Related polymorphic F-box protein genes between haplotypes clustering in the BAC contig sequences around the S-RNase of Japanese pear

Kazuma Okada1,†, Nozomi Tonaka1, Tomio Taguchi1, Takehiko Ichikawa1, Yutaka Sawamura2, Tetsu Nakanishi1 and Takeshi Takasaki-Yasuda1,*

1 Graduate School of Agricultural Science, Kobe University, Kobe, Hyogo 657-8501, Japan
2 National Institute of Fruit Tree Science, Tsukuba, Ibaraki 305-8605, Japan
† Present address: National Institute of Fruit Tree Science, Morioka, Iwate 020-0123, Japan.
* To whom correspondence should be addressed. E-mail: taka@kobe-u.ac.jp

Abstract

Most fruit trees in the Rosaceae exhibit self-incompatibility, which is controlled by the pistil S gene, encoding a ribonuclease (S-RNase), and the pollen S gene at the S-locus. The pollen S in Prunus is an F-box protein gene (SLF/SFB) located near the S-RNase, but it has not been identified in Pyrus and Malus. In the Japanese pear, various F-box protein genes (PpSFBB1–4) located near the S-RNase genes of Japanese pear were constructed, and 649 kb around S2-RNase and 378 kb around S3-RNase were sequenced. Six and 10 pollen-specific F-box protein genes (designated as PpSFBB4-u1-u4, 4-d1-d2 and PpSFBB2-g5, respectively) were found, but PpSFBB4-u1-u5 and PpSFBB2-g5 were absent. The PpSFBB4 genes showed 66.2–93.1% amino acid identity with the PpSFBB2 genes, which indicated clustering of related polymorphic F-box protein genes between haplotypes near the S-RNase of the Japanese pear. Phylogenetic analysis classified 36 F-box protein genes of Pyrus and Malus into two major groups (I and II), and also generated gene pairs of PpSFBB genes and PpSFBB/Malus F-box protein genes. Group I consisted of gene pairs with 76.3–94.9% identity, while group II consisted of gene pairs with higher identities (>92%) than group I. This grouping suggests that less polymorphic PpSFBB genes in group II are non-S pollen genes and that the pollen S candidates are included in the group I PpSFBB genes.

Key words: BAC contig, F-box protein, pollen S gene, Pyrus pyrifolia, self-incompatibility, S-locus, S-RNase.

Introduction

Self-incompatibility (SI) is a genetic system that prevents self-fertilization in flowering plants by the recognition and rejection of self-pollen (de Nettancourt, 2001). In the Rosaceae, Solanaceae, and Plantaginaceae families, SI is classified as gametophytic SI (GSI), and is controlled by a single S-locus with multiple S-haplotypes. Each S-haplotype contains two genetically linked genes, the pistil S gene and the pollen S gene, which determine the S-haplotype specificity of the pistil and pollen, respectively (McCubbin and Kao, 2000). The pistil S encodes a ribonuclease known as S-RNase (McClure et al., 1989; Ishimizu et al., 1996; Xue et al., 1996). The RNase activity of S-RNases is essential for rejection of self-pollen, and the degradation of rRNA by S-RNases inside the self-pollen tube results in inhibition of pollen growth (McClure et al., 1990; Huang et al., 1994). Thus, it is thought that the self S-RNase inhibits growth of the self-pollen tube via degradation of pollen rRNAs. On the other hand, the identity and function of the pollen S remained unknown for a long time. Recently, F-box protein genes were identified as the pollen S genes by
sequence analyses of cosmid and bacterial artificial chromosome (BAC) contigs around S-RNase in Prunus species of the Rosaceae, in Petunia inflata of the Solanaceae, and in Antirrhinum hisppanicum of the Plantaginaceae. These F-box protein genes were termed SLF (S-locus F-box) or SFB (S-haplotype-specific F-box protein) (Lai et al., 2002; Entani et al., 2003; Ushijima et al., 2003; Sijacic et al., 2004). Transformation experiments in P. inflata and analyses of pollen-part self-compatible (SC) mutants in Prunus species provided evidence that SLF/SFB genes are the pollen S genes (Sijacic et al., 2004; Ushijima et al., 2004; Sonneveld et al., 2005; Hauck et al., 2006; Tsukamoto et al., 2006; Vilanova et al., 2006). Generally, F-box proteins function as one of the four major subunits (CUL1, SKP1, RBX1, and F-box) that make up the SCF complex, which regulates protein stability through the ubiquitin–proteasome system (Lechne et al., 2006). The model for S-RNase degradation proposes that the non-self-interaction between S-RNase and SLF/SFB leads to S-RNase ubiquitylation and degradation by the 26S proteasome (McClure and Franklin-Tong, 2006).

In Rosaceae, the pollen S has been identified only in Prunus (almond, apricots, and cherry), but not in Pyrus (pear) and Malus (apple). The Rosaceae comprises three subfamilies: Rosoideae, Dryadoideae, and Spiraeoideae. Prunus, Pyrus, and Malus are all included in Spiraeoideae (Potter et al., 2007). Therefore, it is likely that the pollen S genes in Pyrus and Malus are also F-box protein genes. Recently, S-locus-linked and pollen-specific polymorphic F-box protein genes were isolated from apple (Malus×domestica) and Japanese pear (Pyrus pyrifolia), and these have been proposed as good candidates for the pollen S genes. Cheng et al. (2006) cloned two S-locus-linked F-box protein genes (MdSLF1 and MdSLF2) from apple by reverse transcription-PCR (RT-PCR) with degenerate primers designed from the conserved SLF/SFB sequences. Sassa et al. (2007) found several pollen-specific polymorphic F-box protein genes termed SFB (S locus F-box brothers) in BAC contig sequences around apple S-RNase genes. These SFB genes include MdSFBB1,2 and MdSFBB1,8 around S2-RNase, and MdSFBB2,8 and MdSFBB2,14 around S3-RNase. Using RT-PCR, they also cloned various PpSFBB genes (PpSFBB2, PpSFBB3, and PpSFBB10) that are linked to S-RNase genes of the Japanese pear; PpSFBB1,2, PpSFBB3,8, and PpSFBB4,7 are linked to S4-RNase, and PpSFBB2,8, PpSFBB3,8, and PpSFBB7,14 are linked to S1-RNase. PpSFBB7 genes that are linked to another eight S-RNase genes have been cloned. They show high amino acid sequence identities (97.5–99.7%) among the 10 S-haplotypes (Kakui et al., 2007). However, it is not clear whether PpSFBB genes are located near the S-RNase, like MdSFBB genes, or whether they are the pollen S genes. To identify the pollen S genes in the Japanese pear, a previously constructed BAC library from an S1 homozygote was used and a BAC contig of ~570 kb around S4-RNase was assembled. Sequence analysis of the 240 kb spanning 51 kb upstream to 189 kb downstream of S4-RNase revealed a pollen-specific F-box protein gene (S4F-box0; S4-haplotype F-box protein gene) that differed from PpSFBB1,8, S4F-box0 is located 127 kb downstream of S4-RNase (Okada et al., 2008). The SC cultivar ‘Osa Nijisseiki’ (S2S3sm) is a natural mutant derived from ‘Nijisseiki’ (S2S4). The S4-haplotype of ‘Osa Nijisseiki’ lacks the pistil S function but retains the pollen S function, and is termed the S4sm-haplotype, where ‘sm’ stands for ‘stilar-part mutant’ (Sato, 1993). The S4sm-haplotype has a 236 kb deletion, which includes S2-RNase and S4F-box0, suggesting that the pollen S allele is located outside of the region spanning 48 kb upstream to 188 kb downstream of S2-RNase—that is, outside the region that is deleted in the S4sm-haplotype (Okada et al., 2008).

In this study, the sequence outside of the deleted region in S4sm was analysed, and the 649 kb region from 290 kb upstream to 359 kb downstream of S2-RNase was determined; six PpSFBB genes were found. To evaluate the S-haplotype polymorphism of PpSFBB genes, a BAC library was constructed from the Japanese pear cultivar ‘Choujuuro’ (S2S3) to assemble a BAC contig around S4-RNase. A 378 kb region from 166 kb upstream to 212 kb downstream of S4-RNase was sequenced, and 10 PpSFBB genes were found. Relationships among 36 F-box protein genes of Pyrus and Malus were analysed by comparing their amino acid sequences and by phylogenetic clustering.

Materials and methods

Plant materials

One cultivar and three S homozygotes of the Japanese pear were used: ‘Choujuuro’ (S2S3), and S2, S3, and S4 homozygotes. The S2 and S3 homozygotes were selected from bud-selfed progeny of ‘Choujuuro’ (S2S3) (Terai et al., 1999). The S2 homozygote was segregated from bud-selfed progeny of ‘Nijisseiki’ (S2S4) (Okada et al., 2008). The leaves, mature pollen, and pistils were frozen in liquid nitrogen, and stored at −80 °C until use.

