QTL mapping of pomological traits in peach and related species breeding germplasm

Jonathan Fresnedo-Ramírez, Marco C. A. M. Bink, Eric van de Weg, Thomas R. Famula, Carlos H. Crisosto, Terrence J. Frett, et al.

Molecular Breeding
New Strategies in Plant Improvement

ISSN 1380-3743
Volume 35
Number 8

Mol Breeding (2015) 35:1-19
DOI 10.1007/s11032-015-0357-7

Springer

Mol Breeding (2015)35:166
DOI 10.1007/s11032-015-0357-7

QTL mapping of pomological traits in peach and related species breeding germplasm

Jonathan Fresnedo-Ramírez · Marco C. A. M. Bink · Eric van de Weg · Thomas R. Famula · Carlos H. Crisosto · Terrence J. Frett · Ksenija Gasic · Cameron P. Peace · Thomas M. Gradziel

Received: 10 November 2014 / Accepted: 17 July 2015

Abstract Peach is an economically important fruit tree crop that exhibits high phenotypic variability yet suffers from diversity-limited gene pool. Genetic introgression of novel alleles from related species is being pursued to expand genetic diversity. This process is, however, challenging and requires the incorporation of innovative genomic and statistical tools to facilitate efficient transfer of these exotic alleles across the multiple generations required for introgression. In this study, pedigree-based analysis (PBA) in a Bayesian QTL mapping framework was applied to a diverse peach pedigree introgressed with almond and other related *Prunus* species. The aim was to investigate the genetic control of eight commercially important fruit productivity and fruit quality traits over two subsequent years. Fifty-two QTLs with at least positive evidence explaining up to 98 % of the phenotypic variance across all trait/year combinations were mapped separately per trait and year. Several QTLs exhibited variable association with traits between years. By using the peach genome sequence as a reference, the intrachromosomal positions for several QTLs were shown to differ from those

Electronic supplementary material The online version of this article (doi:10.1007/s11032-015-0357-7) contains supplementary material, which is available to authorized users.

J. Fresnedo-Ramírez · C. H. Crisosto · T. M. Gradziel (✉)
Department of Plant Sciences, University of California, Davis, CA 95616, USA
e-mail: tmgradziel@ucdavis.edu

J. Fresnedo-Ramírez
e-mail: jfresnedo@cornell.edu

C. H. Crisosto
e-mail: chcrisosto@ucdavis.edu

Present Address:
J. Fresnedo-Ramírez
BRC Bioinformatics Facility, Institute of Biotechnology, Cornell University, Ithaca, NY 14853, USA

M. C. A. M. Bink
Biometris, Wageningen University and Research Centre, P.O. Box 16, 6700 AA Wageningen, The Netherlands
e-mail: marco.bink@wur.nl

E. van de Weg
Wageningen UR Plant Breeding, Wageningen University and Research Centre, P.O. Box 386, 6700 AJ Wageningen, The Netherlands
e-mail: eric.vandeweg@wur.nl

T. R. Famula
Department of Animal Science, University of California, Davis, CA 95616, USA
e-mail: trfamula@ucdavis.edu

T. J. Frett · K. Gasic
Department of Agricultural and Environmental Sciences, Clemson University, Clemson, SC 29634, USA
e-mail: tjfrett@email.uark.edu

K. Gasic
e-mail: kgasic@clemson.edu

Published online: 28 July 2015

previously reported in peach. The inclusion of introgressed germplasm and the explicit declaration of the genetic structure of the pedigree as covariate in PBA enhanced the mapping and interpretation of QTLs. This study serves as a model study for PBA in a diverse peach breeding program, and the results highlight the ability of this strategy to identify genomic resources for direct utilization in marker-assisted breeding.

Keywords *Prunus persica* (L.) Batsch · Germplasm introgression · Bayesian · SNPs · Pedigree correction · Genetic structure

Introduction

Peach [*Prunus persica* (L.) Batsch] is a self-compatible fruit tree crop that exhibits high phenotypic variability yet suffers from restricted genetic diversity. The restricted genetic diversity is the result of significant genetic bottlenecks which occurred during domestication, as well as the extensive inbreeding used to develop most of today's modern European and American cultivars (Scorza et al. 1985; Font i Forcada et al. 2012). Most peach cultivars developed during the last 200 years were primarily bred using mass selection and intraspecific hybridization (Okie et al. 2008).

Recently, the introgression of valuable alleles from related species has been pursued to bring in new alleles and increase genetic diversity (Martinez-Gomez et al. 2003). However, introgressions are often associated with linkage drag, which may jeopardize the integrity of the desirable traits of the commercial crop species (Flint-Garcia 2013). Despite this drag, breeding progress has been made through the procurement, identification and incorporation of accessions with unique trait variants. Thus, novel phenotypes have resulted from genetic introgression of unique alleles from introductions collected from the origins of domesticated peach (China, Japan and Korea) as well as introgressions from related species (Gradziel 2002, 2003).

Information on the genetic control of complex traits has been pursed through quantitative trait loci (QTLs) analysis of biparental populations resulting from crosses between peach and related species (see Olukolu and Kole 2012, for a summary). QTL analysis of biparental introgression populations of peach can be limited, however, by patterns of inheritance which are often progeny specific (T. Gradziel, pers. comm.) and which may restrict the inferences about the penetrance, breeding value and architecture of novel traits to stack in breeding germplasm. A promising approach utilizes the pedigrees of breeding selections to uncover the full extent of genetic (QTL) variation of within a breeding program, while enabling the assessment of a wider genetic base to capture additional genetic information. Such an approach is addressed through the pedigree-based analysis (PBA) under the Bayesian framework (Bink et al. 2008, 2014).

PBA performs QTL discovery and subsequent characterization by utilizing the pedigrees of breeding programs (Iezzoni 2010; Peace et al. 2010), as pursued in the RosBREED initiative (www.robreed.org). The identification and quantification of QTLs through PBA account for those chromosomal segments shared in identity by descent (IBD) through multiple (small or large) families with known pedigree and genotypic data and also connected to recent common ancestors. This approach empowers a higher degree of certainty, even for those genetic factors with only moderate effects (Jannink et al. 2001). In addition, it facilitates the estimation of genetic parameters such as heritability and breeding values (Bink et al. 2008, 2014). The estimation of these parameters is particularly relevant for breeding decisions such as crossing design as it enables breeders to simultaneously consider introgressed loci from different genetic backgrounds.

PBA is facilitated through the dedicated software FlexQTL^{TM}, which uses the IBD concept to track QTL and marker alleles and uses a Bayesian framework to map QTLs as well as estimate the genomic breeding values of germplasm generated from single and reciprocal test crosses to complex pedigree structures (van de Weg et al. 2005; Bink et al. 2008, 2014). Results provide insights into each individual's genetic merit, including its capability to pass on additive genetic effects

Present Address:
T. J. Frett
Horticulture Department, University of Arkansas, Fayetteville, AR 72701, USA

C. P. Peace
Department of Horticulture, Washington State University, Pullman, WA 99164, USA
e-mail: cpeace@wsu.edu

influencing a given trait. Thus, it allows exploitation of the full genetic potential contained in a breeding program. While PBA has been successfully applied in sweet cherry (Rosyara et al. 2013) and apple (Guan et al. 2015), no studies have been reported for peach.

The breeding program at UC Davis uses directed introgression from different species to expand genetic diversity in peach (Gradziel 2002, 2003). Although pedigree records are available, many of the earlier generations no longer exist. Several studies on QTL have been performed in progenies of peach and relative species. However, in spite of its narrow genetic base, studies using the pedigree of a peach breeding germplasm including introgressions of related species have not yet been pursued.

This study thus represents the first application of PBA under a Bayesian framework to analyze and identify QTLs for peach. Eight traits were investigated over two consecutive years to uncover useful QTLs within the diverse germplasm utilized. In addition, this study extends previous PBA approaches by determining the genetic structure within the breeding germplasm analyzed.

Methods

Plant material

The breeding germplasm investigated was provided by the UC Davis Processing Peach Breeding Program which included 553 individuals. This germplasm encompassed old and modern cultivars, pure peach selections as well as breeding selections possessing introgressions from related species such as almond [*Prunus dulcis* (Mill.) D.A.Webb], *Prunus argentea* (Lam.) Rehder, *Prunus davidiana* (Carrière) Franch and *Prunus mira* Koehne. The germplasm investigated is part of the US reference set for peach germplasm (Peace et al. 2014) of the RosBREED project (Iezzoni 2010; Iezzoni et al. 2010) and represents the portion of RosBREED germplasm developed for improved processing quality (Gradziel et al. 1993).

Pedigree verification and correction

Pedigree verification and correction was performed through the software FRANz 2.0 (Riester et al. 2009). This software was used to interrogate the relationships of 409 individuals (out of the 553 individuals in the UC Davis germplasm) possessing, in variable extent, genotypic information from 490 SNPs and five SSRs. The parameters used were 500 000 burn-in iterations, three million MCMC iterations, a sample frequency of 10, a $(MC)^3$ swap frequency of 25, a $(MC)^3$ number of swaps of eight (number of threads in microprocessor) and a $(MC)^3$ temperature of 0.2. Correction of parentages was performed for those accessions in which the probability of a parentage distinct from that previously known was accounted above 0.95 and was supported by breeder's expertise (trees are located close each other, crossing or harvested seed labeling happening on same dates, etc.). Prior information about time of introduction for particular accessions, the parental role of each accession (female/male), known relationships in portions of the germplasm and locations of the individuals was provided. The final output was compared with the empirically known relationships within the germplasm.

Pedigree identification and phenotypic and genotypic information

The most informative pedigree, i.e., the set of individuals with most recorded filial relationships along several generations, as well as phenotypic and genotypic data, was determined through ordering and trimming using functions available in the package 'Pedigree' (Coster 2012) for R 2.15.3. Finally, pedigree charts were constructed through PediMap 1.2 (Voorrips et al. 2012).

