

A Unique Mutation in a MYB Gene Cosegregates with the Nectarine Phenotype in Peach

Elisa Vendramin^{1,9}, Giorgio Pea^{2,9}, Luca Dondini³, Igor Pacheco⁴, Maria Teresa Dettori¹, Laura Gazza¹, Simone Scalabrin⁵, Francesco Strozzi², Stefano Tartarini³, Daniele Bassi⁴, Ignazio Verde^{1*}, Laura Rossini^{2,4*}

1 Consiglio per la Ricerca e la Sperimentazione in Agricoltura – Centro di Ricerca per la Frutticoltura (CRA-FRU), Rome, Italy, **2** Parco Tecnologico Padano, Via Einstein, Loc. C.na Codazza, Lodi, Italy, **3** Università di Bologna, DipSA, Bologna, Italy, **4** Università degli Studi di Milano, DiSAA, Milan, Italy, **5** IGA Technology Services, Udine, Italy

Abstract

Nectarines play a key role in peach industry; the fuzzless skin has implications for consumer acceptance. The peach/nectarine (*G/g*) trait was described as monogenic and previously mapped on chromosome 5. Here, the position of the *G* locus was delimited within a 1.1 cM interval (635 kb) based on linkage analysis of an F_2 progeny from the cross 'Contender' (C, peach) x 'Ambra' (A, nectarine). Careful inspection of the genes annotated in the corresponding genomic sequence (Peach v1.0), coupled with variant discovery, led to the identification of MYB gene *PpeMYB25* as a candidate for trichome formation on fruit skin. Analysis of genomic re-sequencing data from five peach/nectarine accessions pointed to the insertion of a LTR retroelement in exon 3 of the *PpeMYB25* gene as the cause of the recessive glabrous phenotype. A functional marker (indelG) developed on the LTR insertion cosegregated with the trait in the CxA F_2 progeny and was validated on a broad panel of genotypes, including all known putative donors of the nectarine trait. This marker was shown to efficiently discriminate between peach and nectarine plants, indicating that a unique mutational event gave rise to the nectarine trait and providing a useful diagnostic tool for early seedling selection in peach breeding programs.

Citation: Vendramin E, Pea G, Dondini L, Pacheco I, Dettori MT, et al. (2014) A Unique Mutation in a MYB Gene Cosegregates with the Nectarine Phenotype in Peach. PLoS ONE 9(3): e90574. doi:10.1371/journal.pone.0090574

Editor: Cameron Peace, Washington State University, United States of America

Received: November 25, 2013; **Accepted:** February 1, 2014; **Published:** March 3, 2014

Copyright: © 2014 Vendramin et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the Ministero delle Politiche Agricole Alimentari e Forestali–Italy (MiPAAF www.politicheagricole.it) through the project 'DRUPOMICS' (grant DM14999/7303/08) and by an Italian grant to DB funded by private and public agencies "MAS.PES: apricot and peach breeding by molecular-assisted selection". The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Dr. Simone Scalabrin, one of the authors of the manuscript, is currently affiliated with IGA Technology Services. This does not alter the authors' adherence to all PLOS ONE policies on sharing data and materials.

* E-mail: ignazio.verde@entecra.it (IV); laura.rossini@unimi.it (LR)

These authors contributed equally to this work.

Introduction

Peach (*Prunus persica* L. Batsch) is one of the most important fruit crops in temperate regions with about 21 million tons produced worldwide and Italy, with over 1.6 million tons, is the second producer after China (FAOSTAT 2011, <http://faostat.fao.org/>). The presence or absence of skin pubescence (fuzziness) is one of the commercial characteristics used to classify peach fruits along with flesh features (color, adhesion and texture) and fruit shape. Nectarines, characterized by the absence of fruit trichomes, are widely cultivated and play an important role in world peach production (30% in Italy, <http://agri.istat.it/>, 2013; 20% in USA, <http://www.nass.usda.gov/>, 2013) and may be associated with decreased allergenic properties. In *P. persica* two major allergens, Pru p 1 and Pru p 3, are known as responsible for the oral allergy syndrome (OAS) [1]. Indeed, the Pru p 3 protein was undetectable in the nectarine 'Rita Star' suggesting that this may be considered as a hypoallergenic variety [2]. Interestingly, in *Humulus lupulus* and in *Nicotiana tabacum*, genes encoding proteins highly similar to Pru p 1 and Pru p 3 are mainly expressed in trichomes [3,4].

The peach/nectarine character is monogenic (*G/g*) with nectarine recessive to fuzzy fruit [5]. The *G* locus was mapped in the distal part of linkage group (LG) 5 [6,7] spanning a region

from 15,126,681 to 16,315,341 (1.189 Mb) of pseudomolecule 5 of the peach reference genome (Peach v1.0) [8]. Peach originated in North-West China and was domesticated there about 4,000–5,000 years ago [8]. From China it spread westwards reaching Persia following the Silk Road, was introduced to Rome in the first century BC and then disseminated to all the Roman Empire [9]. Nectarines have been known in China for over 2,000 years [10] and have been reported in most of the oases of the Tarim Basin (China) and along the Silk Road trade routes in Central Asia and the Caucasus [9,11]. The means and timing of their introduction in Europe are not clear. Likely, Romans did not know this type of peaches [9], but nectarines have been described by several botanists in Europe since the Renaissance period [9]. Old European nectarine varieties include 'Lord Napier', 'Precoce di Croncels' and 'Galopin'. In Southern Italy, traditional local white nectarines, called 'Sbergie' (Sicily) or 'Merendelle' (Calabria), have been cultivated since the 16th century [12]. Cluster analysis suggested that these local accessions are distinct from the western nectarine germplasm, pointing to a putative different origin of this group of cultivars [13]. Historically, nectarines have had little impact in China's peach industry [11], and nowadays there are no reports of traditional nectarine cultivars available in China [14]. The timing of introduction of nectarines to the United States (US)

is controversial: their cultivation is reported in the early 20th century, although a newspaper article (New York Gazette March 28, 1768, p. 3) described nectarines being grown in the US prior to the War of Independence. Modern nectarine breeding started in the US in the middle of the 20th century. In 1942, Anderson introduced the nectarine ‘Le Grand’ using the accession ‘Quetta’, discovered near the homonymous city in India (now part of Pakistan) in 1906 [15], as the source of the nectarine trait. Other known sources of the nectarine trait used in modern western breeding programs were ‘Goldmine’ and ‘Lippiatt’ discovered in New Zealand in 1900 and 1916, respectively [15]. These latter three genotypes are acknowledged as donors of most of the current nectarine cultivars widespread in US and Europe. Modern Japanese breeding programs have extensively used two old European nectarines, ‘Precoce di Croncels’ and ‘Lord Napier’, and modern US cultivars [16]. In the last decades, the trait was introduced to Chinese breeding programs directly from western accessions or indirectly using Japanese material [17].

Trichomes are hair-like appendages that derive from the differentiation of epidermal cells and are classified based on their morphology (unicellular or multicellular), and secretory abilities (glandular or non-glandular) [18,19]. Trichomes may develop on several plant organs (leaf, fruit, seed, etc.). They play an important role in protecting plants against biotic and abiotic stresses [20–22] and can also hold a direct economic relevance. Aromatic substances are often synthesized by glandular trichomes, for example in aromatic plants, such as peppermint (*Mentha piperita*) [23] and basil (*Ocimum basilicum*) [24]. Cotton (*Gossypium hirsutum*) seed fibers are classified as non-glandular trichomes and represent one of the most highly expanded plant cell types [25]. In peach fruit, trichomes are non-glandular and unicellular and first appear on the ovary as early as four weeks before anthesis as observed in the peach ‘Contender’ [26]. By the time of physiological ripening most fruit skin trichomes are dead cells [27]. In *Arabidopsis* a number of genes involved in trichome formation and development have been identified by mutant analyses [28] and transcriptome profiling [29], revealing a complex regulatory network. Several transcription factors interact during trichome initiation and formation: in particular members of the R2R3-MYB class are known to act as positive regulators [30–33], while single-repeat MYB proteins function in negative control [34–36]. Mutations in the R2R3-MYB gene *GLABRA1* (*GL1*) result in glabrous plants in *A. thaliana* [37], *A. lyrata* [37,38] and other Brassicaceae species [39]. In *Gossypium hirsutum*, *GhMYB25*, which encodes an R2R3-MYB factor, is involved in the differentiation of ovule epidermal cells into cotton fibers, as well as in the formation of leaf trichomes [40].

The aims of the present study were to precisely map the *G* locus, identify a candidate gene and develop a reliable marker for the nectarine phenotype (glabrous fruit). To these ends, we used an F₂ population from a cross between the peach ‘Contender’ (C) and the nectarine ‘Ambra’ (A) [41,42] to develop a Single Nucleotide Polymorphism (SNP) map around the *G* locus. Analysis of the corresponding region in the peach genome sequence (Peach v1.0) [8] led to the identification of an R2R3-MYB gene as a candidate for trichome formation in peach fruit. A functional marker (indelG) developed on this gene provides a useful tool for early seedling selection for the peach/nectarine trait in breeding programs.

Materials and Methods

Plant materials & DNA extraction

An F₂ population of 305 seedlings derived from the cross between the peach ‘Contender’ (C) and the nectarine ‘Ambra’ (A),

segregating for the peach/nectarine trait, [41,42] (CxA F₂), was used to develop a SNP map around the *G* locus. The trees were located in a farm belonging to the Municipality of Castel San Pietro (Bologna, Emilia Romagna, Italy) leased to ASTRA (latitude: from 44°24′44.18″N to 44°24′30.08″N; longitude: from 11°35′47.21″E, to: 11°36′2.00″E). No specific permission was required because Daniele Bassi is the curator of the peach germplasm collection grown there and no endangered or protected species were involved.

Trees were planted on their own roots with a spacing of 1 m within and 4 m between rows and trained as slender spindle (one stem with short lateral scaffolds). Pruning was performed yearly and standard cultural practices were applied. Scoring of the peach/nectarine trait was carried out in two seasons to confirm correct scoring of the phenotype.

Ninety-five *P. persica* genotypes, 46 peaches and 49 nectarines grown at the CRA-FRU experimental farm (Rome, Italy) (except for ‘Galopin’ and ‘Lord Napier’ grown at Ivalsa-CNR, Follonica, Italy), were analyzed to validate the functional marker. For each accession, phenotype, pedigree, geographical origin and putative donor of the nectarine trait are reported in Table 1. DNA was extracted from leaf tissue using the DNeasy Plant Mini Kit (Qiagen GmbH, Hilden, Germany) as per manufacturer’s protocol and quantified with NanoDrop (Thermo Scientific, Waltham, MA, USA).