Construction and characterization of an S2S3 BAC library

An S2S3 BAC library was constructed and characterized according to the method of Okada et al. (2008). High molecular weight DNA was isolated from leaf tissue (3 g) of ‘Choujuuro’ (S2S3), partially digested with HindIII, and size-selected twice by pulsed field gel electrophoresis (PFGE). In the first size selection, an agarose slice containing DNA fragments of 60–210 kb was excised and embedded into a new 1% SeaPlaque GTG agarose gel (Cambrex, http://www.cambrex.com). In the second size selection, two size fractions (145–185 kb and 185–205 kb) were recovered by digestion of agarose slices with β-agarase I. DNA from each fraction was separately ligated into HindIII-digested CopyControl pCC1BAC Cloning-Ready vector (EPICENTRE, http://www.epi-bio.com/) and transformed into Escherichia coli strain TransMax EPI300 (EPICENTRE). Equal numbers of transformed cells were picked from each fraction and a total of 61 440 colonies were pooled in 64 individual 96-well plates with 12 columns and eight rows (10 colonies per well) and stored at −80 °C. The BAC plasmid was extracted from the randomly chosen BAC clones by the standard alkaline lysis method, digested with NotI, and separated by PFGE. Insert size was estimated by comparison with a PFGE lambda ladder (New England Biolabs, http://www.neb.com/).
Chromosome walking

Chromosome walking in the region around \( S-R\text{N} \) was performed by PCR screening of the \( S_2S_4 \) BAC library and the previously constructed \( S_2S_5 \) BAC library (Okada et al., 2008). The PCR screening was performed in three consecutive steps as described by Okada et al. (2008). Chromosome walking around the \( S_2-R\text{N} \) was initiated by PCR screening of the \( S_2S_5 \) BAC library with an \( S-R\text{N} \)-specific primer pair, ‘FTQQYQ’ and ‘anti-IIWPNV’ (Ishimizu et al., 1999).

BAC plasmids were isolated from positive BAC clones using the Plasmid Midi Kit (Qiagen, http://www1.qiagen.com/). Both ends of the BACs (~600 bp) of the positive clones were sequenced using T7 and RP vector primers, and a primer pair was designed from each BAC-end sequence (Supplementary Table S1 available at JXB online). For chromosome walking, non-repetitive primer pairs were selected from the BAC-end primer pairs located at the outer ends of the contig by PCR amplification of plate pool templates, which were prepared by mixing all 960 BAC clones in each plate. Furthermore, \( S_2 \)-specific primer pairs were identified from among the non-repetitive primer pairs by PCR, using genomic DNA of the \( S_2 \) and \( S_5 \) homozygotes as templates. These \( S_2 \)-specific primer pairs were used for PCR screening of the BAC library (Supplementary Table S1). For PCR, genomic DNA was isolated from leaves of the \( S_2 \) and \( S_5 \) homozygotes by the modified cetyltrimethylammonium bromide (CTAB) method (Castillo et al., 2001).

To estimate insert sizes and compare restriction patterns, BAC plasmids were digested with NotI and subjected to PFGE. Based on overlapping of the BAC clones, their insert sizes, and restriction patterns, the physical distance was calculated to construct a BAC contig.

BAC subcloning

The plasmids of BAC clones were completely digested with HindIII or EcoRI, separated on a 0.7% agarose gel, and purified from the gels using GENECLEAN Kit III (Qbiogene, http://www.qbiogene.com/). Each fragment was ligated into pBluescript-II SK (+) and transformed into E. coli strain TOP10F\(^+\) (Invitrogen, http://www.invitrogen.com/). Inserts from subclones that were smaller than 7 kb were sequenced by primer walking, and those that were larger than 7 kb were sequenced after subcloning using other restriction enzymes. A primer was designed from each insert-end sequence. Using these primers, the regions outside the subclones in the BAC plasmids were sequenced. The sequences from subclones and the outside sequences were assembled to construct contigs for each BAC clone. Gap regions, for which no sequence data were obtained, were amplified from BAC plasmids by PCR with the Expand High Fidelity PCR system (Roche Diagnostics, http://www.roche-diagnostics.jp/), and directly sequenced.

Nucleotide and amino acid sequence analysis

Nucleotide sequences were determined with the BigDye Terminator v3.1 Cycle Sequencing Kit using an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, http://www3.appliedbiosystems.com/). The sequence data were analysed using GENETYX-MAC Ver. 13 and ATGC Ver. 4 software packages (Genetyx, http://www.sdc.co.jp/genetyx/). Protein-coding sequences were predicted using the GENSCAN program (Burge and Karlin, 1997). Homology searches were carried out using the BLASTX program (Altschul et al., 1997). Deduced amino acid sequences were analysed by Pfam (http://pfam.janelia.org/) to search for protein motifs. Amino acid sequences were aligned using ClustalW (http://clustalw.ddbj.nig.ac.jp/top-j.html) and manually optimized. A phylogenetic tree was constructed by the Neighbor–Joining method (Saitou and Nei, 1987). A Harr plot analysis was performed using GENETYX-MAC Ver. 13 software.

RT-PCR

Total RNAs extracted from pollen, pistils, and leaves were subjected to first-strand cDNA synthesis using ReverTra Ace \( \alpha \) (TOYOBO, http://www.toyoobo.co.jp/). Using the Expand High Fidelity PCR system (Roche Diagnostics), PCR was then carried out with gene-specific primer pairs (Supplementary Table S2 at JXB online). PCR products were separated by electrophoresis on 1.5% or 0.8% agarose gels and visualized by ethidium bromide staining. The PCR products were purified from the gels, and sequenced to confirm gene specificity.

Sequence data

The 649 kb and 378 kb sequences around \( S_4 \) and \( S_2-R\text{N} \) genes have been deposited in the EMBL/GenBank Data Libraries under accession nos AB545981 and AB545982, respectively.

Results

Sequence analysis of 649 kb around the \( S_4-R\text{N} \) gene

Previously, an \( S_4 \) BAC contig spanning 202 kb upstream (18F9 T7-end) to 359 kb downstream (33G4 T7-end) of \( S_4-R\text{N} \) was constructed. In addition, the complete sequences of two BAC clones (17C7 and 5D3) and the partial sequences of two other BAC clones (32D11 and 4D9) were analysed to determine a 240 001 bp sequence spanning 51 kb upstream to 189 kb downstream of \( S_4-R\text{N} \) (Okada et al., 2008). In this study, 32D11 and 4D9 were sequenced in their entirety. To extend the \( S_4 \) BAC contig upstream, chromosome walking was resumed using a non-repetitive BAC-end primer pair (33H11-T7). PCR screening of the \( S_2S_4 \) BAC library yielded five BAC clones: 12B8, 15C1, 31C7, 31F1, and 36B6. As a result, chromosome walking from \( S_2-R\text{N} \) produced a set of overlapping BAC clones (36B6, 18F9, 32D11, 17C7, 5D3, 4D9, and 33G4) covering ~649 kb, spanning 290 kb upstream to 359 kb downstream of \( S_2-R\text{N} \) (Fig. 1A). Three BAC clones (36B6, 18F9, and 33G4) were subcloned and completely sequenced. The sequence assembly of the seven BAC clones (36B6, 18F9, 32D11, 17C7, 5D3, 4D9, and 33G4) yielded a 648 516 bp sequence spanning 290 kb upstream to 359 kb downstream of \( S_2-R\text{N} \).

