Sex Plant Reprod (2004) 16:235–243
DOI 10.1007/s00497-003-0200-x

ORIGINAL ARTICLE

Kazuo Ikeda · Boris Igic · Koichiro Ushijima · Hisayo Yamane · Nathanael R. Hauck · Ryohei Nakano · Hidenori Sassa · Amy F. Iezzoni · Joshua R. Kohn · Ryutaro Tao

Primary structural features of the *S* haplotype-specific F-box protein, SFB, in *Prunus*

Received: 28 August 2003 / Accepted: 14 November 2003 / Published online: 12 December 2003

Abstract The gene *SFB* encodes an F-box protein that has appropriate *S*-haplotype-specific variation to be the pollen determinant in the S-RNase-based gametophytic self-incompatibility (GSI) reaction in *Prunus* (Rosaceae). To further characterize *Prunus SFB*, we cloned and sequenced four additional alleles from sweet cherry (*P. avium*), SFB^1, SFB^2, SFB^4, and SFB^5. These four alleles showed haplotype-specific sequence diversity similar to the other nine *SFB* alleles that have been cloned. In an amino acid alignment of *Prunus* SFBs, including the four newly cloned alleles, 121 out of the 384 sites were conserved and an additional 65 sites had only conservative replacements. Amino acid identity among the SFBs ranged from 66.0% to 82.5%. Based on normed variability indices (NVI), 34 of the non-conserved sites were considered to be highly variable. Most of the variable sites were located at the C-terminal region. A window-averaged plot of NVI indicated that there were two variable and two hypervariable regions. These variable and hypervariable regions appeared to be hydrophilic or at least not strongly hydrophobic, which suggests that these regions may be exposed on the surface and function in the allele specificity of the GSI reaction. Evidence of positive selection was detected using maximum likelihood methods with sites under positive selection concentrated in the variable and hypervariable regions.

Keywords Pollen-*S* · Rosaceae · *SFB* · S-RNase · Self-incompatibility

K. Ikeda and B. Igic contributed equally to this paper

Nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ nucleotide sequence databases under the accession numbers AB111518, AB111519, AB111520, and AB111521, for SFB^1, SFB^2, SFB^5, and SFB^4, respectively

K. Ikeda · K. Ushijima · H. Yamane · R. Tao (✉)
Laboratory of Pomology, Graduate School of Agriculture,
Kyoto University, 606–8502 Kyoto, Japan
e-mail: rtao@kais.kyoto-u.ac.jp
Tel.: +81-75-7536053
Fax: +81-75-7536497

B. Igic · J. R. Kohn
Section of Ecology, Behavior, Evolution,
Division of Biological Sciences,
University of California San Diego,
La Jolla, CA 92093, USA

N. R. Hauck · A. F. Iezzoni
Department of Horticulture,
Michigan State University,
East Lansing, MI 48824, USA

R. Nakano
Laboratory of Postharvest Agriculture, Faculty of Agriculture,
Okayama University,
700–8530 Okayama, Japan

H. Sassa
Kihara Institute for Biological Research,
Yokohama City University,
244-0813 Yokohama, Japan

Introduction

Self-incompatibility (SI) is a widespread mechanism in flowering plants that prevents self-fertilization and the deleterious effects of inbreeding (de Nettancourt 2001). In the most widespread form of SI, gametophytic SI (GSI), the haploid pollen is rejected if it carries an *S*-allele that matches either *S*-allele in the diploid pollen recipient. Several lines of evidence indicate that different *S*-locus genes confer pollen and pistil specificity (McCubbin and Kao 2000).

The pistil determinant of the SI reaction is known to be an S-ribonuclease (S-RNase) in three plant families, the Rosaceae (Sassa et al. 1992), Solanaceae (Anderson et al. 1986; McClure et al. 1989) and Scrophulariaceae (Xue et al. 1996). On the other hand, the pollen determinant has eluded characterization for almost two decades (McCubbin and Kao 2000). Recently a good candidate for the pollen-*S*, an *S* haplotype-specific F-box gene (*SFB*) was identified by sequencing the *S*-locus region of several *Prunus* species (Rosaceae), such as almond (*P. dulcis*; Ushijima et al. 2003), Japanese apricot (*P. mume*; Entani

et al. 2003; Yamane et al. 2003b), and both sweet and sour cherry (*P. avium* and *P. cerasus*; Yamane et al. 2003a). Features of *SFB*, such as pollen-specific expression and a high level of allelic polymorphism, are appropriate characteristics for the pollen determinant in the S-RNase-based GSI system (Entani et al. 2003; Ushijima et al. 2003, Yamane et al. 2003a, 2003b). In addition to these characteristics, *S-RNase* and *SFB* may be inherited as a unit because *SFB* is present in the *S*-locus region where DNA sequence is highly divergent and thus recombination is believed to be suppressed (Entani et al. 2003; Ushijima et al. 2001, 2003). The physical distance between *SFB* and *S-RNase* varies from 380 bp to 30 kb depending on the *S*-haplotype (Entani et al. 2003; Ushijima et al. 2003; Yamane et al. 2003a).

The presence of F-box genes physically linked to the *S-RNase* and exhibiting pollen-specific expression was first reported in *Antirrhinum* (Lai et al. 2002). However, none of the *S*-locus F-box genes (termed *SLF*) identified in *Antirrhinum* exhibit the level of polymorphism that is predicted for the pollen-*S* gene (Zhou et al. 2003). Subsequently, polymorphic and non-polymorphic *S*-locus F-box genes physically close to the *S-RNase* were simultaneously identified by Ushijima et al. (2003) and Entani et al. (2003) in *P. dulcis* and *P. mume*, respectively. Entani et al. (2003) named the F-box gene that exhibited allele polymorphism *SLF*, and the three F-box genes that did not exhibit allele polymorphism, *SLFL*. Ushijima et al. (2003) also used nomenclature that differentiated between the polymorphic and non-polymorphic F-box genes, *SFB* and *SLF*, respectively. Yamane et al. (2003b) also identified the *S*-locus F-box gene of *P. mume* that exhibited allele polymorphism and named this gene *SFB*. The DNA sequences of the two *SFB* alleles of *P. mume* reported by Yamane et al. (2003b) were identical to the *SLF* alleles reported by Entani et al. (2003). It is likely that the *Prunus SFB* and *SLF* (Entani et al. 2003) genes are ortholgous based on their close phylogenetic relationship (Zhou et al. 2003). However, because *AhSLF* is likely not an orthologue of *SFB*, and because *SFB* is the only highly polymorphic *S*-locus F-box gene that has been identified to date in *Prunus*, we use *SFB* for the *Prunus* candidate pollen-*S* gene to recognize these differences.

