Patterns of evolution at the gametophytic self-incompatibility
Sorbus aucuparia (Pyrinae) S pollen genes support the non-self recognition by multiple factors model

Bruno Aguiar1, Jorge Vieira1, Ana E. Cunha1, Nuno A. Fonseca2, David Reboiro-Jato4, Miguel Reboiro-Jato4, Florentino Fdez-Riverola4, Olivier Raspé3 and Cristina P. Vieira1,∗

1 Instituto de Biologia Molecular e Celular (IBMC), University of Porto, Rua do Campo Alegre 823, 4150–180 Porto, Portugal
2 CRACS-INESC Porto, Rua do Campo Alegre 1021/1055, 4169-007 Porto, Portugal
3 National Botanic Garden of Belgium, Domein van Bouchout, B-1860 Meise, Belgium
4 Escuela Superior de Ingeniería Informática, Edificio Politécnico, Campus Universitario As Lagoas s/n, University of Vigo, 32004 Ourense, Spain

∗ To whom correspondence should be addressed. Email: cgvieira@ibmc.up.pt

Received 12 December 2012; Revised 5 March 2013; Accepted 11 March 2013

Abstract

S-RNase-based gametophytic self-incompatibility evolved once before the split of the Asteridae and Rosidae. In Prunus (tribe Amygdaloideae of Rosaceae), the self-incompatibility S-pollen is a single F-box gene that presents the expected evolutionary signatures. In Malus and Pyrus (subtribe Pyrinae of Rosaceae), however, clusters of F-box genes (called SFBBs) have been described that are expressed in pollen only and are linked to the S-RNase gene. Although polymorphic, SFBB genes present levels of diversity lower than those of the S-RNase gene. They have been suggested as putative S-pollen genes, in a system of non-self recognition by multiple factors. Subsets of allelic products of the different SFBB genes interact with non-self S-RNases, marking them for degradation, and allowing compatible pollinations. This study performed a detailed characterization of SFBB genes in Sorbus aucuparia (Pyrinae) to address three predictions of the non-self recognition by multiple factors model. As predicted, the number of SFBB genes was large to account for the many S-RNase specificities. Secondly, like the S-RNase gene, the SFBB genes were old. Thirdly, amino acids under positive selection—those that could be involved in specificity determination—were identified when intra-haplotype SFBB genes were analysed using codon models. Overall, the findings reported here support the non-self recognition by multiple factors model.

Key words: gametophytic self-incompatibility, molecular evolution, positively selected amino acid sites, SFBB, Sorbus aucuparia, S-RNase.

Introduction

Self-incompatibility (SI) is a genetic barrier to self-fertilization in which the female reproductive cells discriminate between genetic relative and non-relative pollen, and reject the former (De Nettancourt, 1977). In order to maintain functional incompatibility, the S genes, those determining the pistil and pollen specificities, must co-evolve as a genetic unit (for details on co-evolution see, Newbigin et al., 2008). In gametophytic SI (GSI), the pollen is rejected when it expresses a specificity that matches either of those expressed in the style. In this system, because of frequency-dependent selection, many specificities are maintained in natural populations (Wright, 1939).

The self-incompatibility S-pistil gene product in Rosaceae, Rubiaceae, Solanaceae, and Plantaginaceae is an extracellular ribonuclease, called S-RNase (Roalson and McCubbin, 2003; Nowak et al., 2011). Phylogenetic analyses suggest
that S-RNase-based GSI has evolved only once, before the split of the Asteridae and Rosidae, about 120 million years ago (MYA) (Igic and Kohn, 2001; Steinbachs and Holsinger, 2002; Vieira et al., 2008a). Because of the single origin of this system, in principle, similarities are expected when comparing the GSI players in these plant families.

In Rosaceae, for the pistil gene, studies at the molecular level have been performed in species of the tribe Amygdaloideae (Prunus) and subtribe Pyrinae (Malus, Pyrus, Sorbus, and Crataegus; see references in Vieira et al., 2010). The pistil gene shows the expected features for a gene determining GSI specificity, namely, high levels of synonymous and non-synonymous divergence, as well as positively selected amino acid sites that account for the many specificities known to be present in natural populations (Vieira et al., 2007).

The pollen component, always an F-box protein, has been identified as one gene in Prunus (called SFB, S-haplotype-specific F-box gene) (Ushijima et al., 2001; Entani et al., 2003; Ushijima et al., 2003; Ikeda et al., 2004; Romero et al., 2004; Sonneveld et al., 2005; Nunes et al., 2006; Vieira et al., 2008b), but multiple genes in Malus, Pyrus, (called SFBBs, S-locus F-box brothers), Petunia (Solanaceae), and Nicotiana (Solanaceae; called SLFs, S-locus F-boxes) (Cheng et al., 2006; Kakui et al., 2007; Sassa et al., 2007; Wheeler and Newbiggin, 2007; Kubo et al., 2010; Minamikawa et al., 2010). In Prunus, the non-self S-RNases taken up by the growing pollen tube are postulated to be inactivated by a general inhibitor and only the self S-RNases are protected from inactivation (Luu et al., 2001; Sonneveld et al., 2005). In the multiple S pollen genes system, found in Malus, Pyrus, Petunia, and Nicotiana, within an S haplotype, each SLF/SFBB is predicted to interact with a subset of non-self S-RNases, and mediates their degradation by the ubiquitin–26S proteasome system. Multiple SLF/SFBB genes are thus required for the recognition of the large repertoire of non-self S-RNases (Kubo et al., 2010) under this protein degradation model (Hua and Kao, 2006; Hua et al., 2007). Therefore, SLF/SFBB genes could represent a case of paralogous gene expansions that are often the substrate for adaptive change.

The very different mechanisms of S-RNase inhibition in systems with one or multiple S pollen genes imply remarkable differences at the S-locus region. Indeed, in Prunus, as expected, if a single S pollen gene is involved, the S-locus region is smaller than 45 kb (Vieira et al., 2008e). The S-RNase gene is flanked by two F-box like genes, namely, SFB and SLFL1 (SLF-like gene 1). Nevertheless, in contrast to the SLFL1 gene, the SFB gene is expressed only in pollen, and its average divergence is similar to that observed for the S-RNase gene (Ushijima et al., 2001; Entani et al., 2003; Ushijima et al., 2003; Vieira et al., 2008a). Despite evidence for specific associations between SLFL1 and S-RNase - SFB genes, as well as the relatively old age of SLFL1 alleles (Vieira et al., 2008a), amino acid sites showing strong evidence for positive selection have been identified only in the SFB gene (Nunes et al., 2006; Vieira et al., 2008b,d). Despite the lack of congruent tree topologies for the S-RNase and SFB genes (Nunes et al., 2006; Tsukamoto et al., 2008), the two genes show evidence for a partially co-evolved history (Tsukamoto et al., 2008).