The genotypic information consisted of 215 genome-wide biallelic SNPs not exhibiting missing data and low rate of genotypic errors which were identified during the development of the IPSC peach 9K SNP array v1 (Verde et al. 2012).

The phenotypic traits were evaluated for individuals located at Davis, California, for two consecutive years (2011 and 2012). The traits were days to bloom (DTB); fruit diameter (FD) in millimeters; fruit development period (FDP) in Julian days; fruit weight (FW) in grams; Julian days to fruit maturity (RD); fruit pH; soluble solids concentration (SSC) in degrees Brix and titratable acidity (TA) expressed as milliequivalents of malic acid content. The eight traits exhibited normally distributed phenotypes and residuals across the entire phenotyped germplasm. The number of records varied from year to year with the

traits pH, SSC and TA being the most unbalanced, ranging from approximately 68 % to approximately 86 % of the total number of individuals available for phenotyping. For the remaining traits, more than 190 phenotypic records were available for each of the two years (Frett et al. 2012).

Genetic structure

The set of 258 individuals containing complete genotypic and phenotypic information was used for the characterization of genetic structure through factor analysis for mixed data (FAMD) (Abascal et al. 2006) with the use of the library FactoMineR 1.25 (Le et al. 2008) implemented in the R package version 2.15.3 (R Development Core Team 2012). Three coordinate dimensions were requested for the 3D scatterplot of the results. FAMD was applied to keep the original scales and nucleotide allelic conformations (which were considered as categorical data). This method allows the combining of data with both continuous (phenotypic traits) and categorical (molecular markers) variables while giving equal weight to both variables in determining the dimensions of variability without making assumptions on the genetic background of the samples (e.g., linkage equilibrium, drift, admixture, isolation by distance).

Results from analysis of the genetic structure of the germplasm were used to assign each of the 258 individuals to one of three groups distinguished in the analysis. This assignment was recorded as a nuisance variable (group) that was subsequently integrated into the data file entered in FlexQTL^{TM} (Bink et al. 2008, 2014) in the form of a vector indicating the cluster membership per individual determined through FAMD. With the 258 individuals included in the FAMD as a base, a pedigree of 355 individuals providing complete genetic information at the pedigree level was entered into FlexQTL^{TM} for mapping of genetic components influencing the eight traits studied. The inclusion of the genetic structure was intended as a measure to add information about the origin of alleles in related germplasm within the pedigree, because the related species used cannot be tracked to previous generations, thus preventing problems of mixing of the Markov chains.

Molecular markers

Two hundred and fifteen biallelic markers were used for this study because they showed no missing data for the genotyped progenies and furthermore showed $\leq$5 % genotypic errors and no major deviations from Mendelian segregation ratios (with $p \leq 0.01$). These markers were also polymorphic and showed 0.15 minor allele frequencies. Markers were previously used for the calculation of genomic realized numerical relationship matrices for the estimation of breeding values in the UC Davis germplasm. Thus, the genetic map used for the location of the QTLs per trait consisted of 215 SNP markers, spread over eight linkage groups (G) representing the eight peach chromosomes, and covering a total genome-wide genetic distance of 692 cM with an average distance between markers of 3.34 cM (Online Resource 1 Fig. S1). The longest observed gap was 36.5 cM between markers ss_3621 and ss_31469 on G1. The highest number of SNP markers, 64, was located on G2, and the lowest number was 10, located on G5.

Bayesian QTL mapping

Through the application of FlexQTL^{TM} (Bink et al. 2014), the eight traits considered in this study were analyzed individually and on a year to year basis through the implementation of the linear model described by Bink et al. (2014). Such a linear model consists of an intercept, which is the phenotypic mean of the trait being analyzed, a design matrix and a vector of genetic group effects, a design matrix of a vector of regressions on the QTL covariates and a model residual error. This model was evaluated through Bayesian modeling following the Markov chain Monte Carlo (MCMC) algorithms described by Bink et al. (2014). The number of QTL was considered a random variable, and the assignation of priors per vector and variances was done as described by Bink et al. (2014).

Preliminary MCMC simulations were performed to identify the most convergent and stable genetic model for explaining each trait for 2011 data. These preliminary simulations consisted of three MCMC chains of one million iterations for each prior number of QTLs, which were: 1, 3 and 5. Thus, nine simulations were performed per trait to assess sensitivity of posterior inferences to prior assumptions. At least 100 effective chain samples (ECS, Sorensen and Gianola 2002) for phenotypic mean, residual variance and number of QTLs were required to draw sound inferences and conclusions. The required lengths of the Markov chain

simulations varied among traits as well as among priors on the number of QTL with a range between 300 000 up and 3 000 000 iterations. To save disk space and to reduce auto-correlation among samples, thinning was applied and, for each simulation, 1000 samples were stored and thus available for statistical inference.

A pairwise comparison of models differing by one QTL from each other was used to infer the number of QTL. Twice the natural logarithm of Bayes factors (BF, Kass and Raftery 1995), denoted $2ln(BF)$, was employed as it allowed easier interpretation since this transformed statistic has a similar scale to the likelihood ratio test. Values for $2ln(BF)$ that are greater than 2, 5 and 10 indicate positive, strong and decisive evidence, respectively, for favoring the larger QTL model. The main criteria for the determination of major QTLs per trait included: explanation of at least 10 % of the phenotypic variation, the exhibition of the QTL with at least strong evidence for both years of evaluation on the same linkage group (G) and co-localization within ±25 cM for identified regions for both years.

Inferences on QTL positions, QTL contributions and posterior probabilities of QTL genotypes were estimated using same thresholds as described by Bink et al. (2008, 2014).

Values for broad-sense heritability (H^2) and narrow-sense heritability (h^2) were calculated from the values of phenotypic variance and the variance of the residual error for each trait and the weighted additive variance of the trait for the genomic region with strong evidence of being a QTL, as reported in the output from FlexQTL™. Note that in this study, the genetic variance is composed of the variance given by the genetic structure plus the weighted additive variance of a given trait.

The QTL genotype probabilities, together with QTL intensity and QTL effect sizes, were also used to predict genomic breeding values (GBVs), as the product of the individual's QTL genotype probability, the QTL allele effect and the QTL intensity (Bink et al. 2008, 2014). Thus, an aggregate genomic breeding value per individual was obtained by the summation of the positional breeding values along the genome.

Finally, identity by descent (IBD) probabilities were estimated using all marker information and pedigree data. Haplotypes for individuals in the pedigrees were estimated using the marker linkage phase information combined with the QTL alleles. In one case, five SNPs were used for an analysis of QTL transmission based on IBD probabilities through pedigree as performed in FlexQTL™.

The assignment of names to identified QTLs followed the conventions for the Genome Database for Rosaceae (Jung et al. 2008, 2014).

Results

Pedigree correction

The pedigree correction identified 64 individuals (~20 %, Online Resource 1 Table S1) with incorrect parentage, 19 of which had one parent not present in the materials genotyped, thus making their pedigrees incomplete. Several accessions with corrected parentage originated from outcrosses. Overall, the parentages of 45 individuals were corrected, which included one breeding selection and four cultivars ('Halford', 'Hesse', 'Rizzi' and 'Woltemade'), which are presented in the Online Resource 1 Table S1 (cultivars in italics).

Pedigree identification

The pedigree consisted of 11 small pedigreed breeding families (numbers of progenies varied from 3 to 35), 67 founders, 53 commercial cultivars (four almond cultivars and 49 peach cultivars with some cultivars also considered as founders), three plant introductions ('Bolinha', 'Bolinha 6' and 'Yumyeong'), three related species (almond, *P. argentea* and *P. mira*), eight interspecific hybrids (*P. persica* × *P. argentea*, *P. persica* × *P. davidiana*, *P. persica* × *P. dulcis* and *P. persica* × *P. mira*) and 262 breeding selections. The selected germplasm is shown in Online Resource 1 Fig. S2A, while the lineages with related *Prunus* species are highlighted in Online Resource Fig. S2B, Fig. S2C, Fig. S2D and Fig. S2E.

Genetic structure

The genetic structure identified in the pedigree of 355 individuals, including peach and almond cultivars, interspecific hybrids and introgression breeding lines, showed that the clustering of the germplasm is related to 'stone-adhesion/flesh-texture' (Online Resource 1

Fig. S3), but was not completely driven by these characteristics. In the first two coordinate dimensions (Fig. S3A), one cluster included clingstone-non-melting accessions (group 1), while the other cluster included freestone-melting accessions (group 2). By graphing three coordinate dimensions, it was possible to distinguish the outlying and distinct genetic pool of 'Yumyeong' (Fig. S3B), as well as the small group of almond accessions (group 3, Fig. S3C). With the addition of information from markers and phenotypic data (Fig. S3D), it was possible to determine that group 1 was strongly influenced by the clingstone-non-melting trait, while group 2 did not exhibit as strong an influence, since the categories 'freestone' and 'melting' were located between groups 1 and 2.

Mapped QTLs

Table 1 shows the results of QTL mapping for the eight traits evaluated during 2011 and 2012. Evidence for 52 QTLs was estimated as positive [$2ln(BF) > 2$]. Several QTLs were not mapped in consecutive years, and the number of QTLs per trait varied from one year to the next, with the traits FD, FW, pH, SSC and TA exhibiting the most variation from year to year.

For most traits, the prior number of QTLs that gave the most stable models ranged from one to three. The broad-sense heritability (H^2) values were above 0.35, with the three lowest being for a phenology-related trait, FDP-2012 (0.36), and the fruit biochemistry traits, SSC-2011 (0.39), pH-2011 (0.44) and TA-2011 (0.61). Known complex traits such as FD and FW had H^2 values between 0.72 and 0.99, and phenological traits such as DTB and RD showed H^2 values above 0.90.

