agriculture

MDPI

Article

A Japanese Plum Breeding Core Collection Capturing and Exploiting Genetic Variation

María Osorio [1], Sebastián Ahumada [2], Rodrigo Infante [2], Igor Pacheco [1], Arnau Fiol [1,*] and Paulina Ballesta [3,*]

[1] BVISA—Nutribreeding, Instituto de Nutrición y Tecnología de Alimentos Dr. Fernando Monckeberg Barros (INTA), Universidad de Chile, El Líbano #5524, Macul 7830490, Chile; maria.osorio@inta.uchile.cl (M.O.); igor.pacheco@inta.uchile.cl (I.P.)

[2] Departamento de Producción Agrícola, Facultad de Ciencias Agronómicas, Universidad de Chile, Santa Rosa #11315, La Pintana 8820808, Chile; sebastian.ahumada.h@ug.uchile.cl (S.A.); rinfante@uchile.cl (R.I.)

[3] Núcleo de Investigación en Sustentabilidad Agroambiental (NISUA), Facultad de Medicina Veterinaria y Agronomía, Universidad de las Américas, Manuel Montt #948, Providencia 7500975, Chile

* Correspondence: arnau.fiol@inta.uchile.cl (A.F.); pballesta@udla.cl (P.B.)

Abstract

The optimal exploitation of genetic variability is essential for the success of breeding programs and for identifying quantitative trait *loci* (QTLs) in genetic association studies. These benefit from populations with a high number of individuals; however, they are expensive since extensive plant maintenance, characterization, and evaluation are required. Core collections offer a practical solution by reducing the number of individuals while representing the original diversity of the population. This study aimed to construct a core collection for Japanese plum to serve as pre-breeding material and enable genetic association studies for traits that are difficult to evaluate. Starting from a population of 1062 individuals genotyped by sequencing, genetic distance and allele coverage metrics were applied to construct several core collections. Genetic parameters and phenotype distribution comparisons allowed for the selection of a core collection of 108 individuals that maximized genetic variability while representative of the original population, confirmed by linkage disequilibrium and population structure analyses. Its usefulness was validated by successfully mapping flowering and maturity dates through marker–trait association. The core collection constructed here will help in the study of fruit quality traits and biotic and abiotic responses, ultimately generating molecular markers to assist the crop's molecular breeding.

Keywords: *Prunus salicina*; genetic diversity; population structure; marker–trait association; SNP

Academic Editor: Zhike Zhang

Received: 20 May 2025
Revised: 13 June 2025
Accepted: 18 June 2025
Published: 26 June 2025

Citation: Osorio, M.; Ahumada, S.; Infante, R.; Pacheco, I.; Fiol, A.; Ballesta, P. A Japanese Plum Breeding Core Collection Capturing and Exploiting Genetic Variation. *Agriculture* **2025**, *15*, 1369. https://doi.org/10.3390/agriculture15131369

1. Introduction

Japanese plum (also Chinese plum or Asian plum; *Prunus salicina Lind*. and hybrids) is a deciduous tree from the *Rosaceae* family that is highly cultivated to consume its fruits. With a total of 2.6 million hectares of harvested area, plums are the most cultivated stone fruit trees within the *Prunus* genus. However, in total production, they rank second (12.4 million tons), surpassed by peaches [1]. Japanese plums are primarily consumed fresh and possess good nutritional and organoleptic properties, with an average sugar content of 15%, 1% of dietary fiber, and 255 kJ 100 g^{-1} of energy value, and also being particularly rich in

phenolic compounds [2,3]. Moreover, plums have health-promoting effects mainly due to their high antioxidant activity, which is linked to anti-inflammatory, anticarcinogenic, and antiadipogenic impact, as well as to the prevention of cardiovascular and neurological diseases [4,5].

The commercial cultivation of Japanese plum is tightly linked to the interspecific crosses developed by Luther Burbank, who between the late 19th and early 20th centuries bred the Asian *P. salicina* with other diploid Asian (*P. simonii*), Eurasian (*P. cerasifera*), and American (*P. americana* and others) plums seeking to improve plum fruit size, flavor, aroma, keeping ability, flesh firmness, acidity, and adaptability [6,7]. The cultivars resulting from these crosses and their descendants formed the basis of the global plum shipping industry and set up the founding genetic material for most modern plum breeding programs [8,9]. These programs now seek to generate elite cultivars that meet expectations with consumers' growing demands, all while being competitive in the market and adapting to the actual environmental challenges. Among the objectives of these programs is the development of high-quality, large fruits that are visually appealing, with optimal flavor based on a balanced sweet-to-sour ratio, and with a juicy but firm flesh. Additionally, some programs aim for an excellent post-harvest shelf life to support long-distance shipment, the increased health benefits of the fruits, and an extended harvest season [9]. Some specific objectives involve a further hybridization of the crop, such as introducing apricot as a donor of PPV (plum pox virus) resistance and self-compatibility traits [10]. Most varieties are self-incompatible and thus require a pollinator with compatible alleles to generate progeny [11], which, added to the breeding history, led to reaching high heterozygosity levels compared to other *Prunus* crops like peach or apricot [12].

The genetic complexity of the crop has resulted in a wide phenotypic diversity, hindering its breeding by reducing the chances of pyramiding a high number of advantageous alleles in the individuals generated from selected crosses [13]. In this context, genomics-assisted breeding is a key tool to efficiently generate the cultivars of the future [14]. Some efforts have been made to detect quantitative trait *loci* (QTLs) on the genome of Japanese plum aiming at assisting its breeding, mainly focused on the trait segregation in the bi-parental <'Angeleno' × 98–99> progeny [15–18]. More recently, genomic prediction has been proven effective for the selection of fruit weight, soluble solid content, acidity, and harvest date in a multi-parent Japanese plum collection, proving that genomics at the service of breeding will increase the odds of obtaining new high-quality plum cultivars [19]. However, the phenotypic evaluation of advanced traits (e.g., bioactive metabolite content, post-harvest behavior, disease resistance, and abiotic stress tolerance) becomes increasingly challenging for large populations because these are difficult to estimate, time-consuming, and/or cost-limited. In such cases where phenotyping generates a bottleneck, the feasibility of identifying the QTLs that regulate trait variability is significantly reduced [20].

Core collections are subsets of individuals from a whole collection, selected by diverse criteria but with the common objective of minimizing sample redundancy and represent the genetic diversity of the original collection with minimal resources [21], which is helpful to deal with genetic studies of advanced traits. The criteria (e.g., genotype, phenotype, origin), as well as how the original collection is represented (e.g., maintaining or maximizing original diversity), are selected according to the ultimate goal of each core [22]. Many plant core collections have been established for a broad range of species over the years, either for germplasm conservation [23,24], pre-breeding selection [25–27], genotype–phenotype association [28,29], or more general purposes by combining the former objectives [30–32]. While no core collection has been established yet for Japanese plum, several have been constituted for other *Prunus* crops to assist their breeding. For example, the core collections of peach, apricot, and sweet cherry served to map traits related to pest resistance and fruit

quality traits such as cracking susceptibility, shape, and firmness, through genome-wide association [33–35], and also proved helpful in validating molecular markers previously associated with flowering date and flower type [36–38].

In this study, we aimed to construct a Japanese plum core collection suitable for forward genetic studies. We evaluated the structure and genetic diversity of the population from the Stone Breeding Program (University of Chile) genotyped in Ballesta et al. (in press) [19], used genotypic data as selection criteria to construct several theoretical core collections, and we selected one maximizing genetic distance between individuals that increased the observed heterozygosity. We demonstrated the usefulness of the selected collection by mapping flowering and harvesting dates through association analysis, two traits critical to improve the crop's adaptability to different climates and extend its harvest season [39]. The core collection constructed in this study will enable further association studies for traits that are complex to phenotype and serve as pre-breeding material for the breeding program.

2. Materials and Methods

2.1. Plant Material and DNA Isolation

The plant material consisted of 31 parental lines, which combined cultivars from the United States, South Africa, and Chile, together with advanced breeding lines from the Stone Breeding Program of the University of Chile (Table S1). These were used to form different cross designs. In the last seven years, at least 44 crosses have been generated from the initial parental lines, conforming half- and full-sib families. In September and October 2022, leaf buds were collected from 1062 trees (parentals and progenies), which were used to extract their DNA according to the CTAB method [40] with modifications from Healy et al. [41].

2.2. Sequencing and Genotyping

Genotyping of the 1062 plum individuals was performed following the ddRADseq protocol described in Peterson et al. (2012) [42], using *NspI* and *MboI* restriction enzyme combination. IGA Technology Services (Udine, Italy) performed library preparation and sequencing using a NovaSeq 6000 instrument (Illumina, San Diego, CA, USA), generating 150 bp paired-end reads. Stacks v2.61 [43] was run to demultiplex the reads, and FastQC v.11.9 to assess their quality. All 1062 samples passed the mean and per-position sequence quality thresholds, averaging 5.1 million high-quality reads per sample. These were aligned to the reference genome of *P. salicina* cv. 'Sanyueli' v.2.0 [44] using BWA-MEM2 v.2.2. 1 software [45]. The SNP calling was performed with Stacks v2.64 with the Maruki-high model on correctly paired-end reads with a minimum mapping quality of 20, and SNPs with a genotype quality below 20 were filtered out. A stringent filter was then applied with VCFtools [46], keeping only biallelic SNPs with a minor allele frequency (MAF) above 10%, with a missing data rate of less than 3% and in Hardy–Weinberg equilibrium (>0.05). To avoid SNP redundancy, the resulting 3584 SNPs were pruned based on linkage disequilibrium using Plink v1.90 [47], removing one SNP for each pair of SNPs within a 10 kb window when their value exceeded $r^2 > 0.5$. Sequencing, mapping, and genotyping details for each sample are available in Supplementary Table S2.