Analysis using GENSCAN software predicted 89 open reading frames (ORFs) in the \( S_2 \) BAC 649 kb contig sequence (Fig. 2A). Among the 89 ORFs, 34 (ORF33–ORF66) were included in the 236 kb deleted region of the \( S_2 \)-haplotype that spans 48 kb upstream to 188 kb downstream of \( S_2-R\text{N} \) (Okada et al., 2008). The other 55 ORFs (ORF1–ORF32 and ORF67–ORF89) were located outside the deleted region. A BLASTX search of the 89 ORFs yielded 61 ORFs with significant similarity (E-value <e-4) to sequences of known proteins in the database (Table 1). Of the 89 ORFs, 40 showed similarity to a (retro) transposon, and 10 were similar to a hypothetical or predicted protein. ORF14 was similar to a zinc knuckle family protein, ORF43 to a zinc finger, ORF50 to a transposon, and ORF62 to an unknown protein. ORF42 and ORF54 corresponded to...
S4-RNase and S4F-box0, respectively. Five ORFs (ORF3, ORF9, ORF15, ORF25, and ORF79) were located outside the deleted region, and showed similarity to MdSFBBs and PpSFBBp9-1. Using GENETYX-MAC Ver. 13 software, the predicted ORFs were reanalysed to determine the precise ORFs from the start (ATG) to the stop codon. ORF3, ORF9, ORF15, ORF25, and ORF79 encoded 410, 390, 403, 394, and 393 amino acid residues, respectively. A Pfam motif search predicted that these five proteins had an F-box domain at the N-terminus and an FBA_1 domain in the
centre (Fig. 3). These ORFs showed pairwise deduced amino acid sequence identities ranging from 62.9% to 94.4% when compared with reported \textit{PpSFBB}, \textit{MdSFBB}, and \textit{MdSLF} genes (Supplementary Table S3 at JXB online). They represent F-box protein genes that differ from \textit{PpSFBB}$^{4-a-c}$, which are linked to \textit{S4-RNase} (Sassa \textit{et al.}, 2007). Thus, these five ORFs were assigned as new \textit{PpSFBB}$^{4}$ genes. ORF3, ORF9, ORF15, and ORF25 were located; 284, 245, 199, and 113 kb upstream of \textit{S4-RNase}. ORF54 (\textit{S4F-box0}) and ORF79 were located; 127 kb and 275 kb downstream of \textit{S4-RNase}. These \textit{PpSFBB}$^{4}$ genes upstream and downstream of \textit{S4-RNase} were named \textit{PpSFBB}$^{4-u}$ and \textit{PpSFBB}$^{4-d}$, respectively, and lower case numbers were assigned to the \textit{PpSFBB}$^{4-u}$ and \textit{PpSFBB}$^{4-d}$ located close to \textit{S4-RNase}; therefore, ORF25, ORF15, ORF9, and ORF3 were designated as \textit{PpSFBB}$^{4-u1}$, \textit{PpSFBB}$^{4-u2}$, \textit{PpSFBB}$^{4-u3}$, and \textit{PpSFBB}$^{4-u4}$, and ORF54 (\textit{S4F-box0}) and ORF79 were (re)named as \textit{PpSFBB}$^{4-d1}$ and \textit{PpSFBB}$^{4-d2}$, respectively. These \textit{PpSFBB}$^{4}$ genes around \textit{S4-RNase} shared the same transcriptional orientation, except for \textit{PpSFBB}$^{4-d1}$ (Fig. 1A). Using ATGC Ver. 4 software, the \textit{S4} 649 kb BAC contig sequence was searched for \textit{SFBB}-like sequences. The analysis revealed a pseudogene (\textit{WpSFBB}$^{4-u1}$) encoding a truncated F-box protein at \textasciitilde239 kb upstream of \textit{S4-RNase} (Fig. 1A). \textit{PpSFBB}$^{4-u1}$–\textit{u4} and \textit{PpSFBB}$^{4-d1}$–\textit{d2} shared 67.2–86.2% amino acid sequence identities with each other (Table 3).

To examine expression of \textit{PpSFBB}$^{4-u1}$–\textit{u4} and \textit{PpSFBB}$^{4-d1}$–\textit{d2}, total RNA was extracted from pollen, pistils, and leaves of the \textit{S4} homozygote. RT-PCR analyses were conducted using gene-specific primer pairs (Supplementary Table S2 at JXB online). \textit{PpSFBB}$^{4-u1}$–\textit{u4} and \textit{PpSFBB}$^{4-d1}$–\textit{d2} were all specifically expressed in pollen, but not in pistils or leaves (Supplementary Fig. S1A). The \textit{PpSFBB}$^{4-d2}$-specific primer pair yielded fragments of 1373 bp and 1142 bp, which both were derived from the \textit{PpSFBB}$^{4-d2}$ transcript, because the forward primer annealed to the 5' untranslated region (UTR) and the coding region of \textit{PpSFBB}$^{4-d2}$ (Supplementary Table S2, Fig. S1A). Thus, in the 649 kb sequence around \textit{S4-RNase} there were six F-box protein genes (\textit{PpSFBB}$^{4-u1}$–\textit{u4} and \textit{PpSFBB}$^{4-d1}$–\textit{d2}) with pollen-specific expression. The three \textit{PpSFBB} genes previously shown to be linked to the \textit{S4-RNase}, \textit{PpSFBB}$^{4-a-c}$, were not within the sequenced region.

**Fig. 2.** ORF maps of the region around \textit{S4-RNase} (A) and \textit{S2-RNase} (B). Arrowheads indicate the location and transcriptional direction of genes predicted by GENSCAN software. Open arrowheads indicate genes showing no significant homology to proteins in databases. Grey arrowheads represent transposable elements. Black arrowheads indicate non-tranposon-like genes. The 649 kb and 378 kb sequences around \textit{S4-RNase} and \textit{S2-RNase} have been deposited with the EMBL/GenBank Data Libraries under accession nos AB545981 and AB545982, respectively.
| ORFs | Homologous protein | Species | Amino acid identity | Score (bits) | E-value* | Accession no. |
|------|--------------------|---------|---------------------|-------------|---------|----------------|
| ORF1 | None               |         |                     |             |         |                |
| ORF2 | Transposon protein | Oryza sativa | 64/122 (52%)       | 134         | 9e-30   | DP000011       |
| ORF3 | (SFBB<sup>4-uδ</sup>) | PpSFBB<sup>5-7</sup> | Arabidopsis thaliana | 24/36 (66%) | 50.8   | AC006920       |
| ORF4 | GAG-POL precursor  | Vitis vinifera | 34/124 (27%)       | 61.2         | 3e-08   | AB111100       |
| ORF5 | None               |         |                     |             |         |                |
| ORF6 | None               |         |                     |             |         |                |
| ORF7 | Retrotasposon protein | Oryza sativa | 45/107 (42%)       | 79.7         | 9e-14   | DP000009       |
| ORF8 | Retrotasposon protein | Oryza sativa | 45/158 (28%)       | 82.4         | 1e-14   | DP000009       |
| ORF9 | (SFBB<sup>4-uδ</sup>) | MdSFBB<sup>3-6</sup> | Malus domestica | 350/390 (89%) | 726    | AB270792       |
| ORF10 | None               |         |                     |             |         |                |
| ORF11 | None               |         |                     |             |         |                |
| ORF12 | Retrotasposon protein | Oryza sativa | 31/78 (39%)       | 54.3         | 6e-08   | DP000011       |
| ORF13 | None               |         |                     |             |         |                |
| ORF14 | None               |         |                     |             |         |                |
| ORF15 | (SFBB<sup>4-uδ</sup>) | MdSFBB<sup>3-6</sup> | Malus domestica | 300/315 (96%) | 590    | AC120534       |
| ORF16 | None               |         |                     |             |         |                |
| ORF17 | None               |         |                     |             |         |                |
| ORF18 | Predicted protein  | Populus trichocarpa | 44/134 (32%) | 69.7     | 2e-10   | DS017968       |
| ORF19 | None               |         |                     |             |         |                |
| ORF20 | Retrotasposon protein | Oryza sativa | 31/81 (38%)       | 60.1         | 1e-09   | DP000086       |
| ORF21 | Hypothetical protein | Vitis vinifera | 43/132 (32%) | 70.5   | 1e-10 | AM426737       |
| ORF22 | Retrotasposon protein | Beta vulgaris | 61/173 (35%) | 82.0     | 3e-13   | EF101866       |
| ORF23 | Retrotasposon gag protein | Asparagus officinalis | 418/767 (54%) | 772   | 0.0 | AC183435       |
| ORF24 | Retrotasposon gag protein | Oryza sativa | 487/1302 (47%) | 921    | 0.0 | AC120534       |
| ORF25 | (SFBB<sup>4-uδ</sup>) | MdSFBB<sup>3-6</sup> | Malus domestica | 306/394 (77%) | 628    | AC270792       |
| ORF26 | None               |         |                     |             |         |                |
| ORF27 | None               |         |                     |             |         |                |
| ORF28 | Reverse transcriptase | Vigna radiata | 23/37 (62%) | 48.5   | 2e-04 | AY694634       |
| ORF29 | Retrotasposon protein | Oryza sativa | 521/962 (54%) | 1028    | 0.0 | DP000011       |
| ORF30 | Retrotasposon gag protein | Asparagus officinalis | 48/129 (37%) | 91.3   | 2e-16  | AC183436       |
| ORF31 | Retrotasposon gag protein | Asparagus officinalis | 827/1032 (60%) | 1272  | 0.0 | AC120534       |
| ORF32 | Integrase          | Populus trichocarpa | 48/149 (32%) | 84.0   | 3e-14  | DQ581600       |
| ORF33 | Predicted protein  | Populus trichocarpa | 34/92 (36%) | 62.4    | 3e-08  | EQ134071       |
| ORF34 | None               |         |                     |             |         |                |
| ORF35 | Retrotasposon gag protein | Asparagus officinalis | 95/194 (48%) | 172    | 4e-41  | AC183435       |
| ORF36 | Retrotasposon gag protein | Asparagus officinalis | 117/405 (28%) | 141    | 2e-31  | AC183435       |
| ORF37 | None               |         |                     |             |         |                |
| ORF38 | None               |         |                     |             |         |                |
| ORF39 | None               |         |                     |             |         |                |
| ORF40 | Transposon protein Pong subclass | Zea mays | 31/120 (25%) | 60.1    | 3e-07  | EU964924       |
| ORF41 | Transposon protein Pong subclass | Zea mays | 185/380 (48%) | 355 | 8e-96 | EU962682       |
| ORF42 | (S<sub>x</sub>RNase) | S<sub>x</sub>RNase | PpSFBB<sup>5-7</sup> | 49/50 (68%) | 115 | 2e-24 | AB014072       |
| ORF43 | Zinc finger        | Medicago truncatula | 23/63 (36%) | 47.8 | 6e-04 | AC148290       |
| ORF44 | None               |         |                     |             |         |                |
| ORF45 | Retrotasposon gag protein | Asparagus officinalis | 65/291 (22%) | 59.7   | 5e-07  | AC183435       |
| ORF46 | Retroelement pol polyprotein-like | Arabidopsis thaliana | 667/1331 (50%) | 1272 | 0.0 | AB024037       |
| ORF47 | Retrotasposon gag protein | Asparagus officinalis | 793/1644 (48%) | 1496 | 0.0 | AC183435       |
| ORF48 | Retrotasposon protein | Oryza sativa | 352/728 (48%) | 673  | 0.0 | DP000011       |
| ORF49 | None               |         |                     |             |         |                |
| ORF50 | Chromosome-associated kinesin KIF4A | Ricinus communis | 37/96 (38%) | 60.5   | 2e-07  | EQ974117       |
| ORF51 | None               |         |                     |             |         |                |
| ORF52 | None               |         |                     |             |         |                |
| ORF53 | (SFBB<sup>4-dω</sup>) | S<sub>x</sub>-F-box0 | PpSFBB<sup>5-7</sup> | 400/400 (100%) | 834 | 0.0 | AB308360       |
| ORF54 | None               |         |                     |             |         |                |
| ORF55 | None               |         |                     |             |         |                |
| ORF56 | Hypothetical protein | Vitis vinifera | 45/154 (29%) | 72.0   | 2e-11  | AM429787       |
| ORF57 | Hypothetical protein | Vitis vinifera | 50/135 (37%) | 81.3   | 2e-13  | AM467140       |
Construction of a BAC contig around the S2-RNase gene

