



OPEN

# The identification of the *Rosa* S-locus and implications on the evolution of the Rosaceae gametophytic self-incompatibility systems

J. Vieira<sup>1,2</sup>, J. Pimenta<sup>1,2</sup>, A. Gomes<sup>1,2</sup>, J. Laia<sup>1,2</sup>, S. Rocha<sup>1,2</sup>, P. Heitzler<sup>3</sup> & C. P. Vieira<sup>1,2</sup>✉

In Rosaceae species, two gametophytic self-incompatibility (GSI) mechanisms are described, the *Prunus* self-recognition system and the Maleae (*Malus/Pyrus/Sorbus*) non-self-recognition system. In both systems the pistil component is a *S-RNase* gene, but from two distinct phylogenetic lineages. The pollen component, always a F-box gene(s), in the case of *Prunus* is a single gene, and in Maleae there are multiple genes. Previously, the *Rosa* S-locus was mapped on chromosome 3, and three putative *S-RNase* genes were identified in the *R. chinensis* 'Old Blush' genome. Here, we show that these genes do not belong to the S-locus region. Using *R. chinensis* and *R. multiflora* genomes and a phylogenetic approach, we identified the *S-RNase* gene, that belongs to the *Prunus* S-lineage. Expression patterns support this gene as being the S-pistil. This gene is here also identified in *R. moschata*, *R. arvensis*, and *R. minutifolia* low coverage genomes, allowing the identification of positively selected amino acid sites, and thus, further supporting this gene as the *S-RNase*. Furthermore, genotype–phenotype association experiments also support this gene as the *S-RNase*. For the S-pollen GSI component we find evidence for multiple F-box genes, that show the expected expression pattern, and evidence for diversifying selection at the F-box genes within an S-haplotype. Thus, *Rosa* has a non-self-recognition system, like in Maleae species, despite the S-pistil gene belonging to the *Prunus* *S-RNase* lineage. These findings are discussed in the context of the Rosaceae GSI evolution. Knowledge on the *Rosa* S-locus has practical implications since genes controlling floral and other ornamental traits are in linkage disequilibrium with the S-locus.

Gametophytic self-incompatibility (GSI), the most common pre-zygotic self-incompatibility genetic mechanism, prevents self-fertilization between genetically related individuals, where the genotype of the haploid pollen determines its incompatibility type<sup>1</sup>. This mechanism is determined by a single locus, the S-locus, and in the most frequent eudicot system<sup>2</sup>, the S-pistil gene codes for a protein with ribonuclease activity, called *S-RNase*<sup>3–5</sup>, and the S-pollen gene (s) code(s) for a F-box protein(s)<sup>6–26</sup>. In incompatible crosses, the cytotoxic *S-RNases* lead to pollen tube RNA degradation, that causes the cessation of pollen tube elongation. Two mechanisms of pollen recognition have been proposed, the self and non-self-recognition systems<sup>6–26</sup>. In the self-recognition mechanism, described in the *Prunus* genus of the Rosaceae family<sup>6, 7, 11, 17, 24, 27</sup>, only one F-box protein is the S-pollen GSI specificity determinant, called SFB, that interacts with the self *S-RNase* (reviewed by<sup>28</sup>). In this system, as expected, the *S-RNase* and *SFB* genes show patterns of co-evolution<sup>29</sup>. In the non-self-recognition mechanism, characterized in Solanaceae, Plantaginaceae, and in the Rosaceae Maleae tribe (*Malus/Pyrus/Sorbus*), there are multiple F-box proteins, called SLFs in Solanaceae and Plantaginaceae, and SFBs in Maleae<sup>13–16, 18–26, 30–32</sup>, that interact with all *S-RNases* but the self-*S-RNase*<sup>15, 16, 18, 20–26, 30–32</sup>. Even though different recognition interactions occur due to different evolutionary paths, the *S-RNase* retained the same feature of cytotoxicity to self-pollen in both mechanisms<sup>28, 30</sup>.

<sup>1</sup>Instituto de Biologia Molecular e Celular (IBMC), Rua Alfredo Allen, 208, 4200-135 Porto, Portugal. <sup>2</sup>Instituto de Investigação e Inovação em Saúde (I3S), Universidade do Porto, Rua Alfredo Allen, 208, 4200-135 Porto, Portugal. <sup>3</sup>Institut de Biologie Moléculaire Des Plantes, CNRS, Université de Strasbourg, UPR 2357, 67000 Strasbourg, France. ✉email: cgvieira@ibmc.up.pt

Rosa genome	Number of sequences						
	Total	Excluding pseudogenes	Presenting				
			Both motifs	Both motifs and IP ≥ 7	Only 2nd motif	Only 2nd motif and IP > 7	Only 1st motif
<i>R. chinensis</i> Assembly-CDS (Rchinensis1)	12	12	10	5	–	–	–
<i>R. chinensis</i> Assembly-ORF (Rchinensis2)	33	26	26	19	–	–	–
<i>R. chinensis</i> GDR-CDS (Rchinensis3)	18	18	17	11	–	–	–
<i>R. chinensis</i> GDR-ORF (Rchinensis4)	44	22	22	13	–	–	–
<i>R. multiflora</i>	34	25	18	14	4	4	3

**Table 1.** Summary of the *S-RNase* like sequences identified in the *R. chinensis*, and *R. multiflora* genomes.

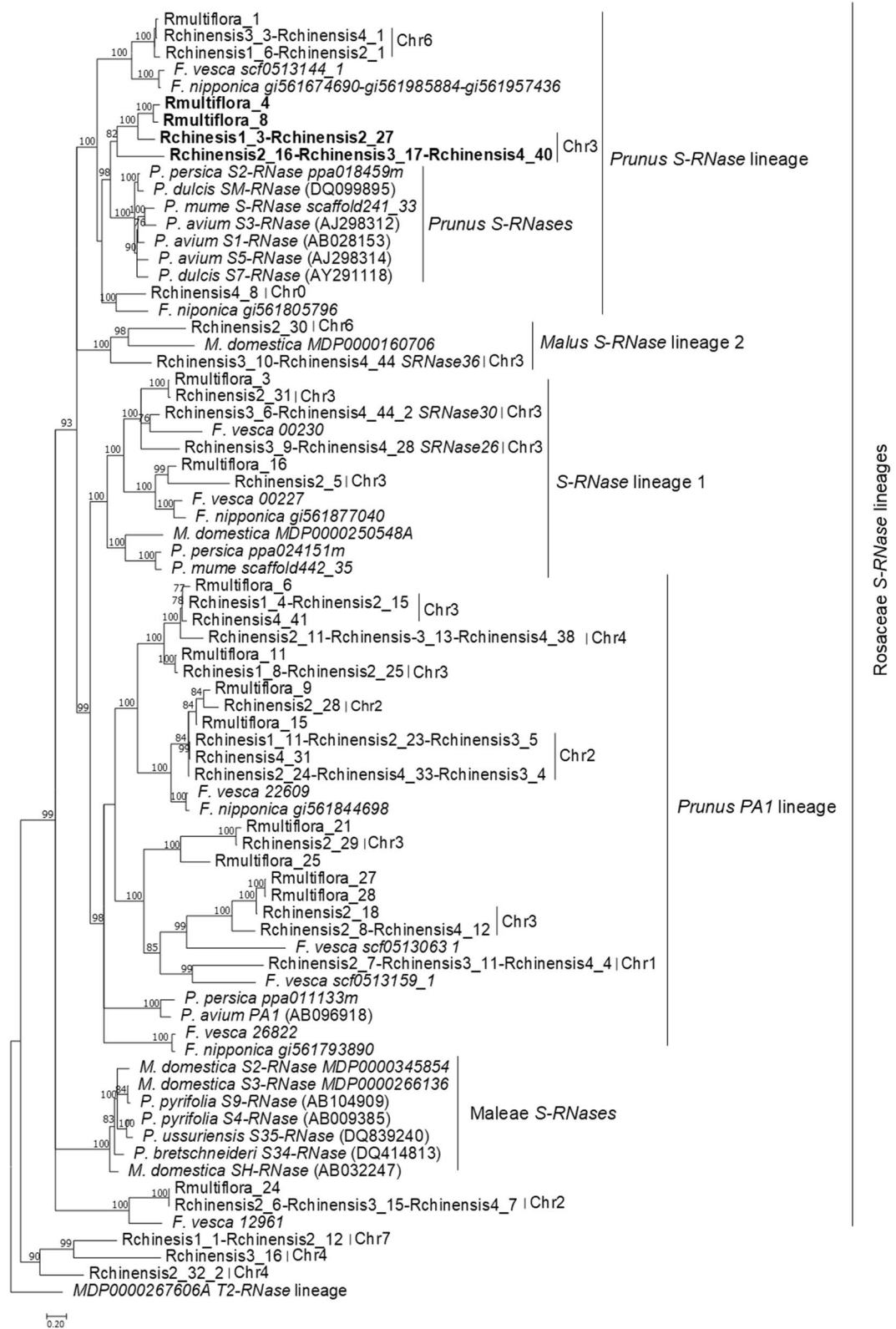
Based on phylogenetic inferences, the *S-RNase* GSI system evolved once in core eudicots<sup>33–37</sup>, and has been the subject of multiple duplication events during evolution. In Rosaceae, different duplicates have been recruited for GSI function in *Prunus* and *Malus/Pyrus/Sorbus*<sup>24,37</sup>. Convergent evolution has also been proposed for the evolution of the *S*-pollen determinant since the *Prunus* *S*-pollen gene, belongs to a different gene lineage than that of the *Malus/Pyrus/Sorbus* *S*-pollen genes. Therefore, it is not surprising the presence of different recognition systems in *Prunus* (self-recognition<sup>6,7,11,12,17,24,27,28,38</sup>) and in *Malus/Pyrus/Sorbus* (non-self-recognition mechanism<sup>13,16,18–20,24,25</sup>). How the two systems evolved is still debated. The observation that *Prunus S-RNase* and *SFB* lineage genes are present in *Fragaria* species (an outgroup), suggests that the ancestral Rosaceae *S*-locus was of the self-recognition type, and that the *Malus* *S*-locus region has a “de novo” evolution<sup>24</sup>. An alternative hypothesis postulates that the divergence between *SFB* and *SFBB* genes occurred early in the establishment of eudicots, and that *Prunus* species started using *SFB* gene as a *S*-pollen factor around the time of the *Prunus* divergence<sup>39</sup>. Under this evolutionary hypothesis, the common ancestor of *Prunus* and *Malus/Fragaria* would present a non-self-recognition mechanism. In this work, by characterizing the *Rosa S*-locus region we clarify the Rosaceae *S*-locus ancestral state.