2.3. Phenotypic Evaluation

The above-described population was phenotyped in two consecutive seasons (2022–2023 and 2023–2024) for the following traits: flowering date (FWD), fruit maturity date (MD), and fruit development period (FDP). All values were recorded in Julian days (JD). The FWD was recorded when at least three floral buds on different branches

were in the full bloom stage, and 80% of the remaining buds had reached the balloon stage [11]. The MD was determined when at least 80% of the fruits in the canopy had a pressure between 18–22 N, measured using a manual fruit penetrometer FT011 (Turoni, Forlì, Italy). The FDP was calculated as the difference between the MD and FWD. The phenotypic distributions of the original population and each core collection constructed (detailed below) were compared with the LaplaceDemon package R library [48].

2.4. Construction of Core Collections

Before building a core collection, trees that did not produce mature fruits in at least one season were removed from the dataset, leaving 720 trees for the analysis. Two software programs were used to construct theoretical core collections: GenoCore [49] and Core Hunter 3 [50]. For Genocore, two core collections were constructed to represent 99% and 100% of the allele coverage, CV99 and CV100, respectively. For CoreHunter, two distinct approaches were used based on genetic distances. The Entry-to-Nearest-Entry (EN) distance is designed to maximize the average distance between the selected individual and others selected within the core. Conversely, the Accession-to-Nearest-Entry (AN) distance minimizes the average distance between each individual from the dataset and the closest selected ones in the core. The genetic distances were computed as Rogers' distance (default parameter). The sizes of the core collections were set as 15 and 20% of the filtered data, resulting in four core collections built with CoreHunter: ANMR15, ANMR20, ENMR15, and ENMR20.

2.5. Genetic Diversity, Population Structure, and Linkage Disequilibrium

The genetic diversity parameters, including the number of effective alleles, Shannon's information index, observed heterozygosity and expected heterozygosity, were calculated using the software GenAlex 6.5 for the original population and all constructed core collections [51].

Population structure analysis was conducted for the original population without prior population assumptions using Structure v.2.3.4 [52]. The study employed an admixture ancestry model with independent allele frequencies between populations, setting a burn-in period and a total Markov Chain Monte Carlo of 10,000 and 100,000 iterations, respectively. The optimal number of genetic clusters was selected according to Evanno's method [53] by testing K values from 1 to 10 with the software Structure Harvester v0.7 [54]. Additionally, a genetic principal component analysis (PCA) was carried out in Tassel 5, and the results were plotted using ggplot2 v3.5.2.

Tassel 5 [55] was used to calculate the linkage disequilibrium (LD) between all pairs of SNPs (full matrix) and estimate the LD decay pattern. The default parameters were employed, and a critical value was calculated following Breshegello and Sorrells (2006) [56]. The output of linkage disequilibrium from Tassel 5 was utilized to calculate a critical value. This was achieved by filtering all missing values (NA) and non-significant values (with a >0.01 threshold). The 95th percentile was then calculated as the critical value.

The LD decay curves were adjusted using a non-linear least squares (NLS) regression in R 4.4.1 [57] (Data S1).

2.6. Validation of the Selected Core Collection Through Association Analysis

The selection criteria of the definitive core collection were based on genetic diversity by prioritizing the maximization of observed heterozygosity, and ensuring adequate representation of the three phenotypic traits evaluated and the genetic clusters of the original population.

An association analysis for the three studied traits was performed to assess the feasibility of the core collection for future breeding selection strategies. The Bayesian information and Linkage-Disequilibrium Iteratively Nested Keyway (BLINK; [58]) model was used

for this purpose, which was implemented in GAPIT 3.4 software [59] from the R environment. The association analyses were conducted for both the core collection and the original population, where the results were compared to assess the concordance of marker–trait associations (MTAs) between both populations. Two *p*-values were established to determine whether the MTAs are considered statistically significant: (i) a Bonferroni correction ($p < 0.05$) based on the number of independent tests according to the total of number SNPs ($n = 2342$; total SNPs; Bonferroni correction: $0.05/2342 = p$ adjusted $< 2.13 \times 10^{-5}$); and (ii) a suggestive *p*-value < 0.00625 based on the assumption that chromosomes can be considered as independent testing units ($n = 8$; total chromosomes; 8 independent association tests; Bonferroni correction: 0.05/8 = suggestive *p* adjusted < 0.00625).

3. Results

3.1. Genome-Wide Distribution of SNPs and Genetic Diversity

3.1.1. SNPs Distribution

The final SNP panel comprised 2342 markers, which were distributed on the eight chromosomes of *P. salicina* (Figure 1) with an average of 292.75 SNPs per chromosome, and an average minimum and maximum distances between SNPs of 4.375 bp and 33.06 Mbp, respectively (Table 1). The highest SNP count was observed on chromosome 1 (Chr1; $n = 400$ SNPs), which accounted for approximately 17% of the total SNPs, while the lowest marker count was located on chromosome 4 (Chr4; 9.2%; $n = 216$ SNPs).

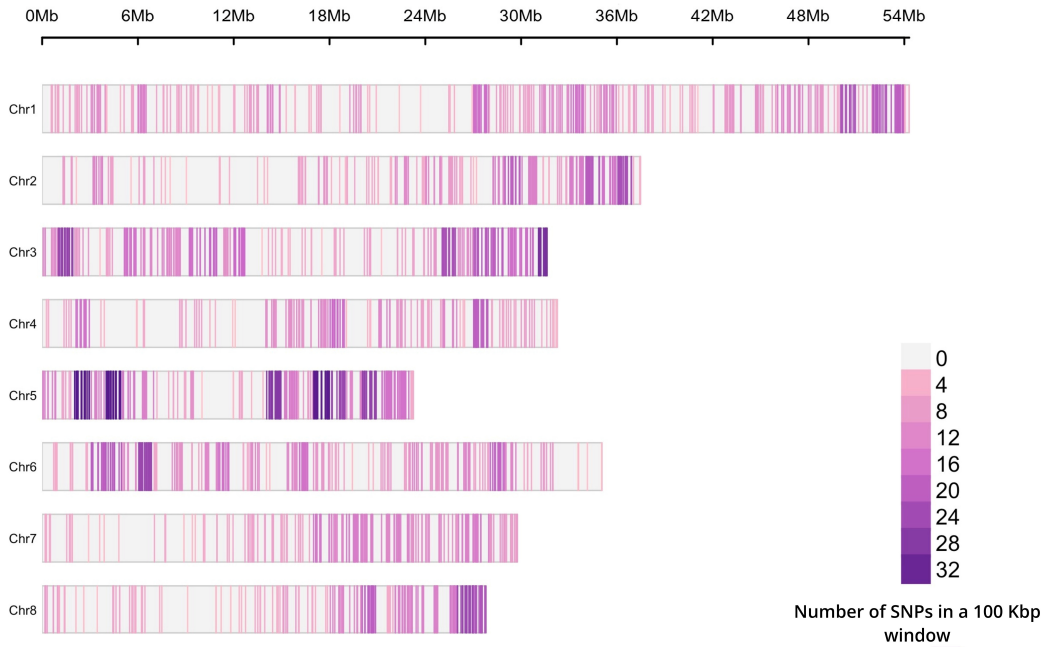

Figure 1. Density plot of the 2342 single nucleotide polymorphisms (SNPs) identified in the 8 chromosomes (represented by the horizontal bars, in Mega-bases) of Japanese plum. The SNPs were obtained using double digest restriction-site associated DNA sequencing (ddRAD-seq). The color intensity indicates the number of SNPs in a 100 Kb window.

Table 1. Summary of SNP markers and linkage disequilibrium distribution across the eight chromosomes of Japanese plum.

Chromosome	Chr. Size (Mbp)	N° Markers	Min Distance (bp)	Max Distance (Mbp)	r^2 min	r^2	r^2 max
Chr1	54.5	400	2	53.60	2.50×10^{-35}	0.08	1
Chr2	37.6	261	1	36.16	5.96×10^{-9}	0.12	1

Table 1. *Cont.*

Chromosome	Chr. Size (Mbp)	N° Markers	Min Distance (bp)	Max Distance (Mbp)	r^2 min	r^2	r^2 max
Chr3	31.8	361	1	31.57	9.07×10^{-9}	0.09	1
Chr4	32.3	216	19	32.00	1.02×10^{-34}	0.10	1
Chr5	23.6	347	1	23.23	1.52×10^{-7}	0.13	1
Chr6	36.3	319	3	34.34	2.4×10^{-8}	0.23	1
Chr7	29.9	220	4	29.58	4.71×10^{-9}	0.12	1
Chr8	28.1	218	4	27.62	1.00×10^{-34}	0.16	1

Chr. Size (Mbp): chromosome size; N° Markers: marker count; Min Distance (bp): minimum distance between markers; Max Distance (Mbp): maximum distance between markers; r^2, r^2 min and r^2 max: average, minimum, and maximum of linkage disequilibrium values per chromosome, respectively.

3.1.2. Genetic Diversity

The genetic diversity parameters of the original collection and the constructed core collections are summarized in Table 2. The two core collections constructed with criteria relying on allele coverage (CV99 and CV100) retain 99% and 100% of the alleles from the original population, selecting a minimum set of 37 and 60 individuals, respectively. The four remaining core collections (ENMR15, ENMR20, ANMR15, and ANMR20) had a higher number (108 and 144 individuals, corresponding to 15 and 20% of the original population, respectively), defined before the analysis and the specific individuals then selected by their genetic distances. The number of effective alleles and Shannon's information index varied slightly between the core collections, ranging from 1.549 to 1.594 and 0.511 to 0.536, respectively. All the core collections except ANMR15 and ANMR20 slightly increased these values compared to the original collection (Ne = 1.564; I = 0.52), with the CV cores displaying the highest values. The observed heterozygosity (Ho) of the whole collection was 0.346. The two CV core collections had a lower value (0.324–0.330) and a greater (0.352–0.356) expected heterozygosity (He). On the contrary, the ENMR and ANMR collections had higher Ho values (0.348–0.369), which were consistently higher than their He (0.335–0.348). Among them, ENMR15 exhibited the highest Ho and the most considerable difference between Ho and He.

Table 2. Genetic diversity parameters for the original population and the six theoretical core collections constructed in this work.

