Document donwnloaded from:

[http://redivia.gva.es/handle/20.500.11939/4780]

This paper must be cited as:

[Zuriaga, E., Molina, L., Badenes, M.L., Romero, C. (2012). Physical mapping of a pollen modifier locus controlling self-incompatibility in apricot and synteny analysis within the Rosaceae. Plant Molecular Biology, 79(3), 229-242.]

ivia
Institut Valencià
d'Investigacions Agràries

The final publication is available at

[http://dx.doi.org/10.1007/s11103-012-9908-z]

Copyright [Springer]

Physical mapping of a pollen modifier locus controlling self-incompatibility in apricot and synteny analysis within the Rosaceae

Elena Zuriaga, Laura Molina, María Luisa Badenes and Carlos Romero*

Instituto Valenciano de Investigaciones Agrarias (IVIA), Apartado Oficial 46113 Moncada, Valencia, Spain.

Elena Zuriaga	garcia_zur@gva.es
Laura Molina	lauravillas@hotmail.com
María Luisa Badenes	mlu_badenes@gva.es
Carlos Romero* (corresponding author)	romero_carsal@gva.es

Phone number: +34 96 342 40 00; Fax number: +34 96 342 40 01

.

33 **Abstract**

34 *S*-locus products (S-RNase and F-box proteins) are essential for the gametophytic
35 self-incompatibility (GSI) specific recognition in *Prunus***.** However, accumulated
36 genetic evidence suggests that other *S*-locus unlinked factors are also required for
37 GSI. For instance, GSI breakdown was associated with a pollen-part mutation
38 unlinked to the S-locus in the apricot (*Prunus armeniaca* L.) cv. 'Canino'. Fine-
39 mapping of this mutated modifier gene (*M*-locus) and the synteny analysis of the
40 *M*-locus within the Rosaceae are here reported. A segregation distortion loci
41 (SDL) mapping strategy, based on a selectively genotyped population, was used
42 to map the *M*-locus. In addition, a bacterial artificial chromosome (BAC) contig
43 was constructed for this region using overlapping oligonucleotides probes, and
44 BAC-end sequences (BES) were blasted against Rosaceae genomes to perform
45 micro-synteny analysis. The *M*-locus was mapped to the distal part of chr.3
46 flanked by two SSR markers within an interval of 1.8 cM corresponding to ~364
47 Kb in the peach (*Prunus persica* L. Batsch) genome. In the integrated genetic-
48 physical map of this region, BES were mapped against the peach scaffold_3 and
49 BACs were anchored to the apricot map. Micro-syntenic blocks were detected in
50 apple (*Malus* × *domestica* Borkh.) LG17/9 and strawberry (*Fragaria vesca* L.)
51 FG6 chromosomes. The *M*-locus fine-scale mapping provides a solid basis for
52 self-compatibility marker-assisted selection and for positional cloning of the
53 underlying gene, a necessary goal to elucidate the pollen rejection mechanism in
54 *Prunus*. In a wider context, the syntenic regions identified in peach, apple and
55 strawberry might be useful to interpret GSI evolution in Rosaceae.

56

57 **Key words**

58 Apricot, mapping, modifier gene, Rosaceae synteny, self-incompatibility

59

60

61

62

63

64

65

66

67 **Introduction**

68

69 Gametophytic self-incompatibility (GSI) is a common reproductive barrier in
70 flowering plants, mostly controlled by the so-called *S*-locus, that prevents self-
71 fertilization (De Nettancourt 2001). The *S*-locus contains at least two linked genes
72 coding for S-RNase and *S*-locus F-box proteins in Plantaginaceae, Solanaceae and
73 Rosaceae. S-RNases are specifically expressed in the style, being essential to
74 reject self-pollen by their cytotoxic activity (McClure et al. 1989; Boskovic and
75 Tobutt 1996; Xue et al. 1996). The *S*-locus F-box proteins, generally termed SLF
76 in Plantaginaceae and Solanaceae (Lai et al. 2002; Sijacic et al. 2004) and SFB in
77 Rosaceae (Ushijima et al. 2003), are expressed in pollen and predicted to be part
78 of a SCF E3 ubiquitin ligase complex, targeting non-self S-RNases for
79 degradation by the 26S proteasome proteolytic pathway in *Petunia* and
80 *Antirrhinum* (Hua and Kao 2006; Huang et al. 2006). An alternative mechanism in
81 *Nicotiana* was suggested by Goldraij et al. (2006) where S-RNases are prevented
82 from exerting their cytotoxic function by sequestration in vacuolar compartments.
83 The role of *S*-locus products is essential for the GSI specific recognition
84 mechanism, but genetic and molecular evidence also shows that non *S*-locus
85 factors are necessary for GSI in Solanaceae and Plantaginaceae. McClure et al.
86 (2000) classified these modifiers into three groups, those affecting the expression
87 of *S*-specific genes, those required for pollen rejection without a wider role in
88 pollination, and factors involved in pollen rejection and other pistil-pollen
89 interactions. Modifier factors have been identified on both pollen and pistil sides.
90 Among the pistil factors, a small asparagine-rich protein, termed HT-B, was
91 shown to be required for *S*-allele-specific pollen rejection in *Nicotiana* (McClure
92 et al. 1999). HT-B is suggested to be stabilized in turn by a Kunitz-type proteinase
93 inhibitor termed NaStEP (Busot et al. 2008; F. Cruz-García pers. comm.). In
94 *Nicotiana* other pistil proteins are required for *S*-specific pollen rejection, such as
95 the 120kDa glycoprotein (Hancock et al. 2005), or have been found to interact
96 with S-RNases, such as the arabinogalactan proteins NaTTS and NaPELPIII
97 (Cruz-García et al. 2005) and the thiorredoxin h (Juárez-Díaz et al. 2006), but
98 their role in SI is less clearly understood. Modifiers have also been identified on
99 the pollen side. For instance, PhSSK1 is a pollen-expressed Skp1-like protein
100 required for cross-pollen compatibility in *Petunia* acting as an adaptor in an SCF

101 complex (Zhao et al. 2010). Other pollen proteins, such as the ubiquitously-
102 expressed SBP1 (Sims and Ordanic 2001) and the pollen endomembrane-
103 associated protein NaPCCP (Lee et al. 2009) interact with pistil proteins but their
104 function remains largely unknown.
105 In Rosaceae, most of the numerous SC accessions found are related to
106 mutations in pistil and pollen *S*-locus factors (Yamane and Tao 2009). However,
107 mutations in non *S*-locus factors have also been associated with SC in sweet
108 cherry (*Prunus avium* L.) (Wünsch and Hormaza 2004), almond (*Prunus*
109 *amygdalus* Batsch) (Fernández et al. 2009) and diploid strawberries (*Fragaria*
110 *spp*.) (Boskovic et al. 2010). In apricot (*Prunus armeniaca*) the cultivar 'Canino'
111 (S_2S_C *Mm*) was found to contain both type of mutations conferring SC. On one
112 side, the S_C-haplotype shows an insertional mutation in the SFB_C gene that
113 produces a truncated protein leading to the loss of pollen-*S* function. On the other,
114 a recessive mutation of the modifier (*m*) gene unlinked to the *S*-locus was shown
115 to independently cause loss of pollen-*S* activity in this cultivar (Vilanova et al.
116 2006). Genetic evidence suggests that, similar to Solanaceae and Plantaginaceae,
117 these factors are essential for the GSI system in Rosaceae but their nature remains
118 unknown. In *Brassica rapa*, a positional cloning strategy was successfully used by
119 Murase et al (2004) to identify a sporophytic SI modifier gene from a self-
120 compatible cultivar, revealing that it encodes a membrane-anchored protein
121 kinase. In this work, we provide solid basis for future identification of the
122 'Canino' pollen-part modifier gene following a similar strategy by fine-mapping
123 the *M*-locus to the distal part of apricot chr. 3. and constructing a BAC contig
124 encompassing this locus. In addition, micro-synteny of this region between apricot
125 and other Rosaceae including peach (*Prunus persica* L. Batsch) (International
126 Peach Genome Initiative – IPGI; http://www.rosaceae.org/peach/genome), apple
127 (*Malus* × *domestica* Borkh.) (Velasco et al. 2010), and strawberry (*Fragaria vesca*
128 L.) (Shulaev et al. 2010) was studied.
129
130 **Materials and Methods**
131
132 *Plant Material*
133 An F_1 population obtained by crossing apricot cvs. 'Goldrich' × 'Canino' ('G×C-
134 01') (*N*=171) segregating for a pollen-part mutation (PPM) conferring SC was