Rose (*Rosa* sp., Rosaceae) species are known by their ornamental value and the production of essential oils used in perfume and cosmetic industry. There are about 200 species in this genus, but only 10 species (*R. canina*, *R. chinensis* ‘Old Blush’ OB), *R. foetida*, *R. gallica*, *R. gigantea*, *R. moschata*, *R. multiflora*, *R. phoenicia*, *R. rugosa*, and *R. wichuraiana* ‘Basye’s Thornless’) have contributed for most of the modern species, throughout processes of hybridization and polyploidization<sup>40</sup>. Based on the absence of seed set after self-pollination, the presence of a GSI system has been postulated for most *Rosa* diploid species<sup>41–43</sup>. Analyses of one of the *R. chinensis* genomes<sup>44</sup>, revealed three putative *S-RNase* genes (*SRNase26*, *SRNase30*, and *SRNase36*) located on chromosome 3, as a candidate region of the *Rosa S*-locus<sup>44</sup>. In this work we show that this region is not the *S*-locus. This is the reason why segregation analyses using *SRNase30* as a marker, shows co-segregation with the *S*-locus at a distance of 4.2 cM<sup>44</sup>. Nevertheless, these segregation analyses showed that the *S*-locus region is located on chromosome 3, as previously reported<sup>45</sup>. According to chromosome location, we also exclude a putative *S*-pollen gene –*RrSLF*, cloned from *R. rugosa* pollen RNA, that in phylogenetic analyses clusters with *Prunus SFB*<sup>46</sup>, since the orthologous *R. chinensis* gene is located on chromosome 2.

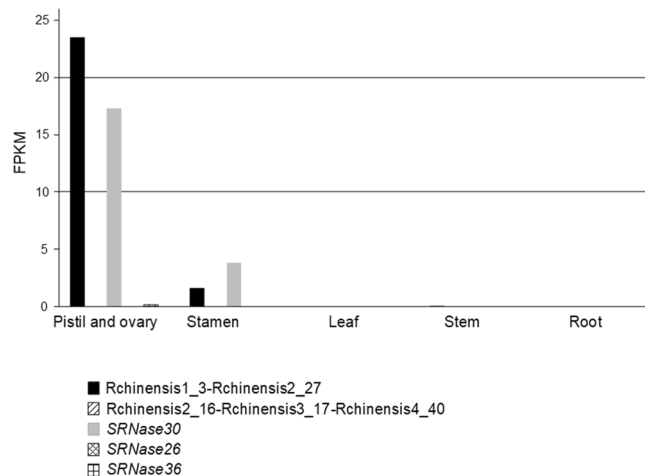
*Rosa* diploid genomes are of relatively small size (560 Mb<sup>47,48</sup>), and for *R. chinensis* ‘Old Blush’<sup>44,49</sup>, and *R. multiflora*<sup>50</sup> there are assembled genomes publicly available (<https://www.ncbi.nlm.nih.gov/sra>). Here, we use these datasets to perform phylogenetic analyses to identify the *Rosa S*-locus genes. Transcriptomic analyses using *R. chinensis* pistil and ovary, stamen, leaf, stem, and root tissue (available in SRA database), support the identification of the *Rosa S*-locus genes. To show that the *Rosa S-RNase* gene shows evidence of positive selection, we used low coverage *Rosa* genomes (*R. moschata*, *R. laevigata*, *R. rugosa*, *R. persica*, *R. xanthina*, *R. minutifolia*, *R. odorata*, *R. arvensis*, and *R. majalis*) available at SRA database, to identify further *S-RNase* alleles. Furthermore, we use *R. arvensis* genotype–phenotype association experiments to confirm that the identified gene is the *S-RNase*. We also identify F-box genes in the vicinity of the *S-RNase*, to determine which GSI system is present in *Rosa*. Expression, phylogenetic analyses, polymorphism levels, and evidence for selection favoring diversification of F-box genes within an *S*-haplotype, suggest multiple F-box genes as the *S*-pollen component. Rosaceae GSI evolution is discussed in the context of these findings.

## Results

***R. chinensis* and *R. multiflora* Rosaceae *S-RNase* lineage genes.** *S-RNases* code for basic proteins (isoelectric point (IP) above 8)<sup>5,24,33,34,37</sup>, that present two conserved amino acid patterns (pattern 1 and 2<sup>36</sup>), are expressed in pistils (where the rejection of self-pollen occurs during the growth of pollen tubes), stigma, and flowering buds, and present signs of diversifying selection<sup>5,24,33,34,36,37</sup>. Of the 80 *S-RNase* like sequences identified in *R. chinensis* (55) and *R. multiflora* (25) genomes (Table 1; Supplementary Table S1), 41 code for putative proteins with a IP above 7 (Table 1; Supplementary Table S2). Phylogenetic analyses of these sequences, and 34 reference sequences from<sup>24</sup> (Supplementary File S1), revealed that the sequences labeled Rchinensis3\_16, Rchinensis2\_32\_2, and Rchinensis1\_1-Rchinensis2\_12, do not belong to the Rosaceae *S-RNase* lineage (Fig. 1). None of the *Rosa S-RNase* lineage sequences cluster with *Malus/Pyrus S-RNases*. Two groups of sequences Rmultiflora\_4, Rmultiflora\_8, and Rchinensis1\_3-Rchinensis2\_27 (gene 1), and Rchinensis2\_16-Rchinensis3\_17-



**Figure 1.** Bayesian phylogenetic tree, showing the relationship of the *R. chinensis* and *R. multiflora* S-RNase like sequences with *Fragaria*, *Prunus*, and *Malus* S-RNase lineage genes. For the *R. chinensis* sequences the chromosomal (Chr) location is given. The tree was rooted with *MDP0000267606A T2-RNase*, not involved in GSI<sup>24</sup>. In bold are the *Rosa* sequences that cluster with *Prunus* S-RNases, that could represent the S-RNase gene. Numbers below the branches represent posterior credibility values above 70.

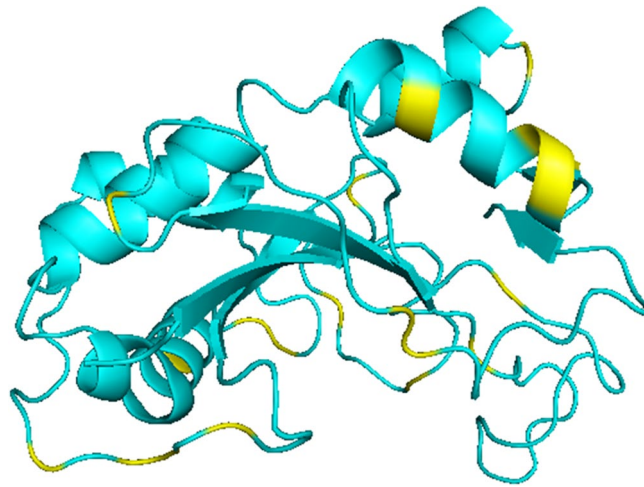


**Figure 2.** *R. chinensis* expression levels (FPKM) for the two *Prunus* *S-RNase* lineage genes, as well as the three previous putative *S-RNases*, in pistil and ovary, stamen, leaf, stem, and root tissues.

