HAL
open science

Prunus genetics and applications after de novo genome sequencing: achievements and prospects.

Maria Jose Aranzana, Véronique Decroocq, Elisabeth Dirlewanger, Iban Eduardo, Zhong Shan Gao, Ksenija Gasic, Amy Iezzoni, Sook Jung, Cameron Peace, Humberto Prieto, et al.

To cite this version:

Maria Jose Aranzana, Véronique Decroocq, Elisabeth Dirlewanger, Iban Eduardo, Zhong Shan Gao, et al.. Prunus genetics and applications after de novo genome sequencing: achievements and prospects.. Horticulture research, 2019, 6, pp.58. 10.1038/s41438-019-0140-8 . hal-02622092

HAL Id: hal-02622092

https://hal.inrae.fr/hal-02622092

Submitted on 26 May 2020

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

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

Distributed under a Creative Commons Attribution 4.0 International License

Aranzana et al. *Horticulture Research* (2019)6:58
https://doi.org/10.1038/s41438-019-0140-8

Horticulture Research
www.nature.com/hortres

REVIEW ARTICLE Open Access

Prunus genetics and applications after de novo genome sequencing: achievements and prospects

Maria José Aranzana[1], Véronique Decroocq[2], Elisabeth Dirlewanger[2], Iban Eduardo[1], Zhong Shan Gao[3], Ksenija Gasic[4], Amy Iezzoni[5], Sook Jung[6], Cameron Peace[6], Humberto Prieto[7], Ryutaro Tao[8], Ignazio Verde[9], Albert G. Abbott[10] and Pere Arús[1]

Abstract

Prior to the availability of whole-genome sequences, our understanding of the structural and functional aspects of *Prunus* tree genomes was limited mostly to molecular genetic mapping of important traits and development of EST resources. With public release of the peach genome and others that followed, significant advances in our knowledge of *Prunus* genomes and the genetic underpinnings of important traits ensued. In this review, we highlight key achievements in *Prunus* genetics and breeding driven by the availability of these whole-genome sequences. Within the structural and evolutionary contexts, we summarize: (1) the current status of *Prunus* whole-genome sequences; (2) preliminary and ongoing work on the sequence structure and diversity of the genomes; (3) the analyses of *Prunus* genome evolution driven by natural and man-made selection; and (4) provide insight into haploblocking genomes as a means to define genome-scale patterns of evolution that can be leveraged for trait selection in pedigree-based *Prunus* tree breeding programs worldwide. Functionally, we summarize recent and ongoing work that leverages whole-genome sequences to identify and characterize genes controlling 22 agronomically important *Prunus* traits. These include phenology, fruit quality, allergens, disease resistance, tree architecture, and self-incompatibility. Translationally, we explore the application of sequence-based marker-assisted breeding technologies and other sequence-guided biotechnological approaches for *Prunus* crop improvement. Finally, we present the current status of publically available *Prunus* genomics and genetics data housed mainly in the Genome Database for Rosaceae (GDR) and its updated functionalities for future bioinformatics-based *Prunus* genetics and genomics inquiry.

Introduction

Stone fruit (peaches, cherries, apricots, plums) and almond belong to the *Prunus* genus, which encompasses approximately 250 species[1] that share a small (250–300 Mbp) and highly syntenic genome[2,3] with eight basic chromosomes. Most of the cultivated species are diploid with two exceptions, the hexaploid European plum (*P. domestica*) and the tetraploid tart cherry (*P. cerasus*). Sexual compatibility is frequent, particularly within members of the same subgenus: *Amygdalus* (peach-almond), *Prunus* (apricot-plum), and *Cerasus* (cherries), although certain crosses between species belonging to *Amygdalus* and *Prunus* subgenera are also possible.

The deployment of basic molecular technology during the past half century advanced our understanding of important aspects of *Prunus* genetics, including the consolidation of its phylogenetic relationships, the analysis of its genetic variability, the development of genetic markers and the construction and comparison of linkage maps, the integration of a large number of major genes and

Correspondence: Pere Arús (pere.arus@irta.cat)
[1]IRTA, Centre de Recerca en Agrigenòmica CSIC-IRTA-UAB-UB, Campus UAB, Edifici CRAG, Cerdanyola del Vallès (Bellaterra), 08193 Barcelona, Spain
[2]UMR 1332 BFP, INRA, University of Bordeaux, A3C and Virology Teams, 33882 Villenave-d'Ornon Cedex, France
Full list of author information is available at the end of the article.

© The Author(s) 2019
Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

quantitative trait loci (QTLs) in these maps, and the identification of markers tightly linked to some traits enabling their selection in plant breeding[4]. Plausible hypotheses about candidate genes for certain characters were established, such as the genes responsible for the melting vs. nonmelting fruit flesh consistency (*M*/*m*) and freestone vs. clingstone flesh adherence to the stone (*F*/*f*)[5,6], the gene for evergrowing (*Evg*/*evg*) for which the recessive homozygote is associated with continuous leaf growth and a failure to enter dormancy[7], and the two genes involved in the self-incompatibility system[8] (Table 1). However, positional cloning of trait-controlling genes was hampered by long intergeneration times, time and space constraints on progeny numbers in mapping crosses, marker availability in specific chromosomal regions, and finally by the lack of efficient transformation methods in most species other than hexaploid plum[9].

The public availability of a peach high-quality DNA sequence in 2010[10] and later those of other *Prunus* species[11–13] unlocked the *Prunus* genome, paving the way to build new knowledge for basic and applied purposes. Our objective in this paper is to review the major achievements that have taken place in the genetics of peach and other *Prunus* crops as a consequence of the availability of whole-genome sequences (WGSs), emphasizing the new genes discovered, and in future targets that will allow us to reach a much deeper understanding of the genetics of a group of species that are important sources of high-quality and healthy nutrients and serve as models for tree species.

The peach genome sequence and other *Prunus* genomes

The plan to sequence the peach genome emerged in the middle of the past decade. Peach was an excellent candidate for shotgun sequencing having a small genome estimated at 265 Mb[14] and extended genomic and genetic resources, such as linkage maps, EST collections, and BAC libraries[4]. The decision to sequence the peach genome was announced by the Joint Genome Institute in 2007 and soon after an Italian consortium (DRUPOMICS) joined the US efforts to form the International Peach Genome Initiative (IPGI) that included scientific institutions from Chile, Spain, and France as well.

