



Development and Evaluation of an Axiom™ 60K SNP Array for Almond (*Prunus dulcis*)

Henri Duval, Eva Coindre, Sebastian E Ramos-Onsins, Konstantinos G Alexiou, Maria J Rubio-Cabetas, Pedro J Martínez-García, Michelle Wirthensohn, Amit Dhingra, Anna Samarina, Pere Arús

► To cite this version:

Henri Duval, Eva Coindre, Sebastian E Ramos-Onsins, Konstantinos G Alexiou, Maria J Rubio-Cabetas, et al.. Development and Evaluation of an Axiom™ 60K SNP Array for Almond (*Prunus dulcis*). *Plants*, 2023, 12 (2), pp.242. 10.3390/plants12020242 . hal-03935771

HAL Id: hal-03935771

<https://hal.science/hal-03935771>

Submitted on 13 Jan 2023



HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Development and Evaluation of an Axiom™ 60K SNP Array for Almond (*Prunus dulcis*)

Henri Duval ^{1,*} , Eva Coindre ¹ , Sebastian E. Ramos-Onsins ² , Konstantinos G. Alexiou ^{2,3} , Maria J. Rubio-Cabetas ⁴ , Pedro J. Martínez-García ⁵ , Michelle Wirthensohn ⁶ , Amit Dhingra ⁷ , Anna Samarina ⁸ and Pere Arús ^{2,3}

- ¹ Unité de Génétique et Amélioration des Fruits et Légumes (GAFL), INRAE (French National Research Institute for Agriculture, Food and Environment), 84143 Montfavet, France
 - ² Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Carrer de la Vall Moronta, Edifici CRAG, Campus UAB, Cerdanyola del Valles, 08193 Barcelona, Spain
 - ³ IRTA (Institute of Agrifood Research and Technology), Campus UAB, Edifici CRAG, Cerdanyola del Valles (Bellaterra), 08193 Barcelona, Spain
 - ⁴ CITA (Agrifood Research and Technology Centre of Aragon), Department of Plant Science, Avda. Montañana 930, 50059 Zaragoza, Spain
 - ⁵ CEBAS (Centro de Edafología y Biología Aplicada del Segura), CSIC, Department of Plant Breeding, Campus Universitario de Espinardo, 30100 Espinardo, Spain
 - ⁶ Waite Research Institute, University of Adelaide, PMB 1 Glen, Osmond, SA 5064, Australia
 - ⁷ Department of Horticulture, Washington State University, Pullman, WA 99164-6414, USA
 - ⁸ Thermo Fisher Scientific, Frankfurter Str. 129B, 64293 Darmstadt, Germany
- * Correspondence: henri.duval@inrae.fr; Tel.: +33-432722667

Abstract: A high-density single nucleotide polymorphism (SNP) array is essential to enable faster progress in plant breeding for new cultivar development. In this regard, we have developed an Axiom 60K almond SNP array by resequencing 81 almond accessions. For the validation of the array, a set of 210 accessions were genotyped and 82.8% of the SNPs were classified in the best recommended SNPs. The rate of missing data was between 0.4% and 2.7% for the almond accessions and less than 15.5% for the few peach and wild accessions, suggesting that this array can be used for peach and interspecific peach × almond genetic studies. The values of the two SNPs linked to the *RMja* (nematode resistance) and *SK* (bitterness) genes were consistent. We also genotyped 49 hybrids from an almond F2 progeny and could build a genetic map with a set of 1159 SNPs. Error rates, less than 1%, were evaluated by comparing replicates and by detection of departures from Mendelian inheritance in the F2 progeny. This almond array is commercially available and should be a cost-effective genotyping tool useful in the search for new genes and quantitative traits loci (QTL) involved in the control of agronomic traits.

Keywords: *Prunus dulcis*; *Prunus persica*; SNP; Axiom array; genotyping



Citation: Duval, H.; Coindre, E.; Ramos-Onsins, S.E.; Alexiou, K.G.; Rubio-Cabetas, M.J.; Martínez-García, P.J.; Wirthensohn, M.; Dhingra, A.; Samarina, A.; Arús, P. Development and Evaluation of an Axiom™ 60K SNP Array for Almond (*Prunus dulcis*). *Plants* **2023**, *12*, 242. <https://doi.org/10.3390/plants12020242>

Academic Editor: Alexey A. Dmitriev

Received: 2 November 2022

Revised: 24 December 2022

Accepted: 29 December 2022

Published: 5 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The detection of single nucleotide markers (SNPs) has been enhanced with the Next-Generation Sequencing (NGS) technologies [1,2]. SNPs used as molecular markers in linkage mapping studies and genome wide association studies (GWAS) have enabled the detection of novel genes or quantitative trait loci (QTL). The SNP array technologies have facilitated the SNP genotyping by providing complete and reliable genotypic data at an acceptable cost [3].

Several high density SNP arrays have been developed in fruit species, either with the Illumina technology in the 9K [4] and 18K SNP peach array [5] and the 15K cherry SNP array [6], or with the Axiom Affymetrix technology in the 480K apple SNP array [7], the 70K and 200K *Pyrus* SNP array [8,9] or the 700K SNP walnut array [10].

Based on the strong synteny between peach (*Prunus persica*) and almond (*Prunus dulcis*), the Illumina 18K peach SNP array was used in a GWAS analysis of a collection of Italian almond germplasm [11]. The authors obtained an average marker density of 1 SNP per 424 kb across the almond genome. With the development of an almond-specific array, this should improve the marker density and cover the whole almond genome, in particular its almond-specific regions.

The whole genome sequences of *Prunus dulcis* cv “Texas” [12] and *Prunus dulcis* cv “Lauranne” [13] were published and are available on the GDR database (<https://www.rosaceae.org/analysis/295> accessed on 1 January 2021), and recently another whole genome cv “Nonpareil” was also published [14]. In addition to the sequencing of the “Texas” genome, the “almond genome” consortium has resequenced several almond cultivars using paired-end Illumina sequencing, allowing the detection of a high number of SNPs and InDel variants.

In this study, we present the development of an Axiom 60K almond SNP array (PdSNP_V1) based on the SNP polymorphism of a set of 81 resequenced almond accessions and the validation and accuracy of the genotyping data generated from the array.