Under the multiple S pollen genes system, SFBB genes are expected to be in linkage with the S-RNase gene and to have pollen expression only. In Malus, the extent of the S-locus region is unknown but is larger than 317 kb (corresponding to the BAC contig s analysed by Sassa et al., 2007, and Minamikawa et al., 2010). In this region, two F-box genes, MdSFBB9-alpha and MdSFBB9-beta (located 42 kb upstream and 93 kb downstream of the S-RNase, respectively), with expression restricted to pollen, and in linkage with the S-RNase gene, were initially reported as S pollen genes (Sassa et al., 2007). The number of these genes in the Malus S-locus region is currently greater than ten (Minamikawa et al., 2010; Sassa et al., 2010). In Pyrus, the size of the S-locus region is also unknown but is larger than 649 kb (Okada et al., 2011), and the number of F-box genes with expression restricted to pollen and in linkage with the S-RNase gene is greater than eight (Minamikawa et al., 2010; Sassa et al., 2010; De Franceschi et al., 2011a,b; Kakui et al., 2011; Okada et al., 2011). The order of the SFBB genes in different S haplotypes is not conserved (Minamikawa et al., 2010; Okada et al., 2011). Therefore, depending on the haplotype analysis in the segregation experiments, an SFBB gene can show linkage with an S-RNase allele but incomplete linkage with another (Pyrus SFBB alpha, and Pyrus SFBB-gamma in De Franceschi et al., 2011a, and SFBB6 in Kakui et al., 2011). It is possible that SFBB alleles not showing linkage to the S-RNase may have no target S-RNase allele and thus are not being constrained by selection (Kakui et al., 2011). These genes have been assigned as not being involved in S pollen specificity (De Franceschi et al., 2012).

Kakui et al. (2007) reported low polymorphism for the PpSFBB-gamma gene compared with the S-RNase gene. This pattern seems to be a common feature of all SFBB genes (Minamikawa et al., 2010; De Franceschi et al., 2011a; Kakui et al., 2011; Okada et al., 2011) and SLFs (Zhou et al., 2003; Wheeler and Newbiggin, 2007; Newbiggin et al., 2008; Kubo et al., 2010). Identical alleles at one SFB gene have been reported from two different S haplotypes (Minamikawa et al., 2010). Another pattern is the high divergence between SFBB genes, comparable to the allelic diversity of the S-RNase gene (Minamikawa et al., 2010; De Franceschi et al., 2011a; Kakui et al., 2011; Okada et al., 2011), but age estimates for these genes have not been obtained. Phylogenetic analyses have shown that diversification of SFBB genes pre-dates speciation of Pyrus and Malus (Minamikawa et al., 2010; Kakui et al., 2011). These data are compatible with the scenario where a large repertoire of non-self S-RNases are targeted and detoxified by multiple SFBB genes, each of which recognizes a subfraction of S-RNases (Kakui et al., 2011). In Petunia, a non-self recognition pollen rejection mechanism has also been proposed (Kubo et al., 2010; Wang and Kao, 2011). Although in the initial protein degradation model the S pollen genes were assumed to inhibit all S-RNases except that of the corresponding S haplotype (Hua and Kao, 2006; Hua et al., 2008), in vivo functional assays and protein-interaction assays revealed that each SLF protein functions as a pollen determinant and recognizes a subset of non-self S-RNases (Kubo et al., 2010; Wang and Kao, 2011).
As expected for the S pollen, evidence for positive selection acting on amino acid sites located in two (V1 and V2) of the four variable regions was found by calculating the ratio of non-synonymous substitutions per non-synonymous site (Ks) divided by the ratio of synonymous substitutions per synonymous site (Ks), using Malus (two SFBB-alpha and two SFBB-beta sequences) and Pyrus (two SFBB-alpha, two SFBB-beta, and two SFBB-gamma sequences) SFBB genes (Sassa et al., 2007). Nevertheless, positively selected amino acid sites have not been identified, mainly due to the small sample size. Therefore, it is not known how these amino acids vary among SFBB genes. Hence, it has been suggested that all SFBB genes could act together as the pollen determinant (Sassa et al., 2007). Under the non-self recognition by multiple factors, the high intra-haplotypic diversity of SFBB is the result of natural selection favouring diversification of SFBB genes within an S haplotype (Kakui et al., 2011). Under this model, no strong evidence for positive selection is expected when individual SFBB genes are considered, as observed for Pyrus SFBB-gamma (Vieira et al., 2009; De Franceschi et al., 2011a), SFBB-alpha, SFBB-beta, SFBB-delta, and SFBB-epsilon genes (De Franceschi et al., 2011a).

Another line of evidence in support of Japanese pear GSI non-self recognition by a multiple factors rejection mechanism comes from analyses of loss of function of two SFBB genes on Si phenotypes (Kakui et al., 2011). In the mutant haplotype Sδlh (a mutant derived from the S4 haplotype that shows st ylar self-compatibility because of deletion of S4-RNase; Okada et al., 2008), deletion of the SFBB1 gene specifically affects recognition of S1-RNase (Okada et al., 2008). Nevertheless, no effect of loss of function of SFBB1 gene from the S5 haplotype was observed on the Si phenotype in the pistils with S1-, S2-, S3-, S4-, and S9-RNases (Kakui et al., 2011, and references therein). Moreover in the S1 haplotype, the Si pollen is rejected and the SFBB1 gene from the S1 haplotype is not truncated. Therefore, the S1-RNase is not always targeted by the SFBB1 gene for degradation and other factors are involved in detoxification of the S1-RNase (Kakui et al., 2011).