The use of whole-genome shotgun sequencing with Sanger chemistry and a completely homozygous genotype, the "Lovell" double haploid PLov2-2n, were crucial for the success of the initiative. Sequencing concluded in late 2009 with the deposition of 3,729,679 paired-end Sanger sequences corresponding to 8.47× genome coverage. Assembly and annotation of genes and repeated sequences were finished in early 2010 and the peach genome sequence was released, under a Fort Lauderdale agreement on April 1, 2010, being the first freely available Rosaceae genome. The assembly[10] was arranged in 2720 contigs and 234 scaffolds covering 227.3 Mb. Scaffolds were assigned to chromosomes using an improved version of the *Prunus* reference map[2,15] to form a chromosome-scale assembly with 218.3 Mb (96%) of sequences in 8 pseudomolecules and 194 unmapped scaffolds (Table 2), providing a sequence accuracy of 99.96% and a completeness of 99%. A combination of de novo identification and structural tools together with similarity searches was used for the annotation of the 84.4 Mb (37.14%) of repetitive sequences. Several homology-based predictors, incorporating transcript assemblies from Rosaceae and *Prunus* ESTs and protein homology with the major sequenced plants, were used to derive 27,852 protein-coding genes and 28,689 protein-coding transcripts. Comparative analyses with other sequenced species and the detection of duplicates within the genome itself highlighted that peach has not undergone recent whole-genome duplication after the ϒ Eudicot hexaploidization[10].

Further analyses of the peach WGS suggested several areas for improvement, including unmapped and randomly oriented scaffolds; correction of putative misassemblies; and better base accuracy, contiguity, and gene prediction. These goals were achieved in the second version of the peach genome (Peach v2.0)[16] using a combination of high-density linkage maps (4 maps and 3576 markers) and high-throughput DNA and RNA sequencing. The improved chromosome-scale assembly is now 58 scaffolds spanning 225.7 Mb (99.2%) arranged in 8 pseudomolecules, 52 of which are oriented (223.3 Mb; 98.2%). Sixty-four-fold Illumina whole-genome sequencing of the reference genotype was used to correct 859 false single-nucleotide polymorphisms (SNPs) and 1347 indels. Moreover, the assembled next-generation sequencing (NGS) contigs enabled the closing of 212 gaps with a 19.2% improvement in the contig L50. Gene annotation was improved using a large amount of RNA-seq data from various peach tissues and organs, as well as annotated repeats that include low copy repeats and a complete set of Helitron transposons (Table 2).

Two other *Prunus* WGSs (Table 2) were obtained using high-throughput NGS Illumina sequencing: (1) Japanese apricot (*P. mume)*[11], producing a fragmented assembly of 237 Mb with 31,390 protein-coding genes and 45% (106.8 Mb) of repeated sequences; and (2) cherry (*P. avium*)[12] with a total length of 272.4Mb, containing 43,349 protein-coding genes and 119.4 Mb (43.8%) of repetitive sequences. Third-generation sequencing methods based on single-molecule sequencing (such as Pacific Biosciences, Oxford Nanopore, and Moleculo) capable of obtaining long reads, up to 50 kb, coupled with short read NGS technologies have provided new inputs to genome sequences. Recently, using a combination of ultra-high coverage (1265.6-fold sequences) Illumina and PacBio long read methodologies, the wild cherry *P. yedoensis*

Table 1 *Prunus* major genes cloned or fine mapped with strong candidate genes

Trait	Origin	Position[a]	Gene description[b]	Strategy	Causal polymorphism
Evergrowing	*P. persica*	G1 (*Evg*)	MADS box transcription factors/*DAM1-DAM6*	Mapping mutants, positional cloning, candidate gene (CG) functional classification and expression analysis[7]	Deletion in the DAM-box gene region of the *evg* mutant
Maturity date/slow ripening	*P. persica*	G4 (*MD*, *Sr*)	NAC transcription factor (*ppa008301m v1.0; Prupe.4G186800 v2.0)*	Fine mapping, whole-genome sequence analysis, and CG functional classification[48]/mapping, CG functional classification and cultivar genotyping[102]	In-frame insertion of 9 bp in the last exon for *MD*/gene deletion for *Sr*
Fruit size	*P. avium*	G2	Cell number regulator (*CNR*) (*PavCNR12)*	CG approach, segregation and association analysis[73]	–
Fruit size	*P. avium; P. persica*	G6	Cell number regulator (*PavCNR20; Prupe.6G240600)*	CG approach, segregation and association analysis[240]/pedigree-based QTL mapping[74]	–
Fruit shape	*P. persica*	G6 (*Sh*)	Leucine-rich repeat receptor-like kinase *(Prupe.6G281100)*	Linkage and association mapping, whole-genome sequencing, CG functional classification and expression analysis[79]	Large deletion affecting the promoter and part of the gene
Flesh color (white/yellow)	*P. persica*	G1 (*Y*)	Carotenoid cleavage dioxygenase (*PpCCD4; ppa006109; Prupe.1G255500*)	CG approach, whole-genome resequence analysis of cultivars and somatic mutants, gene expression and volatile compound analysis[81,82]	Three independent polymorphisms causing non-functional genes: 1 SNP, an SSR, and a transposable element (TE)
Flesh and skin color	*P. persica; P. avium; P. salicina*	G3 (*H*)	MYB transcription factor *(PavMYB10 ; PpMYB10.1)*	QTL analysis and CG functional classification[70,85–87]. Whole-genome sequence analysis[86]	SNPs in *P. avium*;[85] INDEL in *P. persica*[86]
Blood flesh	*P. persica*	G5 (*DBF; Bl*)	NAC transcription factor *(ppa022238m; Prupe.5G006200)*	CG approach, comparative transcriptome and functional analysis[88]	–
Green vs. red leaf color	*P. persica*	G6 (*Gr*)	MYB transcription factor *(PpMYB10.4)*	Transcriptome comparison; transient expression in tobacco and peach leaves[241]	–
Fruit acidity	*P. persica*	G5 (*D*)	Auxin efflux carrier family protein (*ppa006339m; Prupe.5G004300*)	Fine mapping[93], GWAS[28,94], and expression analysis[28,121]	SNP
Melting vs. nonmelting flesh and clingston vs. freestone	*P. persica*	G4 (*M* and *F*)	Endopolygalacturonase *(Prupe.4G261900* and *Prupe.4G262200)*	Fine mapping and CG functional classification[5,6]	SNPs and small indels
Stony hard texture	*P. persica*	G6 (*Hd*)	*PpYUC11-like* (*ppa008176m*)	Fine mapping, GWAS, and CG approach[33,104]	TE
Hairy vs. glabrous skin	*P. persica*	G5 (*G*)	R2R3 MYB transcription factor (*PpeMYB25; ppa023143m; Prupe.5G196100*)	Fine mapping/whole-genome sequence analysis and CG functional classification[167]	TE
Root-knot nematode resistance	*P. cerasifera*	G7 (*Ma*)	Toll/interleukin1 receptor-nucleotide binding site-leucine-rich repeat (*TNL1*)	Fine mapping, positional cloning, and functional validation[131]	–