135 used for mapping. Two sets of seedlings derived from crosses performed with the
136 same parents in 2007 'G×C-07' (N=58) and 2008 'G×C-08' (N=94), as well as a
137 set derived from 'Canino' self-pollination in 2005 'C×C-05' (N=80), were also
138 used in this study. All these trees are maintained at the collection of the Instituto
139 Valenciano de Investigaciones Agrarias (IVIA) in Valencia (Spain). Additionally,
140 143 independent F_2 seed populations (ranging from N=8 to N=192) were obtained
141 after self-pollination of 'G×C-01' trees.
142
143 *Generation of F_2 populations for M-locus genotyping*
144 All trees of the 'G×C-01' population were self-pollinated in the field for three
145 consecutive springs (2006, 2007 and 2008) to obtain new or to complete already
146 existing F_2 seed populations. Before anthesis, insect-proof bags were put over
147 several branches, containing 200-250 flower buds per seedling, approximately, to
148 prevent cross-pollination. Fruits were collected about 3 months later and embryos
149 were dissected from the rest of the seed tissue and stored at –20ºC. The minimal
150 population size genotyped at each F_2 population to obtain at least one individual
151 homozygous for the *S*-locus (carrying the PPM) with risk α was calculated using
152 the equation $N = \ln(\alpha) / \ln(1-P)$ (Hospital 2003), where P denotes the probability
153 that an individual has the desired *S*-genotype (S_1S_1 or S_2S_2). *S*-genotyping of the F_2
154 offspring was performed by PCR-based amplification of the *S-RNase* first intron
155 with the primer pair SRc-F (5'-CTC GCT TTC CTT GTT CTT GC-3') and SRc-
156 R (5'-GGC CAT TGT TGC ACA AAT TG-3') following the protocol described
157 by Vilanova et al. (2005).
158
159 *DNA isolation*
160 DNA was extracted from 50 mg of young leaves following the method described
161 by Doyle and Doyle (1987). DNA quantification was performed measuring with
162 the NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific,
163 Wilmington, DE) and by comparison with lambda DNA (Promega, Madison, WI,
164 USA). Embryo DNA was extracted by incubating for 10 min at 95ºC with 20 µl of
165 TPS (100 mM Tris-HCl, pH 9.5; 1 M KCl; 10 mM EDTA) isolation buffer
166 (Thomsom and Henry 1995).
167
168 *SSR marker analysis*

169 A total of 180 SSR markers, spread over the 8 *Prunus* chromosomes, were tested
170 to perform a genome-wide screen for the PPM (Supplemental Table 1). One
171 hundred thirty-two of them were directly available from the GDR website (Jung et
172 al. 2004) and 48 were identified from the peach genome sequence (peach v1.0
173 International Peach Genome Initiative 2010
174 http://www.rosaceae.org/peach/genome) (Supplemental Table 2) using the
175 Tandem Repeats Finder software (Benson 1999). Primer pairs flanking
176 microsatellite repeat motifs were designed using Primer3 (Rozen and Skaletsky
177 2000). Similarly, 123 additional SSRs were tested to construct a high-density map
178 of the 'Canino' LG3. One hundred two of them were identified from the peach
179 genome scaffold_3 (Supplemental Table 3) and 21 (belonging to the series UDAp
180 (4), EPPCU/EPDCU (8), UDA (2), MA/MD (5), UCDCH (1) and AMPA (1))
181 were available from the GDR website (Jung et al. 2004).

182 SSR amplifications were performed in a GeneAmp® PCR System 9700
183 thermal cycler (Perkin–Elmer, Freemont, CA, USA) in a final volume of 20 µl,
184 containing 75 mM Tris–HCl, pH 8.8; 20 mM $(NH_4)_2SO_4$; 1.5 mM $MgCl_2$; 0.1 mM
185 of each dNTP; 20 ng of genomic DNA and 1 U of Taq polymerase (Invitrogen,
186 Carlsbad, CA). Each polymerase chain reaction was performed by the procedure
187 of Schuelke (2000) using three primers: the specific forward primer of each
188 microsatellite with M13(-21) tail at its 5' end at 0.4 µM, the sequence-specific
189 reverse primer at 0.8 µM, and the universal fluorescent-labeled M13(-21) primer
190 at 0.4 µM. The following temperature profile was used: 94°C for 2 min, then 35
191 cycles of 94°C for 45 s, 50–60°C for 1 min, and 72°C for 1 min and 15 s,
192 finishing with 72°C for 5 min. Allele lengths were determined using an ABI Prism
193 3130 Genetic Analyzer with the aid of GeneMapper software, version 4.0
194 (Applied Biosystems).
195

196 *M-locus fine mapping*

197 Segregation distortion locus (SDL) associated with the PPM was detected using
198 JoinMap 3.0 software (Van Ooijen and Voorrips 2001) by analyzing the χ^2 values
199 of selected SSRs spread over the *Prunus* genome in a subset of the 'G×C-01'
200 (*N*=46) where all individuals carry the PPM. Using this population subset, genetic
201 maps for each linkage group were estimated following the procedure detailed

202 below for linkage group 3 (LG3) except for the logarithm of odds (LOD)
203 grouping threshold established in this case at 5.0.
204 The high density linkage map of 'Canino' chr.3 was constructed following the
205 "two-way pseudo test-cross" model of analysis (Grattapaglia and Sederoff 1994)
206 and using SSR markers segregating in the 'G×C-01' population. Calculations
207 were performed by JoinMap 3.0 software (Van Ooijen and Voorrips 2001) setting
208 "cross-pollinator" data type and using the Kosambi mapping function (Kosambi
209 1944) to convert recombination units into genetic distances. LG3 was established
210 under a LOD grouping threshold of 8.0 and a recombination frequency parameter
211 below 0.4. Segregation of the markers was analyzed one by one to correct
212 possible mistakes in the JoinMap 3.0 output and to develop a graphical ordering.
213
214 *BAC library hybridization*
215 BAC clones identification was made using digoxygenin labelled overlapping
216 oligoncleotides (overgo) probes hybridized in pools (Hilario et al. 2007) against
217 an apricot BAC library developed from the apricot cv. 'Goldrich' (Vilanova et al.
218 2003). A total of 22 overgo probes were designed, from the peach genome
219 sequence corresponding to the scaffold_3, using Overgo1.02i software (Cai et al.
220 1998) and following the website: http://www.mouse-
221 genome.bcm.tmc.edu/webovergo. Sequences without homology with repetitive
222 motifs were selected using the GDR website BLASTN tool (Altschul et al. 1990;
223 Jung et al. 2004). Hybridization was performed in four rounds using different
224 pools comprising up to 10 probes. Positive BACs were verified by PCR using
225 SSR markers in all cases and, occasionally, assigned to individual probes by re-
226 hybridization to colony dot blots.
227
228 *BAC-end sequencing*
229 BAC clones were inoculated into 96-deep well microplates and grown for 20 hrs
230 at 37ºC. Cells were harvested by centrifugation and BACs were purified in 96-
231 well plates by a standard alkaline lysis protocol. BAC DNA was precipitated with
232 isopropanol and washed with 70% ethanol. Sequencing was carried out on
233 ABI3730 equipment using the Big Dye Terminator v.3.1. cycle sequencing kit
234 (Applied Biosystems, Foster City, CA).

235

236 *Sequence analysis*

237 Sequences were edited with the Staden package v.1.4 (Bonfield 2004). Repetitive
238 DNA was identified with the RepeatMasker software (Smit et al. 1996), using the
239 viridiplantae section of the RepBase Update (Jurka et al. 2005) as database. The
240 non-repetitive fraction of the apricot BAC-ends was compared with peach genome
241 v1.0 (IPGI) and strawberry genome v1.0 (Shulaev et al. 2010) using the
242 standalone version of BLAST (Altschul et al. 1990), and with apple genome v1.0
243 (Velasco et al. 2010) using the BLASTN tool of the GDR website (Jung et al.
244 2004). BLASTN was performed with a cut-off value of $1e^{-5}$ in all cases. Parsing
245 of the BLAST results was performed with the Bio::SearchIO module from the
246 Bioperl package (Stajich et al. 2002).

247

248 *Contig construction*

249 BAC-end sequences were mapped against the peach reference genome using the
250 BWA-SW algorithm through the Burrows-Wheeler Aligner (BWA) software (Li
251 and Durbin 2010). In addition, SSRs developed from the peach genome sequence
252 were used to confirm the assigned positions and to anchor BACs into the apricot
253 genetic map by PCR.

254

255 **Results**

256

257 *S-locus unlinked PPM conferring SC in 'Canino' is located on linkage group 3*

258

259 SC in 'Canino' was shown to be associated with an *S*-locus unlinked pollen-part
260 mutation (PPM) (Vilanova et al. 2006), referred as the *M*-locus. To map the *M*-
261 locus, 141 trees from the 'G×C-01' segregating progeny were genotyped for the
262 mutation. 'Canino' has a defective *S*-haplotype and, under the proposed genetic
263 model, is considered heterozygous for the *M*-locus being therefore designated
264 S_2S_C *Mm* (Vilanova et al. 2006). Thus, according to the *S*- and *M*-locus genotypes,
265 the 'G×C-01' cross can be notated as 'Goldrich' (S_1S_2 *MM*) × 'Canino'(S_2S_C *Mm*).
266 Progeny *S*-genotypes fell into four classes: GC-S_1S_2 (28), S_2S_2 (22), S_1S_C (54) and
267 S_2S_C (37). Since trees showing S_1S_2 and S_2S_2 genotypes might only be derived
268 from pollen gametes with genotype S_2m they all were assigned the *Mm* genotype.

269 However, S_1S_C and S_2S_C trees might be derived from pollen gametes S_CM or S_Cm.
270 Thus, their *M*-locus genotype had to be inferred by screening of F_2 offspring
271 (Table 1), assigning the *MM* genotype to those 'G×C-01' S_1S_C and S_2S_C trees that
272 produced F_2 progeny with no embryos of genotypes S_1S_1 or S_2S_2. The minimal
273 sample size (≥ 25 seedlings) to analyze for this genotyping strategy was
274 established with an α risk below 0.01 in 98% of the cases. Consequently, the *Mm*
275 genotype was assigned to those 'G×C-01' S_1S_C and S_2S_C trees found to produce at
276 least one S_1S_1 or S_2S_2 homozygous embryo in the F_2 offspring, and therefore in
277 some of these cases less than 25 F_2 embryos were analyzed (Table 1).

