

RESEARCH ARTICLE

Construction of a highly saturated linkage map in Japanese plum (*Prunus salicina* L.) using GBS for SNP marker calling

Basilio Carrasco^{1*}, Máximo González^{1‡}, Marlene Gebauer¹, Rolando García-González^{2,3}, Jonathan Maldonado⁴, Herman Silva^{4*}

1 Pontificia Universidad Católica de Chile, Facultad de Agronomía e Ingeniería Forestal, Departamento de Ciencias Vegetales, Macul, Santiago, Chile, **2** Sociedad BioTECNOS Ltda, R&D Department Camino a Pangal Km 2 1/2, San Javier, Región del Maule, Chile, **3** Facultad de Ciencias Agrarias y Forestales, Centro de Biotecnología de los Recursos Naturales (CENBio), Universidad Católica del Maule, Talca, Chile, **4** Universidad de Chile, Facultad de Ciencias Agronómicas, Departamento de Producción Agrícola, Laboratorio de Genómica Funcional & Bioinformática, La Pintana, Santiago, Chile

‡ Current address: Centro de Estudios Avanzados en Zonas Áridas (CEAZA), La Serena, Chile
* bcarrasco@uc.cl (BC); hesilva@uchile.cl (HS)



OPEN ACCESS

Citation: Carrasco B, González M, Gebauer M, García-González R, Maldonado J, Silva H (2018) Construction of a highly saturated linkage map in Japanese plum (*Prunus salicina* L.) using GBS for SNP marker calling. PLoS ONE 13(12): e0208032. <https://doi.org/10.1371/journal.pone.0208032>

Editor: Anil Kumar Singh, ICAR-Indian Institute of Agricultural Biotechnology, INDIA

Received: June 20, 2018

Accepted: November 9, 2018

Published: December 3, 2018

Copyright: © 2018 Carrasco et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All sequence data used in this study have been deposited in the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) under the BioProject accession number PRJNA475160.

Funding: This research was supported by Fondo Nacional de Desarrollo Científico y Tecnológico (<http://www.conicyt.cl/fondecyt/>), FONDECYT/Regular no 1160600. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

This study reports the construction of high density linkage maps of Japanese plum (*Prunus salicina* Lindl.) using single nucleotide polymorphism markers (SNPs), obtained with a GBS strategy. The mapping population (An x Au) was obtained by crossing cv. “Angeleno” (An) as maternal line and cv. “Aurora” (Au) as the pollen donor. A total of 49,826 SNPs were identified using the peach genome V2.1 as a reference. Then a stringent filtering was carried out, which revealed 1,441 high quality SNPs in 137 An x Au offspring, which were mapped in eight linkage groups. Finally, the consensus map was built using 732 SNPs which spanned 617 cM with an average of 0.96 cM between adjacent markers. The majority of the SNPs were distributed in the intragenic region in all the linkage groups. Considering all linkage groups together, 85.6% of the SNPs were located in intragenic regions and only 14.4% were located in intergenic regions. The genetic linkage analysis was able to co-localize two to three SNPs over 37 putative orthologous genes in eight linkage groups in the Japanese plum map. These results indicate a high level of synteny and collinearity between Japanese plum and peach genomes.

Introduction

The genus *Prunus* (family *Rosaceae*) contains more than 200 species, which include some relevant agricultural stone fruit crops such as peach, apricot, cherry, Japanese plum, myrobalan and European plum, among others [1, 2]. Japanese plum is a self-incompatible, diploid species ($2n = 2X = 16$) that produces an edible drupe; it has been cultivated for 4,000 years [2] for fruit production and ornamental purposes. Different centers around the world have begun to perform breeding programs to obtain new cultivars. Usually the new cultivars show a high level of variability for some important traits such as harvest date, fruit size, flesh color and shape [3].

Author Rolando García-González is employed by Sociedad BioTECNOS Ltda. This affiliation commenced towards the end of this study. Sociedad BioTECNOS Ltda provided support in the form of salary for author RG-G, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific role of this author is articulated in the 'author contributions' section.

Competing interests: Author Rolando García-González is employed by Sociedad BioTECNOS Ltda. This affiliation commenced towards the end of this study. There are no patents, products in development or marketed products to declare. This does not alter our adherence to all the PLOS ONE policies on sharing data and materials.

Some genetic studies have been carried out to investigate the inheritance pattern and heritability of flowering date, ripening date, fruit size, higher total soluble solids, sweetness and flavor in the fruit [2, 3], while molecular studies have been few and focused only on the ethylene response and skin and flesh colors [4–10]. The construction of a linkage map is an important tool for genetic studies. This map will allow establishing the organization of a genome as well as a relationship between markers and polymorphisms. The mapping information offers evidence about polymorphisms strongly associated with agronomically relevant traits in segregating populations [11, 12], allowing the genetic basis of the trait of interest to be determined, and also can be a powerful tool in breeding programs for choosing parental lines and implementing marker-assisted selection strategies. The availability of a consensus reference map built for Rosaceae has facilitated the order and the physical orientation of maps as well as the localization of new consensus molecular markers for assisted selection in different *Prunus* species. Several genetic maps have been published for Rosaceae species, including peach, apple, pear, raspberry, and cherry. A comparative linkage map has even been built for Rosaceae [13, 14]. Peach has been considered as a model species for the genus *Prunus* [15, 16]. Verde et al. [17] recently published a new version of a high-quality draft genome of peach, which is a baseline for comparative analysis with *Prunus* species. In addition, Zhang et al. [18] have published the genome of mei (*Prunus mume*).

However, to date only a few genetic linkage maps have been published for plums and related species. Dirlwanger et al. [13] developed the first genetic map for mirobalan plum (*P. ceracifera*). Later, 144 microsatellites (or simple sequence repeats, SSR) were mapped in *Prunus mume* [19]. Only two genetic maps have been published for *Prunus salicina*. Vieira et al. [20] built a map using AFLPs and more recently, Salazar et al. [21] described QTLs using SNPs.

Next generation sequencing approaches have recently allowed the release of whole genome reference sequences for many plant species. This technique also allows the identification of thousands of single nucleotide polymorphisms (SNPs), the most abundant type of DNA marker found in eukaryotic genomes [22–24]. SNPs have become tremendously important as markers for genetics research in plants because they have been found to be in high frequency, display a lower mutation rate compared to SSR-based markers and they are uniformly distributed across the genome [25].

The distribution pattern of SNPs in segregating populations plus the available reference genomes have allowed analyzing the linkage relationships and the physical genome distribution of those SNPs, permitting comparisons of the collinearity between related species [26]. These properties make SNPs relevant to carry out genetic studies such as phylogenetics, genetic diversity, association analysis and genetic mapping [27].

