An integrative evolutionary–genomic analysis reveals the factors that shape the sexual diversity and molecular specificity of gametophytic self-incompatibility in *Prunus species*

Goonetilleke et al

Supplementary figures

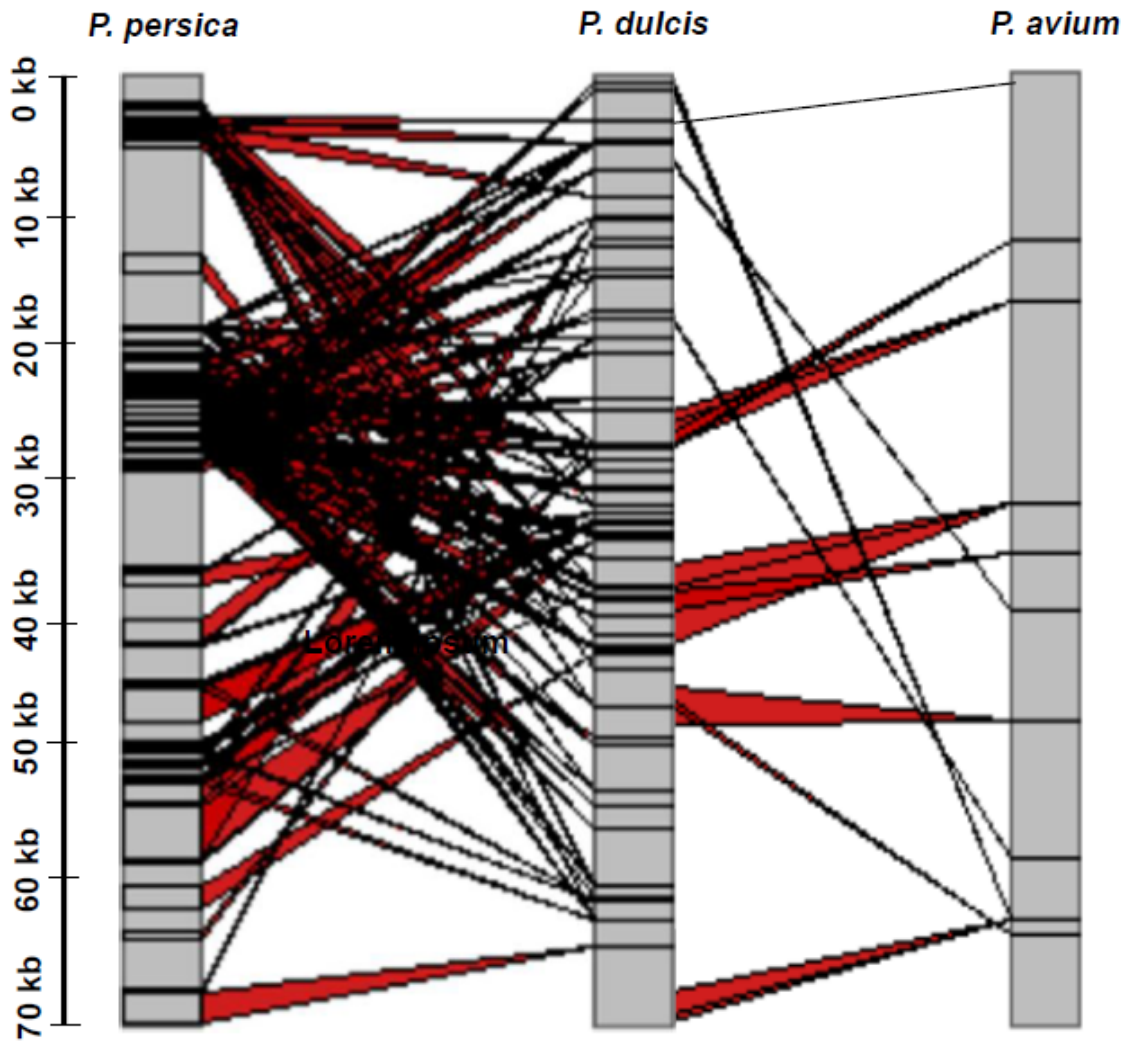


Fig. S1 Comparison of the *P. dulcis* (centre) S-locus with those of *P. persica* (left) and *P. avium* (right). Regions with sequence identity ≥90% are indicated by red lines. The horizontal bar on the left denotes the size of the *S-loci*

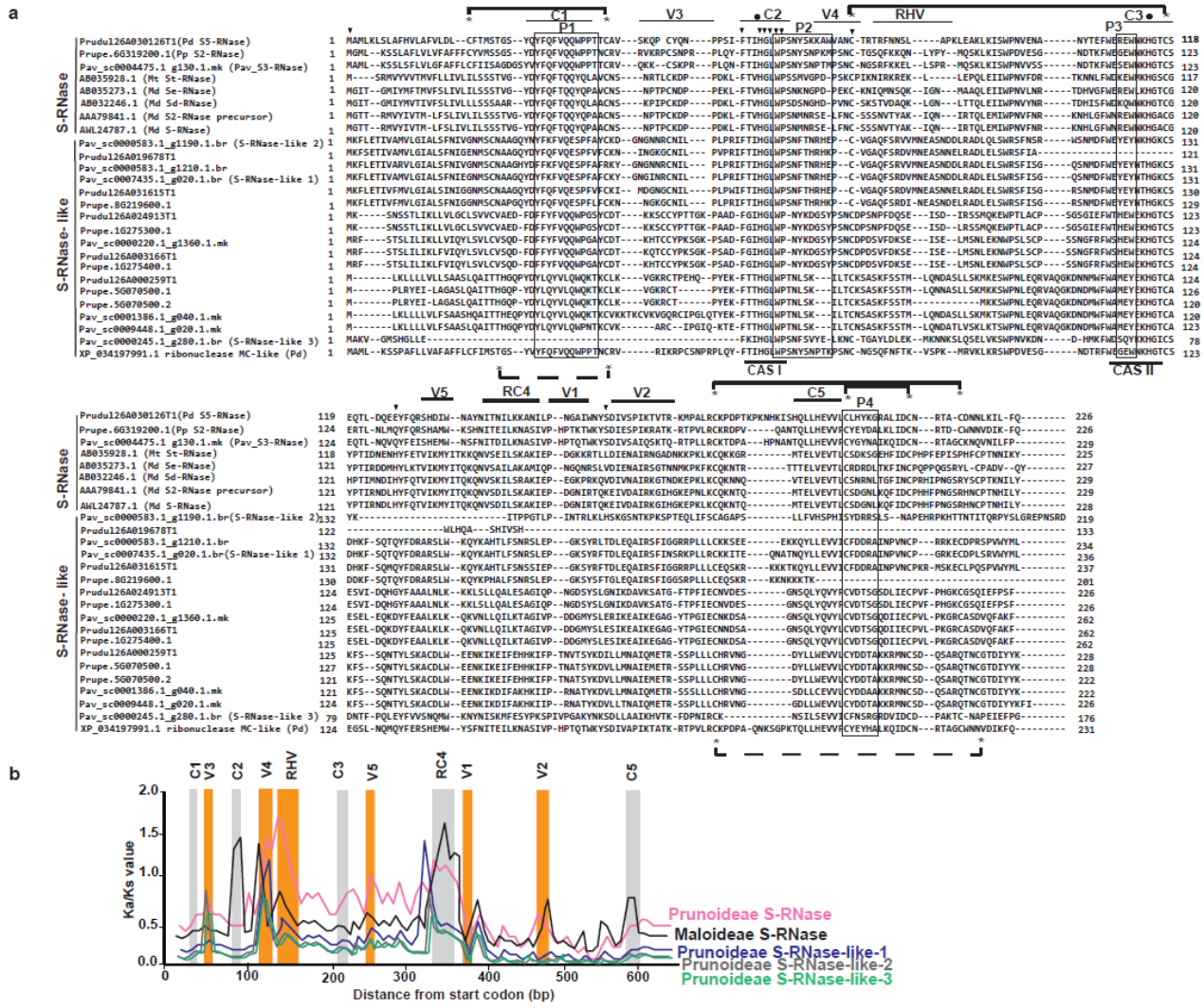


Fig. S2 Sequence alignment and selective pressure analysis of S-RNase and S-RNase-like genes in Rosaceae using *Prunus* and *Malus*. (a) Alignment showing conserved regions (C1–C5 and RC4), variable regions (V1–V5), the hypervariable region (RHV), and conserved active segments (CAS I and CAS II). Absolutely conserved residues, conserved histidine residues, and conserved cysteine residues are indicated by arrowheads, circles, and asterisks, respectively. Predicted disulphide bridges in S-RNases are shown as solid lines, while variation in cysteine positions and disulphide bridges in S-RNase-like proteins are shown as dashed lines. T2-RNase amino acid patterns (P1–P4) are labelled following Green (1994) and Vieira et al. (2008). Five *Malus* S-RNases were included for comparative analysis. (b) Sliding-window Ka/Ks analysis calculated using 100-bp windows with a 20-bp step size.