2. Results–Discussion

2.1. Array Design

2.1.1. SNPs Array Selection

Alignment of the NGS reads of the 81 almond accessions to the “Texas” reference genome [12] and SNP calling with GATK (Section 3.2) allowed us to detect 11,683,655 SNPs. We then applied three different steps (Figure 1), to select SNPs for the chip. In the first step with the filters described in Section 3.2, we obtained 576,968 SNPs with the threshold of less than 20% missing data (NA) and 71,843 SNPs for less than 10% and we retained the 10% NA filter. We did not discard the A/T and C/G SNPs although they require two probes per SNP in the Axiom probe sets. In the second step, we included the three additional SNPs published as trait markers of interest and detailed in Section 3.2 and Table 1.

Table 1. Position of the three SNPs linked to different traits.

SNP Name	Chr	Position_V2	Position_V1	SNP_Ref	SNP_Alt	Trait	Gene
AX-599403222	Pd07	9,019,871	7,763,091	A	G	RK Nematode R	<i>RMja</i>
AX-599403226	Pd01	3,296,885	2,496,687	T	C	Kernel weight	
AX-599403227	Pd05	11,867,647	11,521,869	C	T	Bitter taste	<i>Sk</i>

V1: *Prunus dulcis* Lauranne Genome v1.0. a1; V2: *Prunus dulcis* Texas Genome v2.0.

The final filter on step 3 was based on the p-convert score estimated by Affymetrix. Prior to chip production, an in-silico validation was performed to estimate the p-convert value for the forward and reverse probes for each SNP. The p-convert value is derived from a random forest model to predict the probability of SNP conversion on the array. A high p-convert value means that we expect a high probability of conversion for the SNP probes on the array. SNPs with a p-convert of less than 0.666 were removed, discarding 11,265 SNPs. In addition, Affymetrix had classified SNPs into three categories, recommended, not recommended and neutral (Figure 2). Only SNPs classified as recommended were retained for chip production. In the final selection, 60,581 SNPs that met all the criteria were used to construct the almond chip.

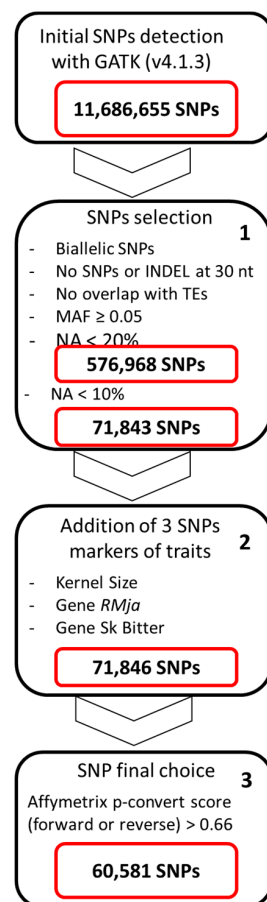


Figure 1. Pipeline of SNP detection, selection and final choice with all filters applied to design the Almond SNP array. (TE: Transposon Element, NA: missing data, MAF: Minor Allele Frequency).

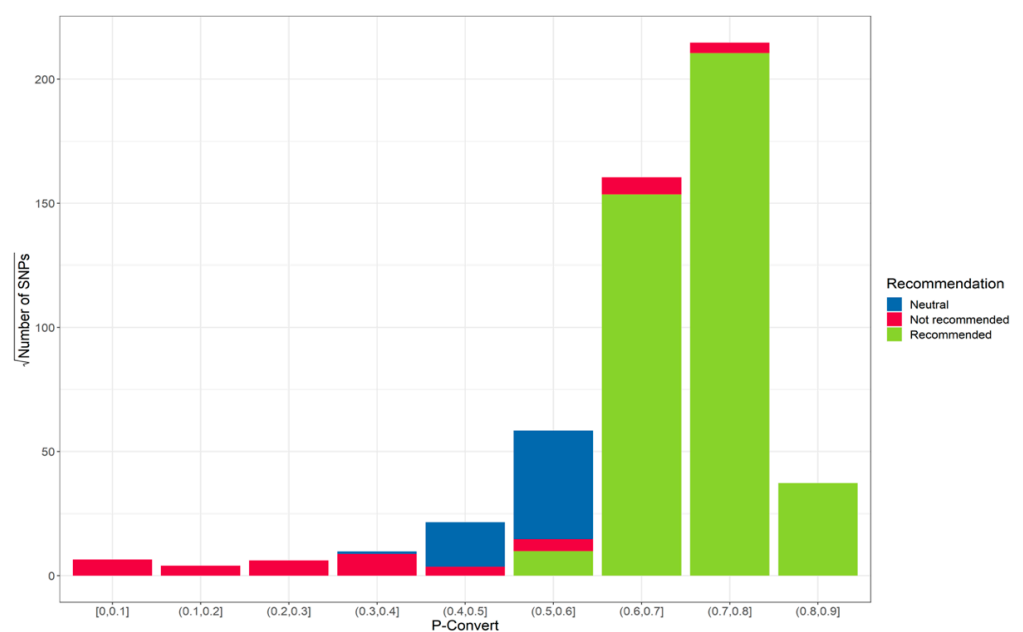


Figure 2. Distribution of the square root of the number of SNPs based on p-convert score and colored by recommendation from Affymetrix in-silico analyses. (P-convert = Probability of conversion based on the best selected strand of the probe sequence.)

2.1.2. Distribution of Array SNPs

To explore the distribution of the 60,581 selected SNPs, SNP frequency was plotted at every one million base pairs along each chromosome. (Figure 3). We have also plotted the distribution of all 576,968 SNPs detected after step 1 of the pipeline selection. The density of SNPs seems highest along the chromosome arms and lowest in the centromeric regions for both distributions. Over these million-base-pair intervals, the number of SNPs on the chip ranges from 22 to 799 (Table S1) and there are no SNP-free intervals. The average distance between two SNPs is 2891 bp.