For seven traits, a total of 10 QTLs were mapped in 2011 (Fig. 1), while 13 QTLs were mapped for eight traits in 2012 (Fig. 2), with all QTLs showing at least strong evidence [$2ln(BF) > 5$]. Six QTLs were mapped for six traits with strong to decisive evidence for both years of phenotypic evaluation and were considered as major QTLs contributing to the phenotypic exhibition of their respective traits. These QTLs are summarized in Table 2 and named according to the conventions of the Genome Database for Rosaceae (Jung et al. 2008, 2014): linkage group (G) number and the symbol for the trait(s). Major QTLs include *G3FD* (although 2011 results had a different location than 2012), *G2FW*; *G5pH*; *G1RD;* and *G4RD* (where *G4RD* had a greater contribution to the phenotypic variance).

Fruit size (FD and FW)

For the analysis of traits in 2011 (Fig. 1), a very promising QTL for FW (*G2FW*) was located near the mid-portion of the linkage group (G) 2. One QTL for FD was identified at the mid-point of G3, (*G3FD*), which spanned ~10 Mb. In 2012 (Fig. 2), *G2FW* again showed decisive evidence for a co-localized region on G2 similar to in 2011; however, the trace plot did not show only a single simulation chain. *G3FD* was located in the upper region in comparison with 2011. *G3FD* showed high values of h^2 in both years (0.93 and 0.98 in 2011 and 2012, respectively). While the locus was not exactly determined in either year, the QTL was located within five SNPs (ss_316025, ss_320900, ss_339562, ss_341291 and ss_345419). An additional QTL for FD was located on G7 having evidence of 2.90 in 2011 and 4.27 in 2012.

In addition, there were some QTLs showing a similar degree of evidence over the two years. One QTL for FD on G2 was located upstream of *G2FW,* with an average evidence of 2.27 in 2011 and 2.97 in 2012, thus being positive in both years.

Phenological traits (DTB, FDP and RD)

For QTLs mapped in 2011 (Fig. 1), two were placed on G1: *q*RD.1 near the top and *q*FDP.1 at the bottom with both separated by 10.5 Mb, although *q*RD.1 exhibited a dispersed trace plot. Major QTLs for the phenological traits DTB and RD (*G4DTB* and *G4RD*, respectively) were located directly below the middle of G4. The top of G6 was a truncated region that contained a QTL for FDP (*q*FDP.6). A QTL for FDP (*q*FDP.8) was also located directly above the middle of G8 within a gap of ~3.5 Mb.

In 2012 (Fig. 2), two QTLs for RD (*q*RD.1.1 and *q*RD.1.2) were found on the upper part of G1. Similarly, the presence of two QTLs for DTB (*q*DTB1.1 and *q*DTB.2, respectively) was also identified. Thus, *q*RD.1.1 was flanked by *q*DTB1.1 and *q*DTB.2, respectively, while *q*DTB.1.2 was flanked by *q*RD.1.1 and *q*RD.1.2. Supporting evidence for *q*FDP.1, detected as strong in 2011, was only positive in 2012. A QTL for DTB was located on G3 (*q*DTB.3), which extended in a region including *G3FD* and a QTL for TA (*q*TA.3). Evidence for *q*FDP.6 was only positive in 2012. Finally, the major QTLs for DTB and RD located in G4 were basically the same in 2011 and 2012.

Table 1 QTL mapped for the eight traits evaluated during 2011 and 2012

Trait	Year	MCMC run length	Records	μ_P	σ_P^2	σ_G^2	H^2	Linkage group	Evidence [$2ln$(BF)]
DTB	2011	1,000,000	193	172.16	333.65	315.99	0.95	4	5.10
								1	4.10
								3	2.70
								8	2.53
	2012	650,000	191	171.27	347.25	328.51	0.95	4	14.57
								1	10.50
FD	2011	500,000	227	73.20	156.81	148.97	0.95	3	31.90
								7	2.90
								2	2.27
	2012	1,000,000	226	62.13	120.75	119.54	0.99	3	27.20
								7	4.27
								2	2.97
								1	2.03
FDP	2011	2,000,000	202	61.28	6.03	5.34	0.88	1	28.10
								8	18.87
								6	13.47
	2012	500,000	192	55.48	77.82	28.17	0.36	3	3.83
								1	2.53
								6	2.27
FW	2011	500,000	232	164.68	4894.88	3681.94	0.75	2	24.17
	2012	2,000,000	222	120.64	2583.09	1860.57	0.72	2	15.37
								5	4.17
								4	2.20
pH	2011	250,000	225	3.95	0.04	0.02	0.44	5	33.03
	2012	500,000	171	4.02	0.05	0.04	0.70	5	6.63
								6	4.90
								1	2.00
RD	2011	3,000,000	233	232.16	341.79	317.01	0.93	4	7.63
								1	5.37
								3	2.87
								5	2.77
								7	2.63
	2012	2,000,000	224	226.41	386.44	355.8	0.92	4	5.80
								1	4.50
SSC	2011	1,000,000	222	13.92	6.42	2.47	0.39	7	7.80
								2	2.13
								1	2.03
	2012	300,000	174	14.85	3.26	1.87	0.57	6	4.47
								3	2.80
								5	2.33
								1	2.10
								2	2.00

Table 1 continued

Trait	Year	MCMC run length	Records	μ_P	σ_P^2	σ_G^2	H^2	Linkage group	Evidence [$2ln$(BF)]
TA	2011	3,000,000	223	0.61	0.05	0.03	0.61	1	3.93
								7	2.80
								5	2.33
	2012	3,000,000	176	0.52	0.04	0.03	0.85	3	30.60
								1	5.43
								5	3.83
								2	2.87

Phenotypic mean (μ_P), phenotypic variance (σ_P^2), genetic variance (σ_G^2) and broad-sense heritability (H^2) values per trait, per year are shown, as well as the linkage groups on which the QTLs were mapped and the evidence for each QTL. The evidence value is the average of the $2ln$(BF) values from three distinct runs per trait per year

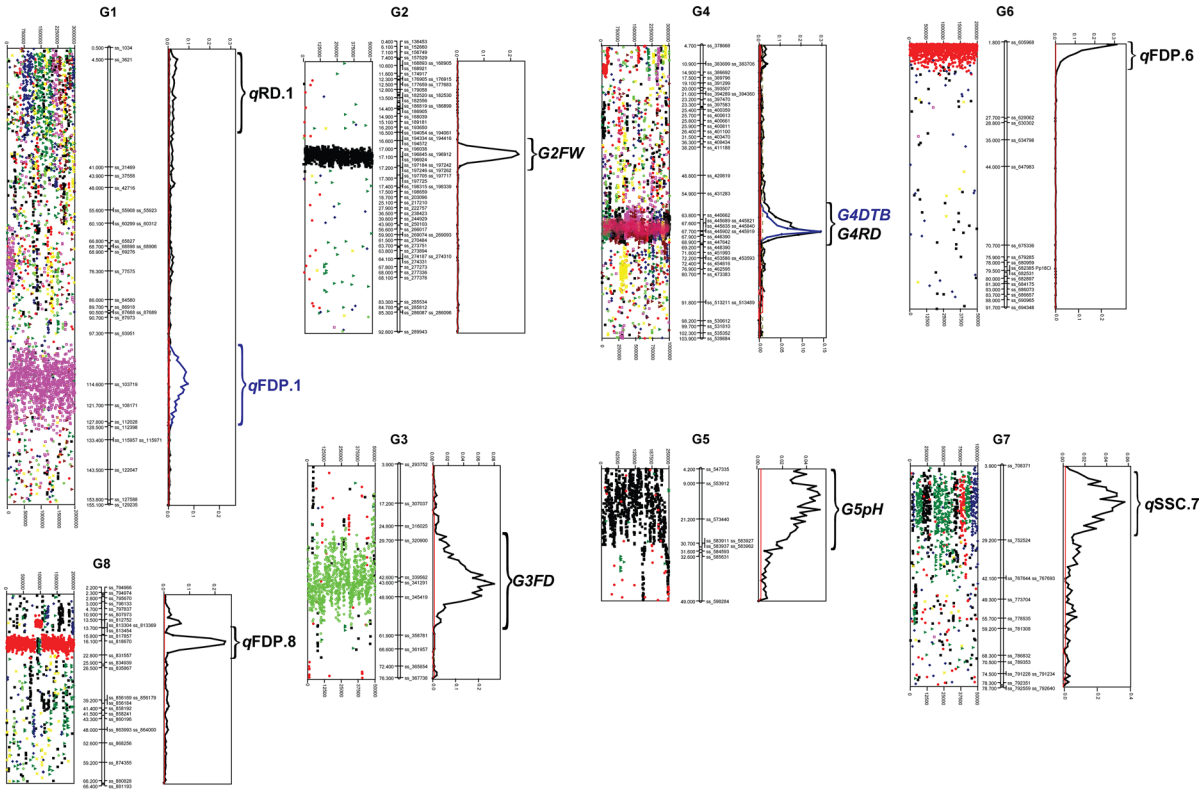

Fig. 1 The QTLs exhibiting strong evidence [$2ln$(BF) > 5] for the year 2011. The linkage groups (G) are shown with a trace plot on the *left* and an intensity plot on the *right*. Linkage map compositions were generated through MapChart version 2.2 (Voorrips 2002)

Biochemical traits (pH, TA and SSC)

In 2011 (Fig. 1), a major QTL for pH (*G5pH*) spanned ~7.2 Mb of the first half of G5. QTLs for SSC (*q*SSC.7) located in a gap region of ~7 Mb at the top of G7. A QTL for TA, *q*TA.1, was identified on the lower part of G1. A QTL with strong evidence for SSC (*q*SSC.7) was located in the upper region of G7, between the markers ss_708371 (1,125,816) and ss_752524 (8,336,521). However, no QTLs with