Table 1 continued

Trait	Origin	Position[a]	Gene description[b]	Strategy	Causal polymorphism
Brachytic dwarf	*P. persica*	G6 (*Dw*)	Giberellic acid (GA) receptor *(PpeGID1c; Prupe.6G332800)*	Sequencing-based mapping strategy, sequence analysis, and functional validation in *P. domestica*[152]	Nonsense mutation
Weeping growth habit	*P. persica*	G3 (*Pl*)	SAM domain gene (*ppa013325; Prupe.3G200700*)	Whole-genome sequencing of mutant and functional validation in *P. domestica*[155–157]	1.8 kb deletion spanning the gene's 5′ end
Broomy (pillar)	*P. persica*	G2 (*Br*)	Tiller angle control (*PpeTAC1*, *ppa010082; Prupe.2G194000*)	Sequencing-based mapping strategy, whole-genome sequence analysis, expression analysis, and CG homology with Arabidopsis mutant[156]	Insertion producing a nonsense mutation
Gametophytic self-incompatibility (pistil)	*Various Prunus*	G6 (*S*)	S-Ribonuclease (*S-RNAse*)	Allele cloning, genotyping, and self-compatibility screening[160,161]	–
Gametophytic self-incompatibility (pollen)	*Various Prunus*	G6 (*S*)	F-box protein (*SFB*)	Allele cloning, genotyping, and self-compatibility screening[165,166]	–
Gametophytic self-incompatibility (pollen)	*P. armeniaca*	G3 (*M*)	Disulfide bond A-like oxidoreductase (*ParMDO*)	Fine mapping, gene expression analysis, whole-genome resequence analysis[163]	Miniature inverted-repeat transposable element (MITE) insertion
Double flower	*P. persica*	G6 (*Di2*)	EU-AP2 transcription factor, TOE-type (*Prupe.6G242400*)	Fine mapping, whole-genome resequence analysis, and CG functional classification and expression analysis[242]	Deletion of an miRNA target site

GWAS genome-wide association study, *QTL* quantitative trait locus, *SNP* single-nucleotide polymorphism, *SSR* simple sequence repeat
[a]In parenthesis, major gene name
[b]Genes starting with *ppa* and *Prupe* correspond to the annotation of the v1.0 and v2.0 versions of the peach genome, respectively; *Pav* are genes from the sweet cherry genome v1.0

Table 2 Comparison of the assembly and statistics among the *Prunus* genomes so far published

	Peach v2.0	Peach v1.0	*P. mume*	PAV_r1.0 (*P. avium*)	*P. yedoensis*
Sequencing strategy	Sanger–NGS Illumina	Sanger	NGS Illumina	NGS Illumina	NGS Illumina–PacBio
Genome coverage	8.47×–NGS 64×	Sanger 8.47×	180×	130.34×	1265.6× (1192× + 73×)
Scaffold number total	191 (241)[a]	202 (234)	1449	10,148	3185
Contig number total	2525	2730	10,894	–	4292
Scaffold sequence total	227.4 Mb	227.3	237.1 Mb	272.4 Mb	323.8 Mb
Mapped scaffold sequence total	225.7 Mb (99.2%)	218.4 Mb (96%)	198.9 (83.9%)	191.7 Mb (70.4%)	281.6 Mb (87.0%)
Oriented scaffold sequence	223.3 Mb (98.2%)	194.6 Mb (85.6%)	–	–	–
Contig sequence total	224.6 Mb	224.6 Mb	219.9 Mb	–	318.7 Mb
Scaffold N50/L50	10/7.3 Mb	9/8.9 Mb	120/577.8 kb	–/219.6 kb	519/198.9 kb
Scaffold N50/L50 (chromosome-scale assembly)	4/27.4 Mb	4/26.8 Mb	–	–	–
Contig N50/L50	250/255.4 kb	294/214.2 kb	2009/31.8 kb	–/276 bp	706/132.6 kb
Number of scaffolds >50 kb	11	21	–	–	–
% main genome in scaffolds >50 kb	99.4%	99.4	–	–	–
Repetitive sequences	79.38 Mb (35.34%)	84.41 Mb (37.14%)	106.8 Mb (44.9%)	119.4 Mb (43.8%)	150.8 Mb (47.2%)
Protein-coding genes	26,873	27,852	31,390	43,673	41,294
Protein-coding transcripts	47,089	28,689	–	–	45,581
Alternative transcripts	20,216	837	–	–	4287

[a]In brackets, assembled scaffolds prior the build of pseudomolecules

genome was sequenced[13], producing an assembly of 323.8 Mb with 41,294 protein-coding genes (Table 2). The sequence of "Texas" almond (P. Arús, unpublished results) is already publicly available at the Genome Database for Rosaceae (GDR).