To date, two main models, an inhibitor model and a receptor model, and several modifications of these models, have been proposed for pollen-pistil recognition in S-RNase-based GSI (Thompson and Kirch 1992; Wang et al. 2003). According to the inhibitor model, a pollen-*S* protein is an intracellular S-RNase inactivator and inhibits all the S-RNases apart from the cognate S-RNase. In the receptor model, the pollen-*S* protein is a membrane receptor that recognizes the cognate S-RNase. Currently, the inhibitor model is favored. Luu et al. (2000) described the *S*-haplotype-independent uptake of S-RNase by pollen tubes and Golz et al. (2001) described the essential role of products of pollen-*S* genes in pollen tube growth.

Recent identification of *SFB* has raised the possibility that S-RNase stability rather than its activity is controlled by the SI reaction. F-box proteins, of which more than 500 have been identified in *Arabidopsis thaliana* (Gagne et al. 2002; Kuroda et al. 2002), are involved in the ubiquitin/26S proteasome pathway of protein degradation (Deshaies 1999; Craig and Tyers 1999). The F-box protein associates with Skp1 through the F-box motif to form the SCF complex and binds specifically to the target protein that is to be polyubiquitinated (Deshaies 1999; Craig and Tyers 1999). The polyubiquitinated protein is recognized and degraded by the 26S proteasome. Therefore, SFB might function in the polyubiquitination of all non-self S-RNases leading to their degradation, and may also specifically interact with its self S-RNase to protect it from degradation, leading to the arrest of self-pollen tubes (Ushijima et al. 2003; Yamane et al. 2003a).

Comparisons of different *S-RNase* alleles in rosaceous and solanaceous species revealed a single variable (RHV) and two hypervariable (HVa and HVb) regions, respectively (Ioerger et al. 1991; Kheyr-Pour et al. 1990; Ushijima et al. 1998). The hypervariable regions appear to be under positive selection (Clark and Kao 1991, Ishimizu et al. 1998) and are believed to be important for allele-specific recognition and triggering of the GSI response. Since we anticipate that S-RNase and SFB interact in an allele-specific manner, we expect the presence of variable regions, which would interact with the variable regions of the S-RNases, to be under positive selection in SFB. Although systematic and comprehensive sequence comparisons of *Prunus SFB* with a reasonably large number of alleles is needed to determine the primary structural features of *Prunus* SFBs, at least two variable regions were found in *Prunus* SFBs (Ushijima et al. 2003; Yamane et al. 2003a, 2003b).

Our objectives were: (1) to clone the *SFB* genes that are linked to the *S-RNase* gene in four additional sweet cherry *S*-haplotypes (S^1, S^2, S^4 and S^5); (2) to compare the structural features and amino acid similarities of the four additional *SFB* alleles with the nine other alleles previously cloned from sweet cherry, almond, and Japanese apricot; and (3) to look for evidence of regions under positive selection among the known *SFB* genes from *Prunus*.

Materials and methods

Plant materials

Ten sweet cherry (*P. avium*) cultivars, 'Early Rivers' (S^1 S^2), 'Van' (S^1 S^3), 'Rainier' (S^1 S^4), 'Seneca' (S^1 S^5), 'Rockport Bigarreau' (S^1 S^6), 'Velvet' (S^2 S^3), 'Bing' (S^3 S^4), 'Emperor Francis' (S^3 S^4), 'Hedelfingen' (S^3 S^5), and 'Satonishiki' (S^3 S^6), and a selection 'NY54' (S^2 S^6) were used in this study. Young leaves were collected in the spring and used for total DNA extraction, as described elsewhere (Tao et al. 1999).

PCR-cloning of partial sequences for SFB^1, SFB^2, SFB^4, and SFB^5

Total DNA from 'Van' (S^1 S^3), 'Early Rivers' (S^1 S^2), 'Bing' (S^3 S^4), and 'Hedelfingen' (S^3 S^5) was PCR-amplified with the *SFB* gene-specific primer set, SFB-C1F and SFB-C5R (Table 1), as

Table 1 Sequences[a] and amino acid locations for the oligonucleotide primers used in this study

Primer	Sequence	Location[b]
SFB-C1F[c]	5′-RTTCGRTTTCTDTTTACRTG-3′	30–37
SFB-C2F	5′-CCWATACAMATATGGAACCC-3′	129–136
SFB-C2R[d]	5′-GGGTTCCATATKTGTATWGG-3′	129–136
FB3R[c]	5′-CCATCGTTTTTCTTGCAGAACCC-3′	287–294
SFB-C5F	5′-TACCAYWTMATTGAGAAAGGTCC-3′	221–228
SFB-C5R[c]	5′-RCTGMATRWRGGACCTTTCTC-3′	225–232

[a] Primer sequences were designed from the alignment of four *Prunus dulcis* SFB sequences (*SFB*^{a}, *SFB*^{b}, *SFB*^{c}, *SFB*^{d})
[b] Numbers refer to the location of the primer on the amino acid alignment shown in Fig. 1
[c] From Yamane et al. (2003a)
[d] From Ushijima et al. (2003)

described by Yamane et al. (2003a). PCR-amplified products were cloned into the TA cloning vector (pGEM-T Easy Vector System, Promega, Madison, Wis.) and sequenced using an ABI PRISM 310 Genetic Analyzer (Applied Biosystems, Foster City, Calif.).

DNA blot analysis

A 5-μg sample of total DNA from young leaves of the sweet cherry cultivars was digested with *Dra*I, run on a 0.9% agarose gel, transferred to a nylon membrane (Hybond-N, Amersham, Piscataway, N.J.), and probed with an *SFB* gene-specific probe under high stringency conditions to avoid cross hybridization with DNA sequences for F-box proteins other than SFB. The *SFB* gene-specific probes were obtained by PCR-labeling of partial sequences for *SFB*^{1}, *SFB*^{2}, *SFB* ^{4} and *SFB*^{5} with DIG-dUTP using either the SFB-C1F and SFB-C5R primers (Table 1) for *SFB*^{1}, *SFB*^{2}, and *SFB*^{5}, or SFB-C2F and SFB-C5R primers (Table 1) for *SFB*^{4}. The alternative primers were used for *SFB*^{4} because of the presence of a *Dra*I restriction site between the SFB-C1F and SFB-C2F sequences in the partial sequence of *SFB*^{4}. Hybridization signal was detected using the anti-DIG-alkaline phosphatase conjugate and the chemiluminescent substrate CSPD (Roche Diagnostics, Mannheim, Germany) after high stringency washes, as described previously (Yamane et al. 2003a).