As genes determining GSI specificity are under frequency-dependent selection, they are maintained for long periods of time (Wright, 1939). In Pyrinae, the oldest S-RNase gene specificity lineage is about 23 million years old (Vieira et al., 2010). Therefore, it is not surprising that most of the S-RNase allele lineages are found in Malus, Pyrus, Sorbus, and Crataegus species (Vieira et al., 2010). In this work, we showed that all SFBB genes described in Malus and Pyrus are also present in Sorbus aucuparia. Phylogenetic inferences of Pyrinae SFBB genes suggested the presence of 16 S. aucuparia SFBB genes. The age of the SFBB duplications also supported the involvement of these genes in GSI. Furthermore, amino acids under positive selection—those that could be involved in specificity determination—were identified when intra-haplotype SFBB genes were analysed.

### Material and methods

**Plant material and DNA extraction**

*S. aucuparia* is a self-incompatible species that has been characterized at the molecular level for the S-RNase gene (Raspé and Kohn, 2002, 2007). Due to the ecology of this bird-dispersed, insect-pollinated species, little population structure is found (Raspé et al., 2000; Raspé and Kohn, 2007). Leaves were collected from seven individuals from a natural population located in Bragança, Portugal (assigned as B). Furthermore, for the segregation experiments, 74 individuals from the progeny of the cross between individual Belgium5 (S17-RNaseS20-RNase) and Belgium6 (S2-RNaseS10-RNase) (Raspé and Kohn, 2007), assigned as D, were used. Genomic DNA was extracted from leaves of individual plants using the method of Ingram et al. (1997) or the Puregene® DNA Purification System (Gentra Systems, Minneapolis, MN, USA). No specific permits were required for the field collection, as the plant location was not privately owned or protected, and *S. aucuparia* is not an endangered or protected species.

**Amplification of S-RNases**

Three primer combinations (SorbusRNaseF and SorbusRNaseR, S-RNaseFT-F and S-RNase(UT)W-R, and Mac2/3R1+) were used for the amplification of S-RNases (Supplementary Table S1 at JXB online). Genomic DNA from individuals B2, B4, B5, B6, B8, B10, and B13 was used as template. Standard amplification conditions were 35 cycles of denaturation at 94 °C for 30 s, primer annealing according to Supplementary Table S1 for 30 s, and primer extension at 72 °C for 3 min. All amplification products were cloned using a TA Cloning kit (Invitrogen, Carlsbad, CA, USA). For each individual and amplification product, the insert of an average of 20 colonies was cut separately with RsAI, AluI, AvaII, and Sau3AI restriction enzymes. For each individual and restriction pattern, one colony was sequenced. For those sequences that showed similarity to the S-RNase gene when using BLASTn, two more colonies were sequenced in order to obtain a consensus sequence. An ABI PRISM BigDye Cycle Sequencing kit (Perkin Elmer, Foster City, CA, USA) 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).

**Amplification of SFBB genes**

Primers SFBBgenF and SFBBgenR (Supplementary Table S1) were designed on the basis of the 12 SLF and SFBB sequences of Cheng et al. (2006) and Sassa et al. (2007) (SFBBa, AB270795; SFBBa, AB270793; SFBBβ, AB270794; SFBBβ, AB270796; SLF1, DQ422810; SLF2, DQ422811; SFBB4, AB270797; SFBBβ, AB270800; SFBBβ, AB270801; SFBBβ, AB270798; SFBBγ, AB270799; SFBB5y, AB280702). These primers, although designed based on a small set of sequences, are, however, present in 65.5% of Malus and Pyrus SFBB sequences available in GenBank (n=165, see Fig. 1 and legend for accession numbers). Genomic DNA from individuals B5 and B6 was used as template. Standard amplification conditions were 35 cycles of denaturation at 94 °C for 30 s, primer annealing at 48 °C for 30 s, and primer extension at 72 °C for 2 min. The amplification products were cloned as described above. For each individual and amplification product, the insert of an average of 100 colonies was cut separately with RsAI, AluI, AvaII, and Sau3AI restriction enzymes. Sequencing reactions were performed as described above.

*S. aucuparia* SFBB1–SFBB3, SFBB5–SFBB9, SFBB11, and SFBB13–SFBB15 genes of the S22 and Sa haplotypes

Genomic DNA of the seven B individuals was used to amplify the SFBB1–SFBB3, SFBB5–SFBB9, SFBB11, and SFBB13–SFBB15 genes using specific primers. For all genes except SFBB6, based on the sequences obtained for B5 and B6 individuals, single-nucleotide polymorphisms were used to find restriction fragment length polymorphisms (RFLPs) that allowed the identification of the S22- and Sa- alleles of these 11 SFBB genes (Supplementary Table S2 at JXB
S. aucuparia SFBB1–SFBB3, SFBB5–SFBB9, SFBB11, and SFBB13–SFBB15 allele sequences in individuals from the progeny of the Belgium5 (S17-RNase/S20-RNase) and Belgium6 (S2-RNase/S10-RNase) cross

For 12 of the 16 SFBB genes studied, we were able to infer the allele that went with both the S22-RNase and the Sa-RNase (see Results). Nevertheless, we wanted to show that these 12 SFBB genes were located in the S-locus region. Therefore, the 74 individuals (assigned as D) from the progeny of the cross between individuals Belgium5 (S17-RNase/S20-RNase) and Belgium6 (S2-RNase/S10-RNase) were genotyped using specific primers (Supplementary Table S1) for the four segregating S-RNase alleles. For each of the 12 SFBB genes, the amplification product obtained using specific primers (Supplementary Table S1) and genomic DNA of individuals D28 (S2-RNase/S20-RNase), D34 (S10-RNase/S17-RNase), D69 (S10-RNase/S20-RNase), and D82 (S2-RNase/S17-RNase) was cloned. These individuals showed the four possible S-RNase combinations, and therefore all segregating SFBB alleles must be present in this sample. For each gene, ten randomly chosen colonies were sequenced to identify the SFBB alleles that were segregating in this cross. For SFBB15, only one allele was identified that was present in all four individuals analysed. Therefore, this gene was not studied further. For the SFBB3, SFBB5–SFBB9, and SFBB14 genes, all four alleles were identified (Supplementary Table S1) and genomic DNA of individuals D28 (S2-RNase/S20-RNase), D34 (S10-RNase/S17-RNase), D69 (S10-RNase/S20-RNase), and D82 (S2-RNase/S17-RNase) was cloned.

