




Method

Mapping the genomic landscape of *Prunus* spp. with PrunusMap

Najla Ksouri ¹, María Ángeles Moreno², Bruno Contreras-Moreira ^{3,+} and Yolanda Gogorcena ^{1,+*}

¹Group of Genomics of Fruit Trees and Grapevine, Department of Pomology, Estación Experimental de Aula Dei-Consejo Superior de Investigaciones Científicas, Avenida de Montañana 1005, E50059 Zaragoza, Spain

²Group of Fruit Tree Breeding and Fruit Quality, Department of Pomology, Estación Experimental de Aula Dei-Consejo Superior de Investigaciones Científicas, Avenida de Montañana 1005, E50059 Zaragoza, Spain

³Laboratory of Computational and Structural Biology, Department of Genetics and Plant Breeding, Estación Experimental de Aula Dei-Consejo Superior de Investigaciones Científicas, Avenida de Montañana 1005, E50059 Zaragoza, Spain

*Corresponding author. E-mail: aoiz@eead.csic.es

†Senior authors

Abstract

Next-generation sequencing has fueled significant advancement in plant breeding tools, such as genome-wide association studies and single-nucleotide polymorphism (SNP) analysis. In this dynamic landscape, plant databases housing SNP markers have evolved into hubs facilitating breeding initiatives and genomic research. PrunusMap, accessible at <https://prunusmap.eead.csic.es> is an open-source Web application tailored for the *Prunus* community. Featuring a user-friendly interface, PrunusMap empowers users to seamlessly align and locate markers across multiple genome versions of *Prunus* species and cultivars, supporting different queries and formats. Beyond locating marker positions, it provides a comprehensive list of annotated nearby genes and proteins. This streamlined process, driven by four intuitive features 'Find markers', 'Align sequences', 'Align proteins', and 'Locate by position', significantly reduces workload and boosts efficiency, particularly for users with limited bioinformatics expertise. Moreover, PrunusMap's versatility is underscored by its commitment to incorporate additional *Prunus* genome sequences, annotations, and markers upon user request.

Introduction

In the Rosaceae family, the availability of 74 whole genomes and annotations provide a robust foundation for marker development (https://www.plabipd.de/plant_genomes_pa.ep, last accessed July 2024). Restriction fragment length polymorphism (RFLP) markers were instrumental in producing a saturated map for almond, the first and most comprehensive in the stone fruit genus *Prunus*. Subsequently, randomly amplified polymorphic DNA (RAPD) markers found widespread use in germplasm diversity studies in peach and other *Prunus* species, helping map loci controlling traits such as flesh color and fruit texture [1]. The development of amplified RFLP (AFLP) markers revealed associations with traits such as resistance root-knot nematodes [2] and evergreen [3]. However, the low reproducibility (RAPDs) and high cost (RFLPs, AFLPs) of these markers led to their replacement by SSRs and SNPs. While SSR markers find frequent utility in *Prunus* breeding, SNP markers have gained prominence due to their cost-effectiveness, high throughput, and genome-wide coverage [4, 5]. Moreover, the accurate prediction of marker positions and the identification of nearby genes are critical for understanding the genetic mechanisms underlying target traits and accelerating modern breeding cycles. For instance, research on Sharka disease in peach identified three highly significant associated SNPs on

chromosomes 2 and 3, conferring a reduction in susceptibility to Plum pox virus. The *Prupe.2G065600* gene on chromosome 2, encoding an RTM2-like protein was selected as a major effect candidate gene [6]. The emergence of SNP markers has indeed revolutionized *Prunus* breeding with two main repositories cataloging the genetic variants: the PeachVar-DB portal [7] and the Genome Database for Rosaceae (GDR) [8–10]. The PeachVar-DB portal provides different tools to retrieve SNPs and Indels extracted from whole genome sequence libraries of peach and wild relatives [7]. Users can conveniently access these variants by specifying either a specific gene identifier or a genomic region of interest, with all coordinates extracted from the peach reference genome version 2.0. In contrast, the GDR stands out as a more multifaceted database, encompassing a broad spectrum of genomic and genetic data within the Rosaceae family. It provides a diverse range of tools aimed at exploring these resources. However, it lacks in capturing high-density genomic data obtained from resequencing projects [7]. Moreover, browsing the GDR often requires switching between multiple pages, which can be cumbersome.

A friendly tool to map the location of genetic markers rapidly and accurately, along with information about nearby genes, could greatly assist breeders, particularly those with limited bioinformatics expertise. To address this need, we developed PrunusMap.

Received: 26 March 2024; Accepted: 16 October 2024; Published: 24 October 2024; Corrected and Typeset: 1 February 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of Nanjing Agricultural University. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