Construction and screening of cosmid and fosmid libraries

Cosmid libraries were constructed from 'Emperor Francis' (S^3 S^4) and 'NY54' (S^2 S^6), using the pWEB::TNC cosmid cloning kit (Epicentre, Madison, Wis.), while a fosmid library was constructed from 'Seneca' (S^1 S^5) using the CopyControl fosmid library production kit (Epicentre). The genomic libraries from 'Seneca', 'Emperor Francis', and 'NY54' were screened for *SFB*^{1} and *SFB*^{5}, *SFB*^{3} and *SFB*^{4}, and *SFB*^{2} and *SFB*^{6}, respectively, with DIG-dUTP labeled *SFB* gene-specific probes, as previously described (Ushijima et al. 2001). The DIG-labeled *SFB* gene-specific probes were obtained by PCR-labeling of partial sequences for *SFBs* that were cloned in this study and *SFB*^{3} and *SFB*^{6} of 'Satonishiki' (Yamane et al. 2003a) with SFB-C1F and SFB-C5R primers (Table 1). The clones were sequenced to determine the entire coding sequence of the *SFBs* by primer walking using essentially SFB-C1F, SFB-C2R, SFB-C5F, and FB3R primers (Table 1) with several allele-specific primers and an ABI PRISM 310 Genetic Analyzer (Applied Biosystems).

Construction of a phylogenetic tree

DNA sequences of 13 *Prunus SFBs* (Fig. 1), including the four sweet cherry *SFB* alleles cloned in this study, were aligned using Clustal X (Thompson et al. 1997) and adjusted manually. The best fit model of nucleotide sequence evolution was first determined in Modeltest (Posada and Crandall 1998) and then used for a maximum likelihood tree search in PAUP* (Swofford 2001). The average amino acid identity among *SFB* and *S-RNase* alleles was also calculated. Since we have identified a *P. mume* S^9 allele (Tao et al. 2003) that is different from the S^9 allele found by Entani et al. (2003), we used the name S^{9b} for the S^9 allele found by Entani et al. (2003). We used SFB^{9b} but not the S^{9b}-RNase for amino acid sequence comparisons because only a partial sequence has been reported for *S*^{9b} *-RNase*.

Identification of variable sites

Normed variability indices (NVI) for each site of the aligned *SFB* sequences (Kheyr-Pour et al. 1990) were calculated. This index ranges from −1 (perfectly conserved) to +1 (every residue different), regardless of the number of residues at the site. For sites with one residue, the NVI is defined to be 0. The sites that had NVI of more than −0.25 were considered to be variable. NVI at each site was averaged with its neighbors in a sliding-window of size 11 to generate a window-averaged plot of NVI.

Generation of hydrophobicity plot

Average amino acid hydrophobicity indexes (Kyte and Doolittle 1982) at each site were calculated over the 13 SFBs. A hydrophobicity plot of the 13 SFBs was generated by averaging the value at each site with its neighbors in a sliding-window of size 11.

Identification of sites under positive selection

We used codeml implemented in PAML 3.13 (Yang 1997) for all likelihood calculations and estimates of nonsynonymous/synonymous substitution rate ratio (ω). Recent application of maximum likelihood methods and Bayes' theorem to studies of nucleotide polymorphism (see Nielsen and Yang 1998; Yang 1998) have led to the development and widespread use of methods to assess posterior probabilities that a particular codon has been under positive selection. Sites with values of ω significantly smaller than 1 indicate an excess of synonymous substitutions, and therefore are inferred to be constrained by purifying selection. Alternatively, those sites for which a significant excess of nonsynonymous substitutions is recorded (ω>1) suggest evolution under positive selection pressure. When ω is not significantly different from 1, neutrality cannot be rejected. Here, we calculate the average values of ω across *SFB* sites and present their variation across the molecule in a centered 11-codon sliding window. In addition, we calculate the posterior probabilities that each codon is under positive selection.

We analyzed the SFB data for molecular selection by comparing the maximum likelihood evaluated fit of a number of models (see review by Yang and Bielawski 2000). Our methodology and nomenclature follow those of Yang and colleagues (Yang and Bielawski 2000; Yang and Swanson 2002). First, we compared the null models that do not allow for positive selection (models M0 and M1) with successively more complex models that do allow for