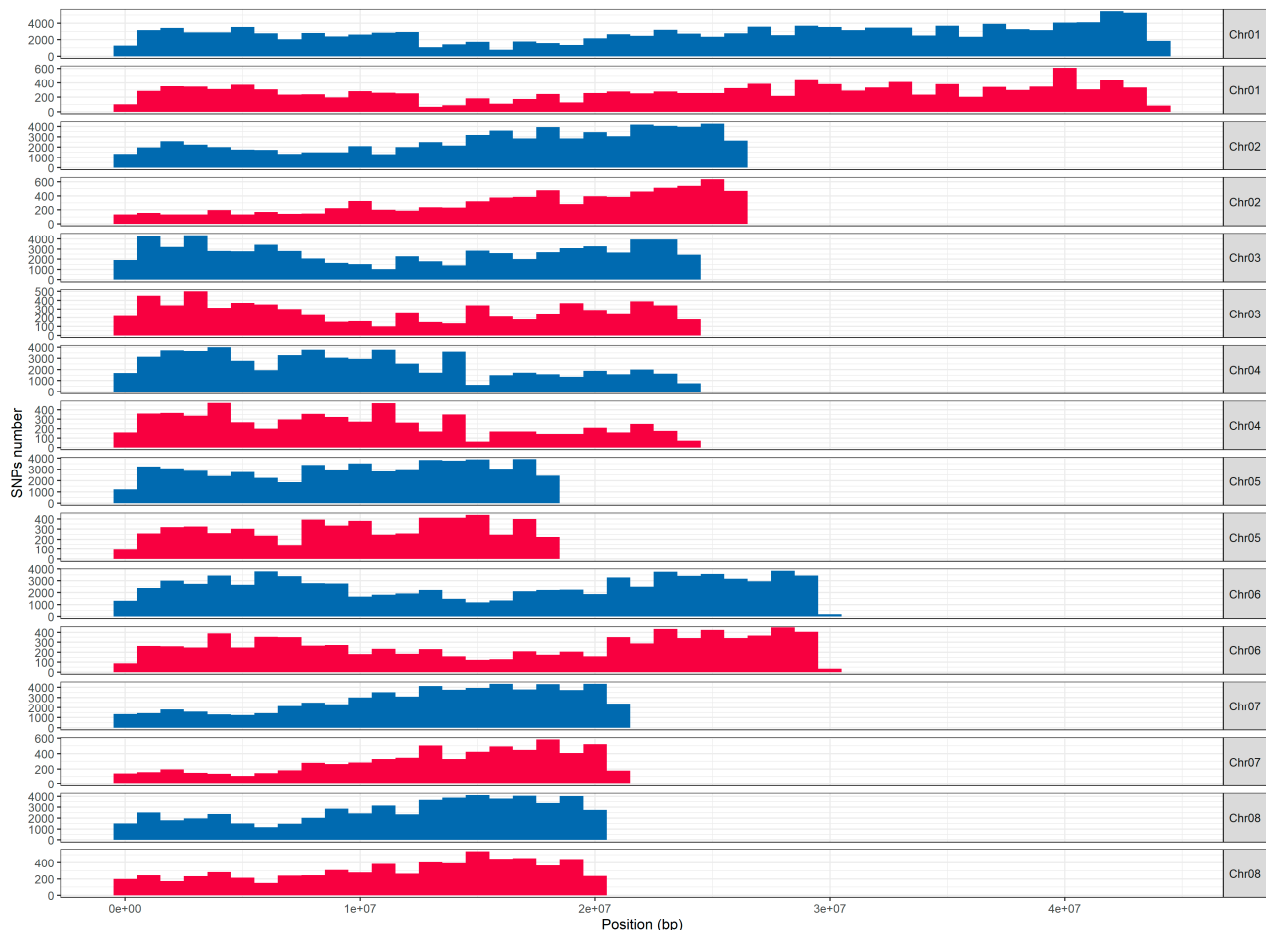


Figure 3. Distribution of SNPs along each chromosome. Each bar represents an interval of 1 Mbp. All the 576,968 SNPs filtered after step 1 are plotted in blue and the 60,581 array SNPs are plotted in red. The abscissa represents the physical position.

2.2. Validation of the Axiom 60K Almond Array

2.2.1. Validation of the Array on the Diversity Panel

To evaluate and validate the Axiom 60K almond SNP array, we genotyped a diversity panel composed of 210 accessions including 187 genotypes of *P. dulcis* and 23 genotypes of species of the subgenus *Amygdalus* including some peach genotypes and interspecific hybrids. A total of 47,012 SNPs (77.6%) were classified in the Poly High Resolution (PHR) Affymetrix category, considered the most informative with good cluster resolution (Table 2). Only 0.1% (85 SNPs) were classified in the “Mono High Resolution” (MHR) category, which means that all samples are monomorphic for these SNPs. A total of 5.1% were classified in the “No Minor Homozygous” (NMH) category which includes SNPs with good cluster resolution, but no minor homozygous genotypes. The other categories “Call Rate Below Threshold” (CBT), “Off-Target Variant” (OTV), Other (low quality), not recommended for downstream analysis, accounted for 17.2% of SNPs.

Table 2. Repartition of SNPs in the six categories.

SNP Category	Number	Percent
Poly High Resolution (PHR)	47,012	77.6
Mono High Resolution (MHR)	85	0.1
No Minor Hom (NMH)	3074	5.1
Call Rate Below Threshold (CBT)	2435	4
Off-Target Variant (OTV)	3425	5.7
Other	4550	7.5
Total	60,581	

For all 60,581 SNPs, we observed between 262 (0.4%) and 1665 (2.7%) missing data for the 185 *P. dulcis* accessions (Table 3). In comparison, in the data generated by NGS resequencing of the 55 almond accessions resequenced by the consortium, we obtained between 1.6% and 7.6% missing data, showing a higher rate of missing data than in the chip data. For the wild almond species (*P. bucharica*, *P. fenziiana*, *P. kuramica*, *P. dehiscens*, *P. webbii*), the array data are also valuable with a similar frequency of missing data as *P. dulcis*, but with a lower percentage of heterozygous SNPs. For accessions of the two peach species, *P. persica* and *P. davidiana*, the rate of missing data was slightly higher (11.7–15.5%), but it is still a low rate. With a relatively high number of heterozygous SNPs, this suggests that the array has the potential to be used for diversity analysis and linkage mapping of peach trees, and specifically for linkage mapping of peach × almond progenies, useful for gene transfers from almond to peach or vice versa [15].

Table 3. Genotyping results of some *Prunus* species and interspecific hybrid accessions with the almond array. (NA: Missing data, He: Heterozygote).