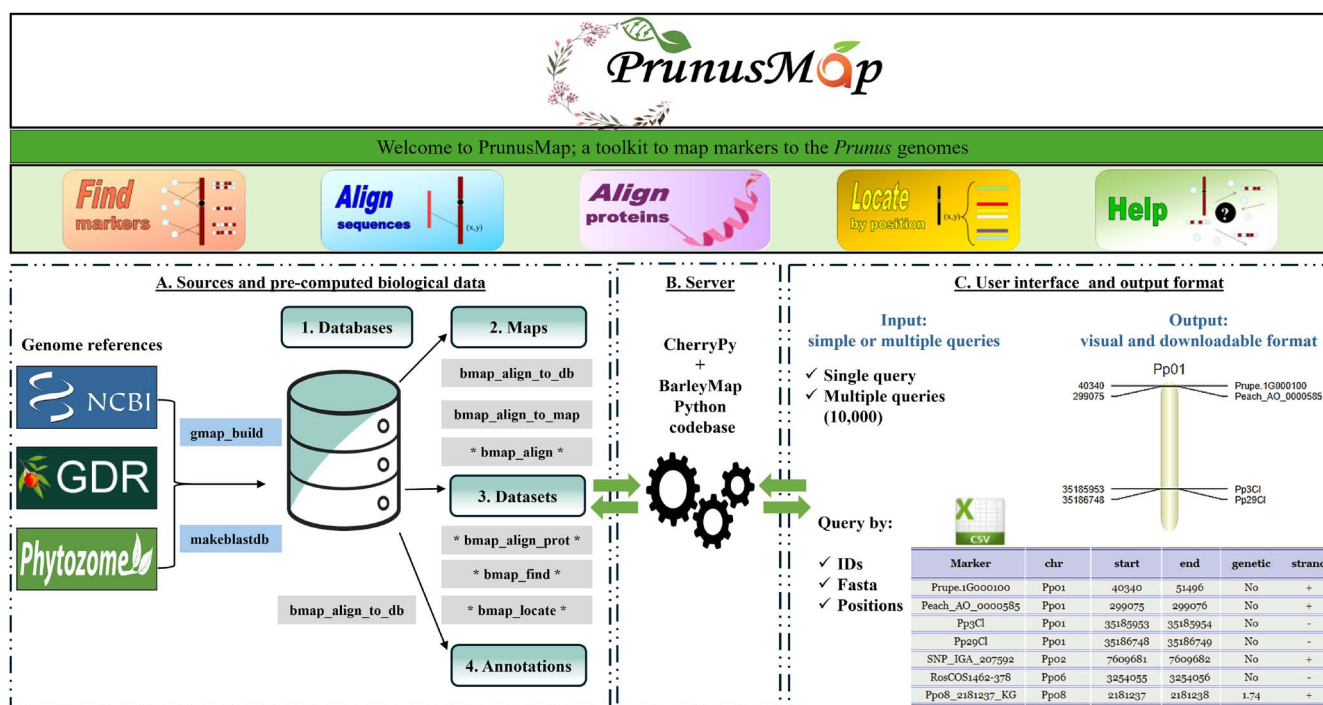


Figure 1. Workflow of PrunusMap Web application. The reference genome sequences of *Prunus* species and cultivars were downloaded from NCBI, GDR, and Phytozome. Gmap_build and makeblastdb command line executables were used to build the corresponding databases from the FASTA files. Databases, maps, datasets, and annotations represent the essential types of biological data required for PrunusMap configuration. Gray-highlighted boxes correspond to tools that are activated once each of the biological resources is properly set up (exp: “bmap_align_to_db” tool is activated once the database is well configured). Those tagged with asterisks correspond to web-accessible tools, while the rest correspond to the standalone commands. PrunusMap accepts both simple or multiple queries as input, and the results are displayed in visual and downloadable formats (CSV: Comma-Separated Values).

Its objective is to streamline the process of locating *Prunus* markers on both genetic and physical maps, accommodating various input formats. Currently, PrunusMap curates SNPs for a variety of *Prunus* species, including relevant peach cultivars (e.g. Chinese-Cling, Zhongyoutao14, 124Pan, and Sovetskiy) and wild relatives such as *Prunus davidiana*, *P. ferganensis*, *P. kansuensis*, and *P. mira*. Additionally, it supports other economically valuable species like almond (*P. dulcis*), sweet cherry (*P. avium*), apricot (*P. armeniaca*), and Japanese apricot (*P. mume*).

PrunusMap is accessible at <https://prunusmap.eead.csic.es> and offers four key features for data retrieval:

- 1) ‘Find markers’: to retrieve the position of markers by providing their identifiers.
- 2) ‘Align sequences’: to obtain the position of FASTA sequences (FAST-All, It is a format for sequences) by pairwise alignment.
- 3) ‘Align proteins’: to determine the position of FASTA amino acid sequences considering splicing and frameshift variations.
- 4) ‘Locate by position’: to examine specific loci based on their map position, to identify which genes, markers, or other loci are present in those regions.

PrunusMap is a fork of BarleyMap, a tool originally designed to serve the barley community [11], with two key enhancements: (i) it accommodates a range of *Prunus* species and cultivars, catering to the diverse needs of *Prunus* research community, and (ii) it introduces the novel functionality of mapping protein sequences, which are usually more conserved than nucleotide sequences. The data flow and architecture are summarized in Fig. 1.

Results

Comparing GMAP and BLASTN

The performance of Genomic Mapping and Alignment Program (GMAP) and Nucleotide BLAST (BLASTN) was assessed by aligning marker sequences against various *Prunus* databases. As shown in Fig. 2 and Fig. S1, GMAP consistently outperformed BLASTN in terms of mapped marker counts, regardless of the SNP array or the reference database used. As expected, the alignment rate was consistently higher for maps targeting the same species as the array. For instance, the peach 9 K and 18 K arrays exhibited a greater number of successfully aligned SNPs against peach databases compared to those of more distantly related species. Notably, the alignment rate was even higher when these arrays were aligned against the *P. persica* cv. Lovell, which was used as the reference genome to develop these arrays (Figs 2A and S1A). Similarly, SNP markers from the 60 K almond array mostly matched the Texas_v2 databases compared to Texas v3, Lauranne, and Nonpareil cultivars (Fig. 2B). This is likely due to the array design, which involved the alignment of resequenced raw reads against Texas_v2 prior to SNP calling [12].

The 6 K cherry SNP array exhibited a bias toward the peach genome (cv. Lovell) (Fig. S1B). This discrepancy is likely due to the use of the Lovell peach reference for SNP discovery [13].

For clarity, the ‘Adafuel’ dataset was mapped only to the peach genome cv. Lovell from the National Center for Biotechnology Information (NCBI) and Joint Genome Institute (JGI) (Fig. 2C). Interestingly, both aligners exhibited comparable accuracy, yielding similar mapping results.

GMAP also excelled in transcript mapping (Fig. 2C). It successfully aligned a substantially higher number of transcripts