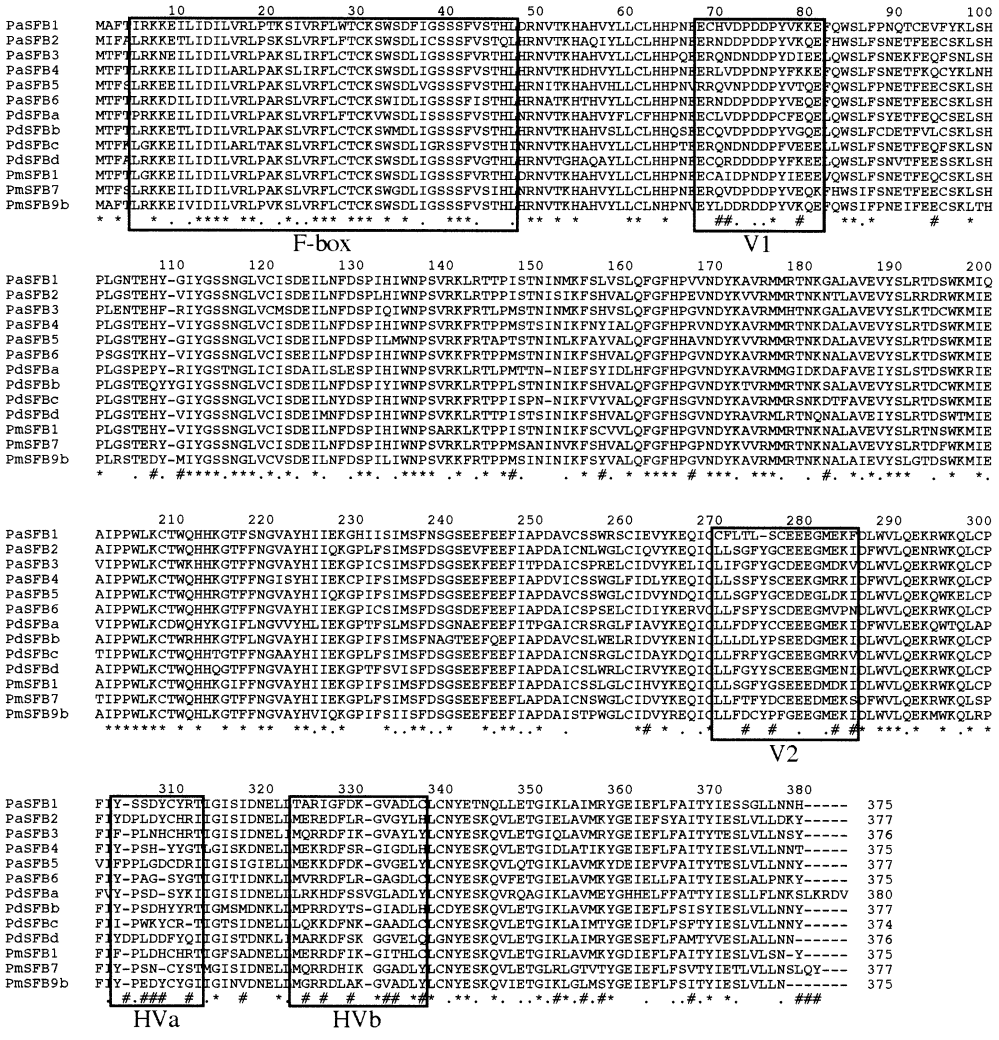

Fig. 1 Amino acid sequence alignment of 13 *Prunus* SFBs. The amino acid sequences of SFBs were aligned using Clustal X (Thompson et al. 1997). *Dashes* Gaps, *asterisks* conserved sites, *dots* conservative replacements [amino acid groups defined in Dayhoff et al. (1979): C, STAPG, MILV, HRK, NDEQ, FYW], *hash* hypervariable sites. F-box and (hyper)variable regions, V1, V2, HVa and HVb are *boxed*. Aligned SFBs: *Pa* sweet cherry (*Prunus avium*), *Pd* almond (*P. dulcis*), *Pm* Japanese apricot (*P. mume*). EMBL/GenBank/DDBJ accession numbers: *Pa-SFB*^{1} (AB111518), *Pa-SFB*^{2} (AB111519), *Pa-SFB*^{3} (AB096857), *Pa-SFB*^{4} (AB111521), *Pa-SFB*^{5} (AB111520), *Pa-SFB*^{6} (AB096858), *Pd-SFB*^{a} (AB092966), *Pd-SFB*^{b} (AB092967), *Pd-SFB*^{c} (AB079776), *Pd-SFB*^{d} (AB081648), *Pm-SFB*^{1} (AB101440), *Pm-SFB*^{7} (AB101441), *Pm-SFB*^{9b} (AB092646)

positive selection (M2 and M3). We also used a second method that involves testing another null model (M7) and a more parameter-rich model (M8). In this case, the null model allows a beta distribution of ω values across sites, but is limited to values between 0 and 1. The more complicated model (M8) adds an extra site class with a free ω ratio that may take values >1 and is estimated from the data. Finally, when positive selection is detected, one can use Bayes' Theorem to identify particular sites under positive selection. This is achieved by calculating the posterior probability that each site belongs to one of the selection classes. For example, in model M3, three different classes of codons evolving under different selective regimes (values of ω) are assumed. The first, ω_0, may be near 0, ω_1, near 1, and ω_2 greater than one. Those sites with a large posterior probability (>95%) of being in class ω_2 would be considered positively selected. If *SFB* is the pollen-part determinants of the SI reaction, exposed sites under positive selection are expected to interact with those of the S-RNase from the same haplotype.

Results

Cloning of four *SFB* alleles of sweet cherry

Partial sequences for *SFB*^{1} and *SFB*^{3}, *SFB*^{1} and *SFB*^{2}, *SFB*^{3} and *SFB*^{4}, and *SFB*^{3} and *SFB*^{5}, were cloned from 'Van', 'Early Rivers', 'Bing', and 'Hedelfingen', respectively (Fig. 1). Partial sequences for *SFB*^{3} from 'Van', 'Bing', and 'Hedelfingen' were exactly the same as the corresponding region of the previously reported *SFB*^{3} of 'Satonishiki' (Yamane et al. 2003a). The probes from partial sequences for *SFB*^{1} of 'Van', *SFB*^{2} of 'Early Rivers', *SFB*^{4} of 'Bing', and *SFB*^{5} of 'Hedelfingen' specifically hybridized to cultivars with the respective alleles when used in DNA blot analysis under high stringency conditions, confirming that the probes were from the corresponding *SFB* allele (Fig. 2). The faint signals observed in lanes 5 and 6 of the *SFB*^{2} panel (Fig. 2) were from *SFB*^{3} since the molecular weight of the signals were identical to those of signals from *SFB*^{3} (Yamane et al. 2003a). These probes were used to identify *SFB*-containing cosmid or fosmid clones. The entire coding sequences for *SFB*^{1} and *SFB*^{5}, *SFB*^{3} and *SFB*^{4}, and *SFB*^{2} and *SFB*^{6} from 'Seneca', 'Emperor Francis', and 'NY54', were determined from these clones. The sequences of the genomic clones of *SFB*^{1}, *SFB*^{2}, *SFB*^{4} and *SFB*^{5} were identical to the partial sequences used as probes and the sequences for *SFB*^{3} and *SFB*^{6} were identical to the previously reported sequences for *SFB*^{3} and *SFB*^{6} of 'Satonishiki' (Yamane et al. 2003a).

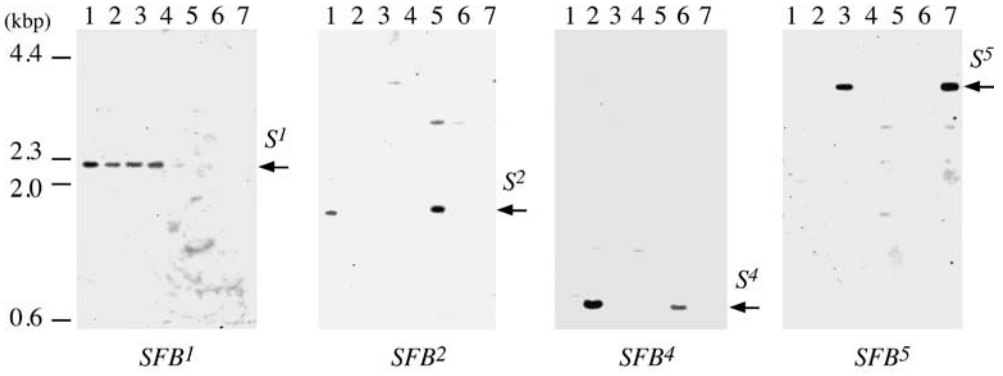

Fig. 2 Genomic DNA blot analysis for *SFBs*. Genomic DNAs from seven sweet cherry cultivars were digested with *Dra*I, separated on a 0.9% agarose gel, and transferred to a nylon membrane. The membrane was hybridized with probes for each *SFB* allele. Arrows S^1, S^2, S^3 and S^5 mark *S*-haplotype-specific signals. Lanes: *1* 'Early Rivers' (S^1S^2), *2* 'Ranier' (S^1S^4), *3* 'Seneca' (S^1S^5), *4* 'Rockport Bigarreau' (S^1S^6), *5* 'Velvet' (S^2S^3), *6* 'Bing' (S^3S^4), *7* 'Hedelfingen' (S^3S^5)

