 2, possibly ASK11 (Risseeuw *et al.*, 2003). As for the other AtSFLs, including two from the Class A and all of the Class C proteins tested, no physical interactions were detected although it is possible that the assay method was not sensitive enough to detect a weak interaction. In *Arabidopsis*, the ASK family contains 19 members (Arabidopsis Genome Initiative, 2000; Farras *et al.*, 2001). These F-box proteins may interact with other members or may have other function rather than that in targeted degradation of polyubiquitinated protein.

In general, F-box proteins contain two regions. The F-box motif in the N-terminus interacts with Skp1 proteins, while the C-terminal region determines the substrate specificity. One or more protein–protein interacting domains are often found in their C-terminal regions. The sequence alignments showed that all the AtSFL family members share similar structure patterns with C1–C4 domains downstream the F-box motif. Thus, the Class A, B, C and S proteins may adopt a similar way in the target recognition. Now, nine F-box proteins have reported target proteins in plants (reviewed in Smalle and Vierstra, 2004). Most discovered target proteins are transcriptional factors, and the F-box proteins recognizing them often have LRR or Kelch domain at C-terminal regions. For examples, TIR1, which contains LRRs, interacts with AUX/IAA (Gray *et al.*, 2001); ZTL contains PAS/PAC domains and recognizes the TOC specifically (Somers *et al.*, 2000; Mas *et al.*, 2003); EBF1 and EBF2, also LRRs-encompassing proteins, have been identified to recruit EIN3 for degradation (Guo and Ecker, 2003; Potuschak *et al.*, 2003). SLY1, with an unknown C-terminal domain, interact with DELLA proteins (McGinnis *et al.*, 2003). All of them belong to family C in the classification system of Gagne *et al.* (2002). However, little is known about the substrate of AtSFL proteins. Recently,

Qiao *et al*. (2004b) reported that AhSLF-S_2 physically interact with S-RNases in *Antirrhinum*, indicating that the C1–C4 domains are, indeed, capable of interacting with certain substrate and the Class A, B, C and S proteins may recognize target proteins in a similar manner. Nevertheless, their targets should not be limited to S- or S-like RNases because only five T2-type S-like RNases are found in *Arabidopsis*. It will be interesting to identify the targets of AtSFLs.

In the microarray analysis, both *AtSFL40* and *53* are clustered together with the genes encoding ubiqutin/26S proteasome-mediated proteolysis-related proteins, such as subunits of COP9 complex, ubiquitin-specific protease 3 and SKP1 interacting partners and protein kinases, providing further support that the AtSFLs may function in the ubiqutin/26S proteasome pathway.

A functional diversity is associated with AtSFL proteins

The F-box-proteins are known to play important roles in yeast, animal and plant (Hershko and Ciechanover, 1998). In *Arabidopsis*, F-box proteins play critical roles in various aspects including embryogenesis, hormonal responses, floral development, photomorphogenesis, circadian rhythms, senescence, and pathogen invasion (Smalle and Vierstra, 2004). Except *AtSFL31* (*SON1*) (Kim and Delaney, 2002), no known mutations have been detected so far for *AtSFL* genes during previous screens. It is possible that these genes are functionally redundant or their mutations are lethal or conditional. The first two possibilities are supported by the initial analyses of T-DNA insertion lines of 37 *AtSFL* genes. They showed no obviously abnormal morphological phenotypes except none homozygous plants were detected in two of them. Observation of embryogenesis indicates that the two genes may be involved in the development of embryo or female gametophyte, but their detailed functions require further investigation. Functional redundancy is also found in several other F-box proteins such as EBF1/EBF2 (Guo and Ecker, 2003). The third possibility could be tested for these lines under specific conditions like biotic and/or abiotic stresses.