Next generation sequencing is rapidly becoming a low-cost technology to carry out massive genetic studies, which is allowing important advances in plant genetics and breeding. Elshire et al. [28] developed a robust and low-cost genotyping method based on partial genome sequencing called genotyping-by-sequencing (GBS). This approach uses restriction enzymes to digest the genome in order to reduce its complexity. The DNA fragments are sequenced by high-throughput methods, obtaining hundreds of thousands of SNPs simultaneously. The GBS approach has been shown to be suited to do genetic analysis and linkage mapping of *Rosaceae* species such as peach [29], sweet cherry [30, 31], raspberry [32], and apple [33]. The objective of the present study was the development of a high-density linkage map which will provide important information about the genetics and genomics of *Prunus salicina* and its relationships with other *Prunus* species. To date this is the most saturated map available for this species.

Material and methods

Parental lines and F₁ segregating population

During the spring of 2013 we crossed the cultivar “Angelino” (An) (late harvest variety; 179 days after full blossom) as a maternal line with the cultivar “Aurora” (Au) (early harvest variety; 112 days after full blossom) as the pollen donor. An was selected as a parental line because is the most important cultivar of Japanese plum for the Chilean fresh fruit exportation industry [34]. An is a late harvest cultivar and has a high level of productivity and a long period of shelf life. Its fruits can be stored at 4 °C for a long period of time (at least 45 days) without expressing physiological disorders such as woolliness or internal breakdown [35]. To obtain seedlings after pollination and then fruit set, the seed were collected and germinated at 4 °C for 75 days. Then 137 F₁ seedlings were established at the experimental station of the Fundación Agro UC located at Curacaví, Chile (33° 26' South, 71° 01' West) during September, 2014. Each seedling was planted directly in the field without grafting and maintained following standard agronomic protocols (canopy pruning, fruit thinning, drip irrigation, fertilization and phytosanitary control).

Genotyping by sequencing (GBS)

GBS protocols were carried out by Institute of Biotechnology, Cornell University, Ithaca, NY, USA according to [28] (more details can be obtained from <http://www.biotech.cornell.edu/brc/genomics-facility>). Young leaves from the 137 segregating F₁ and parental lines An and Au were collected for DNA isolation. High quality DNA was extracted from leaves using a standard CTAB approach modified by Carrasco et al. [36]. The double-stranded DNA concentrations were measured with a Qubit 3.0 Fluorometer (Thermo Fisher Scientific). We extracted at least 70 ng/ul of double-stranded DNA per sample according to the protocol suggested by the BRC Genome Facility (Cornell University Biotechnology Resource Center, USA). GBS libraries were developed using the restriction enzyme ApeKI (GCWGC) and two different adapters according to protocols from the Institute for Genome Diversity (IGD) at Cornell University. ApeKI was selected from a panel of restriction enzymes because it is able to produce hundreds of thousands of fragments between 150 and 500bp. Several studies of the genomes of *Prunus* species have used this criterion successfully to select the restriction enzyme for GBS [21; 28; 31; 37; 38]. GBS sequencing libraries were prepared by ligating the digested DNA to nucleotide adapters (barcodes), followed by standard PCR. Sequencing was performed using Illumina HiSeq2000. The DNA of the parental lines was sequenced three times (independent samples) in order to reduce missing data and errors during SNP calling.

The separate FASTQ (raw files) were aligned to Peach v2.1 [17] (http://www.rosaceae.org/gb/gbrowse/prunus_persica_v2.1) using the Burrows-Wheelers alignment tool version 0.7.8-r441 [39]. Alignments were converted to the SAM format, then merged and sorted into one master binary alignment file with SAMtools 0.1.18 [39]. A ‘master’ TagCounts file was produced, which was aligned to the peach genome V2.1, and a Tags on Physical Map (TOPM) file was built, containing the best genomic position of each tag. The barcode information in the original FASTQ files is used to tally the number of times each tag in the master tag list is observed in each sample (‘taxon’) and these counts are stored in a TagsByTaxa file. The information recorded in the TOPM and TBT is then used to discover SNPs and filter them based upon the proportion of taxa covered by the TagLocus, minor allele frequency and inbreeding coefficient (F_{IT}). More details about the pipeline to SNP calling can be found in Glaubitz et al. [40] and IGD.

Finally, the identified SNPs corresponding to each genotype were stored as a filtered VCF file for posterior filtering with Tassel 5.2.16 [40, 41] and mapping analysis using Joinmap v4.1 software [42].

SNP analysis

SNPs were identified following the nomenclature defined in the peach genome v2.1, where each SNP was located in the pseudomolecules (scaffold) s1 to s8, followed by the physical position in base pairs (bp). Total SNPs were filtered using the VCFtools software [43]. Only biallelic SNPs with minimum allele frequency of 0.05 were used. SNP markers with more than 10% missing data were removed, and only those SNPs that showed an even distribution along the main eight peach scaffolds were used for posterior analysis (high quality SNPs). The location of each SNP within intergenic and genic regions (exon, intron and untranslated regions (UTR)) was determined using the Perl script (www.perl.org), determining the positions from the *Prunus persica* v2.1 GFF annotation.

Linkage map construction

The VCF file of clean SNPs was converted to JoinMap format using ngsTools software [44]; therefore the SNPs that were classified as heterozygous in one of the parents were scored as segregation types <lmxll> or <nnxnp>, while SNPs heterozygous in both parents were scored as segregation type <hkxhk>.

Map construction was performed using JoinMap 4.1 software [42] following a two-step strategy. Parental maps were first constructed using SNPs classified as <lmxll> or <nnxnp>. Then for consensus map construction, the integration of the parental maps was performed including the SNPs classified as <hkxhk>.

The segregation distortion of SNPs was determined by calculating chi-square (χ^2) using JoinMap, and SNPs with a major distortion of $p < 0.001$ were removed. In addition, in those cases in which one SNP had *Similarity of Loci* = 1 with another SNP, only one of them was used for map construction.

Maternal, paternal and consensus map constructions were performed using the regression mapping algorithm. For maternal and paternal map construction, markers were grouped using a minimum independence LOD (logarithm of the odds) score of 10.0 and linkage groups were established at a minimum LOD score of 3.0 and maximum recombination frequency of 0.40. Map distance was estimated using the Kosambi mapping function [45]. Consensus map construction was performed considering the same parameters used for parental map constructions.

Results

Linkage mapping

Using the GBS approach we obtained 454,912,276 reads and 4,995,919 tags; 2,367,021 tags (47.4%) were aligned to a unique position, 200,809 tags (4.0%) were aligned to multiple positions and 2,428,089 tags (48.6%) could not be aligned to the peach genome v2.1 that was used as a reference. The HapMap file containing 49,826 filtered SNPs was used for subsequent analysis. After applying filters (MAF >0.05 and 10% as maximum missing data) a total of 12,720 SNPs were obtained to be analyzed with the software JoinMap 4.1. Chi-square analyses were performed on informative SNPs to evaluate their conformity to the expected Mendelian segregation ratios. SNPs with identical segregation patterns were also discarded. Finally, 1,441 high quality SNPs were available for map construction.