Other de novo assemblies of *Prunus* genomes are underway, such as that of another sweet cherry cultivar, "Regina" (E. Dirlewanger, unpublished results), while those for other *Prunus* crops, plum, apricot, and other wild crop relatives are expected to be released in the next several years. These new sequences provide an opportunity to unravel genetic diversity—still present in *Prunus* wild species—that had been lost in these crops during thousands of years of selection and breeding. This genomic information could be exploited by introgressing new desirable alleles into the cultivated species via classical or marker-assisted breeding (MAB) methods. It could also be the starting point for applying gene editing and cisgenesis to *Prunus* crops in the efforts to create new targeted variability in genes controlling relevant agronomic traits, such as biotic and abiotic stress resistance, for which the cultivated species, peach in particular, are notably lacking.

Soon after the release of the peach genome sequence, the International Peach SNP consortium developed the first peach 9K SNP array v1[17]. Furthermore, the extensive *Prunus* synteny[2] made the peach genome a valuable reference to identify SNPs in related crops[10,18] enabling the construction of a cherry 6K SNP array[19]. Recently the peach and cherry arrays have been updated with 9000 additional SNPs covering previous gaps (I. Verde, unpublished results). In addition to SNP arrays, resequencing data from sets of individuals has enabled unprecedented in-depth analyses of variability and linkage disequilibrium (LD)-based analysis that resulted in the first genome-wide association studies (GWASs). Such applications of the *Prunus* WGSs are reviewed below.

Whole-genome analysis of *Prunus* diversity

Prunus variability, domestication, and crop evolution

The most comprehensive *P. persica* diversity analysis was performed with the 9K peach SNP array[17] and a collection of 1580 accessions (including peaches and few closely related *Prunus* species)[20]. SNP data identified some unknown somatic mutants and confirmed others that were already reported. In agreement with the known breeding history of peach and previous results with simple sequence repeats (SSRs)[21,22], the unique accessions showed a relatively high kinship coefficient corresponding to an average relationship of grandparent–grandchild or half-sibs and were divided into three main subpopulations: two based on geographic distribution (oriental and occidental accessions) and the third, a separation within occidental materials as traditional landraces or those derived from breeding programs. The cherry 6K SNP array[19] was used to characterize 210 cherry accessions. As for peach, variability in sweet cherry was differentially distributed between cultivars and landraces and, in the latter, among geographic locations[23].

Whole-genome resequencing of germplasm collections has also been used for variability analysis, particularly in peach, including cultivars with variability underrepresented in the cultivar panel used for the development of the 9K SNP array (such as those from Asian germplasm collections) and *Prunus* wild species. In addition, such high level of genome definition has provided relevant knowledge on domestication and crop evolution processes. Most *Prunus* cultivated species originated in central Asia, although their recent evolution seems to have followed substantially different patterns. For example, peach and almond, two of the genetically closest relatives, seem to have arisen from a common ancestor in central Asia by the time of the uplift of the Central Asian Massif around 8 MYA[24,25], and then followed completely different paths, almond in the dry environments of the central and western Asian steppes and peach in the humid, subtropical climate of southwestern China[24]. Important trait differences accumulated such as dry vs. fleshy fruit and outcrossing vs. selfing, with many derived biological and populational consequences[25]. Recent comparison of genome-wide diversity patterns in almond and peach has identified regions with signatures of selection that occurred during domestication that interestingly often overlap in both species despite having occurred independently in each around 5000 years ago[26]. Likewise, during *P. persica* domestication, which probably took place through a single domestication event from wild peaches[27], some genes were subjected to artificial selection as deduced from comparisons of the regions showing selective sweeps among domesticated peaches (including landraces, edible and ornamental cultivars) and their wild relatives[27–29]. Such regions contain genes that may have contributed to peach domestication and cultivar differentiation and warrant further research. Although interspecific hybridization events, such as those accompanying creation of some *Prunus* crops like Japanese plum[30], might be associated with increased mutation rates, low evolutionarily rates reported for perennial plants, peach in particular[31], may explain the close relationships among *Prunus* species.

Although SSRs and SNPs are and will remain the principal genotyping tools in *Prunus* species lacking a reference genome, the continuous cost reduction of sequencing techniques forecasts the routine use of resequencing data in future *Prunus* variability studies. Currently, increased sequencing efforts are feeding the databases and providing important sequence resources for diversity and domestication analyses. In a near future, we

may expect an increasing number of studies focused on the identification of particular genomic regions, such as those showing differential patterns of diversity as a consequence of domestication or breeding, gene families, or DNA features such as transposable elements (TEs). Such studies may answer fundamental questions concerning when, where, and how domestication events took place in the development of certain *Prunus* species. However, although promising, relying on available sequence data has a major drawback since most has been done with resequencing techniques that yield short reads and with low depth. Their alignment against the "Lovell" reference genome may fail to map genetic regions that have been lost or duplicated during the domestication process. The recent development of new long-range sequencing technologies should resolve this problem.

Whole-genome diversity has been exploited in GWAS using both SNPs and whole-genome resequencing data with a minor or major depth coverage. Understanding the extension and decay of LD in the species of interest, and in particular in the panel of genotypes used for association analysis, is essential when considering GWAS. WGS data has identified fast LD decay in apricot, spanning <100 bp[18]. Similarly, LD decays fast in ornamental *P. mume* accessions ($r^2 \leq 0.2$ at 50 kb to few hundreds of base-pairs depending on the population)[32] and a little more moderately in cherry ($r^2 \leq 0.2$ at 100 kb)[23]. In contrast, large LD extensions were observed in peach ($r^2 \leq 0.2$ between 0.8 and 1.4 Mb depending on the population[20]). These findings have practical implications: while a large number of SNPs will be required to find some associated with the allele of interest in apricot, lower SNP density should be sufficient in peach. However, as a trade-off, associated SNPs in the former will be most likely closer to the gene than in peach. As a proof of concept, GWAS for several qualitative traits was performed in peach[20], identifying haplotypes linked to the loci or genes already known to be responsible for the trait. GWAS using resequence data from 129 peach accessions[28] identified association signals at various loci for a set of 12 agronomic characters and proposed candidate genes for two of them. Recently, the stony hard flesh (*Sh*) locus was mapped in a panel of peach genotypes using the 9K SNP array[33]. A highly associated SNP was found that, due to the extended LD of peach at this chromosomal region consequence of its location close to the centromere, was organized in a haplotype about 1.9-Mb long. In this case, higher SNP density would not have translated to higher precision.

