Fine mapping of the peach pollen sterility gene (*Ps*/*ps*) and detection of markers for marker-assisted selection

Eduardo I[1,2], de Tomás C[1,2], Alexiou KG[1,2], Giovannini D[5], Pietrella M[5], Carpenedo S[6], Bassols Raseira MC[6], Batlle I[4], Cantin CM[3,7], Aranzana MJ[1,2], Arús P[1,2]

[1]IRTA, Institut de Recerca i Tecnologia Agroalimentàries, Barcelona, Spain

[2]Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Campus UAB, Cerdanyola del Vallès (Bellaterra), 08193 Barcelona, Spain.

[3]IRTA. FruitCentre. Parc Cientific i Tecnològic Agroalimentari de Lleida (PCiTAL). Edifici Fruitcentre, Parc de Gardeny, 25003 Lleida, Spain.

[4]IRTA. Centre Mas de Bover. Crta. Reus - El Morell Km 3.8. 43120 Constantí, Tarragona, Spain.

[5]CREA. Council for Agricultural Research and Economics, Research Centre for Olive, Fruit and Citrus Crops, Forlì, Italy.

[6]Embrapa Clima Temperado, 96.010-971, Pelotas, RS, Brazil.

[7]Current address: ITA-ARAID. Agrifood Research and Technology Centre of Aragón. Avda. Montañana 930, 50059-Zaragoza, Spain.

e-mail address: iban.eduardo@irta.cat

ORCID: 0000-0002-9963-2934

Abstract

In peach, pollen sterility, expressed as absence of pollen in the anthers, segregates as an undesired trait in breeding programs. Pollen fertility screening in progenies is not a common practice mainly because it does not affect fruit set since cross-pollination is frequent. It is also a time-consuming activity that coincides with the busy pollination season. Segregation for this trait could be avoided by using molecular markers to identify appropriate parents or male sterile plants for early culling in progenies expected to segregate, thus increasing breeding efficiency. In peach, pollen sterility is determined by a recessive allele in homozygosis of the major gene, *Ps/ps*, located on chromosome 6. In this work, using a conventional mapping approach combined with bulked segregant analysis using resequencing data, we fine mapped *Ps* to a region of almost 160 kb and developed molecular markers for marker-assisted breeding. These markers were validated in plant materials from three peach breeding programs, including progenies, advanced selections and cultivars,

allowing us to determine that the frequency of the *ps* allele is high (0.23) and also to infer the genotypes of a large collection of cultivars and advanced breeding lines.

Keywords

Prunus persica, male sterility, marker assisted breeding, breeding program

Acknowledgements

We acknowledge financial support from the Spanish Ministry of Economy and Competitiveness, through the ‘‘Severo Ochoa Programme for Centres of Excellence in R&D 2016–2019 (SEV-2015-0533) and Project AGL2015-68329-R, from the CERCA Programme (Generalitat de Catalunya), and the ‘Stony Hard’ project co-funding the CREA-Forlì breeding program. We also thank to C. Fontich (IRTA) for his help with the plant populations in Lleida.

Introduction

Male sterile individuals are sometimes found in segregating progenies of peach [*Prunus persica* L. (Batsch.)]. Flowers from male sterile individuals have no pollen and the anthers are smaller and with a paler color than in normal flowers. Sterile individuals should be discarded from breeding programs because they have low fruit production when grown in monovarietal plots. However, pollen sterility is usually observed in breeding programs when heterozygous individuals are used in crosses. The male or pollen sterility (PS) trait in peach was found to be determined by a single gene (*Ps*/*ps*) by Scott and Weinbeger (1944), where the recessive *ps* allele in homozygosis causes the phenotype. While identification of male sterile individuals can be straightforward, distinguishing whether fertile parents carry the *ps* allele is a long-term activity that requires progeny test analysis. When the PS genotype of the parents is unknown, it is likely that this trait will segregate in various breeding programs. During the flowering season, peach breeders concentrate on crossing, and pollen fertility is seldom assessed in the progenies, meaning male sterile individuals will only be detected by chance, often at the end of the selection process, when advanced selections are evaluated in larger monovarietal plots. At this stage, male sterile cultivars have low production and the problem is easily identified, but only after a substantial and unprofitable investment in space and time. The availability of efficient molecular markers tightly linked to the *Ps* gene, will allow peach breeders to detect and eliminate male sterile material, early and easily, from their breeding programs, and to plan crosses taking into consideration the possible segregation of this trait in their progenies.

A marker linked to the *Ps* locus has been described by Jun et al. (2004) using a bulked segregante analysis (BSA) approach with RAPD markers, and shortly after this, the gene was mapped to the proximal end of linkage group 6 (G6) by Dirlewanger et al. (2006). Other genetic causes of male sterility described in peach, include another gene (*Ps2*), identified in the cultivar 'White Glory' (Chaparro et al 1994; Werner and Creller 1997), and cytoplasmic male sterility induced by almond cytoplasm, with fertility recovered by two independent restorer factors from almond located on G2 and G6 (Donoso et al. 2015). The restorer gene on G6 and *Ps* are close on the chromosome, although they appear to be determined by different genes (Donoso et al. 2015).