278 After the *M*-locus genotyping, 180 SSR markers (Supplemental Tables 1 and
279 2) were selected according to their positions distributed across the eight *Prunus*
280 chromosomes (ranging from 18 in LG2 to 26 in LG8). All these SSR markers
281 were tested in the 'G×C-01' population parents and 42 (23%) of them displayed
282 polymorphism in 'Canino' and were, thus, useful to look for associations with the
283 *M*-locus. Polymorphic SSRs were subsequently tested in a selected subset of the
284 'G×C-01' progeny that comprise 46 trees with the *Mm* genotype (Table 2). In this
285 particular subset, SSR markers linked to the *M*-locus should be highly distorted,
286 since only 'Canino' pollen gametes carrying the *m*-allele are 'represented'. The
287 expected ratio of pollen alleles for an SSR marker unlinked to the *M*-locus is 1:1
288 (Table 2). Accordingly, distorted markers were mainly found in LG3 and LG6
289 (χ^2> 3.84 and P< 0.05). The maximum genetic estimated distance between any
290 pair of markers was no more than 40cM (except for distal end of LG1 ~55cM)
291 (Table 2). Consequently, in the most unfavorable scenario, distance to expected
292 *M*-locus should be lower than 20cM and recombination frequency lower than 0.2.
293 For this case, the expected ratio of pollen alleles for an *M*-locus linked SSR
294 would be 4:1 and only markers located on LG3 and LG6 fulfill this prediction.
295 The *M*-locus segregates independently of the *S*-locus (Vilanova et al. 2006), thus,
296 since distortion of LG6 markers is directly associated with the *S*-locus, LG3 is the
297 most likely location for the *M*-locus. An additional distortion, observed in the
298 CPSCT006 marker on LG5, proved to be associated with pistil alleles and
299 therefore unrelated to the PPM (data not shown).
300
301 *High-density mapping of the M-locus on chr.3*

302

303 As a first step toward constructing a high-density map of the *M*-locus region on
304 chr.3, 102 SSRs from the peach scaffold_3 sequence (Supplemental Table 3) and
305 21 additional SSRs available from the GDR website (Jung et al. 2004) were tested
306 on the 'G×C-01' population parents. A similar percentage of these SSRs (28-30%)
307 did not amplify or produced multi-band patterns in both 'Goldrich' and 'Canino',
308 but polymorphism of amplified SSRs was much higher in 'Goldrich' (55%) than
309 in 'Canino' (23%) (Supplemental Tables 2 and 3). All SSRs polymorphic in
310 'Canino' were tested on the 141 'G×C-01' trees previously genotyped for the
311 PPM. Twenty-five mapped to LG3 consistent with a genetic map of 88cM and an
312 average marker density of 0.28 marker/cM (Fig. 1). However, marker density
313 increased up to 0.75 marker/cM in the region flanked by the most distorted
314 markers UDAp493 and EPPCU7190 (Table 2) according to the selective SSR
315 design criteria. In this map, the *M*-locus is flanked by PGS3_71 and PGS3_96
316 markers within an interval of 1.8 cM. PGS3_62 marker co-segregates with *M*-
317 locus being also within this interval. Graphical ordering of genotype data enable
318 the positioning of recombination breakpoints to confirm map order (Fig. 2).

319 The *M*-locus region map was confirmed by analyzing 152 additional seedlings
320 derived from the same cross scheme ('G×C'). Thirty of them, also belonging to
321 the 'G×C-01' progeny, have been tested for all LG3 markers and the *S*-locus, but
322 still have not been genotyped for the *M*-locus (F_2 offspring). The rest, later
323 produced by two consecutive crosses 'G×C-07' (N=45) and 'G×C-08' (N=77),
324 were tested for a subset of 6 SSRs encompassing the *M*-locus
325 (PGS3_34/PGS3_25 interval) and the *S*-locus. As a whole, the *S*-locus genotypes
326 fell into the four expected classes S_1S_C (40), S_2S_C (45), S_1S_2 (19) and S_2S_2 (18), and
327 this ratio fits a model where the pollen parent is heterozygous for the unlinked *M*-
328 locus (χ^2 = 0.81 and P >0.84). Furthermore, order and distances estimated
329 between SSRs in the *M*-locus region were fully supported (data not shown). Three
330 new recombinants within the *M*-locus region were found in this additional set of
331 'G×C' seedlings but their *M*-locus genotype still has not been determined. Finally,
332 a similar analysis performed on a different population 'C×C-05' (N=80) also
333 confirmed previous results (data not shown).

334

335 *Construction of a ~364-kb contig containing the M-locus*

336

337 Based on the high degree of marker colinearity between apricot and peach
338 genome maps (Table 2) and according to the peach genome sequence (IPGI), the
339 smallest apricot region containing the *M*-locus was estimated to be ~364-Kb in
340 size (Figs. 2 and 3). To construct a contig covering this region, 22 overgo probes
341 were designed from the peach genome sequence (Supplemental Table 4) and
342 hybridized against the apricot genomic BAC library developed by Vilanova et al.
343 (2003). A total of 166 positive BAC clones were detected and 26 (16%) of them
344 were re-confirmed by PCR (Supplemental Table 4). Subsequently, BAC-end
345 sequences (BES) were mapped against the peach genome sequence, and SSRs
346 distributed throughout the region were analyzed to confirm BAC positions and to
347 anchor them into the 'Goldrich' genetic map. Fig. 3 shows the contig covering the
348 *M*-locus region constructed with a subset of 12 selected BACs based on genetic
349 and physical mapping.

350 Markers and BES analysis pointed out that apricot and peach are also highly
351 collinear within the *M*-locus region, except for a ~ 81Kb subregion between the
352 18.67 and 18.75 Mb positions (according to the peach genome). Mis-alignment
353 between apricot and peach was roughly delimited by the 253J12_Sp6 and
354 161F24_Sp6 BES positions (Fig. 3) and confirmed by PCR failed amplification of
355 peach SSRs located within this interval (PGS3_90, 91, 93 and 94) from apricot
356 BACs (251L05, 95D02, 164D09 and 160J21) (data not shown). Interestingly, this
357 peach subregion shows a high repeat content including two long tandem repeats
358 (LTR) of 7.2 Kb (LTR_1774) and 4.8 Kb (LTR_1775) in size (IPGI). Moreover,
359 apricot BESs mis-aligned with peach scaffold_3 show significant homology
360 mainly with two close regions on peach scaffold_2 located around positions 21.84
361 Mb (9.9 Kb in length; 251L05_T7, 160J21_T7 and 159P08_T7) and 24.32 Mb
362 (9.1 Kb in length; 65P22_T7, 164D09_Sp6 and 95D02_T7) (Fig. 3 and
363 Supplemental Table 5).

364

365 *Synteny analysis*

366

367 A total of 52 BAC-end reads were obtained after partial sequencing of the 26
368 positive BACs identified (Supplemental Tables 4 and 5). All BES remained as
369 singletons and were blasted as single query sequences (QS) against the nucleotide
370 genome sequences of peach (IPGI), apple (Velasco *et al*., 2010) and strawberry
371 (Shulaev et al. 2010) with a cut-off value of 1E-05. To map BES unambiguously
372 on the heterologous genomes only first hits were considered (Fig. 4). BLASTN
373 analysis revealed that all 56 apricot BES have significant homology to the peach
374 genome (45 to scaffold_3 and 7 to scaffold_2), 35 to the apple genome (30 to
375 chr.9 and/or chr.17 and 5 to other chrs.) and 30 to the strawberry genome (25 to
376 LG6 and 5 to other LGs) (Fig. 4 and Supplemental Table 5). When compared with
377 peach scaffold_3 and on average, first hit length was 65% of the QS length with
378 92% id and E-values were below $1e^{-49}$ in all cases. Comparison with apple and
379 strawberry showed a lower but still significant degree of similarity. First hit length
380 was 27% of the QS length with 89% id and E-values below $1e^{-7}$ when compared
381 with apple homeologous chrs. 9 and 17, and 26% of the QS length with 86% id
382 and E-values below $1e^{-5}$ when compared with strawberry LG6.

383 As a whole, apricot BES, mapped on the peach chr.3, were collinear with the
384 apricot LG3 genetic map, with the strawberry LG6 and with the inversely oriented
385 apple chrs. 9 and 17 (Fig. 4). The genomic landscape of the 364 Kb peach region
386 syntenic to the apricot *M*-locus contains 59 predicted gene transcripts as annotated
387 by IPGI. Velasco et al. (2010) found 83 predicted genes in the apple syntenic
388 region of chr. 17 (380 Kb between 4.94-5.32 Mb positions) and, according to
389 Shulaev et al. (2010), 54 predicted genes reside in the strawberry syntenic region
390 of chr. 6 (300 Kb between 31.65-31.95 Mb positions). A significant conservation
391 of gene content between the three genomes was found. BLASTX analysis of the
392 apple syntenic region against the peach predicted proteins database (IPGI) with an
393 exp. value cut-off $<1e^{-6}$ found 38 homologous open reading frames (ORF) (64%
394 of the predicted genes in peach). The same analysis performed with strawberry
395 detected 22 homologous ORFs (37% of the predicted genes in peach).
396 Furthermore, gene order was fully preserved in both cases (data not shown).