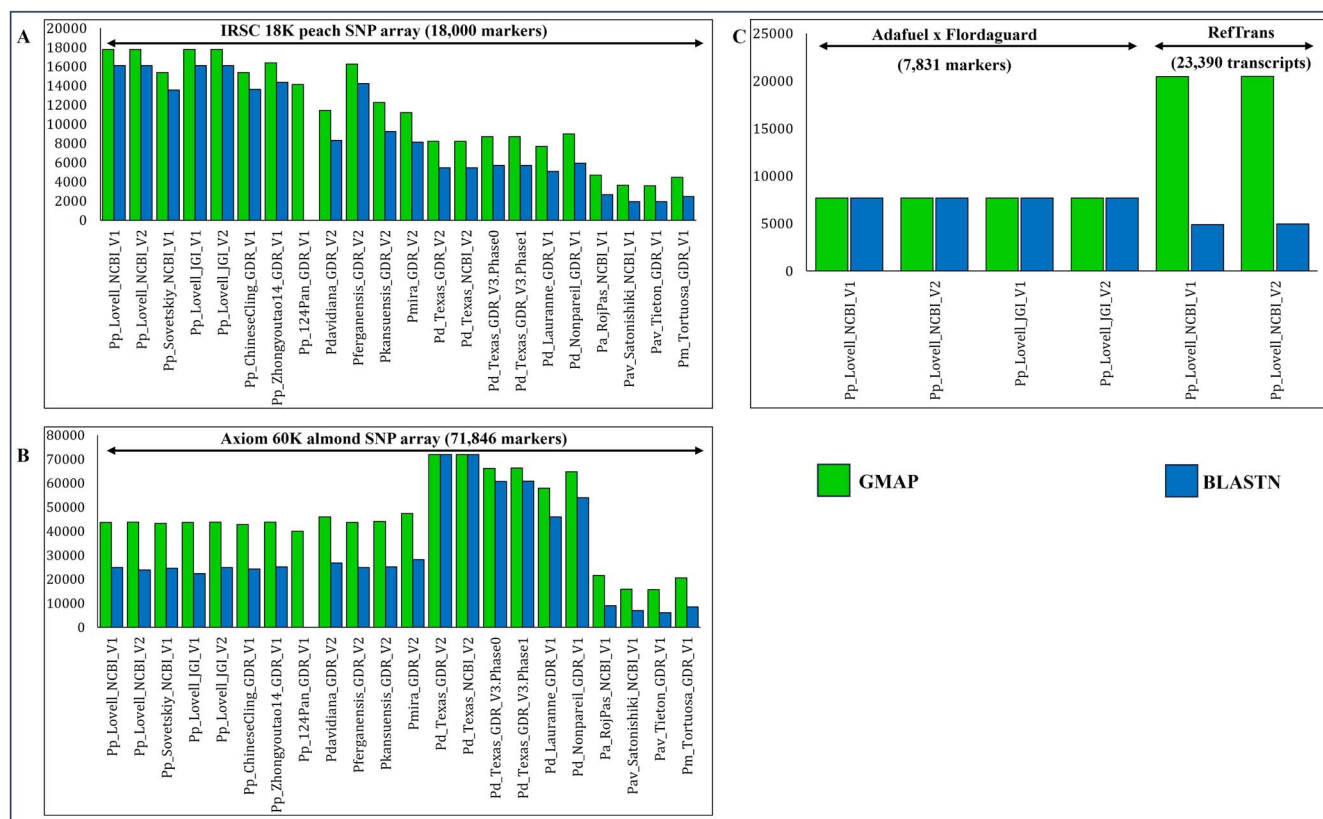


Figure 2. Performance comparison of GMAP and BLASTN aligners. (A) and (B) correspond to the alignment of the IRSC 18 K peach and 60 K almond arrays against all PrunusMap databases, respectively. (C) corresponds to the alignment of genetic markers and transcripts. The x-axis refers to the different databases (maps according to PrunusMap terminology) used for the alignment, while the y-axis corresponds to the number of sequences.

from the peach reference transcriptome (RefTrans_V1; 23 390) compared to BLASTN (20 482 vs. 4902).

Altogether, GMAP consistently surpassed BLASTN across all tested scenarios, achieving significantly higher mapping rates and faster processing times (seconds vs. minutes) compared to BLASTN. Unmapped queries can be attributed to the stringent sequence identity and query coverage thresholds, set by default at 98% and 95%, respectively. Hits below these cutoffs are deemed unreliable and are discarded from the datasets.

To measure mapping accuracy, we aligned the nucleotide probes of the peach 18 K SNP array to the Pp_Lovell_NCBI_V2 reference and checked how often the genomic regions returned by GMAP and BLASTN contained the cognate SNP position. Among 16 095 probes aligned by both algorithms, 16 093 contained the correct SNP location. Only two discrepancies were observed, due to probes mapping to multiple genomic regions with similar sequence identity. The same experiment was carried out with almond 60 K and cherry 6 K SNP arrays, finding that all mapped SNPs (71 835 and 5492) were accurately positioned within the aligned intervals identified by both GMAP and BLASTN.

Benchmarking the accuracy of PrunusMap

'Adafuel' marker sequences were aligned against the different databases of *P. persica* cv. Lovell using 'bmap_align'. The resulting physical positions (in megabase pairs) were plotted against their genetic positions (in centimorgans), as reported by [14]. Notably, high collinearity was evident across all chromosomes, with 'Pearson' correlation coefficients ≥ 0.96 for reference Pp_Lovell_JGI_V2 (Fig. 3A). Overlapping gaps between genetic

and physical positions were mainly observed in chromosomes Pp02, Pp07, and Pp08. These are explained by the uneven distribution of markers mapped in the Adafuel \times Flordaguard population [14]. Note that there are no markers mapping on the short arm of Pp04. Interestingly, when Pp_Lovell_JGI_V1 was used as reference, inversions in scaffolds 1, 6, and 7 were revealed, as confirmed by Verde et al. [15] (Fig. 3B). Similarly, a misplaced contig appears near the centromere of scaffold 4. According to our results, these assembly conformations were corrected in V2, and 60 chromosomal SNPs previously located on unplaced scaffolds were also rectified. Comparable results were obtained for NCBI references, plotted in Figs S2 and S3.

Highlighting protein alignment as a new key feature

PrunusMap extends BarleyMap by introducing a novel feature for protein sequence alignment. Indeed, the novel 'Align_proteins' tool drives protein-to-genome alignments. To benchmark 'Align_proteins', stringent criteria (identity $\geq 98\%$ and coverage $\geq 95\%$) were employed to align complete protein sequences along with short peptides of 50 and 100 bp, against their respective reference genomes. Remarkably, this tool demonstrated high reliability, successfully aligning over 85% of protein queries (Fig. S4). We further assessed cross-species protein mapping. Here, we attempted to map protein sequences from a single species (e.g. peach cv. Lovell_NCBI_V2) against all reference genomes within PrunusMap (Fig. 4). This approach allows researchers to leverage well-annotated proteins for functional insights in species with limited annotations.