Up to now, GWAS has mainly been performed for agronomic and fruit traits evaluated in a single location, usually with small-scale phenotyping techniques. In the near future, we may expect to see GWAS of genetic traits related to crop adaptation, probably involved in epigenetic variability and genome instability, as already evident in some large-scale *Arabidopsis* studies[34,35]. In addition, thanks to the current inclusion of new technologies in the life sciences field, we may expect that new high-throughput phenotyping methods such as those based on three-dimensional imaging[36,37] or large-size plant mobile screening platforms[38] will soon provide enormous amounts of data as substrate for future GWAS.

Disadvantages of the multi-year intergenerational period of *Prunus* species, the seasonality of crop production, and large plant sizes are balanced by the ability to clonally propagate and maintain heterozygous genotypes for many years. This enables evaluation of the same genotype under different climatic and management conditions. Future GWAS strategies should consider the analysis of existing genotype collections replicated in different locations to incorporate environmental effects in the analysis.

Haploblocking *Prunus* genomes

The use of high-density SNP and resequencing data allows high-resolution genome analysis such as "haploblocking", or haplotype blocking, a means of dividing the genome of *Prunus* crop individuals into informative genetic segments. Each segment (called a haploblock) has variants (called haplotypes) within and across individuals. Alleles of the loci within each haplotype tend to be co-inherited owing to their close linkage. Functional utility can be assigned to haplotypes, such as trait effects and ancestral origin. The managers of haploblocked genomes are geneticists and breeders, and so the patterns of co-inheritance described by haploblocking are designed to capture features of value to such users (Fig. 1). Haploblocking provides a fascinating overview of the genetics of any individual examined.

At the broadest level, whole chromosomes are haploblocks, described by their genetic length in centiMorgans (cM), but a finer level of compartmentalization is possible and desirable. In practice, haploblocking divides up chromosomes, establishing systematic criteria for helpfully delimiting each haplotyped region. The series of haploblocks along each chromosome act as an easily manageable number of linked multi-allelic loci, carrying all the information of the underlying genotypic data[39]. A requirement for haploblocking is phased, high-resolution genotypic data, which is currently obtained by genotyping individuals with high-density SNP assays and correct phasing for each individual that is achieved, for example, via inheritance analyses using the software FlexQTL™[40].

There are several ways that *Prunus* crop haploblocks have been or are being delimited. The pedigree-based approach identifies positions of historical recombination in the pedigreed germplasm and uses these positions to delimit haploblocks. In U.S. breeding germplasm of sweet cherry[41] and peach (C. Da Silva Linge, pers. comm.), recombination positions used to define haploblock

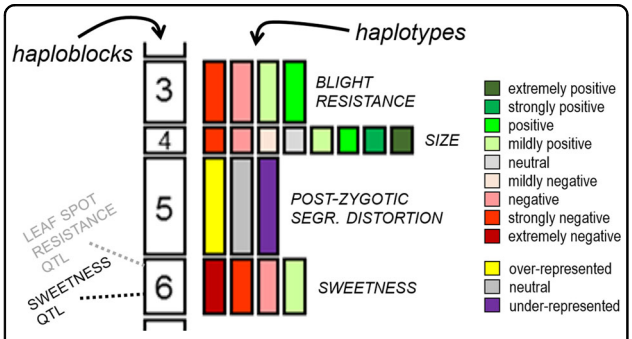

Fig. 1 Dividing a crop's genome into qualitative, cognitively manageable segments by haploblocking adjacent sets of loci can be done in several ways, such as the pedigree-based approach, in which loci within haploblocks have not recombined throughout the pedigree of known progenitors of cultivars. Haplotypes are the variants of haploblocks—sets of co-inherited alleles. To each haplotype can be assigned trait influences, ancestry, and other genetic features. If a breeding parent does not have coupling-phase linkage for desirable alleles within a haploblock containing multiple quantitative trait loci, such tight linkage might be targeted for recombination in the next generation

boundaries were all of those detected in the known progenitor generations leading to cultivars of the germplasm set. Recombination events were positioned with VisualFlexQTL[40], and finally haplotypes were assigned to each haploblock using PediHaplotyper[39]. As germplasm managers using pedigree-based haploblock information examine haplotypes through ancestral generations, they do not have to track complicating recombinations within haplotypes—there is 100% LD within haploblocks. Note that this coupling-phase linkage within a haploblock can be broken into subsequent generations, with a probability determined by the genetic distance between loci. Other ways to establish haploblocks include using regular, arbitrary lengths for each haploblock, such as 1 or 5 cM; the diversity-based approach, whereby the observed degree of LD along each chromosome at increasing levels of genetic diversity among sets of individuals, from families to across species and even across genera, are used to define segments that have rarely recombined (A. Lawton Rauh, pers. comm.); and the aggregate-marker approach, which combines into haploblocks sets of markers positioned within very short physical distances, enabled by purposeful SNP clustering around focal points as done for design of the 20K apple SNP array[42] and the upcoming 9 + 9K peach and 6 + 9K cherry SNP arrays. WGSs are a critical foundation of haploblocking, as they determine the genomic positioning of SNP markers to ensure that each chromosome is covered with genetic polymorphism data. Although haploblocks are described in cM, physical anchoring to the WGS ensures ready connection to associated DNA sequences, sequence motifs, and annotated genes. Research community consensus on the positions and labels of a standardized set of haploblocks for each crop should provide a framework for comparative genetics that is easier to use than a reference genetic map of common markers. A proposed basis for such a haploblock framework is the pedigree-based haploblocks derived from the set of 50–100 ancestors that are common to most cultivars of a crop.