397 BLASTP analysis of the 59 ORFs comprised within the peach syntenic region
398 against The Arabidopsis Information Resource (TAIR) database, with an exp.
399 value cut-off $<1e^{-6}$, was used by IPGI to predict gene functions based on
400 homology to *Arabidopsis* (Supplemental Table 6). This table also indicates those

401 *Prunus*/*Arabidopsis* gene pairs that are best-reciprocal BLASTP hits (blasting
402 *Arabidopsis* proteins against the peach predicted peptides annotated by IPGI)
403 identifying putative orthologues. According to the large-scale gene expression
404 analysis performed by Wang et al. (2008) in *Arabidopsis* mature pollen, hydrated
405 pollen and pollen tubes using Affymetrix ATH1 Genome Arrays, up to 22 of these
406 *Arabidopsis* homologues were found to be pollen-expressed (Supplemental Table
407 6). Our preliminary results, based on RT-PCR performed on a few of these genes,
408 suggest that tissue-specific gene expression in apricot roughly matches that in
409 *Arabidopsis* for their corresponding homologues (data not shown).
410
411 **Discussion**
412
413 *Dysfunction in modifier loci and GSI breakdown in Prunus*
414
415 Dysfunction in the *S*-locus genes is the main cause of SC in *Prunus* (Yamane and
416 Tao 2009). Thus, mutations producing low *S-RNase* transcription levels (Yamane
417 et al. 2003) and mutations disrupting SFB function (Ushijima *et al*., 2004) both
418 have been shown to confer SC. Particularly in cultivated apricots, SC has been
419 mostly associated with the S_C-haplotype carried by numerous cultivars (Halász et
420 al. 2007), where a 358-bp insertion is found in the SFB_C gene resulting in the
421 expression of a truncated protein (Vilanova et al. 2006). However, genetic and
422 molecular analysis of the apricot selection 'Canino' (S_2S_C) revealed an additional
423 mutation unlinked to the *S*-locus independently conferring SC (Vilanova et al.
424 2006). Evidence for this came from analysis of *S*-genotypes among 'Goldrich'
425 (S_1S_2) × 'Canino' (S_2S_C) progeny. On one side, two unexpected classes were
426 found, S_1S_2 and S_2S_2, since pollen tubes carrying the 'Canino ' S_2-haplotype
427 should be incompatible on 'Goldrich' styles. On the other, segregation of the *S*-
428 genotypes fitted a genetic model where the pollen parent carries an *S*-locus
429 unlinked mutation. Moreover, other possible causes for SC were all discarded
430 such as mutations or indels affecting the S_2-haplotype, polyploidy and *S*-allele
431 duplications. The mutation does not seem to affect *S*-locus *F-box* gene expression
432 and, therefore, this GSI mutated modifier gene might be grouped with those
433 required for pollen rejection but with no wider role in pollination as defined by
434 McClure et al. (2000). Dysfunction in modifier loci has also been associated with

435 GSI breakdown in other *Prunus*. For instance, similar to the case of 'Canino', an
436 *S*-locus unlinked pollen-part modifier gene conferring SC was found in the sweet
437 cherry (*P. avium*) cv. 'Cristobalina' (Wünsch and Hormaza 2004). Moreover,
438 gene duplications and modified transcription levels of *S*-locus genes were also
439 discarded as the cause of SC in this case (Wünsch et al. 2010). In almond (*P.*
440 *amygdalus*), a stylar-part modifier affecting *S-RNase* expression has also been
441 suggested to confer SC (Fernández et al. 2009) and, more recently, Boskovic et al.
442 (2010) provided preliminary data for a non-*S* locus stylar factor essential for GSI
443 in diploid *Fragaria*. Nevertheless, despite of the ample genetic evidence
444 accumulated, no GSI modifier gene has yet been cloned in Rosaceae.
445
446 *Paving the way for positional cloning of the M-locus modifier gene*
447
448 In this work, the 'Canino' GSI mutated modifier gene locus (*M*-locus) was
449 mapped as a way to facilitate identification and cloning. For mapping, 141 'G×C-
450 01' progeny were genotyped for the mutation by analyzing F_2 offspring, and
451 observed segregations supported independent inheritance of the *M*- and *S*-loci.
452 Simultaneously, a genetic strategy based on segregation distortion was initiated to
453 identify the *M*-locus genomic region. A chromosomal region that causes distorted
454 segregation ratios is referred to as segregation distortion locus (SDL) (Zhu and
455 Zhang 2007). In this case, pre-zygotic selection of gametes carrying the PPM
456 should produce distortion in the segregation of marker loci close to the SDL.
457 Following this approach, a set of 'G×C-01' PPM-carrying trees (S_1S_2 and S_2S_2)
458 was selected for testing genome-wide distributed SSRs to detect SDL by
459 examining changes in genotypic frequencies. At odds with 'Goldrich', 'Canino' is
460 highly homozygous and therefore a high number of SSRs had to be tested to
461 perform the SDL screening efficiently. Attending to segregation of pollen alleles,
462 two SDL were found on LG3 and LG6. Distortion on LG6 is caused by the *S*-
463 locus and, as previously stated, the *M*-locus is predicted to be unlinked to the *S*-
464 locus, thus, LG3 is the most likely location for the *M*-locus.
465 Screening the whole 'G×Ca-01' population with chromosome specific SSRs
466 allowed fine mapping of the *M*-locus on the 'Canino' chr.3 distal end.
467 Interestingly, Cachi and Wünsch (2011) also recently mapped the non *S*-locus
468 PPM conferring SC to the *P. avium* cv. 'Cristobalina' on the LG3. The closest

469 marker (EMPaS02) was positioned 3.2 cM above the PPM (to centromere
470 direction), although map accuracy was not sufficient to confirm marker order in
471 that region. According to the peach genome sequence EMPaS02 is located at
472 ~20,0 Mb position on chr. 3 while the apricot *M*-locus is found in an interval
473 between ~18,40-18,76 Mb positions. A different map location for each PPM
474 would support different genes as responsible for SC in 'Canino' and
475 'Cristobalina', but this point still requires confirmation. SC is a desired trait for
476 apricot breeding programs and, interestingly, 'Canino' provides an *S*-locus
477 independent source of this trait. Moreover, the small size of the marker bracket
478 flanking the *M*-locus (1.8 cM) guarantees a satisfactory control for marker-
479 assisted selection of SC.
480 Once the *M*-locus position was identified, first steps towards the map-based
481 cloning were undertaken. First, an apricot BAC library (Vilanova et al. 2003) was
482 hybridized using overgo probes to identify BACs within the *M*-locus
483 encompassing region. On average, 7.5 BACs per probe were detected but this
484 average decreased to 1.2 after rejecting BACs unconfirmed by PCR. However, the
485 BAC library was predicted to have a 22-fold genome coverage, though the
486 observed coverage was found to be only 8 after RFLP screening (Vilanova et al.
487 2003). The disparity with our results might be explained by unspecific
488 hybridization due to the non-stringent conditions used. The 26 positive BACs
489 identified formed a single contig covering the *M*-locus region, where the apricot
490 genetic interval of 1.8 cM corresponds to a physical interval of ~364 Kb in the
491 peach genome sequence. Apricot and peach were highly collinear within the *M*-
492 locus region except for an ~81 Kb subregion where peach scaffold_3 shows a
493 high repeat content and apricot BES show homology with peach scaffold_2. A
494 minor chromosomal translocation between chrs. 2 and 3 during *Prunus* speciation
495 might be suggested as the cause of this mis-alignment. Indeed, chromosomal
496 rearrangements do not seem to be unusual within the Rosaceae (Illa et al. 2011).
497
498 *M-locus synteny in the Rosaceae*
499
500 The comparative analysis of the available rosaceous genomes suggests a common
501 hypothetical ancestral chromosome (A5) for apple LG17/LG9, peach PG3 and
502 strawberry FG6 (Illa et al. 2011). Consistent with this conclusion, micro-synteny

503 analysis of the apricot/peach PG3 *M*-locus region allowed to identify synteny
504 blocks in apple LG9/LG17 and strawberry FG6, where gene content and
505 colinearity are basically preserved between genomes. Interestingly, the *Fragaria*
506 RNase *T*-locus also maps on FG6 (Boskovic et al. 2010) and the *Malus S*-locus on
507 LG17 (Maliepaard et al. 1998), and these two chromosomes do not share a
508 common ancestor with the *S*-locus bearing *Prunus* PG6 (Illa et al. 2011).
509 Therefore, co-localization in syntenic chromosomes might suggest a sort of
510 connection between the self-incompatibility related apricot *M*-locus, the apple *S*-
511 locus and the strawberry *T*-locus, but the *M*-locus syntenic blocks mapped at
512 opposite chromosome ends to these two latter loci. Moreover, Boskovic et al.
513 (2010) suggested that the *Fragaria S* and *T*-loci might be paralogues resulting
514 from a duplication in a common ancestor of *Fragaria* and *Prunus* that coalesced
515 in a single *S*-locus in the lineage leading to *Prunus*. Altogether, the *M*-locus does
516 not seem to have any evolutionary relationship with a putative Rosaceae ancestral
517 *S*-locus.