The DNA sequences were deposited in GenBank (accession numbers KC701674–KC701712). For these genes, in order to determine which individuals had a given SFBB allele, specific primers, as well as RFLPs, were developed (Supplementary Table S3). Because alleles of these genes had low levels of diversity, it was often not possible to develop a diagnostic marker for all observed alleles.

Phylogenetic analyses, summary statistics, and testing for positive selection of the S. aucuparia SFBB genes

The DNA sequences were deposited in GenBank (accession numbers KC701614–KC701663). Translated amino acid sequences were aligned using the accurate CLUSTALW algorithm as implemented in DAMBE (Xia and Xie, 2001). This amino acid alignment was used as a guide to obtain the corresponding nucleotide alignment. Analyses of DNA polymorphisms were performed using DnaSP (version 4.1) (Rozas et al., 2003). Using 216 SFBB sequences, minimum evolutionary trees were built with MEGA5 (Tamura et al., 2011), using CNI (level=1) and complete deletion. For the identification of sites under positive selection, we used ADOPS (Reboiro-Jato et al., 2012) and two datasets of 11 SFBB gene sequences that showed linkage with the S22-RNase and Sa-RNase. We compared the M2–M1 and M8–M7 models. We only considered as positively selected those amino acid sites that showed a probability >90% for both naive empirical Bayes (NEB) and Bayes empirical Bayes (BEB) methods, and that were identified in at least two of the three alignment methods used. Fourteen divergent F-box sequences from Fig. 1 in Vieira et al. (2009) (Medicago truncatula G161806856; Antirrhinum hicianum G138229882; Petunia integrifolia G1162134184; Populus trichocarpa G1158749689, G1167963539, G1159885773, G1116734897, G1159674948, G1116734897; Malus domestica G190103253; Prunus avium G133354144; Oryza sativa G1115487495; and Arabidopsis thaliana G130692063, G118404533) were also used. Moreover, we used 16 Malus domestica F-box sequences with known location that are not on chromosome 17 (the location of the S-locus region), obtained from BLASTp at the Genome Database for Rosaceae (http://www.rosaceae.org), using Md SFBB3-beta (BAF47180) as the query.

Results

S. aucuparia S-RNase genes

Because of the S-RNase intron size variation, the amplification products of the seven individuals analysed varied from 505 to 1819 bp (Table 1). Individuals B2, B5, and B6 had the same S-RNase allele, which was identical in the coding region to the S. aucuparia S22-RNase (EF494760). Individuals B2 and B13 had the S. aucuparia S21-RNase allele (EF494759), whilst individual B10 had the S26-RNase allele (EF494764). Individuals B4, B6, and B8 had one S-RNase allele, called Sa-RNase, which shared 99% amino acid identity with Malus domestica Sf-RNase (D50837) and Pyrus pyrifolia S12-RNase (AB426604). Individuals B5 and B13 presented the same S-RNase allele, called Sc-RNase, which shared 98% amino acid identity with Malus domestica S10-RNase (AF239809). The other allele of individual B8 was S20-RNase (AF504272).

For individuals B4 and B10, although three different primer combinations were used and all amplification products obtained were cloned and sequenced, only one S-RNase allele could be characterized.

SFBB genes in S. aucuparia

For B5 and B6 individuals, both having in common the S22-RNase allele, the 900 bp (expected size) amplification product obtained using primers SFBBgenF and SFBBgenR revealed 23 (called B5c1–B5c23) and 27 (B6c1–B6c27) sequences, respectively. Fig. 1 shows the phylogenetic relationship of the 50 S. aucuparia SFBB sequences and the 165 available Malus and Pyrus SFBB–SLF sequences. As S. aucuparia is a diploid species (Castroviejo and Real Jardín Botánico, 1986), the presence of more than two sequences from the same individual means that sequences from different genes are being amplified. Thus, the 27 sequences obtained for B6 individual implied at least 14 genes in S. aucuparia. Nevertheless, using the phylogenetic position of the S. aucuparia sequences and

Table 1. S-RNase haplotypes in the studied individuals NA, Not applicable.

<table>
<thead>
<tr>
<th>Individuals</th>
<th>S-RNases</th>
</tr>
</thead>
<tbody>
<tr>
<td>B2</td>
<td>S21-RNase (523 bp)</td>
</tr>
<tr>
<td>B4</td>
<td>Sa-RNase (684 bp)</td>
</tr>
<tr>
<td>B5</td>
<td>S22-RNase (505 bp)</td>
</tr>
<tr>
<td>B6</td>
<td>S22-RNase (505 bp)</td>
</tr>
<tr>
<td>B8</td>
<td>Sa-RNase (684 bp)</td>
</tr>
<tr>
<td>B10</td>
<td>S26-RNase (721 bp)</td>
</tr>
<tr>
<td>B13</td>
<td>Sa-RNase (510 bp)</td>
</tr>
</tbody>
</table>

Belgium5*  | S17-RNase | S20-RNase |
Belgium6*  | S2-RNase  | S10-RNase |

Multiple S pollen genes determine GSI in S. aucuparia

Although the levels of polymorphism were low (Table 2), for the SFBB1–SFBB3, SFBB5, SFBB7–SFBB9, SFBB11, SFBB13, and SFBB14 genes RFLPs were obtained that allowed identification of the sequences of the S22 and Sα haplotypes (see Materials and method, and Supplementary Table S2).

For these genes, the allele assigned as S22 was present in individuals B2, B5, and B6 (all having S22-RNase), but not in individuals B4, B8, B10, and B13 (Table 1). Furthermore, for SFBB1–SFBB3, SFBB5, SFBB7–SFBB9, SFBB11, SFBB13, and SFBB14, the allele assigned as Sα was only present in individuals B4, B6, and B8 that presented Sa-RNase (Table 1). For the SFBB6 gene, the sequences obtained with specific primers revealed that the S22-allele was only present in B2, B5, and B6 individuals, and the Sα-allele was only present in B4, B6, and B8 individuals only.