Of the 1,441 high quality SNPs, 714 (50%) SNPs were classified as female parent segregation (“Angelino”; <nn x np>); 320 (22%) as male parent segregation (“Aurora”; <lm x ll>) and 407 (28%) SNPs as heterozygous segregation in both parents (<hk x hk>).

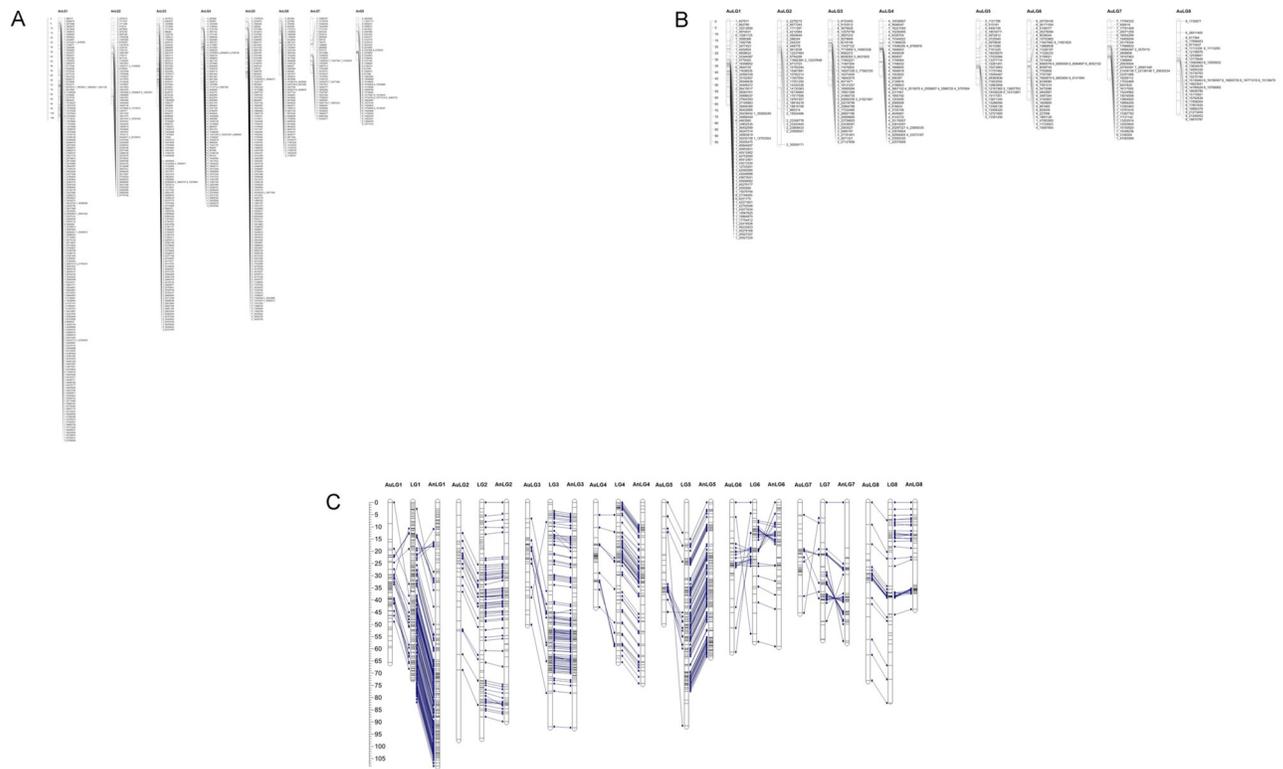


Fig 1. Genetic map. a. “Angelino” (female line). b. “Aurora” (male line). c. Consensus.

<https://doi.org/10.1371/journal.pone.0208032.g001>

The linkage analysis showed that the 1,441 SNPs were spread over eight linkage groups, which corresponds to the reported haploid chromosome number in Japanese plum [2]. The linkage maps spanned 588 cM and 490 cM for An (female line) and Au (male line), respectively (Fig 1a and 1b; S1 Table).

The consensus linkage map was constructed using SNPs mapped in both parental maps plus the SNPs in heterozygous configuration. The consensus map contained 732 non-redundant SNPs and spanned a distance of 617 cM (Table 1; Fig 1c; S1 and S2 Tables). The number of SNPs mapped, the total map length, the average distance and maximum gap (Table 1) ranged from 47 SNPs, 56 cM, 1.22 cM/SNP and 8.0 cM in LG7 to 89 SNPs, 100cM, 1.13 cM/ SNP and 10.2 cM for LG2. The average distance between markers was 0.96 cM/SNP.

Table 1. Characterization of 732 SNPs distributed in the consensus genetic map of Japanese plum.

Linkage Group	SNPs (N ^o)	Total distance (cM)	Average distance (cM/SNP)	Maximum gap (cM)
LG1	161	73	0.45	5.3
LG2	89	100	1.13	10.2
LG3	125	92	0.74	13.8
LG4	82	66	0.81	4.3
LG5	121	92	0.76	12.4
LG6	48	57	1.21	8.4
LG7	47	56	1.22	8.0
LG8	59	82	1.42	9.7
Total	732	617	0.96	9.0

<https://doi.org/10.1371/journal.pone.0208032.t001>

Region	G1	G2	G3	G4	G5	G6	G7	G8	TOTAL
Exonic	95	58	84	54	79	26	33	28	457
Intronic	17	9	12	10	13	13	3	12	89
5' UTR	12	5	3	1	4	4	6	3	38
3' UTR	14	3	7	3	6	1	2	5	41
Intergenic	23	14	19	14	19	4	3	11	107
total	161	89	125	82	121	48	47	59	732

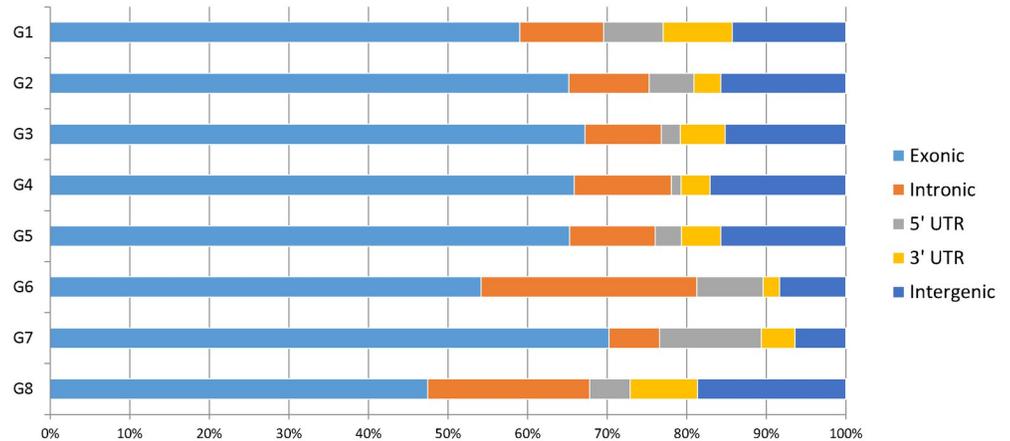


Fig 2. Genomic distribution of 732 SNPs in *P. salicina* based on physical positions identified in peach genome v2.1.