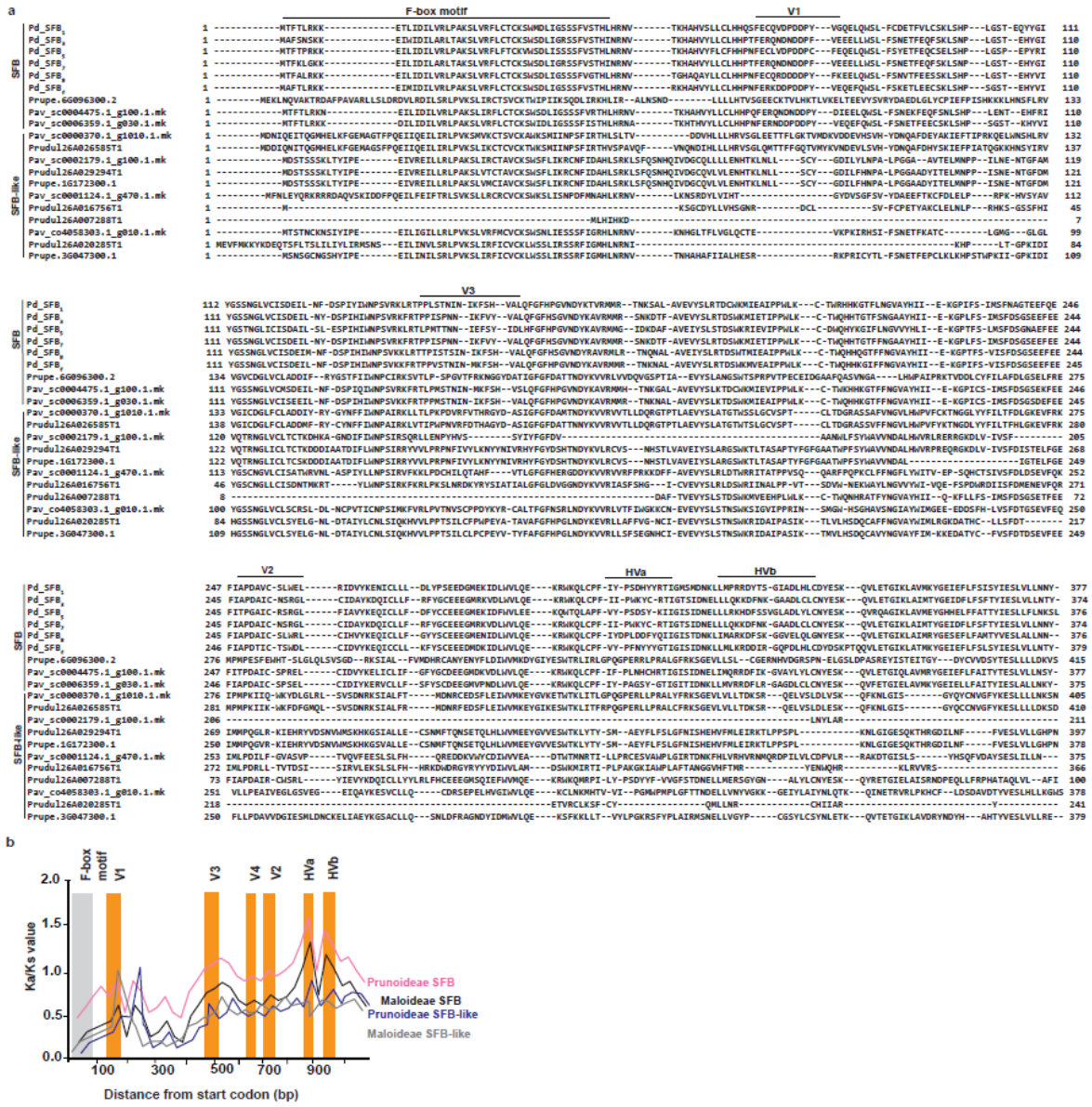


Fig. S3 Sequence alignment and selective pressure analysis of SFB and SFB-like genes in Rosaceae using *Prunus* and *Malus*. (a) Alignment showing the F-box motif, variable regions (V1–V4), and hypervariable regions (HVa and HVb) identified in SFB and SFB-like proteins. (b) Sliding-window Ka/Ks analysis calculated using 100-bp windows with a 20-bp step size

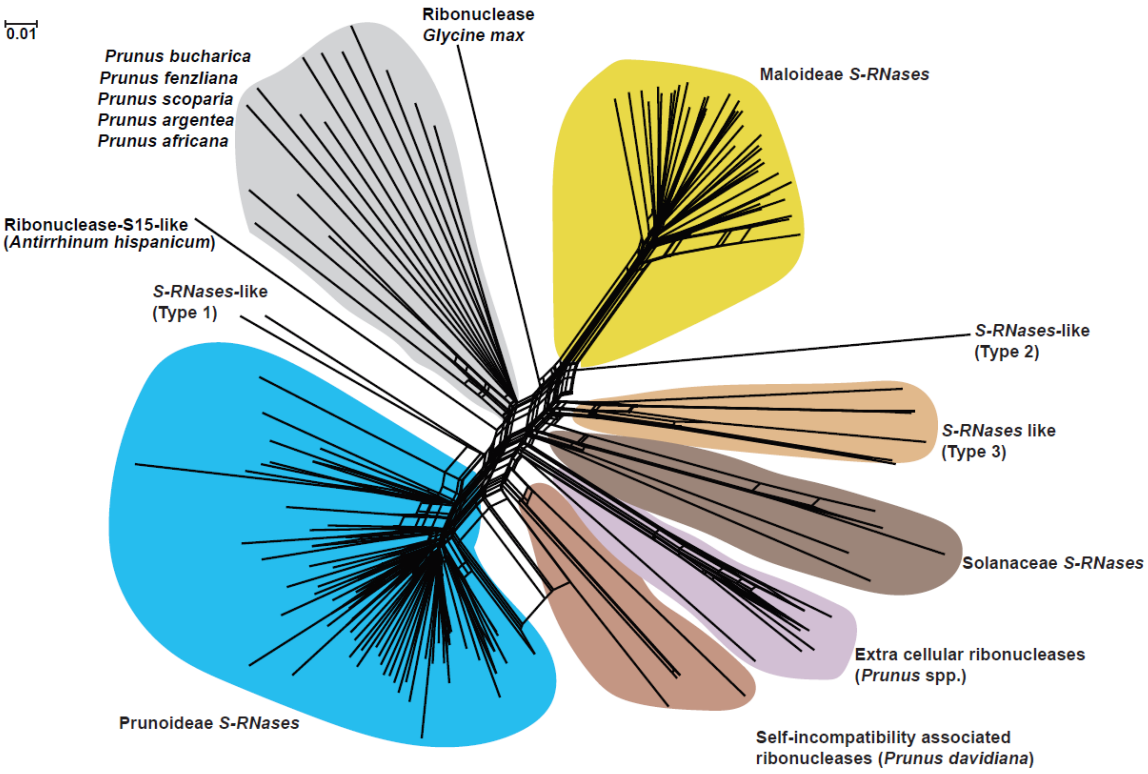


Fig. S4 Neighbor-Net analysis of 154 S-RNase and S-RNase-like sequences. Prunoideae and Maloideae S-RNases are highlighted in blue and yellow, respectively. Type I, Type II, and Type III S-RNase-like sequences and extracellular ribonucleases are also indicated. Cross-links represent alternative possible topologies among groups