Haplotypic variation across haploblocks is powerful genetics information. Haplotypes are genotypically distinguished by differences in their set of alleles (ultimately DNA sequence variation), such as ABBAABB vs. ABABBBB vs. BBBBBBB. These haplotypes can also be functionally distinguished. Where major genes or QTLs for traits of interest lie within (or span) a haploblock, the associated effects of each haplotype on such traits can be calculated and the information attached. Each haplotype can also be traced through the pedigree from the earliest known ancestor that contributed the haplotype through to all descendants carrying that identical-by-descent haplotype. Visualizing the "haplotype mosaics" of this shared ancestry[43] can provide a valuable experience of elite genomes (section "Applications of genome information in *Prunus* improvement: MAB, MAI, genomic selection, and visualization of the genome-wide genetics of elite individuals"). Identity-by-state among germplasm of one or more haplotypes extending over large genetic distances can be used to infer recent shared common ancestry. Other features relevant to breeding and biodiversity management can also be assigned, such as haplotypes that are overrepresented or under-represented in offspring that thereby contribute to segregation distortions, excessive heterozygosity, and excessive homozygosity. Examining the ancestral origins and germplasm distribution of trait locus alleles or segregation-distorting alleles is a relatively simple task using haploblocks.

Haploblocking is a newly available approach to make sense of the vast quantity of genome-wide genotypic data that can now be readily obtained on germplasm individuals. Critically, it involves the consideration of each individual in genetic comparisons, examining the haplotypes that each carries, rather than simply using scaled-up quantitative analyses. New software in data curation and graphical genotyping is needed to facilitate a multitude of practical applications that beckon in breeding and biodiversity management.

Genetics of agronomic characters

During the recent decades, many efforts have been directed to identify loci and the underlying genes that explain the phenotypic variability of a number of agronomic traits. Thanks to the availability of reference genomes, once the chromosomal position for a major gene or QTL is established, it is possible to easily discover and develop additional markers that enable reducing the size

of the target region, where a small set of genes can be identified. This approach, supported by many RNA-seq datasets available for most *Prunus* species[10] and epigenetic analysis, has led to the discovery of several causal genes for certain traits and of strong candidates for others in peach and other *Prunus* (Table 1) that are reviewed below.

Phenology (blooming and maturity dates) and climate change adaptation

In *Prunus* species, as for woody perennial plants in temperate and boreal zones, survival and production depend on precise timing of growth and rest periods in synchrony with seasonal changes in temperature[44]. Flower bud differentiation takes place at the end of the summer and buds continue their development until mid-autumn, when they enter into the dormant stage[45]. The first step of dormancy is growth cessation and meristem acclimatization to cold. Dormancy can be separated into two main phases: (1) endodormancy, when meristems are unable to initiate growth under favorable conditions, followed by (2) ecodormancy, when meristems can resume growth if temperatures are optimal[46]. During endodormancy, a certain amount of cold temperatures, defined as chilling requirement (CR), must accumulate prior to the bud being released from endodormancy. Subsequently, the bud will respond to heat, leading to flowering when heat requirements (HR) are satisfied[46]. Consequently, flowering date is determined by CR and HR, and fruit/nut maturity timing will be associated with flowering date and to the length of the fruit development period.

In the context of global climate change, flowering phenology of deciduous tree species is crucial as it may affect commercial productivity. In fruit tree orchards, flowering phenology has an indirect influence on spring frost damage, pollination, maturity, and fruit production. Climate warming delays the endodormancy release while it accelerates bud growth. Therefore, depending on the balance between these two antagonistic effects, budbreak can be either advanced or delayed and can even be compromised owing to insufficient cold temperature during winter. Additionally, warm temperatures in spring and summer are responsible for an advance of the maturity date. Consequently, it is crucial to anticipate the impact of climate warming on phenological processes with the aim to establish prospective strategies to create new cultivars that are well adapted to future climate. Hence, numerous studies are aimed at deciphering the molecular mechanisms underlying the regulation of bud dormancy and flowering and maturity dates.

Many *Prunus* major genes and QTLs have been described for phenology-related traits such as CR, HR, and flowering and maturity dates. For bloom date, major QTLs were detected on linkage group 6 (G6) for peach and on G4 for apricot and sweet cherry, whereas for maturity date, a major QTL was detected on G4 for all three crops[47]. Fine mapping of the major locus controlling the maturity date in peach identified a precise QTL interval of 220 kb and enabled the identification of a NAC transcription factor as the candidate gene[48] (Table 1) confirming results previously described[47]. With a multi-progeny mapping strategy, 18 peach progenies in total, additional QTLs were detected for flowering date (G4, G6), maturity date (G2, G4, G6), and for fruit development period (G4, G6)[49]. In sweet cherry, one stable QTL for CR and bloom date was detected in the same region on G4[50]. Candidate genes underlying this QTL were investigated using the peach genome sequence and key genes were identified for these two traits[51]. Most of the genes were those involved in chromatin remodeling (*ARP4*, *EMF2*, *PIE1*) and in gibberellin homeostasis (*KS* and *GA2ox*). Thanks to the high synteny existing among *Prunus* species[2], it was possible to use the peach and the sweet cherry genomes to compare homologous regions such as the QTL on G4 controlling CR in peach, sweet cherry, and Japanese apricot[52].

Expression of the dormancy-associated MADS-box (*DAM*) genes has been extensively analyzed and their key role in the endodormancy establishment and maintenance was demonstrated. *DAM5* and *DAM6* genes are upregulated during growth cessation and downregulated by cold exposure during winter in peach[53], Chinese cherry[54], sweet cherry[55], and Japanese apricot[52,56,57]. In peach, histone modifications of *DAM* gene expression were reported[58]. In sweet cherry, DNA methylations and small interfering RNAs involved in the silencing of *PavMADS1* during cold accumulation and dormancy release were analyzed[59], and research on the epigenetic regulation of the *DAMs*, through chromatin immunoprecipitated-sequencing (ChIP-seq) analyses, is in progress[55].

New methods that resolve phase-switches and reconstruct contig-length phase blocks[60] make it possible to identify the haplotypes for each individual, which are especially useful for highly heterozygous genotypes. This ability is important for phenological traits, which are often complex, highly polymorphic, and may involve epigenetic regulation. Genomic selection can be a useful breeding strategy for the phenological traits and promising results were obtained for apricot[61] and sweet cherry[62] for flowering and maturity dates. In addition to these tools, spontaneous mutants are highly valued for identifying genes responsible for phenotypic variation. Many spontaneous (epi-)mutations were found in *Prunus* species[63] and several of them were associated with flowering date. In sweet cherry, the very early-flowering cultivar "Cristobalina" was issued from a spontaneous mutation of "Temprana de Sot"[64]. In apricot, early-flowering "Rojo Pasión Precoz" was derived from the recently released

"Rojo Pasión" apricot cultivar[65]. To date, no clues about the determinism of these extreme phenotypes are available. Using reference WGSs and resequencing data of these mutants or of cultivars with contrasting phenotypes, it should be possible to identify the causal genes.