Recent advances in genome analysis in peach and other *Prunus* (Aranzana et al. 2019) facilitates the adoption of marker-assisted selection in the improvement of these species. For peach, marker assisted selection has been used in several peach breeding programs worldwide for various major genes: melting flesh (Peace et al. 2005, Gu et al. 2016); fruit shape (Picañol et al 2013); fruit flesh color (Falchi et al. 2013,

Adami et al. 2013); acidity (Eduardo et al. 2014); glabrous skin (Vendramin et al. 2014); slow ripening (Eduardo et al. 2015); peach external color (Bretó et al. 2017, Sandefur et al. 2017); stony hard (Pan et al. 2015, Cirilli et al. 2018) and brachytic dwarfism (Hollender et al. 2015, Cantín et al. 2018). Using linkage mapping and whole genome sequence analysis, in this paper we fine mapped the *Ps* gene and validated two tightly linked markers, a simple sequence repeat (SSR) and a single nucleotide polymorphism (SNP), that were genotyped in a large set of breeding progenies and a broad germplasm collection. These markers can be useful for the early selection of this character in peach breeding programs and to identify heterozygous parents that when crossed may produce male-sterile plants in their offspring.

Material and Methods

Plant material and male-sterility phenotyping

For map construction we used two populations segregating for PS from the breeding program of IRTA (Lleida, Spain) consisting of the self pollinated progenies of two selections, PN732 and PN788, with 84 and 75 individuals respectively. For marker validation we used 935 individuals, including 373 genotypes (197 cultivars and 176 advanced selections) from the Cultivar and Advanced Selections Collection (CASC; Supplementary Table S1), and 562 individuals from 13 different segregating progenies (Supplementary Table S2). The CASC materials come from three research organizations: IRTA (Experimental Station at Gimenells, Spain), CREA (Experimental Station at Forli, Italy) and EMBRAPA (Experimental Station of Pelotas, Brazil) and those from segregating progenies from IRTA and CREA. All trees were cultivated using standard agricultural practices in their respective growing regions and data from PS phenotypes were collected in the springs between 2012 and 2019.

Male sterile phenotypes (Figure 1) were identified by visual inspection and manually squeezing mature anthers in recently opened flowers or in the stage immediately preceding opening in the field or in the lab. The male sterile phenotype does not produce pollen grains, and has smaller, empty anthers, white or cream colored, compared to the yellow of normal, fertile anthers. The phenotype is somewhat more difficult to identify in anthocyanic anthers, although their size and complete absence of yellow coloration is recognizable. Uncertain cases were examined with a binocular lens to identify the presence or absence of pollen grains.

Molecular markers

All DNA extractions were done from young leaf tissue using the Doyle and Doyle (1990) protocol in Eppendorf tubes or adapted to 96 well plates. For SSR genotyping, PCR reactions were carried out in a total volume of 10 µL using a PE9700 Thermal Cycler (PE/Applied Biosystems, Foster City, Calif., USA) under the following conditions: 1 min at 95 °C, 30 cycles of 15 s at 95 °C, 15 s at the appropriate annealing temperature, and 1 min at 72 °C, followed by a 5-min extension at 72 °C. Products were analyzed by capillary electrophoresis using the ABI/Prism 3130*xl* (PE/Applied Biosystems) automatic sequencer as in Aranzana et al. (2003), whereas 5% non-denaturing polyacrylamide gels were used to detect CPP21395 alleles in CREA germplasm.

Based on the known position of *Ps* on G6, we designed primers for 13 new SSRs located at the beginning of G6 using Primer3 software (Untergrasser et al. 2012) and the *Prunus* genome sequence and SSR annotation. Primer sequences are given in Table S3. InDels were identified visually using the IGV software (Thorvaldsdóttir et al. 2013) and the alignment files obtained from resequencing data. InDels were called manually when in a specific region there were no reads or around 50% of reads compared to the surrounding regions. PCR conditions were the same as that for SSRs but with extension time adapted to the size of each InDel. Genotyping was done visually using ethidium bromide 2% agarose gels.

InDel024900 was genotyped using a Kompetitive Allele Specific PCR (KASP) SNP assay following the instructions provided by the LGC group (https://www.biosearchtech.com). The primers used were: A1-GAAGGTGACCAAGTTCATGCTCAGTTGAAGATGTGATAACAGGGC, A2-GAAGGTCGGAGTCAACGGATTCCAGTTGAAGATGTGATAACAGGGT and C1-GTTGGGTTGCAATACACCGATTTCCAT. An 8 µL sample mix containing 4 µL of KASPar 2x Mastermix, 0.11 µL of the KASP assay primer mix (allele specific primers at 12 µM and the common reverse primer at 30 µM), 1.89 µL of water and 2 µL of DNA at 20-40 ng/ µL was used. Samples were analyzed using a Lightcycler 480 Instrument II (Roche Life Science) and the Lightcycler 480 software.

Linkage map

JoinMap v.4.1 (van Ooijen et al. 2011) was used for map construction for both PN732 and PN788 segregating populations that were treated as F_2 progenies. Grouping was performed with a threshold of LOD≥3.0 and distance was calculated with the Kosambi function. Linkage group terminology was according to the *Prunus* reference map (TxE) (Dirlewanger et al. 2004). The MapChart 2.1 software (Voorrips 2002) was used to draw the maps.