To analyse the sequence polymorphism of *PpSFBB*\(^{4d1–d2}\) and *PpSFBB*\(^{4u1–u4}\) in another haplotype, a BAC library was constructed from the Japanese pear cultivar ‘Choujuuro’ (S\(_2\)S\(_3\)). The BAC library consisted of two sublibraries derived from two DNA size fractions. One sublibrary, which was derived from the 145–185 kb size fraction, consisted of 30 720 clones with an average insert size of 111 kb. The other sublibrary, which was derived from the 185–205 kb size fraction, consisted of 30 720 clones with an average insert size of 127 kb. The average insert size of the whole BAC library was 119 kb. The haploid genome size of pear is estimated to be 496–536 Mb (Arumuganathan and Earle, 1991). Therefore, the BAC library represented 14-fold genome coverage, giving a >99% theoretical probability of recovering any single-copy DNA sequences in the genome.

To construct a BAC contig around *S2-RNase*, chromosome walking was initiated from *S2-RNase*. PCR screening of the BAC library of ‘Choujuuro’ with an *S-RNase*-specific primer pair yielded 10 BAC clones containing *S2-RNase*: 2E10, 5B5, 13C10, 15E3, 21F7, 37F11, 41A10, 48F8, 53E3, and 57B1. These BAC clones were aligned by PCR analysis with primer pairs designed from each BAC-end sequence, and a first contig was constructed based on the insert size and restriction pattern of the BAC plasmids (Fig. 1B).

For chromosome walking, two non-repetitive and *S2*-haplotype specific primer pairs, 13C10-RP and 2E10-T7, were selected from the BAC-end primer pairs located at the outer ends of the first contig (Supplementary Table S1 at JXB online). PCR screening of the BAC library with 13C10-RP yielded two BAC clones (43D6 and 55A9) upstream of *S2-RNase*. PCR screening of the BAC library with 2E10-T7 yielded five BAC clones (3A1, 7F3, 10H7, 27C4, and 52C1) downstream of *S2-RNase*. Finally, chromosome walking from *S2-RNase* yielded a total of 17 BAC clones. These were aligned to construct a BAC contig of 391 kb spanning 166 kb upstream to 225 kb downstream of *S2-RNase* (Fig. 1B).

Table 1. Continued

| ORFs | Homologous protein | Species | Amino acid identity | Score (bits) | E-value* | Accession no. |
|------|-------------------|---------|---------------------|--------------|--------|---------------|
| ORF58 | RNase H family protein | Asparagus officinalis | 35/63 (55%) | 78.6 | 2e-18 | AC183436 |
| ORF59 | Retrotransposon protein | Oryza sativa | 67/121 (55%) | 137 | 1e-30 | DP000009 |
| ORF60 | Retrotransposon protein | Oryza sativa | 127/297 (42%) | 234 | 2e-59 | DP000009 |
| ORF61 | Retrotransposon gag protein | Asparagus officinalis | 37/129 (28%) | 74.7 | 1e-11 | AC183435 |
| ORF62 | Unknown protein | Arabidopsis thaliana | 51/74 (68%) | 102 | 6e-20 | AK117191 |
| ORF63 | Retrotransposon gag protein | Asparagus officinalis | 119/299 (39%) | 195 | 3e-47 | AC183435 |
| ORF64 | None | | | | | |
| ORF65 | Retrotransposon protein | Beta vulgaris | 35/44 (79%) | 76.3 | 2e-18 | EF101866 |
| ORF66 | Hypothetical protein | Vitis vinifera | 39/110 (35%) | 55.1 | 5e-06 | AM489256 |
| ORF67 | RNase H family protein | Asparagus officinalis | 58/121 (47%) | 115 | 2e-24 | AC183436 |
| ORF68 | Retrotransposon gag protein | Asparagus officinalis | 229/692 (33%) | 341 | 2e-91 | AC183435 |
| ORF69 | Hypothetical protein | Vitis vinifera | 68/227 (29%) | 84.7 | 3e-14 | AM451669 |
| ORF70 | None | | | | | |
| ORF71 | None | | | | | |
| ORF72 | Retrotransposon gag protein | Asparagus officinalis | 714/1299 (54%) | 1440 | 0.0 | AC183435 |
| ORF73 | Retrotransposon gag protein | Asparagus officinalis | 401/608 (65%) | 821 | 0.0 | AC183435 |
| ORF74 | None | | | | | |
| ORF75 | Integrase | Populus trichocarpa | 212/535 (39%) | 357 | 2e-96 | DG561718 |
| ORF76 | Retrotransposon gag protein | Asparagus officinalis | 167/506 (33%) | 223 | 1e-55 | AC183435 |
| ORF77 | None | | | | | |
| ORF78 | None | | | | | |
| ORF79 | (SFBB\(^{4d1–d2}\)) | Malus × domestica | 293/393 (74%) | 561 | 4e-158 | AB270796 |
| ORF80 | Retrotransposon protein | Oryza sativa | 175/487 (35%) | 225 | 9e-57 | DP000009 |
| ORF81 | None | | | | | |
| ORF82 | None | | | | | |
| ORF83 | Hypothetical protein | Vitis vinifera | 27/46 (58%) | 56.2 | 1e-06 | AM482339 |
| ORF84 | None | | | | | |
| ORF85 | Retrotransposon protein | Oryza sativa | 242/432 (56%) | 471 | 4e-130 | DP000009 |
| ORF86 | None | | | | | |
| ORF87 | None | | | | | |
| ORF88 | None | | | | | |
| ORF89 | None | | | | | |