518 Distinct genetic and molecular features are exhibited by the S-RNase based
519 GSI system in *Prunus*, *Malus* and *Fragaria* (Tao and Iezzoni 2010; Boskovic et
520 al. 2010). Unlike *Prunus*, multiple, instead of single *SFB*, are located at the *Malus*
521 *S*-locus (Minamikawa et al. 2010) and two independent *S*-loci, instead of a single
522 one, control GSI in *Fragaria* (Boskovic et al. 2010). However, in spite of these
523 differences, there are major similarities, the pistil *S*-determinant is an S-RNase in
524 all three genera and the pollen *S*-determinant is an F-box protein at least in *Prunus*
525 and *Malus*. This suggests that other factors essential for GSI, such as the modifier
526 genes, might also be preserved across Rosaceae species. Following this reasoning,
527 it is tempting to speculate that the apricot/peach *M*-locus modifier gene
528 orthologues in apple and strawberry, if present, should be located within those
529 synteny blocks in chrs. 17/9 and chr. 6, respectively. Furthermore, under the same
530 hypothesis, those predicted genes conserved in all three species (only 18) should
531 be taking into special consideration to search for candidate genes. In fact,
532 orthologues of the *Nicotiana* HT-B (McClure et al. 1999) were identified in
533 *Solanum* (Kondo et al. 2002) and *Petunia* (Puerta et al. 2009), and SSK1
534 equivalent proteins have been found in the distantly related genera *Antirrhinum*
535 and *Petunia* (Zhao et al. 2010), suggesting that these modifier genes are conserved
536 across genera.

537 Beside the *M*-locus, other flower related traits, such as anther color, petals
538 color and number of developing carpels, also map to the distal part of peach chr.3
539 (Dirlewanger et al. 2004). Furthermore, two predicted genes found within the
540 peach *M*-locus region are homologous to the FERONIA receptor-like kinase that
541 mediates male-female interactions during pollen tube reception in *Arabidopsis*
542 *thaliana* (Escobar-Restrepo et al. 2007). Regarding the apricot modifier gene,
543 there is yet no sound evidence on its putative function and therefore candidate
544 genes can not be selected solely by the gene function annotation. On the basis of
545 sequence similarity, the *Prunus*/*Arabidopsis* gene pairs reported in this work
546 might be considered orthologues with high confidence (Zheng et al., 2005) and
547 consistently a high tissue-specific expression conservation should be expected in
548 general (Movahedi et al., 2011). According to this premise, a significant number
549 of the genes distributed throughout the *M*-locus region might be pollen-expressed,
550 being good candidate genes for *m*. However, the rest cannot be fully discarded
551 since orthologues inferred through sequence similarity do not share similar
552 biological functions in many cases (Movahedi et al., 2011). Preliminary gene
553 expression analysis performed in apricot seems to support the predictions based
554 on homology but further analysis is needed to confirm these results. In this
555 context, narrowing the region containing the *M*-locus along with the screening for
556 PPM-associated polymorphisms and a more detailed gene expression analysis,
557 will be necessary to isolate the modifier gene required for GSI in 'Canino'. The
558 identification of new factors contributing to control of pollen-pistil interactions in
559 *Prunus* would be a valuable step to elucidating the molecular mechanisms
560 underlying GSI and also to provide new tools to understand evolutionary forces
561 behind this trait.
562
563 **Acknowledgements**
564
565 This work was supported by grants from the 'Ministerio de Educación y Ciencia',
566 Research projects AGL2007-60709 and AGL2010-19208. We thank Bruce
567 McClure for review and helpful comments on the manuscript.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403-410

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucl Acids Res 27:573-580

Bonfield J (2004) Staden package, version 1.4. URL http://staden.sourceforge.net Accessed 27 July 2011

Boskovic R, Tobutt KR (1996) Correlation of stylar ribonuclease zymograms with incompatibility alleles in sweet cherry. Euphytica 90:245-250

Boskovic R, Sargent DJ, Tobutt KR (2010) Genetic evidence that two independent *S*-loci control RNase-based self-incompatibility in diploid strawberry. J Exp Bot 61:755-763

Busot GY, McClure B, Ibarra-Sánchez CP, Jiménez-Durán K, Vázquez-Santana S, Cruz-García F (2008) Pollination in *Nicotiana alata* stimulates synthesis and transfer to the stigmatic surface of NaStEP, a vacuolar Kunitz proteinase inhibitor homologue. J Exp Bot 59:3187-3201

Cai W, Reneker J, Chow C, Vaishnav M, Bradley A (1998) An anchored framework BAC map of mouse chromosome 11 assembled using multiplex oligonucleotide hybridization. Genomics 54:387–397

Cachi AM, Wünsch A (2011) Characterization and mapping of non-*S* gametophytic self-compatibility in sweet cherry (*Prunus avium* L.). J Exp Bot 62:1847-1856

Cruz-Garcia FC, Hancock N, Kim D, McClure B (2005) Stylar glycoproteins bind to S-RNase *in vitro.* Plant J 42:295-304

De Nettancourt D (2001) Incompatibility and incongruity in wild and cultivated plants. Springer-Verlag, Berlin

Dirlewanger E, Graziano E, Joobeur T, Garriga-Calderé F, Cosson P, Howad W, Arús P (2004) Comparative mapping and marker assisted selection in Rosaceae fruit crops. Proc Natl Acad Sci USA 101:9891–9896

Doyle JJ, Doyle JL (1987) A rapid isolation procedure for small quantities of fresh leaf tissue. Phyto Bull 19:11-15

Escobar-Restrepo JM, Huck N, Kessler S, Gagliardini V, Gheyselinck J, Yang W-C, Grossniklausi U (2007) The FERONIA receptor-like kinase mediates male-female interactions during pollen tube reception. Science 317:656-660.

Fernández A, Hanada T, Alonso JM, Yamane H, Tao R, Socías i Company R (2009) A modifier locus affecting the expression of the S-RNase gene could be the cause of breakdown of self-incompatibility in almond. Sex Plant Reprod 22:179-186

Goldraij A, Kondo K, Lee CB, Hancock CN, Sivaguru M, Vazquez-Santana S, Kim S, Phillips TE, Cruz-García F, McClure B (2006) Compartmentalization of S-RNase and HT-B degradation in self-incompatible *Nicotiana*. Nature 439:805-810

Grattapaglia D, Sederoff RR (1994) Genetic linkage maps of *Eucalyptus grandis* and *E. urophylla* using a pseudotest-cross strategy and RAPD markers. Genetics 137:1121–1137

Halász J, Pedryc A, Hegedüs A (2007) Origin and dissemination of the pollen-part mutated S_C haplotype which confers self-compatibility in apricot (*Prunus armeniaca*). New Phytol 176:792-803

Hancock CN, Kent L, McClure BA (2005) The 120kDa glycoprotein is required for *S*-specific pollen rejection in *Nicotiana*. Plant J 43:716-723

Hilario E, Bennell TF, Rikkerink E (2007) Screening a BAC library with nonradioactive overlapping oligonucleotide (overgo) probes. In: Hilario E, Mackay J (eds) Protocols for nucleic acid analysis by nonradioactive probes *(Methods in molecular biology vol. 353).* Humana press Inc, Totowa, NJ, pp 79-91

Hospital F (2003) Marker-assisted breeding. In: Newbury HJ (ed) Plant Molecular Breeding. CRC Press Blackwell Publishing Ltd., Oxford, UK, pp 30-59

Hua Z, Kao T-h (2006) Identification and characterization of components of a putative *Petunia S*-locus F-box-containing E3 ligase complex involved in S-RNase based self-incompatibility. Plant Cell 18:2531-2553

Huang J, Zhao L, Yang Q, Xue Y (2006) AhSSK1, a novel SKP1-like protein that interacts with the S-locus F-box protein SLF. Plant J 46:780-793

Illa E, Sargent DJ, López-Girona E, Bushakra J, Cestaro A, Crowhurst R, Pindo M, Cabrera A, van der Knaap E, Iezzoni A, Gardiner S, Velasco R, arús P, Chagné D, Troggio M (2011) Comparative analysis of rosaceous genomes and the reconstruction of a putative ancestral genome for the family. BMC Evol Biol 11:9

Juárez-Díaz JA, McClure B, Vázquez-Santana S, Guevara-García A, León-Mejía P, Márquez-Guzmán J, Cruz–García F (2006) A novel thioredoxin *h* is secreted in *Nicotiana alata* and reduces S-RNases *in vitro*. J Biol Chem 281:3418-3424

Jung S, Jesudurai C, Staton M, Du Z, Ficklin S, Cho I, Abbott A, Tomkins J, Main D (2004) GDR (Genome Database for Rosaceae): integrated web resources for Rosaceae genomics and genetics research. BMC Bioinformatics 5: 130

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase Update, a database of eukaryotic repetitive elements. Cytogen Genome Res 110:462-467

Kondo K, Yamamoto M, Itahashi R, Sato T, Egashira H, Hattori T, Kowyama Y (2002) Insights into the evolution of self-compatibility in *Lycopersicon* from a study of stylar factors. Plant J 30:143-153

Kosambi DD (1944) The estimation of map distance from recombination values. Ann Eugen 12:172–175

Lai Z, Ma W, Han B, Liang L, Zhang Y, Hong G, Xue Y (2002) An F-box gene linked to the self-incompatibility (*S*) locus of *Antirrhinum* is expressed specifically in pollen and tapetum. Plant Mol Biol 50:29-42

Lee CB, Kim S, McClure B (2009) A pollen protein, NaPCCP, that binds pistil arabinogalactan proteins also binds phosphatidylinositol 3-phosphate and associates with the pollen tube endomembrane system. Plant Physiol 149:791-802

Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler Transform. Bioinformatics 26:589-595

McClure B, Mou B, Canevascini S, Bernatzky R (1999) A small asparagine-rich protein required for *S*-allele-specific pollen rejection in *Nicotiana*. Proc Natl Acad Sci USA 96:13548-13553

McClure BA, Cruz-García F, Beecher B, Sulaman W (2000) Factors affecting inter- and intra-specific pollen rejection in *Nicotiana*. Ann Bot 85:113-123