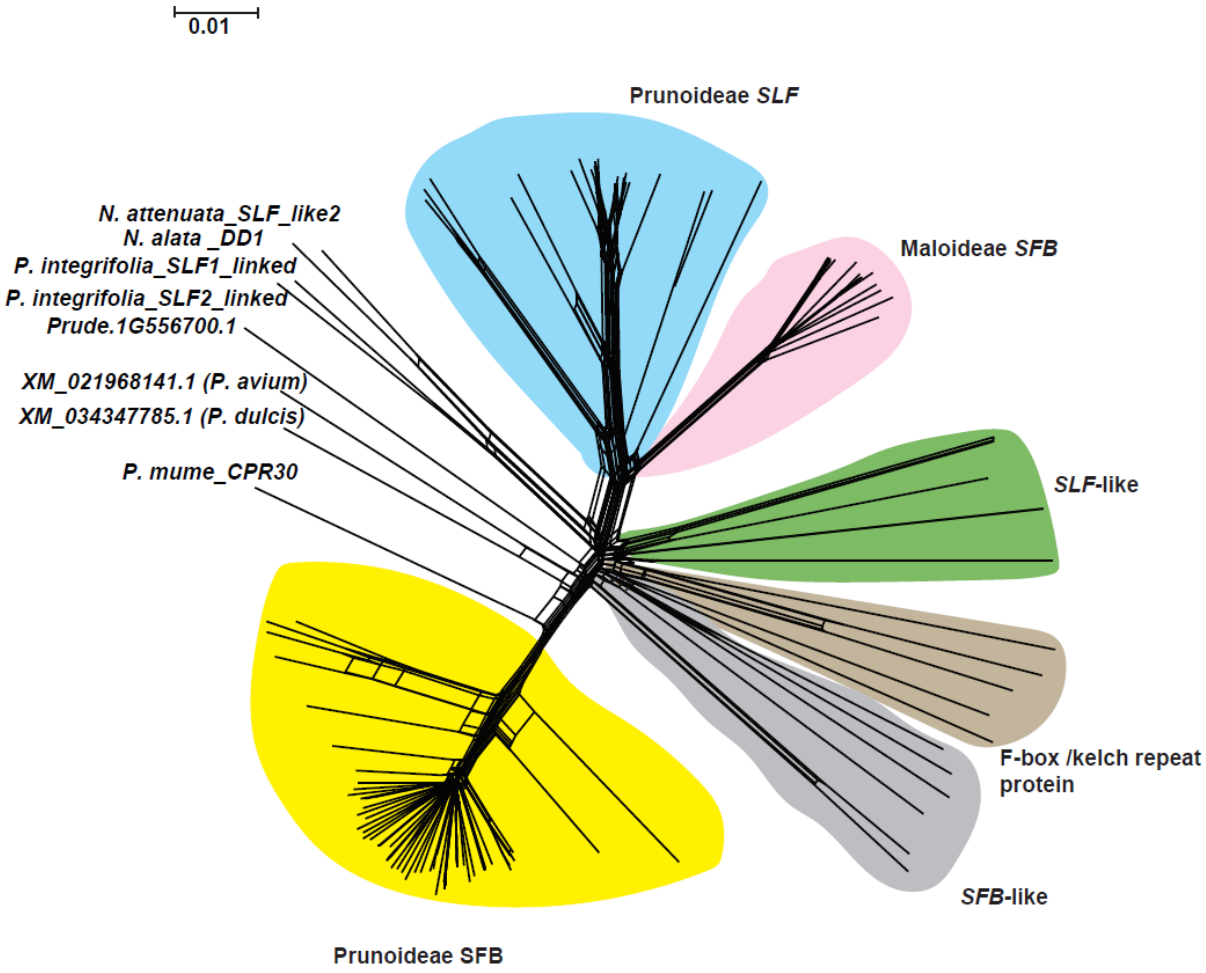


Fig. S5 Neighbor-Net analysis of 113 SLF, SFB, SLF-like, SFB-like, and F-box–kelch repeat sequences. Prunoideae SLFs are highlighted in blue, while Prunoideae and Maloideae SFBs are highlighted in yellow and pink, respectively. SLF-like, SFB-like, and F-box–kelch repeat sequences are shown in green, ash, and light brown, respectively. Cross-links represent alternative possible topologies among groups.

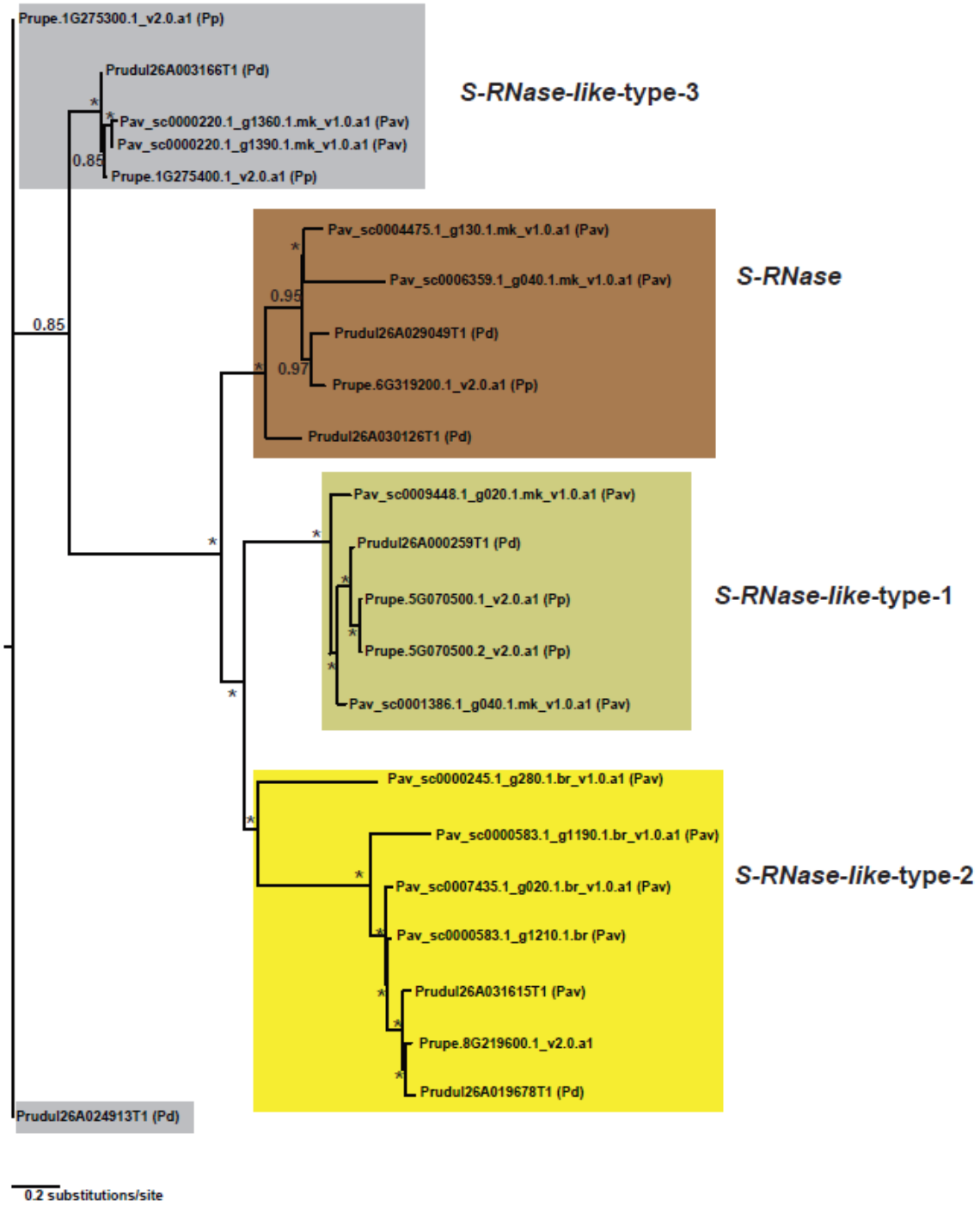


Fig. S6 Bayesian phylogenetic tree of 23 S-RNase and S-RNase-like sequences from sweet almond (*P. dulcis*, Pd), peach (*P. persica*, Pp), and sweet cherry (*P. avium*, Pav) inferred using MrBayes. Node support is shown as Bayesian posterior probabilities (PP), with an asterisk indicating PP = 1.0. S-RNase, S-RNase-like type I, type II, and type III clades are shown in brown, olive, yellow, and ash, respectively