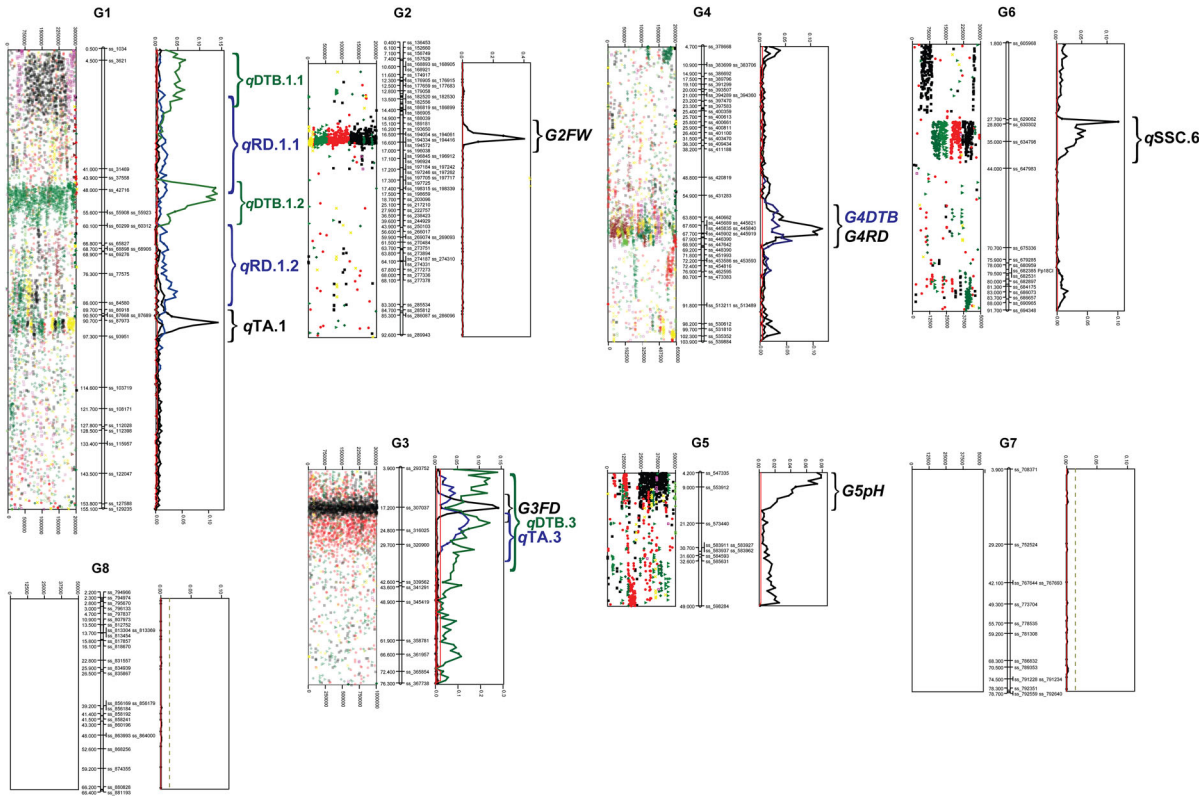

Fig. 2 The QTLs exhibiting strong evidence [$2ln(BF) > 5$] for the year 2012. The linkage groups (G) are shown with a trace plot on the *left* and an intensity plot on the *right*. Linkage map compositions were generated through MapChart version 2.2 (Voorrips 2002)

strong evidence were mapped for TA, showing regions on G1, G7 and G5 with positive evidence only.

In 2012 (Fig. 2), the *G5pH* QTL was limited to a narrower region, showing high association at the bottom of G5. A QTL for SSC (*q*SSC.6) was located close to the mid-point of G6, between the markers ss_629062 (7,918,349) and ss_630302 (12,571,791), a region of ~4.5 Mb in length. In addition, *q*SSC.7 mapped in 2011 did not show any evidence in 2012. The upper middle portion of G3 showed decisive evidence for a QTL for TA (*q*TA.3), while an additional QTL for TA was found on G1, which exhibited strong evidence.

A QTL for SSC showed positive evidence in 2011 and 2012 (2.13 and 2.00, respectively) and was located at the bottom of G2, although plots showed it was close to a gap of ~4.2 Mb between the markers ss_277273 and ss_285534. An additional QTL for SSC with positive evidence in both years (2.03 in 2011 and 2.10 in 2012) was mapped on G1. However, in 2011, this QTL was identified near the bottom, and in 2012, it was closer to the middle of G1. Interestingly, *q*TA.1, which exhibited positive evidence in 2011 (3.93) and strong evidence in 2012 (5.43), showed the same locations and same 'movement' as the QTLs for SSC. Also on G1, a 2012 QTL for pH was identified close to the QTLs for SSC (with barely positive evidence) and *q*TA.1. On G5, a QTL for TA with positive evidence over both years (2.33 in 2011 and 2.87 in 2012) was located at the same location as *G5pH*.

Genomic breeding values

The GBVs were calculated for all individuals at the pedigree level per trait per year (Online Resource 2). In general, GBVs showed high positive correlations with the observed phenotypes for all traits per year, suggesting high accuracy. In 2011, pH exhibited the lowest accuracy, 0.69, followed by SSC with 0.74, FD with 0.85, TA with 0.87, FW with 0.90, DTB with 0.91, RD with 0.96 and FDP with 0.99. In 2012, FDP

Table 2 The QTLs with consistently strong to decisive evidence across two years

QTL	Year	Prior N_{QTL}	σ_P^2	σ_e^2	σ_A^2	h^2	Flanking SNPs	Physical position	*QQ* Genotype	SNP name	Physical position
G4DTB	2011	3	333.65	17.65	70.54	0.21	ss_431283	15 694 219	CC	ss_431283	15 694 219
	2012	3	347.25	18.75	83.80	0.24	ss_447642	19 688 796	GG	ss_445689	19 321 509
									AA	ss_445919	19 329 444
									AA	ss_446390	19 397 033
G3FD	2011	1	156.81	39.59	146.30	0.93	ss_307037	4 903 399	AA	ss_341291	12 454 268
	2012	1	120.75	25.38	118.20	0.98	ss_345419	13 967 230	GG	ss_345419	13 967 230
G2FW	2011	1	4894.88	1212.94	2738.70	0.56	ss_196038	4 855 952	AA	ss_196038	4 855 952
	2012	1	2583.09	722.53	1313.78	0.51	ss_196912	4 895 580	GG	ss_196845	4 893 020
G5pH	2011	1	0.04	0.02	0.04	0.85	ss_547335	1 203 296	GG	ss_547335	1 203 296
	2012	1	0.05	0.02	0.03	0.50	ss_573440	6 070 663	AA	ss_553912	2 580 963
									GG	ss_573440	6 070 663
G1RD	2011	3	341.79	24.77	145.48	0.43	ss_1034	155 117	GG	ss_1034	155 117
	2012	3	386.44	30.61	89.22	0.23	ss_31469	11 708 514	GG	ss_31469	11 708 514
G4RD	2011	3	341.79	24.77	97.11	0.28	ss_445689	19 321 509	AA	ss_445689	19 321 509
	2012	3	386.44	30.61	101.15	0.26	ss_445840	19 327 740	GG	ss_445821	19 327 423
									AA	ss_445840	19 327 740

Details of each QTL are listed including year of evaluation, prior used for the minimum number of QTLs (N_{QTL}) allowed in the model, phenotypic variance $\left(\sigma_P^2\right)$, residual genetic variance $\left(\sigma_e^2\right)$, additive genetic variance $\left(\sigma_A^2\right)$, narrow-sense heritability (h^2), as well the QTLs flanking SNP makers and their physical positions. The allelic combinations of SNP markers in coupling phase to the positive QTL allele (*QQ*) are additionally, shown, with their corresponding names and physical positions

exhibited the lowest accuracy of 0.76, followed by SSC with 0.83, FW with 0.86, pH with 0.90, FD with 0.92, RD, TA with 0.95 and DTB with 0.98. Note that the accuracy for FDP decreased from 2011 to 2012 (shown as the negative value of -0.19 in Online Resource 2 Fig. S4), and although from 2011 to 2012, DTB showed an increased correlation between GBVs and observed phenotypes, the correlation between 2011 and 2012 GBVs was negative (-0.36). For the remaining traits, correlations between 2011 and 2012 GBVs were positive and high with the exception of SSC, with a correlation of 0.31.

Discussion

Pedigree correction

Several accessions (~20 %) were identified to have an incorrect parentage record and thus originated from selfing or outcrossing. In contrast to the expectation of high rates of unintended self-fertilization prior to emasculation and controlled crossing, owing to the tendency in peach for self-fruitfulness and occasional cleistogamy (Gradziel and Weinbaum 1999), the high outcrossing occurrence rate indicates that pollen contamination may have occurred. The interspecies origin of many accessions may have contributed to these pollination errors. For example, lower pollen fecundity has recently been identified in the selection '2001_7_180', an interspecific hybrid between peach (cv. 'Andross') and *P. argentea*, which may have reduced the intended self-fertilizations to obtain F_2 progeny. Similarly, while peach flowers are only occasionally visited by insect pollinizers, flowers of the introgressed genotypes often show stronger attraction to pollinizers which would promote unintended outcrossing. For example, unlike peach, almond nectar is low in astringent cyanoglucosides and, as a consequence, nectar-collecting insects are much more commonly observed visiting even previously emasculated flowers. In most cases, the corrected parentage was consistent with the physical proximity of the proposed male parent in the breeding block.

Genetic structure

The identified genetic structure shows that the clustering of the germplasm was influenced by 'stone-adhesion/flesh-texture' characteristics which is consistent with previous reports by Aranzana et al. (2010) for non-melting and melting-flesh peach cultivars, with the addition of a cluster for the almond-related germplasm, in which the genetic structure is mainly driven by selection of nut and kernel characteristics (Zeinalabedini et al. 2012). The peach breeding program at UC Davis emphasizes clingstone-non-melting materials for canning, resulting in less differentiation, in comparison with the more diverse sets of phenotypes analyzed by Aranzana et al. (2010, 2012) and Cao et al. (2012). Interestingly, introgression selections with little almond resemblance clustered with peach, possibly a result of strong artificial selection for peach development types.