AtSFL36, 37 and *40* displayed a similar expression pattern to the F-box gene *ZTL*, suggesting that they may function in the regulation of circadian rhythm, although they share no sequence homology to *ZTL* apart from the F-box motif. The high correlation ratio of *AtSFL40* and *53* with other F-box genes other than *AtSFLs* also indicates that several F-box genes function together in a particular biological process. Thus, the functional redundancy may even exist among some non-homologous F-box genes.

The AtSFL proteins are plant specific and the SLF proteins likely share a common origin

Initial analysis of several SLF proteins showed that they appear to be plant-specific and no apparent homolog was found in yeast, *Drosophila* and human (Lai *et al*., 2002). Recently, Gutiérrez *et al*. (2004) reported 3848 plant-specific proteins, which are likely to play important roles in processes that are unique and of significance to plants (http://genomebiology.com/2004/5/8/R53). We found there are 16 AtSFLs (AtSFL2, 3, 5, 16, 40, 42, 43, 44, 52, 59, 71, 72, 73, 75, 79 and 81) in the 115 plant-specific F-box proteins described by Gutiérrez *et al*. (2004). In these proteins, 4 are AtSFLs of the Class S (AtSFL40, 75, 79 and 81), 5 the Class A (AtSFL2, 3, 43, 44, 71, 73), 1 the Class B (AtSFL5) and 3 the Class C (AtSFL 15, 42, 72), covering all the classes of AtSFLs. It is clear that 83% (76/92) of AtSFLs are not found in their database since the list of putative plant-specific proteins are not exhaustive for several reasons, such as rather conservative criteria used, ESTs comprising the largest pool of sequence data for many plant species. Together, these results strongly support that AtSFL proteins only occur in the plant lineage.

The phylogenetic analyses suggest that all of the SLF proteins from *Antirrhinum, Prunus and Petunia*, and several SFLs from the non-RNase based systems belong to a monophyletic group (the Class S), implying that the common ancestor of many dicots including *Arabidopsis* possessed S-RNase/F-box-based GSI (gametophytic self-incompatibility). This result provides further support on the monophyletic origin of the S-RNases (Xue *et al*., 1996; Igic and Kohn, 2001). Nevertheless, all of the SFBs (*S*-haplotype–specific F-box proteins) from the *Prunus* species form a sub-lineage distinct from the functionally known SLF from *Antirrhinum* and *Petunia* and other *S*-locus linked F-box proteins, suggesting that they

are likely diverged earlier during the S-RNase-based SI evolution.

Although it is unclear how ancestral *SLF* and *S-RNase* genes were combined together to function during early period of SI evolution, it is clear that coevolution happened after they were recruited into the SI system. The tree topology was not significantly different between SLFs and S-RNases (Xue *et al*., 1996; Igic and Kohn, 2001; Qiao *et al*., 2004b), suggesting a tight linkage of male and female SI determinants during the evolutionary course of these haplotypes. Similar coevolution also happened in *SRK* and *SP11/SCR* genes controlling sporophytic SI in *Brassica* (Sato *et al*., 2002). It is no doubt that in-depth structural analysis of *S*-haplotypes among *S*-RNase-based SI species is expected to shed light on the origin and evolution of *SLF* genes.

Acknowledgements

We thank ABRC (Arabidopsis Biological Resource Center) and TMRI (Torrey Mesa Research Institute) for providing T-DNA insertion lines. The work was supported in part by Chinese Academy of Sciences, the Ministry of Science and Technology of China (2003CB114306) and the National Natural Science Foundation of China (30221002) to Y.X. and a grant from the National Science Foundation of USA (NSF 2010 MCB-0115870) to X.W.D.

References

Alabadi, D., Yanovsky, M.J., Mas, P., Harmer, S.L. and Kay, S.A. 2002. Critical role for CCA1 and LHY in maintaining circadian rhythmicity in *Arabidopsis*. Curr. Biol. 12: 757–761.

Arabidopsis Genome Initiative 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408: 796–815.

Bai, C., Sen, P., Hofmann, K., Ma, L., Goebl, M., Harper, J.W. and Elledge, S.J. 1996. SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. Cell 86: 263–274.