Rchinensis4\_40 (gene 2), cluster with *Prunus S-RNases* (Fig. 1), suggesting that one of them may represent the *Rosa S-RNase* gene. In *R. chinensis* these sequences are located on chromosome 3, where the *S*-locus has been identified<sup>44,45</sup>. These sequences code for putative proteins with IP above 8, they have two putative introns, and amino acid pattern 2 conserved, features typical of *S-RNases*. Gene 1 sequences also present high levels of synonymous polymorphism (27.9%) as the *S-RNases*<sup>36</sup>. Although in this group of sequences pattern 1 shows a violation at the first pattern position (Y instead of [FST]), an identical pattern is observed in Solanaceae for the *S-RNase Nicotiana tomentosiformis* XP\_018625910\_1 (this sequence clusters with other Solanaceae *S-RNases* in phylogenetic analyses; data not shown). Because gene 2 is represented by one sequence, we cannot address levels of diversity. Nevertheless, the presence of identical sequences in the two *R. chinensis* genomes analyzed, suggests low levels of diversity. The putative *SRNase26*, *SRNase30*, and *SRNase36* present IP below 8, and the first two sequences cluster with *Prunus* and *Malus S*-lineage 1 genes, and *SRNase36* clusters with the *Malus S*-lineage 2 gene. In *Prunus* and *Malus* GSI, these sequences are not involved in specificity determination.

**Rchinensis1\_3–Rchinensis2\_27 is expressed in pistil and ovary tissue, like the *S-RNases*.** The *S-RNase* gene is highly expressed in pistils, in stigmas and styles of flowers at anthesis, but also shows low expression in entire flower buds. It should be noted that *S-RNase* lineage genes, can also show a similar pattern of expression, since this expression is inferred to be the ancestral expression of the *S-RNase* lineage genes<sup>24</sup>. Therefore, it is not surprising that *SRNase30*, and *SRNase36*, show expression in pistil and ovary (Fig. 2), as previously reported<sup>44</sup>. *SRNase26* shows no expression in the tissues here analyzed, as well as those used in<sup>44</sup>. Rchinensis1\_3–Rchinensis2\_27, shows a similar expression to *SRNase30*, but with high levels of expression in pistil and ovary, and low expression in stamen (Fig. 2). Rchinensis2\_16–Rchinensis3\_17–Rchinensis4\_40 is not expressed in the tissues here analyzed. These results are compatible with Rchinensis1\_3–Rchinensis2\_27 gene, being the *S*-pistil gene determining GSI specificity.

***Rosa S-RNases* show evidence for positively selected amino acid positions.** To address if the *Rosa* sequences here identified as the *S*-locus pistil gene are the subject of diversifying selection, a feature of the *S-RNase* gene, we first identified and annotated sequences similar to *S-RNases* from other self-incompatible *Rosa* genomes with low coverage (*R. moschata*, *R. laevigata*, *R. rugosa*, *R. persica*, *R. xanthina*, *R. minutifolia*, *R. odorata*, *R. arvensis*, and *R. majalis*; Supplementary Table S1). Phylogenetic analyses using sequences presenting an IP above 7 covering at least the exon where motif 2 is located (Supplementary Fig. S1A,B; Supplementary File S2), together with the *Rosa S-RNase* sequences here identified, and the reference sequences from<sup>24</sup>, revealed four additional *Rosa S-RNases* (Rarvensis\_11, Rminutifolia\_20, Rmoschata17\_14, and Rmoschata17\_30; Supplementary Fig. S1A,B). Similar analyses with the sequences covering the motif 1, revealed four sequences (Rodorata07\_39; Rmoschata08\_12; Rminutifolia\_7; and Rarvensis\_6; Supplementary Fig. S2; Supplementary File S3) that also cluster with *Rosa S-RNases* sequences. Because the phylogenetic relationship of sequences Rarvensis\_11 and Rarvensis\_6 is similar, they may represent two exons of the same gene from the same *S*-haplotype, and thus were treated as such. The same applies to sequences Rminutifolia\_20 and Rminutifolia\_7, as well as sequences Rmoschata17\_14 and Rmoschata17\_34. Using these sequences together with Rchinensis1\_3–Rchinensis2\_27, Rmultiflora\_4, and Rmultiflora\_8, we identify 21 amino acid sites under positive selection (results available at <http://bpositive.i3s.up.pt/> in the project named *Rosa S*-locus genes; BP2018000004) by performing codeML<sup>51</sup> analyses. These amino acid sites are, in principle, responsible for GSI specificity<sup>24,35,52</sup>. The location of these sites at the predicted 3D structure is mostly around the active site pocket region (Fig. 3), as observed in other Rosaceae and Solanaceae species<sup>25,52</sup>.



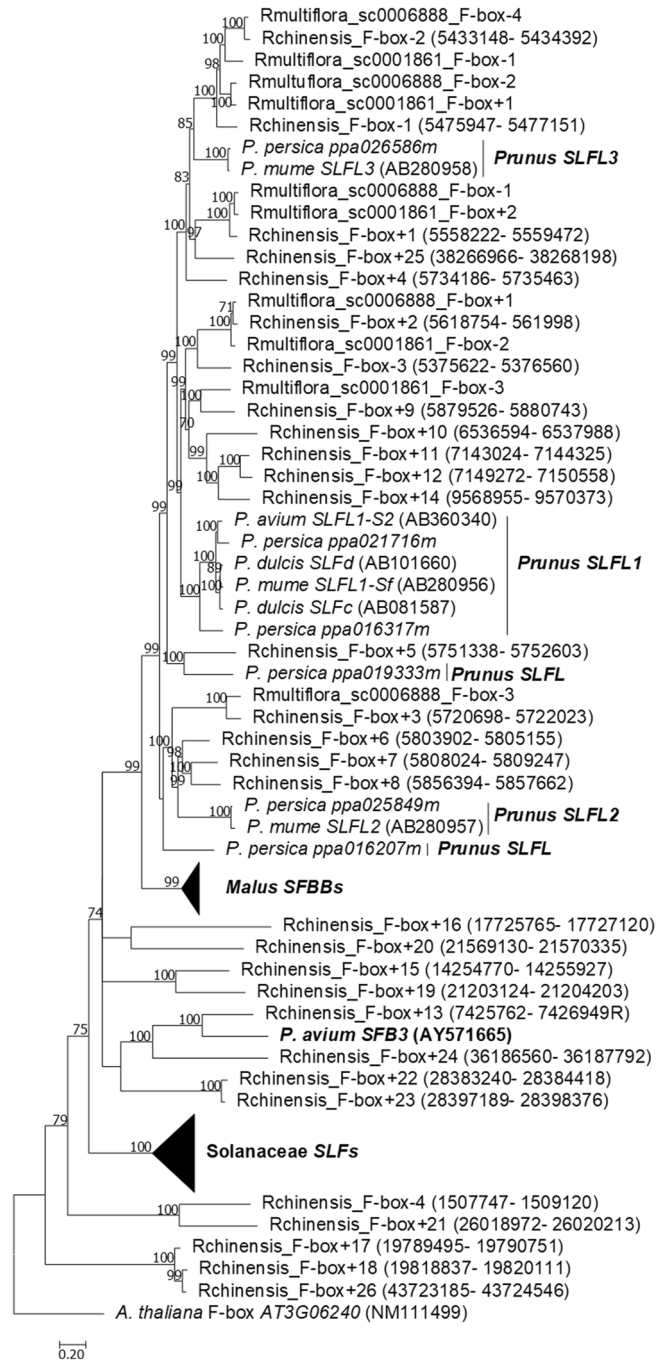
**Figure 3.** Positively selected amino acid sites, highlighted in yellow on the predicted 3D structure of the *R. chinensis* S-RNase.