Bulked Segregant Analysis using resequencing data

In parallel with conventional linkage mapping we attempted a second strategy consisting of obtaining resequencing data from three groups of individuals with different genotypes at the *Ps* locus and comparing the observed polymorphisms searching for a genomic region encompassing *Ps* with the approach described as Bulked Segregant Analysis (BSA) by Paran et al. (1993). For that we selected six cultivars from IRTA collection: two, 'Ghiaccio-2' and 'BtxNr-6' ('BigTop' × 'Nectaross') were male sterile (*psps*), two were *Psps* heterozygotes ('BigTop' and 'Nectaross'), and two were *PsPs* homozygotes ('Armking' and 'Tifany'). In the case of 'Big Top' and 'Nectaross' we know that they are heterozygous for *Ps* because their progeny segregates for male sterility, whereas progeny of 'Tifany' and 'Armking' cultivars crossed with *ps* carriers have been identified in the IRTA breeding program as not segregating for male sterility, therefore we assumed they are not *ps* carriers. After the variant calling analysis described below, we looked for regions of the genome that were homozygous in the two male sterile cultivars, heterozygous in the two cultivars heterozygous for *ps,* and had the reference alleles in the two cultivars that did not carry the *ps* allele.

DNA was extracted from 'BtxNr-6', 'Ghiaccio-2', 'BigTop', 'Nectaross', 'Tifany' and 'Armking' and sent to CNAG (Centre Nacional d'Anàlisi Genòmica, Barcelona) for library preparation and sequencing. Paired-end sequencing (2 × 100) was run using an Illumina HiSeq 2000 sequencer (Illumina Inc, San Diego, CA, USA). The average depth of sequencing coverage was 35×. Adapter removal and quality-based trimming of the raw data was done with Trimmomatic v0.36 (Bolger *et al.* 2014). FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) was used for read quality control before and after trimming. High quality reads were mapped to the peach genome version 2.0 (ftp://ftp.bioinfo.wsu.edu/species/Prunus_persica/Prunus_persica-genome.v2.0.a1/assembly/Prunus_persica_v2.0.a1_scaffolds.fasta.gz) using the BWA-MEM algorithm (v0.7.16a-r1181; http://bio-bwa.sourceforge.net/bwa.shtml) with default parameters. Alignment file sorting, indexing and filtering of multi-mapped reads was done using samtools v1.5 (Li 2011). Raw Illumina data for the six peach individuals (BtxNr-6, Ghiaccio-2, BigTop, Nectaross, Tifany and Armking) can be found at the European Nucleotide Archive (ENA) under the accession numbers ERS4539604, ERS4539605, ERS3508163, ERS1801615, ERS1801617 and ERS1801609.

Three software tools were used for variant calling on the filtered alignment data: BCftools v1.5 (Danecek et al. 2016), platypus v0.8.1 (Rimmer et al. 2014) and GATK v3.8.0 (McKenna et al. 2010). We removed reads with mapping quality of less than 10 and bases with quality of less than 20 and discarded variants with less than eight supporting reads. Additionally, for the bcftools dataset, we discarded variants with less than eight supporting reads and genotype quality of less than 30. For the GATK data set we discarded variants with less than eight supporting reads, RMSMappingQuality of less than 30 and FisherStrand bias of more than 60. From the filtered variants we kept only those that were called with all three tools used. Functional annotation of the variants was performed using the SnpEff software v4.3 (Cingolani et al. 2012).

Results

Pollen sterility phenotyping

Pollen sterility was scored in all individuals from breeding programs and for all progenies of the mapping populations. Out of the 373 cultivars and accessions, 139 gave information about the PS genotype (Supplementary Table S1), whereas we had expected most of those with no PS phenotype would be fertile, particularly in the case of cultivars. Phenotypes were easily identified except for a few Brazilian accessions that had a pollen sterile phenotype but some pollen grains when looked at under the binocular microscope. Occasional pollen grains in sterile individuals has been reported previously (Hesse 1975), and these individuals were considered as pollen sterile for this study.

Mapping the Ps locus

Dirlewanger et al. (2006) found that the closest marker to *Ps* was FG40, an RFLP located 4.8 cM apart. However, it could not be physically located in the reference peach genome sequence because the DNA probe has not been sequenced. Instead we used another RFLP marker, FG215, 8.8cM from *Ps* and concentrated at Pp06:4,178,846 on the peach genome sequence, as a reference to search for SSRs concentrated in the region near the *Ps* locus (Supplementary Table S3). Only four markers segregated in the PN732 and PN788 populations and were succesfully mapped. All individuals had genotypes compatible with their origin as selfed progeny from their parents (Supplementary Figure S1). *Ps* was located between markers CPP21395 (Pp06:1,130,915) and CPP21490 (4,723,809) in both populations. The closest marker was CPP21395 (3.4-4.5 cM).

CPP21395 marker validation in peach breeding programs

CPP21395 was genotyped in 359 genotypes of the CASC using capillary electrophoresis (Table S1). CPP21395 had 6 alleles (187, 190, 194, 196, 198, 200) and 10 different genotypes (Table 1). Fourteen additional samples that were genotyped using acrylamide gels, where only allele 187 could be clearly assigned, were excluded from the allele frequency analysis. Assuming that individuals with a single allele were homozygotes, the frequency of this allele being associated with sterility was 0.23. For the other alleles, 200 and 198 were very abundant (frequencies of 0.46 and 0.30, respectively) and the remaining (190, 194, 196) were rare with frequencies ranging from 0.001 to 0.003.

We obtained reliable phenotypic and genotypic information regarding pollen fertility of 139 cultivars or advanced selections. Of the 22 male sterile individuals, 20 had only the 187 allele (187/-) while two from Brazil were 187/200 and 196/198. As expected from a male sterile cultivar, 'JH Hale' had the 187/- genotype. Four out of the 117 individuals with fertile pollen were homozygous for the 187 allele, three of them from Brazil and one from Spain. Of the 27 individuals known to be carriers of male sterility (*Psps*), 21 had allele 187 and another allele, one was 187/- and the remaining five did not carry the 187 allele.