Bariola, P.A., Howard, C.J., Taylor, C.B., Verburg, M.T., Jaglan, V.D. and Green, P.J. 1994. The *Arabidopsis* ribonuclease gene *RNS1* is tightly controlled in response to phosphate limitation. Plant J. 6: 673–685.

Bechtold, N. and Pelletier, G. 1998. *In planta Agrobacterium*-mediated transformation of adult *Arabidopsis thaliana* plants by vacuum infiltration. Methods Mol. Biol. 82: 259–266.

Chandler, J.S., McArdle, B. and Callis, J. 1997. AtUBP3 and AtUBP4 are two closely related *Arabidopsis thaliana* ubiquitin-specific proteases present in the nucleus. Mol. Gen. Genet. 255: 302–310.

Deshaies, R.J. 1999. SCF and Cullin/RING H2-based ubiquitin ligases. Annu. Rev. Cell Dev. Biol. 15: 435–467.

Entani, T., Iwano, M., Shiba, H., Che, F.S., Isogai, A. and Takayama, S. 2003. Comparative analysis of the self-incompatibility (*S*-) locus region of *Prunus mume*: identification of a pollen-expressed F-box gene with allelic diversity. Genes Cells 8: 203–213.

Farras, R., Ferrando, A., Jasik, J., Kleinow, T., Okresz, L., Tiburcio, A., Salchert, K., del Pozo, C., Schell, J. and Koncz, C. 2001. SKP1-SnRK protein kinase interactions mediate proteasomal binding of a plant SCF ubiquitin ligase. EMBO J. 20: 2742–2756.

Gagne, J.M., Downes, B.P., Shiu, S.H., Durski, A.M. and Vierstra, R.D. 2002. The F-box subunit of the SCF E3 complex is encoded by a diverse superfamily of genes in *Arabidopsis*. Proc. Natl. Acad. Sci. USA 99: 11519–11524.

Gray, W.M., Kepinski, S., Rouse, D., Leyser, O. and Estelle, M. 2001. Auxin regulates SCF^{TIR1}-dependent degradation of AUX/IAA proteins. Nature 414: 271–276.

Guo, H. and Ecker, J.R. 2003. Plant responses to ethylene gas are mediated by $SCF^{EBF1/EBF2}$-dependent proteolysis of EIN3 transcription factor. Cell 115: 667–677.

Gutiérrez, R.A., Green, P.J., Keegstra, K. and Ohlrogge, J.B. 2004. Phylogenetic profiling of the *Arabidopsis thaliana* proteome: what proteins distinguish plants from other organisms? Genome Biol. 5: R53.

Harlow, E. and Lane, D. 1988. Antibodies: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, p. 349.

Herr, J.M. Jr. 1971. A new clearing squash technique for the study of ovule development in angiosperms. Am. J. Bot. 58: 894–790.

Hershko, A. and Ciechanover, A. 1998. The ubiquitin system. Annu. Rev. Biochem. 67: 425–479.

Igic, B. and Kohn, J.R. 2001. Evolutionary relationships among self-incompatibility RNases. Proc. Natl. Acad. Sci. USA 98: 13167–13171.

Kim, H.S. and Delaney, T.P. 2002. *Arabidopsis* SON1 is an F-box protein that regulates a novel induced defense response independent of both salicylic acid and systemic acquired resistance. Plant Cell 14: 1469–1482.

Kipreos, E.T. and Pagano, M. 2000. The F-box protein family. Genome Biol. 1: reviews 3002.1–3002.7.

Kuroda, H., Takahashi, N., Shimada, H., Seki, M., Shinozaki, K. and Matsui, M. 2002. Classification and expression analysis of *Arabidopsis* F-box-containing protein genes. Plant Cell Physiol. 43: 1073–1085.