***R. arvensis* genotype–phenotype association experiments.** For 12 *R. arvensis* accessions, obtained from a breeding program of siblings<sup>53</sup>, *S*-haplotype was deduced either through hand-pollinated tests (Supplementary Fig. S3A,B), allowing to perform co-segregation experiments of the *S2*-allele. An amplification product with the expected size (300 bp; see Material and Methods) was obtained in six (Ose (S1 S2), Url (S4 S5), E200 (Ose x Wid (S3 S6); S2 S3), E404 (E200 x Ose; S1 S2), E435 (E200 x Wid; S2 S6), and E893 (E459 x Url; S1 S5) individuals. The sequence of these amplification products, revealed two types of identical sequences. One obtained from Ose, E200, E404, and E435 individuals, those individuals having the *S2-RNase* (GenBank acc. Numbers MW452856–MW452859). The other sequence type was obtained from Url and E893, those individuals having the *S5-RNase* (GenBank acc. numbers MW452860 and MW452861). These results support co-segregation of the *S*-locus with the *S2-RNase* genotypes here surveyed. Therefore, the *S-RNase* gene here identified is on the *S*-locus region.

**Identification of the *Rosa* *S*-pollen F-box genes.** The *R. rugosa* *S*-locus F-box gene (*RrSLF*; KY446808), expressed in pollen tissue and phylogenetically related with *Prunus* *SFB* gene, was reported as a putative *Rosa* *S*-pollen gene<sup>46</sup>. It should be noted that other F-box genes not involved in *S*-pollen specify determination are also expressed in pollen, as well as in other tissues<sup>24</sup>. The expression of *RrSLF* gene in other tissues has not, however, been addressed<sup>46</sup>. This gene has 99% homology at nucleotide level with a *R. chinensis* sequence (CM009583.1) located on chromosome 2, but the *Rosa* *S*-locus is located on chromosome 3<sup>(44,45, and this work)</sup>. Furthermore, all *S*-pollen genes described in Rosaceae, Solanaceae, and Plantaginaceae are intronless genes, and the *R. chinensis* *RrSLF* (PRQ51373) and *R. multiflora* (Rmu\_sc0016061.1;BDJD01015883.1) orthologous genes have one intron. The orthologous gene has been identified in all low coverage *Rosa* genomes, except *R. majalis*, covering the entire coding region. Low levels of divergence (0.036 for synonymous and 0.009 for non-synonymous divergence respectively, after Jukes and Cantor correction; N = 10) are obtained for this gene. This is in contrast with the *Prunus* *SFB* gene, that presents levels of variability above 20%<sup>6,7,11,36</sup>. Therefore, *RrSLF* is not the *S*-pollen gene.

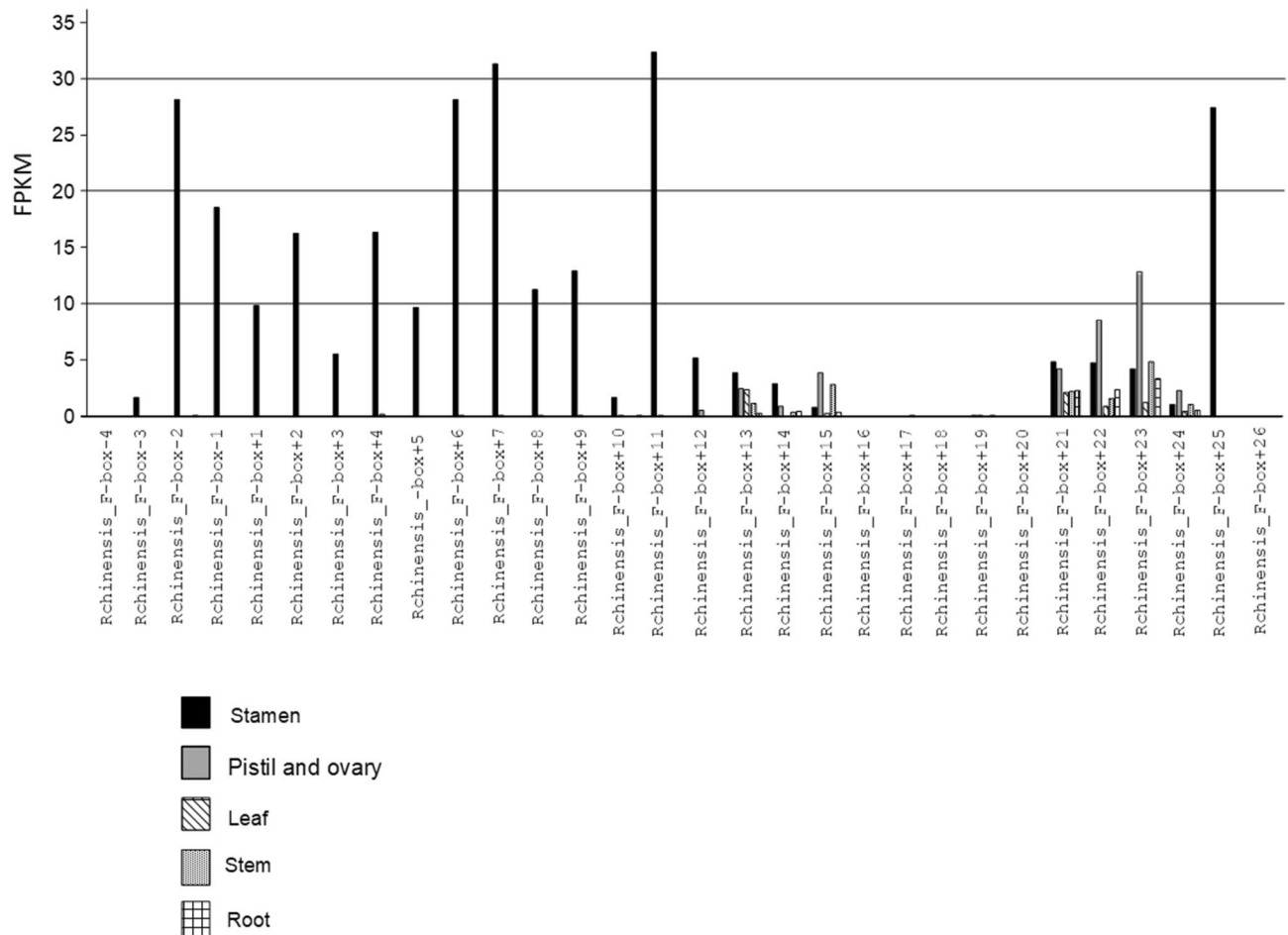
In *R. chinensis* chromosome 3 there are 30 *SFB/SLFL/SFBB* like genes (called Fbox – and + according to the 5' or 3' position relative to the *S-RNase*, respectively; Supplementary Table S3; Supplementary File S4). In the two *R. multiflora* scaffolds where the *S-RNase* is located, there are five such genes (Supplementary Table S3). The phylogenetic analyses of the *R. chinensis* and *R. multiflora* *SFB/SLFL/SFBB* like genes together with *Prunus* *SFB* and *SLFL* genes, *M. domestica* *S1-SFBBs*, *Petunia* and *Nicotiana* *SLF* sequences, and *A. thaliana* F-box/kelch-repeat, shown in Fig. 4 (Supplementary File S4), revealed that the Rchinensis\_F-box + 13 gene clusters with *Prunus* *SFB* gene. This gene is expressed in all tissues here analyzed (Fig. 5), but the *S*-pollen gene(s) are mainly expressed in anthers/pollen only<sup>7,10,12–16,21,23–25</sup>, and thus is unlikely to be the *S*-pollen gene. Furthermore, using a 1188 bp gene region obtained from eight *Rosa* genomes (*R. arvensis*, *R. laevigata*, *R. moschata*, *R. xantina*, *R. rugosa*, *R. odorata*, *R. multiflora*, and *R. chinensis*), low average levels of divergence (0.0321 for synonymous and 0.0128 for non-synonymous divergence respectively, after Jukes and Cantor correction) are observed. Moreover, this gene is the neighbor F-box gene of the *T2-RNase* Rchinensis1\_8-Rchinensis2\_25, that does not cluster with *Prunus* or Maleae *S-RNases* (Fig. 1). Therefore, Rchinensis\_F-box + 13 gene is not involved in *S*-pollen GSI specificity determination.

The F-box genes in the vicinity of the *R. chinensis* and *R. multiflora* *S-RNase* gene cluster, with high support, with different *Prunus* *SLF* genes (Fig. 4), that are a sister group of *Malus* *SFBBs*<sup>24</sup>. In *R. chinensis* 14 of these genes (Rchinensis\_F-box-3 up to Rchinensis\_F-box + 11) show expression in stamen only, compatible with being *S*-pollen genes (Fig. 5). Furthermore, the *R. multiflora* orthologs of *R. chinensis* are not in the same order relatively to the *S-RNase* gene (Fig. 4), showing that this region is highly rearranged, as observed in the *Malus*/



**Figure 4.** Bayesian phylogenetic tree, showing the relationship of the SFB/SLFL/SFBB like genes (called Fbox – and + according to the 5' or 3' position relative to the *S-RNase*) from *R. chinensis* chromosome 3, *R. multiflora* sc0006888, and *R. multiflora* sc0001861, with *M. domestica* SFBBs, *Prunus* SFBs and SLFLs, and Solanaceae SLFs. The tree was rooted with *A. thaliana* F-box/kelch-repeat, not involved in GSI. Numbers below the branches represent posterior credibility values above 70.