Linkage map

In total 305 individuals from the CxA F₂ progeny were analyzed to map the *G* locus using an upgraded version of the CxA map [41,42] covering LG 5 from base 10,192,138 to base 17,544,073 of the peach genome sequence pseudomolecule 5 and spanning the already known *G* interval [7]. To refine the position of the *G* locus, SNPs located in this region were selected from those identified through analysis of re-sequencing data from the CxA F₁ individual [42] (biosample SRS335631, run SRR502997). About 300 bp of the SNP-flanking sequence were downloaded from the IGA peach Gbrowse (<http://www.appliedgenomics.org/>) and the Mass ARRAY Assay Design 3.1 software was used to design multiplex assays for SNP analysis [43]. SNP genotyping was performed using the iPLEX Gold technology available for Sequenom platforms (Sequenom, Inc., San Diego, CA, USA). SNP markers and the *G* locus (scored as a dominant phenotypic marker) were mapped using Joinmap 3.0 [44] with a minimum LOD score of 10 for grouping; the Kosambi mapping function [45] was used to convert recombination frequencies into map distances. Based on this map, three new SNP markers (S5_15988499, S5_15865556 and S5_15866258) were developed in the *G* locus interval (scaffold_5, from 15,853,006 to 16,488,104; see Results and Discussion) and genotyped on eight informative recombinants by sequencing 200 bp encompassing the SNP (primer sequences shown in Table 2). Standard PCRs were performed using GoTaq Green Master Mix (Promega, Madison, WI, USA). Each PCR reaction contained 1 X GoTaq Green Master Mix, 0.4 μM of each primer, 20 ng template DNA and sterile Milli-Q water to a final volume of 25 μl. The PCR protocol consisted in an initial step at 95°C (5 min), followed by 40 cycles at 95°C (30 s), 60°C (30 s) and 72°C (1 min), and a final elongation at 72°C (5 min). PCR products were purified with ExoSapIT (Amersham PharmaciaBiotech, Uppsala, Sweden) and sequenced with the Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA). After ethanol precipitation, sequencing products were mixed with 15 μl of HiDi formamide and subjected to capillary electrophoresis in an ABI Prism 3730 DNA Analyzer (Applied Biosystems, Foster City, CA, USA). Genotyping was performed by

Table 1. List of 95 peach/nectarine varieties representing the diversity of cultivated *P. persica*.

| Accession name | Genotype | Phenotype | Pedigree | Geographical origin ^a | Putative donor of nectarine allele ^b |
|--------------------------|----------|-----------|--|----------------------------------|---|
| Ambra | 197-197 | nectarine | Stark Red Gold x Mayfire | Italy (B) | Lippiatt/Goldmine |
| Aniversario | 197-197 | nectarine | – | Argentina (B) | -/- |
| Big Top | 197-197 | nectarine | – | USA (B) | -/- |
| Branca | 197-197 | nectarine | Goldmine x Pala | Brazil (B) | Goldmine/Goldmine |
| California | 197-197 | nectarine | P60-30 x Fantasia | Italy, 1994 (B) | Quetta/- |
| Centenaria | 197-197 | nectarine | Docura 2 op | Brazil (B) | Goldmine/Goldmine |
| Chiyodared | 197-197 | nectarine | Hirakutsared x Nectared 5 | Japan (B) | Lord Napier or Precoce di Croncels/- |
| Crasimmolo Rosso | 197-197 | nectarine | – | Italy (L) | -/- |
| Crimson Gold | 197-197 | nectarine | Nectarine selection x July Elberta | USA, 1967 (B) | -/- |
| Sabrina | 197-197 | nectarine | – | Spain (B) | -/- |
| Fairlane | 197-197 | nectarine | (Le Grand x Sun Grand) x Fantasia | USA, 1973 (B) | Lippiatt/Quetta |
| Fantasia | 197-197 | nectarine | Gold king x p101-24 (=Red King op) | USA, 1969 (B) | Quetta/Quetta |
| Firebrite | 197-197 | nectarine | Flavortop x (Red King F ₂) | USA, 1974 (B) | Quetta/- |
| Fuzador | 197-197 | nectarine | ((Greensboro op x Elberta) x JHHale) F ₂ | France, 1973 (B) | -/- |
| Galopin | 197-197 | nectarine | – | England, 1869 (L) | -/- |
| Gold Mine | 197-197 | nectarine | – | New Zeland, 1900 (B) | |
| Golden Grand | 197-197 | nectarine | Le Grand x (Le Grand x Kim) | USA, 1954 (B) | Quetta/Lippiatt |
| Jacquotte | 197-197 | nectarine | – | France, 1979 (B) | -/- |
| Laura | 197-197 | nectarine | – | USA, 1995 (B) | |
| JinXia | 197-197 | nectarine | (Okubo x Okitsu) x Okitsu F ₂ | China (B) | Lord Napier/Precoce di Croncels |
| Lord Napier | 197-197 | nectarine | – | Belgium, 1859 (L) | -/- |
| Madonna Di Agosto | 197-197 | nectarine | – | Italy (L) | -/- |
| Madonna Di Giugno | 197-197 | nectarine | – | Italy (L) | -/- |
| Magali | 197-197 | nectarine | – | France, 1988 (B) | -/- |
| Max | 197-197 | nectarine | – | Italy, 1995 (B) | -/- |
| Mayfire | 197-197 | nectarine | Armking op | USA, 1984 (B) | Goldmine/Goldmine |
| Nectaross | 197-197 | nectarine | Stark Red Gold x Le Grand | Italy, 1983 (B) | Lippiatt/Quetta |
| Nettarina Pendula | 197-197 | nectarine | – | Italy (L) | -/- |
| Nico | 197-197 | nectarine | – | Italy (B) | -/- |
| Phn 91-12 | 197-197 | nectarine | – | USA (B) | -/- |
| Phn 91-14 | 197-197 | nectarine | – | USA (B) | -/- |
| Phn 91-17 | 197-197 | nectarine | – | USA (B) | -/- |
| Quetta | 197-197 | nectarine | – | Pakistan, 1906 (L) | |
| Ricci 2 | 197-197 | nectarine | Stark Red Gold x Tastyfree | Italy, 1993 (B) | Lippiatt/Quetta |
| Romamer 1 | 197-197 | nectarine | – | Romania, 1983 (B) | -/- |
| Romamer 2 | 197-197 | nectarine | City 29-245 x RR48-153 | Romania, 1983 (B) | -/- |
| Russian Nectarine 592-81 | 197-197 | nectarine | – | Ukrayne, 1980 (L) | -/- |
| Russian Nectarine 598-81 | 197-197 | nectarine | – | Ukrayne, 1980 (L) | -/- |
| Shizukured | 197-197 | nectarine | Okitsu x NJN17 | Japan (B) | Lord Napier or Precoce di Croncels/- |
| Silver Lode | 197-197 | nectarine | (Goldmine x Rio Oso Gem) x (Goldmine x July Elberta) | USA, 1951 (B) | Goldmine/Goldmine |
| Sirio | 197-197 | nectarine | Flamekist x Fantasia | Italy, 1987 (B) | Quetta/Quetta |
| Snow Queen | 197-197 | nectarine | – | USA, 1975 | -/- |
| Souvenir Nikitski | 197-197 | nectarine | Lola op | Ukrayne, 1988 (B) | -/- |
| StarkRedgold | 197-197 | nectarine | Sun Grand op | USA (B) | Lippiatt/Lippiatt |
| Summer Beauty | 197-197 | nectarine | Red Diamond x Sun Grand | USA, 1979 (B) | Quetta/Lippiatt |
| Vania | 197-197 | nectarine | – | Italy, 1990 (B) | -/- |
| ZeeGlo | 197-197 | nectarine | (Red Grand op) x (Sun Grand x Merrill Gem) | USA, 1988 (B) | Quetta/Lippiatt |
| Zephyr | 197-197 | nectarine | – | France, 1992 (B) | -/- |

Table 1. Cont.

| Accession name | Genotype | Phenotype | Pedigree | Geographical origin ^a | Putative donor of nectarine allele ^b |
|--------------------|----------|-----------|---|----------------------------------|---|
| Zincal 5 | 197-197 | nectarine | – | Spain (B) | -/- |
| Acireale | 941-941 | peach | – | Italy (L) | |
| Amber Gold | 941-941 | peach | Red Grand x Royal May | USA, 1973 (B) | |
| Amsden | 941-941 | peach | | USA, 1868 (B) | |
| Autumnglo | 941-197 | peach | (Candoka x Tennessee Natural) x Merrill Fiesta | USA, 1977 (B) | - |
| BaekmiJosaeng | 941-941 | peach | Mishima x Nunome Wase | Republic of Korea, 1983 (B) | |
| Baldagenais | 941-197 | peach | – | France | - |
| Changbang Mutant | 941-941 | peach | – | Republic of Korea, 1986 (B) | |
| Chiyomaru | 941-941 | peach | – | Japan, 1988 (B) | |
| Chui Huang Tao | 941-941 | peach | – | China (B) | |
| Ciccio Petrino | 941-941 | peach | – | Italy | |
| City 32-82 | 941-197 | peach | – | USA (B) | - |
| Contender | 941-941 | peach | Wiblo x [Norman x (Candor x (Summercrest x Redhaven))] | USA, 1987 (B) | |
| Cp 88/2 | 941-941 | peach | – | Mexico (B) | |
| Early Gold | 941-941 | peach | – | Japan, 1980 (B) | |
| Elberta | 941-941 | peach | Chinese Cling op (perhaps x Early Crawford) | USA, 1889 (B) | |
| CxA F ₁ | 941-197 | peach | Contender x Ambra | Italy (B) | Lippiatt or Goldmine |
| Fairtime | 941-197 | peach | (Rodeo x Kirkman Gem) op | USA, 1968 (B) | |
| Fayette | 941-941 | peach | Fay Elberta x (Fireglow x (Fireglow x Hiley)) | USA, 1966 (B) | |
| FeiChing Bai Li 17 | 941-941 | peach | – | China (B) | |
| FeiChing Tao | 941-941 | peach | – | China, 1909 (B) | |
| Fidelia | 941-197 | peach | (O'Henry x Giant Babcock) x (May Grand x Sam Houston) | USA, 1986 (B) | - |
| Frau Maria Rudolf | 941-941 | peach | Top Red Delicious op | Germany (B) | |
| Grosse Mignonne | 941-941 | peach | – | France, 1667 (L) | |
| Higama | 941-941 | peach | Japanese seed op | France, 1970 (B) | |
| Hwando 1 | 941-941 | peach | – | Republic of Korea, 1977 (B) | |
| J H Hale | 941-941 | peach | putative selfpollination of Elberta | USA (B) | |
| JingYu | 941-197 | peach | Okubo x Okitsu | China (B) | Lord Napier or Precoce di Croncels |
| KurakataWase | 941-941 | peach | – | Japan (B) | |
| PLove2- 2N | 941-941 | peach | di-haploids of Lovell | USA, 1882 (B) | |
| O'Henry | 941-197 | peach | Merrill Bonanza op | USA, 1970 (B) | - |
| Pantao 20-58 | 941-941 | peach | – | China, 1869 (B) | |
| Pillar | 941-941 | peach | – | Japan (B) | |
| Redhaven | 941-941 | peach | Halehaven x Kalhaven | USA, 1940 (B) | |
| Reginella II | 941-941 | peach | – | Italy (B) | |
| Rou Tao | 941-941 | peach | – | China (L) | |
| Russotto | 941-941 | peach | – | Italy (B) | |
| Sahua Hong Pantao | 941-941 | peach | – | China (L) | |
| Sanguinella | 941-941 | peach | – | Italy (L) | |
| Shen Zhou Mi Tao | 941-941 | peach | – | China (B) | |
| Siberian C | 941-941 | peach | selection of a China seedling | Canada, 1967 (B) | |
| Summer Pearl | 941-197 | peach | [(Candoka x Tennessee natural op) x ((Candoka x Flaming Gold) x NJN5 op)] x (P1119844 x White Hale) | USA, 1979 (B) | - |
| TaturaDawn | 941-941 | peach | Levis selfpollination | Australia, 1960 (B) | |
| Yoshihime | 941-197 | peach | (Nakatsu Hakuto x Nunome Wase) x Akatsuki | Japan, 1990 (B) | - |
| Yu Bai | 941-941 | peach | – | China (L) | |
| Yumyeong | 941-941 | peach | Yamato-Wase x Nunome Wase | Republic of Korea, 1977 (B) | |

Table 1. Cont.

| Accession name | Genotype | Phenotype | Pedigree | Geographical origin ^a | Putative donor of nectarine allele ^b |
|-----------------|----------|-----------|----------|----------------------------------|---|
| ZansetsuShidare | 941-941 | peach | – | Japan (B) | |

^a(B) = breeding materials (L) = landrace.