Table 2 Identities of the putative amino acid sequences of the *Prunus S* locus genes. The upper half presents amino acid sequence identities (%) between *Prunus* SFBs; the lower half between the S-RNases. S^{9b}-RNase is not listed as its published sequence is partial (AB092646). *Pa P. avium*, *Pd P. dulcis*, *Pm P. mume*. EMBL/GenBank/DDBJ accession numbers: *Pa-SFB*^{1} (AB111518), *Pa-SFB*^{2} (AB111519), *Pa-SFB*^{3} (AB096857), *Pa-SFB*^{4} (AB111521), *Pa-SFB*^{5} (AB111520), *Pa-SFB*^{6} (AB096858), *Pd-SFB*^{a} (AB092966), *Pd-SFB*^{b} (AB092967), *Pd-SFB*^{c} (AB079776), *Pd-SFB*^{d} (AB081648), *Pm-SFB*^{1} (AB101440), *Pm-SFB*^{7} (AB101441), *Pm-SFB*^{9b} (AB092645), *Pa-S*^{1}-RNase (AB028153), *Pa-S*^{2}-RNase (AB010304), *Pa-S*^{3}-RNase (AB010306), *Pa-S*^{4}-RNase (AB028154), *Pa-S*^{5}-RNase (PAV298314), *Pa-S*^{6}-RNase (AB010305), *Pd-S*^{a}-RNase (AB026836), *Pd-S*^{b}-RNase (AB011469), *Pd-S*^{c}-RNase (AB011470), *Pd-S*^{d}-RNase (AB011471), *Pm-S*^{1}-RNase (AB101438), and *Pm-S*^{7}-RNase (AB101439)

	Prunus avium						*P. dulcis*				*P. mume*		
	PaS1	PaS2	PaS3	PaS4	PaS5	PaS6	PdSa	PdSb	PdSc	PdSd	PmS1	PmS7	PmS9b
PaS1	–	76.4	76.6	77.4	75.1	75.3	66.0	77.5	75.0	76.4	78.7	76.3	75.0
PaS2	86.5	–	77.5	80.1	79.8	80.1	67.6	78.0	76.9	79	82.5	80.5	77.7
PaS3	74.4	71.9	–	77.4	77.5	79.5	67.8	77.2	78.5	76.4	82.2	77.7	76.1
PaS4	80.5	81.8	70.3	–	78.5	81.1	70.6	79.3	78.9	77.2	82.2	80.1	77.7
PaS5	72.8	75.2	69.7	72.6	–	76.4	66.3	76.2	77.2	75.6	79.8	77.0	76.7
PaS6	80.0	85.9	72.9	79.3	72.8	–	68.2	78.2	76.5	77.5	80.6	79.4	78.5
PdSa	56.3	55.8	57.4	55.6	57.8	54.6	–	67.9	68.9	67.4	69.7	70.3	67.3
PdSb	82.4	83.9	73.7	75.1	75.6	79.0	57.8	–	75.6	76.2	80.6	78.9	76.9
PdSc	80.5	81.8	73.5	77.8	73.3	78.3	55.6	77.1	–	75.6	80.3	80.1	74.5
PdSd	74.8	73.8	71.3	71.1	73.1	71.3	58.5	74.6	71.3	–	79.8	76.3	75.8
PmS1	76.0	80.8	73.1	73.6	73.0	74.7	57.8	74.1	72.7	66.2	–	81.0	77.4
PmS7	73.0	75.0	67.8	76.9	69.2	74.5	56.6	71.7	68.0	72.1	74.9	–	79.4

Comparisons of SFB and S-RNase sequences

Amino acid sequences of six sweet cherry *SFB* alleles, *SFB*^{1}, *SFB*^{2}, *SFB*^{3}, *SFB*^{4}, *SFB*^{5} and *SFB*^{6}, were aligned with four almond alleles, *SFB*^{a}, *SFB*^{b}, *SFB*^{c}, and *SFB*^{d}, and three Japanese apricot alleles, *SFB*^{1}, *SFB*^{7}, and *SFB*^{9b} (Fig. 1). Amino acid identity among the 13 *SFB* alleles ranged from 66.0% to 82.5% while identity among the linked *S-RNase* alleles, except for the S^{9b} *-RNase* allele, ranged from 54.6% to 86.5% (Table 2). The S^{9b} *-RNase* allele was not incorporated into this analysis because only a partial sequence has been reported. The average amino acid identity among *SFB* alleles was 76.5%, while the average among *S-RNase* alleles was 72.0%. *SFB*^{a} always showed the lowest identity with a given *SFB*, as in the case of identities between S^a *-RNase* and the other 11 *S-RNases* (Table 2). Inter-specific amino acid identities of SFBs were often higher than intra-specific identities. Because amino acid divergence can be affected by differences in the proportions of the proteins under positive or purifying selection, we also compared levels of synonymous change in *SFB* and *S-RNase* DNA sequences. Using the method of Nei and Gojobori (1986), the average percent of synonymous substitutions per site (D_s) for *SFB* and *S-RNase* sequences were 0.278 (SD 0.026) and 0.284 (SD 0.026), respectively, and not significantly different.

Phylogenetic reconstruction

The phylogenetic tree of 13 *Prunus SFB* alleles was poorly resolved, with no group receiving more than 72% bootstrap support (Fig. 3). The tree is consistent with

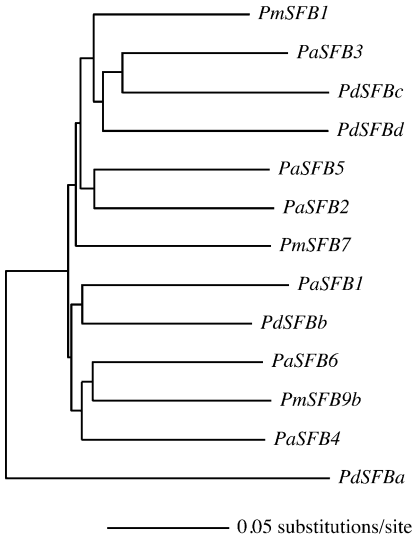

Fig. 3 Phylogenetic tree of *Prunus SFBs*. All phylogenetic reconstruction was conducted in PAUP* (Swofford 2001) as described in the text. Despite considerable variation among sequences, little phylogenetic signal was recovered. The highest bootstrap support for any grouping was only 72% for the node that joins *PaSFB3* and *PdSFBc*