Lai, Z., Ma, W., Han, B., Liang, L., Zhang, Y., Hong, G. and Xue, Y. 2002. An F-box gene linked to the self-incompatibility (*S*) locus of *Antirrhinum* is expressed specifically in pollen and tapetum. Plant Mol. Biol. 50: 29–42.

Liu, Y., Mitsukawa, N., Oosumi, T. and Whittier, R.F. 1995. Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. Plant J. 8: 457–463.

Mansfield, S.G. and Briarty, L.G. 1991. Early embryogenesis in *Arabidopsis thaliana*: 1. The developing embryo. Can. J. Bot. 69: 461–476.

Mas, P., Kim, W.Y., Somers, D.E. and Kay, S.A. 2003. Targeted degradation of TOC1 by ZTL modulates circadian function in *Arabidopsis thaliana*. Nature 26: 567–570.

McGinnis, K.M., Thomas, S.G., Soule, J.D., Strader, L.C., Zale, J.M., Sun, T.P. and Steber, C.M. 2003. The *Arabidopsis SLEEPY1* gene encodes a putative F-box subunit of an SCF E3 ubiquitin ligase. Plant Cell 15: 1120–1130.

Patton, E.E., Willems, A.R. and Tyers, M. 1998. Combinatorial control ubiquitin-dependent proteolysis: don't Skp the F-box hypothesis. Trends Genet. 14: 236–243.

Potuschak, T., Lechner, E., Parmentier, Y., Yanagisawa, S., Grava, S., Koncz, C. and Genschik, P. 2003. EIN3-dependent regulation of plant ethylene hormone signaling by two *Arabidopsis* F- box proteins: EBF1 and EBF2. Cell 115: 679–689.

Printen, J.A. and Sprague, G.F., Jr. 1994 Protein interactions in yeast pheromoneresponse pathway: Ste5p interacts with all members of the MAP kinase cascade. Genetics 138: 609–619.

Qiao, H., Wang, H., Zhao, L., Zhou, J., Huang, J., Zhang, Y. and Xue, Y. 2004b. The F-Box protein AhSLF-S_2 physically interacts with S-RNases that may be inhibited by the ubiquitin/26S proteasome pathway of protein degradation during compatible pollination in *Antirrhinum*. Plant Cell 16: 582–595.

Qiao, H., Wang, F., Zhao, L., Zhou, J., Lai, Z., Zhang, Y., Robbins, T.P. and Xue, Y. 2004a. The F-Box protein AhSLF-S_2 controls the pollen function of S-RNase-based self-incompatibility. Plant Cell 16: 2307–2332.

Risseeuw, E.P., Daskalchuk, T.E., Banks, T.W., Liu, E., Cotelesage, S.L., Hellmann, H., Estelle, M., Somer, D.E. and Crosby, W.L. 2003. Protein interaction analysis of SCF ubiquitin E3 ligase subunits from *Arabidopsis*. Plant J. 34: 753–767.

Sakata, K., Nagamura, Y., Numa, H., Antonio, B.A., Nagasaki, H., Idonuma, A., Watanabe, W., Shimizu, Y., Horiuchi, I., Matsumoto, T., Sasaki, T. and Higo, K. 2002. RiceGAAS: an automated annotation system and database for rice genome sequence. Nucleic Acids Res. 30: 98–102.

Sato, K., Nishio, T., Kimura, R., Kusaba, M., Suzuki, T., Hatakeyama, K., Ockendon, D.J. and Satta, Y. 2002. Coevolution of the *S*-locus genes *SRK*, *SLG* and *SP11*/*SCR* in *Brassica oleracea* and *B. rapa*. Genetics 162: 931–940.

Serino, G., Tsuge, T., Kwok, S., Matsui, M., Wei, N. and Deng, X.W. 1999. *Arabidopsis cop8* and *fus4* mutations define the same gene that encodes subunit 4 of the COP9 signalosome. Plant Cell 11: 1967–1980.