^b- = unknown donor of the nectarine allele.

For each accession the following information is reported: name, genotype at the indelG marker (941 bp reference peach allele, 197 bp nectarine allele carrying the retrotransposon insertion), phenotype, pedigree, country of origin and year of release/discovery when known, putative donor of the nectarine allele in *P.persica* if pedigree information was available.

doi:10.1371/journal.pone.0090574.t001

visual inspection of the resulting electropherograms using 4PEAKS freeware (Nucleobytes Inc.).

Variant discovery from NGS data

In order to identify genetic variants putatively involved in the control of the nectarine trait publicly available paired-end (PE) whole-genome re-sequencing data of *P. persica* accessions from study SRP013437 [8] were downloaded from the NCBI Sequence Read Archive (SRA) [46]. Five accessions were considered for this study: ‘Bolero’ (biosample SRS335629, run SRR501836), ‘OroA’ (biosample SRS335635, run SRR502986), ‘Lovell’ Clone PLov2-2N (biosample SRS335634, run SRR502985), ‘Quetta’ (biosample SRS335636, runs SRR502989 and SRR502987) and F₁ ‘Contender’ × ‘Ambra’ (biosample SRS335631, run SRR502997). ‘Quetta’ was included as the reference nectarine accession. The CxA F₁ individual originated the CxA F₂ population used to map the *G* locus. The peach ‘Lovell’ Clone PLov2-2N is the doubled haploid used to generate the reference peach genome sequence (Peach v1.0) [8], providing an internal control for false variant calling. Finally, the peaches ‘Bolero’ and ‘OroA’ were chosen as controls for nectarine segregation, as it has been demonstrated that the nectarine trait does not segregate in the ‘Bolero’ × ‘OroA’ F₁ population [42].

SRA data of each run were dumped in fastq format using the *fastq-dump* tool of NCBI sratoolkit v2.1.16 software (<http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software>), splitting forward and reverse paired reads for each sample into two separate files. Reads were quality filtered on a single sample basis using Trimmomatic v0.22 [47], first trimming leading and trailing bases below a quality threshold of 20, and then removing trimmed reads shorter than 24 bp or having an average quality below 20 (calculated on 8 bp long sliding windows). For each sample, only reads passing the quality filtering as matching pairs were retained and aligned to the whole *P. persica* reference genome Peach v1.0 using the Burrows-Wheeler Alignment Tool (BWA v0.6.2) [48]. The *aln* (IS linear-time algorithm) and *sampe* (all default options except *-n 25 -N 25*) commands were applied, respectively, for finding suffix array (SA) coordinates of each individual read and to convert them to chromosomal coordinates and to pair the reads. The resulting alignment SAM files were converted by Picard Tools version 1.77 (<http://picard.sourceforge.net/>) to sorted BAM files compliant to the Genome Analysis Toolkit (GATK) format, using the tools *CleanSam*, *SamFormatConverter* and *AddOrReplaceReadGroups*. GATK-compliant BAM files were submitted to GATK version 2.3-3 [49] for pre-processing procedures, i.e. indel realignment, duplicate removal and base quality score recalibration (BQSR). The data table needed for the recalibration step in BQSR was manually generated upon validated SNP data from the Peach 9K chip array [50]. Variant discovery procedures were then applied using whole-genome recalibrated alignments of all five samples

simultaneously. Genotypes for SNP and small INDEL variants were called through the GATK *HaplotypeCaller* tool applying hard filtering parameters [51]. Structural variants were also independently called on the same recalibrated alignment data by Pindel software v0.2.4t [52] following standard procedures.

Reads from the resequencing of ‘Quetta’ (biosample SRS335636, runs SRR502989 and SRR502987) were also analyzed using the CLC genomic workbench (6.0.1). The 75 bp pair-end fragments were trimmed for quality retaining only nucleotides with Phred values higher than 30, and trimmed reads were aligned against the *P. persica* reference genome (Peach v1.0) [8] using the read mapping tool. Only reads with over 90% identity over at least 92% of their length were mapped on the reference. All variant discovery searches were limited to the locus *G* mapping interval defined by informative recombinants in the C×A F₂ mapping population (scaffold_5, from 15,853,006 to 16,488,104, see Results and Discussion).

Validation of candidate variant

To validate the putative variant individuated among the resequenced genotypes long-range PCRs were performed on five nectarines genotypes (‘Quetta’, ‘Goldmine’, ‘Madonna di Agosto’, ‘Stark Red Gold’ and ‘Ambra’) and on the peach ‘Contender’, with primers Seq16F and Seq4R (Table 3) designed flanking the putative insertion, using Herculase DNA Polymerase (Agilent Technologies, Santa Clara, CA, USA). Each reaction contained 1x Herculase reaction buffer, 0.3 mM dNTP mix, 0.5 μM each Seq16F and Seq4R primer, 3% DMSO, 2.5 U Herculase polymerase, 300 ng template DNA, and sterile Milli-Q water to a final volume of 50 μl. The following PCR protocol was performed on a Esco Swift Maxi thermocycler (Esco GB Ltd, Downton, UK) or an Applied Biosystems 2720 Thermal Cycler (Applied Biosystems, Foster City, CA, USA): 95°C for 5 min; 28 cycles of 95°C (1 min), 59°C (1 min), 72°C (12 min) followed by a step at 72°C for additional 12 min. All PCR amplicons were checked on 1% agarose gel in an overnight run in SB buffer. A standard ethidium bromide staining was used for band visualization.

For restriction analysis, 5 μl of long-range PCR products from ‘Quetta’, ‘Goldmine’, ‘Madonna di Agosto’, ‘Stark Red Gold’ and ‘Ambra’ were digested with *EcoRI* and *HindIII* (Fermentas, Vilnius, Lithuania) in a single overnight reaction at 37°C. Master mix was calculated by double digest tool (<http://www.thermoscientificbio.com/webtools/doubledigest/>) with 0.5 U/sample of each restriction enzyme and buffer R. Restriction products were separated on 1% agarose gels and stained with ethidium bromide for band visualization.

The ‘Quetta’ long-range PCR product was first purified with the Macherey-Nagel PCR clean-up kit (Carlo Erba reagents, Italy) and quantified by Picogreen (Quant-iT PicoGreen dsDNA kit,

Table 2. SNP markers.

| Markers ^a | cM ^b | Peach v1.0 position (nt) | SNP allele ^c | Forward primer | Reverse primer | SNP Extension primer ^d |
|----------------------|-----------------|--------------------------|-------------------------|----------------------|------------------------|-----------------------------------|
| S5_10192138 | 0 | 10,192,138 | G/A | GATGAATGGGTGAAGGTAAG | TTCCGCAAAAAAAAAACATATC | ttaagGGGTGAAGGTAAGTTTCACA |
| S5_11640083 | 4.2 | 11,640,083 | T/G | CCTACTACACAATTGCCTTA | GTCGGTCGTCAGTTTTTTTG | cgacCTTGATAGATTCTAATGGAAGTA |
| S5_12847567 | 11.4 | 12,847,567 | T/A | TGGCGATTTTTCTAGCTAC | CTCTTCCCAATCTCAATCG | AATCGCATTGTTGAGAC |
| S5_13449464 | 14.3 | 13,449,464 | G/C | CGGTGATTGATATGATGATG | CCACTCAAATTGCCTTTCCC | GTGATTGATATGATGATGATT-TATAT |
| S5_13852617 | 16.7 | 13,852,617 | A/C | ATACCTATGTTCACTCCCG | TTGGCTGGTAAGGTTGTTGG | TGTAAGTGTGTAATCTAATG |
| S5_14584095 | 24.3 | 14,584,095 | G/C | AAGTTGTTCCAGTGGCAACC | TATAGTGGGGCTGGAATCTG | cccGGCTGGAATCTGTTCTCTCAGAC |
| S5_14894563 | 25.5 | 14,894,563 | C/A | GCAGAGGAATTTTCCCTAC | TTAGGGAGGGAGCTATGTTG | gTTTGAGGATTTTGGGC |
| S5_14949332 | 26.0 | 14,949,332 | T/C | AGAAGATGTGGTCCAGAGG | TTCTCGATCCGGAAGGAGAT | TCCCAGATCCAAGACCC |
| S5_15731107 | 28.4 | 15,731,107 | T/G | CTGTTGAAGACAGGTTTGG | AACATGCTTGGCGTTCGTC | GGCTTCGTCATACGCC |
| S5_15853006 | 28.8 | 15,853,006 | T/A | TAGTTGTCTGCAACCCG | CCGAGAAGACTGAAGAGTTG | GCTATTAAGACATTAGAGATGA |
| S5_15865556 | 29.1 | 15,865,556 | A/T | GGTTGGGGCTCTGATTCT | TAAAGCAACACATTGCAG | |
| S5_15866258 | 29.1 | 15,866,258 | A/G | TCAGCTTGTCCATGGCATT | GCCGTAAGGCTTTCCTCTC | |
| S5_15988499 | 29.4 | 15,988,499 | C/T | GCCGTGAAGTGGAGTTCT | GATTCTACTCTGCTCTGTCT | |
| S5_16488104 | 29.9 | 16,488,104 | A/G | TTCGATTATTAGTTCAC | CAGGTTGTGAGTTTGCTTG | TACTAAACGGAAGCTATGT |
| S5_17544073 | 36.2 | 17,544,073 | G/A | GCCATCTCTGTTTCTCTG | GTAGTATCAGCCGACTGTAG | cTAACATACATGACATGACATACA-CCC |

^aSNP detection was performed by Sequenom MassArray technology, except for the SNPs in italics that were genotyped by Sanger sequence.

^bthe genetic distances estimated by the analysis of informative recombinant plants are reported in italics.

^cthe first nucleotide correspond to the reference allele in the peach genome (Peach v1.0).

^dlower case bases correspond to the tails added to the SNP extension primer for Sequenom MassArray analysis.