---

**Fig. 1.** Maximum parsimony tree showing the relationship of the Pyrinae SFBB genes. Md, Malus domestica; Pp, Pyrus pyriforma; Pb, Pyrus bretscheri; Pu, Pyrus ussuriensis; Ps, Pyrus sinkiangensis; Pc, Pyrus communis; Sa, Sorbus aucuparia (sequences are indicated by a filled circle). Bx-cx represents a Bragança S. aucuparia population (B) and the individual name (x), with cx representing the colony type. Numbers below the branches represent bootstrap values above 60. Genes of known location in Malus are: Md-X1, Md-X2, Md-X3, Md-X5, Md-X7, Md-X8, Md-X9, Md-X14, Md-X18, Md-X20, MsdSBBS3u, and MsdSBBS3β for the haplotype S3; and Md-X6, Md-X10, Md-X12, Md-X13, Md-X15, Md-X16, Md-X17, Md-X19, MsdSBBS9u, and MsdSBBS9β for the haplotype S9 (Sassa et al., 2007; Minamikawa et al., 2010). Genes of known location in Pyrus are: PpSBBS4-d1, PpSBBS4-d2, and PpSBBS4-u1–PpSBBS4-u4 for haplotype S4; and PpSBBS2-d1–PpSBBS2-d5, and PpSBBS2-u1–PpSBBS2-u5 for haplotype S5 (Okada et al., 2011). GenBank numbers are: Pyrus 1 sequences: AB621615, 293337907, EU979310, EU979317, AB621617, EU979315, EU979311, EU979313, 293337889, 293337854, AB270800, AB621616, 293337905, 293337893, 293337887, EU979316, 293337891, AB270797, 293337911, 293337895, EU979309, 293337909; Pyrus 2 sequences: 293337814, 293337810, AB545981, 293337808, 293337806, 293337804; Pyrus 3 sequences: 293337784, 293337782, 293337780, 293337786, 293337778, EU422961, EU422960, 293337788, 293337776, EU081892, EU422958, AB270799, 293337802, EU081890, AB297937, EU422956, AB297935, 293337798, EU418249, EU979327, AB297939, AB297934, 293337938, EU081887, AB297933, 293337796, EU418248, AB297940, AB270802, EU081894, 293337800, EU081891, AB297936, EU081893, EU422959, EU422962; Pyrus 4 sequences: EU979319, EU979320, 293337923, 293337878, 293337844, EU979324, Pyrus 5 sequences: AB621602, 293337876, 293337858, 293337846, 293337842; Malus 1 sequences: 293337901, 293337897, AB539848, 293337903, 293337899, AB539868, FJ008668.
Table 2. Synonymous (above the diagonal) and non-synonymous (below the diagonal) divergence levels in S. aucuparia SFBB genes. Synonymous and non-synonymous diversity is presented in the diagonal. The estimated age, in million years, for the SFBB genes split is given in square brackets. –, Only one sequence is available.