<https://doi.org/10.1371/journal.pone.0208032.g002>

Most of the SNPs identified using the peach genome were co-localized in the correct linkage group after linkage analysis. However, some SNPs did not follow the same physical order relative to the scaffolds of peach genome v2.1. We found that 69 SNPs (9.5%) of 732 SNPs mapped to different linkage groups compared to the peach genetic map.

SNP distribution through the linkage groups

The location of each SNP in the intragenic and intergenic regions (Fig 2) was determined by using the physical position in peach genome v2.1. Table 2 shows the distribution of SNPs in the intragenic as well as intergenic regions of the eight linkage groups.

Considering all linkage groups together, 85.6% of the SNPs (625 SNPs) were located in intragenic regions (62.6% in exons; 12.2% in introns and 10.8% in UTR), and only 14.4% (107 SNPs) were located in intergenic regions (Fig 2).

Table 2. Distribution of SNPs in linkage groups.

Linkage Group (LG)	SNPs in intragenic regions (%)	SNPs in intergenic regions (%)
LG1	85.7	14.3
LG2	85.2	14.8
LG3	85.5	15.2
LG4	82.9	17.1
LG5	84.3	15.7
LG6	91.7	8.3
LG7	93.6	6.4
LG8	81.4	18.6

<https://doi.org/10.1371/journal.pone.0208032.t002>

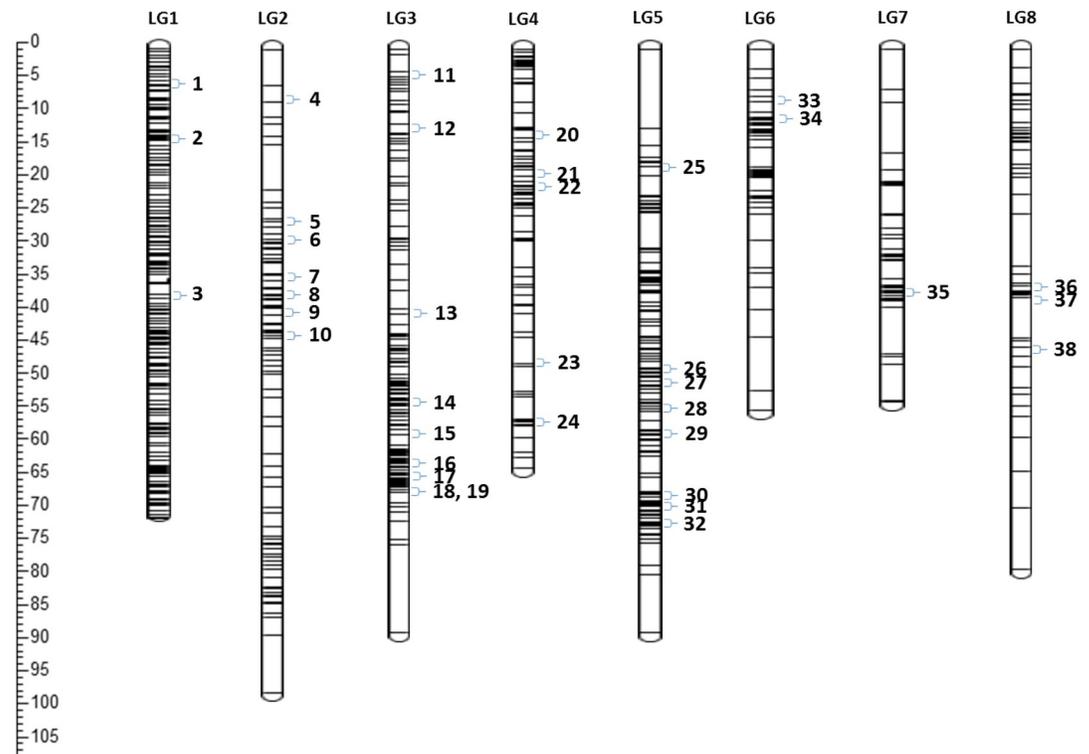


Fig 3. Japanese plum consensus genetic map. The numbers indicate the map position of 38 putative orthologous genes in eight linkage groups of Japanese plum (see supplementary material for details). These genes were identified according to the annotation of peach genome v2.1.

<https://doi.org/10.1371/journal.pone.0208032.g003>

The genetic linkage analysis was also able to co-localize at least 1 SNP in 460 putative orthologous genes identified in the peach genome. These genes were distributed at the linkage groups as follows: LG1 = 112; LG2 = 50; LG3 = 65; LG4 = 52; LG5 = 79; LG6 = 34; LG7 = 34 and LG8 = 34 (for more details about these SNPs and genes see [S2 Table](#)). Using linkage analysis we were able to co-localize two to three SNPs simultaneously in 38 genes in the eight linkage groups of the genetic map of Japanese plum ([Fig 3](#)).

Discussion

Genetic linkage maps are used as a tool for primary localization of important genomic regions associated with the genetic control of both qualitative and quantitative traits, which can help to support a breeding program [46]. The huge amount of genetic information that can be obtained from high throughput sequencing approaches today have allowed whole-genome sequencing and linkage analysis to become powerful methodologies for the identification of genetic polymorphisms related to complex traits in different species [47, 48]. GBS is able to generate abundant SNPs for a large number of individuals of a species with a relatively low cost per sample [49]. It is also an effective source of sequence tags that can be used as genetic anchors to direct contig/scaffold assembly and to map genomic fragments using a reference genome.

The GBS approach used in this study identified 49,826 SNPs to initiate the mapping analysis. However, most of them (97.1%) were not used for the construction of a linkage genetic map and were eliminated due to low minor allele frequency (<0.05), excess of missing data

(>10%), SNPs with incorrect genotype, more than two alleles for one position and distortion of segregation. It is important to highlight that according to Bastien et al., [50] the restriction enzyme ApeKI used to generate the genomic library in GBS approach [28] helps to identify more SNPs than other restriction enzymes, but with a lower read coverage per marker and resulting in more missing data. After applying the filtering criteria, 1,441 high quality SNPs were available to build the genetic map in Japanese plum. Those 1,441 SNPs represented only 2.9% of the total SNPs identified, which shows the limitations of the GBS approach. Davey et al., [49] and Jiang et al., [51] suggested that the main limitation of GBS is the presence of missing data due to low coverage genotyping, inconsistency in the number of sites sequenced per sample and number of reads per site.

New methodologies to discover SNPs that can bypass these difficulties could give a more consistent number of useful SNPs for carrying out genetic studies [51–53]. In contrast to traditional GBS [28], Specific-Locus Amplified Fragment Sequencing (SLAF-seq) [51–53] is a new and efficient method for SNP discovery and genotyping that does not require a reference genome, uses two restriction enzymes (HaeIII and MseI), selecting fragments between 370–450bp. SLAF-seq is able to produce a larger number of high quality SNPs for mapping purposes [53].