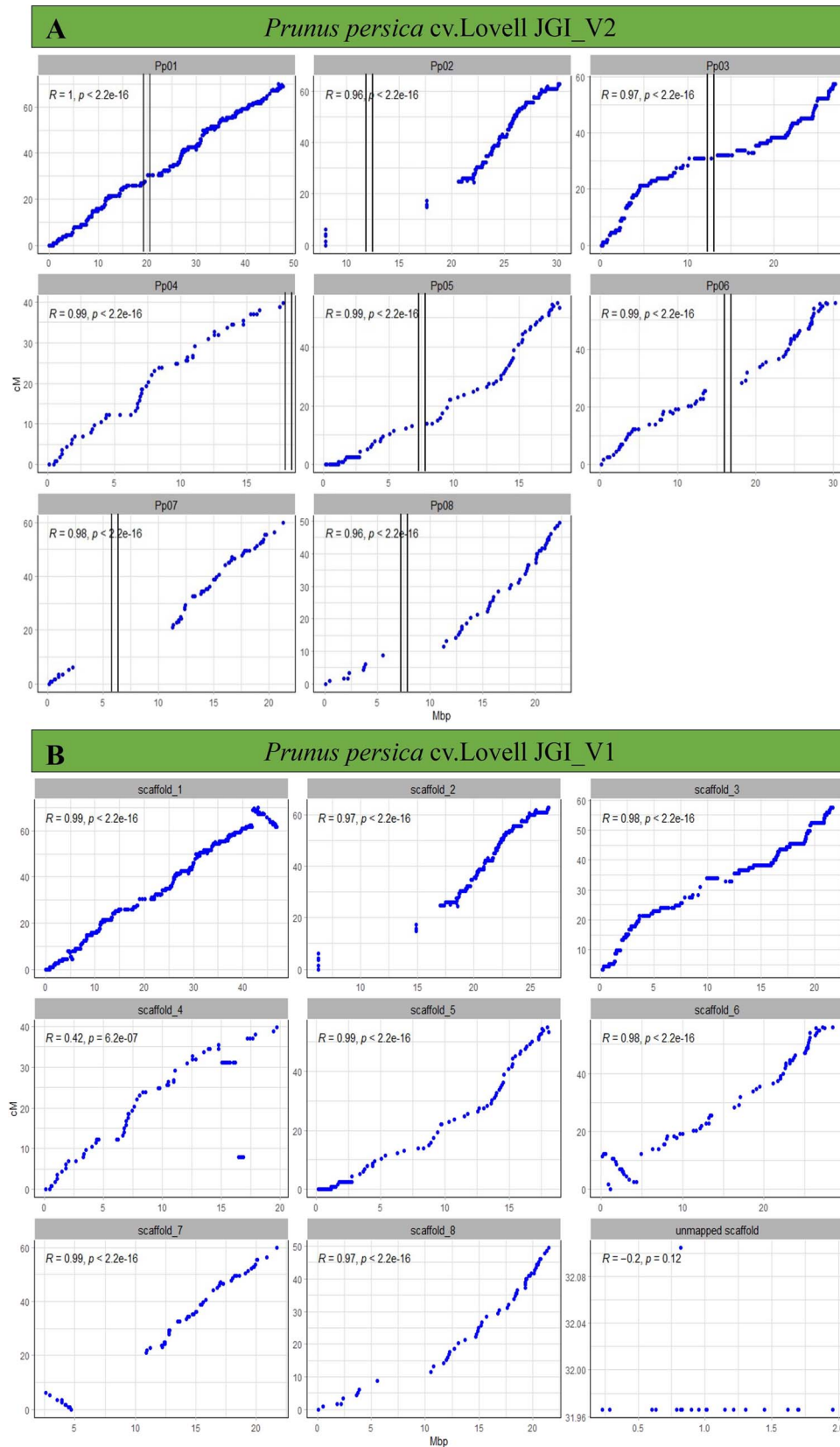


Figure 3. Relationship between the genetic and physical position of 'Adafuel' SNP markers within each pseudomolecule (chromosomes). For *P. persica* Pp_Lovell_JGI_V2 (A), pseudomolecules are referred to as Pp01 to Pp08, while in *P. persica* Pp_Lovell_JGI_V1 (B), they are labeled as scaffold_1 to scaffold_8. SNP markers were plotted according to their physical position on the peach genome reference (x-axis), and their genetic position was retrieved from the 'Adafuel' linkage map (y-axis). Vertical bars indicate the putative position of the centromeres, and R values correspond to the 'Pearson' correlation.

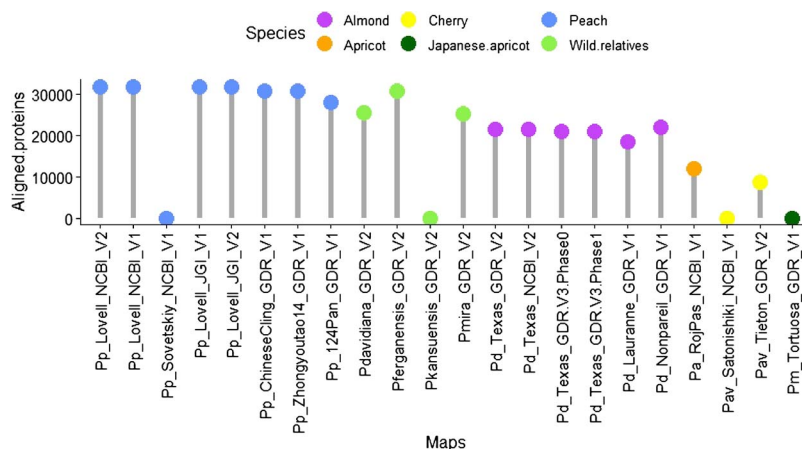


Figure 4. Cross-species protein mapping in PrunusMap. Protein sequences from peach cultivar Lovell (NCBI_V2) were mapped against reference genomes (maps) (x-axis) within PrunusMap. The y-axis represents the number of protein sequences with high-fidelity alignments.

Examining the insights provided by PrunusMap

The ‘Find markers’ tool successfully mapped 10 out of 11 user-provided markers linked to peach bacterial spot resistance to Pp_Lovell_JGI_V2 (Fig. 5). The unmapped marker (SNP_IGA_46754) failed to meet minimum alignment quality thresholds. Mapped markers reside on chromosomes Pp01 and Pp06, flanking genes annotated with disease resistance protein domains (LRR and NB-ARC) (*Prupe.1G165300*, *Prupe.6G243700*, and *Prupe.6G243800*). These genes are known as major disease-resistance genes in plants and have been reported to be involved in pathogen recognition and innate immune responses in peach [16].

Discussion

In the following sections, we compare PrunusMap and other tools offering similar functionalities.

Genome Database for Rosaceae and PrunusMap

The GDR serves as a central hub of the Rosaceae family. As a public repository, it provides access to multiple versions of genome assemblies enriched with gene descriptions, InterPro domains, and Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway terms [10]. Moreover, it hosts collections of expressed sequence tags, full-length transcripts, metabolic pathways, maps, and quantitative and Mendelian trait loci linked to agronomically significant traits [10]. Beyond its role as a repository, the GDR provides analytical tools to explore genetic and genomic data (search, sequence retrieval, BLAST, synteny viewer, map viewer, Breeding Information Management System (BIMS)).

While sharing similar purposes, PrunusMap is specifically designed to facilitate the identification of *Prunus* markers on both physical and genetic maps, providing a streamlined analysis that includes a detailed list of nearby genes and proteins. Although this functionality may seem to be fulfilled by GDR, there are key differences incorporated in PrunusMap.

First, regarding sequence homology searches, the GDR relies exclusively on BLAST, whereas PrunusMap employs the GMAP as its default aligner complemented with BLASTN. Based on our findings, GMAP consistently outperformed BLASTN in locating marker sequences (Figs 2 and S1). This performance gain is further accentuated when mapping transcripts, which require

intron-aware alignments. These results agree with those reported in barley [11].

Second, when searching for molecular markers, the GDR supports an extensive list of parameters such as marker type, name, array, organism, chromosome, map, trait, and citation. PrunusMap further refined its search capabilities to align with our vision, offering a graphical visualization of marker locations on their corresponding chromosomes. This view is accompanied by a table summarizing the start and end positions of the markers, the strand, and, crucially, a list of annotated genes and proteins residing nearby. The strand orientation, absent in GDR, is particularly valuable for breeders as it can assist in designing Polymerase Chain Reaction (PCR) primers or Kompetitive allele specific PCR (KASP) markers, predicting missense mutations, and defining haplotypes. Finally, by displaying nearby genes on the same results table, PrunusMap eliminates the need to conduct multiple searches and navigate from one webpage to another. This streamlined approach enhances efficiency and simplifies the process for researchers and breeders alike.

Third, PrunusMap offers a significant advantage by enabling cross-species and intraspecies marker mapping across diverse genome versions and cultivars. This capability addresses the gene nomenclature discrepancy that already exists between different databases, such as Pp_Lovell_NCBI and Pp_Lovell_JGI. Additionally, many published studies often tend to focus on a single cultivar, resulting in other cultivars being overlooked. For instance, a marker closely linked to a desirable trait in one cultivar may exhibit a different position or allele in another cultivar. PrunusMap can handle this imbalance enabling breeders to explore several references in the same search. This, in turn, facilitates the identification of new sources of desirable traits. For example, the ability to map the 60 K Axiom markers on three different almond genome references (Texas, Lauranne, and Nonpareil) would enable breeders to easily explore the genetic diversity and relationships among these three cultivars.

Fourth, PrunusMap streamlines retrieving the physical and genetic positions of all markers from a given genetic map in a single operation, as shown here with ‘Adafuel’ markers, which support the analysis of genetic maps in a single step.

Finally, PrunusMap further enhances user experience by enabling the option to send search results via email. This feature is particularly helpful for large-scale analyses that may require extended processing times.