Population	Individuals	Data Type	N	Ne	I	Ho	He
Original	1062	Mean	1043	1.564	0.520	0.346	0.343
		SE	0.114	0.005	0.003	0.002	0.002
CV100	60	Mean	59.031	1.580	0.531	0.330	0.352
		SE	0.021	0.005	0.002	0.002	0.002
CV99	37	Mean	36.539	1.594	0.536	0.324	0.356
		SE	0.014	0.005	0.002	0.002	0.002
ENMR15	108	Mean	106.756	1.575	0.526	0.369	0.348
		SE	0.025	0.005	0.003	0.003	0.002
ENMR20	144	Mean	142.289	1.574	0.525	0.365	0.347
		SE	0.030	0.005	0.003	0.002	0.002
ANMR15	108	Mean	106.224	1.549	0.511	0.348	0.335
		SE	0.029	0.006	0.003	0.003	0.002
ANMR20	144	Mean	142.133	1.551	0.514	0.350	0.337
		SE	0.030	0.005	0.003	0.002	0.002

SE: standard error; N: individuals genotyped per locus; Ne: number of effective alleles; I: Shannon's information index; Ho: observed heterozygosity; He: expected heterozygosity.

3.2. Phenotype Distribution

The phenotype distribution on the original population and the six constructed core collections is shown in Figure 2. In the original population, the FWD exhibited a trimodal distribution in both seasons, while the MD and FDP followed a bimodal distribution in 2022–2023 and monomodal in the 2023–2024 season. Overall, the two CV collections did not follow the distribution pattern of the complete set and exhibited a monomodal distribution for the MD and FDP evaluated in 2022–2023, and the FWD in 2023–2024. The ANMR collections exhibited a similar distribution trend from the original collection, except for the FWD (2023–2024) and ANMR15 on the FDP (2022–2023), which displayed bimodal and monomodal distributions, respectively. The ENMR collections also tend to follow the distribution of the original population, except for the MD in both seasons and ENMR15 for the FDP (2022–2023), which were trimodally distributed, and ENMR20 for the MD (2023–2024), showing a bimodal distribution.

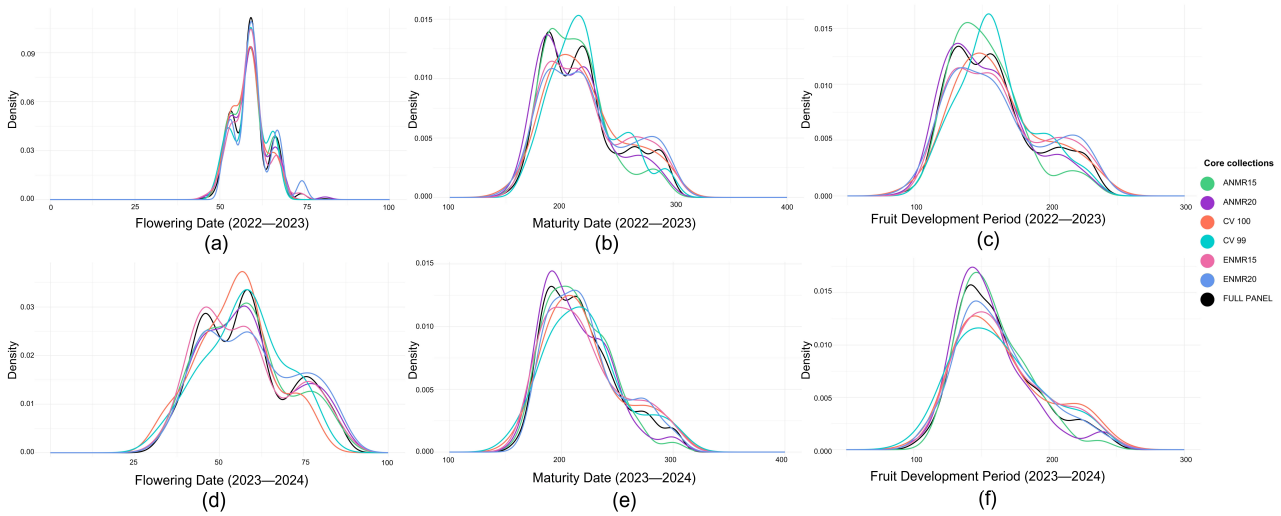

Figure 2. Distribution plot of all phenotypic traits. The *x*-axis of the graph represents the Julian days for each trait measured for the individuals in the full panel and the core collections (CV99, CV100, ANRM15, ANRM20, ENMR15, and ENMR20), represented by different colors, while the *y*-axis shows the relative density of the samples. The FWD, MD, and FDP were evaluated on the original population and compared with the constructed core collections. On the top, season 2022–2023; bottom, season 2023–2024. (**a**,**d**) Flowering date (FWD); (**b**,**e**) maturity date (MD); (**c**,**f**) fruit development period (FDP).

The 10th and 90th percentiles of all traits over the seasons were assessed. The average for the 10th and 90th percentile of the FWD were 52 and 66 JD, respectively, in the first season, and 43 and 78 JD for the subsequent season, respectively. In the case of the MD, the average for the 10th percentile was 179 for both seasons. The mean value for the 90th percentile was 265 in the first and 255 JD in the second season. The 10th and 90th percentiles of the FDP in the 2022–2023 season were of 120 and 204 JD, respectively; in the 2023–2024 season, these values increased to 128 and 197, respectively.

3.3. Population Structure

Considering the genetic diversity and phenotypic distribution results, the ENMR15 core collection was chosen over the others because of its higher genetic diversity and optimal representation of the phenotypic distribution of the original population. Further analyses were conducted on this specific collection to assess the correct representation of the original population structure and validate its usefulness for genetic association analyses.

According to Evanno's method, the most probable number of genetic clusters in the original population was two. The results indicated that the two clusters are not entirely isolated, with only 34.2% of the individuals being strongly assigned (Q > 0.9) to one of the groups, indicating a high level of admixture (Figure 3a). The first two components of the PCA explained 14.3 and 5.8% of the genetic variability, respectively, with the first being sufficient to represent the membership probability of individuals in their assigned group (Figure 3b). The population structure was represented in a PCA, where members of the same genetic group are genetically closer to each other than to members of families generated from different progenitors (Figure 3c). For the ENMR15 core collection, the first two principal components of the genetic PCA explained 9.4 and 5.6%, respectively, of the variability (Figure 3d), and their members represented the two clusters defined for the original population.

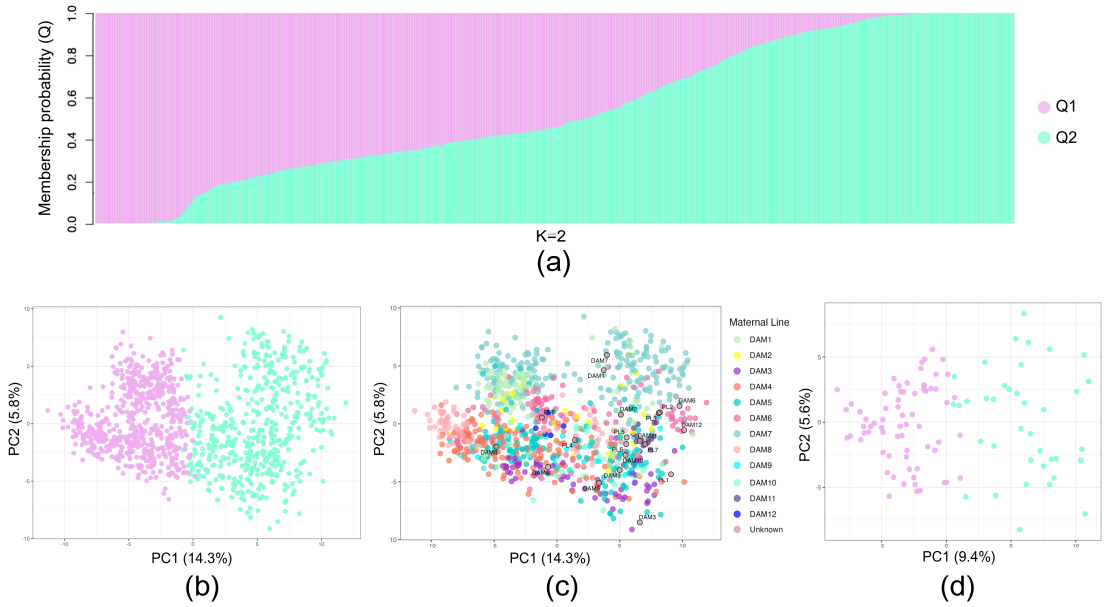

Figure 3. Population structure and principal component analysis (PCA) results based on 2342 high-quality SNPs. (**a**) The inferred population structure of the 1062 individuals of the original population for two genetic clusters (K = 2). The *x*-axis displays the individuals (1062) arranged according to their membership probability (*y*-axis; Q), with Q1 (light purple) and Q2 (aquamarine) representing the two distinct clusters. (**b**) The genetic PCA for all individuals (1062), distributed in the coordinates generated on the first two components. Each individual is assigned to Q1 or Q2, in concordance with the previous plot. (**c**) The former PCA, highlighting the parentals (PL: paternal line) used in the breeding program to generate 44 families, with their descendants colored by their maternal origin (DAM). (**d**) Genetic PCA of the 108 individuals comprising the ENMR15 core collection, colored by their cluster assignment (as illustrated in plot a).

3.4. Patterns of Linkage Disequilibrium

The original population showed an average LD value of $r^2 = 0.128$ for all chromosomes. Linkage disequilibrium values for all chromosomes ranged from values close to zero (r^2 min = 2.5×10^{-35}; chromosome 1) to complete disequilibrium (r^2 max = 1; for all chromosomes). Notably, the average LD value on chromosome 6 was higher than any other chromosome ($r^2 = 0.228$), while the average LD values on chromosomes 1 and 3 were the lowest ($r^2 = 0.0839$ and 0.0869, respectively). The LD decay pattern across the eight chromosomes for the original population is shown in Figure 4. The genome-wide LD decayed rapidly within 3103.66 Kbp, considering a critical value of $r^2 = 0.12$ ($p < 0.01$). For the ENMR15 core collection, the LD rapidly decayed within 1177.25 Kbp, with an estimated critical value of $r^2 = 0.427$ (Figure 4).