Linkage maps using SNPs have been published for several species of the Rosaceae, including raspberry [32], apple [33, 54], pear [55], peach [29, 56], sweet cherry [30, 31], Chinese plum [57], Japanese plum [21] and almond [37]. These linkage maps were able to identify thousands of molecular markers; nevertheless, after filtering only a reduced number of SNPs were used to build a genetic map.

Few genetic linkage maps have been published for plum species. The first genetic map for plum was published by Dirlerwanger et al. [13]. They reported a map of a myrobalan clone (*P. ceracifera*) using 93 markers (two SCARs plus 91 SSR) which expanded over eight linkage groups, covering 524.8 cM. Later, Vieira et al. [20] reported a linkage map for parental lines of Japanese plum using 56 to 84 AFLP which covered 905.5 to 1,349.6 cM with an average distance between markers of 16.1 cM to 16.2 cM.

Recently Salazar et al., [21] published a genetic map using SNPs obtained by GBS. They mapped a total of 981 SNPs, 479 SNPs for the female line and 502 SNPs for the male line (cv. “Angeleno”), covering 688.8 cM and 647.03 cM, respectively. The average distance between SNPs was two cM. However, Salazar et al., [21] did not report a consensus map for the two parental lines, therefore it is not possible to distinguish redundant SNPs.

In contrast, we analyzed 1,441 high quality SNPs. Of these, 714 SNPs were mapped in the female parent (cv. “Angeleno”) and 320 in the male parent (cv. “Aurora”) expanded over 578cM and 472 cM, respectively. Our consensus map was built using 732 non-redundant SNPs, covering 617 cM with an average distance between SNPs of 0.96 cM. This saturation represents at least twice that reported by Salazar et al. [21] and is considered a highly saturated map according to Slate et al. [58]. These authors indicated that a high-density map refers an average interval between SNPs less than two cM. It is important to note that both our study and that of Salazar et al., [21] used the “Angeleno” cultivar as a parental line, which could be favorable for future studies.

Comparing our linkage map with those previously built for Rosaceae species (Table 2), the map length and density of the SNP were similar to those published for *P. persica* [56] and *P. avium* [30, 31]. Up to now only the *P. mume* map has higher saturation, with a density of 0.15 to 0.24 cM/SNP [57]. However, the *P. mume* map has large gaps in LG1 (24.13cM); LG6 (12.79cM) and LG8 (24.13cM). In contrast, the Japanese plum map reported in this study showed a maximum gap of 13.8cM, which was a little more than half that of *P. mume*. The presence of gaps in a linkage map needs to be analyzed carefully, because they could indicate

Table 3. Coverage and density reported in available genetic maps for some Rosaceae species developed using SNPs.

Species	SNPs (Nr)	Map length (cM)	Density (cM/SNP)
<i>Prunus salicina</i> ^a	730	626	0.96
<i>Prunus salicina</i> [21]	479–502	647.03–688.8	2.0
<i>Prunus pérsica</i> [49]	588	454	0.81
<i>Prunus avium</i> [30]	687–723	639.9–752.9	0.9–1.1
<i>Prunus avium</i> [31]	987	731.3	0.7
<i>Prunus mume</i> [50]	8,007	1,550.62	0.15–0.24
<i>Rubus ideaus</i> [32]	2,391–4,521	376.6–462.7	0.1–0.16
<i>Malus pumila</i> [51]	15,417	1,267	0.37
<i>Pyrus communis</i> [52]	3,853	3,266	0.64

^aThis study<https://doi.org/10.1371/journal.pone.0208032.t003>

an assembly problem in the reference genome. Plant genomes have proved to be difficult to assemble because of high heterozygosity, large numbers of repeat sequences and being large and complex [59]. Eichler et al., [60] suggested that genome duplication will be the most important cause of gaps in a physical and genetic map because they are difficult to assemble due to mapping of markers to multiple regions. Reduction of the gaps by increasing the number of offspring and molecular markers will improve the density and resolution of linkage maps, enhancing the ability to identify QTLs and markers for assisted selection.

Table 3 indicates that the density of our Japanese plum map (cM/SNPs) was rather superior to those obtained for previous maps developed for other species of Rosaceae such as raspberry, apple, and pear. The number of offspring (137) and the number of SNPs analyzed (730) in the segregating populations (An x Au) could explain this result.

The Japanese plum map revealed that most of the SNPs (90.5%) were located in the linkage groups according to their physical locations in the scaffold of the peach genome. However, some discrepancies were observed, such as differences in the correlative order of the SNPs in the genetic map and differences between physical locations of the SNPs in the peach genome and their positions in the linkage groups. Such discrepancies in SNP order have been previously reported in almond [61], sweet cherry [30, 31] and mei (*P. mume*) [57]. Also, some few SNPs (9.5%) located physically in the peach scaffold mapped to a different linkage group in the Japanese plum map. These discrepancies could be explained by assembly errors in the peach genome [30], mapping errors due to the number of individuals analyzed, and some evolutionary process such as genomic translocations and deletions.

Genus *Prunus* has shown conserved intraspecific and intragenic co-linearity in the Rosaceae [13], also even with other genera such as *Populus* [13]. Conserved synteny is important to plant breeding because it can be exploited to identify molecular markers linked to agronomically relevant traits between related species. A recent example of synteny-based development of molecular markers was done between peach and almond [62]. Therefore, based on the synteny pattern observed between *Prunus* species we tried to identify SNPs mapped in Japanese plum in genes localized in the V2.1 peach genome. We were able to find 510 putative orthologous genes by using SNPs from the peach genome as a reference. 38 genes out of the 510 putative orthologues were co-localized with two to three SNPs after genetic linkage analysis was carried out in our segregating F₁ population of Japanese plum. Our result suggests synteny between the Japanese plum and peach genomes (See Fig 3).

We found that 85.4% of the SNPs analyzed were located in intragenic regions. Pootakham et al., [63] and Bastien et al., [50] suggested that methylation-sensitive enzymes such as ApeKI increase the degree of enrichment in coding regions of the genome. More than 50% of the SNPs were located within or close to genic regions.

The number of SNPs identified in our study was superior to those reported by Klagges et al. [30] and Guajardo et al. [31] in sweet cherry (See Fig 2). We have obtained a highly saturated linkage map for *P. salicina* (Japanese plum) that will be important in breeding studies as well as for the genome sequencing of this species being developed around the world (Pacheco, I. and Silva, H., unpublished data; Fernandez, A., unpublished data).