* Significant similarity corresponds to an E-value < e\(^{-4}\).
Table 2. Open reading frames (ORFs) predicted by GENSCAN in the 378 kb region around S2-RNase

| ORFs | Homologous protein | Species | Amino acid identity | Score (bits) | E-value$^a$ | Accession no. |
|------|-------------------|---------|---------------------|--------------|-------------|---------------|
| ORF1 | Serine-threonine protein kinase | Ricinus communis | 262/411 (63%) | 506 | 2e-141 | EC974075 |
| ORF2 | Retrotransposon protein | Oryza sativa | 189/479 (39%) | 355 | 1e-95 | DP000011 |
| ORF3 | S2-locus F-box | Malus x domestica | 280/312 (89%) | 602 | 9e-170 | DO422811 |
| ORF4 | DNA glycosylase DEMETER | Arabidopsis thaliana | 351/1051 (33%) | 385 | 1e-110 | DQ332423 |
| ORF5 | None | None | None | None | None | None |
| ORF6 | DNA glycosylase | Populus trichocarpa | 196/291 (67%) | 365 | 4e-99 | CM000346 |
| ORF7 | (SFBB$^{2-u5}$)S2-u5 | Malus x domestica | 280/312 (98%) | 752 | 0.0 | AB270798 |
| ORF8 | DNA glycosylase | Oryza sativa | 369/394 (93%) | 747 | 0.0 | DP000010 |
| ORF9 | Retrotransposon protein | Oryza sativa | 78/248 (31%) | 104 | 2e-20 | DP000010 |
| ORF10 | Retrotransposon protein | Oryza sativa | 132/267 (49%) | 249 | 8e-64 | DP000010 |
| ORF11 | None | None | None | None | None | None |
| ORF12 | None | None | None | None | None | None |
| ORF13 | Retrotransposon protein | Oryza sativa | 189/479 (39%) | 355 | 1e-95 | DP000011 |
| ORF14 | GAG-POL precursor | Vitis vinifera | 65/208 (31%) | 105 | 2e-20 | AB111080 |
| ORF15 | None | None | None | None | None | None |
| ORF16 | (SFBB$^{2-u4}$)MdSFBB$^{3-b}$ | Malus x domestica | 369/394 (93%) | 735 | 0.0 | AB270792 |
| ORF17 | Retrotransposon protein | Oryza sativa | 332/820 (40%) | 570 | 4e-160 | DP000009 |
| ORF18 | Retrotransposon protein | Oryza sativa | 369/394 (93%) | 747 | 0.0 | DP000010 |
| ORF19 | Retrotransposon protein | Oryza sativa | 269/693 (38%) | 466 | 8e-129 | DP000009 |
| ORF20 | Retrotransposon protein | Oryza sativa | 369/394 (93%) | 747 | 0.0 | DP000010 |
| ORF21 | Retrotransposon protein | Oryza sativa | 369/394 (93%) | 747 | 0.0 | DP000010 |
| ORF22 | Retrotransposon protein | Oryza sativa | 78/248 (31%) | 104 | 2e-20 | DP000010 |
| ORF23 | None | None | None | None | None | None |
| ORF24 | Retrotransposon protein | Oryza sativa | 189/479 (39%) | 355 | 1e-95 | DP000011 |
| ORF25 | None | None | None | None | None | None |
| ORF26 | Retrotransposon protein | Oryza sativa | 189/479 (39%) | 355 | 1e-95 | DP000011 |
| ORF27 | None | None | None | None | None | None |
| ORF28 | Retrotransposon protein | Oryza sativa | 189/479 (39%) | 355 | 1e-95 | DP000011 |
| ORF29 | None | None | None | None | None | None |
| ORF30 | None | None | None | None | None | None |
| ORF31 | (SFBB$^{2-d4}$)S2-u4 | Malus x domestica | 369/394 (93%) | 735 | 0.0 | AB270792 |
| ORF32 | None | None | None | None | None | None |
| ORF33 | Hypothetical protein | Vitis vinifera | 69/195 (35%) | 88.2 | 2e-07 | AM423348 |
| ORF34 | None | None | None | None | None | None |
| ORF35 | Hypothetical protein | Vitis vinifera | 69/195 (35%) | 88.2 | 2e-07 | AM423348 |
| ORF36 | None | None | None | None | None | None |
| ORF37 | None | None | None | None | None | None |
| ORF38 | None | None | None | None | None | None |
| ORF39 | None | None | None | None | None | None |
| ORF40 | None | None | None | None | None | None |
| ORF41 | None | None | None | None | None | None |
| ORF42 | None | None | None | None | None | None |
| ORF43 | None | None | None | None | None | None |
| ORF44 | None | None | None | None | None | None |
| ORF45 | None | None | None | None | None | None |
| ORF46 | None | None | None | None | None | None |
| ORF47 | None | None | None | None | None | None |
| ORF48 | None | None | None | None | None | None |

$^a$Significant similarity corresponds to an E-value < e$^{-4}$. 

---

Downloaded from https://academic.oup.com/jxb/article-abstract/62/6/1887/59468 by guest on 30 July 2018
F-box protein genes around the S-RNase of Japanese pear
Sequence analysis of 378 kb around the S2-RNase gene

To identify the genes around S2-RNase, three overlapping BAC clones, 55A9, 48F8, and 10H7, were subcloned and completely sequenced (Fig. 1B). Sequence assembly of the three BAC clones yielded a 378 419 bp sequence. Analysis using GENSCAN software predicted 57 ORFs in the 378 kb region (Fig. 2B). A BLASTX search of these ORFs yielded 41 ORFs with significant similarity (E-value <e-4) to sequences of known proteins in the database (Table 2). ORF22 corresponded to S2-RNase. Among the 57 ORFs, 20 were similar to (retro) transposons and four were similar to a hypothetical protein. ORF1 was similar to a serine-threonine protein kinase, ORF4 to the DNA glycosylase DEMETER, ORF6 to a DNA glycosylase, ORF40 and ORF42 to TIR-NBS-LRR-type disease resistance proteins, and ORF53 to a cyclin-like F-box. Ten ORFs (ORF3, ORF7, ORF8, ORF16, ORF19, ORF24, ORF31, ORF43, ORF46, and ORF49) showed high sequence similarity to MdSFBB genes, MdSLF genes, PpSFBBa–d, or S-Fbox0. Using GENETYX-MAC Ver. 13 software, the predicted ORFs were reanalysed to determine the precise ORFs from the start (ATG) to the stop codon. ORF3, ORF7, ORF8, ORF16, ORF19, ORF24, ORF31, ORF43, ORF46, and ORF49 encoded 393, 396, 394, 392, 392, 394, 395, 400, 393, and 390 amino acid residues, respectively. A Pfam motif search predicted that these proteins had an F-box domain at the N-terminus and an FBA_1 domain in the centre (Fig. 3). When compared with reported PpSFBB, MdSFBB, and MdSLF genes, these ORFs showed pairwise deduced amino acid sequence identities ranging from 62.0% to 94.9% (Supplementary Table S3 at JXB online). The F-box protein genes differed from PpSFBBc–e, which was reported to be linked to S2-RNase (Kakui et al., 2007). Thus, these ORFs were assigned as new PpSFBB genes. ORF3, ORF7, ORF8, ORF16, and ORF19 were located ~146, ~108, ~102, ~53, and ~22 kb, respectively, upstream of S2-RNase. ORF24, ORF31, ORF43, ORF46, and ORF49 were located ~10, ~56, ~119, ~137, and ~158 kb, respectively, downstream of S2-RNase. These new PpSFBB genes upstream and downstream of S2-RNase were named PpSFBBa–u and PpSFBBa–d, respectively, and lower case numbers were assigned to the PpSFBBb–u and PpSFBBb–d located close to S2-RNase. Therefore, ORF19, ORF16, ORF8, ORF7, and ORF3 were designated as PpSFBBa–u, PpSFBBa–d, PpSFBBb–u, PpSFBBb–d, PpSFBBc–u, PpSFBBc–d, PpSFBBd–u, and PpSFBBd–d, and ORF24, ORF31, ORF43, ORF46, and ORF49 were designated as PpSFBBa–d, PpSFBBb–d, PpSFBBc–d, PpSFBBd–d, and PpSFBBd–d, respectively. These PpSFBB genes around S2-RNase shared variable transcriptional orientations (Fig. 1B). Using ATGC Ver. 4 software, the 378 kb S2 BAC contig sequence was searched for SFBB-like sequences. The analysis revealed two pseudogenes (ΨPpSFBBa–u1 and ΨPpSFBBb–u2) encoding truncated F-box proteins that were located ~46 kb upstream and ~159 kb upstream of S2-RNase, respectively (Fig. 1B). PpSFBBa–u2 and PpSFBBb–d1–d5 shared 66.3–86.0% amino acid sequence identity with each other, and showed 66.2–93.1% identity with PpSFBBa–u4 and PpSFBBb–d1–d5 (Table 3).