The information content of our marker set was limited due to germplasm introgression from related species in our pedigree. This limited the use of biallelic markers without missing data because some SNPs did not amplify, despite the fact that during the development and validation of the IPSC peach 9K SNP array, few almond and interspecific hybrids were considered (Verde et al. 2012). Furthermore, marker information was generally absent for individuals in the earlier generations of the pedigree due to the non-availability of plant material (DNA). This hampered the accurate tracing of marker alleles from mapping progeny to founder individuals which consequently resulted in less distinct identity by descent (IBD) probabilities. Chromosome regions with low marker information content may also give rise to spurious QTLs as there is an opportunity for segregation patterns of QTL alleles to correlate with phenotypic variation.

The addition of the genetic structure as a nuisance variable in the genetic model helped to prevent false-positive QTLs that may arise from the substructure among founders when different species are involved, even when germplasm pools are closely related (Calboli et al. 2008). This is particularly important when species are closely related but have divergent reproductive contexts which may influence genomic features such as the extent of linkage disequilibrium (LD) (Charlesworth and Charlesworth 1979). LD in self-fruiting peach is long (Aranzana et al. 2010) but in almond is expected to be short due to occurrence of self-sterility via gametophytic self-incompatibility. Such differences in genomic features may result in differentiated recombination rates (Arús et al. 2010;

Tanksley et al. 1992), particularly where introgression occurred recently, as it may affect the ability to predict whether individuals within a diverse pedigree are carrying same alleles because of IBD.

In addition, the declaration of genetic structure reduced the number of iterations needed for convergence and stability of the genetic models proposed through FlexQTL™. For example, the analysis of TA in 2011 required five million iterations for convergence when the genetic structure was not considered. When genetic structure was considered as a nuisance variable, the analysis of TA required only three million iterations. This outcome suggests that use of the vector of membership in the genetic model as an indicator of genetic structure worked as a meiosis indicator or descent indicator to prevent irreducibility of the Markov chain (Cannings and Sheehan 2002). Complex pedigree structures may present problems for mixing of the chains in the MCMC (Lee and van der Werf 2005), consequently affecting the performance of the reversible-jump method implemented in FlexQTL™ and thus complicating the convergence needed to make sound statistical inferences.

Mapped QTLs

Eight traits were investigated within a representative pedigree of the Processing Peach Breeding Program at UC Davis, which includes genetic introgression from related species including almond, *P. argentea*, *P. davidiana* and *P. mira*.

In QTL studies using biparental populations, the size of the progeny is critical to the detection of minor QTLs effect size given the limited number of recombination events in the progeny. In this PBA study, the average family size was 13 individuals, with the largest being 35 individuals from the self-pollination of '2000_16_133' and the second largest being 29 individuals from a cross between 'Dr. Davis'(♀) and 'D62_193'(♂). However, the PBA approach takes advantage of combining data from multiple families. This was accounted for in the design where care was taken for a balanced allelic representation of important breeding parents captured in the RosBREED peach breeding crop reference set (Peace et al. 2014). Thus, QTL-segregating family sizes ranged effectively between 10 and 50 progenies, which were interrogated simultaneously for QTLs for each trait. The detection of modest QTL effects is possible as a result of the increased number of individuals that can be considered in the breeding germplasm.

PBA allows analysis of several IBD-connected variable-sized families simultaneously, so that a large number of alleles are considered and fitted in the genetic models through the Bayesian framework under PBA. Because the number of meiosis considered is similar to that captured in a large biparental family, the statistical power is maintained.

In addition, the lack of phenotypic and especially genotypic data for some ancestors is compensated through the well-tracked pedigree connections achieved through the pedigree correction performed here, as well as the determination of genetic structure and the design of the US reference set. Thus, in this study, genotypic data were available for 258 individuals out of 355 individuals in the pedigree (72.9 %), and phenotypic data for 74.8 to 90.3 % for 2011 and 66.3 to 88 % for 2012, ensuring that all the phenotyped germplasm as well as their parents possessed genotypic information. Having genotypic information of individuals from earlier generations in the pedigree enables the tracking by IBD of genetic factors of importance within breeding germplasm through PBA regardless of the availability of phenotypic data of individuals in earlier generations (Peace et al. 2014).

The genetic linkage map density in this study was low, with some linkage groups exhibiting gaps greater than 20 cM. This was, in part, a consequence of inclusion of related species in the pedigree as it limited the number of fully informative SNP markers across the whole pedigree. These markers were also used in a previous study for the calculation of genomic realized numerical relationship matrices for the estimation of breeding values, where the procedures applied do not allow missing data. Regardless of the low-density marker data, the chosen markers possessed a high degree of certainty in their traceability across the studied pedigree. Thus, the 215 SNPs used were fully informative for our germplasm. Although environmental fluctuations between years could have affected trait exhibition and therefore QTL detection, only small negative deviation from the normal climatological average temperature (−0.83 and −0.36 °C, respectively) occurred during the fruit development period (March to September) in both 2011 and 2012. Moreover, only a slight positive difference was recorded from the normal climatological precipitation records during the fruit development period (+22.04

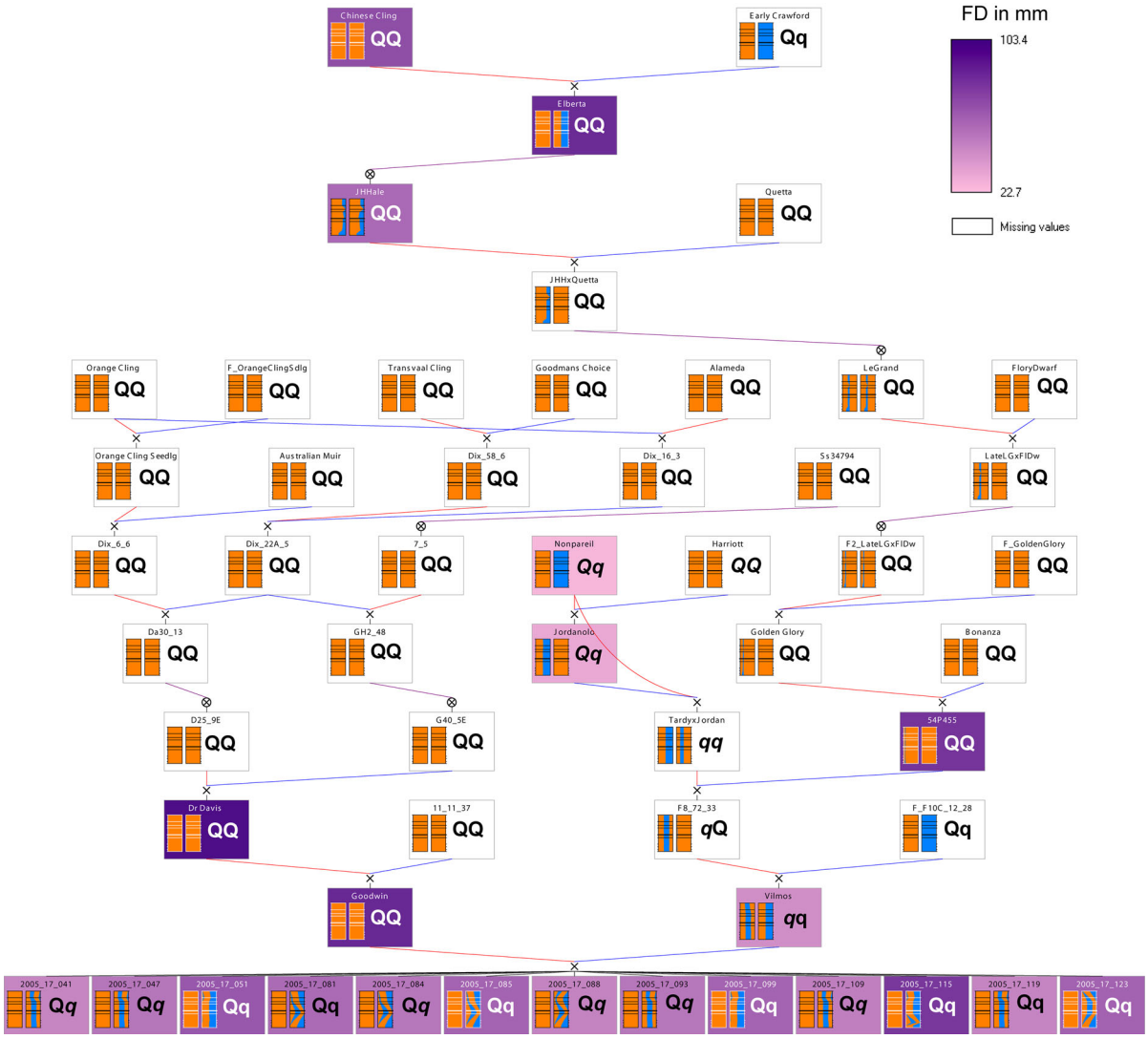

Fig. 3 Prediction of the segregation of *G3FD* based on genotype probabilities with positive evidence calculated through FlexQTL™. The *rectangles* on the *left of each box* are graphical representations of each allele based on probabilities given by the proportion of the *rectangle* filled with a certain *color*, in this case *orange* for Q and *blue* for q. Note that Q and q do not specify any dominance relationship. Also, note that the male parent of the final full sib family ('Vilmos') has almond background carrying the negative allele (*q*) from an almond source. The original q allele from the peach pedigree was lost since 'Early Crawford' passed on its Q allele. Figure generated through PediMap 1.2 (Voorrips et al. 2012). (Color figure online)

and +2.82 mm, respectively) in both years. Despite these fluctuations, we were able to identify statistically well-supported QTLs for at least five of the eight traits in both years, providing a higher degree of certainty in comparison with a QTL analysis which only included the average phenotype across both years of study. These detected QTLs were in agreement with chromosomal locations described previously in peach (discussed below). A priori, these QTLs are of immediate relevance for the germplasm studied here, because the allelic diversity considered represents the breeding parents for the UC Davis processing peach breeding program, which uses interspecific introgression from almond and relative species, and exemplifies the situation discussed by Peace et al. (2014). Thus, several small families are considered for QTL detection, but these families are restricted to represent distinct breeding parent alleles (such as those used for

the breeding of processing vs. fresh consumption cultivars). The discussion of QTLs depicting these traits is presented based on groupings for fruit size and appearance as well as phenological and biochemical characteristics.