For each marker the following information is reported: marker name, position in cM with respect to the map in Figure 1, position in bp with respect to the peach genome sequence (Peach v1.0), SNP allele and the primer sequences used.

doi:10.1371/journal.pone.0090574.t002

Life Technology, US) in preparation for sequencing using the Illumina MiSeq platform with a 150 bp paired end sequencing strategy. Preparation of Nextera XT library was performed with 1 ng of genomic DNA according to the Nextera XT protocol (Ver. Oct 2012, rev C). Briefly, the DNA was fragmented in 5 µl of Amplicon Tagment Mix and 10 µl of Tagment DNA buffer (Illumina, San Diego, CA, USA). Tagmentation reactions were performed by incubation at 55°C for 5 min followed by neutralization with 5 µl of Neutralize Tagment Buffer for 5 min. Tagmented DNA (25 µl) was used as the template in a 50 µl limited-cycle PCR (12 cycles) and processed as outlined in the Nextera XT protocol. Amplified DNA was purified using 90 µl of

AMPure XP beads then normalized with 45 µl of combined Library Normalization beads/additives. In preparation for cluster generation and sequencing, the normalized library was diluted in hybridization buffer and heat denatured. Due to the low diversity of the library a phiX spike-in (30%) was added to the final denatured 10 pM library. The sample was sequenced using the MiSeq Personal Sequencer (Illumina Inc., San Diego, CA, USA) running MiSeq Control Software Version 2.0.

A total of 20 M reads were obtained and analyzed using the CLC genomic workbench (6.0.1). Trimming and De Novo assembly tools were used with default parameters. The assembly obtained was filtered retaining only contigs with a length greater

Table 3. List of Primers.

| Name | Sequence 5'→3' | Start Position on Scaffold 5 (nt) | Len | Notes |
|--------------|---------------------------|-----------------------------------|-----|---|
| Seq16F | ATCCGCTCCTCATGTAGTACA | 15,898,340 | 22 | upstream the LTR insertion |
| Seq4R | CAAATAAACACCACCTACTCTGTTA | 15,899,297 | 27 | downstream the LTR insertion |
| Seq15F | GCTGTAGGCTAAGGTGGAGAA | 15,898,078 | 21 | RT-PCR primer designed on exon 2 of <i>PpeMYB25</i> |
| Seq15R | ACTGAGCAATGTGGCTGAGA | 15,898,586 | 20 | RT-PCR primer designed on exon 3 of <i>PpeMYB25</i> |
| ppaRPII_RT-F | TGAAGCATAACCTATGATGATGAAG | | 26 | RT-PCR housekeeping, [53](Tong et al, 2009) |
| ppaRPII_RT-R | CTTTGACAGCACAGTAGATTCC | | 23 | RT-PCR housekeeping, [53](Tong et al, 2009) |
| indelG-F | CTTGACCTGAGTTCGATTCCG | 15,898,324 | 22 | upstream the indel in <i>PpeMYB25</i> |
| indelG-1R | GGCTTCAATGGCAGAACAAGG | | 21 | within left border of the indel in <i>PpeMYB25</i> |
| indelG-2R | GCAGGTGGTGGAGATTTCATCAT | 15,899,264 | 23 | downstream the indel in <i>PpeMYB25</i> |

Name, sequence, position on Peach v1.0 and length of primers used to perform long-range PCR, RT-PCR and indelG assay.

doi:10.1371/journal.pone.0090574.t003

than 200 bp and formed by more than 500 reads. The filtered contigs were mapped using BWA v0.7.5a [53] with the MEM algorithm against the Peach v1.0 genome. Identification of conserved domains on contigs was performed using HMMER v3.1 [54] against the PFAM [55] database v27.0. Blast search of contigs was done using NCBI-BLAST+v2.2.27 [56] against the NT database downloaded from NCBI FTP site on Oct 23 2013.

Expression analysis

Total RNA was extracted from floral buds collected from 'Contender' and 'Ambra' at different developmental stages (seven, five, four and one week before anthesis) using the RNeasy Plant Mini Kit (Qiagen GmbH, Hilden, Germany) and treated with DNase I (Sigma-Aldrich, St. Louis, MI, USA) following manufacturers' instructions; 1 µg of RNA was reverse-transcribed using the GoScript Reverse Transcription System (Promega, Fitchburg, WI, USA) with oligo (dT)₁₅ according to the manufacturer's protocol. For reverse transcription analysis, primers were designed on exon 2 (Seq15F) and exon 3 (Seq15R) of the *PpeMYB25* gene and the RNA polymerase II sequence was used as reference gene [57] (Table 3). For RT-PCR, 1 µL of cDNA was used with 1x GoTaq Green Master Mix, 0.4 µM of each primer and sterile Milli-Q water to a final volume of 10 µL. The PCR protocol consisted of an initial denaturation at 95°C for 2 min, followed by 35 cycles at 95°C (20 s), 62°C (20 s) and 72°C (30 s), followed by a final elongation at 72°C (5 min). PCR products were checked on 1% agarose gel at 5 V/cm in TBE buffer. A standard ethidium bromide staining was used for band visualization.

Functional marker design and genotyping

To perform association studies for the nectarine trait and provide a tool for marker assisted breeding (MAB) a codominant marker (indelG), consisting of a three-primer PCR assay (primers indelG-F, indelG-1R and indelG-2R; Table 3), was developed for genotyping of the candidate insertion: two outer primers (one forward and one reverse), designed on opposite sides of the insertion were combined in a single reaction with an inner reverse primer (designed on the reconstructed left sequence of the insertion).

PCRs were carried out in 10 µL containing 10 ng of template DNA, 1x PCR buffer, 1.5 mM MgCl₂, 200 µM each dNTP, 0.2 µM each primers and 0.5 U of Platinum Taq DNA polymerase (Life Technologies, Carlsbad, CA, USA). Amplifications were performed on a Veriti thermal cycle (Life Technologies, Carlsbad, CA, USA) with the following temperature profile: 95°C (5 min) followed by 35 cycles at 94°C (30 s), 61°C (30 s), 72°C (30 s) and a final extension at 72°C (10 min). PCR products were separated on an ethidium bromide stained 1% agarose gel. The 305 seedlings of the CxA F₂ population, the parents and the hybrid CxA F₁, as well as 46 peach and 49 nectarine accessions, were analyzed (Table 1).

Results and Discussion

Mapping of the G locus

Scoring of the CxA F₂ progeny revealed the presence of 246 peach and 59 nectarine plants, indicating a slight distortion from the expected 3:1 segregation ($\chi^2 = 5.41$, $p < 0.05$ with 1 d.f.). In agreement with a previous report [7], the peach/nectarine phenotype was regarded as a dominant trait (*G* locus). A total of 12 SNPs, covering about 7 Mb of LG 5 around the *G* locus, were mapped on the whole progeny (Figure 1). Consistent with the observed distortion of the phenotypic trait, a skewed segregation was also found for all the markers around the *G* locus from

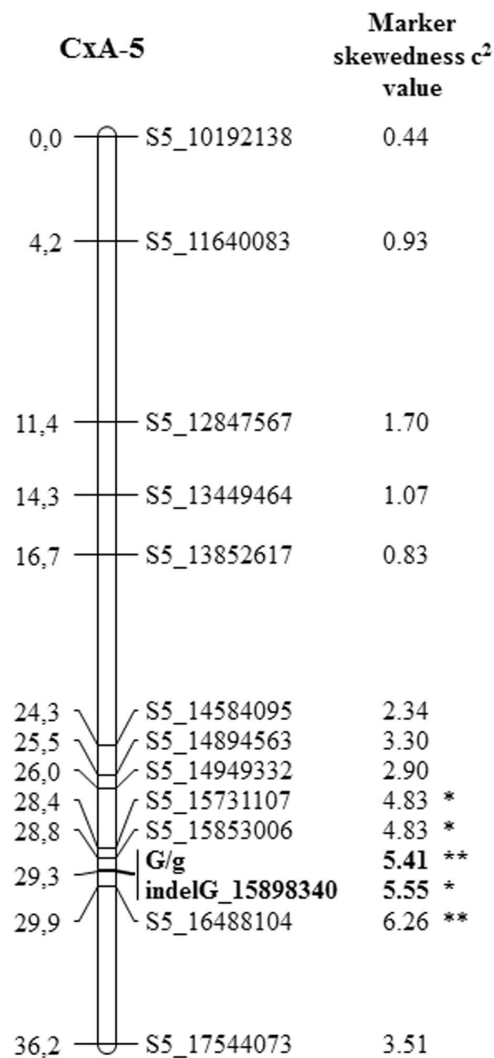


Figure 1. LG 5 CxA map around the *G* locus. Linkage map obtained from analysis of the CxA F₂ progeny. On the left side distances are indicated in cM; on the right the marker name, the physical position on Peach v1.0 and marker skewedness are reported. The peach/nectarine locus and the indelG marker are shown in bold. doi:10.1371/journal.pone.0090574.g001

S5_15731107 to S5_16488104 (Figure 1). The *G* locus was placed between markers S5_15853006 and S5_16488104 within an interval 1.1 cM (635 kb, Figure 1). Three additional SNPs inside the *G* region (S5_15865556, S5_15866258 and S5_15988499) were also successfully mapped by genotyping of informative recombinant plants. However none of them were useful to refine the position of *G* locus (data not shown).

Variant identification in the locus G genomic region

In order to identify variants putatively underlying the nectarine phenotype, genome-wide recalibrated alignment data from five *P. persica* accessions were examined in detail around the *G* locus (scaffold_5, from 15,853,006 to 16,488,104). In this region, GATK *HaplotypeCaller* detected 291 SNP and indel variants above the chosen minimum phred-scaled score quality threshold of 200, out of which 67 mapped within predicted genes in the peach genome (Peach v1.0) [8]. Of the latter, 20 variants, distributed in seven genes, were heterozygous in CxA F₁ with non-segregating joint

genotype combinations in ‘Bolero’ and ‘OroA’. In the nectarine ‘Quetta’ the same analysis identified two variants, both indels, homozygous for the non-reference allele and located in two distinct genes (*ppa023682m* and *ppa023143m*). Both of them were heterozygous in CxA F₁, where no other variants were found in these two genes, and were thus considered as putative candidates for the *G* locus. The first candidate variant is a deletion of 3 motives in a known (AAAC)₆ microsatellite at position 16,463,040 which maps within the only intron of the predicted gene *ppa023628m* (best *Arabidopsis thaliana* blastx match AT3G29575.1, ABI five binding protein *AFP3*, e-value 5×10^{-18}). ABI Five Binding Proteins (AFPs) are members of a small plant-specific protein family, characterized by three conserved domains of unknown function. AFPs act as negative regulators of ABA signaling [58] and have no known involvement in trichome formation. Next we focused on the second variant, reported by *HaploTYPE Caller* as two distinct insertions of 76 and 49 bp on the left and right side, respectively, of an (AC)₃ motif at position 15,898,458. This INDEL variant maps to the last exon (exon 3) of the predicted gene *ppa023143m* (best *Arabidopsis thaliana* blastx match AT5G15310.1, MYB domain protein 16, AtMYB16, e-value 1×10^{-73}). Similarity with R2R3-MYB transcription factors known to control epidermal cell differentiation [40,59] (see below) pointed to this gene as a likely candidate for the peach/nectarine trait [54]. This second variant was also detected by Pindel software in CxA F₁ and ‘Quetta’ samples only, as a large insertion compared to the reference sequence. In particular, this large insertion at position 15,898,458 was supported in Pindel by a total of 31 reads, 17 overhanging on the left side of the insertion (10 in CxA F₁ and 7 in ‘Quetta’) and 14 on its right side (5 in CxA F₁ and 9 in ‘Quetta’). The presence of this insertion in the nectarine allele is also supported by analysis of ‘Quetta’ resequencing data using CLC Genomic Workbench. Within the considered mapping interval (from 15,853,006 to 16,488,104) a total of 94,789 reads (9.15 million nucleotides) were aligned against the reference genome sequence, 61% of which in pairs and the remaining as single reads due to unexpected mapping distances, mate inversion, unmapping or mapping in other contigs. In agreement with Pindel results, the third exon of the *ppa023143m* gene showed a dramatic reduction of paired-end distances and an increase of single reads at

position 15,898,458 (Figure 2), compatible with a large insertion in ‘Quetta’ compared to the ‘Lovell’ reference sequence. Due to this insertion only single reads could align in the region; the software reports the lack of paired reads assigning the value zero to the paired-end distance and increasing to 100% the percentage of single reads (Figure 2).