Sijacic, P., Wang, X., Skirpan, A.L., Wang, Y., Dowd, P.E., McCubbin, A.G., Huang, S. and Kao, T.H. 2004. Identification of the pollen determinant of S-RNase-mediated self-incompatibility. Nature 429: 302–305.

Smalle, J. and Vierstra, R.D. 2004. The ubiquitin 26S proteasome proteolytic pathway. Annu. Rev. Plant Biol. 55: 555–590.

Somers, D.E., Schultz, T.F., Milnamow, M. and Kay, S.A. 2000. *ZEITLUPE* encodes a novel clock-associated PAS protein from *Arabidopsis*. Cell 101: 319–329.

Sullivan, J.A., Shirasu, K. and Deng, X.W. 2003. The diverse roles of ubiquitin and the 26s proteasome in the life of plants. Nat. Rev. Genet. 4: 948–958.

Takahashi, N., Kuroda, H., Kuromori, T., Hirayama, T., Seki, M., Shinozaki, K., Shimada, H. and Matsui, M. 2004. Expression and interaction analysis of *Arabidopsis Skp1*-related genes. Plant Cell Physiol. 45: 83–91.

Taylor, C.B., Bariola, P.A., del Cardayré, S.B., Raines, R.T. and Green, P.J. 1993. RNS2: a senescence-associated RNase of *Arabidopsis* that diverged from the S-RNases before speciation. Proc. Natl. Acad. Sci. USA 90: 5118–5112.

Ushijima, K., Sassa, H., Dandekar, A.M., Gradziel, T.M., Tao, R. and Hirano, H. 2003. Structural and transcriptional analysis of the self-incompatibility locus of almond: identification of a pollen-expressed F-box gene with haplotype-specific polymorphism. Plant Cell 15: 771–781.

Ushijima, K., Yamane, H., Watari, A., Kakehi, E., Ikeda, K., Hauck, N.R., Iezzoni, A.F. and Tao, R. 2004. The *S* haplotype-specific F-box protein gene, *SFB*, is defective in self-compatible haplotypes of *Prunus avium* and *P. mume*. Plant J. 39: 573–586.

Wang, H., Huang, J., Lai, Z. and Xue, Y. 2002. F-box proteins in flowering plants. Chin. Sci. Bull. 47: 1497–1501.

Wang, Y., Tsukamoto, T., Yi, K.W., Wang, X., Huang, S., McCubbin, A.G. and Kao, T.H. 2004. Chromosome walking in the *Putunia inflate* self-incompatibility (*S*-) locus and gene identification in a 881-kb contig containing *S_2-RNase*. Plant Mol. Biol. 54: 727–742.

Wang, Y., Wang, X., McCubbin, A.G. and Kao, T.H. 2003. Genetic mapping and molecular characterization of the self-incompatibility (*S*) locus in *Petunia inflata*. Plant Mol. Biol. 53: 565–580.

Weigel, D. and Galzebrook, J. 2002. *Arabidopsis*: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 243–248.

Winston, J.T., Koepp, D.M., Zhu, C., Elledge, S.J. and Harper, J.W. 1999. A family of mammalian F-box proteins. Curr. Biol. 9: 1180–1182.

Xue, Y., Carpenter, R., Dickinson, H.G. and Coen, E.S. 1996. Origin of allelic diversity in *Antirrhinum S* locus RNases. Plant Cell 8: 805–814.

Yamane, H., Ikeda, K., Ushijima, K., Sassa, H. and Tao, R. 2003. A pollen-expressed gene for a novel protein with an F-box motif that is very tightly linked to a gene for S-RNase in two species of cherry, *Prunus cerasus* and *P. avium*. Plant Cell Physiol. 44: 764–769.

Zhou, J., Wang, F., Ma, W., Zhang, Y., Han, B. and Xue, Y. 2003. Structural and transcriptional analysis of *S*-Locus F-box genes in *Antirrhinum*. Sex. Plant Reprod. 16: 165–177.