Fruit size (FD and FW)

PBA identified *G3FD* as a novel genetic element affecting fruit diameter in peach and also a locus which interacts with QTLs for FD on G2 and G7. These results agree with findings from linear parameter analysis by Fernández i Martí et al. (2013) for an F_1 almond population derived from a cross between 'Vivot' × 'Blanquerna,' in which QTLs related to fruit size were identified on G3, G2 and G7.

The transmission of *G3FD* alleles was tracked across the pedigree analyzed (Fig. 3), revealing positive and negative peach alleles (Q and q, respectively) that can be traced from several of the early nineteenth-century commercial US peach founders, including 'Chinese Cling' (homozygous for Q) and 'Early Crawford' (heterozygous). This finding suggests that the many breeding programs using 'Early Crawford' and its descendants might be generating progenies carrying the negative allele for FD in peach. In the case of the UC Davis breeding program, almond lineages also may have made contributions, because QTLs for characteristics such as width, spherical index, thickness and kernel weight were also located on G3 (Fernández i Martí et al. 2013).

The directed transmission of QTL alleles from founders through domestication and breeding is a consequence of artificial selection promoting the accumulation of favorable commercial alleles (stacking), a model described for tomato domestication by Tanksley (2004). Even changes in gene dosage, occurring from chromosomal rearrangements when divergent genomes are combined, may contribute to the accumulation of the favorable alleles. The inclusion of almond-derived introgressions in germplasm analyzed through PBA is believed to have enhanced the ability to distinguish the association of the *G3FD* genomic region with a fruit-size parameter, such as fruit diameter, given that the introduction of the allele from almond contrasted with the peach allele (which is supported by the absence of transgressive segregation in hybrid progenies). The availability of a full sib family segregating for both alleles (e.g., Qq × Q*q*, because homozygous genotypes for q or *q* are not likely to be available in breeding programs) within the pedigree, and the use of denser genetic maps might improve the resolution for the genetic dissection of *G3FD*. The genotype probability estimate allowed by FlexQTL™ will facilitate crossing designs that more accurately target the molecular characterization of this trait.

Findings that G2 contains a QTL for FD concur with those from Quilot et al. (2004) in peach and Zhang et al. (2010) and Rosyara et al. (2013) in cherry. However, the chromosomal location is different. In the present study, the QTL was located in the upper G2 region, at ~4.5 Mb. Quilot et al. (2004) located it close to the RFLP marker CC115 which, based on the C-Map for *Prunus*, is located approximately at 22 Mb. Furthermore, Zhang et al. (2010) and Rosyara et al. (2013) located a QTL associated with FD on the upper region (~16 Mb) of G2 in cherry. In this study, *q*FD.7 is located between 21 and 22.5 Mb on G7, the same linkage group where Fernández i Martí et al. (2013) located QTLs for length of nut and kernel in almond. However, the location is different (based on the peach genome sequence) since QTLs for almond were located close to the SSR CPPCT033, which is located at 16.7 Mb on G7.

The situation for *G2FW* is similar to the one for *q*FD.2, since they co-located at 4.5 Mb on G2. Quilot et al. (2004) also found a QTL for FW at ~22 Mb on G2. In contrast, Rosyara et al. (2013) concluded that cherry and peach do not share co-located QTLs for FW. In the case of almond, QTLs for nut weight, thickness, geometric diameter, spherical index and size were located between markers UDP98-025 and BPPCT002 (Fernández i Martí et al. 2013), which in the peach genome sequence are flanking a region of ~6 Mb between 10.87 and 16.6 Mb. Although the QTLs identified for FD and FW were located on G2, they did not co-localize in similar chromosomal positions with those from previous studies in peach, cherry and almond. However, the identification of genomic regions influencing FD and FW suggests a conserved genetic scheme influencing fruit size in *Prunus* species, but with different locations along G2.

In a protein homology search of cell number regulator (*CNR*) genes, first performed on the peach reference genome sequence for later localization in cherry, De Franceschi et al. (2013) reported four homologs located at ~15.6 Mb of G2. The *CNR* genes

modulate cell proliferation in the ovary carpel, as validated in various crop species including tomato, maize, eggplant and peppers (Guo and Simmons 2011). For 2012 data, in addition to *G2FW*, regions on G4 and G5 showed positive evidence for FW, but for 2011 data neither region co-localized with cherry, although *q*FW.5 was located in the same linkage group, but in a different position with respect to a QTL for nut weight on the 'Vivot' × 'Blanquerna' map (particularly for a QTL related to kernel weight) (Fernández i Martí et al. 2013). Additionally, the QTLs *q*FW.4 and *q*FW.5 agreed with the linkage group location of *CNR* homologs identified on G4 and G5 of peach by De Franceschi et al. (2013); however, the exact chromosomal positions did not co-localize because these regions were not well defined in our study, while in De Franceschi et al. (2013), there are located at 1.5 Mb on G4 and 17.1 on G5.

Phenological traits (DTB, FDP and RD)

Given the complexity and economic importance of fruit ripening, it is a prime candidate for marker-assisted breeding. Several putative genomic regions have been mapped on G1, G2, G4 and G7 (Eduardo et al. 2011; Fan et al. 2010). The QTLs for DTB and RD located on G1 and G4 matched those previously reported chromosomes, yet not at the same locations. The *G1RD* was located within the first 10 Mb of G1, while qBD1d from Fan et al. (2010) was located close to the marker BPPCT028, at approximately 45.68 Mb. Likewise, the *G4DTB* and *G4BD* were located around 18 and 19 Mb, while *q*MD4.1 was located at ~11 and 11.2 Mb on G4 (Pirona et al. 2013). The overlapping of *G4DTB* and *G4BD* in our study matched with the pleiotropic locus reported by Eduardo et al. (2011) and fine mapped by Pirona et al. (2013), as well as the minor multi-year QTLs for chilling requirement (qCR4.b) from Fan et al. (2010) located between 10 and 12.7 Mb. The co-localization of *G4DTB* and *G4BD* was also associated with QTLs related to flowering and fruit maturity times. A similar interacting region associated with flowering time and chilling requirement in almond (Sánchez-Pérez et al. 2012), which had a long confidence region with a peak located close to the marker UDP96-003, was nearby 8.5 Mb on the peach genome sequence. Although positions of QTLs for DTB and RD on G4 varied among studies, the detection of a genomic region influencing flowering and ripening time is constant; therefore, we propose to designate this region as *G4Mat*.

While FDP is often associated with RD (Etienne et al. 2002), in this study, QTLs for FDP were located on four different linkage groups when considering both years, with none co-localizing with QTLs for DTB or RD. Two QTLs for FDP were identified in both years, and in both cases, the evidences moved from decisive in 2011 to merely positive in 2012, and with no defined region for 2012. Thus, *q*FDP.1 and *q*FDP.6 were not co-localized with the QTL on G4 as reported by Etienne et al. (2002) near marker CPPCT003B at ~9.8 Mb, but rather overlapped with a QTL for maturity date in a region related to *G4Mat*. In 2011, QTLs with positive evidence for DTB and RD overlapped at the top of G3, the same region in which a QTL with positive evidence for FDP was located in 2012. Yamamoto et al. (2001) also reported QTLs for flowering, maturation and fruit dropping times on G3; however, their positions are uncertain in the current peach genome sequence since they used SSR marker UDP96, and currently, only UDP96-008 is located between 17 and 17.5 Mb on G3, while, in our study, *q*DTB.3 and *q*RD.3 were located between 4.9 and 8.5 Mb.

The lack of a strong association for FDP with DTB and RD in our study suggests that an additional and independent genetic element may be influencing *G4Mat,* possibly introduced in the introgression lines, and leading to the exhibition of extra-late fruit ripening genotypes. Delayed fruit ripening has been an important UC breeding goal and a main reason for the introgression of new germplasm.

Biochemical traits (pH, TA and SSC)

QTLs for biochemical traits related to pH, organic acids and sucrose were co-localized in a region on the proximal position of G5 (Etienne et al. 2002; Quilot et al. 2004), close to the location of *G5pH* identified in this study (between 1 and 6 Mb). However, the low density of markers for G5 in our study limited the detection of a more proximal overlap. In addition, QTLs with positive evidence for SSC and TA in 2012 co-localized with *G5pH* for a genomic complex that we propose to as *G5Flav*. The occurrence of *G5Flav* suggests that the gene composition of this locus is showing pleiotropy, as with *G4Mat*. The *G5Flav* also

appears to be associated with the *D* locus, associated with low acidity (Boudehri et al. 2009), which has been reported to be located near CPPCT040, at 1 Mb on G5. Chromosome localization of both, *G5Flav* and *D* locus, on G5 suggests that they are the same locus influencing components involved in peach flavor. We cannot conclude *G5Flav* to be the *D* locus since the pedigree studied here did not include peach accessions categorized as low acid. Thus, further studies are needed to dissect the allelic composition of *G5Flav* where a broader pool of peach accessions, including low-acid types segregating for the *D* locus, is considered.

Some of the year-to-year inconsistencies in the QTLs for SSC and TA on G5 may have resulted from the loss of about 40 juice samples for pH, SSC and TA testing in 2012, but which were available in 2011. The degree of uncertainty associated with these traits would increase in their respective analyses, as the likelihood function for the trait is directly affected. The missing data may have also negatively affected the power for the detection of effects of another co-localized locus for pH, SSC and TA on G1 a possible candidate for a *q*Flav.1 locus. This is because evidence for such a locus is suggested by the identification of co-localized QTLs: for TA, which had positive evidence in 2011 and strong evidence in 2012; for pH, which showed positive evidence only in 2012; and for SSC, which showed positive evidence in both years.

Genomic breeding values

The GBVs are an efficient tool for translating QTL mapping findings into breeding applications because they provide an intuitive and quantitative scale for the selection of genotypes to advance in the breeding objectives. The accuracy indicated by the high correlation for 2011 and 2012 GBVs and observed phenotypes suggests a major contribution from additive genetic effects. However, in the case of phenological traits such as DTB and FDP, the observed negative correlation between 2011 and 2012 GBVs suggests that even though additive genetic effects greatly influenced the variation in both traits, effects related to environmental factors (e.g., year, location, climatic conditions) and agricultural management are also relevant in the exhibitions of the traits.