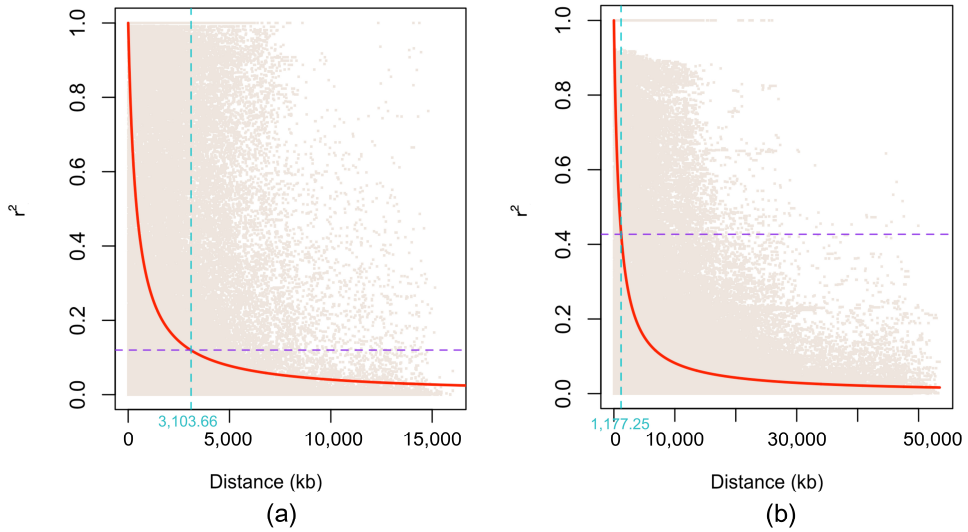

Figure 4. Linkage disequilibrium measures (r^2) against physical distance between pairs of SNP markers in kilobases. The purple dotted line denotes the critical r^2 value, while the turquoise line indicates the physical distance the critical value reaches. (**a**) The LD decay pattern for the original population and (**b**) the pattern for the ENMR15 core collection.

3.5. Association Analysis for Core Collection Validation

The association analysis for the original population revealed 69 marker–trait associations ($p < 2.13 \times 10^{-5}$) for the three traits evaluated in the two consecutive seasons (Table 3). In addition, 300 marker–trait associations were detected considering a suggestive *p*-value of 0.00625 (Supplementary Table S3, Figure S1). A lower number of marker–trait associations were detected for the core collection, of which 59 and 12 were significant using *p*-values of 0.00625 and 2.5×10^{-5}, respectively.

Table 3. Summary of the most significant marker–trait associations ($p < 2.13 \times 10^{-5}$) in the entire population and core collection of Japanese plum based on field performance in two consecutive seasons.

Population	Trait	Season	SNP	Chr	Pos (Mbp)	PVE (%)
Original	FWD	2022–2023	4_14579882	4	14.58	21.43
			4_30218694	4	30.22	14.67
Original	FWD	2023–2024	2_17677588	2	17.68	4.73
			2_17810972	2	17.81	2.51
			2_29360599	2	29.36	12.49
			3_29679394	3	29.68	2.23
			4_29671885	4	29.67	5.18
			4_30832109	4	30.83	4.30
			6_28964130	6	28.96	17.63
Original	MD	2022–2023	1_6041549	1	6.04	2.57
			1_8537798	1	8.54	3.90
			1_10657148	1	10.66	2.08
			1_27497096	1	27.50	2.81
			2_34062566	2	34.06	5.78
			3_5568519	3	5.57	0.86
			3_15529224	3	15.53	1.25
			4_2123398	4	2.12	5.79
			4_2749012	4	2.75	0.51
			4_9009431	4	9.01	5.34
			4_16271618	4	16.27	5.80
			4_17776021	4	17.78	5.42
			5_14843112	5	14.84	1.68
			6_4538253	6	4.54	3.36
			6_23796481	6	23.80	1.24
			7_29346449	7	29.35	29.04
			8_2117948	8	2.12	1.99
			8_20238405	8	20.24	2.17
			8_25944299	8	25.94	1.52

Table 3. *Cont.*

Population	Trait	Season	SNP	Chr	Pos (Mbp)	PVE (%)
Original	MD	2023–2024	1_7582789	1	7.58	2.74
			1_12071117	1	12.07	1.08
			1_39963944	1	39.96	0.25
			2_34062566	2	34.06	2.37
			3_15415536	3	15.42	1.73
			3_20574499	3	20.57	1.82
			4_2123398	4	2.12	1.47
			4_2749012	4	2.75	1.65
			4_9009431	4	9.01	10.50
			4_14086328	4	14.09	11.05
			4_14379875	4	14.38	1.52
			4_15557481	4	15.56	10.86
			4_16271618	4	16.27	4.67
			4_16847730	4	16.85	4.07
			4_17776021	4	17.78	14.03
			4_22490748	4	22.49	2.88
			5_9438804	5	9.44	2.49
			6_11673835	6	11.67	2.33
			6_31840696	6	31.84	2.34
			7_27485481	7	27.49	4.97
			8_19266695	8	19.27	1.84
Core	MD	2022–2023	4_9990227	4	9.99	40.56
			4_18831526	4	18.83	13.97
			4_22490748	4	22.49	13.37
Core	MD	2023–2024	4_16271618	4	16.27	27.93
			4_25103389	4	25.10	41.27
Original	FDP	2022–2023	1_6041549	1	6.04	7.80
			1_12071117	1	12.07	5.95
			2_34062566	2	34.06	8.75
			4_2423446	4	2.42	2.53
			4_8617235	4	8.62	7.88
			4_9009431	4	9.01	6.84
			4_14011592	4	14.01	1.42
			4_16271618	4	16.27	4.82
			4_17776021	4	17.78	17.84
			4_18293367	4	18.29	8.35
			5_14956833	5	14.96	1.23
			6_4538253	6	4.54	7.32
Original	FDP	2023–2024	4_2123398	4	2.12	2.62
			4_9009431	4	9.01	9.27
			4_16271618	4	16.27	2.31
			4_17455555	4	17.46	17.63
			4_17776021	4	17.78	26.58
			4_22490748	4	22.49	7.53
			6_13073537	6	13.07	5.46
			7_21679670	7	21.68	2.75
Core	FDP	2022–2023	2_34062566	2	34.06	17.95
			4_18831526	4	18.83	18.19
			4_22490748	4	22.49	15.71
			5_2130737	5	2.13	9.90
Core	FDP	2023–2024	2_29453403	2	29.45	14.78
			4_9990227	4	9.99	40.48
			4_18831526	4	18.83	13.70

Chr: chromosome; Pos (Mbp): physical position expressed in Mbp; PVE: phenotypic variation explained.

According to the association study considering the original population, two and seven SNP markers were significantly associated with the FWD evaluated in the 2022–2023 and 2023–2024 seasons ($p < 2.13 \times 10^{-5}$), respectively, explaining between 2.23 and 21.43% of the total phenotypic variation. On the other hand, 92 marker–trait associations (Table S3) were detected considering a suggestive *p*-value of 0.00625 for the two seasons. In this context, four markers on chromosomes 1 (SNP 1_50892252), 2 (SNP 2_4372972), 3 (SNP 3_9303984), and 8 (SNP 8_7359085) were significantly associated with the FWD evaluated for the 2022–2023 and 2023–2024 seasons in the original population. No marker–trait associations

were found for the FWD in the core collection, considering a p value of 2.13×10^{-5}, although 27 markers were significantly associated with the FWD, considering a suggestive p value of 0.00625. Of these 27 associations, four markers (SNPs 2_27203744, 2_29437650, 2_29453403, and 2_29487535) located on chromosome 2 were significantly associated with the FWD in both 2022–2023 and 2023–2024. Most of the common markers associated with the FWD for the original population and the core collection were only detected for the 2022–2023 season (three markers located on chromosomes 2, 6, and 7).

Fifty-two and sixty-three marker–trait associations were detected for the MD assessed in 2022–2023 and 2023–2024 (original population), respectively, considering a suggestive p-value of 0.00625. The markers significantly associated with the MD, considering a p-value of 2.13×10^{-5}, explained up to 29.04% of the phenotypic variation (SNP 7_29346449; 2022–2023 season). Eight markers were associated with the MD in both seasons, located on chromosomes 2 (SNP 2_34062566), 4 (SNPs 4_2123398, 4_2749012, 4_9009431, 4_16271618, 4_17776021, and 4_22490748), and 6 (SNP 6_4538253) for the original population. Only eight marker–trait associations were detected for the core collection, even considering a suggestive p-value. Of these eight SNP markers, seven were located on chromosome 4. Only two markers on chromosome 4 (SNPs 4_22490748 and 4_16271618) were significantly associated with MD for the core collection and the original population in a single or both seasons.

For the FDP evaluated in the 2022–2023 and 2023–2024 seasons, 56 and 35 marker–trait associations were found, respectively, considering a suggestive p-value of 0.00625, while twelve and eight associations were detected, respectively, with a p-value of 2.13×10^{-5} in the original population. Seven markers located on chromosomes 2 (SNPs 2_34062566 and 2_35099722), 3 (SNP 3_20506903), and 4 (SNPs 4_9009431, 4_16271618, and 4_17776021; and one additional SNP 4_2423446 $p < 0.000625$) were statistically associated with the FDP evaluated in both the 2022–2023 and 2023–2024 years, which explained up to 26.58% of the phenotypic variation. When implementing the association analysis for the core collection, only three markers on chromosome 4 (SNP 4_9990227, 4_18831526, and 4_22490748; $p < 0.00625$) were statistically associated with the FDP for the 2022–2023 and 2023–2024 seasons (24 total marker–trait associations). One SNP marker located on chromosome 1 (SNP 1_6041549) was consistently detected in the core collection and the original population for the 2022–2023 season but not for the 2023–2024 season. One marker located on chromosome 2 (SNP 2_34062566) was consistently detected in the original population in both seasons but was only detected in the core collection in the 2022–2023 season. Two markers (SNPs 4_18831526 and 4_22490748) located on chromosome 4 were statistically associated, considering both the core collection and the original population.