Species	Accessions	NA Number	NA Percent	He Number	He Percent
<i>P. dulcis</i>	185 *	262–1665	0.4–2.7	7077–14,065	11.9–23.4
<i>P. bucharica</i>	1	3099	5	3242	6
<i>P. fenziiana</i>	9	1448–1937	2.4–3.2	3516–4267	5.9–7.3
<i>P. kuramica</i>	1	4212	7	3599	6
<i>P. dulcis</i> × <i>P. dehiscens</i>	2	576–1208	1–2	10,126–11,792	17.1–19.7
<i>P. dulcis</i> × <i>P. webbii</i>	1	570	1	14,976	25
<i>P. persica</i>	2	7103–9412	11.7–15.5	4965–5897	9.3–11.5
<i>P. davidiana</i>	2	5610–5671	9.3–9.4	4598–4822	8.4–8.8
<i>P. persica</i> × <i>P. dulcis</i>	3	1240–2003	2–3.3	6928–11,277	11.8–19
<i>P. bucharica</i> × <i>P. persica</i>	1	2130	4	4643	8
<i>P. pedunculata</i>	1	12,806	21	6809	14

* There were 2 other almond accessions with 7.3% and 12.1% of missing data.

2.2.2. Validation of the Three Additional SNPs

In the almond diversity panel analysis, the three additional SNPs were classified in the category Poly High Resolution (PHR) and were considered “best recommended” (Figure S1). For the SNP marker AX-599403222 of the *RMja* gene [16], only the two resistant accessions R1107 (Alnem-1) and R1109 (Alnem-88) were genotyped Homozygote with the alternative allele, which was consistent with their resistant phenotypes [17]. Three wild accessions and seven almond accessions were genotyped heterozygote for this SNP marker. As the *RMja* gene is a dominant gene, these genotypes would be resistant, which will be validated by an inoculation test with root-knot nematodes. All other almond accessions were genotyped homozygous susceptible. This is consistent with published data [18] and results obtained in resistance tests of four almond accessions (Ferragnès, Ferrastar, Dorée and Pointue d’Aureille) against *M. javanica* (data not published). The *RMja* gene is an orthologue of the plum gene *Ma* [19] and the genotyping with the almond chip of the three resistant clones P1079 (*MA/MA*), P2175 (*Ma/ma*) and P2980 (*Ma/ma*) was well determined by the SNP AX-599403222.

For the SNP AX-599403226, including in the *Prudu_000307-V1.0* gene, which could be associated with kernel weight [20], we found a segregation AA/AB/BB of 13/45/152 in the

panel (Figure S1B). With our kernel weight data available on some almond accessions, we did not observe statistically significant differences between the three classes, so we could not confirm the association. However, this SNP is classed in the Poly High Resolution class h justifying it to be a good SNP marker.

For the SNP AX-599403227 located in the region of the *Sk* gene involved in the bitter taste of almond [13], we found segregation AA/AB/BB of 70/120/20 for the panel of 217 accessions. The majority of almond varieties are sweet according the genotyping score (AA or AB), confirming previous studies. As previously observed by Sanchez et al. [21], the percentage of bitter phenotypes was very low, as a result of the selection against the recessive allele of the sweet kernel (*Sk*) in the breeding program. Twenty accessions were scored BB (Table S2) and amongst them, there were the five bitter accessions of the panel, but also 15 sweet accessions. This discrepancy in “Atocha” was observed previously by Sánchez-Pérez et al. [13]; a different SNP in the same candidate gene (*bHLH2*) could be the main reason for the sweet phenotype according these authors. In the case of “Texas” and the other sweet accessions that scored ‘BB’, the observed discrepancy could be explained by a different origin of the sweet phenotype, as result of a different SNP mutation in the same or other unknown gene; this still remains unclear. In conclusion, this marker can confirm that with the genotype AA or AB, the variety is sweet, but with the genotype BB we could not confirm it is bitter. However, this SNP has a good profile and high resolution, and a good marker of this region (Figure S1C).

2.2.3. Validation of the Array with a F2 Progeny

To construct a genetic map for the F2 almond progeny, we first selected heterozygous SNPs of the F1 parent “Penta” from the files generated by the almond chip and analyzed with “Axiom Analyze Suite”. These SNPs were classified in the two categories Poly High Resolution (PHR) (10,624 SNPs) and No Minor Hom (50 SNPs). We retained only 7234 intercross SNPs with a co-dominant 1:2:1 segregation, with a p-value less than or equal to 0.01 for a chi-squared goodness-of-fit test [22]. Following the hypothesis of no dominance between alleles, we coded these markers as *hkxhk*, as in a cross-pollination (CP) in JoinMap® 4.1 [23]. For grouping, an LOD ≥ 4 was used. The Maximum Likelihood (ML) mapping algorithm and the Kosambi mapping function implemented in JoinMap® were applied to order and determine the genetic location of markers. In order to keep the most reliable markers, we chose to remove markers mapping far from any other marker with a threshold of 5 cM (centimorgan). Finally, the selection and the construction of the genetic map were performed in JoinMap® 4.1, resulting in a final set of 1159 SNPs (Table 4). The map size was 876 cM and the mean distance between two SNP markers was 0.76 cM.

Table 4. Efficient markers to create genetic map for a F2 biparental population. (LG: Linkage group, He = Heterozygote, Nb = Number.)

LG	SNP Nb on “Penta”		Map	
	He	$p\text{-Value } (\chi^2) \leq 0.01$	Marker Nb	Size (in cM)
1	2055	1541	261	208
2	1339	534	81	69
3	1146	1146	166	108
4	1302	574	96	89
5	1042	451	87	96
6	1456	1402	190	114
7	1176	841	145	99
8	1158	745	133	94
Total	10,674	7234	1159	876

Our results show that the almond chip PdSNPV1 can be useful for constructing high-density linkage maps for almond and peach \times almond populations. Some linkage maps with SNP markers have already been constructed in almond and peach \times almond

populations using the SNP Illumina peach chip [24] or genotyping by sequencing (GBS) technology [25,26], but this almond chip could generate more specific almond SNP markers with greater ease and lower cost.