CPP21395 was also validated as a potential marker for marker assisted selection (MAS) for *Ps* in several progenies of a Spanish and an Italian peach breeding programs. All the individuals from unselected progenies were genotyped with CPP21395 and phenotyped for pollen fertility. A total of 562 individuals from 13 different progenies were analyzed (Supplemetary Tables S2, S4). Fourteen individuals homozygous for the 187 allele had fertile pollen while the other 122 were male sterile. In the case of the heterozygous individuals for the 187 allele, six were male sterile while the remaining 250 individuals were fertile. Finally, two individuals that did not have the 187 allele were male sterile, while the remaining 168 were fertile. These results indicate that CPP21395 could predict the *Ps* phenotype with an efficiency of 96.1% (22 mistakes with 562 individuals). Considering that the unexpected genotypes were caused by crossovers between *Ps* and CPP21395, the recombination frequency estimated from these data is r=0.02 equivalent to ~2 cM Kosambi, lower than the estimation based on the linkage map, which is consistent with the fact that some crossovers could not be identified due to the dominance of the pollen fertility character.

Ps fine mapping with BSA

Using the resequencing data from six individuals with contrasting *Ps*/*ps* genotypes, we used two strategies to identify additional markers for the selection of the pollen sterility locus. The first consisted of finding SSRs and InDels that were polymorphic in the resequenced genotypes and located in the region of the CPP21395 marker. With this approach we identified five SSRs and three InDels (see Supplementary Table 5S) that were genotyped in the parents of PN732 and PN788 populations. Four segregating markers, InDel1860, CPP21215, CPP21230 and CPP21245, were added to the genetic maps (Figure 2). In population PN732, *Ps* cosegregated with marker CPP21245 and was flanked by Indel1860 and CPP21230, defining a genomic region of 1.6 cM, spanning ~232 Kb, between Pp06: 1,860,920 and Pp06: 2,192,629. In PN788 neither CPP21245 nor CPP21215 segregated and *Ps* was located between InDel1860 and CPP21490, and cosegregated with CPP21230.

BSA analysis using resequencing data indicated two target regions that were homozygous in the two male sterile cultivars, heterozygous in the two cultivars heterozygous for *ps* and had the reference alleles in the two cultivars that did not carry the *ps* allele. One was on G5 (13,622,325 to 14,502,877bp) and one on G6, coinciding with the chromosomal fragment where *Ps* mapped to (1,956,981 to 2,116,368bp). Twenty-seven annotated genes were identified in the latter region (Table 2), and we estimated the effects of the 90 SNP and InDel variants detected. Only one polymorphism, located in gene *Prupe6G024900* (Pp06:1,996,812), had an insertion polymorphism with a predicted high impact variant, producing a splicing variation (G>GC) that would cause a loss of function. This gene is annotated in the Peach genome v2.0 as a cellulose synthase. Sixty cellulose synthase genes are annotated in the peach genome, four of them forming a cluster containing gene *Prupe6G024900*. A blastx analysis showed that our gene is similar to proteins of other species annotated as cellulose synthase-like protein E1.

Cellulose is an important component of the pollen cell wall intine (Wu et al. 2015) and male sterile cultivars with defects in pollen grain cell walls have been described in several species including cotton (Wu et al. 2015), rice (Xu et al. 2017) and sorghum (Petti et al. 2015), so we selected this gene as a possible candidate for the pollen sterile phenotype. We cannot discard the possibility that PS phenotype is a case of cytoplasmic male sterility (CMS). In the 27 candidate genes listed in Table 2, there are two genes (*Prupe6G024500* and *Prupe6G026400*) annotated as pentatricopeptide repeat proteins (PPR), which are usually restorer genes of CMS (Gaborieau et al. 2016).

Genotyping with InDel024900 and comparison with CPP21395

A marker based on the variant of *Prupe6G024900* was developed (InDel024900) and used to genotype 259 individuals from the CASC (Supplementary Table 1S). This marker was diallelic and the individuals were classified as A (homozygous for the insertion), B (homozygous for the alternative allele) and H (heterozygous), where we expected A to be associated with pollen sterility. For 92 of the individuals studied we had the genotype of InDel024900 and the PS phenotype and the test was as expected in 86 of them (93% success rate). All six erroneous cases were fertile individuals (with the A genotype). Based on the information from the CASC we obtained an approximate value for the recombination frequencies between the three loci (CPP21395, InDel024900 and *Ps*) of 0.12 for CPP21395-InDel024900, 0.05 for CPP21395-*Ps* and 0.06 for InDel024900-*Ps*, indicating that *Ps* is flanked by the two markers at similar genetic distance. The recombination fractions were higher than those estimated with conventional segregating progenies as they are produced with a collection of individuals where more generations have likely occurred during their history. From these data, the sum of the recombination fractions of the two marker-*Ps* combinations (0.11) was smaller than that of the two markers (0.12).