Validation and reconstruction of a long insertion in exon 3 of gene *ppa023143m*

The physical presence of the long insertion within exon 3 of gene *ppa023143m* was confirmed by long-range PCR using a primer pair designed on intron 2 (Seq16F) and exon 3 (Seq4R) flanking the insertion site. An amplification product of about 7 kb was obtained in five nectarines, ‘Madonna di Agosto’, ‘Quetta’, ‘Stark Red Gold’, ‘Goldmine’ and ‘Ambra’ (Figure 3). In contrast, in a peach genotype (‘Contender’) the same primer pair gave an amplification product of 960 bp (data not shown). ‘Quetta’ and ‘Goldmine’ are two donors of the trait in modern breeding and ‘Stark Red Gold’ is known to carry the nectarine allele of ‘Lippiat’, the third donor of the trait. ‘Madonna di Agosto’ belongs to a group of landraces not directly related to modern breeding germplasm [12,13]. The double digestion of the five amplicons, with *EcoRI/HindIII*, shows the same restriction pattern for all the accessions (Figure 3) suggesting that a unique mutational event gave origin to the nectarine trait present in the modern nectarine germplasm as well as in the local southern Italian ecotypes.

The amplicon obtained in ‘Quetta’ was sequenced by Next Generation Sequencing (NGS). Following filtering and assembly 90% of the reads were collected in three major contigs. Contig_1 (GenBank accession number KJ150676), formed by a consensus sequence of 5,836 bp, was mapped on scaffold 3 of peach genome v1.0 and showed a perfect match for 5,713 bp (scaffold_3:13,409,926.13,415,638) corresponding to the predicted LTR_1684 region (http://services.appliedgenomics.org/gbrowse/prunus_public/). The missing 123 bp that do not align on scaffold 3 showed a perfect match on scaffold 5 (scaffold_5:15,898,458.15,898,581), confirming that this contig represents an insertion in the third exon of predicted gene *ppa023143m*. When the LTR_1684 region, found in correspondence of Contig_1 alignment on scaffold 3, was submitted to CENSOR [60] (Release

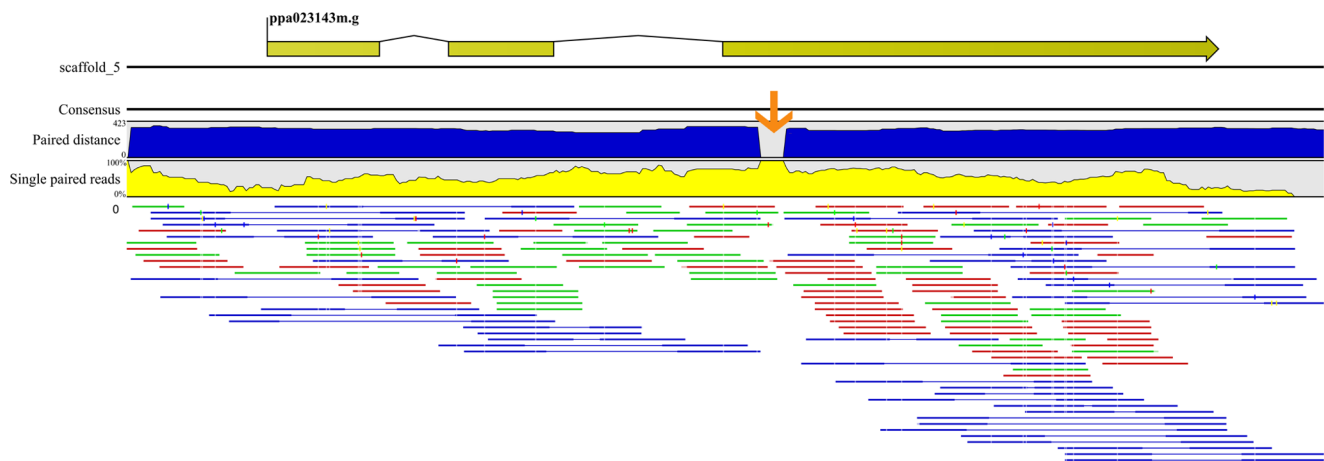


Figure 2. Alignment of Quetta reads against a 635 kb interval of Peach v1.0 pseudomolecule 5. Alignment results of reads, obtained by the resequencing of ‘Quetta’, against the peach genome region identified by the mapping interval in LG5 (from 15,853,006 bp to 16,488,104 bp). Top panel: intron-exon structure of *ppa023143m*. Central panel: plot of ‘Quetta’ paired-end distance (blue) and frequencies of single reads (yellow) at the *ppa023143m* locus. Bottom panel: blue lines are paired reads, green and red lines correspond to single reads with missing mate on the right and left side, respectively. The orange arrow points to the putative insertion inside exon 3 of *ppa023143m*. doi:10.1371/journal.pone.0090574.g002

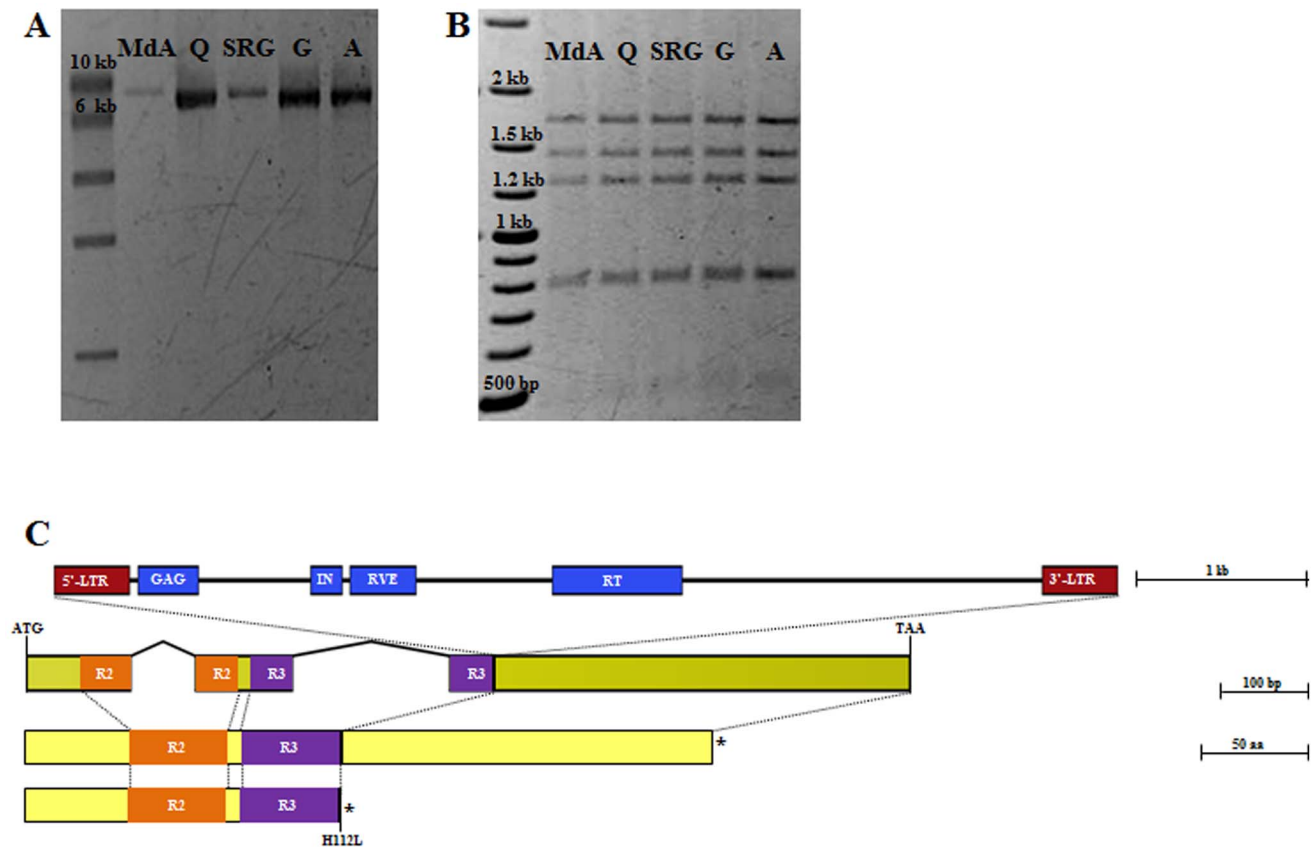


Figure 3. Variant discovery in *PpaMYB25* (annotation refinement of *ppa023143m*). Five nectarine genotypes ('Madonna di Agosto', MdA; 'Quetta', Q; 'Stark Red Gold', SRG; 'Goldmine', G; 'Ambra', A) were analyzed to confirm the presence of the insertion within exon 3 of *PpaMYB25*. (A) Long-range amplification products reveal for all the accessions a fragment of about 7 kb (compared to 960 bp expected from the reference genome). (B) Double digestion results of the long-range PCR products show the same pattern for all the genotypes. (C) Position and structure of the *Ty-copia* retrotransposon deduced by the by the NGS analysis of 'Quetta' long-range amplicon. The insertion results in a truncated version of the R2R3-MYB protein.

doi:10.1371/journal.pone.0090574.g003

18.01, 16 Jan 2013) high similarity was found to a 6,033 bp annotated *Ty1-copia* retrotransposon (Copia-24_FV-I) from strawberry (*Fragaria vesca*). A conserved protein domain analysis on the complete sequence of Contig_1 also revealed the presence of different domains: UBN2 (gag-polypeptide of LTR copia-type) from position 1,354 to 1,689 of Contig_1, GAG-pre-integrase domain from position 2,329 to 2,511, RVE (Integrase core domain) from position 2,551 to 2,928 and RVT2 (Reverse transcriptase domain) from position 3,694 to 4,431. The four domains identified showed a high prediction confidence, with E-values of 3.8e-16, 4.3e-13, 6.8e-28 and 1.6e-98 respectively. All these predicted domains are typical of retrotransposable elements. The remaining two contigs, Contig_2 and Contig_3, 300 and 876 bp respectively, were also mapped on the peach genome. Both were split on scaffold 5 (scaffold_5:15,898,340.15,898,361; scaffold_5: 15,898,458. 15,899,262) in exon 3 of predicted gene *ppa023143m* and on scaffold 3 in the same LTR_1684 region of Contig_1. Thus, these contigs span the point of insertion of the *Ty-copia* retroelement within exon 3 of the gene. Flanking this insertion point, we found a characteristic Target Site Duplication (TSD) (AC)₃ [61] produced by the retroelement upon insertion into a new site. Together these results confirm the insertion of a *Ty1-copia* retrotransposon in the 'Quetta' allele of gene *ppa023143m*.