The calculated GBVs are the result of a Bayesian method in which the selection variable is applied on a finite number of factors, i.e., QTLs discovered through PBA, influencing the exhibition a given trait within pedigreed germplasm. Thus, though the reversible-jump method (Green 1995) implemented in FlexQTL^{TM}, changes in MCMC dimension generate a model where a given number of QTLs is a hypothesis to contrast against an alternative model with a different number of QTLs (one more or one less than the current model). Hence, it assumed that a trait is influenced by a limited number of QTLs within a linkage group (chromosome), which allows consideration of a set of hypotheses in the form of integer numbers and thus use of a Poisson distribution as prior for the number of QTLs as proposed by Sillanpää and Arjas (1999).

On the other hand, model selection targeting genomic selection (Meuwissen et al. 2001) is related to the infinitesimal model (Bulmer 1980), in which the main assumption is that an infinite number of linked loci exist, each with an infinitesimal additive effect and taken as an explanatory variable with a known continuous distribution of their effects. Meuwissen et al. (2001) proposed the construction of models from a training population to predict performance of individuals in a related population through the summation of the effects of several thousands of markers in putative tight LD with the causal genes influencing the trait. The markers with no association with these genes are heavily penalized and effectively excluded from the model (shrinkage of their effects toward zero). In our Bayesian QTL analysis, a discrete distribution (Poisson) was used as a prior for the number of QTL in the model. In some Bayesian genomic prediction approaches, a continuous distribution is used with the variable π being the proportion of markers with nonzero effects (Habier et al. 2011). Genomic selection thus pursues the construction of accurate models without emphasizing the underlying genetic–biological factors (i.e., QTLs) as model variables (Bink et al. 2014).

PBA provides information about number of QTLs, their additive effect size, posterior probabilities of QTL genotypes and QTL-based breeding values, which enable the optimization of information for decision making in plant breeding (Bink et al. 2014). Our Bayesian method is different from the genomic prediction approaches as it explicitly models QTL with the transmission probabilities from parents to offspring being dependent on the marker data. In our method, the marker data are also used to sample the position of the

QTL relative to the marker positions (interval mapping). In genomic prediction methods, only marker positions are considered and the markers themselves enter into the model and have a direct link to the phenotypes (i.e., there is no QTL modeled). In that perspective, genomic selection is much more sensitive to the marker density and length of LD in the training population. Genomic selection strategies emphasize the development and refinement of linear models for the prediction of performance without elucidating the underlying factors influencing the variation of traits.

In summary, pedigree-based analysis, used in this study to map QTLs under the Bayesian framework performed through FlexQTL^{TM} for a diverse peach breeding pedigree, identified several well-supported QTLs for at least five of the eight traits investigated. Many of those QTLs were located on chromosome regions reported in previous independent studies, although the exact positions with respect to the peach genome sequence did not always overlap. The inclusion of introgressed germplasm and the explicit declaration of the genetic structure of the pedigree as nuisance variable enhanced the identification of QTLs such as *G3FD* not previously reported for peach germplasm. Finally, the adoption of the PBA strategy using the genomic resources developed in RosBREED has greatly facilitated the implementation of marker-assisted breeding in the Processing Peach Breeding Program at UC Davis.

Acknowledgments This study was funded by USDA's National Institute of Food and Agriculture—Specialty Crop Research Initiative project, 'RosBREED: Enabling marker-assisted breeding in Rosaceae' (2009-51181-05808). Jonathan Fresnedo Ramírez was supported by a CONACYT-UCMEXUS (Mexican Council of Science and Technology and University of California Institute for Mexico and the United States) Doctoral fellowship at the University of California, Davis. The contributions of M.C.A.M. Bink & E.W. van de Weg were also supported by the EU-FruitBreedomics project funded by the Commission of the European Communities (Contract FP7-KBBE-2010-265582). Thanks to Dr. Pedro J. Martínez-García for his helpful comments about the organization of the manuscript. Special thanks for their valuable help during the phenotyping of the accessions to the field crew of the Processing Peach Breeding Program at UC Davis, supervised by Mary Ann Thorpe, Sabrina Marchand, Helen Chan and Rachel Williams. Likewise, thanks to Palma Lower, writing specialist at UC Davis, for her valuable comments and corrections during early drafts of the manuscript.

Data archiving The genotypic and phenotypic datasets of the UC Davis pedigree-connected germplasm can be accessed through the Breeders Toolbox available at the Genome Database for Rosaceae (http://www.rosaceae.org/breeders_toolbox). The QTL information is accessible through the Trait Loci search tool at GDR. (http://www.rosaceae.org/search/qtl).

Authors' contributions J.F.R. carried out the analyzes and drafted the manuscript, M.C.A.M.B. and E.V.W. provided support for implementation and performing of PBA as well as for the interpretation of the results, also helped in drafting the manuscript, T.R.F. helped to perform pedigree pruning and determination of genetic structure, C.H.C. provided support for phenotypic evaluation and analyzes, T.J.F., K.G. and C.P.P. developed the SNP genotyping and database for the peach set in RosBREED and helped in drafting the manuscript, and T.M.G. provided the genetic materials, coordinated the study and elaborated on manuscripts. All authors read and approved the final and reviewed manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that there are no conflicts of interest.

References

Abascal E, García Lautre I, Landaluce MI (2006) Multiple factor analysis of mixed tables of metric and categorical data. In: Multiple correspondence analysis and related methods. Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences. Chapman and Hall/CRC, pp 351–367. doi:10.1201/9781420011319.ch15

Aranzana MJ, Abbassi EK, Howad W, Arus P (2010) Genetic variation, population structure and linkage disequilibrium in peach commercial varieties. BMC Genet 11:69. doi:10.1186/1471-2156-11-69

Aranzana MJ, Illa E, Howad W, Arús P (2012) A first insight into peach [*Prunus persica* (L.) Batsch] SNP variability. Tree Genet Genomes 8(6):1359–1369. doi:10.1007/s11295-012-0523-6

Arús P, Yamamoto T, Dirlewanger E, Abbott AG (2010) Synteny in the Rosaceae. In: Plant breeding reviews. Wiley, New York, pp 175–211. doi:10.1002/9780470650349.ch4

Bink MCAM, Boer MP, Braak CJF, Jansen J, Voorrips RE, van de Weg WE (2008) Bayesian analysis of complex traits in pedigreed plant populations. Euphytica 161(1/2):85–96

Bink MCAM, Jansen J, Madduri M, Voorrips RE, Durel CE, Kouassi AB, Laurens F, Mathis F, Gessler C, Gobbin D, Rezzonico F, Patocchi A, Kellerhals M, Boudichevskaia A, Dunemann F, Peil A, Nowicka A, Lata B, Stankiewicz-Kosyl M, Jeziorek K, Pitera E, Soska A, Tomala K, Evans KM, Fernández-Fernández F, Guerra W, Korbin M, Keller S, Lewandowski M, Plocharski W, Rutkowski K, Zurawicz E, Costa F, Sansavini S, Tartarini S, Komjanc M, Mott D, Antofie A, Lateur M, Rondia A, Gianfranceschi L, van de Weg WE (2014) Bayesian QTL analyses using pedigreed families of an outcrossing species, with application to fruit

firmness in apple. Theor Appl Genet 127(5):1073–1090. doi:10.1007/s00122-014-2281-3

Boudehri K, Bendahmane A, Cardinet G, Troadec C, Moing A, Dirlewanger E (2009) Phenotypic and fine genetic characterization of the D locus controlling fruit acidity in peach. BMC Plant Biol 9 (Artn 59)

Bulmer MG (1980) The mathematical theory of quantitative genetics. Oxford University Press, Oxford

Calboli FCF, Sampson J, Fretwell N, Balding DJ (2008) Population structure and inbreeding from pedigree analysis of purebred dogs. Genetics 179(1):593–601. doi:10.1534/Genetics.107.084954

Cannings C, Sheehan NA (2002) On a misconception about irreducibility of the single-site gibbs sampler in a pedigree application. Genetics 162(2):993–996

Cao K, Wang L, Zhu G, Fang W, Chen C, Luo J (2012) Genetic diversity, linkage disequilibrium, and association mapping analyses of peach (*Prunus persica*) landraces in China. Tree Genet Genomes 8(5):975–990. doi:10.1007/s11295-012-0477-8

Charlesworth D, Charlesworth B (1979) The evolutionary genetics of sexual systems in flowering plants. Proc R Soc Lond B Biol Sci 205(1161):513–530

Coster A (2012) pedigree: pedigree functions, 1.4 edn. http://CRAN.R-project.org/package=pedigree

De Franceschi P, Stegmeir T, Cabrera A, van der Knaap E, Rosyara UR, Sebolt AM, Dondini L, Dirlewanger E, Quero-Garcia J, Campoy JA, Iezzoni AF (2013) Cell number regulator genes in *Prunus* provide candidate genes for the control of fruit size in sweet and sour cherry. Mol Breed 32(2):311–326. doi:10.1007/S11032-013-9872-6

Eduardo I, Pacheco I, Chietera G, Bassi D, Pozzi C, Vecchietti A, Rossini L (2011) QTL analysis of fruit quality traits in two peach intraspecific populations and importance of maturity date pleiotropic effect. Tree Genet Genomes 7(2):323–335. doi:10.1007/S11295-010-0334-6

Etienne C, Rothan C, Moing A, Plomion C, Bodenes C, Svanella-Dumas L, Cosson P, Pronier V, Monet R, Dirlewanger E (2002) Candidate genes and QTLs for sugar and organic acid content in peach [*Prunus persica* (L.) Batsch]. Theor Appl Genet 105(1):145–159. doi:10.1007/S00122-001-0841-9