McClure BA, Haring V, Ebert PR, Anderson MA, Simpson RJ, Sakiyama F, Clarke AE (1989) Style self-incompatibility gene products of *Nicotiana alata* are ribonucleases. Nature 342:955-957

Maliepaard C, Alston FH, van Arkel G, Brown LM, Chevreau E, Dunemann Evans KM, Gardiner S, Guilford P, van Heusden AW, Janse J, Laurens F, Lynn JR, Manganaris AG, den Nijs APM, Periam N, Rikkerink E, Roche P, Ryder C, Sansavini S, Schmidt H, Tartarini S, Verhaegh JJ, Vrielink-van Ginkel M, King GJ (1998) Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. Theor Appl Genet 97:60–73

Minamikawa M, Kakui H, Wang S, Kotoda N, Kikuchi S, Koba T, Sassa H (2010) Apple S locus region represents a large cluster of related, polymorphic and pollen-specific F-box genes. Plant Mol Biol 74:143-154

Movahedi S, Van de Peer Y, Vandepoele K (2011) Comparative network analysis reveals that tissue specificity and gene function are important factors influencing the mode of expression evolution in *Arabidopsis* and rice. Plant Physiol 156: 1316-1330

Murase K, Shiba H, Iwano M, Che F-S, Watanabe M, Isogai A, Takayama S (2004) A membrane-anchored protein kinase involved in *Brassica* self-incompatibility signalling. Science 303:1516-1519

Puerta AR, Ushijima K, Koba T, Sassa H (2009) Identification and functional analysis of pistil self-incompatibility factor *HT-B* of *Petunia*. J Exp Bot 60:1309-1318

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols *(Methods in Molecular Biology vol. 132)*, Humana Press, Totowa, NJ, pp 365-386

Schuelke M (2000) An economic method for the fluorescent labelling of PCR fragments. Nat Biotechnol 18:233-234

Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkert O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP, Burns P, Davis TM, Slovin JP, Bassil N, Hellens RP, Evans C, Harkins T, Kodira C, Desany B, Crasta OR, Jensen RV, Allan AC, Michael TP, Setubal JC, Celton J-M, Rees DJG, Williams KP, Holt SH, Ruiz Rojas JJ, Chatterjee M, Liu B, Silva H, Meisel L, Adato A, Filichkin SA, Troggio M, Viola R, Ashman T-L, Wang H, Dharmawardhana P, Elser J, Raja R, Priest HD, Bryant Jr DW, Fox SE, Givan SA, Wilhelm LJ, Naithani S, Christoffel A, Salama DY, Carter J, Lopez Girona E, Zdepski A, Wang W, Kerstetter RA, Schwab W, Korban SS, Davik J, Monfort A, Denoyes-Rothan B, Arus P, Mittler R, Flinn B, Aharoni A, Bennetzen JL, Salzberg SL, Dickerman AW, Velasco R, Borodovsky M, Veilleux RE, Folta KM (2010) The genome of woodland strawberry (*Fragaria vesca*). Nature Genet 43:109-116

Sijacic P, Wang X, Skirpan AL, Wang Y, Dowd PE, McCubbin AG, Huang S, Kao T-h (2004) Identification of the pollen determinant of S-RNase-mediated self-incompatibility. Nature 429:302-305

Sims TL, Ordanic M (2001) Identification of a *S*-ribonuclease-binding protein in *Petunia hybrida*. Plant Mol Biol 47:771-783

Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigian C, Fuellen G, Gilbert JGR, Korf I, Lapp H, Lehväslaiho H, Matsalla C, Mungall CJ, Osborne BI, Pocock MR, Schattner P, Senger M, Stein LD, Stupka

E,Wilkinson MD, Birney E (2002) The Bioperl toolkit: Perl modules for the life sciences. Genome Res 12:1611-1618

Tao R, Iezzoni AF (2010) The S-RNase-based gametophytic self-incompatibility system in *Prunus* exhibits distinct genetic molecular features. Sci Hort 124:423-433

Thomson D, Henry R (1995) Single-step protocol for preparation of plant tissue for analysis by PCR. Biotechniques 19:394-400

Ushijima K, Sassa H, Dandekar AM, Gradziel TM, Tao R, Hirano H (2003) Structural and transcriptional analysis of the self-incompatibility locus of almond: Identification of a pollen-expressed F-box gene with haplotype-specific polymorphism. Plant Cell 15:771-781

Ushijima K, Yamane H, Watari A, Kakehi E, Ikeda K, Hauck NR, Iezzoni AF, Tao R (2004) The *S* haplotype-specific F-box protein gene, *SFB*, is defective in self-compatible haplotypes of *Prunus avium* and *P. mume*. Plant J 39:573-586

Van Ooijen JW, Voorrips RE (2001) JoinMap®3.0, Software for the calculation of genetic linkage maps. Plant Research International, Wageningen, The Netherlands

Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D, Salvi S, Pindo M, Baldi P, Castelletti S, Cavaiuolo M, Coppola G, Costa F, Cova V, Dal Ri A, Goremykin V, Komjanc M, Longhi S, Magnago P, Malacarne G, Malnoy M, Micheletti D, Moretto M, Perazzolli M, Si-Ammour A, Vezzulli S, Zini E, Eldredge G, Fitzgerald LM, Gutin N, Lanchbury J, Macalma T, Mitchell JT, Reid J, Wardell B, Kodira C, Chen Z, Desany B, Niazi F, Palmer M, Koepke T, Jiwan D, Schaeffer S, Krishnan V, Wu C, Chu VT, King ST, Vick J, Tao Q, Mraz A, Stormo A, Stormo K, Bogden R, Ederle D, Stella A, Vecchietti A, Kater MM, Masiero S, Lasserre P, Lespinasse Y, Allan AC, Bus V, Chagné D, Crowhurst RN, Gleave AP, Lavezzo E, Fawcett JA, Proost S, Rouzé P, Sterck L, Toppo S, Lazzari B, Hellens RP, Durel C-E, Gutin A, Bumgarner RE, Gardiner SE, Skolnick M, Egholm M, Van de Peer Y, Salamini F, Viola R (2000) The genome of the domesticated apple (*Malus* x *domestica* Borkh.). Nature Genet 42:833-839

Vilanova S, Badenes ML, Burgos L, Martínez-Calvo J, Llácer G, Romero C (2006) Self-compatibility of two apricot selections is associated with two pollen-part mutations of different nature. Plant Physiol 42:629-641

Vilanova S, Romero C, Abernathy D, Abbott AG, Burgos L, Llácer G, Badenes ML (2003) Construction and application of a bacterial artificial chromosome (BAC) library of *Prunus armeniaca* L. for the identification of clones linked to the self-incompatibility locus. Mol Genet Genomics 269:685-691

Vilanova S, Romero C, Burgos L, Llácer G, Badenes ML (2005) Identification of self-(in)compatiblity alleles in apricot (*Prunus armeniaca* L.) by PCR and sequence analysis. J Am Soc Hort Sci 130:893-898

Wang Y, Zhang W-Z, Song L-F, Zou J-J, Su Z, Wu W-H (2008) Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in *Arabidopsis*. Plant Physiol 148: 1201-1211

Wünsch A, Hormaza JI (2004) Genetic and molecular analysis in Cristobalina sweet cherry, a spontaneous self-compatible mutant. Sex Plant Reprod 17:203-210

Wünsch A, Tao R, Hormaza JI (2010) Self-compatibility in 'Cristobalina' sweet cherry is not associated with duplications or modified transcription levels of *S*-locus genes. Plant Cell Rep 29:715-721

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

Yamane H, Ikeda K, Hauck NR, Iezzoni AF, Tao R (2003) Self-incompatibility (*S*) locus region of the mutated S^6-haplotype of sour cherry (*Prunus cerasus*) contains a functional pollen *S* allele and a non-functional pistil *S* allele. J Exp Bot 54:2431-2437

Yamane H, Tao R (2009) Molecular basis of self-(in)compatibility and current status of *S*-genotyping in Rosaceous fruit trees. J Jpn Soc Hort Sci 78:137-157

Zhao L, Huang J, Zhao Z, Li Q, Sims TL, Xue Y (2010) The Skp-like protein SSK is required for cross-pollen compatibility in S-RNase-based self-incompatibility. Plant J 62:52-63

Zheng XH, Lu F, Wang Z-Y, Zhong F, Hoover J, Mural R (2005) Using shared genomic synteny and shared protein functions to enhance the identification of orthologous gene pairs. Bioinformatics 21: 703-710

Zhu C, Zhang YM (2007) An EM algorithm for mapping distortion segregation loci. BMC Genetics 8:82

Table 1.- *M*-locus genotyping of GC-S_1S_C and GC-S_2S_C trees belonging to the 'G×C-01' population. *S*-genotypes were determined by PCR-based amplification of *S-RNase* alleles in the F_2 progenies. Number of embryos falling into each *S*-genotypic class are indicated. Chi-square (χ^2 and *P* values for the expected segregation ratios 1:1 (*MM*) and 1:3:2 (*Mm*) obtained from each independent F_2 population are also shown.