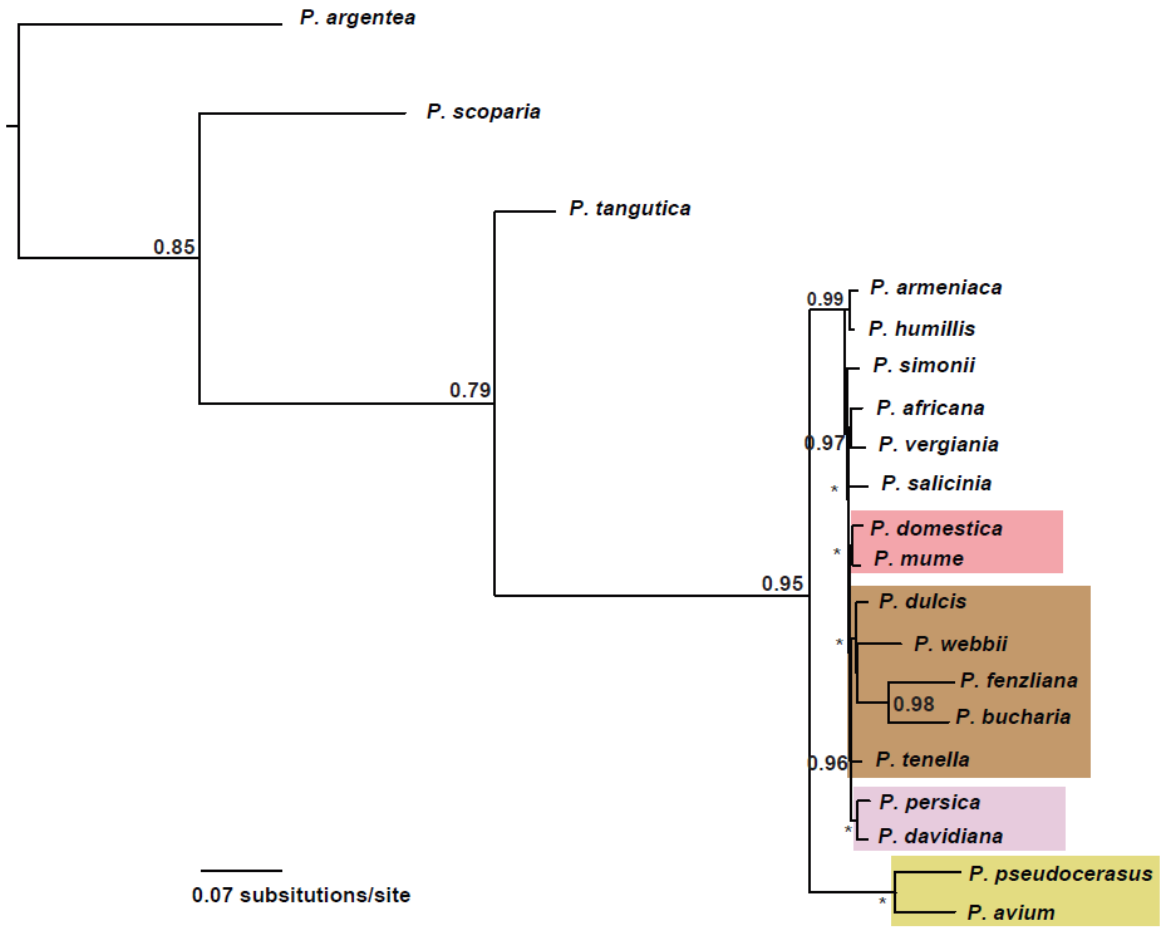


Fig. S7 Bayesian phylogenetic tree of 20 selected S-RNase alleles from Prunoideae inferred using MrBayes. Node support is shown as Bayesian posterior probabilities (PP), with an asterisk indicating PP = 1.0. Phylogenetic relationships within the genus *Prunus* and inferred ancestry of commercially important species are indicated, including Japanese plum (*P. mume*; salmon pink), sweet almond (*P. dulcis*; brown), peach (*P. persica*; light pink), and sweet cherry (*P. avium*; yellow).

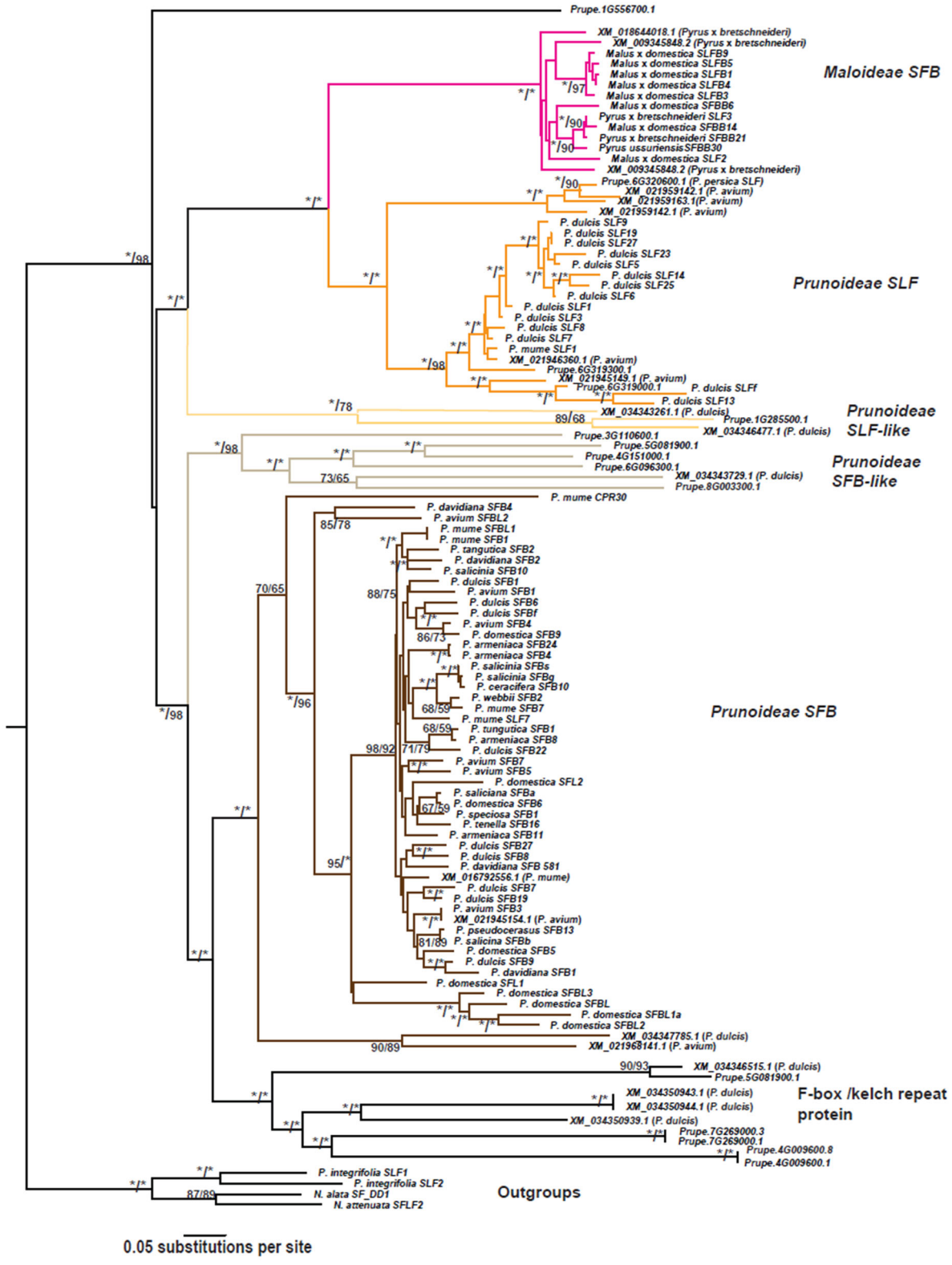


Fig. S8 Bayesian phylogenetic tree of 113 SFB, SLF, SFB-like, SLF-like, and F-box–kelch repeat sequences from Prunoideae and Maloideae inferred using MrBayes. Node support is indicated by Bayesian posterior probabilities (PP) and maximum likelihood bootstrap values (BS); an asterisk denotes PP = 1.0 or BS = 100%.

Branches corresponding to Maloideae SFB, Prunoideae SFB, SLF, SFB-like, SLF-like, and F-box–kelch repeat clades are shown in pink, dark brown, orange, light brown, yellow, and black, respectively.