*Pyrus/Sorbus* *S*-locus region<sup>13, 16, 18–20, 24, 25</sup>. This is surprising since we identified the *Rosa S-RNase* as belonging to the *Prunus S*-lineage, and in *Prunus* there is a single *S*-pollen gene. In *Prunus* the *S*-pollen gene presents levels of diversity similar to the *S-RNase* gene<sup>11, 29</sup>. Therefore, we also determined levels of synonymous and non-synonymous divergence for the F-box genes surrounding the *R. chinensis S-RNase* (*Rchinensis\_F-box-1* and *Rchinensis\_F-box+1*, used as query in a blastn to identify contigs containing the orthologous genes in the low coverage *Rosa* genomes). The low levels of synonymous and non-synonymous divergence for *Rchinensis\_F-box-1* (0.070 and 0.005 respectively), and *Rchinensis\_F-box+1* (0.112 and 0.02129 respectively) are incompatible with the hypothesis that one of them is determining *Rosa S*-pollen specificity. Therefore, multiple F-box genes must be involved in *Rosa* pollen GSI specificity determination, as in *Malus/Pyrus/Sorbus* and Solanaceae species.



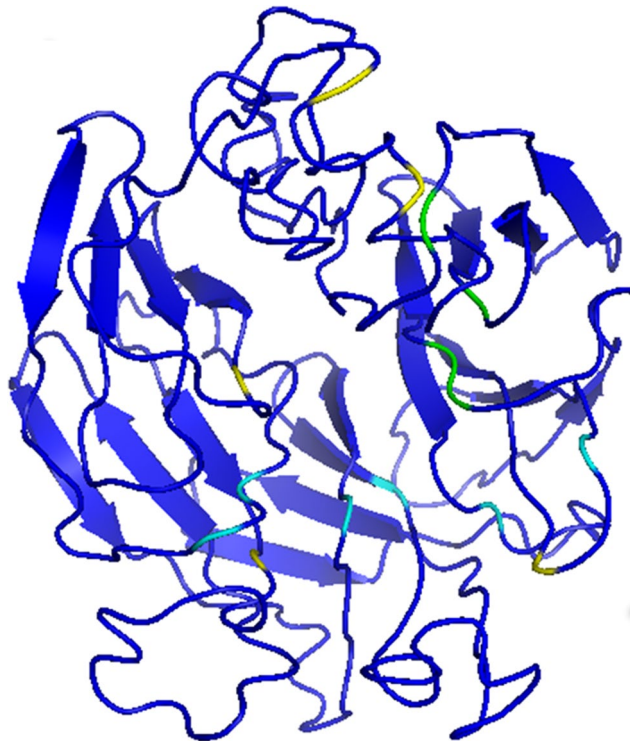
**Figure 5.** *R. chinensis* expression levels (FPKM) for the F-box genes located on chromosome 3, in pistil and ovary, stamen, leaf, stem, and root tissues.

Using the 14 *R. chinensis* F-box sequences in the vicinity of the *S-RNase*, that are expressed in stamen, we find evidence for positively selected amino acid sites, as expected if these genes are involved in *S*-pollen specificity determination (results available at <http://bpositive.i3s.up.pt/> under project Rosa *S*-locus genes; BP2018000004). On the predicted 3D structure, these amino acid sites are located in the same regions (Fig. 6) as those observed for *Petunia S*-pollen genes<sup>26</sup>. Evidence for positive selection is also observed for the five F-box genes of the *R. multiflora* scaffold sca0006888 (results available at <http://bpositive.i3s.up.pt/> under project Rosa *S*-locus genes; BP2018000004). Therefore, the data suggests that *Rosa S*-pollen specificity is determined by multiple F-box genes, like in *Malus/Pyrus/Sorbus* and Solanaceae species.

## Discussion

In *Rosa*, there are very important traits of horticultural interest such as flower development, architecture, senescence, scent biosynthesis and emission, ease of reproduction, and resistance to biotic and abiotic stresses, that have been selected only once during the history of rose selection, and incorporated into many rose varieties<sup>44, 45, 48</sup>. Indeed, only 10 species have contributed to the genetic make-up of most of the modern rose cultivars, and some old and popular cultivars, such as ‘Old Blush’ have dominated the history of rose selection<sup>40</sup>. This Chinese rose from the Song dynasty (960–1279) conveys several desirable characters such as recessive reblooming habit, recessive lack of prickles (stem) and dominant flower doubleness<sup>44, 45, 48</sup>, all co-segregating with the *S*-locus region. The characterization of the *S*-locus region here performed is thus, very important in order to help breeding selection, and the control of genetic diversity.

The *Rosa S*-locus is composed of a *S-RNase* gene that belongs to the *Prunus* lineage (Fig. 1). This gene shows all expected features of a *S*-pistil gene since it shows expression in pistil and ovary (Fig. 2), evidence for diversifying selection (Fig. 3), and co-segregation with the *S*-locus. In *Fragaria*, that also belongs to the Rosoideae subfamily (the two genera have been diverging for about 50 million years<sup>54</sup>), a *Prunus* lineage *S-RNase* gene has been also reported as the putative *S*-pistil gene<sup>24</sup>. This suggests that the Rosoideae GSI system could be of the self-recognition type, with one *S*-pollen gene, as in *Prunus*<sup>24</sup> and references therein). Nevertheless, the *Rosa* F-box gene that clusters with the *Prunus SFB* gene (Fig. 4) is not the neighbor of the *S-RNase* gene, as observed in *Prunus*, and presents expression (Fig. 5) and polymorphism levels incompatible with being the *S*-pollen gene. In *Rosa* there are multiple F-box genes in the vicinity of the *S-RNase*, that show expression compatible with being



**Figure 6.** Positively selected amino acid sites on the predicted 3D structure of the *R. chinensis* F-box located in the 5' region of the *S-RNase*. In yellow are highlighted those sites that appear as positively selected in the two datasets analyzed (14 and 13 *R. chinensis* F-box genes in the vicinity of the *S-RNase* that are expressed in stamen), in green those that are located in a region not analyzed when 14 F-box genes are considered, and in blue those that change due to alignment gaps in the two datasets analyzed.

the *S*-pollen gene (Fig. 5), and evidence for diversifying selection for F-box genes within a *S*-haplotype (Fig. 6). These are the expected features of *S*-pollen genes in a non-self-recognition system, as reported in *Malus/Sorbus/Pyrus*<sup>13, 16, 18–20, 24, 25</sup> and Solanaceae species<sup>15, 22, 23, 26, 30–32</sup>. Therefore, the *Rosa* *S*-locus region encompasses a large region, as reported for other species presenting non-self recognition systems, and this could explain why in rose several traits of horticultural interest are linked to the *S*-locus. It should be noted that the *Rosa* *S*-locus region may be highly rearranged, since we could not align the contig containing the *S-RNase* gene from *R. chinensis* with those from *R. multiflora* as well as the two contigs containing the *S-RNase* gene from *R. multiflora*. This observation has been previously reported for the two available *R. chinensis* genomes, that could not be aligned in this region (conceivably carrying two different *S*-haplotypes; see Fig. 1 of<sup>55</sup>). Pollen transcriptome analyses of multiple *S*-haplotypes are, however needed to determine how many F-box genes in the vicinity of the *Rosa S-RNase* are involved in pollen GSI specificity, as performed in *Malus*<sup>25</sup> and *Petunia*<sup>21</sup>.

Since multiple *S*-pollen genes are determining *Rosa* *S*-pollen specificity, as in Maleae species, this suggests that the ancestral Rosaceae GSI system was of the non-self-recognition system type. This implies that during evolution the *Malus S-RNase* lineage, that does not cluster with the *Rosa S-RNase*, has been recruited de novo from a duplicate of the ancestral *S-RNase* gene. Moreover, it implies that the *Prunus S*-pollen gene, that does not cluster with *Rosa S*-pollen genes, evolved de novo from an unrelated F-box gene.

Low levels of variation have been reported in roses but the *S*-locus region, being large and under balancing selection, may help retain a substantial fraction of the variability. Indeed, the evolution within the genus *Rosa*, occurred via interspecific hybridization, allopolyploidization and genetic reticulation among sympatric species<sup>56</sup>. In Europe, the recent post-glaciation period expedited the process of speciation, as the northward extension of the biotopes increased. The diploid *R. arvensis* holds a special place in this process, since, although it belongs to the synstylae group, has a substantial genomic promiscuity with the polyploid caninae group<sup>56, 57</sup>, that may also help maintain diversity levels. The inbred strains that are being established for *R. arvensis*<sup>53</sup> are a starting point to investigate the diversity of the *S*-locus in European wild roses. Mutants breaking down self-incompatibility, will also help refine the *S*-locus effect on roses molecular diversity.