<table>
<thead>
<tr>
<th>SFBB1</th>
<th>SFBB2</th>
<th>SFBB3</th>
<th>SFBB4</th>
<th>SFBB5</th>
<th>SFBB6</th>
<th>SFBB7</th>
<th>SFBB8</th>
<th>SFBB9</th>
<th>SFBB10</th>
<th>SFBB11</th>
<th>SFBB12</th>
<th>SFBB13</th>
<th>SFBB14</th>
<th>SFBB15</th>
<th>SFBB16</th>
</tr>
</thead>
<tbody>
<tr>
<td>SFBB1</td>
<td>0.0036</td>
<td>0.2322</td>
<td>0.2713</td>
<td>0.2495</td>
<td>0.2351</td>
<td>0.2321</td>
<td>0.2480</td>
<td>0.3218</td>
<td>0.2592</td>
<td>0.3199</td>
<td>0.2274</td>
<td>0.2100</td>
<td>0.1231</td>
<td>0.2207</td>
<td>0.2345</td>
</tr>
<tr>
<td>SFBB2</td>
<td>0.0071</td>
<td>[15.61]</td>
<td>[18.24]</td>
<td>[16.77]</td>
<td>[15.80]</td>
<td>[16.67]</td>
<td>[21.63]</td>
<td>[17.42]</td>
<td>[21.50]</td>
<td>[15.28]</td>
<td>[14.12]</td>
<td>[8.27]</td>
<td>[14.83]</td>
<td>[15.76]</td>
<td>[21.11]</td>
</tr>
<tr>
<td>SFBB3</td>
<td>0.2090</td>
<td>0.0452</td>
<td>0.2555</td>
<td>0.2424</td>
<td>0.2427</td>
<td>0.2175</td>
<td>0.2299</td>
<td>0.3167</td>
<td>0.2462</td>
<td>0.3565</td>
<td>0.2485</td>
<td>0.2328</td>
<td>0.2129</td>
<td>0.2142</td>
<td>0.2369</td>
</tr>
<tr>
<td>SFBB4</td>
<td>0.0146</td>
<td>[17.04]</td>
<td>[16.29]</td>
<td>[16.31]</td>
<td>[14.62]</td>
<td>[15.45]</td>
<td>[21.29]</td>
<td>[16.55]</td>
<td>[23.96]</td>
<td>[16.70]</td>
<td>[15.65]</td>
<td>[14.31]</td>
<td>[14.40]</td>
<td>[15.92]</td>
<td>[21.03]</td>
</tr>
<tr>
<td>SFBB5</td>
<td>0.0166</td>
<td>[14.34]</td>
<td>[14.19]</td>
<td>[13.79]</td>
<td>[12.89]</td>
<td>[18.50]</td>
<td>[15.14]</td>
<td>[16.98]</td>
<td>[14.87]</td>
<td>[12.93]</td>
<td>[16.29]</td>
<td>[14.08]</td>
<td>[10.96]</td>
<td>[18.57]</td>
<td></td>
</tr>
<tr>
<td>SFBB6</td>
<td>0.2027</td>
<td>0.1916</td>
<td>0.0425</td>
<td>0.2134</td>
<td>0.2111</td>
<td>0.2052</td>
<td>0.1917</td>
<td>0.2752</td>
<td>0.2252</td>
<td>0.2566</td>
<td>0.2213</td>
<td>0.1923</td>
<td>0.2423</td>
<td>0.2095</td>
<td>0.1631</td>
</tr>
<tr>
<td>SFBB7</td>
<td>0.1928</td>
<td>0.1720</td>
<td>0.1472</td>
<td>0.0326</td>
<td>0.2355</td>
<td>0.2192</td>
<td>0.1885</td>
<td>0.2702</td>
<td>0.2207</td>
<td>0.2711</td>
<td>0.2183</td>
<td>0.1865</td>
<td>0.2345</td>
<td>0.2185</td>
<td>0.1801</td>
</tr>
<tr>
<td>SFBB8</td>
<td>0.1884</td>
<td>0.1894</td>
<td>0.1765</td>
<td>0.1659</td>
<td>0.0598</td>
<td>0.1517</td>
<td>0.1382</td>
<td>0.2651</td>
<td>0.2690</td>
<td>0.2770</td>
<td>0.2207</td>
<td>0.1773</td>
<td>0.2214</td>
<td>0.2168</td>
<td>0.1780</td>
</tr>
<tr>
<td>SFBB9</td>
<td>0.2119</td>
<td>0.1959</td>
<td>0.1800</td>
<td>0.1911</td>
<td>0.0705</td>
<td>0.0124</td>
<td>0.1437</td>
<td>0.2579</td>
<td>0.2689</td>
<td>0.3033</td>
<td>0.2189</td>
<td>0.1801</td>
<td>0.2141</td>
<td>0.1407</td>
<td>0.1780</td>
</tr>
<tr>
<td>SFBB10</td>
<td>0.1776</td>
<td>0.1782</td>
<td>0.1619</td>
<td>0.1689</td>
<td>0.0671</td>
<td>0.0843</td>
<td>0.0641</td>
<td>0.2501</td>
<td>0.2276</td>
<td>0.2380</td>
<td>0.1994</td>
<td>0.1533</td>
<td>0.2275</td>
<td>0.2078</td>
<td>0.1631</td>
</tr>
<tr>
<td>SFBB11</td>
<td>0.2611</td>
<td>0.2570</td>
<td>0.2039</td>
<td>0.2002</td>
<td>0.1778</td>
<td>0.2092</td>
<td>0.1796</td>
<td>0.0305</td>
<td>0.3208</td>
<td>0.2718</td>
<td>0.2331</td>
<td>0.3016</td>
<td>0.2594</td>
<td>0.2263</td>
<td>0.2459</td>
</tr>
<tr>
<td>SFBB12</td>
<td>0.1709</td>
<td>0.1688</td>
<td>0.1492</td>
<td>0.1215</td>
<td>0.1755</td>
<td>0.1916</td>
<td>0.1664</td>
<td>0.0548</td>
<td>0.2781</td>
<td>0.2376</td>
<td>0.2147</td>
<td>0.2256</td>
<td>0.2137</td>
<td>0.1827</td>
<td>0.3174</td>
</tr>
<tr>
<td>SFBB13</td>
<td>0.1848</td>
<td>0.1709</td>
<td>0.1548</td>
<td>0.1455</td>
<td>0.1768</td>
<td>0.1992</td>
<td>0.1721</td>
<td>0.2182</td>
<td>0.1473</td>
<td>0.1473</td>
<td>0.1473</td>
<td>0.1473</td>
<td>0.1473</td>
<td>0.1473</td>
<td>0.1473</td>
</tr>
<tr>
<td>SFBB14</td>
<td>0.1822</td>
<td>0.1811</td>
<td>0.1677</td>
<td>0.1364</td>
<td>0.1711</td>
<td>0.1918</td>
<td>0.1729</td>
<td>0.2250</td>
<td>0.1378</td>
<td>0.1161</td>
<td>0.0646</td>
<td>0.1658</td>
<td>0.1970</td>
<td>0.2308</td>
<td>0.1699</td>
</tr>
<tr>
<td>SFBB15</td>
<td>0.1677</td>
<td>0.1712</td>
<td>0.1522</td>
<td>0.1290</td>
<td>0.1572</td>
<td>0.1674</td>
<td>0.1540</td>
<td>0.1970</td>
<td>0.1349</td>
<td>0.1155</td>
<td>0.0844</td>
<td>0.0844</td>
<td>0.0844</td>
<td>0.0844</td>
<td>0.0844</td>
</tr>
<tr>
<td>SFBB16</td>
<td>0.1080</td>
<td>0.1808</td>
<td>0.1715</td>
<td>0.1503</td>
<td>0.1666</td>
<td>0.1871</td>
<td>0.1569</td>
<td>0.2285</td>
<td>0.1344</td>
<td>0.1530</td>
<td>0.1561</td>
<td>0.1375</td>
<td>0.0341</td>
<td>0.2129</td>
<td>0.2036</td>
</tr>
</tbody>
</table>

\* Diversity levels calculated from a larger data set indicated in brackets.
Multiple S pollen genes determine GSI in *S. aucuparia*

As described in Material and methods, these genes were sequenced using specific primers (Supplementary Table S1) in individuals that had S2-RNase, S10-RNase, S17-RNase, and S20-RNase. These sequences were used to develop specific RFLP markers (Supplementary Table S3) or specific primers for a particular SFBB allele. These markers were used in the segregation analyses of these genes in 74 individuals from the progeny of the cross between individuals Belgium5 (S17-RNase/S20-RNase) and Belgium6 (S2-RNase/S10-RNase). For all 11 genes, we found linkage of a particular SFBB allele with a specific S-RNase allele (Table 3). Therefore, in *Sorbus*, these genes are located in the S-locus region.

Positively selected amino acid sites in 11 *S. aucuparia* SFBB genes from the S22 and Sa haplotypes

Adaptive evolution is likely to act on a small subset of amino acids, and thus average substitution rates across the gene may not indicate positive selection (Yang and Bielawski, 2000). Under the non-self recognition by multiple factors system, allelic products of the pollen S gene would be highly homologous with each other and would be expected to target similar fractions of S-RNases because allelic divergence of each type of pollen S is not required and would not be favoured by natural selection (Kakui et al., 2011). Fixation of duplicated genes is an adaptive event, and these duplicated genes can act as a source of protein subfunctionalization by evolution of key positions in the protein (Hurles, 2004). In the case of SFBB genes, protein subfunctionalization would imply a change in the amino acids that determine the specificity recognition of S-RNases. SFBBs would thus be a positively evolving gene family. In fact, the high intra-haplotype $K_{a}$ values of SFBB (for the S22 haplotype, $K_{a}$ is 0.172, and for the Sa haplotype, $K_{a}$ is 0.161) suggest that natural selection has favoured diversification of SFBB paralogues to target allelic variants of S-RNases in Pyrinae (Kakui et al., 2011). Adaptive evolution of duplicated paralogous gene families using CodeML (Yang, 1997) has been identified at the genome level to identify genes subject to positive selection (Emes and Yang, 2008). The amino acid sites identified as being positively selected with a posterior probability higher than 90% for the 11 SFBB genes of the S22 and Sa haplotypes, using the Yang (1997) method as implemented in ADOPS (Reboiro-Jato et al., 2012), are shown in Fig. 2. There were 12 amino acid sites identified as being positively selected common to the S22 and Sa haplotypes. Amino acid sites at positions 179 and 197 of the Sa haplotype were, however, assigned as positively selected with probabilities higher than 91 and 83% in BEB, and 75 and 71%