_S_-genotypes of F_2 progenies from S_1S_C 'GxC-01' F_1 trees

Hyb	S_1S_1	S_1S_C	S_CS_C	Total	χ^2 (*P*-value)	Gen	Hyb	S_1S_1	S_1S_C	S_CS_C	Total	χ^2 (*P*-value)	Gen
GC-4	0	12	14	26	0.15 (0.70)	*MM*	GC-93	0	12	16	28	0.57 (0.45)	*MM*
GC-7	2	17	6	25	3.40 (0.18)	*Mm*	GC-95	1	21	6	28	7.57 (0.02)[a]	*Mm*
GC-12	6	13	9	28	0.46 (0.79)	*Mm*	GC-103	3	15	10	28	0.72 (0.70)	*Mm*
GC-14	2	6	5	13	0.16 (0.93)	*Mm*	GC-104	0	16	15	31	0.15 (0.70)	*MM*
GC-24	3	2	3	8	3.14 (0.21)	*Mm*	GC-105	6	4	2	12	9.67 (0.01)[a]	*Mm*
GC-31	0	14	13	27	0.04 (0.85)	*MM*	GC-111	5	14	9	28	0.04 (0.95)	*Mm*
GC-32	0	14	14	28	0.00 (1.00)	*MM*	GC-115	3	8	6	17	0.06 (0.97)	*Mm*
GC-34	0	17	11	28	1.29 (0.26)	*MM*	GC-117	7	9	12	28	3.71 (0.16)	*Mm*
GC-38	6	11	11	28	1.32 (0.52)	*Mm*	GC-119	7	10	11	28	2.60 (0.27)	*Mm*
GC-41	0	15	13	28	0.14 (0.71)	*MM*	GC-121	0	15	13	28	0.14 (0.71)	*MM*
GC-46	0	15	10	25	1.00 (0.32)	*MM*	GC-127	3	14	11	28	0.90 (0.64)	*Mm*
GC-47	2	10	10	22	1.82 (0.40)	*Mm*	GC-131	0	17	11	28	1.29 (0.26)	*MM*
GC-50	0	15	13	28	0.14 (0.71)	*MM*	GC-144	4	9	8	21	0.43 (0.81)	*Mm*
GC-52	3	8	7	18	0.28 (0.87)	*Mm*	GC-148	7	18	8	33	1.36 (0.51)	*Mm*
GC-56	2	16	10	28	1.86 (0.39)	*Mm*	GC-150	0	13	14	27	0.04 (0.85)	*MM*
GC-61	0	15	13	28	0.14 (0.71)	*MM*	GC-160	4	16	8	28	0.57 (0.75)	*Mm*
GC-62	4	6	5	15	1.20 (0.55)	*Mm*	GC-165	3	15	7	25	1.04 (0.59)	*Mm*
GC-64	0	11	16	28	0.93 (0.64)	*MM*	GC-167	7	16	5	28	3.46 (0.18)	*Mm*
GC-65	0	16	12	28	0.57 (0.45)	*MM*	GC-168	0	11	17	28	1.29 (0.26)	*MM*
GC-75	1	22	5	28	9.47 (0.01)[a]	*Mm*	GC-178	0	11	17	28	1.29 (0.26)	*MM*
GC-76	0	18	10	28	2.29 (0.13)	*MM*	GC-181	8	11	9	28	3.03 (0.22)	*Mm*
GC-79	1	15	12	28	3.72 (0.16)	*Mm*	GC-186	4	11	10	25	0.52 (057)	*Mm*
GC-80	4	17	7	28	1.32 (0.52)	*Mm*	GC-191	0	14	15	29	0.03 (0.85)	*MM*
GC-82	4	13	11	28	0.47 (0.59)	*Mm*	GC-193	0	13	15	28	0.14 (0.71)	*MM*
GC-86	0	17	11	28	1.29 (0.26)	*MM*	GC-194	2	8	4	14	0.29 (0.87)	*Mm*
GC-88	0	14	14	28	0.00 (1.00)	*MM*	GC-195	2	12	13	27	3.33 (0.19)	*Mm*
GC-90	0	12	17	29	0.86 (0.35)	*MM*	GC-196	3	6	7	16	1.06 (0.59)	*Mm*

Total no. of genotyped trees: **23 S_1S_C _MM_ and 31 S_1S_C _Mm_**

_S_-genotypes of F_2 progenies from S_2S_C 'GxC-01' F_1 trees

Hyb	S_2S_2	S_2S_C	S_CS_C	Total	χ^2 (*P*-value)	Gen	Hyb	S_2S_2	S_2S_C	S_CS_C	Total	χ^2 (*P*-value)	Gen
GC-8	4	16	8	28	0.57 (0.75)	*Mm*	GC-101	0	17	10	27	1.82 (0.18)	*MM*
GC-10	0	17	11	28	1.29 (0.26)	*MM*	GC-109	4	15	9	28	0.18 (0.91)	*Mm*
GC-13	0	22	6	28	9.14 (0.002)[a]	*MM*	GC-112	0	18	14	32	0.50 (0.48)	*MM*
GC-20	6	10	12	28	2.28 (0.32)	*Mm*	GC-114	2	14	12	28	2.29 (0.32)	*Mm*
GC-22	2	17	9	28	2.18 (0.34)	*Mm*	GC-116	0	15	11	26	0.62 (0.43)	*MM*
GC-23	6	14	8	28	0.57 (0.75)	*Mm*	GC-120	6	10	12	28	2.29 (0.32)	*Mm*
GC-30	0	17	11	28	1.29 (0.26)	*MM*	GC-126	2	3	1	6	1.45 (0.47)	*Mm*
GC-37	0	18	8	26	3.85 (0.05)	*MM*	GC-146	0	18	17	35	0.03 (0.86)	*MM*
GC-48	3	3	3	9	2.00 (0.37)	*Mm*	GC-155	7	15	6	28	2.42 (0.30)	*Mm*
GC-55	0	14	10	24	0.67 (0.41)	*MM*	GC-159	0	20	8	28	5.14 (0.02)[a]	*MM*
GC-57	2	8	6	16	0.25 (0.88)	*Mm*	GC-161	4	5	5	14	1.80 (0.41)	*Mm*
GC-63	0	16	12	28	0.57 (0.45)	*MM*	GC-180	0	17	11	28	1.29 (0.26)	*MM*
GC-74	0	18	12	30	1.20 (0.27)	*MM*	GC-183	5	10	13	28	2.61 (0.27)	*Mm*
GC-77	0	15	13	28	0.14 (0.71)	*MM*	GC-184	5	9	10	24	1.50 (0.47)	*Mm*
GC-78	0	18	10	28	2.29 (0.13)	*MM*	GC-187	0	10	17	27	1.82 (0.18)	*MM*
GC-96	4	12	12	28	1.15 (0.56)	*Mm*	GC-188	0	12	15	27	0.33 (0.56)	*MM*
GC-98	0	18	10	28	2.29 (0.13)	*MM*	GC-189	5	14	9	28	0.04 (0.98)	*Mm*
GC-99	4	9	13	26	3.38 (0.18)	*Mm*	GC-190	4	12	8	24	0.00 (1.00)	*Mm*

	GC-192	3	13	12	27	1.43 (0.49)	*Mm*
Total no. of genotyped trees: **18 S_2S_C *MM* and 19 S_2S_C *Mm***							

[a] Observed ratios differ significantly from expected at $P < 0.05$

Table 2.- Identification of segregation distortion SSR loci distributed throughout the 'Canino' linkage groups (LG) using a subset of the 'G×C-01' population carrying the PPM (*N*=46). χ^2 and *P* values estimated for each SSR, considering the expected segregation ratio are indicated.