2.3. Evaluation of the Error Rates in Different Analyses

2.3.1. Comparison of “Texas” Genotyping with the “Texas” Reference Genome

The genotypic array data from the “Texas” accession was compared to the data of the “Texas” reference genome. Table 5 represents the number of SNPs that belong to one of the three allelic states 0, 1 or 2 with “0” for no common allele with the reference genome (alternative homozygote), “1” for only one common allele with the reference genome (heterozygote), “2” for two common alleles with the reference genome (reference homozygote). With the array data of “Texas”, we found 109 SNPs as alternative homozygotes to the reference genome. They are considered genotyping errors, meaning that the error rate was around 0.2% in the “Texas” array data. This could be also considered clonal variation, between the two different “Texas” clones used in the DNA extraction. The number of heterozygotes (10,416) is high but within the range observed in the interval observed in the 185 almond accessions of Table 3.

Table 5. Number of SNPs for each allelic state for the genotyped “Texas” accession.

Allele State *	Number	Percent
0	109	0.2
1	10,416	17.2
2	49,765	82.1
NA	291	0.5
Total	60,581	

* allele common number with the reference “Texas”, NA: missing data.

2.3.2. Comparison of Array Data from Three Replicates of the “Ferrastar” Variety

To assess the repeatability of the genotyping, we tested three “Ferrastar” samples. The differences of allelic state between each replicate of Ferrastar were the consequence of the natural error rate belonging to the Axiom technology. There were 0.90% and 0.95% error rates between the first replicate and the two other replicates and 1.01% between the second and the third replicate (Table 6), which is considered a low error rate.

Table 6. Error numbers between each replicate of Ferrastar.

Replicate	R2	R3
R1	549 (0.90%)	576 (0.95%)
R2	-	615 (1.01%)

2.3.3. Comparison between Expected and Observed Homozygote F2 Data

For all homozygous SNP types detected in the F1 parent “Penta”, the 49 F2 hybrids were expected to have the same allelic status as the parent. With 20,462 monomorphic markers AA in Penta for the A allele of “Texas”, we expected only AA datapoints for these 20,462 markers in the 49 F2 hybrids, but 494 were genotyped AB and 2167 were BB on the 1,002,203 observed data points corresponding to 0.27% erroneous data (Table 7). For the alternative B allele of “Texas”, there were 22,276 monomorphic markers BB in Penta, and in the 49 hybrids, 1324 data points were AA and 332 were AB on the 1,091,205 data points corresponding to 0.15% erroneous data. The error rate was lower than in the other two previous comparisons and also lower than the results presented in the 70k SNP Pyrus chip [8] where there was an error rate of 3%.

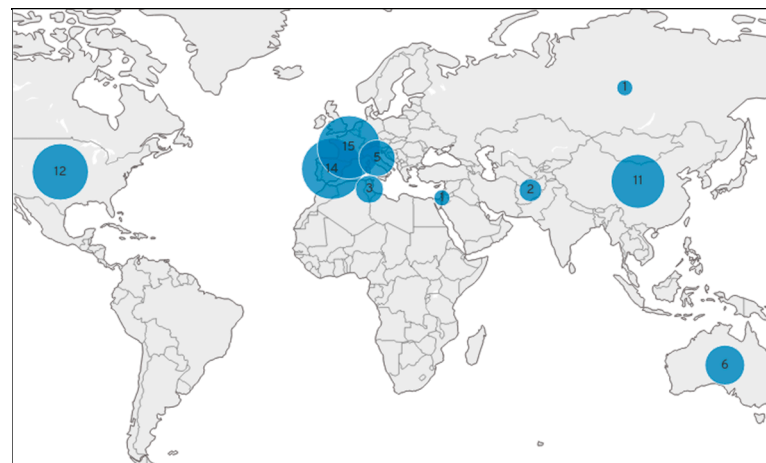
Table 7. Number of monomorphic SNPs in the F1 Penta and erroneous data points in the 49 hybrids of the F2 progeny.

F1 (“Penta”)		F2 Progeny				Erroneous Data Points	
	Marker Nb	AA	AB	BB	Total	Nb	%
AA	20,462	999,542	494	2167	1,002,203	2661	0.27
BB	22,276	1324	332	1,089,549	1,091,205	1656	0.15

3. Materials and Methods

3.1. Plant Material Resequenced

We resequenced 50 accessions of *P. dulcis* and 5 of wild species, two of *P. webbii* and one of *P. bucharica*, *P. kuramica*, and an interspecific hybrid *P. fenziiana* × *P. bucharica*. Forty-two accessions were resequenced on Illumina® HiSeq2000 by the CNAG-CRG Barcelona (<https://www.cnag.crg.eu/> accessed on 1 June 2020) and the 14 others were resequenced on Novaseq 6000 Illumina at the MGX platform at Montpellier (<https://www.mgx.cnrs.fr/> accessed on 1 January 2021)). For increasing the panel, we downloaded the sequences available on NCBI, 14 published by Yu et al. [27], 3 by Koepke et al. [28], and 9 by Velasco et al. [29] (Table S3). This panel of 81 accessions included a wide genetic diversity encompassing American, Australian, French, Italian, Spanish and Russian origins (Figure 4) and the three cultivars: “Tuono”, “Cristomorto” and “Nonpareil” which are the main genitors used in almond breeding programs [30].

**Figure 4.** Origins of the resequenced accessions.

All fastq files were trimmed using Trim Galore! version 0.6.1 [31] and their quality control was checked before and after trimming using FastQC v0.11.5 [32]. The trimmed fastq files for the 81 resequenced accessions were aligned to the “Texas” almond reference genome v2.0 [12] (<https://www.rosaceae.org/> accessed on 1 January 2021) using BWA version 0.7.16a-r1181 [33]. Samtools (v1.9) [34] was used to convert SAM into BAM format, sort and index the BAM files.

3.2. SNP Calling and Selection of SNPs for Array Development

All fastq files were trimmed using Trim Galore! version 0.6.1 [31] and their quality control was checked before and after trimming using FastQC v0.11.5 [32]. The trimmed fastq files for the 81 resequenced accessions were aligned to the “Texas” almond reference genome v2.0 [12] (<https://www.rosaceae.org/> accessed on 1 January 2021) using BWA version 0.7.16a-r1181 [33]. We used GATK v4.1.3 [35] to perform the variant calling with the following VariantFiltration expression: QD (QualityByDepth) < 2.0, SOR (StrandOddsRatio) > 3.0, FS (FisherStrand) > 60.0, MQ(RMSMappingQuality) < 40.0, MQRankSum < −12.5, ReadPosRankSum < −8.0, ExcessHet > 54.69, DP (Depth) > 10.