Discussion

The *Ps* locus was mapped following a conventional mapping strategy using SSR and indel markers distributed along peach chromosome 6, reducing its region to approximately 333 kb, between markers Indel1860 and CPP21230 (Ps06:1.859.488 to Ps06:2.192.785). Then, applying a BSA mapping strategy, we used resequencing data from six cultivars with known pollen fertility phenotype, which independently identified a region of 159 kb (Pp06:1.956.981 to Pp06: 2.116.368) within the region detected using the linkage map. This demonstrates the validity of the BSA approach with DNA resequencing data, identifying a smaller genome fragment than that obtained with linkage mapping. BSA has previously been assayed in peach by Hollender et al. (2018), although we used a much smaller number of individuals, six vs. 74, and obtained a higher resolution (159 kb vs. 2 Mb). Our success with BSA was in part determined by the separate analysis of three genotypic classes, *PsPs*, *Psps* and *psps*, whereas Hollender et al. (2018) used only two classes. Another important aspect is the probable single origin of the pollen sterility mutation in the materials studied, as individuals carrying male sterility alleles from independent mutations would have likely complicated the analysis.

We selected two markers for further analysis, the CPP21395 SSR, at approximately 3 cM from the *Ps* locus based on the linkage maps, and InDel024900, in the region detected by the BSA analysis that produced a single base insertion with a predicted major effect *Prupe.6G024900*, a cellulose synthase E1 that we considered a candidate gene for the PS phenotype. Genotyping with these markers in a large collection of peach cultivars and breeding materials of widely varying origin determined that CPP21395 and InDel024900 flanked the *Ps* locus. The *Prupe.6G024900* gene was discounted as causal for *Ps* as several recombination events were found between *Ps* and InDel024900, leaving the PPR genes in this region as probable candidates for the PS phenotype, which should be confirmed in further experiments. Recently, genome-wide association analysis (Li et al. 2019) in a collection of peach materials detected a strong association between the PS phenotype and a SNP near the region defined by CPP21395 and InDel024900 (Pp06:1.996.812), but outside of it (Pp06:2.014.933), suggesting that this SNP would not improve our prediction of male sterility. Both CPP21395 and InDel024900 could be used as efficient tools to identify pollen sterile individuals in breeding populations, with 96% and 93% accuracy, respectively. Combining the information of both markers, the phenotype was correctly predicted in the CASC collection of cultivars and breeding accessions, except for nine individuals out of 92 (10%) that, based on the two markers, were recombinants with unpredictable phenotype. The fact that recombination fractions of the two marker-*Ps* combinations (0.11) was smaller than that of the two markers (0.12) could be because for nine recombinant individuals between the two markers having the pollen fertile phenotype (corresponding to the *Ps*-genotype) it was not possible to establish the position of the recombination breakpoint(s) within the CPP21395-*Ps*-InDel024900 fragment, and we counted only the recombinations between the two markers. Overall, both markers had a similar resolution for marker-assisted selection.

The *ps* allele was very frequent in the collection of materials studied (0.23). Assuming a distribution of genotypic frequencies similar to a Hardy-Weinberg equilibrium, a reasonable hypothesis in peach materials from modern breeding programs (Aranzana et al. 2003; Li et al. 2013) as most of the genotypes examined here, one would expect a 5 % pollen sterile plants, compared to the 9 % (6 of 64) that we had only in the cultivars with phenotypes obtained in this project. While these results are not significantly different (Chi-squared = 2.11; n.s.), they give an idea of the magnitude of the problem, where having 10% of the progenies being pollen sterile is not unusual.

There is great interest in controlling the generation of PS individuals, and the best option in a breeding program is to know the parental genotypes and either avoid crosses between parents carrying the *ps* allele or, if these crosses are performed, eliminate the *psps* individuals with marker assisted selection using tightly linked markers, such as those developed in this paper. Alternatively, PS individuals have to be identified by their phenotype in the field. Using the two markers, here, in 167 cultivars and advanced lines without a known phenotype, we were able to predict, with high probability, that three were male sterile, 155 male fertile (92 *PsPs*, 39 *Psps*, and 24 *PsPs* or *Psps*) and nine were recombinants with uncertain PS phenotype (Table 1S). Additionally, for the 66 accessions with male-fertile phenotype but lacking information on their genotype, 42 were inferred to be *PsPs*, 14 *Psps* and 10 were either *Psps* or *PsPs*. This information is very helpful for breeders to plan their crosses more efficiently and to be able to apply MAS for this trait in breeding programs.

The high frequency of the *ps* allele in breeding materials comes from the history of this character in the breeding programs of western countries, where most of the accessions studied here come from. Modern breeding in peach started in the US about a century ago from a limited number of founders that are the basis of the breeding gene pool used today in Europe and the US. About a dozen of cultivars were recurrently used in these programs (Hesse 1975; Scorza et al. 1985), some with known or suspected presence of the *ps* allele. Three of the most important founders are 'JH Hale', a male sterile cultivar resulting from self-pollination of another founder, the fertile 'Elberta', and 'Chinese Cling', the seed-parent of 'Elberta', also male sterile (Connors 1927; Werner and Creller, 1997; Okie 1998; Hesse 1975). Another major founder, 'Belle of Georgia', is also progeny of 'Chinese Cling', and others, such as 'Fay Elberta' and 'Early Elberta', were developed from seeds of 'Elberta' and are likely to carry the *ps* allele. The pedigrees of most of these varieties have been validated with molecular markers (Aranzana et al. 2010). These founders determined a high frequency of *ps* in the initial steps that, in spite of the selection against male sterile individuals, has seen no major decrease in the materials currently used. It is possible that this character could only be selected against in homozygosis, and the number of generations from the founders to the modern varieties is also low (3-6) with limited opportunity for selection. In addition, we cannot discard favorable effects of the *ps* allele or alleles of neighboring genes on other aspects of peach production and fruit quality that may have been unwittingly selected for by breeders. The information provided in this paper would be helpful to design experiments to determine the possible value of the *ps* haplotype or to decrease its frequency in breeding materials to avoid unwanted selection of male sterile trees.