## Methods

**Identification of putative *S-RNase* like sequences in *Rosa* genomes.** The annotations (CDS) of two *R. chinensis* genomes (available at NCBI RefSeq database ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) and GDR database (<https://www.rosaceae.org>) Supplementary Table S1) were downloaded, and the sequences showing similarity with reference *S-RNases* assigned as Rchinensis1 and Rchinensis3, respectively. Moreover, the corresponding genome sequences were downloaded as a FASTA file, and the *S-RNase* sequences here annotated labeled as Rchinensis2



and *Rchinensis*<sup>4</sup>, respectively. *R. multiflora* genome (NCBI assembly database, Supplementary Table S1) was also downloaded as a FASTA file, since annotations are not available for this species. To find and extract protein encoding segments larger than 100 bp, we have used getorf, using the emboss Docker image available at pegi3s Bioinformatics Docker images project (<https://pegi3s.github.io/dockerfiles>). Then, we selected the protein encoding segments that show similarity with reference *S-RNase* sequences, using tblastx (Expect value (e) < 0.05), as implemented in SEDA<sup>58, 59</sup>, using as query *Prunus S3-RNase* (AJ298312), *Malus Sh-RNase* (AB032247), and *Fragaria* putative *S-RNase* (gi561957436, gi561674690 and gi561985884)<sup>24</sup>. Based on this information, we manually annotated the corresponding genome region to identify the exons of each gene. For each putative gene we obtained the corresponding amino acid sequence to calculate IP, using ExPASy software<sup>60</sup>.

We used the same protocol for the genomes here assembled using the short reads of nine *Rosa* genomes downloaded from NCBI SRA database (Supplementary Table S1). In this case, we used FastQC to evaluate read quality, and Cutadapt to trim reads<sup>61</sup>, and ABySS 2.0<sup>62</sup> for the de novo assembly, using the Docker images available at pegi3s Bioinformatics Docker images project (<https://pegi3s.github.io/dockerfiles>).

**Identification of the F-box genes located in *R. chinensis* chromosome 3 and *R. multiflora* scaffolds where the *S-RNase* is located.** The protocol presented for the *S-RNase* like sequences was also used to obtain F-box genes located in *R. chinensis* chromosome 3, and for the two *R. multiflora* scaffolds where the *S-RNase* gene has been identified, using as query *P. avium* *SFB3* (AAT72121.1), *P. avium* *SLFL1* (BAG12295.1), and *M. domestica* *SFBB3-beta* (BAF47180.1).

**Phylogenetic analyses.** Phylogenetic analyses of the *R. chinensis* and *R. multiflora* *S-RNase* like sequences and F-box like genes were performed using sequences aligned with MUSCLE alignment algorithm, as implemented in ADOPS<sup>63</sup>. Only codons with a support value above 2 were used for phylogenetic reconstruction. Bayesian trees were obtained using MrBayes 3.1.2<sup>64</sup> as implemented in the ADOPS pipeline<sup>63</sup>. The Generalised Time-Reversible (GTR) model of sequence evolution was implemented in the analyses, allowing for among-site rate variation and a proportion of invariable sites. Third codon positions were allowed to have a gamma distribution shape parameter different from that of first and second codon positions. Two independent runs of 1,000,000 generations with four chains each, were carried out. The average standard deviation of split frequencies was always ~ 0.01 and the potential scale reduction factor for every parameter was ~ 1.00, showing that convergence was achieved. Trees were sampled every 100th generation and the first 5000 samples were discarded (burn-in). The tree was converted to Newick format using the Format Conversion website ([http://phylogeny.lirmm.fr/phylo.cgi/data\\_converter.cgi](http://phylogeny.lirmm.fr/phylo.cgi/data_converter.cgi)) and edited using Mega<sup>765</sup>.

The phylogenetic analyses of the low coverage *Rosa* genomes (Supplementary Table S1) sequences were performed with Mega<sup>765</sup>, using ClustalW alignment algorithm, Neighbor-Joining method, bootstrap test with 10,000 replicates, the p-distance method for computing the evolutionary distances, and pairwise deletion since sequences can have different sizes (Supplementary Table S2).

**Expression of *R. chinensis* *S-RNase* like, and F-box genes located on chromosome 3 in pistil and ovary, stamen, leaf, stem and root transcriptomes.** To estimate expression of the *Rosa S-RNase* like sequences located on chromosome 3, we use RNA-seq data from *R. chinensis* pistil and ovary, stamen, leaf, stem, and root transcriptomes (Supplementary Table S4). We used FastQC to evaluate read quality, and Cutadapt to trim reads<sup>61</sup>. FPKM values were estimated using the RSEM method, as implemented in Trinity<sup>66</sup>, using the *R. chinensis* Refseq CDS, and the *R. chinensis* *S-RNase* like and F-box sequences located on chromosome 3.

***R. arvensis* genotype–phenotype association experiments.** Genomic DNA was extracted from leaves of 12 *R. arvensis* individuals, for which the haplotypes were predicted according to their parents and progeny (Supplementary Fig. S3A,B), using the method of<sup>67</sup>. PCRs were performed using the genomic DNA and primers RA-F 5' GGAAGCCARACTGAAGAT 3' and RA-R 5' AGCATCACAGTYTCGATCA 3', designed for conserved regions of the putative *Rosa S-RNase* sequences here identified. Standard amplification conditions were 35 cycles of denaturation at 94 °C for 30 s, 52 °C for 30 s for primer annealing, and primer extension at 72 °C for 2 min. The amplification products with the expected size (353 bp) for individuals Ose, E200, E404, Url, and E893 were cloned using the TA cloning kit (Invitrogen, Carlsbad, CA). For each individual, the insert of 16 colonies was cut separately with *DdeI*, and *HinfI* restriction enzymes, and only one restriction pattern was observed, and thus three colonies only were sequenced. The ABI PRISM BigDye cycle-sequencing kit (Perkin Elmer, Foster City, CA), and specific primers, or the primers for the M13 forward and reverse priming sites of the pCR2.1 vector, were used to prepare the sequencing reactions. Sequencing runs were performed by STABVIDA (Lisboa, Portugal).

**Identification of positively selected amino acid sites, their location on the crystal structure, and polymorphism levels.** For the six sequences identified as putative *Rosa S-RNases*, we inferred positively selected amino acid sites, using codeML<sup>51</sup>, as implemented in ADOPS<sup>63</sup>, using Muscle as the alignment method. Such analyses were also performed for the 14 *R. chinensis* F-box sequences that are in the vicinity of the *S-RNase*, that cluster with *Prunus SLFL*, and that are expressed in stamen. Since the inclusion of *Rchinensis\_F-box-3* gene sequence excludes from the analyses a large fraction of the 3' region, we performed these analyses after removing this sequence. codeML analyses were also performed for five *R. multiflora* sc0006888 F-box sequences in the vicinity of the *S-RNase*. The details of the analyses can be seen at the B+ database (bpositive.i3s.up.pt<sup>68</sup>; *Rosa S-locus* genes BP2018000004). Model comparisons were M2a-M1a and M8-M7. We consider as

positively selected those amino acid sites that show a probability higher than 90% for both naive empirical Bayes (NEB) or Bayes empirical Bayes (BEB) methods.

To visualize these positions in the 3D structure, for the S-RNase (translation of Rchinensis1\_3–Rchinensis2\_27; Supplementary File S1) we first identified the signal peptide using SignalIP (<http://www.cbs.dtu.dk/services/SignalP/>) website tool available at ExpASY<sup>60</sup>. After removing the signal peptide, the 3D structure was modeled by I-Tasser<sup>69</sup>, and the model with the highest C-score value used. The same methodology was used for the putative S-pollen gene (translation of R chinensis\_F box-1; Supplementary File S4), but in this case after removing the F-box domain (the first 60 amino acid positions). All structural images were produced using PyMOL (The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC.).

Levels of polymorphism were obtained with DnaSp<sup>70</sup>.

## Data availability

All data generated or analyzed during this study are included in this published article (and its Supplementary Data File S1–S4). *R. arvensis* S2- and S5-RNase are deposited in GenBank, acc. Numbers MW452856–MW452861.