After this initial filtering, we applied three more selection steps (Figure 1). In the first step, SNP selection was done for biallelic SNPs, absence of other SNPs around 30 nucleotides, not overlapping with a transposable element, no missing data >10% and a Minor Allele Frequency (MAF) higher than 5%. In the second step, we included 3 additional SNPs published as markers for interesting traits (Table 1). The SNP AX-599403222 is the SNP of the KASP marker SP903 designed in the root knot nematode resistance *RMja* gene [16]. The AX-599403226 SNP is a SNP detected in the region of the mutation of the bHLH (basic helix-loop-helix) transcription factor (gene *Sk*) involved in the bitter taste of almond [13]. The last SNP, AX-599403226 located on the group 1, was detected by a GWAS signal for kernel weight [36]. We were not aware of any other SNP markers linked to other agronomic traits in almond and only these 3 SNPs were included in the chip. In the last selection step, the selected SNPs and corresponding flanking sequences were submitted to Affymetrix for initial probe screening and to estimate a P-converge score for each SNP.

3.3. Plant Material for the 60K SNP Array Validation

To evaluate and validate the Axiom 60K SNP almond array, a diversity panel was composed including 210 accessions made up of 187 of *P. dulcis* from the INRAE and IRTA collections, and other accessions of species of the subgenus *Amygdalus* (Table 3). Three replicates of the accession “Ferrastar” and the reference “Texas” were included in this panel. We also used an F2 progeny of 49 individuals from a self-pollination of the self-fertile Spanish accessions “Penta” issued from the cross “Lauranne” × S5133 [37], and considered as F1 hybrid.

3.4. Genotyping and SNP Analysis

Genomic DNA was extracted from young leaves according to the protocol of Ananaviciute et al. [38]. The amount of DNA required for genotyping with the array was 50 µL/genotype at 10 ng/µL. Genotyping was performed on an Axiom GeneTitan system on the INRAE Gentyane platform at Clermont-Ferrand (<https://gentyane.clermont.inrae.fr/> accessed on 1 November 2021).

After genotyping accessions, genotype calling and QC metrics were performed with SNP Axiom Analysis Suite v5.1 (Affymetrix) using diploid threshold configurations and default DishQC settings (DQC ≥ 0.82 and call rate > 0.97).

The software classified SNPs into six categories: Poly High Resolution (PHR) when we observed a good resolution for the two homozygous clusters and at least two occurrences of the minor allele, Mono High Resolution (MHR) when the SNP passes all thresholds except the number of minor alleles., meaning that all genotyped samples are monomorphic, No Minor Hom (NMH) when one of the homozygote clusters was not observed; Off-Target Variant (OTV) when the SNP has a sequence significantly different from the sequence of the hybridization probes caused for example by double deletions or sequence non-homology, Call Rate Below Threshold (CBT) when the SNP call rate is below the ratio of 97, meaning that the software has failed to identify clusters, “Other” when the SNP does not satisfy all quality thresholds.

4. Conclusions

We have developed and validated a new Axiom 60K almond SNP array (PdSNP_V1) and our studies have shown that the genotyping is reliable, with few missing data and erroneous points. Based on F2 progeny, we found only a few errors indicating departures from Mendelian inheritance of homozygote SNP loci.

The PdSNP_V1 array could be used for almond diversity analysis, association studies and genetic mapping for breeding purposes, as a large number of breeding genitors were included in the set of accessions resequenced for the detection of SNPs. Our data showed that this array is also effective for peach species and some wild *Amygdalus* species. The PdSNP_V1 array would be suitable for *Prunus* rootstock breeding, notably for linkage mapping of peach × almond progenies. The PdSNP_V1 array is the first developed almond

array and is commercially available from ThermoFischer Scientific. SNP genotyping can be performed on INRAE's Gentyane platform or on other platforms equipped with Axiom genotyping instruments.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/plants12020242/s1>; Table S1: Summary of the SNP numbers per chromosome; Table S2: List of accessions scored 'BB' for the SNP AX-599403227; Table S3: list of the resequenced accessions; Figure S1: Cluster plots of the three additional SNPs.

Author Contributions: Conceptualization, methodology: H.D., S.E.R.-O. and K.G.A.; validation: H.D., E.C. and P.J.M.-G.; resources: H.D., M.J.R.-C., P.J.M.-G., M.W., A.D. and P.A. provided resequenced data; A.S. coordinated the creation of the array for Thermo Fischer; writing: H.D. and E.C.; supervision, H.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the INRAE Plant biology and Breeding Division.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The probe set sequences and positions of all SNPs included in the almond chip from this study have been deposited in the GDR database under ID code tfGDR1059 and the data will be available for download at https://www.rosaceae.org/publication_datasets (accessed on 28 December 2022). The Axiom 60K almond SNP array (PdSNP_V1) is commercially available with Thermo Fischer Scientific, with request to the author before 2024.