4. Discussion

Environmental adaptation and phenotypic plasticity in crops are required for their sustained production over time and are essential to obtain advances in their genetic improvement. In this context, germplasm collections provide genetic resources to improve modern commercial genotypes. However, the lack of large-scale phenotypic evaluations hinders the selection of genotypes and the generation of new varieties of high commercial value [20]. Core collections represent a key strategy to exploit genetic variability at an optimal manageability and overcome this limitation, allowing for the minimization of the resources required to phenotype a population for genotype–phenotype association analyses and the maintenance of efficient pre-breeding collections [21].

High-throughput sequencing technologies can generate a vast number of molecular markers over the genome of the target species. In this study, we used ddRAD sequencing, which allows for the further optimization of the sequencing cost per sample but has some

common drawbacks, such as increased data missingness [60]. We applied stringent filters to the sequenced data, significantly reducing the total marker count. Still, we kept the best quality SNPs well distributed over the eight chromosomes of Japanese plum, suitable for population structure analysis, linkage disequilibrium estimation, and the generation of a core collection with increased genetic diversity to enhance Japanese plum breeding.

4.1. Generation of a Core Collection Maximizing Genetic Diversity

Multiple strategies have been employed to construct core collections since their conceptualization by Frankel (1984) [61]. Initially, sample origin and phenotypic data were the common stratification choices because these offered a quick and cost-effective way of sample differentiation. Later, the widespread and systematic use of molecular markers led to their preference due to their stability, co-dominance, and heritability, which served to discard the influence of ambient factors [21]. Still, nowadays many core collections are constructed by combining DNA markers with phenotypic data to ensure adequate representation of traits of interest [62–64]. Since the ultimate goal of the core collection in this study is to serve as pre-breeding material and enable fruit quality studies, individuals without flowering or poor fruit setting were filtered out, as their inclusion would compromise the collection's utility. Subsequently, the selection criteria relied exclusively on the information provided by DNA markers, aiming at maximizing the genetic diversity to make it suitable for genetic association studies targeting a wide range of heritable traits [33].

The CV collections were constructed with a maximization strategy that favors allelic richness, which in this study resulted in a minimum of 37 and 60 individuals required to retain 99% and 100%, respectively, of the alleles of the original collection, similar to those values obtained in peach and apricot core collections [36,65]. Considering that Japanese plum is a highly diverse species, these low numbers still agree with the original population design because, even though it is relatively large, it consisted of half- and full-sibling families generated with recurrent parentals. While ideal for the manageability of a pre-breeding collection, such a limited sample size might lack the resolution to map QTLs in *Prunus* crops successfully [66]. To increase the sample size to at least one hundred genetically diverse individuals while minimizing genetic redundancy, a distance-based strategy recommended by Odong et al. (2013) [22] was incorporated. The entry-nearest-entry metric included in Core Hunter 3 was designed to maximize the genetic diversity of the whole population while avoiding the inclusion of close individuals [22,50]. Although the former would theoretically be the ideal metric to construct the core collection of our objectives, the accession-nearest-entry metric that favors an evenness representation of all individuals was also incorporated, and the optimal core collection was chosen based on comparative analyses.

Among all tested collections, both ENMR collections had the highest observed heterozygosity, increased expected heterozygosity, mean number of effective alleles per locus, and Shannon's information index relative to the original population. Since only genotypic data were used for selection, three traits (flowering date, maturity date, and fruit development period) were used to assess how the core collections represented the distribution of the original population. The ENMR collections faithfully represent the original set of phenotypic distribution in the considered phenological traits and, in some cases, include more individuals at the extremes of the distributions, as expected for the entry-nearest-entry metric [22]. While this effect was not observed in the FWD trait, it was very notable in the MD and FDP traits at the right end of their phenotypic distributions, which was reflected on their averaged 90th percentile values. These values in the ENMR15 collection were higher compared to the original population: for the MD, the values were 20 and 22 days higher for the first and second season, respectively; for the FDP, these were 15 and 22 days

higher, respectively. This indicates that the core collection, only relying on genotypic data, favored the representation of late maturity trees that were underrepresented in the original population. The ENMR core collection with 108 individuals was ultimately selected over the one with 144 due to slightly higher diversity index values and a better cost-efficiency. The entry-nearest-entry metric has also been the optimal method for the generation of core collections in other Rosaceous crops such as apple and strawberry [67,68].

The original population consisted of two genetic groups with high levels of admixture between them. This low level of structure is observed in multi-parent populations, in which most individuals are closely related because of the recurrent use of the parentals in the breeding program [69]. Carrasco et al. (2012) [70] analyzed 29 Japanese plum elite cultivars and, similar to our results, determined that the population consisted of only two groups with low levels of differentiation. Conversely, in studies in which the population consisted of individuals of very different origins, or that included non *P. salicina* plum species and hybrids from the early breeding of the modern Japanese plum, more than four genetic clusters could be identified [12,71]. The two genetic clusters found here were well represented in the ENMR15 collection, further indicating that it faithfully represents the original population.

The core collection constructed here is a representation of the genetic diversity of the breeding program, whose founders were plum cultivars generated after the interspecific hybridization of the species. Typically, core collections have been constructed including landraces and historical cultivars, as in the sweet cherry and apricot core collections [33,35]. Still, some reference populations, such as the PeachRefPop and the apple REFPOP, have also incorporated several multi-parent progenies to represent the diversity of modern cultivated varieties and facilitate further genetic studies [36,72]. While the core collection constructed here might represent the variability of modern plum cultivars, it may fail to represent the full extension of the species-wide genetic diversity. To achieve this goal, a global effort would be required to construct a plum germplasm collection that incorporates, aside from modern cultivars from diverse origin, landraces from China and the early cultivars from the Burbank breeding program, including those non *P. salicina* species used in the early hybridization of the crop [71].

4.2. Genome-Wide Patterns of Linkage Disequilibrium (LD)

In the context of genomics-assisted breeding, linkage disequilibrium (LD) allows for estimating the marker density needed to achieve adequate resolution in the mapping of quantitative trait *loci* (QTLs). As expected, the LD pattern of the core collection decayed faster than in the original population. The LD decayed within ~1100 Kbp, and a critical threshold of $r^2 = 0.427$ was considered necessary to maintain a significant association between *loci*. In contrast, the LD decay was observed within ~3100 Kbp for the original population, supporting the maintenance of the LD over considerably greater physical distances. *P. salicina* has a relatively small genome of approximately 280 Mbp; therefore, based on the observed LD patterns, it is estimated that a minimum of approximately 100 and 250 SNP markers would be necessary to effectively map QTLs in the original population and the core collection, respectively. These results support the feasibility of using both populations for genetic mapping studies.

The LD pattern observed in the original population and the ENMR15 core collection was consistent with the estimated genetic diversity indices (see above). The core collection had slightly greater genetic diversity (in terms of heterozygosity) than the original population, which is expected given that it was designed to conserve individuals that represent the greatest genetic diversity in the population. Consequently, linkage disequilibrium in the core collection is expected to decay rapidly and spread over shorter distances. According to

other studies in natural populations with higher diversity of *Prunus* [73–75], the LD pattern reported in this study extends over relatively long distances, which is consistent with the diversity and genetic structure parameters. For instance, Thurow et al. (2020) [73] identified three clusters conforming to a Brazilian peach breeding germplasm, and consequently, the LD pattern decayed above 38 Kbp. In contrast to the present study, Carrasco et al. (2022) [76] previously examined the genetic diversity of a Japanese plum breeding program, which was defined by two genetically homogeneous groups, and linkage disequilibrium was spread over relatively short distances. In fact, Carrasco et al. (2022) [76] considered only genetically distinct elite cultivars, which explains the discrepancy with the results of this study.

4.3. Core Collection Validation Through Association Analysis

To validate the future usefulness of the core collection for developing selection strategies for the Japanese plum breeding program, a comparative study for detecting marker–trait associations was implemented in the original population and in the core collection (in this case, ENMR15). In general, a small number of SNPs were significantly associated with the studied traits for both the original population and the core collection. Notably, the associations identified in the original population were consistent with previous studies. Chromosomes 1, 4, 6, and 8 have been extensively described to be associated with phenological traits such as the FWD, MD, and FDP in *Prunus* spp. [16,77–80]. Previously, Salazar et al. 2020 [16] identified three QTLs located on chromosomes 6, 1, and 8 associated with the blooming date, and on chromosomes 4 and 2 associated with the FDP in an F1 population from the cross '98–99' × 'Angeleno' of *P. salicina*. In the present study, marker–trait associations detected in the original population, explaining between 12 and 21% of the phenotypic variation in the FWD, could not be detected in the core collection. On the other hand, considering the LD decay pattern for the core collection (LD decay > ~1177 kb), two *loci* significantly associated with the FWD in the core collection could be linked to the *loci* strongly associated with the FWD in the original population. The SNP markers 4_29203059 (which could be related to 4_30218694, which explained 14% of the PV of the FWD) and 2_29363790 (which could be related to 2_29360599, which explained 12.5% of the PV of the FWD) were located less than 162 kb from the markers detected in the original population. In addition, several SNPs located on chromosomes 6 were significantly associated with the FWD in the core collection, considering a suggestive *p*-value, which is consistent with previous studies examining the genetic control of flowering dates in *Prunus* [16,78]. Although the flowering beginning date is considered highly heritable [16,77,78] inter-season variation is recognized as conditioning accurate phenotyping and the detection of stable QTLs for flowering dates. In fact, according to the Chilean Meteorological Division (https://www.meteochile.gob.cl/ (accessed on 28 April 2025)), the rainfall between the seasons evaluated in this study was highly variable (August–September 2023: 173.6 mm and August–September 2022: 23.6 mm), affecting floral structures and the ability to phenotype the appropriate beginning and end dates of the flowering period.