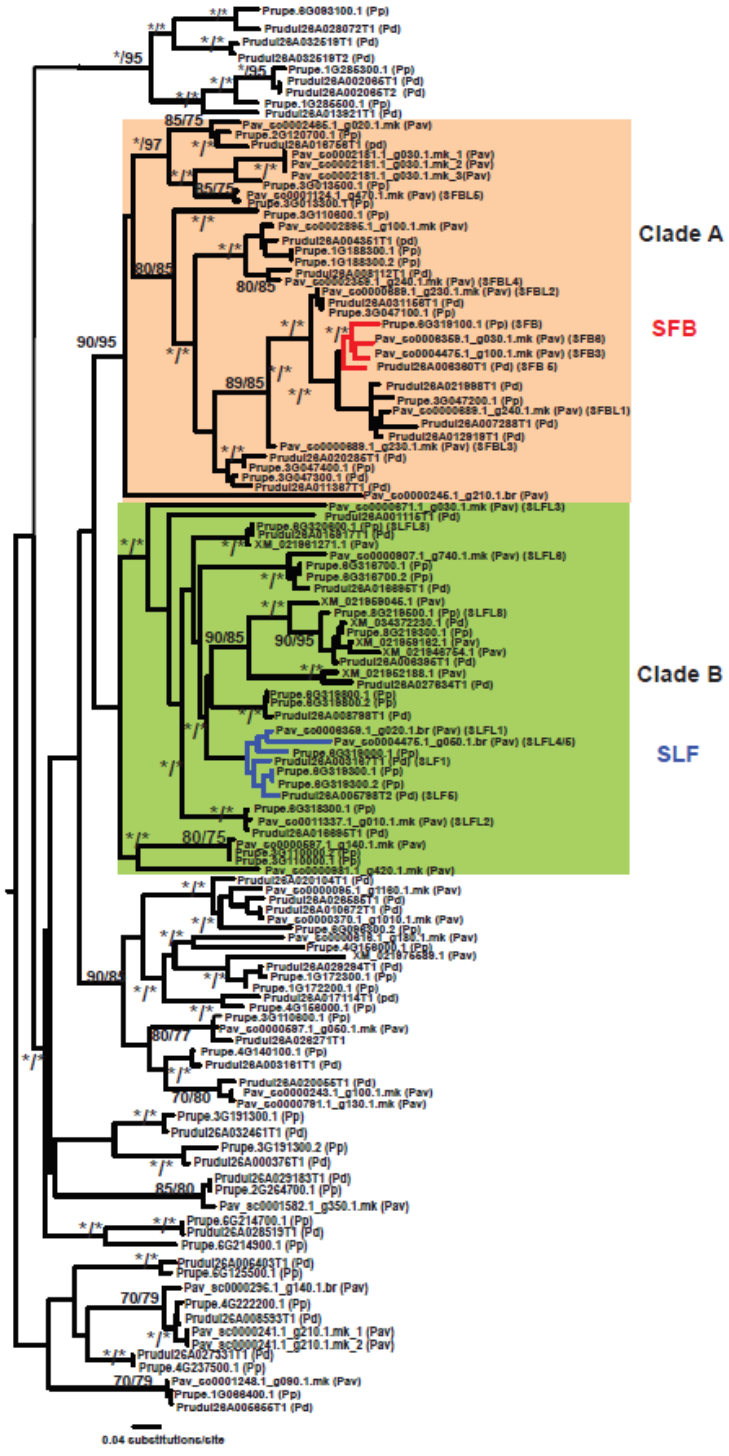


Fig. S9 Bayesian phylogenetic tree of 121 SLF, SLF-like, SFB, and SFB-like sequences from the genomes of *P. persica* (Pp), *P. dulcis* (Pd), and *P. avium* (Pav) inferred using MrBayes. Node support is shown as Bayesian posterior probabilities (PP) and maximum likelihood bootstrap values (BS); an asterisk denotes PP = 1.0 or BS = 100%. Clades corresponding to S-haplotype–specific F-box genes (SFB; red) and S-locus F-box genes (SLF; blue) in *Prunus* genomes are labelled as clades A (orange) and B (green), respectively.

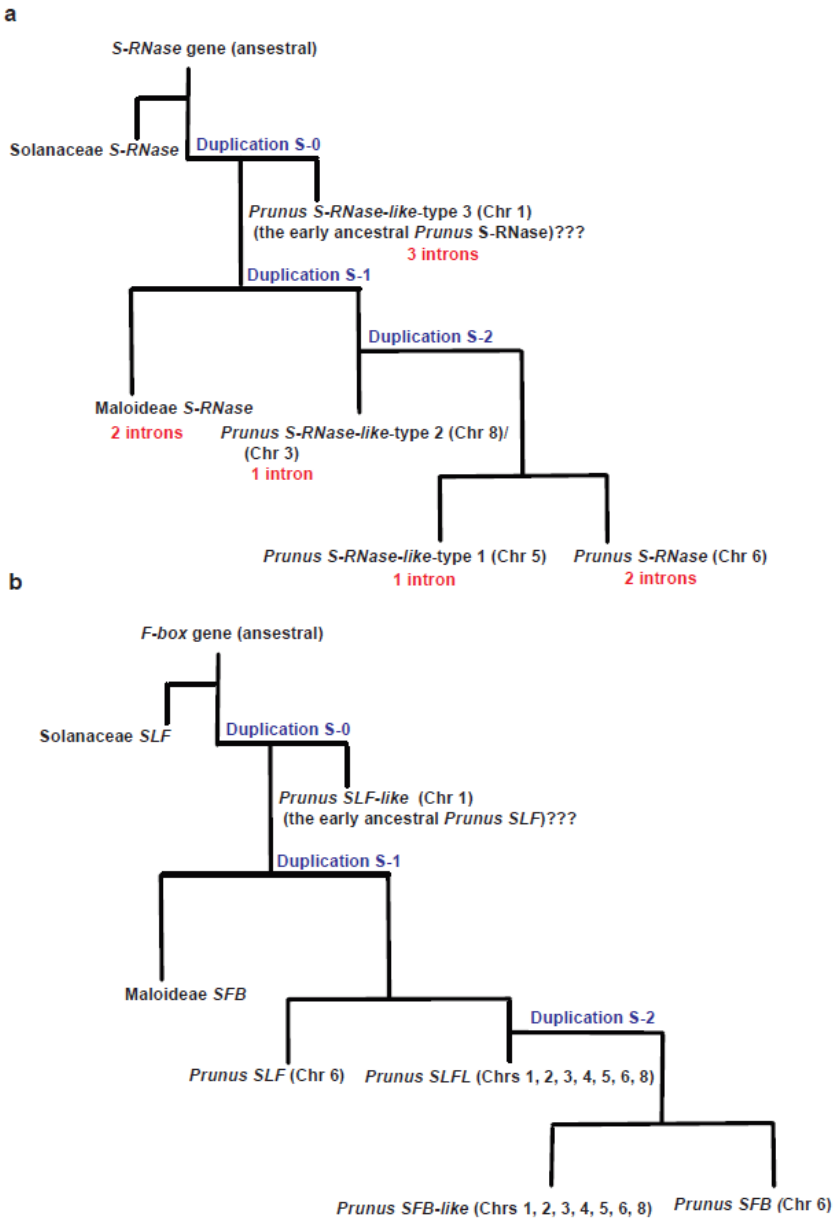


Fig. S10 Proposed model for the establishment of S-RNase (a) and SFB (b) genes in *Prunus*. (a) Following the emergence of Rosids, an initial S-RNase duplication event (Duplication S-1) occurred in the ancestral Rosaceae genome, giving rise to the Maloideae S-RNase lineage and inter-chromosomal duplications of early S-RNase copies in Prunoideae on chromosomes 8 and 3. A subsequent duplication event (Duplication S-2) produced the functional S-RNase on chromosome 6 and S-RNase-like genes on chromosome 5. (b) The establishment of SFB genes appears to follow a similar evolutionary trajectory. The first F-box duplication (Duplication S-1) in the ancestral Rosaceae genome generated the Maloideae SFB lineage and multiple inter-chromosomal F-box copies in Prunoideae (chromosomes 1–6 and 8), with one copy differentiating into SLF on chromosome 6. A second duplication event (Duplication S-2) subsequently gave rise to the Prunus SFB on chromosome 6 and SFB-like genes distributed across chromosomes 1–6 and 8.