Total RNA was extracted from pollen, pistils, and leaves of the S2 homozygote to examine the expression of PpSFBBa–u1–u5 and PpSFBBb–d1–d5. RT-PCR analyses were conducted using gene-specific primer pairs (Supplementary Table S2 at JXB online). PpSFBBa–u1–u5 and PpSFBBb–d1–d5 were all specifically expressed in pollen, but not in pistils or leaves (Supplementary Fig. S1B). Thus, in the 378 kb sequence around S2-RNase there were 10 F-box protein genes (PpSFBBa–u1–u5 and PpSFBBb–d1–d5) with pollen-specific expression. The PpSFBBb–c gene, previously shown to be linked to the S2-RNase, was not within the sequenced region.

Comparison of deduced amino acid sequences between the PpSFBBa and PpSFBBb genes

The pairwise deduced amino acid sequence identities of nine PpSFBBa genes (PpSFBBa–u1–u4, PpSFBBa–d1–d2, and PpSFBBa–c1–c3) and 11 PpSFBBb genes (PpSFBBb–u1–u5, PpSFBBb–d1–d5, and PpSFBBb–e1–e5) were compared within and between haplotypes (Table 3). Sequence identity among the PpSFBBa genes ranged from 62.3% to 86.2%, and among PpSFBBb genes ranged from 63.0% to 86.0%. The PpSFBBa and PpSFBBb genes showed 62.1–99.0% identity between haplotypes. Identities of >90% were found between PpSFBBa–u2 and PpSFBBb–u1 (92.1%), PpSFBBa–u3 and PpSFBBb–u2 (93.1%), PpSFBBa–u4 and PpSFBBb–u4 (94.9%), and PpSFBBb–e1 and PpSFBBb–e1 (99.0%). Identities ranging from 80% to 90% were found between PpSFBBa–u2 and PpSFBBa–u1 (87.8%), PpSFBBa–u3 and PpSFBBa–u4 (84.9%), PpSFBBa–d1 and PpSFBBb–d1 (82.8%), PpSFBBb–d2 and PpSFBBb–d2 (84.2%), PpSFBBb–u1 and PpSFBBb–u5 (81.9%), and PpSFBBb–u2 and PpSFBBb–e1 (81.6%). The other 89 pairwise comparisons showed identities of <80%.

Phylogenetic analysis of the F-box protein genes of Pyrus and Malus

Most PpSFBBa and PpSFBBb genes cloned in this study shared the highest amino acid sequence identities with the F-box protein genes of Malus (MdSFBB and MdSLF genes), although PpSFBBa–u4 and PpSFBBb–u4 showed the highest identities with PpSFBBb–f4 (77.5%) and PpSFBBb–f4 (94.9%) derived from the same species, respectively.

Fig. 3. Alignment of deduced amino acid sequences of PpSFBBa–u4–u4, 4–d1–d2 and PpSFBBb–u1–u5, 2–d1–d5. Amino acid sequences were aligned using ClustalW. Conserved sites and relatively conserved sites are marked with asterisks and dots, respectively. F-box domains and FBA_1 domains of F-box proteins are coloured and underlined, respectively. Accession numbers for the F-box protein genes are as follows: PpSFBBa–u1–u4, 4–d1–d2 (AB545981) and PpSFBBb–u1–u5, 2–d1–d5 (AB545982).
Table 3. Pairwise amino acid sequence identities (%) of PpSFBB<sup>4</sup> and PpSFBB<sup>2</sup> genes

|       | PpSFBB<sup>4-u2</sup> | PpSFBB<sup>4-u3</sup> | PpSFBB<sup>4-u4</sup> | PpSFBB<sup>4-d1</sup> | PpSFBB<sup>4-d2</sup> | PpSFBB<sup>4-a</sup> | PpSFBB<sup>4-b</sup> | PpSFBB<sup>4-c</sup> | PpSFBB<sup>4-d3</sup> | PpSFBB<sup>4-d4</sup> | PpSFBB<sup>4-d5</sup> | PpSFBB<sup>4-u1</sup> | PpSFBB<sup>4-u2</sup> | PpSFBB<sup>4-u3</sup> | PpSFBB<sup>4-u4</sup> | PpSFBB<sup>4-d1</sup> | PpSFBB<sup>4-d2</sup> | PpSFBB<sup>4-a</sup> | PpSFBB<sup>4-b</sup> | PpSFBB<sup>4-c</sup> | PpSFBB<sup>4-d3</sup> | PpSFBB<sup>4-d4</sup> | PpSFBB<sup>4-d5</sup> | PpSFBB<sup>4-u1</sup> | PpSFBB<sup>4-u2</sup> | PpSFBB<sup>4-u3</sup> | PpSFBB<sup>4-u4</sup> | PpSFBB<sup>4-d1</sup> | PpSFBB<sup>4-d2</sup> | PpSFBB<sup>4-a</sup> | PpSFBB<sup>4-b</sup> | PpSFBB<sup>4-c</sup> | PpSFBB<sup>4-d3</sup> | PpSFBB<sup>4-d4</sup> | PpSFBB<sup>4-d5</sup> |
|-------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| PpSFBB<sup>4-u1</sup> | 71.4 | 70.6 | 72.7 | 73.1 | 71.8 | 69.6 | 71.0 | 66.3 | 72.3 | 73.6 | 73.3 | 71.2 | 74.0 | 72.8 | 76.3 | 72.6 | 76.3 | 71.5 | 66.3 |
| PpSFBB<sup>4-u2</sup> | – | 86.2 | 73.7 | 67.3 | 69.7 | 83.2 | 67.0 | 70.8 | 92.1 | 87.8 | 71.5 | 66.5 | 68.5 | 70.3 | 72.4 | 72.0 | 75.1 | 66.2 | 70.3 |
| PpSFBB<sup>4-u3</sup> | – | 70.3 | 67.2 | 69.5 | 80.5 | 67.2 | 67.4 | 84.9 | 93.1 | 70.1 | 66.9 | 68.8 | 70.8 | 70.6 | 72.4 | 73.7 | 67.1 | 66.9 |
| PpSFBB<sup>4-u4</sup> | – | 69.7 | 68.8 | 70.9 | 67.9 | 77.5 | 71.2 | 73.2 | 68.4 | 67.1 | 70.4 | 70.2 | 70.4 | 69.7 | 72.7 | 69.4 | 77.3 |
| PpSFBB<sup>4-d1</sup> | – | 71.2 | 70.7 | 69.6 | 63.6 | 68.7 | 68.7 | 68.8 | 68.6 | 72.5 | 73.9 | 72.3 | 82.8 | 74.3 | 73.1 | 63.6 |
| PpSFBB<sup>4-d2</sup> | – | 69.2 | 67.9 | 63.7 | 67.6 | 70.7 | 69.7 | 68.1 | 71.2 | 77.4 | 73.0 | 74.8 | 84.2 | 73.6 | 63.4 |
| PpSFBB<sup>4-x</sup> | – | 68.7 | 68.2 | 81.9 | 81.6 | 70.2 | 68.7 | 70.7 | 69.4 | 70.5 | 72.5 | 73.0 | 69.2 | 67.7 |
| PpSFBB<sup>4-b</sup> | – | 62.3 | 66.6 | 68.4 | 68.8 | 94.9 | 69.2 | 71.8 | 69.0 | 71.3 | 73.0 | 70.5 | 62.1 |
| PpSFBB<sup>4-d3</sup> | – | 68.2 | 70.0 | 64.9 | 62.5 | 65.5 | 65.6 | 65.1 | 65.1 | 69.2 | 65.5 | 99.0 |
| PpSFBB<sup>4-u1</sup> | – | 86.0 | 69.7 | 66.3 | 69.7 | 70.5 | 72.0 | 70.5 | 73.3 | 67.9 | 67.7 |
| PpSFBB<sup>2-u2</sup> | – | 72.3 | 74.0 | 72.8 | 74.3 | 71.8 | 65.5 |
| PpSFBB<sup>2-u3</sup> | – | 69.5 | 70.7 | 74.1 | 71.1 | 72.3 | 73.8 | 71.0 | 64.9 |
| PpSFBB<sup>2-u4</sup> | – | 68.4 | 70.8 | 69.0 | 70.9 | 72.5 | 70.0 | 63.0 |
| PpSFBB<sup>2-u5</sup> | – | 71.0 | 74.0 | 72.8 | 74.3 | 71.8 | 65.5 |
| PpSFBB<sup>2-d1</sup> | – | 72.8 | 75.4 | 81.2 | 79.0 | 65.4 |
| PpSFBB<sup>2-d2</sup> | – | 74.9 | 77.8 | 73.6 | 64.3 |
| PpSFBB<sup>2-d3</sup> | – | 77.1 | 73.6 | 65.1 |
| PpSFBB<sup>2-d4</sup> | – | 78.7 | 69.4 |
| PpSFBB<sup>2-d5</sup> | – | 65.3 |