Acknowledgments: We thank Charles Poncet, Pauline Izorche and the staff of the Gentyane platform (INRAE Clermont-Ferrand) for the genotyping and Naïma Dlalal (INRAE GAFL) for the DNA extraction of the accession samples. We thank Frederique Bitton for helping with the extraction of the axiom71mer and Véronique Decroocq for the provision of clones of *Prunus feniziana*.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Varshney, R.K.; Nayak, S.N.; May, G.D.; Jackson, S.A. Next-Generation Sequencing Technologies and Their Implications for Crop Genetics and Breeding. *Trends Biotechnol.* **2009**, *27*, 522–530. [CrossRef] [PubMed]
2. Iwata, H.; Minamikawa, M.F.; Kajiyama-Kanegae, H.; Ishimori, M.; Hayashi, T. Genomics-Assisted Breeding in Fruit Trees. *Breed. Sci.* **2016**, *66*, 100–115. [CrossRef] [PubMed]
3. Rasheed, A.; Hao, Y.; Xia, X.; Khan, A.; Xu, Y.; Varshney, R.K.; He, Z. Crop Breeding Chips and Genotyping Platforms: Progress, Challenges, and Perspectives. *Mol. Plant* **2017**, *10*, 1047–1064. [CrossRef]
4. Verde, I.; Bassil, N.; Scalabrin, S.; Gilmore, B.; Lawley, C.; Gasic, K.; Micheletti, D.; Rosyara, U.; Cattonaro, F.; Vendramin, E.; et al. Development and Evaluation of a 9K SNP Array for Peach by Internationally Coordinated SNP Detection and Validation in Breeding Germplasm. *PLoS ONE* **2012**, *7*, e35668. [CrossRef]
5. Gasic, K.; Linge, C.D.S.; Bianco, L.; Troggio, M.; Rossini, L.; Bassi, D.; Aranzana, M.J.; Arus, P.; Verde, I.; Peace, C.; et al. Development and Evaluation of a 9K SNP Addition to the Peach IpSC 9K SNP Array V1. *Hortscience* **2019**, *54*, S188.
6. Vanderzande, S.; Zheng, P.; Cai, L.; Barac, G.; Gasic, K.; Main, D.; Iezzoni, A.; Peace, C. The Cherry 6+9K SNP Array: A Cost-Effective Improvement to the Cherry 6K SNP Array for Genetic Studies. *Sci. Rep.* **2020**, *10*, 7613. [CrossRef]
7. Bianco, L.; Cestaro, A.; Linsmith, G.; Muranty, H.; Denance, C.; Theron, A.; Poncet, C.; Micheletti, D.; Kerschbamer, E.; Di Pierro, E.; et al. Development and Validation of the Axiom (R) Apple480K SNP Genotyping Array. *PLANT J.* **2016**, *86*, 62–74. [CrossRef]
8. Montanari, S.; Bianco, L.; Allen, B.; Martinez-Garcia, P.; Bassil, N.; Postman, J.; Knabel, M.; Kitson, B.; Deng, C.; Chagne, D.; et al. Development of a Highly Efficient Axiom (TM) 70 K SNP Array for Pyrus and Evaluation for High-Density Mapping and Germplasm Characterization. *BMC Genom.* **2019**, *20*, 331. [CrossRef]
9. Li, X.; Singh, J.; Qin, M.; Li, S.; Zhang, X.; Zhang, M.; Khan, A.; Zhang, S.; Wu, J. Development of an Integrated 200K SNP Genotyping Array and Application for Genetic Mapping, Genome Assembly Improvement and Genome Wide Association Studies in Pear (Pyrus). *Plant Biotechnol. J.* **2019**, *17*, 1582–1594. [CrossRef]
10. Marrano, A.; Martinez-Garcia, P.J.; Bianco, L.; Sideli, G.M.; Di Pierro, E.A.; Leslie, C.A.; Stevens, K.A.; Crepeau, M.W.; Troggio, M.; Langley, C.H.; et al. A New Genomic Tool for Walnut (*Juglans Regia* L.): Development and Validation of the High-Density Axiom (TM) J. Regia 700K SNP Genotyping Array. *Plant Biotechnol. J.* **2019**, *17*, 1027–1036. [CrossRef]
11. Di Guardo, M.; Farneti, B.; Khomenko, I.; Modica, G.; Mosca, A.; Distefano, G.; Bianco, L.; Troggio, M.; Sottile, F.; La Malfa, S.; et al. Genetic Characterization of an Almond Germplasm Collection and Volatilome Profiling of Raw and Roasted Kernels. *Hortic. Res.* **2021**, *8*, 27. [CrossRef] [PubMed]