Received: 31 March 2020; Accepted: 19 January 2021

Published online: 12 February 2021

## References

- De Nettancourt, D. *Incompatibility in angiosperms* (Springer, Berlin, 1977).
- Igic, B., Lande, R. & Kohn, J. R. Loss of self-incompatibility and its evolutionary consequences. *Int. J. Plant Sci.* **169**, 93–104 (2008).
- Rosalson, E. H. & McCubbin, A. G. S-RNases and sexual incompatibility: Structure, functions, and evolutionary perspectives. *Mol. Phylogenet. Evol.* **29**, 490–506. [https://doi.org/10.1016/s1055-7903\(03\)00195-7](https://doi.org/10.1016/s1055-7903(03)00195-7) (2003).
- McClure, B. Darwin's foundation for investigating self-incompatibility and the progress toward a physiological model for S-RNase-based SI. *J. Exp. Bot.* **60**, 1069–1081. <https://doi.org/10.1093/jxb/erp024> (2009).
- Nowak, M. D., Davis, A. P., Anthony, F. & Yoder, A. D. Expression and trans-specific polymorphism of self-incompatibility RNases in coffee (Rubiaceae). *PLoS ONE* **6**, e21019. <https://doi.org/10.1371/journal.pone.0021019> (2011).
- Entani, T. *et al.* 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. <https://doi.org/10.1046/j.1365-2443.2003.00626.x> (2003).
- Ushijima, K. *et al.* 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. <https://doi.org/10.1105/tpc.009290> (2003).
- Qiao, H. *et al.* The F-box protein AhSLF-S2 controls the pollen function of S-RNase-based self-incompatibility. *Plant Cell* **16**, 2307–2322. <https://doi.org/10.1105/tpc.104.024919> (2004).
- Tsakamoto, T., Ando, T., Watanabe, H., Marchesi, E. & Kao, T. H. Duplication of the S-locus F-box gene is associated with breakdown of pollen function in an S-haplotype identified in a natural population of self-incompatible *Petunia axillaris*. *Plant Mol. Biol.* **57**, 141–153. <https://doi.org/10.1007/s11103-004-6852-6> (2005).
- Cheng, J., Han, Z., Xu, X. & Li, T. Isolation and identification of the pollen-expressed polymorphic F-box genes linked to the S-locus in apple (*Malus × domestica*). *Sex. Plant Reprod.* **19**, 175. <https://doi.org/10.1007/s00497-006-0034-4> (2006).
- Nunes, M. D., Santos, R. A., Ferreira, S. M., Vieira, J. & Vieira, C. P. Variability patterns and positively selected sites at the gametophytic self-incompatibility pollen SFB gene in a wild self-incompatible *Prunus spinosa* (Rosaceae) population. *New Phytol.* **172**, 577–587. <https://doi.org/10.1111/j.1469-8137.2006.01838.x> (2006).
- Hua, Z., Meng, X. & Kao, T. H. Comparison of *Petunia inflata* S-Locus F-box protein (Pi SLF) with Pi SLF like proteins reveals its unique function in S-RNase based self-incompatibility. *Plant Cell* **19**, 3593–3609. <https://doi.org/10.1105/tpc.107.055426> (2007).
- Sassa, H. *et al.* 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. <https://doi.org/10.1534/genetics.106.068858> (2007).
- Wheeler, D. & Newbigin, E. Expression of 10 S-class SLF-like genes in *Nicotiana glauca* pollen and its implications for understanding the pollen factor of the S locus. *Genetics* **177**, 2171–2180. <https://doi.org/10.1534/genetics.107.076885> (2007).
- Kubo, K. *et al.* Collaborative non-self recognition system in S-RNase-based self-incompatibility. *Science* **330**, 796–799. <https://doi.org/10.1126/science.1195243> (2010).
- Minamikawa, M. *et al.* Apple S locus region represents a large cluster of related, polymorphic and pollen-specific F-box genes. *Plant Mol. Biol.* **74**, 143–154. <https://doi.org/10.1007/s11103-010-9662-z> (2010).
- Tsakamoto, T., Hauck, N. R., Tao, R., Jiang, N. & Iezzoni, A. F. Molecular and genetic analyses of four nonfunctional S haplotype variants derived from a common ancestral S haplotype identified in sour cherry (*Prunus cerasus* L.). *Genetics* **184**, 411–427. <https://doi.org/10.1534/genetics.109.109728> (2010).
- Kakui, H. *et al.* Sequence divergence and loss-of-function phenotypes of S locus F-box brothers genes are consistent with non-self recognition by multiple pollen determinants in self-incompatibility of Japanese pear (*Pyrus pyrifolia*). *Plant J.* **68**, 1028–1038. <https://doi.org/10.1111/j.1365-3113X.2011.04752.x> (2011).
- Okada, K. *et al.* Related polymorphic F-box protein genes between haplotypes clustering in the BAC contig sequences around the S-RNase of Japanese pear. *J. Exp. Bot.* **62**, 1887–1902. <https://doi.org/10.1093/jxb/erq381> (2011).
- Aguiar, B. *et al.* Patterns of evolution at the gametophytic self-incompatibility *Sorbus aucuparia* (Pyrinae) S pollen genes support the non-self recognition by multiple factors model. *J. Exp. Bot.* **64**, 2423–2434. <https://doi.org/10.1093/jxb/ert098> (2013).
- Williams, J. S., Der, J. P., dePamphilis, C. W. & Kao, T. H. Transcriptome analysis reveals the same 17 S-locus F-box genes in two haplotypes of the self-incompatibility locus of *Petunia inflata*. *Plant Cell* **26**, 2873–2888. <https://doi.org/10.1105/tpc.114.126920> (2014).
- Williams, J. S. *et al.* Four previously identified *Petunia inflata* S-locus F-box genes are involved in pollen specificity in self-incompatibility. *Mol. Plant* **7**, 567–569. <https://doi.org/10.1093/mp/sst155> (2014).
- Kubo, K. *et al.* Gene duplication and genetic exchange drive the evolution of S-RNase-based self-incompatibility in *Petunia*. *Nat. Plants* **1**, 14005. <https://doi.org/10.1038/nplants.2014.5> (2015).
- Aguiar, B. *et al.* Convergent evolution at the gametophytic self-incompatibility system in *Malus* and *Prunus*. *PLoS ONE* **10**, e0126138. <https://doi.org/10.1371/journal.pone.0126138> (2015).
- Pratas, M. I. *et al.* Inferences on specificity recognition at the *Malus domestica* gametophytic self-incompatibility system. *Sci. Rep.* **8**, 1717. <https://doi.org/10.1038/s41598-018-19820-1> (2018).
- Vieira, J. *et al.* Predicting specificities under the non-self gametophytic self-incompatibility recognition model. *Front. Plant Sci.* **10**, 879. <https://doi.org/10.3389/fpls.2019.00879> (2019).
- Ikeda, K. *et al.* Primary structural features of the S haplotype-specific F-box protein, SFB *Prunus*. *Sex. Plant Reprod.* **16**, 235–243. <https://doi.org/10.1007/s00497-003-0200-x> (2004).