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Longman, Harlow

Fan S, Bielenberg DG, Zhebentyayeva TN, Reighard GL, Okie WR, Holland D, Abbott AG (2010) Mapping quantitative trait loci associated with chilling requirement, heat requirement and bloom date in peach (*Prunus persica*). New Phytol 185(4):917–930. doi:10.1111/J.1469-8137.2009.03119.X

Fernández i Martí A, Font I Forcada C, R Socias I Company (2013) Genetic analysis for physical nut traits in almond. Tree Genet Genomes 9(2):455–465. doi:10.1007/S11295-012-0566-8

Flint-Garcia SA (2013) Genetics and consequences of crop domestication. J Agric Food Chem 61(35):8267–8276. doi:10.1021/jf305511d

Font i Forcada C, Oraguzie N, Igartua E, Moreno MÁ, Gogorcena Y (2012) Population structure and marker–trait associations for pomological traits in peach and nectarine cultivars. Tree Genet Genomes 9(2):331–349. doi:10.1007/s11295-012-0553-0

Frett TJ, Gasic K, Clark JR, Byrne D, Gradziel T, Crisosto C (2012) Standardized phenotyping for fruit quality in peach [*Prunus persica* (L.) Batsch]. J Am Pomol Soc 66(4):214–219

Gradziel TM (2002) Almond species as sources of new genes for peach improvement. Acta Hortic 592:81–88

Gradziel TM (2003) Interspecific hybridizations and subsequent gene introgression within *Prunus* subgenus *Amygdalus*. Acta Hortic 622:249–255

Gradziel TM, Weinbaum SA (1999) High relative humidity reduces anther dehiscence in apricot, peach, and almond. HortScience 34(2):322–325

Gradziel TM, Beres W, Pelletreau K (1993) Inbreeding in California canning clingstone peach cultivars. Fruit Var J 47(3):160–168

Green PJ (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82(4):711–732. doi:10.1093/Biomet/82.4.711

Guan Y, Peace C, Rudell D, Verma S, Evans K (2015) QTLs detected for individual sugars and soluble solids content in apple. Mol Breeding 35(6):1–13. doi:10.1007/s11032-015-0334-1

Guo M, Simmons R (2011) Cell number counts—the fw2.2 and CNR genes and implications for controlling plant fruit and organ size. Plant Sci 181(1):1–7. doi:10.1016/J.Plantsci.03.010

Habier D, Fernando RL, Kizilkaya K, Garrick DJ (2011) Extension of the Bayesian alphabet for genomic selection. BMC Bioinf 12:186. doi:10.1186/1471-2105-12-186

Iezzoni A (2010) RosBREED: enabling marker-assisted breeding in the Rosaceae. HortScience 45(8):S27–S28

Iezzoni A, Weebadde C, Luby J, Yue CY, van de Weg WE, Fazio G, Main D, Peace CP, Bassil NV, McFerson J (2010) RosBREED: enabling marker-assisted breeding in Rosaceae. Acta Horticulturae 859:389–394

Jannink JL, Bink MCAM, Jansen RC (2001) Using complex plant pedigrees to map valuable genes. Trends Plant Sci 6(8):337–342

Jung S, Staton M, Lee T, Blenda A, Svancara R, Abbott A, Main D (2008) GDR (Genome Database for Rosaceae): integrated web-database for Rosaceae genomics and genetics data. Nucleic Acids Res 36(suppl 1):D1034–D1040. doi:10.1093/nar/gkm803

Jung S, Ficklin SP, Lee T, Cheng C-H, Blenda A, Zheng P, Yu J, Bombarely A, Cho I, Ru S, Evans K, Peace C, Abbott AG, Mueller LA, Olmstead MA, Main D (2014) The genome database for Rosaceae (GDR): year 10 update. Nucleic Acids Res 42(D1):D1237–D1244. doi:10.1093/nar/gkt1012

Kass RE, Raftery AE (1995) Bayes factors. J Am Stat Assoc 90(430):773–795. doi:10.2307/2291091

Le S, Josse J, Husson F (2008) FactoMineR: an R package for multivariate analysis. J Stat Softw 25(1):1–18

Lee SH, Van der Werf JHJ (2005) The role of pedigree information in combined linkage disequilibrium and linkage mapping of quantitative trait loci in a general complex pedigree. Genetics 169(1):455–466. doi:10.1534/genetics.104.033233

Martinez-Gomez P, Arulsekar S, Potter D, Gradziel TM (2003) An extended interspecific gene pool available to peach and almond breeding as characterized using simple sequence repeat (SSR) markers. Euphytica 131(3):313–322. doi:10.1023/A:1024028518263

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Okie WR, Bacon T, Bassi D (2008) Fresh market cultivar development. In: Layne D, Bassi D (eds) The peach: botany, production and uses. CABI, Wallinford, pp 37–60

Olukolu BA, Kole C (2012) Molecular mapping of complex traits. In: Kole C, Abbott AG (eds) Genetics, genomics and breeding of crop plants. CRC Press, Boca Raton, pp 126–157

Peace C, Bassil NV, Bink M, Brown SK, Byrne DH, Clark JR, Davis TM, Evans K, Fazio G, Finn CE, Gasic K, Gradziel T, Hancock JF, Luby J, Main D, Oraguzie N, van de Weg E, Wang DC, Xu KN, Iezzoni A (2010) RosBREED's marker-assisted breeding Pipeline. HortScience 45(8):S54–S54

Peace CP, Luby JJ, van de Weg WE, Bink MCAM, Iezzoni AF (2014) A strategy for developing representative germplasm sets for systematic QTL validation, demonstrated for apple, peach, and sweet cherry. Tree Genet Genomes 10(6):1679–1694. doi:10.1007/s11295-014-0788-z

Pirona R, Eduardo I, Pacheco I, Linge CD, Miculan M, Verde I, Tartarini S, Dondini L, Pea G, Bassi D, Rossini L (2013) Fine mapping and identification of a candidate gene for a major locus controlling maturity date in peach. BMC Plant Biol 13 (Artn 166)

Quilot B, Wu BH, Kervella J, Genard M, Foulongne M, Moreau K (2004) QTL analysis of quality traits in an advanced backcross between Prunus persica cultivars and the wild relative species *P. davidiana*. Theor Appl Genet 109(4):884–897. doi:10.1007/S00122-004-1703-Z

R Development Core Team (2012) R: a language and environment for statistical computing, 2.15th edn. Development Core Team R, Vienna

Riester M, Stadler PF, Klemm K (2009) FRANz: reconstruction of wild multi-generation pedigrees. Bioinformatics 25(16):2134–2139. doi:10.1093/Bioinformatics/Btp064

Rosyara UR, Bink MCAM, van de Weg E, Zhang GR, Wang DC, Sebolt A, Dirlewanger E, Quero-Garcia J, Schuster M, Iezzoni AF (2013) Fruit size QTL identification and the prediction of parental QTL genotypes and breeding values in multiple pedigreed populations of sweet cherry. Mol Breed 32(4):875–887. doi:10.1007/S11032-013-9916-Y

Sánchez-Pérez R, Dicenta F, Martinez-Gomez P (2012) Inheritance of chilling and heat requirements for flowering in almond and QTL analysis. Tree Genet Genomes 8(2):379–389. doi:10.1007/S11295-011-0448-5

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 Am Soc Hortic Sci 110(4):547–552

Sillanpää MJ, Arjas E (1999) Bayesian mapping of multiple quantitative trait loci from incomplete outbred offspring data. Genetics 151:1605–1619

Sorensen D, Gianola D (2002) Likelihood, Bayesian and MCMC methods in quantitative genetics. Statistics for biology and health. Springer, New York

Tanksley SD (2004) The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. Plant Cell 16:S181–S189. doi:10.1105/Tpc.018119

Tanksley SD, Ganal MW, Prince JP, de Vicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, Messeguer R, Miller JC, Miller L, Paterson AH, Pineda O, Roder MS, Wing RA, Wu W, Young ND (1992) High density molecular linkage maps of the tomato and potato genomes. Genetics 132(4):1141–1160

van de Weg WE, Voorrips RE, Finkers HJ, Kodde LP, Meulenbroek EJ, Jansen J, Bink MCAM (2005) Pedigree genotyping: a new pedigree-based approach of QTL identification and allele mining by exploiting breeding material. Acta Hortic 708:483–488

Verde I, Bassil N, Scalabrin S, Gilmore B, Lawley CT, Gasic K, Micheletti D, Rosyara UR, Cattonaro F, Vendramin E, Main D, Aramini V, Blas AL, Mockler TC, Bryant DW, Wilhelm L, Troggio M, Sosinski B, Aranzana MJ, Arus P, Iezzoni A, Morgante M, Peace C (2012) Development and evaluation of a 9K SNP array for peach by internationally coordinated SNP detection and validation in breeding germplasm. PlosOne 7(4) (ARTN e35668). http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0035668

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93(1):77–78. doi:10.1093/jhered/93.1.77

Voorrips RE, Bink MCAM, van de Weg WE (2012) Pedimap: software for the visualization of genetic and phenotypic data in pedigrees. J Hered 103(6):903–907. doi:10.1093/Jhered/Ess060

Yamamoto T, Shimada T, Imai T, Yaegaki H, Haji T, Matsuta N, Yamaguchi M, Hayashi T (2001) Characterization of morphological traits based on a genetic linkage map in peach. Breed Sci 51(4):271–278. doi:10.1270/jsbbs.51.271

Zeinalabedini M, Sohrabi S, Nikoumanesh K, Imani A, Mardi M (2012) Phenotypic and molecular variability and genetic structure of Iranian almond cultivars. Plant Syst Evol 298(10):1917–1929. doi:10.1007/S00606-012-0691-8

Zhang GR, Sebolt AM, Sooriyapathirana SS, Wang DC, Bink MCAM, Olmstead JW, Iezzoni AF (2010) Fruit size QTL analysis of an F1 population derived from a cross between a domesticated sweet cherry cultivar and a wild forest sweet cherry. Tree Genet Genomes 6(1):25–36. doi:10.1007/s11295-009-0225-x