### Table 3. Segregation analyses of *S. aucuparia* SFBB1–SFBB3, SFBB5–SFBB9, SFBB11, SFBB13, and SFBB14.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Allele</th>
<th>S2-RNase (N=39)</th>
<th>S10-RNase (N=35)</th>
<th>S17-RNase (N=34)</th>
<th>S20-RNase (N=40)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SFBB1</td>
<td>S2-SFBB1</td>
<td>39</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SFBB2</td>
<td>S10-SFBB2</td>
<td>0</td>
<td>35</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SFBB3</td>
<td>S17-SFBB3</td>
<td>0</td>
<td>0</td>
<td>34</td>
<td>40</td>
</tr>
<tr>
<td>SFBB5</td>
<td>S2-SFBB5</td>
<td>39</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SFBB6</td>
<td>S20-SFBB5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>40</td>
</tr>
<tr>
<td>SFBB7</td>
<td>S17-SFBB7</td>
<td>0</td>
<td>0</td>
<td>34</td>
<td>0</td>
</tr>
<tr>
<td>SFBB8</td>
<td>S10-SFBB8</td>
<td>0</td>
<td>35</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SFBB9</td>
<td>S17-SFBB9</td>
<td>0</td>
<td>0</td>
<td>34</td>
<td>0</td>
</tr>
<tr>
<td>SFBB11</td>
<td>S20-SFBB9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>40</td>
</tr>
<tr>
<td>SFBB13</td>
<td>S2-SFBB13</td>
<td>39</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SFBB14</td>
<td>S10-SFBB13</td>
<td>0</td>
<td>35</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SFBB15</td>
<td>S2-SFBB14</td>
<td>39</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SFBB16</td>
<td>S10-SFBB14</td>
<td>0</td>
<td>35</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SFBB17</td>
<td>S17-SFBB17</td>
<td>0</td>
<td>0</td>
<td>34</td>
<td>0</td>
</tr>
<tr>
<td>SFBB18</td>
<td>S20-SFBB20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>40</td>
</tr>
</tbody>
</table>
in NEB, respectively. It should be noted that, within each haplotype, the combination of these amino acid sites was different for every Sorbus SFBB gene analysed, and the minimum number of differences between any two SFBB genes from the same haplotype at these amino acid sites was six (between SFBB9 and SFBB14 at the S22 haplotype; Fig. 2).

As expected, in the F-box region, there were no amino acid sites showing evidence for positive selection.

Fig. 2. Schematic representation of the amino acid sites identified as positively selected (shaded) when using at least two different alignment algorithms and the method of Yang (1997) implemented in ADOPS (Reboiro-Jato et al., 2012) with a probability higher than 90 and 95% (bold) in both NEB (naive empirical Bayes) and BEB (Bayes empirical Bayes) and the 11 S. aucuparia SFBB genes of the S22 and Sa haplotype.
When comparing the amino acid sites identified as positively selected between the two S haplotypes, for the SFBB1, SFBB2, SFBB5, SFBB6, SFBB8, and SFBB13 genes, there are no differences (Supplementary Table S4 at JXB online). For the SFBB9 gene, at these amino acid sites there was one amino acid difference between the two S haplotypes. Two amino acid differences at these amino acid sites were observed in the SFBB11 gene between the two S haplotypes. Four amino acid differences at the amino acid sites identified as positively selected were observed at the SFBB3 and SFBB4 genes between the S22 and Sa haplotypes.

Discussion

In S. aucuparia, there are at least 16 SFBB genes. For these genes, Malus and Pyrus orthologues have been described, for which expression has been shown to be pollen restricted (Sassa et al., 2007; Minamikawa et al., 2010; De Franceschi et al., 2011a,b). Indeed, all Pyrinae SFBB-like sequences described so far are expressed in pollen only. Although the 11 SFBB genes here studied were in linkage with the S-RNase, not all Malus and Pyrus SFBB genes are in linkage with the S-RNase gene (De Franceschi et al., 2011a; Kakui et al., 2011), and thus these genes are probably not involved in determining S pollen specificity (see review by Minamikawa et al., 2010; De Franceschi et al., 2011a; Kakui et al., 2011; Okada et al., 2011).

The characterization of a large number of genes in one species, as performed here, will help to establish gene orthologies in different species using a phylogenetic approach. It should be noted that synteny alone cannot be used to establish orthologies, as gene order is not conserved among S haplotypes (Minamikawa et al., 2010; Okada et al., 2011). Because polymorphism levels at SFBB genes were always below 10% (Table 2, and Table 1 in Kakui et al., 2011), this can also be used as a guide for the presence of multiple genes within species. Indeed, the exception of the less than 10% diversity reported for the Pyrus SFBB1 gene (which includes PpSFBB4-d1 and PpSFBB2-d3 sequences; Kakui et al., 2011) is due to the inclusion of two different genes (Fig. 1). The PpSFBB4-d1 sequence is deleted in the mutant haplotype S<sub>dom</sub> (Okada et al., 2008) and specifically affects recognition of the S1-RNase. The non-functional gene (named SFBB1-like) of the S5 haplotype, which shows no effect in the S5 pollen phenotype when crossed with plants having S1-RNase pistils (Kakui et al., 2011), may thus represent a different gene from PpSFBB4-d1.