References

Adami M, Franceschi P, Brandi F, Liverani A, Giovannini D, Rosati C, Dondini L, Tartarini S. 2013. Identifying a carotenoid cleavage dioxygenase (*ccd4*) gene controlling yellow/white fruit flesh color of peach. Plant Molecular Biology Reporter 31:1166-1175.

Aranzana MJ, Carbó J, Arús P (2003) Microsatellite variability in peach [*Prunus persica* (L) Batsch.]: cultivar identification, marker mutation, pedigree inferences and population structure. Theor Appl Genet 106:1341-1352.

Aranzana MJ, El Kadri A., Howad W, Arús P (2010) Genetic variation, population structure and linkage disequilibrium in peach commercial varieties. BMC Genetics 11:69.

Aranzana MJ, Decroocq V, Dirlewanger E, Eduardo I, Gao ZS, Gasic K, Iezzoni A, Jung S, Peace C, Prieto H, Tao R, Verde I, Abbott AG, Arús P (2019) *Prunus* genetics and applications after de novo genome sequencing: achievements and prospects Hortic Res 6:58.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114-2120.

Bretó MP, Cantín CM, Iglesias I, Arús P, Eduardo I (2017) Mapping a major gene for red skin color suppression (highlighter) in peach. Euphytica 213:14.

Cantin CM, Arús P, Eduardo I (2018) Identification of a new allele of the *Dw* gene causing brachytic dwarfing in peach. BMC Res Notes 11:386.

Chaparro JX, Werner DJ, O'Malley D, Sederoff RR (1994) Targeted mapping and linkage analysis of morphological isozyme, and RAPD markers in peach. Theor Appl Genet 87:805-815.

Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118, iso-2 and iso-3. Fly 6(2):80-92.

Cirilli M, Giovannini D, Ciacciulli A, Chiozzotto R, Gattolin S, Rossini L, Liverani A, Bassi D (2018) Integrative genomics approaches validate PpYUC11-like as candidate gene for the stony hard trait in peach (*P. persica* L. Batsch). BMC Plant Biology, 18:88.

Connors CH (1927) Pollen sterility in peaches. Science 66(1710):332

Danecek P, Schiffels S, Durbin R. https://samtools.github.io/bcftools/call-m.pdf.

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.

Dirlewanger E, Cosson P, Boudehri K, Renaud C, Capdeville G, Tauzin Y, Laigret F, Moing A (2006) Development of a second-generation genetic linkage map for peach [*Prunus persica* (L.) Batsch] and characterization of morphological traits affecting flower and fruit. Tree Genet Genomes 3:1-13.

Donoso JM, Picañol R, Eduardo I, BatlIe I, Howad W, Aranzana MJ, Arús P (2015) High-density mapping suggests a cytoplasmic male sterility system with two restorer factors in almond x peach progenies. Hort Res 2, 15016.

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13-15.

Eduardo I, López-Girona E, BatlIe I, Reig G, Iglesias I, Howad W, Arús P, Aranzana MJ (2014) Development of diagnostic markers for selection of the subacid trait in peach. Tree Genet Genomes 10:1695-1709.

Eduardo I, Picañol R, Rojas E, Batlle I, Howad W, Aranzana MJ, Arús P (2015) Mapping of a major gene for the slow ripening character in peach: co-location with the maturity date gene and development of a candidate gene-based diagnostic marker for its selection. Euphytica 205:627-636.

Falchi R, Vendramin E, Zanon L, Scalabrin S, Cipriani G, Verde I, Vizzotto G, Morgante M (2013) Three distinct mutational mechanisms acting on a single gene underpin the origin of yellow flesh in peach. Plant Journal 76:175-187.

Gaborieau L, Brown GG, Mireau H (2016) The Propensity of Pentatricopeptide Repeat Genes to Evolve into Restorers of Cytoplasmic Male Sterility. Front. Plant Sci. 7:1816.

Gu C, Wang L, Wang W, Zhou H, Ma BQ, Zheng HY, Fang T, Ogutu C, Vimolmangkang S, Han YP (2016) Copy number variation of a gene cluster encoding endopolygalacturonase mediates flesh texture and stone adhesion in peach. J Exp Bot 67:1993-2005.

Hesse C.O. (1975) Peaches. Advances in Fruit Breeding:285-335.

Hollender CA, Hadiarto T, Srinivasan C, Scorza R, Dardick C (2015) A brachytic dwarfism trait (dw) in peach trees is caused by a nonsense mutation within the gibberellic acid receptor PpeGID1c. New Phytol. 210(1):227-39.

Hollender CA, Pascal T, Tabb A, Hadiarto T, Srinivasan C, Wang W, Liu Z, Scorza R, Dardick C (2018) Loss of a highly conserved sterile alpha motif domain gene (WEEP) results in pendulous branch growth in peach trees. Proc Natl Acad Sci USA 115:20.

Jun JH, Chung KH, Jeong SB, Lee HJ (2004) An RAPD marker linked to the pollen sterility gene *ps* in peach (*Prunus persica*). The Journal of Horticultural Science and Biotechnology 79(4):587-590.

Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27(21):2987-93.

Li XW, Meng XQ, Jia HJ, Yu ML, Ma RJ, Wang LR, Cao K, Shen ZJ, Niu L, Tian JB, Chen MJ, Xie M, Arús P, Gao ZS, Aranzana MJ (2013) Peach genetic resources: diversity, population structure and linkage disequilibrium. BMC Genetics 14:84.