Values >90% are shown in bold.
Supplementary Table S3 at *JXB* online. The deduced amino acid sequences of the 36 F-box protein genes of *Pyrus* and *Malus* were aligned with *PmSLFS* of *P. mume* using ClustalW, and a rooted phylogenetic tree was constructed by the Neighbor–Joining method with *PmSLFS* as an outgroup (Fig. 4). F-box protein genes of *Pyrus* and *Malus* did not form taxa-independent clusters, and several *PpSFBB* genes were positioned closest to *MdSFBB* and *MdSLF* genes. The F-box protein genes of *Pyrus* and *Malus* were grouped into two major groups: group I (84% bootstrap value) and group II (91% bootstrap value). Group I included *PpSFBB*4-u1–d2, *PpSFBB*2-u3–u5, *PpSFBB*-b genes, *MdSFBB*3 genes, and *MdSLF* genes, while group II included *PpSFBB*4-u2–u4, *PpSFBB*2-u1–u2, *PpSFBB*-a genes, and *MdSFBB*9 genes. Comparing group I with group II, amino acid sequences were conserved in F-box domains, but were divergent in the five regions designated as R1, R2, R3, R4, and R5. In these regions, sequences and/or
insertions/deletions (indels) were relatively conserved within each group (Fig. 3).

**Discussion**

The results of a previous study suggested that the pollen S-allele is distal to the region from 48 kb upstream to 188 kb downstream of S-RNase (Okada et al., 2008). In this study, the BAC contig around S-RNase was extended to 659 kb, and a 648 516 bp region spanning 290 kb upstream to 359 kb downstream of S-RNase was sequenced. Sequence analysis of the 649 kb region predicted five new pollen-specific F-box protein genes (PpSFBB\textsuperscript{a1–a4} and PpSFBB\textsuperscript{d1–d2}). The 649 kb sequence around S-RNase included six PpSFBB\textsuperscript{a1–a4} genes including PpSFBB\textsuperscript{d1–d2} (S-F-box0), but not PpSFBB\textsuperscript{d1–d2}. In addition, a BAC library was constructed from ‘Choujuuro’ (S2S3), and a BAC contig of 391 kb around S-RNase was assembled. Sequence analysis of a 378 419 bp region spanning 166 kb upstream to 212 kb downstream of S-RNase predicted 10 new pollen-specific F-box protein genes (PpSFBB\textsuperscript{a1–a5}, 2-d1–d5). The 378 kb sequence around S-RNase included 10 PpSFBB\textsuperscript{a1–a5} genes, but not PpSFBB\textsuperscript{d1–d2}. The predicted products of PpSFBB\textsuperscript{a1–a5}, 2-d1–d5 showed typical features of F-box proteins: an F-box domain at the N-terminus and an FBA\textsubscript{1} domain in the centre (Fig. 3). These results indicated that F-box protein genes with pollen-specific expression are clustered around the S-RNase region in Japanese pear, and that PpSFBB\textsuperscript{a1–a5} and PpSFBB\textsuperscript{d1–d2}, which are linked to the S-RNase, were located outside the sequenced region.

**Organization of the F-box protein gene cluster around the S-RNase gene of Japanese pear**

Among PpSFBB\textsuperscript{a1–a5}, 2-d1–d5 and PpSFBB\textsuperscript{d1–d2}, some genes may be located more distantly from S-locus regions. Entani et al. (2003) conducted pattern matching analysis of homologies (Harr plot analysis) for the sequences around PmS\textsubscript{T} and PmS\textsubscript{RNases} of P. mume. Their results revealed that highly divergent S\textsubscript{T} and S\textsubscript{RNases} regions are surrounded by co-linear flanking regions, and that S\textsubscript{T} and S\textsubscript{RNases} locus regions are ~27 kb and 15 kb long, respectively. Harr plot analysis of the 649 kb and 378 kb sequences around S\textsubscript{T} and S\textsubscript{RNases} was conducted, and no co-linearity was found between these sequences (data not shown). This result suggests that both the 649 kb and 378 kb sequences are a part of the S-locus region, or that either sequence could contain both the S-locus region and its flanking region. Sequence analysis of the 649 kb and 378 kb regions predicted 40 and 20 transposon-like sequences around S\textsubscript{T}-RNase and S\textsubscript{RNase}, respectively (Tables 1, 2). The S-locus, which controls S-RNase-based GSI, contains many transposon-like sequences. For example, transposon-like sequences were found in three out of 12 ORFs in 72 kb of the P. dules S\textsubscript{Y}-haplotype (Ushijima et al., 2003), in four out of 11 ORFs in 64 kb of the A. hispanicum S\textsubscript{Y}-haplotype (Lai et al., 2002), and in 31 out of 50 ORFs in 328 kb of the P. inflata S\textsubscript{Y}-haplotype (Wang et al., 2004). These transposon-like sequences generate polymorphisms among S-haplotypes, and might contribute to suppression of recombination between S-RNase and SLF/SFB. In the sequenced regions around S\textsubscript{T}-RNase and S\textsubscript{RNase}, the non-co-linearity, the abundant (retro) transposon insertions, and the absence of PpSFBB\textsuperscript{d1–d2} suggest that the 649 kb and 378 kb sequences around S\textsubscript{T}-RNase and S\textsubscript{RNase} are part of the S-locus region, and that the S-locus regions of the Japanese pear are probably larger than those of Prunus species.

The organization of the F-box protein gene clusters around S\textsubscript{T}-RNase and S\textsubscript{RNase} was compared when S\textsubscript{T}-RNase and S\textsubscript{RNase} were fixed in the same transcriptional orientation (Fig. 1). PpSFBB\textsuperscript{a1–a5} and PpSFBB\textsuperscript{d1–d2} are located ~113 kb upstream and ~127 kb downstream of S\textsubscript{T}-RNase, whereas PpSFBB\textsuperscript{a1–a5} and PpSFBB\textsuperscript{d1–d2} are located close to S\textsubscript{T}-RNase (~22 kb upstream and ~10 kb downstream of S\textsubscript{T}-RNase, respectively). The average densities of F-box protein genes were one gene/108 kb around S\textsubscript{T}-RNase and one gene/38 kb around S\textsubscript{RNase}. Together, these results suggest that F-box protein genes are clustered in the region around S\textsubscript{T}-RNase more tightly than in the region around S\textsubscript{RNase}.

F-box protein genes, SLF/SFB and SLF-like genes (SLFL), were identified in cosmid and fosmid contigs around the S-RNase of Prunus species. SLF/SFB genes are the pollen S genes, but SLFL genes are probably not involved in SI recognition (Entani et al., 2003; Ushijima et al., 2003). SLF/SFB and SLFL1–SLFL3 cloned from the same haplotypes show low amino acid sequence identity with each other. For example, PmSLFL\textsubscript{S} is 11.7–16.9% identical to PmSLFL1\textsubscript{S}, PmSLFL2\textsubscript{S}, and PmSLFL3\textsubscript{S}, which share 26.9–45.3% identity with each other (Entani et al., 2003; Matsumoto et al., 2008). PdSFBB\textsuperscript{a1–a5} and PdSFBB\textsuperscript{d1–d2} are 18.7 and 20.2% identical to PdSLF\textsuperscript{a1–a5} and PdSLF\textsuperscript{d1–d2} (orthologues of PmSLFL1 of P. mume), respectively (Ushijima et al., 2003). In contrast to Prunus species, PpSFBB\textsuperscript{a1–a5}, 2-d1–d5 shared 67.2–86.2% and 66.3–86.0% identity within each haplotype, respectively (Table 3). This indicates that the region around an S-RNase of the Japanese pear comprises related F-box protein genes, which is different from the F-box protein gene organization around the S-RNases in Prunus, in which there are clusters of F-box protein genes that show low levels of identity to each other. The amino acid sequence identities between PpSFBB\textsuperscript{a1–a5}, 4-d1–d2 and PpSFBB\textsuperscript{d1–d2} and PpSFBB\textsuperscript{a1–a5}, 2-d1–d5 ranged from 66.2% to 93.1% (Table 3), and were higher than those within each haplotype (66.3–86.2%). These similarities between haplotypes indicate that related polymorphic F-box protein genes between haplotypes were clustered in the regions around S\textsubscript{T}-RNase and S\textsubscript{RNase}.