As expected for genes determining GSI specificity, the SFBB genes predate the appearance of the Malus, Pyrus, and Sorbus genera. Species from these genera may, however, have diverged in the last 5 million years (under the assumption of a molecular clock for five genes; Table 4). It should be noted that this interpretation is far from being consensual. Indeed, when using information from both the fossil record and molecular data, Campbell et al. (2007) suggested that the genera Malus, Pyrus, and Sorbus, among others, are the result of an ancient, rapid radiation associated with a low mutation rate (as discussed by Vieira et al., 2010). The two most closely related SFBB genes are 8.27 million years old (Table 2). On the other hand, the two most divergent SFBB genes are 23.96 million years old (Table 2). This age is, as expected, similar to the age of the oldest Pyrinae S-RNase specificity lineages (about 23 million years old; Vieira et al., 2010). This implies that the genes identified here do not have a Pyrus orthologue (under the assumption of 32 million years for the age of the split between the Amygdaloideae and Pyrinae lineages; Wikstrom et al., 2001).

At the S-RNase gene, 18 S. aucuparia alleles have been characterized (Raspé and Kohn, 2007, and this work), but 40 different specificities have been inferred (Raspé and Kohn, 2007). Furthermore, using a phylogenetic approach, 35 S-RNase specificities have been estimated in Pyrinae (Vieira et al., 2010). Although we cannot be sure that all Sorbus SFBB genes have been characterized, since the approach used here depended on primer specificity, the 16 Sorbus SFBB genes clustered with high support with all SFBB genes described in Malus, and 13 out of 16 with Pyrus sequences obtained from the characterization of BAC libraries using different primers. The three exceptions (PpSFBB2-d2, PpSFBB2-d3, and PpSFBB4-d1; Fig. 1) may represent Sorbus SFBB genes that have not been characterized, or, genes that recognize S-RNase specificities that have been lost in S. aucuparia. Furthermore, we are assuming that the 16 genes may be involved in pollen GSI specificity, although we have evidence of linkage with the S-RNase for 11 genes only. Nevertheless, the number of S-RNase specificities seems to be higher than the number of SFBB genes. This is expected, as, under the non-self recognition by multiple factors rejection model, each S-RNase can be targeted by multiple SFBB genes (Kubo et al., 2010; Wang and Kao, 2011). Moreover, natural selection favours diversification of SFBB genes within an S haplotype (Kakui et al., 2011). When the 11 Sorbus SFBB genes of the same S haplotype were analysed, amino acids that were associated with recognition of the S-RNase specificities—those under positive selection—were observed using codon models. Because selection acts at the gene level, it was not surprising that the same amino acid sites under positive selection were identified in the two S haplotypes (the 11 SFBBs were the same). The amino acid sites under positive selection could, however, represent different SFBB protein functions. Although there is no functional data for most plant F-box genes, phylogenetic divergent F-box genes, in principle, are associated with different functions. Therefore, to test whether the identified positively selected amino acid sites reflected different functions, we selected 14 F-box sequences from the seven divergent groups presented in Fig. 1 in Vieira et al. (2009), most of which were not expected to be involved in GSI. No amino acid sites under positive selection were identified in the region analysed here. An identical result was obtained when only Malus domestica F-box genes, not located in the S-locus region (chromosome 17), were analysed. Thus, the amino acid sites under positive selection identified in the Sorbus SFBB sequences may represent amino acids involved in specificity determination.
The variability at the amino acid sites under positive selection within SFBB genes may account for the differences observed in SI behaviour when different alleles of the same SFBB gene are analysed. The presence of identical amino acids at the positively selected sites between different S haplotypes, in the same SFBB gene, implies that alleles of these genes are recognizing the same S-RNase specificities and marking them for degradation. This was the case for six of the 11 Sorbus SFBB genes analysed. The S22 and Sa haplotypes were from the same population, and thus they must recognize as non-self the same spectrum of S-RNases, except for Sa- and S22-RNase. In a model of one SFBB gene–one S-RNase specificity, we would expect to find, between the two S haplotypes, differences in one SFBB gene only at the amino acid sites under positive selection. There were, however, five SFBB genes that showed differences at the amino acids under positive selection. This again suggests that one SFBB gene can recognize more than one S-RNase specificity. In Petunia, co-immunoprecipitation experiments and transgenic approaches have shown that the SLF1 gene recognized S17-RNase as non-self in four S haplotypes analysed, but only recognized S9-RNase as non-self in two S haplotypes (Kubo et al., 2010). Thus, one SFBB gene of an S haplotype can interact with two or more S-RNases (Kubo et al., 2010; Wang and Kao, 2011). The same was observed when the Pyrus SFBB1 gene was deleted (see Introduction), although another interpretation for this result may be possible (see above). At present, predictions about recognition of an S-RNase specificity by a particular SFBB gene are very difficult to test as neither transformation nor antisense RNA technology used to induce post-transcriptional gene silencing (Lopez-Gomollon and Dalmay, 2010) are available for these species or for this system. Furthermore, the Pyrneae species studied are shrubs or trees that have a minimum 2–3 years juvenile period following planting (Shulaev et al., 2008). Therefore, at present, only characterization of truncated SFBB genes can be used to confirm these predictions. Nevertheless, as the plants having these truncated SFBB gene sequences show no specific phenotype (the SI phenotype will be different only in the presence of the S-RNase that is recognized by this gene), only by performing detailed studies in several individuals will these mutations be identified. At present, there are only two such mutations in Pyrus (De Franceschi et al., 2011a; Kakui et al., 2011). The characterization of these genes in a larger number of individuals is thus needed.

### Supplementary data

Supplementary data are available at JXB online.

**Supplementary Table S1.** Primers used in this work.

**Supplementary Table S2.** RFLPs used to identify the SFBB1–SFBB3, SFBB5, SFBB7–SFBB9, SFBB11, SFBB13, and SFBB14 genes of the S22 and Sa haplotypes.

**Supplementary Table S3.** Patterns used for the identification of alleles from SFBB1–SFBB3, SFBB5–SFBB9, SFBB11, SFBB13, and SFBB14 genes in the segregation experiment.

**Supplementary Table S4.** Polymorphisms at the amino acid sites under positive selection between the S22 and Sa haplotypes.

### Acknowledgements

This work has been funded by the project PTDC/BIA-BEC/100616/2008 comp-01-0124-FEDER-008916, funded by Programa Operacional para a Ciência e Inovação (POCI) 2012, co-funded by Fundo Europeu para o Desenvolvimento Regional (FEDER) funds and Programa Operacional para a Promoção da competitividade (COMPETE; FCOMP-01-0124-FEDER-022718; PEst-CS/SAU/1A0002/2011). B. A. is the recipient of a PhD grant (SFRH/BD/69207/2010) from FCT.

### References