Li Y, Cao K, Zhu G, Fang W, Chen C, Wang X, Zhao P, Guo J, Ding T, Guan L, Zhang Q, Guo W, Fei Z, Wang L (2019) Genomic analyses of an extensive collection of wild and cultivated accessions provide new insights into peach breeding history. Genome Biology 20:36.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 20:1297-303.

Okie WR (1998) Handbook of Peach and Nectarine Varieties: Performance in the Southeastern United States and Index of Names. U.S. Department of Agriculture, Agriculture Handbook No 714, 814 pp.

Pan L, Zeng WF, Niu L, Lu ZH, Liu H, Cui GC, Zhu YQ, Chu JF, Li WP, Fang WF, Cai ZG, Li GH, Wang ZQ (2015) PpYUC11, a strong candidate gene for the stony hard phenotype in peach (*Prunus persica* L. Batsch), participates in IAA biosynthesis during fruit ripening. J Exp Bot 66:7031–7044.

Paran I & Michelmore RW (1993) Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. Theor. Appl. Genet. 85:985-993.

Peace CP, Crisosto CH, Gradziel TM (2005) Endopolygalacturonase: a candidate gene for Freestone and Melting flesh in peach. Mol Breeding 16:21-31.

Petti C, Hirano K, Stork J, DeBolt S (2015) Mapping of a Cellulose-Deficient Mutant Named dwarf1-1 in Sorghum bicolor to the Green Revolution Gene gibberellin20-oxidase Reveals a Positive Regulatory Association between Gibberellin and Cellulose Biosynthesis. Plant Physiology 169(1):705-16.

Picañol R, Eduardo I, Aranzana MJ, Howad W, Batlle I, Iglesias I, Alonso JM, Arús P (2012) Combining linkage and association mapping to search for markers linked to the flat fruit character in peach. Euphytica 190:279–288

Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SRF, WGS500 Consortium, Wilkie AOM, McVean G, Lunter G (2014) Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nature Genetics 46: 912–918.

Sandefur P, Frett T, Clark J, Gasic K, Peace C (2017) A DNA test for routine prediction in breeding of peach blush, Ppe-Rf-SSR. Mol Breeding 37: 11.

Scorza R, Mehlenbacher SA, Lightner GW (1985) Inbreeding and coancestry of freestone peach cultivars of the eastern United States and implications for peach germplasm improvement. J Amer Soc Hort Sci 110(4):547-552.

Scott DH, Weinberger JH (1944) Inheritance of pollen sterility in some peach varieties. Proc Am Soc Hortic Sci USA 45:229–232.

Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in Bioinformatics 14(2): 178–192.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3-new capabilities and interfaces. Nucleic Acids Research 40(15):e115.

Van Ooijen, J.W. (2011). Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. Genetics Research, 93(5):343-349.

Vendramin E, Pea G, Dondini L, Pacheco I, Dettori MT, Gazza L, Scalabrin S, Strozzi F, Tartarini S, Bassi D, Verde I, Rossini L (2014) A unique mutation in a MYB gene cosegregates with the nectarine phenotype in peach. PLoS One 9(3):e90574.

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93(1):77-8.

Werner D.J., Creller M.A. (1997) Genetic Studies in peach: inheritance of sweet kernel and male sterility. J Amer Soc Hort Sci 122(2):215-217.

Wu Y, Min L, Wu Z, Yang L, Zhu L, Yang X, Yuan D, Guo X, Zhang X (2015) Defective pollen wall contributes to male sterility in the male sterile line 1355A of cotton. Scientific reports 5: 9608.

Xu D, Shi J, Rautengarten C, Yang L, Qian X, Uzair M, Zhu L, Luo Q, An G, Waßmann F, Schreiber L, Heazlewood JL, Scheller HV, Hu J, Zhang D, Liang W (2017) *Defective Pollen Wall 2* (*DPW2*) Encodes an Acyl Transferase Required for Rice Pollen Development. Plant Physiology 173(1): 240-255.

Fig 1 Peach flowers with male sterile (left) and fertile (right) anthers

Fig 2 Linkage mapping of *Ps* in the G6 of F_2 populations PN732 and PN788

Table 1 Summary of the CPP21395 genotypes in the cultivar and advanced selections collection studied with capillary electrophoresis