12. Alioto, T.; Alexiou, K.G.; Bardil, A.; Barteri, F.; Castanera, R.; Cruz, F.; Dhingra, A.; Duval, H.; Fernandez i Marti, A.; Frias, L.; et al. Transposons Played a Major Role in the Diversification between the Closely Related Almond and Peach Genomes: Results from the Almond Genome Sequence. *Plant J.* **2020**, *101*, 455–472. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Sanchez-Pérez, R.; Pavan, S.; Mazzeo, R.; Moldovan, C.; Aiese Cigliano, R.; Del Cueto, J.; Ricciardi, F.; Lotti, C.; Ricciardi, L.; Dicenta, F.; et al. Mutation of a BHLH Transcription Factor Allowed Almond Domestication. *Science* **2019**, *364*, 1095–1098. [\[CrossRef\]](#) [\[PubMed\]](#)
14. D’Amico-Willman, K.; Sideli, G.; Allen, B.; Anderson, E.; Gradziel, T.; Fresnedo-Ramirez, J. Identification of Putative Markers of Non-Infectious Bud Failure in Almond [*Prunus Dulcis* (Mill.) DA Webb] Through Genome Wide DNA Methylation Profiling and Gene Expression Analysis in an Almond × Peach Hybrid Population. *Front. Plant Sci.* **2022**, *13*, 804145. [\[CrossRef\]](#)
15. Donoso, J.M.; Picanol, R.; Serra, O.; Howad, W.; Alegre, S.; Arus, P.; Eduardo, I. Exploring Almond Genetic Variability Useful for Peach Improvement: Mapping Major Genes and QTLs in Two Interspecific Almond × Peach Populations. *Mol. Breed.* **2016**, *36*, 16. [\[CrossRef\]](#)
16. Duval, H.; Van Ghelder, C.; Callot, C.; Esmenjaud, D. Characterization of the *RMja* Resistance Gene to Root-Knot Nematodes from the ‘Alnem’ Almond Rootstock. *Acta Hortic.* **2018**, *1219*, 325–330. [\[CrossRef\]](#)
17. Kochba, J.; Spiegelroy, P. Alnem-1, Alnem-88, Alnem-201 Almonds—Nematode-Resistant Rootstock Seed Source. *Hortscience* **1976**, *11*, 270. [\[CrossRef\]](#)
18. Saucet, S.B.; Van Ghelder, C.; Abad, P.; Duval, H.; Esmenjaud, D. Resistance to Root-Knot Nematodes *Meloidogyne* Spp. in Woody Plants. *New Phytol.* **2016**, *211*, 41–56. [\[CrossRef\]](#)
19. Van Ghelder, C.; Esmenjaud, D.; Callot, C.; Dubois, E.; Mazier, M.; Duval, H. Ma Orthologous Genes in *Prunus* Spp. Shed Light on a Noteworthy NBS-LRR Cluster Conferring Differential Resistance to Root-Knot Nematodes. *Front. Plant Sci.* **2018**, *9*, 1269. [\[CrossRef\]](#)
20. Pavan, S.; Delvento, C.; Mazzeo, R.; Ricciardi, F.; Losciale, P.; Gaeta, L.; D’Agostino, N.; Taranto, F.; Sánchez-Pérez, R.; Ricciardi, L.; et al. Almond Diversity and Homozygosity Define Structure, Kinship, Inbreeding, and Linkage Disequilibrium in Cultivated Germplasm, and Reveal Genomic Associations with Nut and Seed Weight. *Hortic. Res.* **2021**, *8*, 15. [\[CrossRef\]](#)
21. Sanchez-Perez, R.; Howad, W.; Garcia-Mas, J.; Arus, P.; Martinez-Gomez, P.; Dicenta, F. Molecular Markers for Kernel Bitterness in Almond. *Tree Genet. Genomes* **2010**, *6*, 237–245. [\[CrossRef\]](#)
22. Yang, S.; Fresnedo-Ramírez, J.; Sun, Q.; Manns, D.C.; Sacks, G.L.; Mansfield, A.K.; Luby, J.J.; Londo, J.P.; Reisch, B.I.; Cadle-Davidson, L.E. Next Generation Mapping of Enological Traits in an F2 Interspecific Grapevine Hybrid Family. *PLoS ONE* **2016**, *11*, e0149560. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Ooijen, J.; Ooijen, J.; van ’t Verlaat, J.W.; Ooijen, J.; Tol, J.; Dale, J.; Buren, J.; Meer, J.M.; Krieken, J.V.; Ooijen, J.; et al. JoinMap®4, Software for the Calculation of Genetic Linkage Maps in Experimental Populations. 2006. Available online: <https://www.kyazma.nl/index.php/JoinMap/> (accessed on 13 July 2013).
24. Donoso, J.M.; Eduardo, I.; Picanol, R.; Batlle, I.; Howad, W.; Aranzana, M.J.; Arus, P. High-Density Mapping Suggests Cytoplasmic Male Sterility with Two Restorer Genes in Almond × Peach Progenies. *Hortic. Res.* **2015**, *2*, 15016. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Goonetilleke, S.N.; March, T.J.; Wirthensohn, M.G.; Arús, P.; Walker, A.R.; Mather, D.E. Genotyping by Sequencing in Almond: SNP Discovery, Linkage Mapping, and Marker Design. *G3 Genes Genomes Genet.* **2018**, *8*, 161–172. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Guajardo, V.; Martínez-García, P.J.; Solís, S.; Calleja-Satrústegui, A.; Saski, C.; Moreno, M.Á. QTLs Identification for Iron Chlorosis in a Segregating Peach–Almond Progeny through Double-Digest Sequence-Based Genotyping (SBG). *Front. Plant Sci.* **2022**, *13*, 872208. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Yu, Y.; Fu, J.; Xu, Y.; Zhang, J.; Ren, F.; Zhao, H.; Tian, S.; Guo, W.; Tu, X.; Zhao, J. Genome Re-Sequencing Reveals the Evolutionary History of Peach Fruit Edibility. *Nat. Commun.* **2018**, *9*, 5404. [\[CrossRef\]](#)
28. Koepke, T.; Schaeffer, S.; Harper, A.; Dicenta, F.; Edwards, M.; Henry, R.J.; Möller, B.L.; Meisel, L.; Oraguzie, N.; Silva, H.; et al. Comparative Genomics Analysis in Prunoideae to Identify Biologically Relevant Polymorphisms. *Plant Biotechnol. J.* **2013**, *11*, 883–893. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Velasco, D.; Hough, J.; Aradhya, M.; Ross-Ibarra, J. Evolutionary Genomics of Peach and Almond Domestication. *G3 Genes Genomes Genet.* **2016**, *6*, 3985–3993. [\[CrossRef\]](#)
30. Pérez de los Cobos, F.; Martínez-García, P.J.; Romero, A.; Miarnau, X.; Eduardo, I.; Howad, W.; Mnejja, M.; Dicenta, F.; Socias i Company, R.; Rubio-Cabetas, M.J.; et al. Pedigree Analysis of 220 Almond Genotypes Reveals Two World Mainstream Breeding Lines Based on Only Three Different Cultivars. *Hortic. Res.* **2021**, *8*, 11. [\[CrossRef\]](#)
31. Krueger, F.; James, F.; Ewels, P.; Afyounian, E.; Schuster-Boeckler, B. FelixKrueger/TrimGalore: V0.6.7. Available online: <https://github.com/FelixKrueger/TrimGalore/tree/0.6.7> (accessed on 23 July 2021).
32. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed on 20 October 2022).
33. Li, H.; Durbin, R. Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *Bioinforma. Oxf. Engl.* **2009**, *25*, 1754–1760. [\[CrossRef\]](#)
34. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. 1000 Genome Project Data Processing Subgroup the Sequence Alignment/Map Format and SAMtools. *Bioinforma. Oxf. Engl.* **2009**, *25*, 2078–2079. [\[CrossRef\]](#) [\[PubMed\]](#)

35. Van der Auwera, G.A.; Carneiro, M.O.; Hartl, C.; Poplin, R.; del Angel, G.; Levy-Moonshine, A.; Jordan, T.; Shakir, K.; Roazen, D.; Thibault, J.; et al. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Curr. Protoc. Bioinforma.* **2013**, *43*, 11.10.1–11.10.33. [[CrossRef](#)]
36. Pavan, S.; Delvento, C.; Ricciardi, L.; Lotti, C.; Ciani, E.; D'Agostino, N. Recommendations for Choosing the Genotyping Method and Best Practices for Quality Control in Crop Genome-Wide Association Studies. *Front. Genet.* **2020**, *11*, 447. [[CrossRef](#)]
37. Dicenta, F.; Cremades, T.; Jose Martinez-Garcia, P.; Martinez-Gomez, P.; Ortega, E.; Rubio, M.; Sanchez-Perez, R.; Lopez-Alcolea, J.; Egea, J. Penta and Makako: Two Extra-Late Flowering Self-Compatible Almond Cultivars from CEBAS-CSIC. *Hortscience* **2018**, *53*, 1700–1702. [[CrossRef](#)]
38. Antanaviciute, L.; Harrison, N.; Battey, N.H.; Harrison, R.J. An Inexpensive and Rapid Genomic DNA Extraction Protocol for Rosaceous Species. *J. Hortic. Sci. Biotechnol.* **2015**, *90*, 427–432. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.