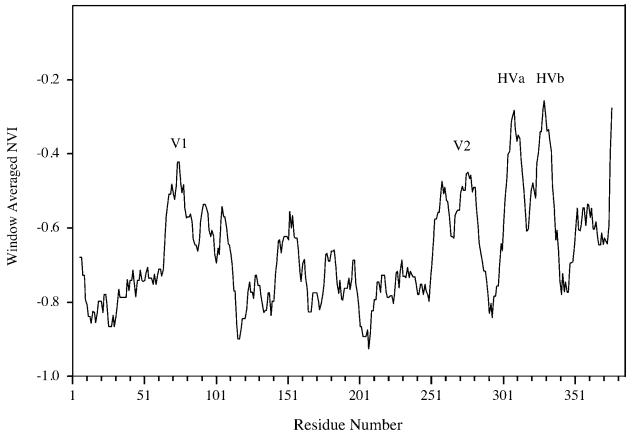

Fig. 4 Window-averaged plot of normed variability index (NVI) at each site in the alignment of the 13 SFB proteins. The NVI for each site was calculated as described by Kheyr-Pour et al. (1990). The NVI at each site was then averaged with its neighbors in a sliding-window of size 11. The last two peaks, HVa and HVb, were hypervariable regions. They corresponded to the variable regions A and B as defined previously (Ushijima et al. 2003). In addition, two new variable regions, V1 and V2, were detected in this study

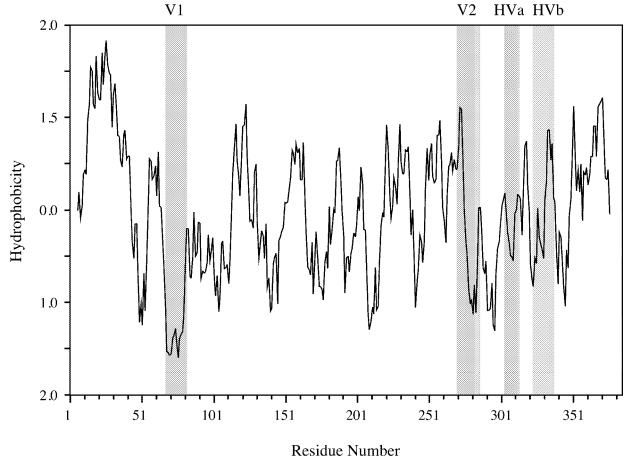

Fig. 5 Window-averaged composite hydrophobicity plot of the 13 SFB proteins. The amino acid hydropathicity indices (Kyte and Doolittle 1982) were averaged at each site over all 13 alleles. A hydrophobicity plot of the 13 SFBs was generated by averaging the average at each site with its neighbors in a sliding-window of size 11. *Shaded regions* (Hyper)variable regions V1, V2, HVa and HVb

trans-specific evolution of *SFBs* since alleles from different species group together; however, there is little support for such groupings. The tree is extremely star-like with all alleles diverging from one another near the base. A phylogenetic tree of the cognate *S-RNases* (not shown) is similarly star-like, making any assessment of whether or not *S-RNases* and *SFBs* share the same phylogenetic history difficult.

Primary structural features of *Prunus* SFB

In an alignment of 13 SFBs (Fig. 1), a total of 121 out of 384 sites were conserved and an additional 65 sites had only conservative replacements, which are scattered throughout the SFB. A total of 34 sites were considered to be variable sites having NVI values of more than −0.25, with most of the variable sites being located at the C-terminal region of SFB, as reported previously (Ushijima et al. 2003; Yamane et al. 2003a, 2003b). A window-averaged plot of NVI identified two variable regions, V1 and V2, and two hypervariable regions, HVa and HVb (Fig. 4). Three of these are located at the C-terminal region of SFB and the other is located directly downstream from the F-box motif (Fig. 4). A hydrophobicity plot of the 13 SFBs showed that the variable and hypervariable regions are not strongly hydrophobic, but there is no overall correlation between an amino acid's hydrophobicity and its NVI (r=−0.04, NS) (Fig. 5).

Evidence of positive selection

Likelihood ratio tests comparing different models of selection on *SFB* are presented in Table 3. Models that allow for positive selection provide a significantly better fit to the data than models that do not (models M2 and M3 tested against models M0 and M1). A more refined test of selection tested a model with a flexible distribution of ω values between 0 and 1 (no selection; M7) against a similar model that allows an additional class with ω>1 (positive selection; M8). This test is significant

Table 3 Likelihood ratio tests of models of selection on the *SFB* gene from *Prunus*. Models M0–M8 follow Yang and Swanson (2002). Model M0 assumes one selective class for all sites. Model M1 assumes two selective classes with conserved (ω_0=0) and neutral (ω_1=1) sites. Model M2 adds a third selective class which is free to vary, allowing for positive selection. Model M3 assumes three free selective classes. Model 7 assumes a beta distribution of ω over sites varying between 0 and 1. Model 8 adds to model 7 an additional selective class with ω free to vary, allowing for positive selection. All likelihood ratio tests are highly significant. In particular, models allowing for positive selection provide a significantly better fit to the data than models which do not (M2 and M3 vs M0 and M1; M8 vs M7)

Model	P[a]	lnL	−2(ΔlnL)[b]					
			M0	M1	M2	M3	M7	M8
M0	1	−6,461.50	–		180.7	261.1		
M1	1	−6,414.36		–	86.42	166.82		
M2	3	−6,371.15			–	80.4		
M3	5	−6,330.95				–		
M7	2	−6,369.97					–	75.6
M8	4	−6,332.17						–

[a] Number of parameters in each model associated with assumptions about the number and type of selective classes among codons
[b] Fits of the models to the data evaluated using twice the difference in the log-likelihoods. This value is tested against the Chi-square distribution with degrees of freedom equal to the difference in the number of free parameters

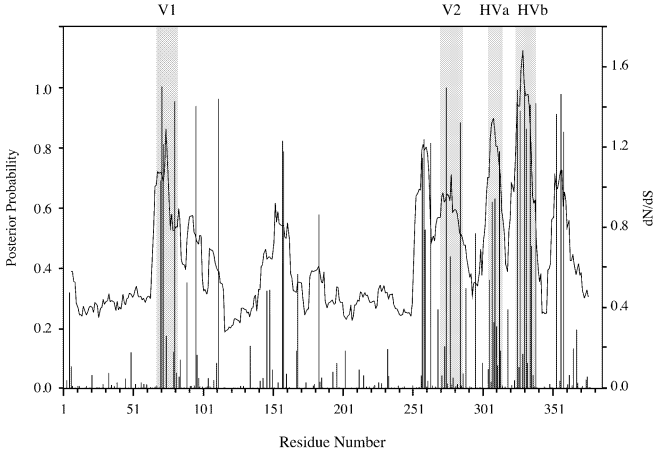

Fig. 6 Posterior probabilities of sites inferred to be under positive selection using PAML 3.13 (Yang 1997). *Vertical lines* posterior probability of positive selection on each codon under model M3, *continuous line* average ratio of non-synonymous (d*N*) to synonymous (d*S*) substitutions per site. Window size was 11 codons with the window centered over each codon

(2Δ*lnL*=75.6, 2 *df*, *P*<0.001). Thus our results suggest that positive selection acts on *SFB*.