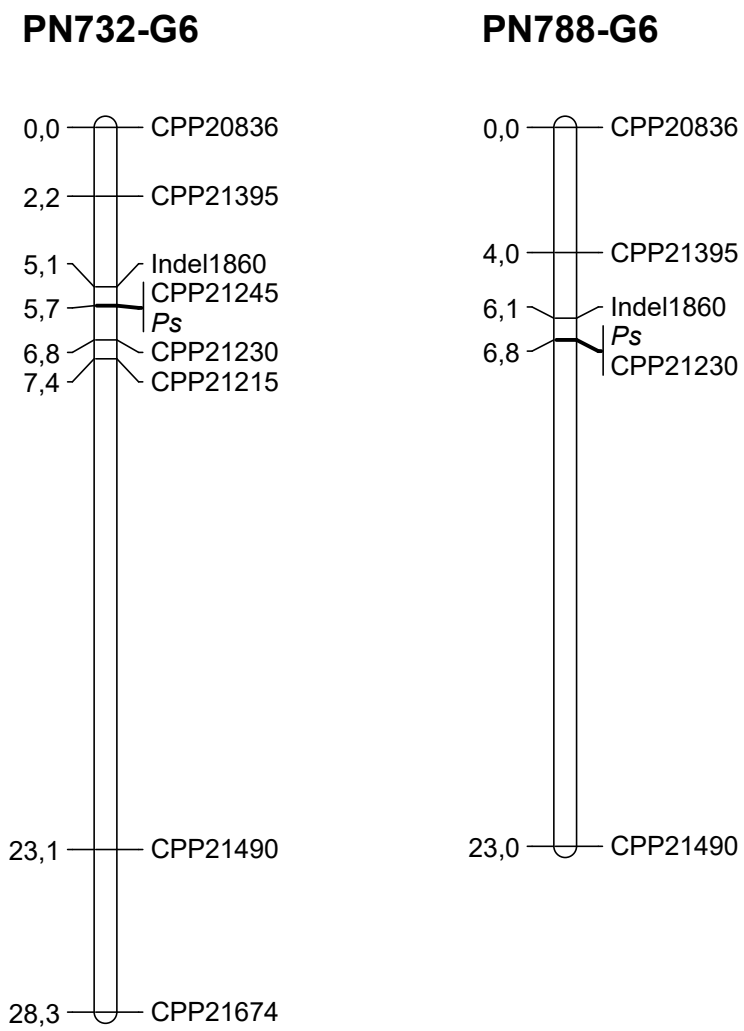
Genotype	Nb. individuals	Pollen sterile	Pollen fertile	Unknown
200/-	96	-	23	73
198/200	72	-	21	51
198/-	49	-	16	33
196/200	1	-	-	1
196/198	1	1	-	-
194/198	1	-	-	1
187/200	64	1	27	36
187/190	1	-	-	1
187/198	45	-	17	28
187/-	29	15	4	10
Total	**359**	**17**	**108**	**234**

Table 2 Genes annotated in the peach genome v2.0, in the region where *ps* was located (Pp06:1.956.981 to Pp06:2.116.368)

Name	Location	BLAST
Prupe.6G024400	Pp06:1958444..1974196	AT1G55860.2 ubiquitin-protein ligase 1
Prupe.6G024500	Pp06:1974865..1979965	AT5G55840.1 Pentatricopeptide repeat (PPR) superfamily protein
Prupe.6G024600	Pp06:1980829..1984139	AT5G55850.3 RPM1-interacting protein 4 (RIN4) family protein
Prupe.6G024700	Pp06:1984407..1986778	AT3G07680.1 emp24/gp25L/p24 family/GOLD family protein
Prupe.6G024800	Pp06:1988969..1992904	AT1G55850.1 cellulose synthase like E1
Prupe.6G024900	Pp06:1993594..1997839	AT1G55850.1 cellulose synthase like E1
Prupe.6G025000	Pp06:2000316..2004155	AT1G55850.1 cellulose synthase like E1
Prupe.6G025100	Pp06:2004661..2008198	AT1G55850.1 cellulose synthase like E1
Prupe.6G025200	Pp06:2015242..2018105	AT1G67500.2 recovery protein 3
Prupe.6G025300	Pp06:2020454..2025425	AT5G55860.1 Plant protein of unknown function (DUF827)
Prupe.6G025400	Pp06:2025597..2029476	AT4G26620.1 Sucrase/ferredoxin-like family protein
Prupe.6G025500	Pp06:2029993..2034716	AT1G55840.1 Sec14p-like phosphatidylinositol transfer family protein
Prupe.6G025600	Pp06:2037432..2041407	AT5G47750.1 D6 protein kinase like 2
Prupe.6G025700	Pp06:2042999..2045381	AT5G27990.1 Pre-rRNA-processing protein TSR2, conserved region
Prupe.6G025800	Pp06:2045776..2050416	AT1G60620.1 RNA polymerase I subunit 43
Prupe.6G025900	Pp06:2054131..2054415	hypothetical protein
Prupe.6G026000	Pp06:2064771..2065342	AT4G26590.1 oligopeptide transporter 5
Prupe.6G026100	Pp06:2069708..2073473	AT4G26590.1 oligopeptide transporter 5
Prupe.6G026200	Pp06:2079104..2083027	AT4G26590.1 oligopeptide transporter 5
Prupe.6G026300	Pp06:2083270..2087321	AT5G55930.1 oligopeptide transporter 1
Prupe.6G026400	Pp06:2087469..2089362	AT5G48910.1 Pentatricopeptide repeat (PPR) superfamily protein
Prupe.6G026500	Pp06:2089883..2092148	AT5G55940.1 Uncharacterised protein family (UPF0172)
Prupe.6G026600	Pp06:2092736..2096757	AT3G13224.2 RNA-binding (RRM/RBD/RNP motifs) family protein
Prupe.6G026700	Pp06:2100008..2107565	AT3G07660.1 Kinase-related protein of unknown function (DUF1296)
Prupe.6G026800	Pp06:2106907..2110035	AT5G55950.1 Nucleotide/sugar transporter family protein
Prupe.6G026900	Pp06:2111266..2114599	AT1G55830.2 unknown protein

Figure 1

Figure 2


PN732-G6
0,0
CPP20836
2,2
CPP21395
5,1
Indel1860
5,7
CPP21245
Ps
6,8
CPP21230
7,4
CPP21215
23,1
CPP21490
28,3
CPP21674
PN788-G6
0,0
CPP20836
4,0
CPP21395
6,1
Indel1860
Ps
6,8
CPP21230
23,0
CPP21490