**Classification of PpSFBB genes based on phylogenetic analysis and sequence polymorphism**

In Prunus species, F-box protein genes around S-RNase genes were grouped into two major classes, the SLF/SFB clade and the SLFL clade, by a phylogenetic analysis.
On 30 July 2018

PpSFBB genes show lower levels of allelic sequence identity (77.8–81.3% for PmSLF genes, 68.4–76.4% for PdSFB genes, and 75.1–81.1% for PmSFBB genes, respectively) than SLFL genes (88.5–92.0% for PmSLFL1, 95.8–98.6% for PmSLFL2, and 95.1% for PdSLF) (Entani et al., 2003; Ushijima et al., 2003; Ikeda et al., 2004; Matsumoto et al., 2008). The sequence differences of the F-box protein genes among haplotypes implied that SLF/SFB genes with lower levels of identity were pollen S candidates, and that SLFL genes with high levels of identity were not (Entani et al., 2003; Ushijima et al., 2003).

The phylogenetic relationships and sequence differences of F-box protein genes of *Pyrus* and *Malus* would be useful for delineating pollen S candidates from *PpSFBB* and *PpSFBB*-like genes (Matsumoto et al., 2008). Phylogenetic analysis based on the deduced amino acid sequences of 36 F-box protein genes of *Pyrus* and *Malus* allowed them to be classified into two major groups, I and II (Fig. 4). *PpSFBB* genes, *PpSFBB*-like genes, and *MdSFB*-type genes were classified into group II, and the other *PpSFB*, *MdSFB*-type, and *MdSLF* genes were in group I. The phylogenetic analysis also generated a *PpSFBB*-like subgroup and 10 gene pairs of *PpSFBB* genes and *PpSFBB/Malus* F-box protein genes. Sequence identities between the paired genes ranged from 76.3% to 96.4% (Table 3, Supplementary Table S3) (Entani et al., 2003; Ushijima et al., 2003; Zhou et al., 2003; Wang et al., 2004; Hua et al., 2007). Therefore, several *PpSFBB* genes in a particular haplotype are probably not pollen S genes. The non-pollen S proteins of *P. inflata*, *P. sinensis*, and *P. suaveolens* either fail to interact with S-RNase or interact much more weakly than PISLFL. When the deduced amino acid sequences of PISLFL and PISLFLs were compared, three PISLFL-specific regions (SR1, SR2, and SR3) that confer on PISLFL its unique function in SI were revealed (Hua et al., 2007). Although the interactions of PpSFBBs with S-RNase have not yet been analysed, five regions (R1, R2, R3, R4, and R5) were identified where amino acid sequences were variable between the group I and II F-box protein genes (Fig. 3). The sequence differences in these regions might account for different interactions with S-RNase between the group I and II F-box protein genes. Therefore, there remains the possibility that the less polymorphic group II F-box protein genes are non-pollen S genes.

**Supplementary data**

Supplementary data are available at *JXB* online.

Figure S1. Expression of *PpSFBB* genes located around *S*<sub>r</sub>-RNase (A) and *S*<sub>r</sub>-RNase (B).

Table S1. Primer pairs used to construct BAC contigs around *S*<sub>r</sub>-RNase.

Table S2. Gene-specific RT-PCR primer pairs.

Table S3. Pairwise amino acid sequence identities (%) of *PpSFBB* genes with previously reported *PpSFBB*, *MdSFB*, and *MdSLF* genes.

**Acknowledgements**

This work was supported in part by Grants-in-Aid for Scientific Research (B) (nos 16380028; 19380020) and (C) (no. 17580027) from the Ministry of Education, Culture, Sports, Science and Technology, Japan. K.O. was supported by research fellowships of the Japan Society for the Promotion of Science for Young Scientists.
Antirrhinum is expressed specifically in pollen and tapetum. Plant Molecular Biology 50, 29–42.
Lechner E, Achard P, Vansiri A, Potuschak T, Genschik P. 2006. F-box proteins everywhere. Current Opinion in Plant Biology 9, 631–638.
Matsumoto D, Yamane H, Tao R. 2008. Characterization of SLFL1, a pollen-expressed F-box gene located in the Prunus S locus. Sexual Plant Reproduction 21, 113–121.
McClure BA, Franklin-Tong V. 2006. Gametophytic self-incompatibility: understanding the cellular mechanisms involved in ‘self’ pollen tube inhibition. Planta 224, 233–245.
McClure BA, Gray JE, Anderson MA, Clarke AE. 1990. Self-incompatibility in Nicotiana alata involves degradation of pollen ribosomal RNA. Nature 347, 757–760.
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, 955–957.
McCubbin AG, Kao TH. 2000. Molecular recognition and response in pollen and pistil interactions. Annual Review of Cell and Developmental Biology 16, 333–364.
Newbiggin E, Paape T, Kohn JR. 2008. RNase-based self-incompatibility: puzzled by pollen S. The Plant Cell 20, 2286–2292.
Okada K, Tonaka N, Moriya Y, Norioka N, Sawamura Y, Matsumoto T, Nakaneishi T, Takasaki-Yasuda T. 2008. Deletion of a 236 kb region around S4-RNase in a stylet mutant S4sm-haplotype of Japanese pear. Plant Molecular Biology 66, 389–400.
Potter D, Eriksson T, Evans RC, et al. 2007. Phylogeny and classification of Rosaceae. Plant Systematics and Evolution 266, 5–43.
Saito T, Sato Y, Sawamura Y, Shoda M, Kotobuki K. 2002. Studies on breeding of self-compatible Japanese pear. 2. Characteristic of pollen of S4tm gene originated from ‘Osa-nijisseiki’. Journal of the Japanese Society for Horticultural Science 71, suppl 2), 123(in Japanese).
Saitou N, Nei M. 1987. The neighbor–joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4, 406–425.
Sassa H, Kaku H, Miyamoto M, Suzuki Y, Hanada T, Ushijima K, Kusaba M, Hirano H, Koba T. 2007. S locus F-box brothers: multiple and pollen-specific F-box genes with S haplotype-specific polymorphisms in apple and Japanese pear. Genetics 175, 1869–1881.
Sijacic P, Wang X, Skirpan AL, Wang Y, Dowd PE, McCubbin AG, Huang S, Kao TH. 2004. Identification of the pollen determinant of S-RNase-mediated self-incompatibility. Nature 429, 302–305.
Sato Y. 1993. Breeding of self-compatible Japanese pear. In: Hayashi T, Omura M, Scott NS, eds. Techniques on gene diagnosis and breeding in fruit trees. 241–247 Tsukuba, Japan: FTRS (Fruit Tree Research Station).
Sonneveld T, Tobutt KR, Vaughan SP, Robbins TP. 2005. Loss of pollen S function in two self-compatible selections of Prunus avium is associated with deletion/mutation of an S haplotype-specific F-box gene. The Plant Cell 17, 57–51.
Terai O, Sato Y, Saito T, Abe K, Kotobuki K. 1999. Identification of homozygotes of self-incompatibility gene (S-gene), as useful tools to determine the S-genotype in Japanese pear, Pyrus pyrifolia Nakai. Bulletin of the National Institute of Fruit Tree Science 32, 31–38.

Tsukamoto T, Hauck NR, Tao R, Jiang N, lezziyi AF. 2006. Molecular characterization of three non-functional S-haplotypes in sour cherry (Prunus cerasus). Plant Molecular Biology 62, 371–383.

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. The Plant Cell 15, 771–781.

Ushijima K, Yamane H, Watari A, Kakehi E, Ikeda K, Hauck NR, lezziyi AF, Tao RT. 2004. The S haplotype-specific F-box protein gene, SFB, is defective in self-compatible haplotypes of Prunus avium and P. mume. The Plant Journal 39, 573–586.

Vilanova S, Badenes ML, Burgos L, Martinez-Calvo J, Llacer G, Romero C. 2006. Self-compatibility of two apricot selections is associated with two pollen-part mutations of different nature. Plant Physiology 142, 629–641.

Wang Y, Tsukamoto T, Yi KW, Wang X, Huang SS, McCubbin AG, Kao TH. 2004. Chromosome walking in the Petunia inflata self-incompatibility (S) locus and gene identification in an 881 kb contig containing S2-RNase. Plant Molecular Biology 54, 727–742.

Xue YB, Carpenter R, Dickinson HG, Coen ES. 1996. Origin of allelic diversity in antirrhinum S locus RNases. The Plant Cell 8, 805–814.

Zhou JL, Wang F, Ma WS, Zhang YS, Han B, Xue YB. 2003. Structural and transcriptional analysis of S-locus F-box genes in Antirrhinum. Sexual Plant Reproduction 16, 165–177.