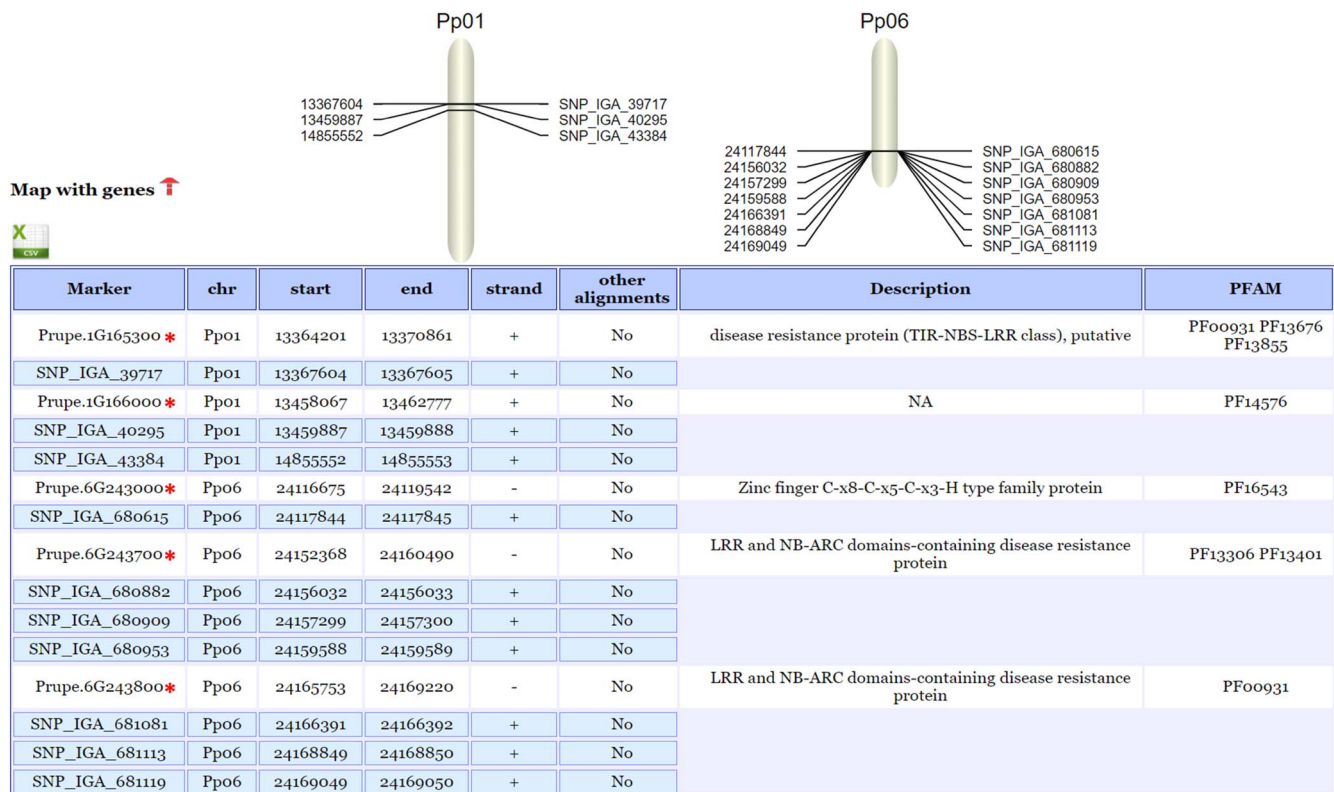


Figure 5. Illustration of PrunusMap functionalities. Genes marked with an asterisk correspond to nearby genes identified through the 'Find marker' tool.

Overall, both the GDR and PrunusMap serve as valuable tools, each offering distinct, yet complementary features tailored to different user preferences.

PeachVar-DB and PrunusMap

PeachVar-DB is a valuable resource for exploring the genetic makeup of a collection comprising 121 peach accessions and 21 wild relatives from the *Amygdalus* subgenus derived from the resequencing [7]. Users can get a broad overview by selecting a specific accession, delving into a specific genome region, or conducting a targeted gene-level analysis by providing the gene ID and features such as 5' Untranslated Region (UTR), 3'UTR, Coding Sequence (CDS), or primary transcript. However, unlike PrunusMap, when selecting an accession or chromosome region, PeachVar-DB does not display information on nearby genes alongside the genetic variants. As with GDR, PeachVar-DB utilizes BLASTN for sequence similarity comparison but lacks graphical visualization of the mapping results. Additionally, PrunusMap introduced the 'Align_proteins' feature to perform protein-to-nucleotide alignments. This functionality complements nucleotide-based sequence searches, as it can map protein sequences annotated in species, which are often more conserved than gene sequences. Furthermore, in contrast to PrunusMap, PeachVar-DB does not support multiquery searches and file upload functionalities. Finally, it exclusively presents information on markers aligned to the peach reference genome v2.0.

Conclusions and future directions

PrunusMap was developed to empower *Prunus* researchers with user-friendly analysis tools to support decision-making and accelerate breeding goals. We anticipate it will serve as a valuable

tool for breeders in combination with GDR and PeachVar-DB. Furthermore, we expect PrunusMap to be continuously updated and expanded to cover other *Prunus* species upon demand from users. We welcome feedback and suggestions at compbio@eed.csic.es.

Material and methods

PrunusMap Web interface

PrunusMap is a freely accessible application created as a fork of Barleymap [11]. Its back-end functionality and interactivity are implemented in Python 2.6, relying on CherryPy to handle the user requests [17]. The front-end interface uses a Perl graphical library for chromosome visualization and intuitive interaction (<https://github.com/pseudogene/genetic-mapper>) (Fig. 1).

Biological resources

PrunusMap stores and categorizes data into four different classes: databases, maps, datasets, and annotations.

Databases

Databases are FASTA-formatted genome sequence files that support sequence alignments. These genome references were sourced from NCBI [18], Phytozome (JGI) [19], and GDR [8–10] and indexed using 'gmap_build' and 'makeblastdb' from GMAP and BLASTN tools [20, 21]. Additionally, for protein mapping, they were indexed using MINIPROT v13.0 [22]. An example of database configuration is provided in Fig. S5.

Currently, PrunusMap hosts a comprehensive collection of 22 genomes of *Prunus* species all described in Fig. 6: several relevant *P. persica* cultivars (e.g. ChineseCling, Zhongyoutao14, 124Pan, and Sovetskiy); wild relatives such as *P. davidiana*, *P. ferganensis*, *P.*