28. Matsumoto, D. & Tao, R. Distinct Self-recognition in the Prunus S-RNase-based gametophytic self-incompatibility system. *Hortic. J.* **85**. <https://doi.org/10.2503/hortj.MI-IR06> (2016).
29. Tsukamoto, T. *et al.* Genetic and molecular characterization of three novel S-haplotypes in sour cherry (*Prunus cerasus* L.). *J. Exp. Bot.* **59**, 3169–3185. <https://doi.org/10.1093/jxb/ern172> (2008).
30. Sun, P., Li, S., Lu, D., Williams, J. S. & Kao, T. H. Pollen S-locus F-box proteins of *Petunia* involved in S-RNase-based self-incompatibility are themselves subject to ubiquitin-mediated degradation. *Plant J.* **83**, 213–223. <https://doi.org/10.1111/tpj.12880> (2015).
31. Li, J. *et al.* Electrostatic potentials of the S-locus F-box proteins contribute to the pollen S specificity in self-incompatibility in *Petunia hybrida*. *Plant J.* **89**, 45–57. <https://doi.org/10.1111/tpj.13318> (2017).
32. Wu, L. *et al.* Use of domain-swapping to identify candidate amino acids involved in differential interactions between two allelic variants of type-1 S-locus F-box protein and S3-RNase in *petunia inflata*. *Plant Cell Physiol.* **59**, 234–247. <https://doi.org/10.1093/pcp/pcx176> (2018).
33. Igic, B. & Kohn, J. R. Evolutionary relationships among self-incompatibility RNases. *Proc. Natl. Acad. Sci. USA* **98**, 13167–13171. <https://doi.org/10.1073/pnas.231386798> (2001).
34. Steinbachs, J. E. & Holsinger, K. E. S-RNase-mediated gametophytic self-incompatibility is ancestral in eudicots. *Mol. Biol. Evol.* **19**, 825–829. <https://doi.org/10.1093/oxfordjournals.molbev.a004139> (2002).
35. Vieira, J., Morales-Hojas, R., Santos, R. A. & Vieira, C. P. Different positively selected sites at the gametophytic self-incompatibility pistil S-RNase gene in the Solanaceae and Rosaceae (*Prunus*, *Pyrus*, and *Malus*). *J. Mol. Evol.* **65**, 175–185. <https://doi.org/10.1007/s00239-006-0285-6> (2007).
36. Vieira, J., Fonseca, N. A. & Vieira, C. P. An S-RNase-based gametophytic self-incompatibility system evolved only once in eudicots. *J. Mol. Evol.* **67**, 179–190. <https://doi.org/10.1007/s00239-008-9137-x> (2008).
37. Ramanaukas, K. & Igic, B. The evolutionary history of plant T2/S-type ribonucleases. *PeerJ* **5**, e3790. <https://doi.org/10.7717/peerj.3790> (2017).
38. Tao, R. & Iezzoni, A. F. The S-RNase-based gametophytic self-incompatibility system in *Prunus* exhibits distinct genetic and molecular features. *Sci. Hortic.* **124**, 423–433. <https://doi.org/10.1016/j.scienta.2010.01.025> (2010).
39. Akagi, T., Henry, I. M., Morimoto, T. & Tao, R. Insights into the prunus-specific S-RNase-based self-incompatibility system from a genome-wide analysis of the evolutionary radiation of S locus-related F-box genes. *Plant Cell Physiol.* **57**, 1281–1294. <https://doi.org/10.1093/pcp/pcw077> (2016).
40. Gudin, S. Rose: genetics and breeding. *Plant Breed. Rev.* **17**, 159–189. <https://doi.org/10.1002/9780470650134.ch3> (2000).
41. Cole, P. & Melton, B. Self-compatibility and cross-compatibility relationships among genotypes and between ploidy of the rose. *J. Am. Soc. Hortic. Sci.* **111**, 122–125 (1986).
42. Heslop-Harrison, Y. & Shivanna, K. R. The receptive surface of the angiosperms stigma. *Ann. Bot.* **41**, 1233–1258. <https://doi.org/10.1093/oxfordjournals.aob.a085414> (1977).
43. Ueda, Y. & Akimoto, S. Cross- and self-compatibility in various species of the Genus *Rosa*. *J. Hortic. Sci. Biotechnol.* **76**, 392–395. <https://doi.org/10.1080/14620316.2001.11511382> (2001).
44. Hibrand Saint-Oyant, L. *et al.* A high-quality genome sequence of *Rosa chinensis* to elucidate ornamental traits. *Nat. Plants* **4**, 473–484. doi:<https://doi.org/10.1038/s41477-018-0166-1> (2018).
45. Debener, T. *et al.* Genetic and molecular analyses of key loci involved in self-incompatibility and floral scent in roses. *Acta Hort.* **870**, 183–190 (2010).
46. Wei, Y. *et al.* Cloning and bioinformatics analysis of *Rosa rugosa* S Locus F-Box Gene (*RrSLE*). *Am. J. Plant Sci.* **8**, 1550–1559. <https://doi.org/10.4236/ajps.2017.87107> (2017).
47. Yokoya, K., Roberts, A. V., Mottley, J., Lewis, R. & Brandham, P. E. Nuclear DNA amounts in roses. *Ann. Bot.* **85**, 557–561. <https://doi.org/10.1006/anbo.1999.1102> (2000).
48. Bendahmane, M., Dubois, A., Raymond, O. & Bris, M. L. Genetics and genomics of flower initiation and development in roses. *J. Exp. Bot.* **64**, 847–857. <https://doi.org/10.1093/jxb/ers387> (2013).
49. Raymond, O. *et al.* The *Rosa* genome provides new insights into the domestication of modern roses. *Nat. Genet.* **50**, 772–777. <https://doi.org/10.1038/s41588-018-0110-3> (2018).
50. Nakamura, N. *et al.* Genome structure of *Rosa multiflora*, a wild ancestor of cultivated roses. *DNA Res.* **25**, 113–121. <https://doi.org/10.1093/dnares/dsx042> (2018).
51. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591. <https://doi.org/10.1093/molbev/msm088> (2007).
52. Vieira, J., Ferreira, P. G., Aguiar, B., Fonseca, N. A. & Vieira, C. P. Evolutionary patterns at the RNase based gametophytic self-incompatibility system in two divergent Rosaceae groups (Maloideae and Prunus). *BMC Evol. Biol.* **10**, 200. <https://doi.org/10.1186/1471-2148-10-200> (2010).
53. Heitzler, P. *Rosa arvensis* as a possible genetic model. *Acta Hort.* **1232**, 25–32. <https://doi.org/10.17660/ActaHortic.2019.1232.5> (2019).
54. Xiang, Y. *et al.* Evolution of Rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication. *Mol. Biol. Evol.* **34**, 262–281. <https://doi.org/10.1093/molbev/msw242> (2017).
55. Smulders, M. J. M. *et al.* In the name of the rose: A roadmap for rose research in the genome era. *Hortic. Res.* **6**, 65. <https://doi.org/10.1038/s41438-019-0156-0> (2019).
56. Zhu, Z. M., Gao, X. F. & Fougere-Danezan, M. Phylogeny of *Rosa* sections *Chinenses* and *Synstylae* (Rosaceae) based on chloroplast and nuclear markers. *Mol. Phylogenet. Evol.* **87**, 50–64. <https://doi.org/10.1016/j.ympev.2015.03.014> (2015).
57. De Cock, K. Genetic diversity of wild roses (*Rosa* spp.) in Europe, with an in-depth morphological study of Flemish populations. *Ghent University. Faculty of Bioscience Engineering, Ghent, Belgium.* (2008).
58. López-Fernández, H. *et al.* Bioinformatics protocols for quickly obtaining large-scale data sets for phylogenetic inferences. *Interdiscip. Sci. Comput. Life Sci.* **11**, 1–9. <https://doi.org/10.1007/s12539-018-0312-5> (2019).
59. López-Fernández, H. *et al.* SEDa: a desktop tool suite for FASTA files processing. *IEEE/ACM Trans. Comput. Biol. Bioinform.* <https://doi.org/10.1109/TCBB.2020.3040383> (2020).
60. Artimo, P. *et al.* ExpASY: SIB bioinformatics resource portal. *Nucleic Acids Res.* **40**, W597–603. <https://doi.org/10.1093/nar/gks400> (2012).
61. Martin, M. CUTADAPT removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* **17**, doi:<https://doi.org/10.14806/ej.17.1.200> (2011).
62. Jackman, S. D. *et al.* ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. *Genome Res.* **27**, 768–777. <https://doi.org/10.1101/gr.214346.116> (2017).
63. Reboiro-Jato, D. *et al.* ADOPS—Automatic Detection Of Positively Selected Sites. *J Integr Bioinform* **9**, 200. <https://doi.org/10.2390/biecoll-jib-2012-200> (2012).
64. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755. <https://doi.org/10.1093/bioinformatics/17.8.754> (2001).
65. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874. <https://doi.org/10.1093/molbev/msw054> (2016).
66. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512. <https://doi.org/10.1038/nprot.2013.084> (2013).

67. Ingram, G. C. *et al.* Dual role for fimbriata in regulating floral homeotic genes and cell division in *Antirrhinum*. *EMBO J.* **16**, 6521–6534. <https://doi.org/10.1093/emboj/16.21.6521> (1997).
68. Vazquez, N. *et al.* Large scale analyses and visualization of adaptive amino acid changes projects. *Interdiscip. Sci.* **10**, 24–32. <https://doi.org/10.1007/s12539-018-0282-7> (2018).
69. Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **12**, 7–8. <https://doi.org/10.1038/nmeth.3213> (2015).
70. Rozas, J., Sanchez-DelBarrio, J. C., Messeguer, X. & Rozas, R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496–2497. <https://doi.org/10.1093/bioinformatics/btg359> (2003).

## Acknowledgements

This work was financed by the National Funds through FCT—Fundação para a Ciência e a Tecnologia, I.P., under the project UIDB/04293/2020, and the Centre National de la Recherche Scientifique (CNRS).

## Author contributions

J.V., and C.P.V. designed the research. P. H. collected the plant material and designed the *R. arvensis* genetics. J.V., J. P., A.G., J. L., S. R., and C.P.V. performed the genomic and transcriptomic analyses. J. P., and C.P.V. performed the molecular work. All authors reviewed and approved the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-83243-8>.

**Correspondence** and requests for materials should be addressed to C.P.V.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021