Fruit-related characters

Fruit size, shape, and color

Fruit appearance highly influences consumer's choice. Therefore, large fruits, with a certain fruit color and shape are breeder's priorities. Fruit size is one of the most important phenotypic traits that distinguishes modern *Prunus* cultivars from their small-fruited wild ancestors. For example, in sweet cherry, wild (syn. mazzard), landrace, and modern bred cultivars typically have fruit sizes of ~2, 6, and 12 g, respectively[66]. Discovering major loci underlying genes that determine fruit size has been a high priority due to the importance of fruit size to growers' profitability and the need for breeders to select against small-fruited alleles when introgressing traits from wild *Prunus* species.

Fruit size is a classic quantitative trait controlled by many loci, highly influenced by the environment, with >50 QTLs identified to date. Many fruit size studies have involved crossing large-fruited modern cultivars with *Prunus* wild relatives, which in peach consisted of crosses with *P. davidiana*, an ornamental small-fruited *P. persica* and almond *P. dulcis*[49,67–69]. A comparison of the nucleotide diversity in genome sequences of these wild species (*P. kansuensis* and *P. davidiana*) with that of cultivated peach identified regions associated with selective sweeps related to domestication and breeding[10]. In sweet cherry, a fruit size QTL region on chromosome 2 was shown to be under positive selection[41]. In certain cases, QTLs co-locate across species such as the G7 QTL discovered in both peach[68] and Japanese plum[70]. In other cases, the co-location of fruit size genes with fruit firmness, such as the QTL on G5 in cherry[71], raises the question of whether this is a pleiotropic locus. However, it will take the discovery of the underlying genes to know whether there are one or more genes responsible for these QTLs.

To date, members of two gene families have been associated with the fruit size increase that accompanied domestication. The first study involved the *FW2.2* cell number regulator (CNR) gene family originally identified in tomato[72]. A total of 23 *FW2.2/CNR* family members were identified in the peach genome, two of these *CNRs* co-located with fruit size QTLs on G2 and G6 in a domesticated sweet cherry × mazzard cross[73] (Table 1). The gene family member on G2, named *PavCNR12*, was hypothesized to contribute to an increase in fruit size by increasing mesocarp cell number. The gene family member on G6, *PpCNR20*, is also a candidate gene for a fruit weight QTL identified in peach[74]. Another gene implicated in the control of fruit size in sweet cherry is a member of the Cytochrome P450 (CYP) subfamily, termed *CYP78A* that has been shown to be involved in plant organ growth and development including tomato fruit[75]. The transcript level for *PaCYP78A9* was significantly higher in a cherry landrace compared to a mazzard cherry[76]. In addition, silencing of *PaCYP78A9* during sweet cherry fruit development caused a reduction in fruit size due to a reduction in both mesocarp cell number and size. To date, this gene has not been shown to be associated with a fruit size QTL.

Prunus fruits show large intraspecific variability for fruit and stone shape, which can be used for germplasm characterization. Similar to fruit size, cell number may determine fruit shape. This is the case of the flat shape of peach fruits, which is primarily determined by the regulation of cell production in the vertical direction during early fruit development[77]. The flat shape locus *S/s* has been mapped on G6, where *S* is a partly dominant allele; *ss* genotypes produce round fruits, *Ss* genotypes produce flat fruits and fruits from *SS* trees abort a few weeks after fruit set[78]. The participation of a second gene producing fruit abortion has been hypothesized, although pistils of *SS* flowers show already a differential phenotype compared to *Ss* and *ss* types (Fig. 2). Association analysis of SNPs and indels in the *S* locus suggests the gene *Prupe.6G281100*, a leucine-rich receptor-like kinase (LRR-RLK), as a candidate for the fruit shape (Table 1). The best protein hit of this gene occurred with *AtRLP12*, which in Arabidopsis functionally complements *CLAVATA2*, a key regulator that controls the stem cell population size[79]. The flat-associated allele of *Prupe.6G281100* is not expressed, resembling a dosage mechanism. However, polymorphisms in this gene do not completely explain the trait in all germplasm, suggesting the participation of additional alleles or genes[77,79]. Apart from the peach flat shape, investigation of genes determining diverse shapes (i.e., oblate or elliptic) is still lacking.

The inheritance of white vs. yellow flesh in peach, one of the first Mendelian traits described in *Prunus* (*Y/y*), maps to G1[80] with white flesh dominant over yellow flesh. The pigments abundant in the yellow peaches are carotenoids, for which the biosynthetic pathway is well known. Subsequently, the gene underlying the *Y* locus was shown to be a carotenoid cleavage dioxygenase 4 (*PpCCD4*) that is considered responsible for carotenoid degradation in white-fleshed peaches[81]. The dominant functional allele was found to confer white flesh while yellow peaches have loss-of-function mutations of this gene[82] (Table 1).

Anthocyanin pigments in *Prunus* are responsible for the red fruit skin and flesh color including the red "blush" on peach fruit. Their biosynthetic pathway is well studied in many crops including apple where a R2R3 MYB