LG	Locus	Peach Mb[a]	Apricot cM[b]	Seg. type[c]	-c	-d	-e	-g	hh	hk	kk	nn	np	Total	M^{M} (*P*-value)[d]
1	Gol051	4,69	00,0 (0,26)	<efxeg>			26	19						45	1,09 (0,30)
1	ssrPaCITA5	11,32	29,4 (0,00)	<efxeg>			24	22						46	0,09 (0,77)
1	EPPCU0027	9,51	29,5 (0,21)	<efxeg>			21	24						45	0,20 (0,65)
1	UDAp414	26,52	52,6 (0,16)	<abxcd>	22	23								45	0,02 (0,88)
1	CITA7	32,02	69,7 (0,00)	<efxeg>			26	18						44	1,46 (0,23)
1	EPPCU1589	31,81	69,7 (0,42)	<efxeg>			19	26						45	1,09 (0,30)
1	Gol004	45,40	124,8	<nnxnp>								24	22	46	0,09 (0,77)
2	ssrPaCITA16	03,76	00,0 (0,19)	<efxeg>			18	27						45	1,80 (0,18)
2	ssrPaCITA19	13,01	20,3 (0,04)	<efxeg>			29	17						46	1,39 (0,24)
2	CPSCT044	17,22	24,0 (0,04)	<efxeg>			17	29						46	1,39 (0,24)
2	UDP98-411	20,17	27,8 (0,11)	<abxcd>	17	29								46	1,39 (0,24)
2	CPSCT021	23,74	39,4 (0,13)	<efxeg>			26	19						45	1,09 (0,30)
2	CPSCT034	26,35	52,3	<efxeg>			26	19						45	1,09 (0,30)
3	ssrPaCITA23	02,70	00,0 (0,02)	<abxcd>	14	32								46	7,04 (0,008)[e]
3	UDAp468	04,85	02,9 (0,02)	<efxeg>			31	13						44	7,36 (0,007)[e]
3	EPPCU2256	06,14	05,8 (0,19)	<efxeg>			32	14						46	7,04 (0,008)[e]
3	UDAp493	15,17	25,9 (0,15)	<efxeg>			41	4						45	30,4 (3e-8)[e]
3	EPPCU7190	19,78	41,9	<hkxhk>					1	14	26			41	34,6 (3e-8)[e]
4	BPPCT040	06,46	00,0 (0,03)	<efxeg>			15	25						40	2,50 (0,11)
4	CPDCT045	06,21	02,5 (0,33)	<efxeg>			19	27						46	1,39 (0,24)
4	CPSCT005	29,88	41,8	<efxeg>			19	26						45	1,09 (0,30)
5	PGS5_03	05,06	00,0 (0,08)	<nnxnp>								20	18	38	0,56 (0,46)
5	ssrPaCITA21	10,78	08,7 (0,05)	<nnxnp>								24	21	45	0,20 (0,65)
5	CPSCT006	11,53	15,5 (0,05)	<hkxhk>					18	24	4			46	8,61 (0,013)[e]
5	BPPCT037	12,31	22,3 (0,00)	<abxcd>	20	23								43	0,21 (0,65)
5	pchgms4	12,67	23,3	<nnxnp>								21	24	45	0,20 (0,65)
6	PGS6_04	04,95	00,0 (0,24)	<efxeg>			21	24						45	0,20 (0,65)
6	UDAp420	08,14	27,0 (0,16)	<hkxhk>					8	21	17			46	3,74 (0,15)
6	Ma027a	20,90	44,4 (0,18)	<abxcd>	35	11								46	12,5 (0,0004)[e]
6	UDP98-412	24,75	62,6 (0,10)	<efxeg>			40	5						45	27,2 (1.8e-7)[e]
6	Locus-*S*	26,45	72,8 (0,02)	<efxeg>			46	0						46	46,0 (0,000)[e]
6	ssrPaCITA12	27,84	75,0	<efxeg>			44	2						46	38,4 (0,000)[e]
7	CPSCT004	6,68	00,0 (0,11)	<nnxnp>								23	23	46	0,00 (1,00)
7	CPPCT022	10,23	12,1 (0,04)	<efxeg>			20	26						46	0,78 (0,38)
7	UDP98_405	10,94	15,9 (0,00)	<nnxnp>								22	23	45	0,02 (0,88)
7	PGS7_05	13,08	23,2 (0,22)	<efxeg>			26	20						46	0,78 (0,38)
7	CPSCT042	17,08	45,6	<efxeg>			22	24						46	0,09 (0,77)
8	CPSCT018	00,12	00,0 (0,25)	<nnxnp>								19	26	45	1,09 (0,30)
8	PGS8_05	07,39	28,0 (0,09)	<nnxnp>								25	20	45	0,56 (0,46)
8	UDAp401	10,50	36,7 (0,04)	<nnxnp>								24	22	46	0,09 (0,77)
8	M6a	15,03	40,6 (0,11)	<nnxnp>								22	24	46	0,09 (0,77)
8	UDP98-409	17,78	49,1	<efxeg>			24	21						45	0,20 (0,65)

[a] Marker position (Mb) within the corresponding peach genome scaffolds which sizes were estimated by IPGI (scaffold_1, 46.88Mb; _2, 26.81Mb; _3, 22.02Mb; _4, 30.53Mb; _5, 18.50Mb; _6, 28.90Mb; _7, 22.79Mb and _8, 21.83Mb)

[b] Map position (cM) and rec. frequencies (in brackets) estimated by JoinMap 3.0 using the apricot 'G×C-01' population subset *N*=46

[c] Segregation type as per JoinMap 3.0

[d] Chi-square test was performed for the expected ratios 1:1 (as per <ab×cd>, <ef×eg> and <nn×np>) and 1:2:1 (<hk×hk>), respectively.

[e] Observed ratios differ significantly from expected at $P < 0.05$

Figure captions

Fig. 1 High-density simple sequence repeat (SSR) map of ‘Canino’ LG3 showing the *M*-locus location. Distances in centimorgan (cM) are shown on the left

Fig. 2 Graphical maps of the recombinant hybrids from the ‘GxC-01’ population at the *M*-locus corresponding to ‘Canino’. The map region between markers PGS3_34 and PGS3_25 is shown. Distances in centimorgan (cM) are shown on the right of the apricot map and their corresponding positions in megabases (Mb) on the peach genome sequence are shown on the left. Black vertical bars represent self-compatible (SC) and white bars self-incompatible (SI) chromosomal regions. Recombinant seedlings are numbered at the top

Fig. 3 Contig constructed with ‘Goldrich’ BACs covering the *M*-locus region on the distal part of apricot chr.3 (not to scale). Aligned BACs showing their BAC-ends Sp6 (S) and T7 (T) are represented by *grey boxes*. Mis-aligned fragments are shown as *white boxes*. SSRs amplified from BACs are indicated by *black dots* and those anchored into the ‘Goldrich’ genetic map are indicated by *white dots*. *Dashed-lines* indicate the SSR positions corresponding to the apricot genetic map and the peach physical map. Distances in centimorgan (cM) are shown at the top for the ‘Goldrich’ genetic map and those in megabases (Mb) are shown down below for the peach physical map. *Nº Rec* indicates the number of recombinants found in ‘GxC-01’ corresponding to ‘Goldrich’

Fig. 4 Comparative analysis of the apricot *M*-locus region with peach, apple and strawberry genomes. *Curly brackets* comprise apricot BAC-ends anchored into the ‘Goldrich’ map and *triangles* indicate syntenic positions on different genomes. Distances in centimorgan (cM) are shown on the left of ‘Goldrich’ genetic map and in megabases (Mb) on the right of peach, apple and strawberry physical maps

Fig. 1

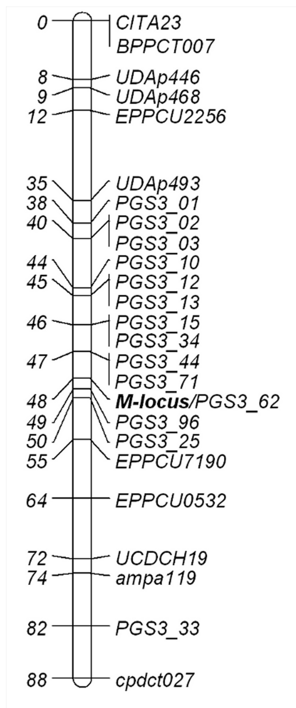

Fig. 2

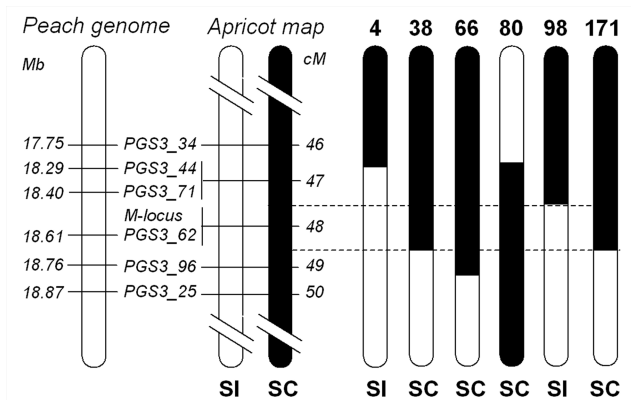

Fig. 3

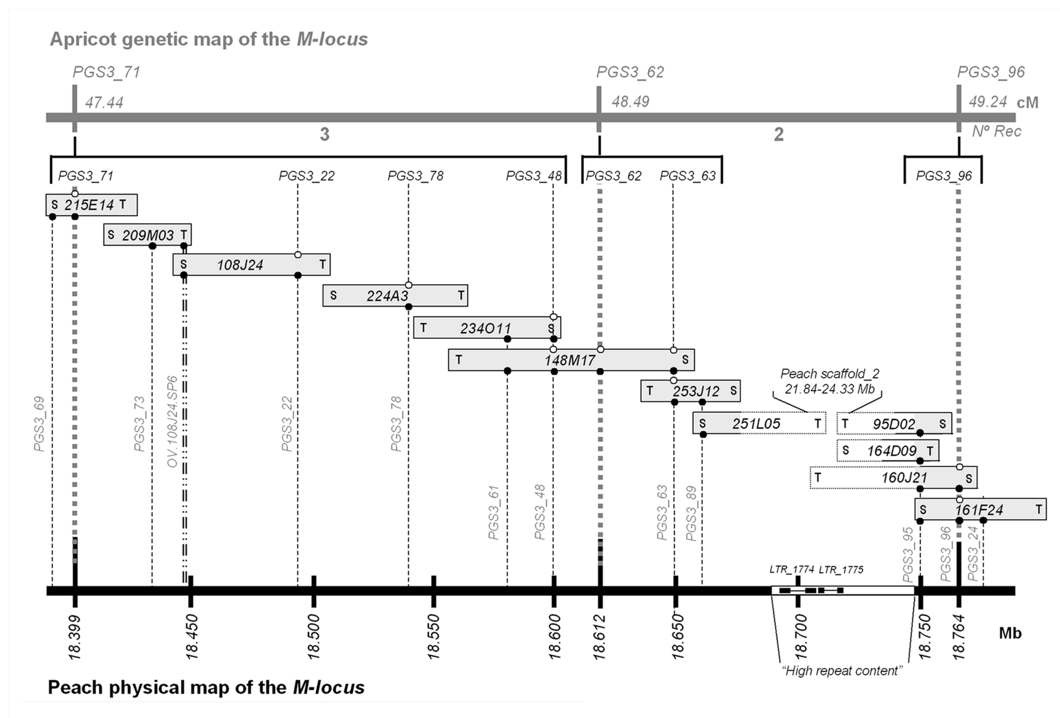

Fig. 4