Supporting information

S1 Table. Genetic distance between SNPs mapped by linkage groups. The information is given for the female, the male and the consensus map according to peach genome version v2.1. (DOCX)

S2 Table. SNPs information related to the linkage map and position on the peach genome version v2.1. Table A: Linkage map of female line cv ‘Angelino’. For each SNP is described his original linkage group, the position on the physical map and the position on the genetic map. Physical map position according to peach genome V2.1. **Table B: Linkage map of male line cv ‘Aurora’.** For each SNP is described its original linkage group, the position on the physical map and the position on the genetic map. Physical map position according to peach genome V2.1. **Table C: Consensus linkage map.** For each SNP is described its original linkage group, the position on the physical map, the position on the genetic map and corresponding orthologous genes that co-localize 2 to 3 SNPs simultaneously (n = 38). Physical map position according to peach genome V2.1. **Table D: Identification of the SNPs that co-localize 2 to 3 SNPs simultaneously over 38 genes.** Physical map position according to peach genome V2.1. **Table E: Functional annotation (Phytozome) of 38 putative orthologous genes that co-localize 2 to 3 SNPs simultaneously.** **Table F: SNPs identification over 460 putative orthologous genes without multiple SNP co-localization.** Physical map position according to peach genome V2.1. (XLSX)

Author Contributions

Conceptualization: Basilio Carrasco, Herman Silva.

Data curation: Basilio Carrasco, Máximo González, Jonathan Maldonado.

Formal analysis: Jonathan Maldonado.

Funding acquisition: Herman Silva.

Investigation: Basilio Carrasco, Herman Silva.

Methodology: Máximo González, Marlene Gebauer, Rolando García-González, Herman Silva.

Project administration: Herman Silva.

Software: Jonathan Maldonado.

Supervision: Herman Silva.

Validation: Jonathan Maldonado.

Writing – original draft: Herman Silva.

Writing – review & editing: Basilio Carrasco.

References

1. Carrasco B, Meisel L, Gebauer M, Gonzalez RG, Silva H. Breeding in peach, cherry and plum: from a tissue culture, genetic, transcriptomic and genomic perspective. *Biological Research*. 2013; 46(3): 219–30. <https://doi.org/10.4067/S0716-97602013000300001> PMID: 24346068
2. Okie WR, Weinberger JH. Plums. In: Janick J, Moore JN, editors. *Fruit breeding, Tree and Tropical Fruits*, 2nd ed. New York: John Wiley & Sons; 1995. p. 559–607.
3. Hancock JF. *Temperate fruit crop breeding: germplasm to genomics*. Hancock JF, editor. Netherland: Springer Science & Business Media; 2008. 1–455 p.
4. El-Sharkawy I, Kim WS, El-Kereamy A, Jayasankar S, Svircev AM, Brown DCW. Isolation and characterization of four ethylene signal transduction elements in plums (*Prunus salicina* L.). *Journal of experimental botany*. 2007; 58(13):3631–43. <https://doi.org/10.1093/jxb/erm213> PMID: 18057041
5. El-Sharkawy I, Sherif S, Mila I, Bouzayen M, Jayasankar S. Molecular characterization of seven genes encoding ethylene-responsive transcriptional factors during plum fruit development and ripening. *Journal of experimental botany*. 2009; 60(3):907–22. <https://doi.org/10.1093/jxb/ern354> PMID: 19213809
6. González M, Carrasco B, Salazar E. Genome-wide identification and characterization of R2R3MYB family in Rosaceae. *Genomics data*. 2016; 9:50–7. <https://doi.org/10.1016/j.gdata.2016.06.004> PMID: 27408811
7. González M, Maldonado J, Salazar E, Silva H, Carrasco B. De novo transcriptome assembly of ‘Angeleno’ and ‘Lamoon’ Japanese plum cultivars (*Prunus salicina*). *Genomics data*. 2016; 9:35–6. Epub 2016/07/14. <https://doi.org/10.1016/j.gdata.2016.06.010> PMID: 27408806
8. González M, Salazar E, Cabrera S, Olea P, Carrasco B. Analysis of anthocyanin biosynthesis genes expression profiles in contrasting cultivars of Japanese plum (*Prunus salicina* L.) during fruit development. *Gene Expression Patterns*. 2016; 21(1):54–62. <https://doi.org/10.1016/j.gep.2016.06.005> PMID: 27378315
9. González M, Salazar E, Castillo J, Morales P, Mura-Jornet I, Maldonado J, et al. Genetic structure based on EST–SSR: a putative tool for fruit color selection in Japanese plum (*Prunus salicina* L.) breeding programs. *Molecular Breeding*. 2016; 36(6):1–15. <https://doi.org/10.1007/s11032-016-0491-x>
10. Minas IS, Forcada CFI, Dangi GS, Gradziel TM, Dandekar AM, Crisosto CH. Discovery of non-climacteric and suppressed climacteric bud sport mutations originating from a climacteric Japanese plum cultivar (*Prunus salicina* Lindl.). *Frontiers in Plant Science*. 2015; 6. <https://doi.org/10.3389/fpls.2015.00316> PMID: 26029222
11. Diez-de-Medina S, Silva H. A molecular marker approach using intron flanking EST-PCR to map candidate genes in peach (*Prunus persica*). *Electronic Journal of Biotechnology*. 2012; 15(5). <https://doi.org/10.2225/vol15-issue5-fulltext-9>
12. Lande R, Thompson R. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics*. 1990; 124(3):743–56. PMID: 1968875
13. Dirlwanger E, Cosson P, Howad W, Capdeville G, Bosselut N, Claverie M, et al. Microsatellite genetic linkage maps of myrobalan plum and an almond-peach hybrid location of root-knot nematode resistance genes. *Theoretical and Applied Genetics*. 2004; 109(4):827–38. <https://doi.org/10.1007/s00122-004-1694-9> PMID: 15241595
14. Dirlwanger E, Graziano E, Joobeur T, Garriga-Caldere F, Cosson P, Howad W, et al. Comparative mapping and marker-assisted selection in Rosaceae fruit crops. *Proceedings of the National Academy of Sciences of the United States of America*. 2004; 101(26):9891–6. Epub 2004/05/26. <https://doi.org/10.1073/pnas.0307937101> PMID: 15159547
15. Arús P, Verde I, Sosinski B, Zhebentyayeva T, Abbott AG. The peach genome. *Tree Genetics & Genomes*. 2012; 8(3):531–47. <https://doi.org/10.1007/s11295-012-0493-8>
16. Verde I, Abbott AG, Scalabrin S, Jung S, Shu SQ, Marroni F, et al. The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nature genetics*. 2013; 45(5):487–94. <https://doi.org/10.1038/ng.2586> PMID: 23525075
17. Verde I, Jenkins J, Dondini L, Micali S, Pagliarani G, Vendramin E, et al. The Peach v2.0 release: high-resolution linkage mapping and deep resequencing improve chromosome-scale assembly and contiguity. *Bmc Genomics*. 2017; 18. <https://doi.org/10.1186/s12864-017-3606-9> PMID: 28284188
18. Zhang QX, Chen WB, Sun LD, Zhao FY, Huang BQ, Yang WR, et al. The genome of *Prunus mume*. *Nature Communications*. 2012; 3. <https://doi.org/10.1038/ncomms2290> PMID: 23271652