Association analysis for the MD detected genomic regions located mainly on chromosome 4 that were statistically related to the MD, both for the original population and the core collection, in both seasons. Chromosome 4 has been extensively studied for its relationship with the ripening date of *Prunus* spp. fruits, identifying several stable molecular markers for more than one evaluation season [16,78,81]. Previously, Pirona et al. [82] described that a NAC transcription factor would be encoded in a genomic region of chromosome 4 of the peach, which is postulated as a master candidate gene for ripening. Notably, some SNPs significantly associated with the MD and located on chromosome 4 (detected in the original population) were not statistically related to the MD in the core collection until the

significance threshold was set to a suggestive value (e.g., SNP 4_22490748). Interestingly, in this study, an SNP significantly associated with the MD located on chromosome 7 was detected, which explained 20% of the phenotypic variation in the 2022–2023 season. Despite the magnitude of its effect, this locus was not detected in the 2023–2024 season, which explains why it was not detected in the core collection.

A major QTL located ~17 Mbp on chromosome 4 was detected, explaining up to 26% of the phenotypic variation in the FDP in both seasons. Controversially, this QTL was not detected in the association analysis for the core collection, considering the significance threshold values of $p < 0.00625$. On the other hand, this SNP marker was associated with the FDP in the core collection with a significance of $p < 0.05$ in both seasons, from which it can be assumed that this locus was not detected due to its low representativeness in the core collection, although an SNP located ~18 Mbp on chromosome 4 explained up to 18% of the variation in the FDP in both seasons.

The traditional association studies are characterized by their extensive sample sizes, often encompassing hundreds or thousands of accessions. This approach guarantees achieving success, as it ensures that the trait under investigation possesses the essential qualities of being a continuous variable, accompanied by numerous observations and a normal distribution. Consequently, at a small sample size (n < 100 individuals), only *loci* that account for a substantial proportion of the phenotypic variance will be identified; meanwhile, *loci* exhibiting a minor effect will need the evaluation of a thousand individuals to be detected [43,83].

In this study, it has been demonstrated that the number of individuals influences the probability of discovering marker–trait associations. On the other hand, marker–trait associations adjacent to those detected in the original population were detected, and were consistent with previous studies; therefore, the results demonstrated that the core collection can be used for the detection of new genomic regions to ultimately assist breeding programs.

5. Conclusions

This study is the first construction of a Japanese plum core collection based on a multi-parental original population, maintaining the representation of several elite cultivars, progenies, and selections. This core collection maximizes genetic diversity and is an optimal representation of the original population with only 10% of the total individuals ($n = 108$), providing access to plant resources that represent the total genetic diversity of the breeding program while minimizing the amount of material that needs to be maintained. We validated its usefulness by association analysis, which revealed stable QTLs for flowering date, fruit development period, and maturity date in two consecutive seasons, consistent with previous findings. This core collection will provide valuable genetic resources for future research and breeding in Japanese plum.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agriculture15131369/s1, Table S1: Parental lines from the Japanese plum breeding program; Table S2: Sequencing and genotyping statistics per sample; Table S3: Marker–trait association for original and core collection populations in two consecutive seasons. Figure S1: Circular Manhattan plots of marker–trait associations (MTAs). Data S1: Linkage disequilibrium curve fitting script.

Author Contributions: Conceptualization, P.B., A.F. and I.P.; methodology, M.O.; phenotyping, M.O. and S.A.; formal analysis, M.O.; resources, R.I.; data curation, M.O., A.F. and P.B.; writing—original draft preparation, M.O., P.B. and A.F.; writing—review and editing, I.P.; supervision, P.B. and A.F.; project administration, I.P.; funding acquisition, P.B., A.F. and I.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by ANID, through the following grants: FONDECYT Postdoctorado N°3220494, FONDECYT Postdoctorado N°3240559, FONDECYT Regular N°1231614, FONDECYT Iniciación N°11250577, and IDeA ID22I10266.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: The authors thank the team at the Rinconada de Maipú experimental station and the CALIFRU laboratory for the maintenance of the orchard and help during the harvest seasons.

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

Abbreviations

The following abbreviations are used in this manuscript:

QTLs	Quantitative trait *loci*
PPV	Plum pox virus
MAF	Minor allele frequency
FWD	Flowering date
FDP	Fruit development period
MD	Maturity date
JD	Julian days
AN	Accession-to-nearest-entry
EN	Entry-to-nearest-entry
LD	Linkage disequilibrium
NLS	Non-linear least squares
BLINK	Bayesian information and linkage disequilibrium iteratively nested keyway
MTAs	Marker–trait associations
Bp	Base-pair
Mbp	Mega base-pair
Kbp	Kilo base-pair
SNPs	Single nucleotide polymorphism
ddRADseq	Double digest restriction-site associated DNA sequencing

References

1. FAOSTAT. Food and Agriculture Organization of the United Nations. 2022. Available online: https://www.fao.org/faostat/ (accessed on 15 October 2024).
2. Lozano, M.; Vidal-Aragón, M.C.; Hernández, M.T.; Ayuso, M.C.; Bernalte, M.J.; García, J.; Velardo, B. Physicochemical and Nutritional Properties and Volatile Constituents of Six Japanese Plum (*Prunus salicina* Lindl.). *Cultivars. Eur. Food Res. Technol.* **2009**, *228*, 403–410. [CrossRef]
3. Vincente, A.R.; Manganaris, G.A.; Ortiz, C.M.; Sozzi, G.O.; Crisosto, C.H. Nutritional Quality of Fruits and Vegetables. In *Postharvest Handling*; Elsevier: Amsterdam, The Netherlands, 2014; pp. 69–122.
4. Khoo, H.E.; Azlan, A.; Tang, S.T.; Lim, S.M. Anthocyanidins and Anthocyanins: Colored Pigments as Food, Pharmaceutical Ingredients, and the Potential Health Benefits. *Food Nutr. Res.* **2017**, *61*, 1361779. [CrossRef] [PubMed]
5. Liu, W.; Nan, G.; Nisar, M.F.; Wan, C. Chemical Constituents and Health Benefits of Four Chinese Plum Species. *J. Food Qual.* **2020**, *2020*, 1–17. [CrossRef]
6. Okie, W.R.; Ramming, D.W. Plum Breeding Worldwide. *Horttechnology* **1999**, *9*, 162–176. [CrossRef]
7. Karp, D. Luther Burbank's Plums. *HortScience* **2015**, *50*, 189–194. [CrossRef]
8. Okie, W.R.; Hancock, J.F. Plums. In *Temperate Fruit Crop Breeding*; Springer: Dordrecht, The Netherlands, 2008; pp. 337–358.
9. Sottile, F.; Caltagirone, C.; Giacalone, G.; Peano, C.; Barone, E. Unlocking Plum Genetic Potential: Where Are We At? *Horticulturae* **2022**, *8*, 128. [CrossRef]
10. Nicolás-Almansa, M.; Ruiz, D.; Salazar, J.A.; Guevara, A.; Cos, J.; Martínez-Gómez, P.; Rubio, M. Phenotypic and Molecular Characterization of New Interspecific Japanese Plum × Apricot Hybrids (*Plumcots*). *Sci. Hortic.* **2023**, *318*, 112131. [CrossRef]
11. Guerra, M.E.; Rodrigo, J. Japanese Plum Pollination: A Review. *Sci. Hortic.* **2015**, *197*, 674–686. [CrossRef]