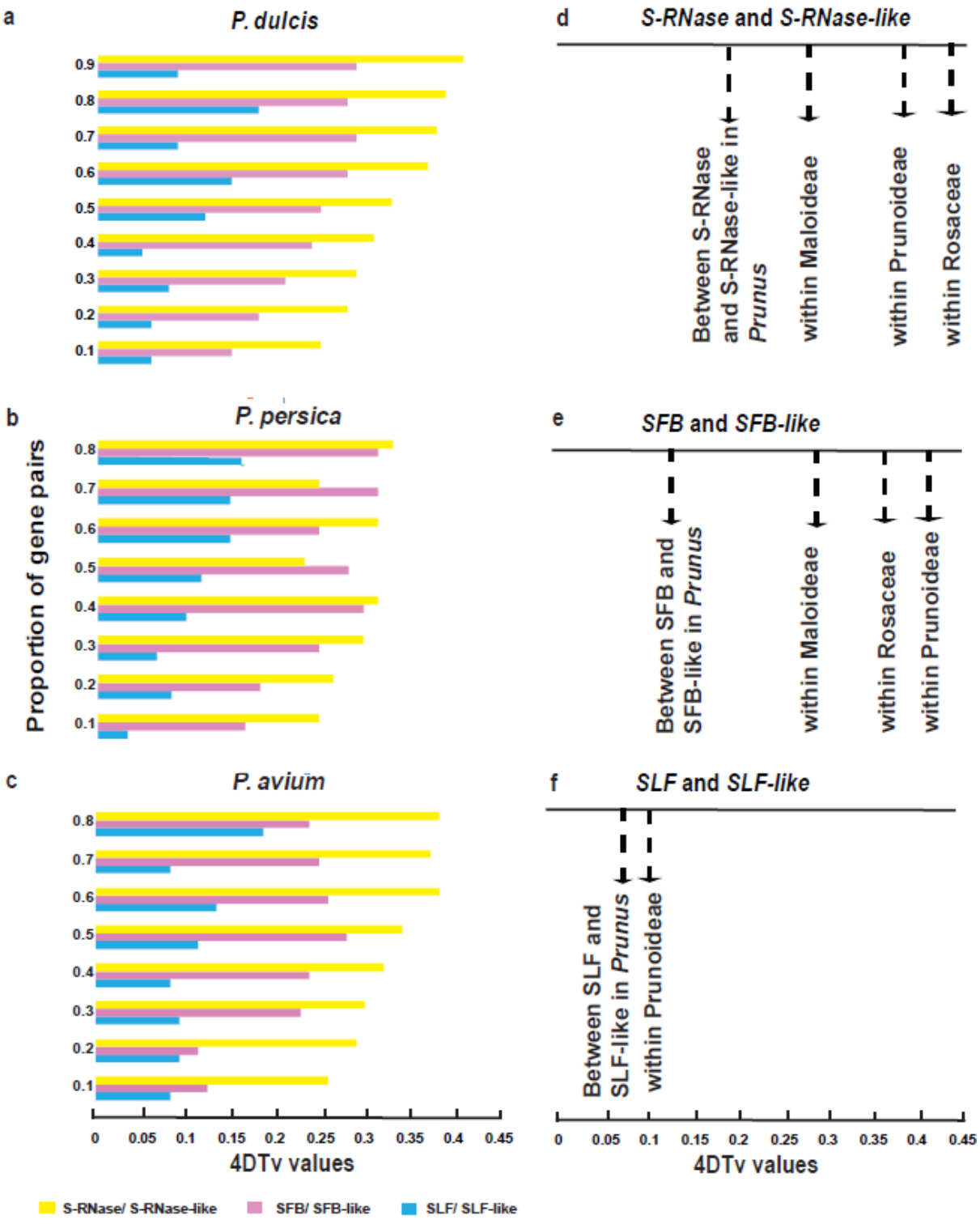


Fig. S11 Fourfold degenerate transversion (4DTv) values for *Prunus* and *Malus*. Left panels show pairwise mean 4DTv values between tandemly duplicated S-associated genes (S-RNase/S-RNase-like, SFB/SFB-like, and SLF/SLF-like) in *P. dulcis* (a), *P. persica* (b), and *P. avium* (c). Right panels show 4DTv values calculated from complete datasets used in phylogenetic analyses for S-RNase and S-RNase-like (d), SFB and SFB-like (e), and SLF and SLF-like (f).

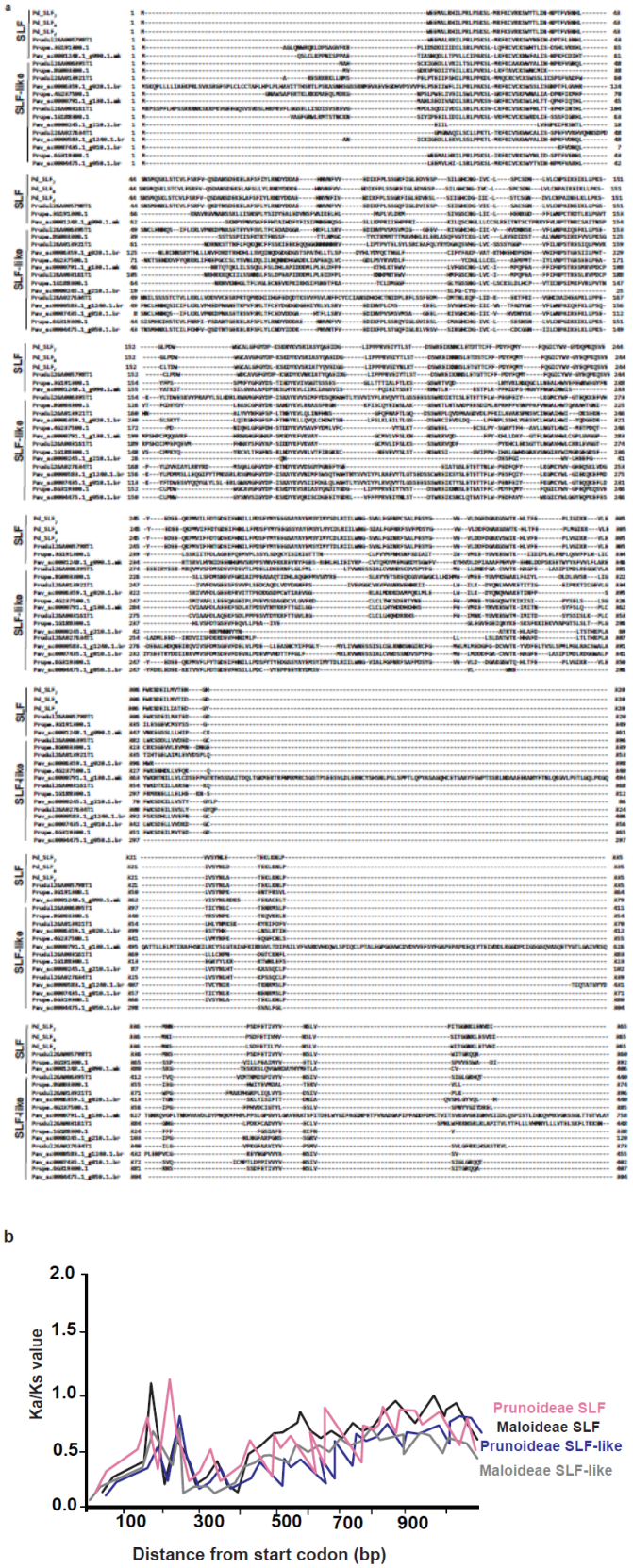


Fig. S12 Sequence alignment and selective pressure analysis of SLF and SLF-like in Rosaceae using *Prunus* and *Malus*. (a) Sequence alignment of SLF and SLF-like proteins. (b) Sliding-window Ka/Ks analysis calculated using 100-bp windows with a 20-bp step size.