19. Sun LD, Yang WR, Zhang QX, Cheng TR, Pan HT, Xu ZD, et al. Genome-Wide Characterization and Linkage Mapping of Simple Sequence Repeats in Mei (*Prunus mume* Sieb. et Zucc.). PLoS one. 2013; 8(3). <https://doi.org/10.1371/journal.pone.0059562> PMID: 23555708
20. Vieira EA, Nodari RO, Dantas A, Ducroquet JPHJ, Dalbó M, Borges CV. Genetic mapping of Japanese plum. Crop Breed Appl Biotech. 2005; 5:29–37. <https://doi.org/10.12702/1984-7033.v05n01a04>
21. Salazar JA, Pacheco I, Shinya P, Zapata P, Silva C, Aradhya M, et al. Genotyping by Sequencing for SNP-Based Linkage Analysis and Identification of QTLs Linked to Fruit Quality Traits in Japanese Plum (*Prunus salicina* Lindl.). Frontiers in Plant Science. 2017; 8. <https://doi.org/10.3389/fpls.2017.00476> PMID: 28443103
22. Batley J, Barker G, O'Sullivan H, Edwards KJ, Edwards D. Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. Plant Physiology. 2003; 132(1):84–91. <https://doi.org/10.1104/pp.102.019422> PMID: 12746514
23. Batley J, Edwards D. SNP applications in plants. In: Oraguzie NC, Gardiner SE, Rikkerink EH, Silva HN, editors. Association mapping in plants. New York: Springer; 2007. p. 95–102.
24. Sonah H, Bastien M, Iqura E, Tardivel A, Legare G, Boyle B, et al. An Improved Genotyping by Sequencing (GBS) Approach Offering Increased Versatility and Efficiency of SNP Discovery and Genotyping. PLoS one. 2013; 8(1). <https://doi.org/10.1371/journal.pone.0054603> PMID: 23372741
25. Dirlwanger E, Cosson P, Tavaud M, Aranzana MJ, Poizat C, Zanetto A, et al. Development of micro-satellite markers in peach [*Prunus persica* (L.) Batsch] and their use in genetic diversity analysis in peach and sweet cherry (*Prunus avium* L.). Theoretical and Applied Genetics. 2002; 105(1):127–38. <https://doi.org/10.1007/s00122-002-0867-7> PMID: 12582570
26. Truong HT, Ramos AM, Yalcin F, de Ruiter M, van der Poel HJA, Huvenaars KHJ, et al. Sequence-Based Genotyping for Marker Discovery and Co-Dominant Scoring in Germplasm and Populations. PLoS one. 2012; 7(5). <https://doi.org/10.1371/journal.pone.0037565> PMID: 22662172
27. Yang HA, Tao Y, Zheng ZQ, Li CD, Sweetingham MW, Howieson JG. Application of next-generation sequencing for rapid marker development in molecular plant breeding: a case study on anthracnose disease resistance in *Lupinus angustifolius* L. BMC Genomics. 2012; 13. <https://doi.org/10.1186/1471-2164-13-318> PMID: 22805587
28. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS one. 2011; 6(5). <https://doi.org/10.1371/journal.pone.0019379> PMID: 21573248
29. Bielenberg DG, Rauh B, Fan SH, Gasic K, Abbott AG, Reighard GL, et al. Genotyping by Sequencing for SNP-Based Linkage Map Construction and QTL Analysis of Chilling Requirement and Bloom Date in Peach *Prunus persica* (L.) Batsch. PLoS one. 2015; 10(10). <https://doi.org/10.1371/journal.pone.0139406> PMID: 26430886
30. Klagges C, Campoy JA, Quero-Garcia J, Guzman A, Mansur L, Gratacos E, et al. Construction and comparative analyses of highly dense linkage maps of two sweet cherry intra-specific progenies of commercial cultivars. PLoS one. 2013; 8(1):e54743. <https://doi.org/10.1371/journal.pone.0054743> PMID: 23382953
31. Guajardo V, Solis S, Sagredo B, Gainza F, Munoz C, Gasic K, et al. Construction of High Density Sweet Cherry (*Prunus avium* L.) Linkage Maps Using Microsatellite Markers and SNPs Detected by Genotyping-by-Sequencing (GBS). PLoS one. 2015; 10(5). <https://doi.org/10.1371/journal.pone.0127750> PMID: 26011256
32. Ward JA, Bhangoo J, Fernandez-Fernandez F, Moore P, Swanson JD, Viola R, et al. Saturated linkage map construction in *Rubus idaeus* using genotyping by sequencing and genome-independent imputation. BMC Genomics. 2013; 14. <https://doi.org/10.1186/1471-2164-14-2> PMID: 23324311
33. Gardner KM, Brown P, Cooke TF, Cann S, Costa F, Bustamante C, et al. Fast and Cost-Effective Genetic Mapping in Apple Using Next-Generation Sequencing. G3-Genes Genomes Genetics. 2014; 4(9):1681–7. <https://doi.org/10.1534/g3.114.011023> PMID: 25031181
34. Elorriaga A. Actualidad de los carozos chilenos. Rev Frut Coopefrut SA. 2008;(3):4–10.
35. Robledo P, Becerra C, Manríquez D, Defillippi B. Caracterización fisiológica de diferentes variedades de ciruela y su implicancia en el manejo de postcosecha para la exportación a mercados distantes. Boletín INIA. 2011; 227:95–104.
36. Carrasco B, Garces M, Rojas P, Saud G, Herrera R, Retamales JB, et al. The Chilean strawberry *Fragaria chiloensis* (L.) Duch.: Genetic diversity and structure. Journal of the American Society for Horticultural Science. 2007; 132(4):501–6.
37. Goonetilleke SN, March TJ, Wirthensohn MG, Arus P, Walker AR, Mather DE. Genotyping by Sequencing in Almond: SNP Discovery, Linkage Mapping, and Marker Design. G3 (Bethesda, Md). 2018; 8(1): 161–72. Epub 2017/11/17. <https://doi.org/10.1534/g3.117.300376> PMID: 29141988