12. Guerrero, B.I.; Guerra, M.E.; Herrera, S.; Irisarri, P.; Pina, A.; Rodrigo, J. Genetic Diversity and Population Structure of Japanese Plum-Type (Hybrids of *P. salicina*) Accessions Assessed by SSR Markers. *Agronomy* **2021**, *11*, 1748. [CrossRef]
13. Topp, B.L.; Russell, D.M.; Neumüller, M.; Dalbó, M.A.; Liu, W. Plum. In *Fruit Breeding*; Springer: Boston, MA, USA, 2012; pp. 571–621.
14. Varshney, R.K.; Bohra, A.; Yu, J.; Graner, A.; Zhang, Q.; Sorrells, M.E. Designing Future Crops: Genomics-Assisted Breeding Comes of Age. *Trends Plant Sci.* **2021**, *26*, 631–649. [CrossRef]
15. Salazar, J.A.; Pacheco, I.; Shinya, P.; Zapata, P.; Silva, C.; Aradhya, M.; Velasco, D.; Ruiz, D.; Martínez-Gómez, P.; Infante, R. Genotyping by Sequencing for SNP-Based Linkage Analysis and Identification of QTLs Linked to Fruit Quality Traits in Japanese Plum (*Prunus salicina* Lindl.). *Front. Plant Sci.* **2017**, *8*, 476. [CrossRef] [PubMed]
16. Salazar, J.A.; Pacheco, I.; Zapata, P.; Shinya, P.; Ruiz, D.; Martínez-Gómez, P.; Infante, R. Identification of Loci Controlling Phenology, Fruit Quality and Post-Harvest Quantitative Parameters in Japanese Plum (*Prunus salicina* Lindl.). *Postharvest Biol. Technol.* **2020**, *169*, 111292. [CrossRef]
17. Valderrama-Soto, D.; Salazar, J.; Sepúlveda-González, A.; Silva-Andrade, C.; Gardana, C.; Morales, H.; Battistoni, B.; Jiménez-Muñoz, P.; González, M.; Peña-Neira, Á.; et al. Detection of Quantitative Trait Loci Controlling the Content of Phenolic Compounds in an Asian Plum (*Prunus salicina* L.) F1 Population. *Front. Plant Sci.* **2021**, *12*, 679059. [CrossRef]
18. Battistoni, B.; Salazar, J.; Vega, W.; Valderrama-Soto, D.; Jiménez-Muñoz, P.; Sepúlveda-González, A.; Ahumada, S.; Cho, I.; Gardana, C.S.; Morales, H.; et al. An Upgraded, Highly Saturated Linkage Map of Japanese Plum (*Prunus salicina* Lindl.), and Identification of a New Major Locus Controlling the Flavan-3-Ol Composition in Fruits. *Front. Plant Sci.* **2022**, *13*, 805744. [CrossRef]
19. Ballesta, P.; Fiol, A.; Ahumada, S.; Osorio, M.; Ibañez, J.; Fresnedo-Ramírez, J.; Mora-Poblete, F.; Infante, R.; Battistoni, B.; Pacheco, I. Genomic Prediction of Phenological and Fruit-Quality Traits in a Multi-Family Japanese Plum Breeding Population. *Hortic Plant J.* **2025**, *in press*.
20. Mir, R.R.; Reynolds, M.; Pinto, F.; Khan, M.A.; Bhat, M.A. High-Throughput Phenotyping for Crop Improvement in the Genomics Era. *Plant Sci.* **2019**, *282*, 60–72. [CrossRef]
21. Gu, R.; Fan, S.; Wei, S.; Li, J.; Zheng, S.; Liu, G. Developments on Core Collections of Plant Genetic Resources: Do We Know Enough? *Forests* **2023**, *14*, 926. [CrossRef]
22. Odong, T.L.; Jansen, J.; van Eeuwijk, F.A.; van Hintum, T.J.L. Quality of Core Collections for Effective Utilisation of Genetic Resources Review, Discussion and Interpretation. *Theor. Appl. Genet.* **2013**, *126*, 289–305. [CrossRef]
23. Oliveira, M.F.; Nelson, R.L.; Geraldi, I.O.; Cruz, C.D.; de Toledo, J.F.F. Establishing a Soybean Germplasm Core Collection. *Field Crops Res.* **2010**, *119*, 277–289. [CrossRef]
24. Wu, J.; Liu, D.; Wariss, H.M.; Zhang, H.; Su, M.; Li, W.; Han, Z. Genetic Diversity and Construction of Core Collection Provides New Insight for the Conservation of Edible *Allium galanthum* in Xinjiang. *Sci. Hortic.* **2025**, *341*, 113961. [CrossRef]
25. Upadhyaya, H.D.; Dwivedi, S.L.; Gowda, C.L.L.; Singh, S. Identification of Diverse Germplasm Lines for Agronomic Traits in a Chickpea (*Cicer arietinum* L.) Core Collection for Use in Crop Improvement. *Field Crops Res.* **2007**, *100*, 320–326. [CrossRef]
26. Mahmoodi, R.; Dadpour, M.R.; Hassani, D.; Zeinalabedini, M.; Vendramin, E.; Micali, S.; Nahandi, F.Z. Development of a Core Collection in Iranian Walnut (*Juglans regia* L.) Germplasm Using the Phenotypic Diversity. *Sci. Hortic.* **2019**, *249*, 439–448. [CrossRef]
27. Yang, W.; Yang, B.; Lu, L.; Zhang, X.; Sun, J.; Wang, L.; Zheng, Z.; Liang, D.; Wang, K.; Yan, X.; et al. Utilizing SSR-Based Core Collection Development to Improve Conservation and Utilization of *Corylus* L. Genetic Resources. *PLoS ONE* **2024**, *19*, e0312116. [CrossRef] [PubMed]
28. Liang, W.; Dondini, L.; De Franceschi, P.; Paris, R.; Sansavini, S.; Tartarini, S. Genetic Diversity, Population Structure and Construction of a Core Collection of Apple Cultivars from Italian Germplasm. *Plant Mol. Biol. Rep.* **2015**, *33*, 458–473. [CrossRef]
29. Guo, Y.; Kuang, L.; Xu, Y.; Yan, T.; Jiang, L.; Dong, J.; Wu, D. Construction of a Worldwide Core Collection of Rapeseed and Association Analysis for Waterlogging Tolerance. *Plant Growth Regul.* **2022**, *98*, 321–328. [CrossRef]
30. Lee, H.-Y.; Ro, N.-Y.; Jeong, H.-J.; Kwon, J.-K.; Jo, J.; Ha, Y.; Jung, A.; Han, J.-W.; Venkatesh, J.; Kang, B.-C. Genetic Diversity and Population Structure Analysis to Construct a Core Collection from a Large *Capsicum germplasm*. *BMC Genet.* **2016**, *17*, 142. [CrossRef]
31. Christov, N.K.; Tsonev, S.; Todorova, V.; Todorovska, E.G. Genetic Diversity and Population Structure Analysis—A Prerequisite for Constructing a Mini Core Collection of Balkan *Capsicum annuum* Germplasm. *Biotechnol. Biotechnol. Equip.* **2021**, *35*, 1010–1023. [CrossRef]
32. Guo, Q.; Liu, J.; Li, J.; Cao, S.; Zhang, Z.; Zhang, J.; Zhang, Y.; Deng, Y.; Niu, D.; Su, L.; et al. Genetic Diversity and Core Collection Extraction of *Robinia pseudoacacia* L. Germplasm Resources Based on Phenotype, Physiology, and Genotyping Markers. *Ind. Crops Prod.* **2022**, *178*, 114627. [CrossRef]

33. Campoy, J.A.; Lerigoleur-Balsemin, E.; Christmann, H.; Beauvieux, R.; Girollet, N.; Quero-García, J.; Dirlewanger, E.; Barreneche, T. Genetic Diversity, Linkage Disequilibrium, Population Structure and Construction of a Core Collection of *Prunus avium* L. Landraces and Bred Cultivars. *BMC Plant Biol.* **2016**, *16*, 49. [CrossRef]
34. Donkpegan, A.S.L.; Bernard, A.; Barreneche, T.; Quero-García, J.; Bonnet, H.; Fouché, M.; Le Dantec, L.; Wenden, B.; Dirlewanger, E. Genome-Wide Association Mapping in a Sweet Cherry Germplasm Collection (*Prunus avium* L.) Reveals Candidate Genes for Fruit Quality Traits. *Hortic. Res.* **2023**, *10*, uhad191. [CrossRef]
35. Serrie, M.; Segura, V.; Blanc, A.; Brun, L.; Dlalah, N.; Gilles, F.; Heurtevin, L.; Le Pans, M.; Signoret, V.; Viret, S.; et al. Multi-environment GWAS uncovers markers associated to biotic stress response and genotype-by-environment interactions in stone fruit trees. *Hortic. Res.* **2025**, *12*, uhaf088. [CrossRef] [PubMed]
36. Cirilli, M.; Micali, S.; Aranzana, M.J.; Arús, P.; Babini, A.; Barreneche, T.; Bink, M.; Cantin, C.M.; Ciacciulli, A.; Cos-Terrer, J.E.; et al. The Multisite PeachRefPop Collection: A True Cultural Heritage and International Scientific Tool for Fruit Trees. *Plant Physiol.* **2020**, *184*, 632–646. [CrossRef] [PubMed]
37. Branchereau, C.; Quero-García, J.; Zaracho-Echagüe, N.H.; Lambelin, L.; Fouché, M.; Wenden, B.; Donkpegan, A.; Le Dantec, L.; Barreneche, T.; Alletru, D.; et al. New Insights into Flowering Date in *Prunus*: Fine Mapping of a Major QTL in Sweet Cherry. *Hortic. Res.* **2022**, *9*, uhac042. [CrossRef]
38. Cirilli, M.; Rossini, L.; Chiozzotto, R.; Baccichet, I.; Florio, F.E.; Mazzaglia, A.; Turco, S.; Bassi, D.; Gattolin, S. Less Is More: Natural Variation Disrupting a MiR172 Gene at the Di Locus Underlies the Recessive Double-Flower Trait in Peach (*P. persica* L. Batsch). *BMC Plant Biol.* **2022**, *22*, 318. [CrossRef]
39. Gaudinier, A.; Blackman, B.K. Evolutionary Processes from the Perspective of Flowering Time Diversity. *New Phytol.* **2020**, *225*, 1883–1898. [CrossRef]
40. Doyle, J.J.; Doyle, J.L. Isolation of Plant DNA from Fresh Tissue. *Focus* **1990**, *12*, 13–15.
41. Healey, A.; Furtado, A.; Cooper, T.; Henry, R.J. Protocol: A Simple Method for Extracting next-Generation Sequencing Quality Genomic DNA from Recalcitrant Plant Species. *Plant Methods* **2014**, *10*, 21. [CrossRef]
42. Peterson, B.K.; Weber, J.N.; Kay, E.H.; Fisher, H.S.; Hoekstra, H.E. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS ONE* **2012**, *7*, e37135. [CrossRef]
43. Catchen, J.; Hohenlohe, P.A.; Bassham, S.; Amores, A.; Cresko, W.A. Stacks: An Analysis Tool Set for Population Genomics. *Mol. Ecol.* **2013**, *22*, 3124–3140. [CrossRef]
44. Liu, C.; Feng, C.; Peng, W.; Hao, J.; Wang, J.; Pan, J.; He, Y. Chromosome-Level Draft Genome of a Diploid Plum (*Prunus salicina*). *Gigascience* **2020**, *9*, giaa130. [CrossRef]
45. Vasimuddin, M.; Misra, S.; Li, H.; Aluru, S. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. In Proceedings of the 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS), Rio de Janeiro, Brazil, 20–24 May 2019; IEEE: New York, NY, USA, 2013; pp. 314–324.
46. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The Variant Call Format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [CrossRef] [PubMed]
47. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.W.; Daly, M.J.; et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Human. Genet.* **2007**, *81*, 559–575. [CrossRef] [PubMed]
48. Hall, B.; Hall, M.; Statisticat, L. LaplacesDemon: Complete Environment for Bayesian Inference. Available online: https://cran.r-project.org/web/packages/LaplacesDemon/LaplacesDemon.pdf (accessed on 19 March 2025).
49. Jeong, S.; Kim, J.-Y.; Jeong, S.-C.; Kang, S.-T.; Moon, J.-K.; Kim, N. GenoCore: A Simple and Fast Algorithm for Core Subset Selection from Large Genotype Datasets. *PLoS ONE* **2017**, *12*, e0181420. [CrossRef] [PubMed]
50. De Beukelaer, H.; Davenport, G.F.; Fack, V. Core Hunter 3: Flexible Core Subset Selection. *BMC Bioinform.* **2018**, *19*, 203. [CrossRef]
51. Peakall, R.O.D.; Smouse, P.E. Genalex 6: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research. *Mol. Ecol. Notes* **2006**, *6*, 288–295. [CrossRef]
52. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **2000**, *155*, 945–959. [CrossRef]
53. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the Number of Clusters of Individuals Using the Software structure: A Simulation Study. *Mol. Ecol.* **2005**, *14*, 2611–2620. [CrossRef]
54. Earl, D.A.; von Holdt, B.M. STRUCTURE HARVESTER: A Website and Program for Visualizing STRUCTURE Output and Implementing the Evanno Method. *Conserv. Genet. Resour.* **2012**, *4*, 359–361. [CrossRef]
55. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for Association Mapping of Complex Traits in Diverse Samples. *Bioinformatics* **2007**, *23*, 2633–2635. [CrossRef]
56. Breseghello, F.; Sorrells, M.E. Association Mapping of Kernel Size and Milling Quality in Wheat (*Triticum aestivum* L.) Cultivars. *Genetics* **2006**, *172*, 1165–1177. [CrossRef]