The next step in our analysis attempts to identify specific sites under positive selection. The posterior probability values are significant (95% level) for six sites (70, 110, 273, 324, 329, 355) under the M3 model, in which ω_3=2.8, indicating positive selection. Model M8 identifies the same sites as M3 as well as an additional six sites (79, 94, 283, 326, 337, 352). Obviously, the 95% significance level is somewhat arbitrary, and although some sites identified as belonging to the selected class may not be significant at the 95% level, they are still potentially under positive selection. Posterior probabilities that sites belong to the positively selected class and a moving average of ω are shown in Fig. 6. Sites under positive selection are particularly common in the C-terminal end of the protein, in accordance with the results of the NVI analysis, and there is a strong correlation between a residue's NVI value and its estimated d*N*/d*S* ratio (r=0.8, P<0.001). Also as expected, the N-terminal F-box region of the molecule is under mostly purifying selection (low d*N*/d*S*).

Discussion

We have cloned and sequenced four *SFB* alleles from sweet cherry and showed that they have allele-specific sequence polymorphism. The characteristics of all *SFB* alleles that have been cloned so far, including the four newly cloned alleles in this study, provide several lines of evidence that *SFB* is a very good candidate for the pollen determinant in the GSI system in *Prunus* (Ushijima et al. 2003; Yamane et al. 2003a). Furthermore, the cloning of four additional alleles in this study made it possible to perform a detailed characterization of the primary structural features of *Prunus SFBs*.

Comparisons of amino acid sequences of the 13 *Prunus SFB* alleles and their linked *S-RNase* alleles, except for S^{9b} *-RNase*, revealed similar levels of variability among SFBs (66.0–82.5%) and S-RNases (54.6–86.5%). For both genes, the allele from the *P. dulcis* S^a-haplotype was least similar to all other alleles. On average, 88.9 amino acid residues (23.2%) differ between two *SFB* alleles, whereas 56.7 (27.5%) differ between two *S-RNase* alleles. Similarity in divergence values of *SFB* and *S-RNase* genes is one piece of evidence that *SFB* plays a role in pollen specificity. Comparisons of the SFB sequences identified both conserved and variable sites. About one-half of the total residues of SFB are conserved or have only conservative replacements. These conserved sites could be important for the structure and function of SFB. For example, many conserved residues at the N-terminal region of SFB belong to the F-box motif, which is essential for forming an SCF complex for protein degradation by the ubiquitin/26S proteasome pathway (Deshaies 1999). Other conserved sites scattered through-

out the SFB molecule could also be critical to SFB structure and function.

Variable sites could be important for determining the allelic specificity of the SFB protein. We calculated NVI values, which were previously used to asses the variable sites of S-RNases (Ioerger et al. 1991; Kheyr-Pour et al. 1990; Ushijima et al. 1998), to quantify the degree of variability for the non-conserved sites of *Prunus* SFBs. The NVI values of *Prunus* SFB sites were lower than those of rosaceous S-RNase sites as reported by Ushijima et al. (1998). This could be partly because a single genus was compared in this study to estimate NVI for SFB sites while several genera from different subfamilies were used to estimate NVI for rosaceous S-RNase sites (Ushijima et al. 1998). In fact, NVI values for *Prunus* S-RNases were comparable to those for *Prunus* SFBs (data not shown). In this study, we used the cut-off value of −0.25 to decide which sites are variable based on previous findings by Ushijima et al. (2003). Most of the variable sites are located at the C-terminal region of SFB, where the two previously identified variable regions A and B are located (Ushijima et al. 2003; Yamane et al. 2003a, 2003b). The HVa and HVb regions determined in this study are several residues narrower than the previously identified variable regions. Comparisons of more *SFB* alleles in this study made it possible to more precisely determine the HVa and HVb regions. The window-averaged plot of NVI that was constructed in this study confirmed that HVa and HVb were hypervariable regions.

We found significant support for the hypothesis that positive selection acts on the SFB molecule and where it acts is largely consistent with hypotheses generated from the NVI analysis. Sites under positive selection are largely, though not entirely, found in the C-terminal region of the protein, overlapping with regions identified as hypervariable. The good correspondence between V1, hydrophilicity, and sites inferred to be under positive selection also points to the region being potentially important in determining specificity. In addition, the N-terminal F-box motif is confirmed to be under purifying selection.

Hypervariability, lack of strong hydrophobicity, and evidence of positive selection of HVa and HVb as well as V2 at the C terminal region of SFB implies that the region may be responsible for the discrimination between self and non-self S-RNases. It is known that the C-terminal region of the F-box protein functions as a receptor to incorporate a target protein into an SCF complex, thus causing the incorporated protein to be polyubiquitinated and then degraded by the 26S proteasome (Deshaies 1999). Since Ushijima et al. (2003) proposed that SFB interacted specifically with the cognate S-RNase to maintain it in an active state, the cognate S-RNase may serve as a pseudosubstrate of SCF^{SFB} as in the case of the mammalian pseudosubstrate hnRNP-U of $SCF^{\beta\text{-TrCP}}$ (Davis et al. 2002). Investigation of the interaction between the (hyper)variable regions of the C-terminal region of SFB (HVa, HVb, and V2) and RHV of S-RNase is intriguing because a hypervariable region, RHV, of *Prunus* S-RNase is hydrophilic (data not shown) and thought to be exposed on the surface of the S-RNase molecule, as is the case with *Pyrus* S-RNases (Matsuura et al. 2001). In addition to the (hyper)variable regions at the C-terminal region of SFB, a window-averaged plot of NVI identified a variable V1 region at the N-terminal half of SFB, which is also hydrophilic and under positive selection, suggesting that this region may have some important function in discrimination between self and non-self S-RNases.

In summary, we have cloned four additional *SFB* alleles from sweet cherry, which made it possible to perform a detailed characterization of primary structural features of *Prunus* SFBs. The analysis revealed two hypervariable regions at the C-terminus of SFBs, which could be responsible for discrimination between self- and non-self S-RNases. Furthermore, codons inferred to be under positive selection are clustered in the hypervariable C-terminal region of the molecule. Levels of divergence between *S-RNases* and between *SFB* genes were similar, and the most divergent allele for each gene came from the same haplotype. All these features are consistent with the hypothesis that *SFB* is the pollen-*S* in GSI of *Prunus*. Although the overall structural features identified based on the analysis of 13 *SFB* alleles in this study likely represent the essential primary structural features of *Prunus SFBs*, a more precise picture could be constructed if comparisons were made with homologous sequences in Solanaceae, Scrophulariaceae and/or the subfamily Maloideae of Rosaceae, providing the *SFB* locus is present in these species.