	Cultivars / Species	Assemblies / BioProjects	Chr/ IDs	Scaffold number	Scaffold IDs
<p><i>P. persica</i> and wild relatives*</p>		NCBI	v1		
	Lovell	GCF_000346465.1	---	202	NW_006760184.1 – NW_006760385.1
	Lovell	GCF_000346465.2	NC_034009.1 – NC_034016.1	183	NW_018027148.1 – NW_018027330.1
	Sovetskyi	GCA_022343065.3	CM039277.1 – CM039284.1	---	JAJDMZ010000009.1 – JAJDMZ010000236.1
	Lovell	GCF_000346465.1	---	202	scaffold_1 ...
	Lovell	GCF_000346465.2	Pp01 – Pp08	183	scaffold_1 ...
	ChineseCling	PRJNA745488	chr1 – chr8	127	contig 1...
	Zhongyoutao14	www.stylebio.cn	G1 – G8	---	---
	124Pan	CNA0019231	Pp01_RaGOO – Pp08_RaGOO	---	scaffold_12_RaGOO ...
	<i>dauidiana</i> *	PRJNA753201	Pda01 – Pda08	556	ptg0000091 ...
	<i>ferganensis</i> *	PRJNA753201	Pfe01 – Pfe08	439	ptg0000041 ...
	<i>kansuensis</i> *	PRJNA753201	Pka01 – Pka08	330	ptg0000141 ...
<i>mira</i> *	PRJNA753201 PRJNA753549	Pmi01 – Pmi08	364	ptg0000111 ...	
<p><i>P. dulcis</i></p>		NCBI	v2		
	Texas	GCF_902201215.1	NC_047650.1 – NC_047657.1	683	NW_023010004.1 – NW_023010686.1
	Texas	GCF_902201215.1	Pd01 – Pd08	683	pdulcis26_s0345 – pdulcis26_s1699
	Texas.Phase0	ERP158378	Chr01 – Chr08	---	---
	Texas.Phase1	ERP158378	Chr01 – Chr08	---	---
	Lauranne	AP019297-AP019304	Pd01 – Pd08	---	---
	Nonpareil	GCA_021292205	CM037988.1 – CM037995.1	95	JAJFAZ020000011.1 – JAJFAZ020000105.1
<p><i>P. armeniaca</i></p>		NCBI			
Rojo Pasión	GCA_903112645.1	---	93	CAEKDK010000011 – CAEKDK010000093.1	
<p><i>P. avium</i></p>		NCBI	v2		
	Satonishiki	GCA_002207925.1	---	10148	NW_018921241.1 – NW_018931388.1
	Tieton	GCA_014155035.1	chr_1 – chr_8	53	contig9 – contig61
<p><i>P. mume</i></p>		GDR	v2		
Tortuosa	...	Chr1 – Chr8	24	scaffold9 – scaffold32	

Figure 6. Illustration of PrunusMap databases (maps according to PrunusMap terminology). Please note that when ‘. . .’ is used, it means that IDs are not in consecutive order. However, when a range is used; for instance, NC_034009.1 – NC_034016.1 means the IDs are named in continued order. Abbreviations are as follows: Pp—*P. persica* (peach), Pd—*P. dulcis* (almond), Pa—*P. armeniaca* (apricot), Pav—*P. avium* (sweet cherry), and Pm—*P. mume* (Japanese apricot). The asterisk (*) refers to closely related wild peach species.

kansuensis, and *P. mira*; and other economically valuable species like almond (*P. dulcis*), sweet cherry (*P. avium*), apricot (*P. armeniaca*), and Japanese apricot (*P. mume*).

Maps

Maps are files designed to store the positional arrangement, whether physical or genetic, of sequences derived from databases.

PrunusMap employs standardized map identifiers, combining species abbreviations, cultivar names (if applicable), genome source, and version information. This format facilitates efficient map identification (e.g. Pp_ChineseCling_gdr1 represents *P. persica* cultivar ‘ChineseCling’, genome v1 downloaded from GDR). A detailed list of maps is available in the help section <https://prunusmap.eead.csic.es/prunusmap/help>.

Table 1. Gene models (A) markers (B) and mapped protein sequences (C) housed in PrunusMap.

Species	Maps	A. Gene models		
		Total	IDs	References
<i>Prunus persica</i> and wild relatives*	Pp_Lovell_NCBI_V1	28 087	PRUPE_ppa019766mg	[15, 23]
	Pp_Lovell_NCBI_V2	25 030	LOC18793189	
	Pp_Lovell_JGI_V1	27 864	ppa000003m.g	
	Pp_Lovell_JGI_V2	26 873	Prupe.1G000100	
	Pp_ChineseCling_GDR_V1	26 335	evm.TU.contig279.5	[24]
	Pp_Zhongyoutao14_GDR_V1	30 181	Pp01G000510	[25]
	Pp_124Pan_GDR_V1	25 155	P124PAN00019	[26]
	Pp_Sovetskiy_NCBI_V1	128	gene-ndhA	[27]
	Pdavidiana_GDR_V2*	27 236	Pda01g0001	[28]
	Pferganensis_GDR_V2*	28 587	Pfe01g0001	
	Pkansuensis_GDR_V2*	26 986	Pka01g0001	
	Pmira_GDR_V2*	28 519	Pmi01g0001	
	<i>Prunus dulcis</i>	Pd_Lauranne_GDR_V1	23 266	Prudu_020920_v1.0
Pd_Nonpareil_GDR_V1		45 581	L3X38_000408	[30]
Pd_Texas_NCBI_V2		25 445	LOC117629531	[31]
Pd_Texas_GDR_V2		27 042	Prudu126A002130	
Pd_Texas_GDR_V3.Phase0		29 145	TexasF0_G5	[32]
Pd_Texas_GDR_V3.Phase1		30 150	TexasF1_G30142	
<i>Prunus armeniaca</i>	Pa_RojPas_NCBI_V1	52 344	Gene-CURHAP_LOCUS15	[33]
<i>Prunus avium</i>	Pav_Satonishiki_NCBI_V1	28 800	LOC110751831	[34]
	Pav_Tieton_GDR_V1	39 984	FUN_000009	[35]
<i>Prunus mume</i>	Pm_Tortuosa_GDR_V1	29 706	PmuVar_Chr2_0001	[36]
SNP Arrays	Total	B. Markers		
		IDs	References	
IRSC 9 K peach	9000	SNP_IGA_134631 snp_scaffold_1_46157131 Pp8Cl RosCOS1338-411	[37]	
IRSC 18 K peach	18 000	SNP_IGA_679 Peach_AO_0000136 snp_scaffold_8_17395002 RosCOS1549-533	https://www.rosaceae.org/organism/26133	
IRSC 6 K cherry	5696	RosCOS1139- 146_snp_sweet_cherry_Pp1_43832684	[13]	
Axiom 60 K	71 846	AX-158803044	[12]	
Adafuel	7831	Pp01_10008318_YC	[14]	
Maps	Total	C. Proteins UniProt		
		IDs	References	
Pp_Lovell_NCBI_V2	38 303	E3WOH3_PRUPE	[15]	
Pdavidiana_GDR_V2	229	A0A385H4F7_9ROSA	[28]	
Pferganensis_GDR_V2	87	A0A6B9IN88_9ROSA		
Pd_Lauranne_GDR_V1	39 030	A0A1W6CB65_PRUDU	[29]	
Pd_Nonpareil_GDR_V1	38 377		[30]	
Pd_Texas_NCBI_V2	42 636		[31]	
Pd_Texas_GDR_V3.Phase0	39 243		[32]	
Pd_Texas_GDR_V3.Phase1	39 222			
Pa_RojPas_NCBI_V1	43 609	A0A6J5TF90_PRUAR	[33]	
Pav_Satonishiki_NCBI_V1	29 660	A0A6P5SIM1_PRUAV	[34]	
Pav_Tieton_GDR_V1	27 102		[35]	
Pm_Tortuosa_GDR_V1	225	A0A2H4N340_PRUMU	[36]	

The asterisk * refers to closely related wild peach species belonging to the genus *Prunus* in the Rosaceae family.