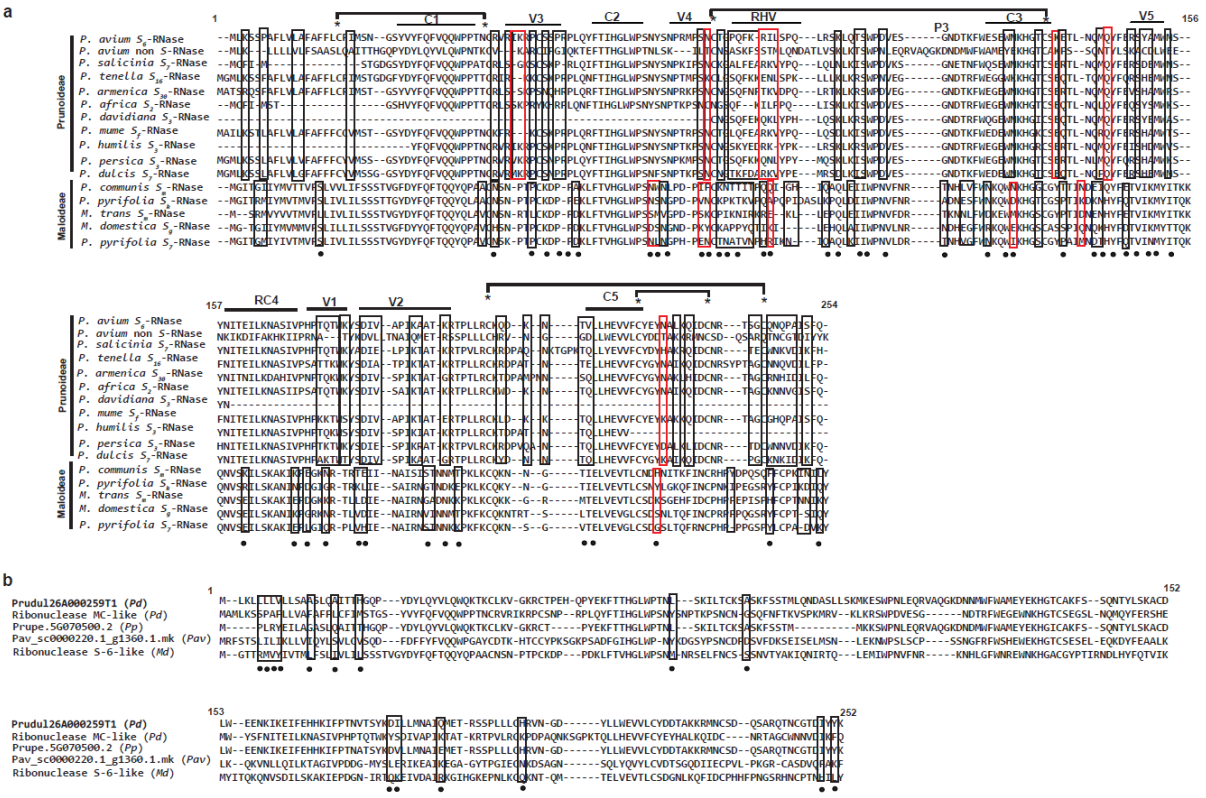


Fig. S13 Schematic representation of the S-RNases and S-RNase-like in Prunoideae and Maloideae in Rosaceae. In S-RNases (a), conserved regions (C1, C2, C3, RC4 and C5), variable regions (V1–V5) a hypervariable region (RHV) are shown. Cysteine residues that are predicted to be linked by disulphide bridges in S-RNases are connected by solid lines. Rectangles indicate the positively selected amino acid residues in the S-RNases (a) and S-RNase-like (b) where red and black rectangles indicate the sites under positive selection with posterior probabilities higher than 0. 99 and posterior probabilities higher than 0.95, respectively. Circles under the alignment indicate the sites that are identified using both gap (using DnaSP) and non-gap (PAML) datasets.

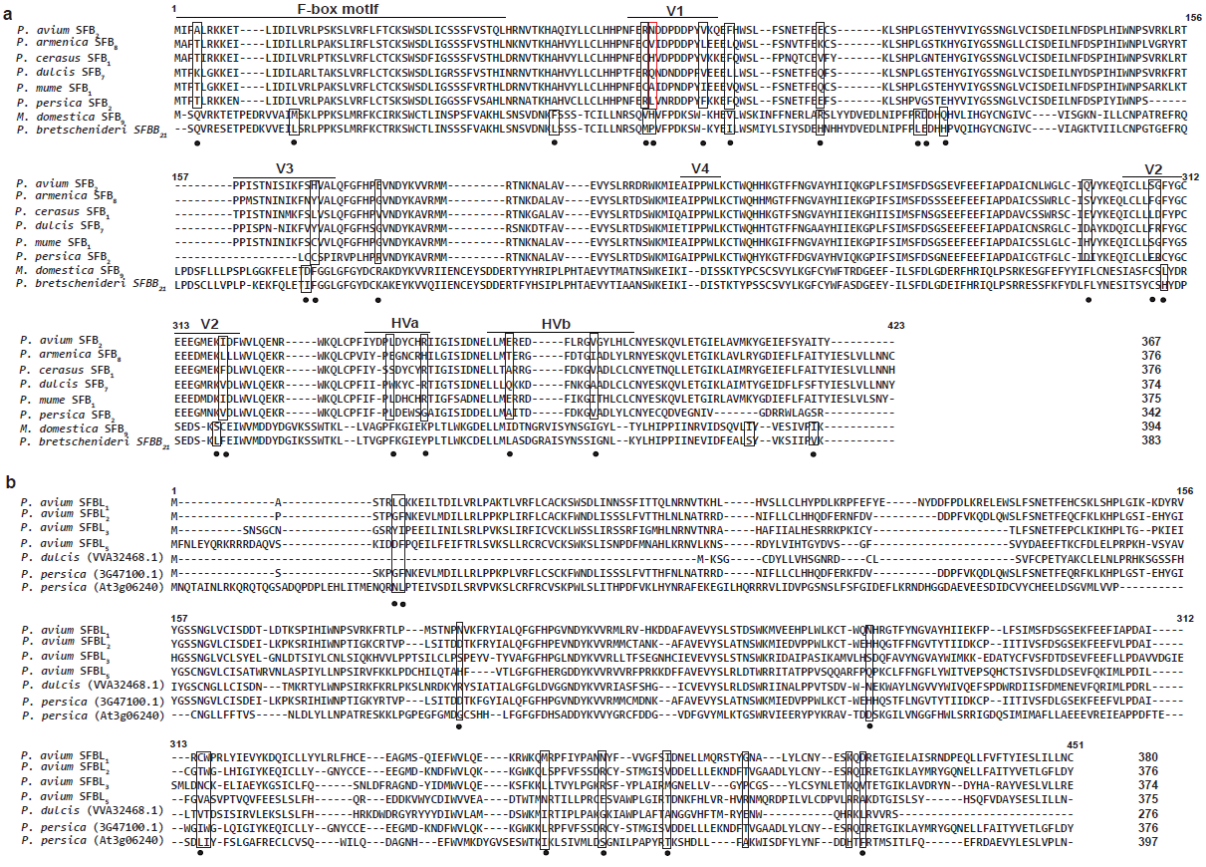


Fig. S14 Schematic representation of the positively selected amino acid residues in the SFB (a) in Prunoideae and Maloideae (a) and SFB-like in Prunoideae (b). In SFB, F-box motif, variable regions (V1–V4) a hypervariable regions (HVa and HVb)) are shown. Rectangles indicate the positively selected amino acid residues in the SFB (a) and SFB -like (b) where red and black rectangles indicate the sites under positive selection with posterior probabilities higher than 0. 99 and posterior probabilities higher than 0.95, respectively. Circles under the alignment indicate the sites that are identified using both gap (using DnaSP) and non-gap (PAML) datasets.



Fig. S15 Schematic representation of the SLF in Prunoideae and Maloideae and SLF-like in Prunoideae. Black rectangles indicate the positively selected amino acid residues in the SLF (a) and SLF -like (b) that posterior probabilities higher than 0.95. Circles under the alignment indicate the sites that are identified using both gap (using DnaSP) and non-gap (PAML) datasets.