Acknowledgements The authors gratefully acknowledge the gifts of plant material from the National Institute of Fruit Tree Science (Morioka, Japan). This work was supported by the Japan-US Cooperative Science Program, a Grant-in-Aid (no. 13460014) for Scientific Research (B) to R.T., and a Grant-in-Aid (no. 15004783) from the Japan Society for the Promotion of Science (JSPS) Research Fellows to K.U., who is a Research Fellow of JSPS.

References

Anderson MA, Cornish EC, Mau S-L, Williams EG, Hoggart R, Atkinson A, Bonig I, Grego B, Simpson R, Roche P, Haley JD, Penschow J, Niall HD, Tregear GW, Coghlan JP, Crawford RJ, Clarke AE (1986) Cloning of cDNA for a stylar glycoprotein associated with expression of self-incompatibility in *Nicotiana alata*. Nature 321:38–44

Clark AG, Kao T-H (1991) Excess nonsynonymous substitution at shared polymorphic sites among self-incompatibility alleles of Solanaceae. Proc Natl Acad Sci USA 88:9823–9827

Craig KL, Tyers M (1999) The F-box: a new motif for ubiquitin dependent proteolysis in cell cycle regulation and signal transduction. Prog Biophys Mol Biol 72:299–328

Davis M, Hatzubai A, Andersen JS, Ben-Shusham E, Fisher Z, Yaron A, Bauskin A, Mercurio F, Mann M, Ben-Neriah Y (2002) Pseudosubstrate regulation of the $SCF^{\beta\text{-TrCP}}$ ubiquitin ligase by hnRNP-U. Genes Dev 16:439–451

Dayhoff MO, Schwartz RM, Orcutt BC (1979) A model for evolutionary change in proteins. In: Dayhoff MO (ed) Atlas of protein sequence and structure, vol 5, Suppl 3. National Biomedical Research Foundation, Washington D.C., pp 345–352

Deshaies RJ (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, 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

Gagne JM, Downes BP, Shiu S-H, Durski AM, Vierstra RD (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

Golz JF, Oh H-Y, Su V, Kusaba M, Newbigin E (2001) Genetic analysis of *Nicotiana* pollen-part mutants is consistent with the presence of an S-ribonuclease inhibitor at the *S* locus. Proc Natl Acad Sci USA 98:15372–15376

Ioerger TR, Gohike JR, Xu B, Kao T-H (1991) Primary structural features of the self-incompatibility protein in Solanaceae. Sex Plant Reprod 4:81–87

Ishimizu T, Endo T, Yamaguchi-Kabata Y, Nakamura KT, Sakiyama F, Norioka S (1998) Identification of regions in which positive selection may operate in S-RNase of Rosaceae: implication for *S*-allele-specific recognition sites in S-RNase. FEBS Lett 440:337–342

Kheyr-Pour A, Bintrim SB, Ioerger TR, Remy R, Hammond SA, Kao T-H (1990) Sequence diversity of pistil S-proteins associated with gametophytic self-incompatibility in *Nicotiana alata*. Sex Plant Reprod 3:88–97

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

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105–132

Lai Z, Ma W, Han B, Liang L, Zhang Y, Hong G, 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

Luu D, Quin X, Morse D, Cappadocia M (2000) S-RNase uptake by compatible pollen tubes in gametophytic self-incompatibility. Nature 407:649–651

Matsuura T, Sakai H, Unno M, Ida K, Sato M, Sakiyama F, Norioka S (2001) Crystal structure at 1.5-angstrom resolution of *Pyrus pyrifolia* pistil ribonuclease responsible for gametophytic self-incompatibility. J Biol Chem 276:45261–45269

McClure BA, Haring V, Ebert PR, Anderson MA, Simpson RJ, Sakiyama F, Clarke AE (1989) Style self-incompatibility gene products of *Nicotiana alata* are ribonucleases. Nature 342:757–760

McCubbin AG, Kao T-H (2000) Molecular recognition and response in pollen and pistil interactions. Annu Rev Cell Dev Biol 16:333–364

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Nettancourt D de (2001) Incompatibility and incongruity in wild and cultivated plants. Springer, Berlin Heidelberg New York

Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148:929–936

Posada D, Crandall KA (1998) MODELTEST: Testing the model of DNA substitution. Bioinformatics 14:817–818

Sassa H, Hirano H, Ikehashi H (1992) Self-incompatibility-related RNases in styles of Japanese pear (*Pyrus serotina* Rehd.). Plant Cell Physiol 33:811–814

Swofford DL (2001) PAUP*. Phylogenetic analysis using parsimony (and other methods), version 4. Sinauer Associates, Sunderland, Mass.

Tao R, Yamane H, Sugiura A, Murayama H, Sassa H, Mori H (1999) Molecular typing of *S*-alleles through identification, characterization and cDNA cloning for S-RNases in sweet cherry. J Am Soc Hortic Sci 124:224–233

Tao R, Namba A, Yamane H, Fuyuhiro Y, Watanabe T, Habu T, Sugiura A (2003) Development of the S^f-RNase gene-specific PCR primer set for Japanese apricot (*Prunus mume* Sieb. et zucc.). Hortic Res (Japan) 2:237–240

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Thompson RD, Kirch HH (1992) The *S* locus of flowering plants: when self-rejection is self-interest. Trends Genet 8:381–387

Ushijima K, Sassa H, Tao R, Yamane H, Dandekar AM, Gradziel TM, Hirano H (1998) Cloning and characterization of cDNAs encoding S-RNases from almond (*Prunus dulcis*): primary structural features and sequence diversity of the S-RNases in Rosaceae. Mol Gen Genet 260:261–268

Ushijima K, Sassa H, Tamura M, Kusaba M, Tao R, Gradziel TM, Dandekar AM, Hirano H (2001) Characterization of the *S*-locus region of almond (*Prunus dulcis*): analysis of a somaclonal mutant and a cosmid contig for an *S* haplotype. Genetics 158:379–386

Ushijima K, Sassa H, Dandekar AM, Gradziel TM, Tao R, 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

Wang Y, Wang X, Skirpan AL, Kao T-H (2003) S-RNase-mediated self-incompatibility. J Exp Bot 54:115–122

Xue Y, Carpenter R, Dickinson HG, Coen ES (1996) Origin of allelic diversity in *Antirrhinum S* locus RNases. Plant Cell 8:805–814

Yamane H, Ikeda K, Ushijima K, Sassa H, Tao R (2003a) 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

Yamane H, Ushijima K, Sassa H, Tao R (2003b) The use of *S* haplotype-specific F-box protein gene, *SFB*, as a molecular marker for *S*-haplotype and self-compatibility in Japanese apricot (*Prunus mume*). Theor Appl Genet 107:1357-1361

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13:555–556

Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol 15:568–573

Yang Z, Bielawski JP (2000) Statistical methods for detecting molecular adaptation. Trends Ecol Evol 15:496–503

Yang Z, Swanson WJ (2002) Codon substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. Mol Biol Evol 19:49-57

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