Standalone version

Secondary tools such as 'bmap_align_to_db' and 'bmap_align_to_map' are only available in the standalone version, which can be installed following the instructions in the GitHub repository.

Datasets

PrunusMap datasets are genes, molecular markers, and UniProt proteins, often associated with AlphaFold structural models [21]. Each dataset is a collection of one of these classes along with

their precomputed map positions, determined through sequence alignment against the reference database. Genes and markers were aligned using GMAP and/or BLASTN, while proteins were mapped with MINIPROT [22]. See Table 1 for a summary of gene models, markers, and lifted-over proteins.

PrunusMap provides a comprehensive collection of SNP markers, including the International Rosaceae SNP Consortium (IRSC) peach 9 K and 18 K arrays, the IRSC 6 K array for sweet cherry, and the Axiom 60 K Chip for almond. These datasets were sourced from the GDR (<https://www.rosaceae.org/organism/26133>) with the commitment to expand the collection based on user requests. The peach 9 K array originated from resequencing 56 peach accessions, with the 18 K chip built upon it. The cherry 6 K array was developed based on a detection panel of 16 sweet and eight sour cherry accessions, while the Axiom 60 K was derived from whole-genome resequencing of 81 almond genomes [12]. Notably, all these markers have been mapped across all reference genomes within PrunusMap, empowering users to efficiently identify the precise location of any marker on any available reference genome.

Genetic markers for the hybrid peach-almond rootstock 'Adafuel' were retrieved from the maps resulting from the analysis of the Adafuel × Flordaguard population [14]. Only SNPs from the 'Adafuel' linkage map were selected as they covered the eight-linkage group. Conversely, only four groups were constructed for 'Flordaguard'. Currently, this is the only dataset in PrunusMap featuring genetic positions.

Annotations

Gene datasets were enriched with functional annotations from InterPro and Pfam databases [38, 39]. To ensure the highest accuracy, annotations were limited to references *P. persica* (Pp_Lovell_JGI_V2) and *P. dulcis* (Pd_Texas_GDR_V2).

PrunusMap commands: navigating the toolkit

PrunusMap offers a variety of Web and standalone tools, which are summarized below. While the former are publicly accessible, standalone tools require local installation and configuration of PrunusMap. Check the repository https://github.com/eead-csic-compbio/prunusmap_web and the help section at <https://prunusmap.eead.csic.es/prunusmap/help> for more details (Fig. 1).

Web-based tools

Markers can be searched using different inputs: FASTA-formatted genomic and protein sequences, IDs, or positions. This is facilitated by the following standalone tools, which can examine up to 10 000 entries on single or multiple maps.

- 'bmap_align': aligns FASTA sequences to reference databases using GMAP, BLASTN, or both [18, 21]. Initially, queries are searched using GMAP with BLASTN taking over if no matches are found. This iteration continues until either all queries are aligned, or no additional databases are available. The default parameters are set as minimum identity = 98% and minimum coverage = 95% but can be customized.

- 'bmap_align_prot': Similar to bmap_align, this tool aligns amino acid sequences considering splicing and frameshifts. The adjustable identity and query coverage cutoffs were set as 98% and 95% by default to ensure accurate protein hits.

- 'bmap_find': takes a list of query IDs and retrieves their alignment positions from the pre-computed datasets listed in Table 1.

- 'bmap_locate': locates features (genes, markers, and/or proteins) based on their chromosomal or scaffold positions. The input

requires a list formatted with chromosome/scaffold names and base-pair/centimorgan positions in-line (e.g. Chr2 12 496 912).

For each tool, users can optionally retrieve lists of nearby genes and/or markers within a customizable radius (in base pairs) based on their specific needs.

PrunusMap output

The output is provided through the Web interface or conveniently sent to an email address for ease of interpretation and storage. Within the Web interface, the results are displayed in two formats: as a graphical representation of the genome, highlighting the query locations and as a downloadable Comma Separated Values (CSV) file (Fig. 1). An additional table showcasing the position of nearby markers, genes, or proteins is also provided. The graphical interface that depicts chromosomes as vertical bars offers functionality comparable to classic genome browsers, allowing users to visualize genomic features relative to their chromosomal positions (Fig. S6). The search radius for neighboring features can be tailored and defined either in base pairs (bp) or in centimorgans (cM), depending on the underlying selected map (Fig. S7). Additional details are provided in the help section: <https://prunusmap.eead.csic.es/prunusmap/help>.

Showcase analysis

To benchmark PrunusMap, we analyzed relevant markers documented in the literature. For instance, [40] reported two quantitative trait loci (QTLs) linked to fruit resistance to bacterial spot infection ('Xap.PpOC-1.2' and 'Xap.PpOC-1.6'), sitting on chromosomes 1 and 6. Eleven SNPs in close proximity to these QTLs were used to design KASP markers, resulting in a 44% reduction in seedling planting [40]. In this context, we tested the 'Find markers' Web tool in order to locate the 11 SNPs that correspond to the following 9 K peach array markers: SNP_IGA_39717, SNP_IGA_40295, SNP_IGA_43384, SNP_IGA_46754, SNP_IGA_680615, SNP_IGA_680882, SNP_IGA_680909, SNP_IGA_680953, SNP_IGA_681081, SNP_IGA_681113, and SNP_IGA_681119.

Acknowledgments

The authors would like to thank Dr Veronica Guajardo for providing data on 'Adafuel' SNPs and Bixente Sehabiague for his technical assistance. This work was funded by CSIC [grants 2020AEP119 & FAS2022_052], the Spanish Research Agency [grants AGL2017-83358-R & PID2023-146749OB-I00, funded by MCIN/AEI/10.13039/501100011033 'A way of making Europe'] and the Government of Aragón [grants A44, A09_23R, A10_23R, and PhD contract to Najla Ksouri 2018-2023], which were cofinanced with FEDER funds.

Author contributions

Y.G. and B.C.-M. conceived the project and its components. M.-A.M. provided genetic data; B.C.-M. created and maintained the PrunusMap virtual machine and added support to protein alignments; N.K. created the datasets, analyzed results, and wrote the draft manuscript; and B.C.-M. and Y.G. discussed and revised the manuscript. All authors read and approved the final manuscript.

Data availability statement

PrunusMap is freely accessible at <https://prunusmap.eead.csic.es>. Instructions for configuring and using the Web

application can be found at https://github.com/ead-csic-compbio/prunusmap_web. All sets of curated SNP markers mapped using PrunusMap are accessible at https://github.com/ead-csic-compbio/prunusmap_web/tree/master/download.

Conflict of interest

None declared.

Supplementary Data

Supplementary data are available at Horticulture Research online.