A BLASTN search of the retrotransposon sequence of Contig_1 against the peach reference genome returned 5 highly similar hits (from 87.1% to 100% sequence identity), two on chromosome 3, and one each on chromosomes 4, 7 and 8. All five hits were about 6 kb long and were precisely delimited by highly similar (from 97.7% to 100.0%) LTR sequences, each flanked by characteristic Target Site Duplications, thus confirming the existence of other copies of this LTR-retroelement in the reference peach genome. In particular LTRs found in the 'Quetta' allele of *ppa023143m* are identical to those present in scaffold 3 (scaffold_3:14,488,093.14,488,522) and scaffold 4 (scaffold_4:27,503,652.27,504,081). These data are consistent with previous analyses showing that 12.6% of LTR-retrotransposons have identical LTRs, indicating recent and ongoing retrotransposition activity in the peach genome [8].

Transposable elements (TEs) are known to cause many kinds of genetic variations in plants and played an important role in plant evolution and domestication [62]. In a survey of allelic variants at 60 genes involved in crop domestication and diversification, 15% were caused by TE insertions [63]. For example, an LTR insertion in a regulatory region of the *teosinte branched1 (tb1)* gene resulted in overexpression of the gene, causing the conversion from highly branched wild teosinte to the single culm architecture of domesticated maize [64,65]. However, the most common effect of TE insertion is the loss of gene function [57] and recessive TE-

| | | |
|----------|----|--|
| PpeMYB25 | 15 | RGPWTPPEEDRKLLEDEIQHGHGWSSSLPOKAGLKRCGKSCRLRWINYLRPDIK |
| GhMYB25 | 14 | KGPWTPEEDQKLLAYIEQHGHGSRALPLKAGLQRCGKSCRLRWINYLRPDIK |
| AmMYBML1 | 14 | RGPWTPPEEDQKLLSYIQEHGHGSRALPSKAGLQRCGKSCRLRWSNYLRLPDIK |
| | | |
| PpeMYB25 | 68 | RGNFSLHEEDQTIHQHALLGNRWSAIAAHLPKRTDNEIKNYWNTHL |
| GhMYB25 | 67 | RGKFSLQEEQTIHQHALLGNRWSAIAATHLPKRTDNEIKNYWNTHL |
| AmMYBML1 | 67 | RGKFSLQEEQAIHQHALLGNRWSAIAATHLPKRTDNEIKNYWNTHL |

Figure 4. Aminoacid alignment of the R2 and R3 MYB repeat sequences. MYB domains (pfam00249) of peach PpeMYB25, cotton GhMYB25 (ACJ07153.1, [39]) and *Antirrhinum* AmMYBML1 (CAB433991.1, [54]) were aligned using the Muscle on line tool at EBI (<http://www.ebi.ac.uk/Tools/msa/muscle/>). Graphic display of the alignment was obtained using BoxShade (http://www.ch.embnet.org/software/BOX_form.html). Black shaded residues are identical, grey shaded residues are similar. Coordinates in the protein sequences are indicated. doi:10.1371/journal.pone.0090574.g004

induced mutations have played an important role in plant domestication [62]; examples include “sticky” foxtail millet [66], Mendel’s wrinkled peas [67], seedless apple [68], fruit color in grape [69–71] and peach flesh [72,73]. In some cases, multiple independent mutational events were selected as demonstrated by insertion of TEs in the *waxy* locus resulting in the “sticky” phenotype in foxtail millet [66]. Another example is the yellow peach phenotype, which is associated with three different mutational events (an LTR insertion, a SNP and a frameshift mutation at a microsatellite locus) that occurred independently in a carotenoid cleavage dioxygenase gene directly involved in pigment degradation [72,73]. In contrast with the situation for “sticky” foxtail millet [66] and peach color [72,73] our results suggest that a unique mutational event has originated the nectarine phenotype, i.e. the loss of trichomes in peach fruit. The occurrence of unique mutations affecting single genes selected by humans during domestication and diversification of crop species is not rare. In a recent review of 60 such genes, 26 display a unique mutational event selected and spread by humans [63].

Gene ppa023143m encodes an R2R3-MYB transcription factor putatively involved in trichome formation

Careful inspection of predicted gene *ppa023143m* led us to reannotate the coding sequence (CDS) extending exon 3 compared to the Peach v1.0 annotation [8]. The reannotated CDS is predicted to encode a peptide of 330 aminoacids showing similarity to the R2R3-MYB transcription factors GhMYB25 from allotetraploid cotton *Gossypium hirsutum* (58.4% similarity) [40] and MIXTA-like1 from *Antirrhinum* (AmMYBML1, 55.3% similarity) [59] (Figure 4). In eukaryotes, MYB factors represent one of the largest and most functionally diverse gene families, which dramatically expanded in plants [74–77] playing a central role in a variety of processes from plant development to responses to biotic and abiotic stresses [74,78,79]. MYB family members share a highly conserved DNA binding domain (the MYB domain) usually composed of up to three amino acid repeats (R1, R2, R3) [80]. GhMYB25 and AmMYBML1 belong to R2R3-MYB subgroup 9 [40,77] along with other genes involved in regulating petal epidermal cell shape [81]. AmMYBML1 plays a role in trichome differentiation in the corolla tube of the *Antirrhinum* flower [59]. GhMYB25, normally expressed at the time of fiber initiation in the outer integument of ovules, is differentially expressed between fibreless mutants and normal lined cotton [40,82,83] and its altered expression affects seed, fiber and trichome development in transgenic cotton [40]. Considering the known role of these homologues in trichome development, *ppa023143m* is a strong candidate for the peach/nectarine phenotype and was named PpeMYB25. The insertion of the Ty1-*copia* retrotransposon in exon 3 of the PpeMYB25 gene introduces an H112L substitution and a

premature stop codon (TAA), resulting in a peptide of 112 aminoacids precisely truncated at the C-terminal end of the R3 MYB domain. GhMYB25 and AmMYBML1 share the distinctive C-terminal motif of R2R3-MYB sub-group 9 [40,59,76] along with PpeMYB25 and other genes involved in regulating epidermal cell shape [81]. Analysis of *Antirrhinum* R2R3-MYB genes from subgroup 9 indicated that the C-terminal domain folds as an amphipathic α -helix with putative transactivation ability [84]. An insertional mutant in the MIXTA gene, resulting in loss of this C-

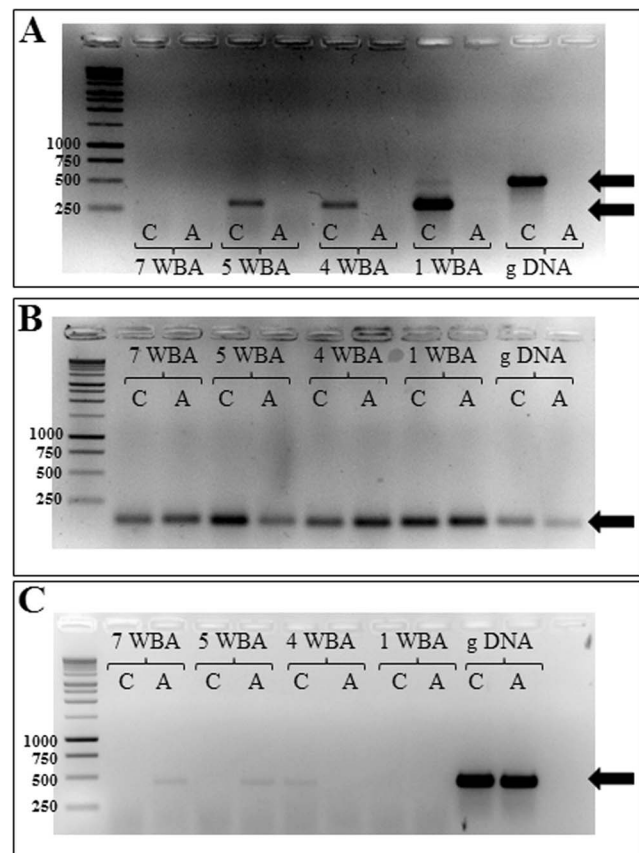


Figure 5. Expression analysis of PpeMYB25 in peach and nectarine flower buds. (A) The expression patterns of the R2R3-MYB gene were evaluated in ‘Contender’ [C] and ‘Ambra’ [A] buds at seven, five, four and one week before anthesis (WBA). Genomic DNA of the two cultivars was also tested as a control. The same samples were analyzed for expression of *RPII* as standard (B) and checked for DNA contamination (C). doi:10.1371/journal.pone.0090574.g005

terminal region has been shown to cause recessive phenotypic alteration in epidermal cell differentiation [85]. The recessiveness of the nectarine trait indicates that it corresponds to a loss-of-function mutation. According to this reasoning and by analogy with observations in cotton, we propose that the observed insertion in the *PpeMYB25* gene results in a non-functional form of the MYB transcription factor that normally promotes trichome formation in fuzzy peaches. If this were correct, all nectarines should be non-functional homozygous mutants at this gene.

In order to evaluate the timing of *PpeMYB25* transcript expression with respect to trichome development, RT-PCR analyses were performed on 'Contender' and 'Ambra' floral buds sampled at seven, five, four and one week before anthesis. A forward primer designed on exon 2 was used in combination with a reverse primer on exon 3 (downstream of the LTR insertion) in order to evaluate the expression profiles of the gene. Expression was evident in 'Contender' from five weeks before anthesis, just before trichomes begin to appear [26] and continued through to one week before anthesis accompanying trichome differentiation (Figure 5). In contrast, the expression of *PpeMYB25* was never visible in 'Ambra' floral buds, consistent with the presence and the position of the large insertion (Figure 5). Together sequence and expression analyses support the proposed involvement of *PpeMYB25* gene in promoting trichome differentiation. Mutational events in regulatory genes have played a major role during crop domestication and breeding; in a recent overview of domestica-

tion and diversification genes, 37 out of 60 (~62%) encoded transcription factors [63].

Association study and origin of the nectarine trait in peach germplasm

The genotype at the LTR retrotransposon insertion in gene *PpeMYB25* was assessed in the CxA F₂ population as well as in a panel of 95 peach and nectarine accessions by means of a functional marker based on a three primers PCR assay (indelG) (Figure 6). As expected, the marker co-segregated with the *G* locus in the CxA F₂ progeny (Figure 1), with all the nectarines displaying a unique fragment of 197 bp. Similarly, all nectarines in the germplasm panel were characterized by a fragment of the same length (Figure 6). Peach accessions fell into two categories: those homozygous for the reference allele (941 bp) and those heterozygous (197 bp, 941 bp). Taking into account pedigree information [15,86], we confirmed all the known heterozygotes ('Autumnglo', 'Fairtime', 'Fidelia', 'Jing Yu', 'O'Henry', 'Summer Pearl' and CxA F₁, Table 1) and we also detected three previously unknown heterozygous genotypes ('Baldagenais', 'City 32-82' and 'Yoshihime') carrying the nectarine allele originated from the LTR insertion. The complete association between the indelG marker and the trait confirms the presence of the Ty1-*copia* retrotransposon within exon 3 of the *PpeMYB25* gene in all the nectarines analyzed.