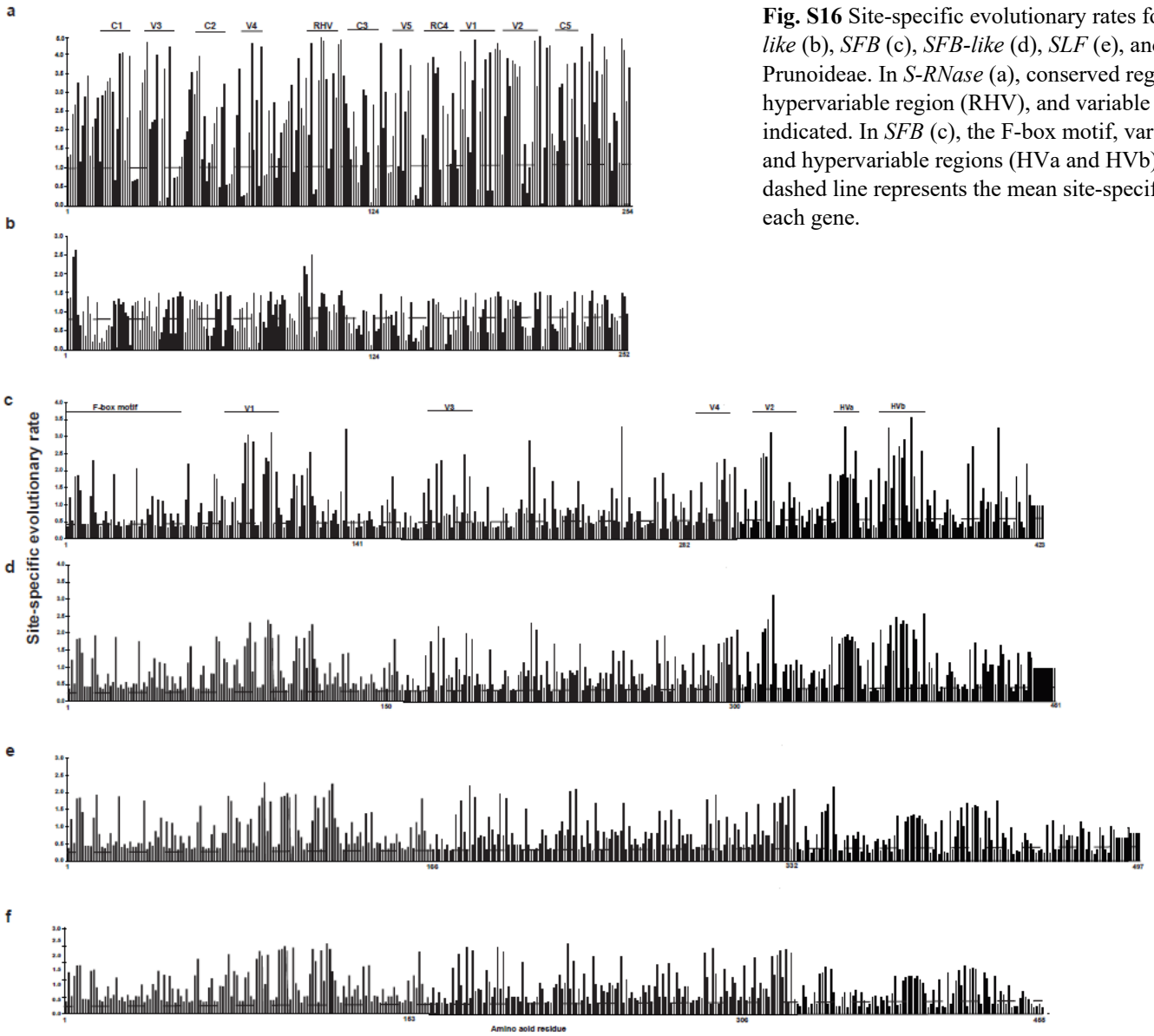


Fig. S16 Site-specific evolutionary rates for *S-RNase* (a), *S-RNase-like* (b), *SFB* (c), *SFB-like* (d), *SLF* (e), and *SLF-like* (f) genes in Prunoideae. In *S-RNase* (a), conserved regions (C1–C5), the hypervariable region (RHV), and variable regions (V1–V5) are indicated. In *SFB* (c), the F-box motif, variable regions (V1–V4), and hypervariable regions (HVa and HVb) are labelled. The black dashed line represents the mean site-specific evolutionary rate for each gene.

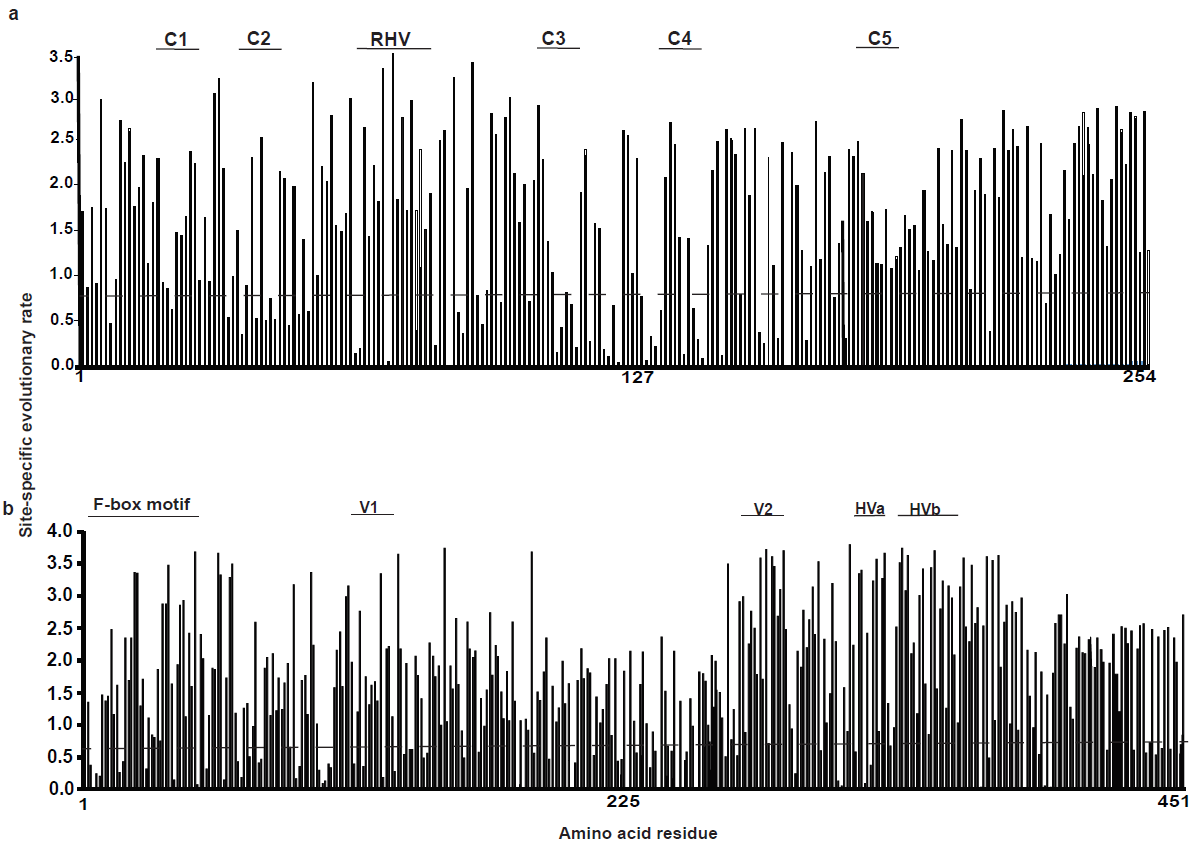


Fig. S17 Site-specific evolutionary rates for S-RNase (a) and SFB (b) genes in Maloideae. In S-RNase (a), conserved regions (C1–C5) and the hypervariable region (RHV) are indicated. In SFB (b), the F-box motif, variable regions (V1–V2), and hypervariable regions (HVa and HVb) are labelled. The black dashed line represents the mean site-specific evolutionary rate for each gene.

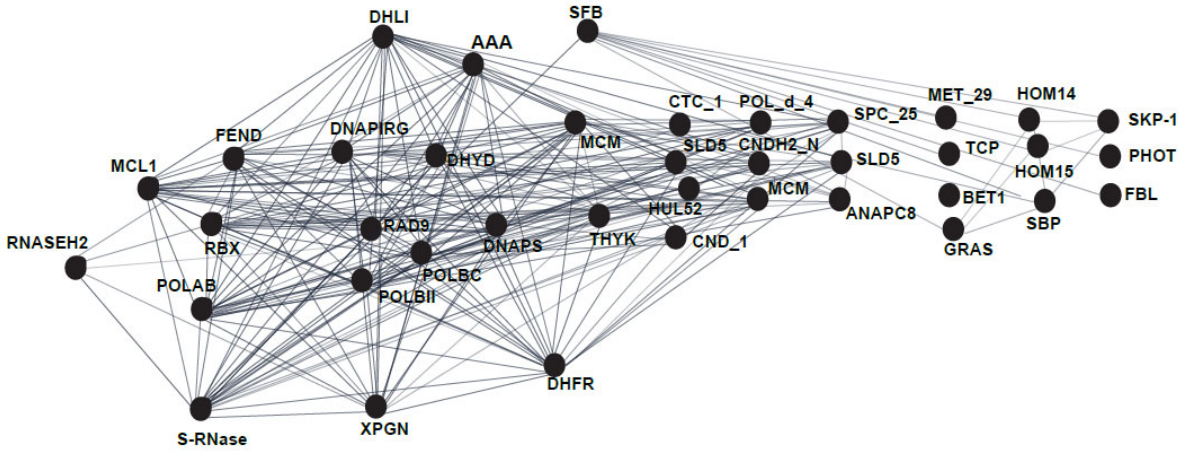


Fig. S18 Backbone protein–protein interaction (PPI) network underlying self-compatibility and self-incompatibility reactions in *Prunus*, inferred using the STRING database with S-RNase and SFB as seed proteins.