57. Lazzeroni, L.C. Linkage Disequilibrium and Gene Mapping: An Empirical Least-Squares Approach. *Am. J. Human Genet.* **1998**, *62*, 159–170. [CrossRef] [PubMed]
58. Huang, M.; Liu, X.; Zhou, Y.; Summers, R.M.; Zhang, Z. BLINK: A Package for the next Level of Genome-Wide Association Studies with Both Individuals and Markers in the Millions. *Gigascience* **2019**, *8*, giy154. [CrossRef] [PubMed]
59. Lipka, A.E.; Tian, F.; Wang, Q.; Peiffer, J.; Li, M.; Bradbury, P.J.; Gore, M.A.; Buckler, E.S.; Zhang, Z. GAPIT: Genome Association and Prediction Integrated Tool. *Bioinformatics* **2012**, *28*, 2397–2399. [CrossRef] [PubMed]
60. Elbasyoni, I.S.; Lorenz, A.; Guttieri, M.; Frels, K.; Baenziger, P.; Poland, J.; Akhunov, E. A Comparison between Genotyping-by-Sequencing and Array-Based Scoring of SNPs for Genomic Prediction Accuracy in Winter Wheat. *Plant Sci.* **2018**, *270*, 123–130. [CrossRef]
61. Frankel, O.H. Genetic Perspectives of Germplasm Conservation. In *Genetic Manipulation: Impact on Man and Society*; Arber, W., Illmensee, K., Peacock, W.J., Starlinger, P., Eds.; Cambridge University Press: Cambridge, UK, 1984; pp. 161–170.
62. Deng, Q.; Zhao, S.; Guan, C.; Fu, G.; Yu, S.; Lin, S.; Wang, Z.; Fu, H.; Lu, X.; Cheng, S. Construction and Evaluation of Pepper Core Collection Based on Phenotypic Traits and SSR Markers. *Appl. Sci.* **2024**, *14*, 7473. [CrossRef]
63. Chen, X.; Liu, Y.-Y.; Zheng, F.-S.; Cheng, G.-X.; Guo, M.; Li, J.-S.; Wang, X.-M. Construction of a Core Collection of Tomato (*Solanum lycopersicum*) Germplasm Based on Phenotypic Traits and SNP Markers. *Sci. Hortic.* **2024**, *339*, 113855. [CrossRef]
64. Wang, X.; Li, X.; Chen, D.; Gao, J.; Hao, S.; Zhang, H.; Zhao, Z.; Shen, M.; Chen, H.; Qi, F.; et al. Core Germplasm Construction of Tea Plant Populations Based on Genome-Wide SNP and Catechins in Shaanxi Province, China. *J. Integr. Agric.* **2025**, *in press*.
65. Krichen, L.; Audergon, J.M.; Trifi-Farah, N. Relative Efficiency of Morphological Characters and Molecular Markers in the Establishment of an Apricot Core Collection. *Hereditas* **2012**, *149*, 163–172. [CrossRef]
66. Aranzana, M.J.; Decroocq, V.; Dirlewanger, E.; Eduardo, I.; Gao, Z.S.; Gasic, K.; Iezzoni, A.; Jung, S.; Peace, C.; Prieto, H.; et al. *Prunus* Genetics and Applications after de Novo Genome Sequencing: Achievements and Prospects. *Hortic. Res.* **2019**, *6*, 58. [CrossRef]
67. Zurn, J.D.; E Hummer, K.; Bassil, N.V. Exploring the Diversity and Genetic Structure of the U.S. National Cultivated Strawberry Collection. *Hortic. Res.* **2022**, *9*, uhac125. [CrossRef]
68. Gilpin, L.; Røen, D.; Schubert, M.; Davik, J.; Rumpunen, K.; Gardli, K.A.; Hjeltnes, S.H.; Alsheikh, M. Genetic Characterization of the Norwegian Apple Collection. *Horticulturae* **2023**, *9*, 575. [CrossRef]
69. Scott, M.F.; Ladejobi, O.; Amer, S.; Bentley, A.R.; Biernaskie, J.; Boden, S.A.; Clark, M.; Dell'aCqua, M.; Dixon, L.E.; Filippi, C.V.; et al. Multi-Parent Populations in Crops: A Toolbox Integrating Genomics and Genetic Mapping with Breeding. *Heredity* **2020**, *125*, 396–416. [CrossRef] [PubMed]
70. Carrasco, B.; Díaz, C.; Moya, M.; Gebauer, M.; García-González, R. Genetic Characterization of Japanese Plum Cultivars (*Prunus salicina*) Using SSR and ISSR Molecular Markers. *Cienc. Investig. Agrar.* **2012**, *39*, 533–543. [CrossRef]
71. Spaeth, R.A.; Pincot, D.D.; Potter, D.; Brown, P.; Gradziel, T.; Preece, J.E. Relatedness of Luther Burbank's Plum (*Prunus* sp.) Introductions Based on Genotyping by Sequencing. *HortScience* **2024**, *59*, 873–880. [CrossRef]
72. Jung, M.; Roth, M.; Aranzana, M.J.; Auwerkerken, A.; Bink, M.; Denancé, C.; Dujak, C.; Durel, C.-E.; i Forcada, C.F.; Cantin, C.M.; et al. The apple REFPOP—A reference population for genomics-assisted breeding in apple. *Hortic. Res.* **2020**, *7*, 1–16. [CrossRef]
73. Thurow, L.B.; Gasic, K.; Raseira, M.D.C.B.; Bonow, S.; Castro, C.M. Genome-Wide SNP Discovery through Genotyping by Sequencing, Population Structure, and Linkage Disequilibrium in Brazilian Peach Breeding Germplasm. *Tree Genet. Genomes* **2020**, *16*, 10. [CrossRef]
74. 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]
75. Huang, X.; Ni, Z.; Shi, T.; Tao, R.; Yang, Q.; Luo, C.; Li, Y.; Li, H.; Gao, H.; Zhou, X.; et al. Novel Insights into the Dissemination Route of Japanese Apricot (*Prunus Mume* Sieb. et Zucc.) Based on Genomics. *Plant J.* **2022**, *110*, 1182–1197. [CrossRef]
76. Carrasco, B.; Ramirez, C.; Gebauer, M.; Meisel, L.A.; Hasbun, R.; Silva, H. Phenotypic and Genetic Analysis of a Peach and a Japanese Plum Core Collection for Pre-Breeding and Distinctness Assessment. *Chil. J. Agric. Res.* **2022**, *82*, 457–468. [CrossRef]
77. Dirlewanger, E.; Quero-García, J.; Le Dantec, L.; Lambert, P.; Ruiz, D.; Dondini, L.; Illa, E.; Quilot-Turion, B.; Audergon, J.-M.; Tartarini, S.; et al. Comparison of the Genetic Determinism of Two Key Phenological Traits, Flowering and Maturity Dates, in Three Prunus Species: Peach, Apricot and Sweet Cherry. *Heredity* **2012**, *109*, 280–292. [CrossRef]
78. Mora, J.R.H.; Micheletti, D.; Bink, M.; Van de Weg, E.; Cantín, C.; Nazzicari, N.; Caprera, A.; Dettori, M.T.; Micali, S.; Banchi, E.; et al. Integrated QTL Detection for Key Breeding Traits in Multiple Peach Progenies. *BMC Genom.* **2017**, *18*, 404. [CrossRef]
79. Castède, S.; Campoy, J.A.; García, J.Q.; Le Dantec, L.; Lafargue, M.; Barreneche, T.; Wenden, B.; Dirlewanger, E. Genetic Determinism of Phenological Traits Highly Affected by Climate Change in *Prunus avium:* Flowering Date Dissected into Chilling and Heat Requirements. *New Phytol.* **2014**, *202*, 703–715. [CrossRef]

80. Branchereau, C.; Hardner, C.; Dirlewanger, E.; Wenden, B.; Le Dantec, L.; Alletru, D.; Parmentier, J.; Ivančič, A.; Giovannini, D.; Brandi, F.; et al. Genotype-by-Environment and QTL-by-Environment Interactions in Sweet Cherry (*Prunus avium* L.) for Flowering Date. *Front. Plant Sci.* **2023**, *14*, 1142974. [CrossRef] [PubMed]
81. Rawandoozi, Z.J.; Hartmann, T.P.; Carpenedo, S.; Gasic, K.; Linge, C.d.S.; Cai, L.; Van de Weg, E.; Byrne, D.H. Mapping and Characterization QTLs for Phenological Traits in Seven Pedigree-Connected Peach Families. *BMC Genom.* **2021**, *22*, 187. [CrossRef] [PubMed]
82. Pirona, R.; Eduardo, I.; Pacheco, I.; Linge, C.D.S.; Miculan, M.; Verde, I.; Tartarini, S.; Dondini, L.; Pea, G.; Bassi, D.; et al. Fine Mapping and Identification of a Candidate Gene for a Major Locus Controlling Maturity Date in Peach. *BMC Plant Biol.* **2013**, *13*, 166. [CrossRef]
83. Korte, A.; Farlow, A. The Advantages and Limitations of Trait Analysis with GWAS: A Review. *Plant Methods* **2013**, *9*, 29. [CrossRef]