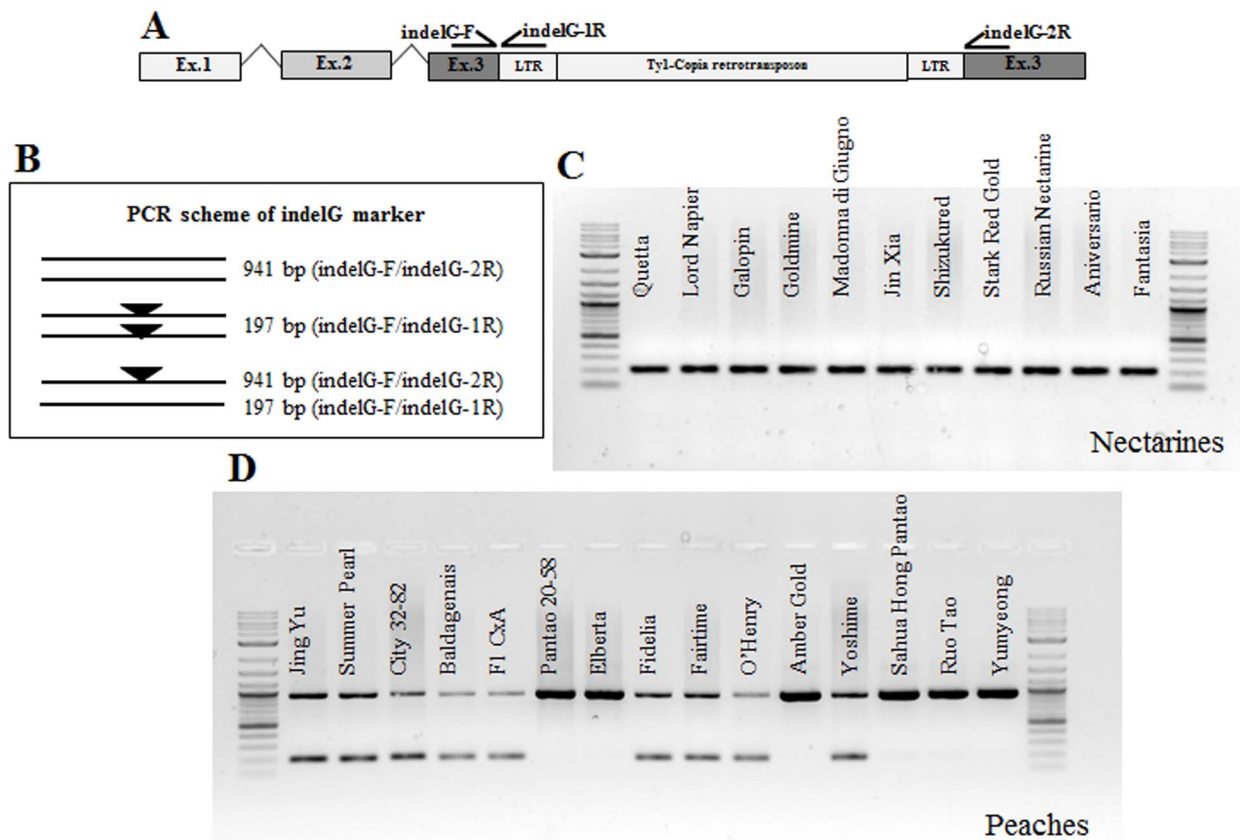


Figure 6. Functional Marker indelG. A marker assay was developed based on sequence information on the *PpeMYB25* gene and the Ty1-*copia* insertion. Three primers were designed to discriminate peach and nectarine genotypes (A, B). A panel of nectarines including the putative donors of the trait, show a unique fragment of about 200 bp (C). A set of peaches, of diverse pedigree and origins (Table 1) (D), shows homozygous or heterozygous patterns.

doi:10.1371/journal.pone.0090574.g006

Modern western nectarines trace back to three founders discovered at the beginning of the last century ('Quetta' in Pakistan, and 'Lippiatt' and 'Goldmine' in New Zealand). The founders and their modern descendants show the presence of the retrotransposon insertion. The insertion is also present in Southern Italian traditional landraces ('Madonna di Giugno' and 'Madonna di Agosto') cultivated since the 16th century [12]. In addition to these genotypes, we confirmed the presence of the retrotransposon in several non-related accessions including old European landraces ('Lord Napier' and 'Galopin') [15], modern Asian cultivars (Chinese and Japanese) and different nectarine cultivars of unknown pedigree (from Italy, East Europe, USA, Mexico and South America). The modern Asian nectarines analyzed in this study ('Chiyodared', 'Shizukured' and 'Jin Xia') have two old European landraces, 'Lord Napier' and 'Precoce di Croncels', as donors of the nectarine trait. No traditional nectarine landraces have been reported in China [14] and all modern Chinese nectarines inherited the trait from western germplasm [17].

Together these results indicate that all known nectarine germplasm derives from a unique mutational event in *PpeMYB25* selected and spread by humans during peach dissemination and breeding.

Conclusions

Nectarines play an important role in the peach industry. In the present study, using a candidate gene approach coupled with fine mapping and NGS-based variant discovery, we provide strong evidence that the transcription factor gene *PpeMYB25* acts as a positive regulator of trichome formation in peach fruit. The

insertion of a *Ty1-copia* retrotransposon within the third exon of *PpeMYB25* was identified as the putative cause of a loss-of-function mutation underlying the nectarine phenotype, further supporting the importance of transposition in plant genome evolution and phenotypic variability in domesticated crops. Finally, the development of a functional marker, *indelG*, provides an efficient diagnostic tool for the early selection of the peach/nectarine trait in marker assisted breeding (MAB).

Acknowledgments

We wish to thank S. Foschi (CRPV, Cesena, Italy) and M. Lama (ASTRA Innovazione, Faenza, Italy) for their valuable contribution in field tree management, as well as C. Cantini (Ivalsa-CNR Follonica, Italy) for providing leaf tissue for DNA analyses of the accessions 'Galopin' and 'Lord Napier'. We are grateful to D.S. Horner (DBS, Università di Milano, Italy) for critical comments and English revision of the manuscript. The authors thank the Centre for Applied Biomedical Research (CRBA) of Bologna for valuable contributions to the Sequenom analyses. The authors also thank Fondazione Cassa di Risparmio in Bologna for supporting CRBA. We acknowledge the Department of Rare Books and Special Collections, Princeton University Library (USA) for providing photo duplication of the New York Gazette March 28, 1768.

Author Contributions

Conceived and designed the experiments: GP ST IV LR. Performed the experiments: EV GP LD IP LG MTD. Analyzed the data: EV GP LD IP SS FS ST IV LR. Contributed reagents/materials/analysis tools: DB ST IV LR. Wrote the paper: EV GP ST IV LR. Critically revised the manuscript: LD IP MTD SS FS DB LG.

References

- Pastorello E, Farioli L, Pravettoni V, Scibilia J, Mascheri A, et al. (2011) Pru p 3-sensitized Italian peach-allergic patients are less likely to develop severe symptoms when also presenting IgE antibodies to Pru p 1 and Pru p 4. *Int Arch Allergy Immunol* 156: 362–372.
- Botton A, Vegro M, De Franceschi F, Ramina A, Gemignani C, et al. (2006) Different expression of Pp-LTP1 and accumulation of Pru p 3 in fruits of two *Prunus persica* L. Batsch genotypes. *Plant Sci* 171: 106–113.
- Wang G, Tian L, Aziz N, Broun P, Dai X, et al. (2008) Terpene biosynthesis in glandular trichomes of hop. *Plant Physiol* 148: 1254–1266.
- Harada E, Kim JA, Meyer AJ, Hell R, Clemens S, et al. (2010) Expression profiling of tobacco leaf trichomes identifies genes for biotic and abiotic stresses. *Plant Cell Physiol* 51: 1627–1637.
- Blake MA (1932) The JH Hale peach as a parent in peach crosses. *Proc Natl Acad Sci USA* 29: 131–136.
- Dirlwanger E, Cosson P, Boudehri K, Renaud C, Capdeville G, et al. (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.
- Le Dantec L, Cardinet G, Bonet J, Fouché M, Boudehri K, et al. (2010) Development and mapping of peach candidate genes involved in fruit quality and their transferability and potential use in other Rosaceae species. *Tree Genet Genomes* 6: 995–1012.
- Verde I, Abbott AG, Scalabrini S, Jung S, Shu S, et al. (2013) The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nat Genet* 45: 487–494.
- Faust M, Timon B (1995) Origin and dissemination of peach. *John Wiley*. Janik J, editor *Horticultural Reviews*, Volume 17.
- Yoon J, Liu D, Song W, Liu W, Zhang A, et al. (2006) Genetic diversity and ecogeographical phylogenetic relationships among peach and nectarine cultivars based on Simple Sequence Repeat (SSR) markers. *J Amer Soc Hort Sci* 131: 513–521.
- Hesse CO (1975) Peach In: *Advances in fruit breeding*. Purdue Uni. Janick, J, Moore J, editor West Lafayette, Ind.
- Venuto A (1516) *De agricultura opusculum*. Edizioni dell'Orso ISBN 978-88-6274-029-6.
- Marchese A, Tobutt KR, Caruso T (2005) Molecular characterisation of Sicilian *Prunus persica* cultivars using microsatellites. *J Hort Sci Biotechnol* 80: 121–129.
- Okie WR, Bacon T, Bassi D (2008) *Fresh Market Cultivar Development in: Peach, Botany, Production and Uses*. CAB Intern. Lane DR and Bassi D, editor.
- Okie WR (1998) *Handbook of peach and nectarine varieties*. U.S. Dept. of Agriculture, Agricultural Research Service; National Technical Information Service, distributor.
- Yoshida M (1994) Horticulture in Japan. In: Kitagawa, H, Iwahori, S, Yakuwa, T, Konishi, K, editor. *Peach*. Tokyo: Asakura Publishing Co., Ltd. pp. 32–37.
- Tian JB, Cheng F, Cheng H, Han F, Dong B (2012) Review and perspective of nectarine breeding in China. *Acta Horticulture* 962: 97–103.
- Esau K (1977) *Anatomy of seed plants*, 2nd edn. Wiley. New York.
- Uphof JCT (1962) *Plant Air, Encyclopedia of plant anatomy*. Grubruer. Zimmerman, W; Ozenda, PG, editor Berlin.
- Jeffree CE (2006) The fine structure of the plant cuticle. In 'Biology of the plant cuticle'. Blackwell. Riederer M, Müller C, editor Oxford.
- Stavrianakou S, Liakopoulos G, Miltiadou D, Markoglou A, Ziogas B, et al. (2010) Antifungal and antibacterial capacity of extracted material from non-glandular and glandular leaf hairs applied at physiological concentrations. *Plant Stress* 4: 25–30.
- Xia Y, Yu K, Navarre D, Seebold K, Kachroo A, et al. (2010) The glabral mutation affects cuticle formation and plant responses to microbes. *Plant Physiol* 154: 833–846.
- Lange BM, Wildung MR, Stauber EJ, Sanchez C, Pouchnik D, et al. (2000) Probing essential oil biosynthesis and secretion by functional evaluation of expressed sequence tags from mint glandular trichomes. *Proc Natl Acad Sci* 97: 2934–2939.
- Iijima Y, Davidovich-Rikanati R, Fridman E, Gang DR, Bar E, et al. (2004) The biochemical and molecular basis for the divergent patterns in the biosynthesis of terpenes and phenylpropenes in the peltate glands of three cultivars of basil. *Plant Physiol* 136: 3724–3736.
- Wilkins TA, Rajasekaran K, Anderson DM (2000) *Cotton Biotechnology*. CRC Crit Rev Plant Sci 19: 511–550.
- Creller MA, Werner DJ (1996) Characterizing the novel fruit surface morphology of 'Marina' Peach Using Scanning Electron Microscopy. 121: 198–203.
- Fernández V, Khayet M, Montero-Prado P, Heredia-Guerrero JA, Liakopoulos G, et al. (2011) New insights into the properties of pubescent surfaces: peach fruit as a model. *Plant Physiol* 156: 2098–2108.
- Hülkamp M (2004) Plant trichomes: a model for cell differentiation. *Nat Rev Mol Cell Biol* 5: 471–480.
- Jakoby MJ, Falkenhan D, Mader MT, Brininstool G, Wischnitzki E, et al. (2008) Transcriptional profiling of mature *Arabidopsis* trichomes reveals that NOECK encodes the MIXTA-like transcriptional regulator MYB106. *Plant Physiol* 148: 1583–1602.
- Oppenheimer DG, Herman PL, Sivakumaran S, Esch J, Marks MD (1991) A myb gene required for leaf trichome differentiation in *Arabidopsis* is expressed in stipules. *Cell* 67: 483–493.