38. Gurcan K, Teber S, Ercisli S, Yilmaz KU. Genotyping by Sequencing (GBS) in Apricots and Genetic Diversity Assessment with GBS-Derived Single-Nucleotide Polymorphisms (SNPs). *Biochemical genetics*. 2016; 54(6):854–85. Epub 2016/07/29. <https://doi.org/10.1007/s10528-016-9762-9> PMID: 27465591
39. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25(14):1754–60. <https://doi.org/10.1093/bioinformatics/btp324> PMID: 19451168
40. Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, et al. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PloS one*. 2014; 9(2). <https://doi.org/10.1371/journal.pone.0090346> PMID: 24587335
41. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*. 2007; 23(19):2633–5. <https://doi.org/10.1093/bioinformatics/btm308> PMID: 17586829
42. Van Ooijen JW. JoinMap 4.0: software for the calculation of genetic linkage maps in experimental population. Kyazma BV. Wageningen, Netherlands; 2006.
43. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011; 27(15):2156–8. <https://doi.org/10.1093/bioinformatics/btr330> PMID: 21653522
44. Fumagalli M, Vieira FG, Linderoth T, Nielsen R. ngsTools: methods for population genetics analyses from next-generation sequencing data. *Bioinformatics*. 2014; 30(10):1486–7. <https://doi.org/10.1093/bioinformatics/btu041> PMID: 24458950
45. Kosambi DD. The Estimation of Map Distances from Recombination Values. *Annals of Eugenics*. 1943; 12(1):172–5. <https://doi.org/10.1111/j.1469-1809.1943.tb02321.x>
46. Tanksley SD, Young ND, Paterson AH, Bonierbale MW. RFLP mapping in plant-breeding—New Tools for and old Science. *Bio-Technology*. 1989; 7(3):257–64. <https://doi.org/10.1038/nbt0389-257>
47. Lu F, Romay MC, Glaubitz JC, Bradbury PJ, Elshire RJ, Wang TY, et al. High-resolution genetic mapping of maize pan-genome sequence anchors. *Nature Communications*. 2015; 6. <https://doi.org/10.1038/ncomms7914> PMID: 25881062
48. Ott J, Wang J, Leal SM. Genetic linkage analysis in the age of whole-genome sequencing. *Nature Reviews Genetics*. 2015; 16(5):275–84. <https://doi.org/10.1038/nrg3908> PMID: 25824869
49. Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*. 2011; 12(7):499–510. <https://doi.org/10.1038/nrg3012> PMID: 21681211
50. Bastien M, Boudhrioua C, Fortin G, Belzile F. Exploring the potential and limitations of genotyping-by-sequencing for SNP discovery and genotyping in tetraploid potato. *Genome*. 2018; 61(6):449–56. Epub 2018/04/25. <https://doi.org/10.1139/gen-2017-0236> PMID: 29688035
51. Jiang ZH, Wang HY, Michal JJ, Zhou X, Liu B, Woods LCS, et al. Genome Wide Sampling Sequencing for SNP Genotyping: Methods, Challenges and Future Development. *International Journal of Biological Sciences*. 2016; 12(1):100–8. <https://doi.org/10.7150/ijbs.13498> PMID: 26722221
52. Ma JQ, Huang L, Ma CL, Jin JQ, Li CF, Wang RK, et al. Large-Scale SNP Discovery and Genotyping for Constructing a High-Density Genetic Map of Tea Plant Using Specific-Locus Amplified Fragment Sequencing (SLAF-seq). *PloS one*. 2015; 10(6). <https://doi.org/10.1371/journal.pone.0128798> PMID: 26035838
53. Sun XW, Liu DY, Zhang XF, Li WB, Liu H, Hong WG, et al. SLAF-seq: An Efficient Method of Large-Scale De Novo SNP Discovery and Genotyping Using High-Throughput Sequencing. *PloS one*. 2013; 8(3):e58700. Epub 2013/03/26. <https://doi.org/10.1371/journal.pone.0058700> PMID: 23527008
54. Di Pierro EA, Gianfranceschi L, Di Guardo M, Koehorst-van Putten HJJ, Kruijselbrink JW, Longhi S, et al. A high-density, multi-parental SNP genetic map on apple validates a new mapping approach for outcrossing species. *Horticulture Research*. 2016; 3. <https://doi.org/10.1038/hortres.2016.57> PMID: 27917289
55. Li LT, Deng CH, Knabel M, Chagne D, Kumar S, Sun JM, et al. Integrated high-density consensus genetic map of *Pyrus* and anchoring of the ‘Bartlett’ v1.0 (*Pyrus communis*) genome. *DNA Research*. 2017; 24(3):289–301. <https://doi.org/10.1093/dnares/dsw063> PMID: 28130382
56. Martinez-Garcia PJ, Fresnedo-Ramirez J, Parfitt DE, Gradziel TM, Crisosto CH. Effect prediction of identified SNPs linked to fruit quality and chilling injury in peach *Prunus persica* (L.) Batsch. *Plant Molecular Biology*. 2013; 81(1–2):161–74. <https://doi.org/10.1007/s11103-012-9989-8> PMID: 23184287
57. Zhang J, Zhang QX, Cheng TR, Yang WR, Pan HT, Zhong JJ, et al. High-density genetic map construction and identification of a locus controlling weeping trait in an ornamental woody plant (*Prunus mume* Sieb. et Zucc). *DNA Research*. 2015; 22(3):183–91. <https://doi.org/10.1093/dnares/dsv003> PMID: 25776277

58. Slate J, Gratten J, Beraldi D, Stapley J, Hale M, Pemberton JM. Gene mapping in the wild with SNPs: guidelines and future directions. *Genetica*. 2009; 136(1):97–107. <https://doi.org/10.1007/s10709-008-9317-z> PMID: 18780148
59. Basantani MK, Gupta D, Mehrotra R, Mehrotra S, Vaish S, Singh A. An update on bioinformatics resources for plant genomics research. *Current Plant Biology*. 2017; 11–12:33–40. <https://doi.org/10.1016/j.cpb.2017.12.002>
60. Eichler EE, Clark RA, She X. An assessment of the sequence gaps: unfinished business in a finished human genome. *Nature reviews Genetics*. 2004; 5(5):345–54. Epub 2004/05/15. <https://doi.org/10.1038/nrg1322> PMID: 15143317
61. Tavassolian I, Rabiei G, Gregory D, Mnejja M, Wirthensohn MG, Hunt PW, et al. Construction of an almond linkage map in an Australian population Nonpareil x Lauranne. *Bmc Genomics*. 2010; 11. <https://doi.org/10.1186/1471-2164-11-551> PMID: 20932335
62. Ricciardi F, Del Cueto J, Bardaro N, Mazzeo R, Ricciardi L, Dicenta F, et al. Synteny-Based Development of CAPS Markers Linked to the Sweet kernel LOCUS, Controlling Amygdalin Accumulation in Almond (*Prunus dulcis* (Mill.) D.A.Webb). *Genes*. 2018; 9(8). Epub 2018/08/02. <https://doi.org/10.3390/genes9080385> PMID: 30065184
63. Pootakham W, Sonthirod C, Naktang C, Jomchai N, Sangsrakru D, Tangphatsornruang S. Effects of methylation-sensitive enzymes on the enrichment of genic SNPs and the degree of genome complexity reduction in a two-enzyme genotyping-by-sequencing (GBS) approach: a case study in oil palm (*Elaeis guineensis*). *Molecular breeding: new strategies in plant improvement*. 2016; 36(11):154. Epub 2016/12/13. <https://doi.org/10.1007/s11032-016-0572-x> PMID: 27942246