References

- Araújo M D C, Rodrigues P, Filho SA. et al. Genetic variability in the peach palm genebank examined with RAPD markers. *Crop Breed Appl Biotechnol.* 2010;**10**:211–7
- Gillen AM, Bliss FA. Identification and mapping of markers linked to the Mi gene for root-knot nematode resistance in peach. *J Am Soc Hortic Sci.* 2005;**130**:24–33
- Wang Y, Georgi LL, Reighard GL. et al. Genetic mapping of the evergrowing gene in peach [*Prunus persica* (L.) Batsch]. *J Hered.* 2002;**93**:352–8
- Aranzana MJ, Decroocq V, Dirlwanger E. et al. *Prunus* genetics and applications after *de novo* genome sequencing: achievements and prospects. *Hortic Res.* 2019;**6**:58–25
- Butiuc-Keul A, Coste A, Postolache D. et al. Molecular characterization of *Prunus* cultivars from Romania by microsatellite markers. *Horticultrae.* 2022;**8**:291–3
- Cirilli M, Rossini L, Geuna F. et al. Genetic dissection of Sharka disease tolerance in peach (*P. persica* L. Batsch). *BMC Plant Biol.* 2017;**17**:192–15
- Cirilli M, Flati T, Gioiosa S. et al. PeachVar-DB: a curated collection of genetic variations for the interactive analysis of peach genome data. *Plant Cell Physiol.* 2018;**59**:e2–9
- Jung S, Jesudurai C, Staton M. et al. GDR (Genome Database for Rosaceae): integrated web resources for Rosaceae genomics and genetics research. *BMC Bioinformatics.* 2004;**5**:1–8
- Jung S, Staton M, Lee T. et al. GDR (Genome Database for Rosaceae): integrated web-database for Rosaceae genomics and genetics data. *Nucleic Acids Res.* 2008;**36**:D1034–40
- Jung S, Lee T, Cheng C-H. et al. 15 years of GDR: new data and functionality in the genome database for Rosaceae. *Nucleic Acids Res.* 2019;**47**:D1137–45
- Cantalapiedra CP, Boudiar R, Casas AM. et al. BARLEYMAP: physical and genetic mapping of nucleotide sequences and annotation of surrounding loci in barley. *Mol Breed.* 2015;**35**:1–13
- Duval H, Coindre E, Ramos-Onsins SE. et al. Development and evaluation of an Axiom™ 60K SNP array for almond (*Prunus dulcis*). *Plants (Basel).* 2023;**12**:1–12
- Peace C, Bassil N, Main D. et al. Development and evaluation of a genome-wide 6K SNP array for diploid sweet cherry and tetraploid sour cherry. *PLoS One.* 2012;**7**:1–15
- Guajardo V, Martínez-García PJ, Solís S. et al. QTLs identification for iron chlorosis in a segregating peach–almond progeny through double-digest sequence-based genotyping (SBG). *Front Plant Sci.* 2022;**13**:1–14
- Verde I, Jenkins J, Dondini L. et al. The Peach v2.0 release: high-resolution linkage mapping and deep resequencing improve chromosome-scale assembly and contiguity. *BMC Genomics.* 2017;**18**:225–18
- Fu W, Da Silva Linge C, Gasic K. Genome-wide association study of Brown rot (*Monilinia* spp.) tolerance in peach. *Front Plant Sci.* 2021;**12**:1–14
- Hellegouarch, S. (2007). *CherryPy: Essentials Rapid Python Web Application Development Design, Develop, Test, and Deploy your Python Web Applications Easily*. Packt publishing Ltd. www.visionwt.com
- Sayers EW, Bolton EE, Brister JR. et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 2022;**50**:D20–6
- Goodstein DM, Shu S, Howson R. et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 2012;**40**:D1178–86
- Boratyn GM, Camacho C, Cooper PS. et al. BLAST: a more efficient report with usability improvements. *Nucleic Acids Res.* 2013;**41**:W29–33
- Wu TD, Watanabe CK. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics.* 2005;**21**:1859–75
- Li H. Protein-to-genome alignment with miniprot. *Bioinformatics.* 2023;**39**:1–6
- Verde I, Abbott AG, Scalabrin S. et al. The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nat Genet.* 2013;**45**:487–94
- Cao K, Yang X, Li Y. et al. New high-quality peach (*Prunus persica* L. Batsch) genome assembly to analyze the molecular evolutionary mechanism of volatile compounds in peach fruits. *Plant J.* 2021;**108**:281–95
- Lian X, Zhang H, Jiang C. et al. *De novo* chromosome-level genome of a semi-dwarf cultivar of *Prunus persica* identifies the aquaporin PpTIP2 as responsible for temperature-sensitive semi-dwarf trait and PpB3-1 for flower type and size. *Plant Biotechnol J.* 2022;**20**:886–902
- Zhang A, Zhou H, Jiang X. et al. The draft genome of a flat peach (*Prunus persica* L. cv. '124 Pan') provides insights into its good fruit flavor traits. *Plan Theory.* 2021;**10**:1–17
- Gladysheva-Azgari M, Petrova K, Tsygankova S. et al. A *de novo* genome assembly of cultivated *Prunus persica* cv. 'Sovetskiy'. *PLoS One.* 2022;**17**:e0269284–13
- Cao K, Peng Z, Zhao X. et al. Chromosome-level genome assemblies of four wild peach species provide insights into genome evolution and genetic basis of stress resistance. *BMC Biol.* 2022;**20**:139–17
- Sánchez-Pérez R, Pavan S, Mazzeo R. et al. Mutation of a bHLH transcription factor allowed almond domestication. *Science.* 2019;**364**:1095–8
- D'Amico-Willman KM, Ouma WZ, Meulia T. et al. Whole-genome sequence and methylome profiling of the almond [*Prunus dulcis* (Mill.) D.A. Webb] cultivar 'nonpareil'. *G3 (Bethesda).* 2022;**12**:1–9
- Alioto T, Alexiou KG, Bardil A. 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**:455–72
- Castanera R, De Tomás C, Ruggieri V. et al. A phased genome of the highly heterozygous "Texas" almond uncovers patterns of allele-specific expression linked to heterozygous structural variants. *Hortic Res.* 2024;**11**:uhae106
- Campoy JA, Sun H, Goel M. et al. Gamete binning: chromosome-level and haplotype-resolved genome assembly enabled by high-throughput single-cell sequencing of gamete genomes. *Genome Biol.* 2020;**21**:306–20

34. Shirasawa K, Isuzugawa K, Ikenaga M. et al. The genome sequence of sweet cherry (*Prunus avium*) for use in genomics-assisted breeding. *DNA Res.* 2017;**24**:499–508
35. Wang J, Liu W, Zhu D. et al. Chromosome-scale genome assembly of sweet cherry (*Prunus avium* L.) cv. Tieton obtained using long-read and Hi-C sequencing. *Hortic Res.* 2020;**7**:122–11
36. Zheng T, Li P, Zhuo X. et al. The chromosome-level genome provides insight into the molecular mechanism underlying the tortuous-branch phenotype of *Prunus mume*. *New Phytol.* 2022;**235**:141–56
37. Verde I, Bassil N, Scalabrin S. 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**:1–13
38. Mistry J, Chuguransky S, Williams L. et al. Pfam: the protein families database in 2021. *Nucleic Acids Res.* 2021;**49**:D412–9
39. Paysan-Lafosse T, Blum M, Chuguransky S. et al. InterPro in 2022. *Nucleic Acids Res.* 2023;**51**:D418–27
40. Fleming MB, Miller T, Fu W. et al. Ppe.XapF: high throughput KASP assays to identify fruit response to *Xanthomonas arboricola* pv. *pruni* (Xap) in peach. *PLoS One.* 2022;**17**:e0264543–17