31. Payne CT, Zhang F, Lloyd AM (2000) GL3 encodes a bHLH protein that regulates trichome development in *Arabidopsis* through interaction with GL1 and TTG1. *Genetics* 156: 1349–1362.
32. Zhang F, Gonzalez A, Zhao M, Payne CT, Lloyd A (2003) A network of redundant bHLH proteins functions in all TTG1-dependent pathways of *Arabidopsis*. *Development* 130: 4859–4869.
33. Wang S, Chen JG (2008) *Arabidopsis* transient expression analysis reveals that activation of GLABRA2 may require concurrent binding of GLABRA1 and GLABRA3 to the promoter of GLABRA2. *Plant Cell Physiol* 49: 1792–1804.
34. Wada T, Tachibana T, Shimura Y, Okada K (1997) Epidermal cell differentiation in *Arabidopsis* determined by a Myb homolog, CPC. *Science* (80) 277: 1113–1116.
35. Schellmann S, Schnitger A, Kirik V, Wada T, Okada K, et al. (2002) TRIPTYCHON and CAPRICE mediate lateral inhibition during trichome and root hair patterning in *Arabidopsis*. *EMBO J* 21: 5036–5046.
36. Kirik V, Simon M, Wester K, Schiefelbein J, Hulskamp M (2004) ENHANCER of TRY and CPC 2 (ETC2) reveals redundancy in the region-specific control of trichome development of *Arabidopsis*. *Plant Mol Biol* 55: 389–398.
37. Hauser MT, Harr B, Schlotterer C (2001) Trichome distribution in *Arabidopsis thaliana* and its close relative *Arabidopsis lyrata*: molecular analysis of the candidate gene GLABROUS1. *Mol Biol Evol* 18: 1754–1763.
38. Kivimäki M, Kärkkäinen K, Gaudeul M, Løe G, Agren J (2007) Gene, phenotype and function: GLABROUS1 and resistance to herbivory in natural populations of *Arabidopsis lyrata*. *Mol Ecol* 16: 453–462.
39. Li F, Zou Z, Yong HY, Kitashiba H, Nishio T (2013) Nucleotide sequence variation of GLABRA1 contributing to phenotypic variation of leaf hairiness in *Brassicaceae* vegetables. *Theor Appl Genet* 126: 1227–1236.
40. Machado A, Wu Y, Yang Y, Llewellyn DJ, Dennis ES (2009) The MYB transcription factor GhMYB25 regulates early fibre and trichome development. *Plant J* 59: 52–62.
41. Pirona R, Eduardo I, Pacheco I, Da Silva Linge C, Miculan M, et al. (2013) Fine mapping and identification of a candidate gene for a major locus controlling maturity date in peach. *BMC Plant Biol* 13: 166.
42. Eduardo I, Pacheco I, Chietera G, Bassi D, Pozzi C, et al. (2011) QTL analysis of fruit quality traits in two peach intraspecific populations and importance of maturity date pleiotropic effect. *Tree Genet Genomes* 7: 323–335.
43. Jurinke C, van den Boom D, Cantor CR, Köster H (2002) The use of MassARRAY technology for high throughput genotyping in: chip technology. Hoheisel J, Brazma A, Büssow K, Cantor CR, Christians FC, et al, editors Berlin, Heidelberg: Springer Berlin Heidelberg.
44. Van Ooijen JW, Voorrips RW (2002) Joinmap 3.0, Software for the calculation of genetic linkage maps. Wageningen, The Netherlands: Plant Research International.
45. Kosambi DD (1943) The estimation of map distances from recombination values. *Ann Eugen* 12: 172–175.
46. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, et al. (2008) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 36: D13–21.
47. Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, et al. (2012) RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res* 40: W622–W627.
48. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760.
49. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20: 1297–1303.
50. Verde I, Bassil N, Scalabrin S, Gilmore B, Lawley CT, et al. (2012) Development and evaluation of a 9K SNP array for peach by internationally coordinated SNP detection and validation in breeding germplasm. *PLoS One* 7: e35668.
51. DePristo MA, Banks E, Poplin R, Garimella K V, Maguire JR, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43: 491–498.
52. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z (2009) Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* 25: 2865–2871.
53. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26: 589–595.
54. Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 39: W29–37.
55. Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, et al. (2012) The Pfam protein families database. *Nucleic Acids Res* 40: D290–301.
56. Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R, et al. (2008) Database indexing for production MegaBLAST searches. *Bioinformatics* 24: 1757–1764.
57. Tong Z, Gao Z, Wang F, Zhou J, Zhang Z (2009) Selection of reliable reference genes for gene expression studies in peach using real-time PCR. *BMC Mol Biol* 10: 71.
58. Lopez-Molina L, Mongrand S, Kinoshita N, Chua NH (2003) AFP is a novel negative regulator of ABA signaling that promotes ABI5 protein degradation. *Genes Dev* 17: 410–418.
59. Perez-Rodriguez M, Jaffe FW, Butelli E, Glover BJ, Martin C (2005) Development of three different cell types is associated with the activity of a specific MYB transcription factor in the ventral petal of *Antirrhinum majus* flowers. *Development* 132: 359–370.
60. Kohany O, Gentles AJ, Hankus L, Jurka J (2006) Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* 7: 474.
61. Kumar A (1996) The adventures of the Tyl-copia group of retrotransposons in plants. *Trends Genet* 12: 41–43.
62. Lisch D (2013) How important are transposons for plant evolution? *Nat Rev Genet* 14: 49–61.
63. Meyer RS, Purugganan MD (2013) Evolution of crop species: genetics of domestication and diversification. *Nat Rev Genet* 14: 840–852.
64. Studer A, Zhao Q, Ross-Ibarra J, Doebley J (2011) Identification of a functional transposon insertion in the maize domestication gene tb1. *Nat Genet* 43: 1160–1163.
65. Doebley JF, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. *Cell* 127: 1309–1321.
66. Kawase M, Fukunaga K, Kato K (2005) Diverse origins of waxy foxtail millet crops in East and Southeast Asia mediated by multiple transposable element insertions. *Mol Genet Genomics* 274: 131–140.
67. Bhattacharyya MK, Smith AM, Ellis THN, Hedley C, Martin C (1990) The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. *Cell* 60: 115–122.
68. Yao JL, Dong YH, Morris BAM (2001) Parthenocarpic apple fruit production conferred by transposon insertion mutations in a MADS-box transcription factor. *Proc Natl Acad Sci USA* 98: 1306–1311.
69. Kobayashi H, Goto-Yamamoto N, Hirochika H (2004) Retrotransposon-induced mutations in grape skin color. *Science* (80) 304: 982.
70. Cadle-Davidson MM, Owens CL (2008) Genomic amplification of the Gret1 retroelement in white-fruited accessions of wild vitis and interspecific hybrids. *Theor Appl Genet* 116: 1079–1094.
71. Shimazaki M, Fujita K, Kobayashi H, Suzuki S (2011) Pink-colored grape berry is the result of short insertion in intron of color regulatory gene. *PLoS One* 6: e21308.
72. Falchi R, Vendramin E, Zanon L, Scalabrin S, Cipriani G, et al. (2013) Three distinct mutational mechanisms acting on a single gene underpin the origin of yellow flesh in peach. *Plant J*: 76 (2):175–187.
73. Adami M, Franceschi P, Brandi F, Liverani A, Giovannini D, et al. (2013) Identifying a carotenoid cleavage dioxygenase (ccd4) gene controlling yellow/white fruit flesh color of peach. *Plant Mol Biol Report* 31: 1166–1175.
74. Dubos C, Stracke R, Gretewold E, Weisshaar B, Martin C, et al. (2010) MYB transcription factors in *Arabidopsis*. *Trends Plant Sci* 15: 573–581.
75. Wilkins O, Nahal H, Foong J, Provart NJ, Campbell MM (2009) Expansion and diversification of the *Populus* R2R3-MYB family of transcription factors. *Plant Physiol* 149: 981–993.
76. Stracke R, Werber M, Weisshaar B (2001) The R2R3-MYB gene family in *Arabidopsis thaliana* R. *Curr Opin Plant Biol* 4: 447–456.
77. Jia L, Clegg MT, Jiang T (2004) Evolutionary dynamics of the DNA-binding domains in putative R2R3-MYB genes identified from rice subspecies indica and japonica genomes. *Plant Physiol* 134: 575–585.
78. Zhang L, Zhao G, Jia J, Liu X, Kong X (2012) Molecular characterization of 60 isolated wheat MYB genes and analysis of their expression during abiotic stress. *J Exp Bot* 63: 203–214.
79. Glover B, Perez-Rodriguez M, Martin C (1998) Development of several epidermal cell types can be specified by the same MYB-related plant transcription factor. *Development* 125: 3497–3508.
80. Rosinski JA, Atchley WR (1998) Molecular evolution of the Myb family of transcription factors: evidence for polyphyletic origin. *J Mol Evol* 46: 74–83.
81. Martin C, Bhatt K, Baumann K, Jin H, Zachgo S, et al. (2002) The mechanics of cell fate determination in petals. *Philos Trans R Soc Lond B Biol Sci* 357: 809–813.
82. Wu Y, Machado AC, White RG, Llewellyn DJ, Dennis ES (2006) Expression profiling identifies genes expressed early during lint fibre initiation in cotton. *Plant Cell Physiol* 47: 107–127.
83. Lee JJ, Hassan OSS, Gao W, Wei NE, Kohel RJ, et al. (2006) Developmental and gene expression analyses of a cotton naked seed mutant. *Planta* 223: 418–432.
84. Jaffe FW, Tattersall A, Glover BJ (2007) A truncated MYB transcription factor from *Antirrhinum majus* regulates epidermal cell outgrowth. *J Exp Bot* 58: 1515–1524.
85. Noda K, Glover BJ, Linstead P, Martin C (1994) Flower colour intensity depends on specialized cell shape controlled by a Myb-related transcription factor. *Nature* 369: 661–664.
86. Aranzana MJ, Abbassi EK, Howad W, Arús P (2010) Genetic variation, population structure and linkage disequilibrium in peach commercial varieties. *BMC Genet* 11: 69.