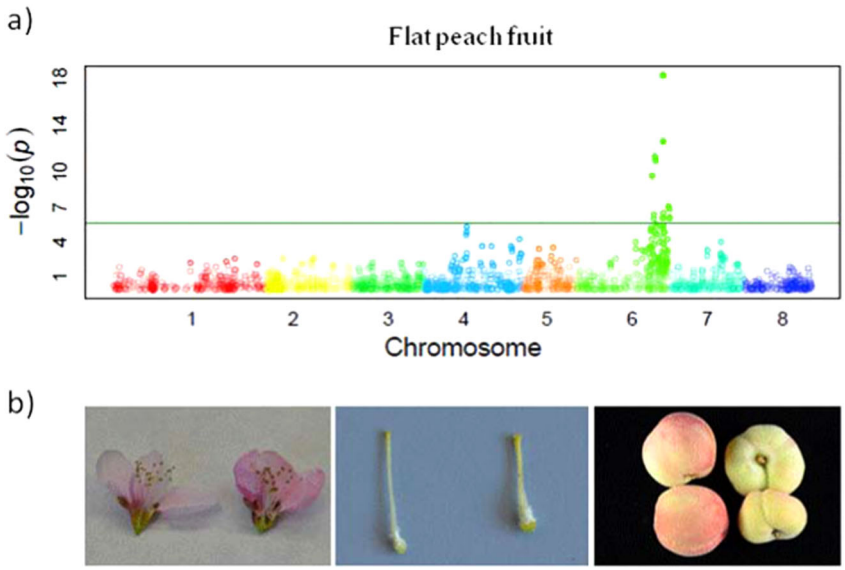

Fig. 2 Flat fruit shape in peach. a Manhattan plot from Micheletti et al.[23] data. Chromosomes are marked with different colors on the horizontal axis. The horizontal green line represents the significance threshold for the association. **b** Images of flowers, pistils, and fruits of a round (left) and a flat (right) fruit cultivar where it can be seen that the flat vs. round character is determined early in flower formation

anthocyanin transcription factor, *MdMYB10*, controls red fruit flesh color[83,84]. Likewise, in *Prunus*, there is strong evidence that this same anthocyanin transcription factor, *MYB10*, on G3 is the major gene responsible for the red color differences (Table 1). In sweet cherry, mahogany fruit (skin and flesh) color is dominant to yellow fruit color and the major QTL on G3 maps to an interval containing *PavMYB10*[85]. This same G3 interval was shown to have major genes/QTLs controlling anther color in almond × peach progenies[69], red blush on the skin in peach[86,87], and skin color in Japanese plum[70]. However, other QTLs controlling the percentage of red skin color in peach were identified on G4, G5, and G6[49]. Some of these QTLs have been shown to exhibit epistatic interactions[85], and this has been extended to interactions among the involved genes. For example, the red-flesh peach type, called "blood type," was found to be controlled by multiple loci with a candidate gene identified on chromosome 5 that regulates the transcription of *PpMYB10*[88] (Table 1).

Owing to an excellent understanding of pigment biosynthesis pathways and the high heritability for color, the molecular characterization of fruit color is the most advanced. Yet, knowledge on the molecular basis of the more subtle differences in color will require the identification of the full set of genes and an exploration of epistatic interactions. For example, a transcriptomic analysis of fruit ripening in plum revealed the co-expression of the MYB transcription factor with another important anthocyanin transcription factor, a basic helix-loop-helix, bHLH[89].

The second trait for which the identification of likely candidate genes is most advanced is mesocarp cell number where the comparisons were made among representatives of wild and domesticated germplasm. Characterizing the alleles for candidate genes is an important next step made possible by the increasing amount of sequence data from a broader range of plant materials. This enables proof-of-function experiments such as those described for the sweet cherry fruit size gene *PaCYP78A9*[76]. These types of experiments are also critical for addressing hypotheses of pleiotropy that have been suggested for a few co-locating QTLs. In addition to narrowing QTL regions to reduce the number of candidate genes, fine mapping studies are needed to dissect the trait-rich QTL regions. Obvious first targets are the trait-rich clusters on chromosome 2 in sweet cherry[41] and chromosome 4 in multiple *Prunus* species[70,90,91] (A. Iezzoni, unpublished results).

Fruit flavor and flesh consistency

Apart from the external appearance, sensorial attributes (like flavor and flesh consistency) are highly relevant for fruit quality. *Prunus* fruit flavor is dependent on the concentrations and balance of sugars and acids. Elucidating the genetic control of these metabolites is difficult due to the metabolic complexity, changes in individual sugar and acid levels during fruit maturation, and the environment. Numerous QTLs have been identified on all eight *Prunus* chromosomes for traits and metabolites associated with sweetness (e.g., soluble solid content (SSC), individual sugars: fructose, glucose, sucrose and sorbitol), and acidity (e.g., pH, titratable acidity, and malic acid). In peach, one QTL mapped to the top of G5

behaves as a major locus, *D*/*d*, determining low acid (subacid) content[92–94]. In general, fruit can be placed into two groups, normal acid and low acid, with low acidity determined by the dominant allele. The physical map location of the *D* locus has been narrowed to a 100-kb region[93] and a candidate gene coding for an auxin afflux carrier family protein (*ppa006339m; Prupe.5G004300*) was proposed based on GWAS[28] (Table 1). For the components of sweetness, many QTLs have been identified most with minor effects. For example, six QTLs for SSC were found in a collection of 18 populations of peach[49], and numerous QTLs for sugar metabolism during peach fruit development along with co-localized candidate genes have been reported[95]. Yet, the determination of the underlying genes and their metabolic functions is in its infancy. Global mRNA and protein profiling results can provide support for candidate genes and identify key metabolic switches and co-expression gene networks[96] controlling the metabolism and accumulation of compounds contributing to fruit flavor. This approach may be particularly useful for sweetness, as the vast majority of the QTLs identified have minor effects. It was successfully used to characterize differences in sugar metabolism pathways between two Japanese plum cultivars, the climacteric "Santa Rosa" and its non-climacteric bud sport mutant[97], resulting in the identification of several highly connected sugar metabolism-associated genes that could be acting as metabolic hubs. These highly connected genes could be targets for future study.

Fruit flesh texture is one of the key traits in most peach breeding programs. On one side, it determines two different fruit typologies: melting flesh (MF) for fresh consumption, and non-melting (NMF), mainly for canning. On the other side, flesh texture also determines postharvest behavior, with other fruit texture types as slow melting (SMF) and stony hard (SH), both associated with longer shelf life. The *M* and *F* genes determining fruit texture (MF/NMF) and flesh adhesion to the stone (clingstone vs. freestone), respectively, were identified very early as endopolygalacturonase (endoPG) genes[5,6] (Table 1). They map to a single locus on G4, where two endoPG genes, *Prupe.4G261900* and *Prupe.4G262200*, are involved in their inheritance. The different haplotypes of these two loci and the phenotypes that they determine have been elucidated[6,98]. Using this information, different molecular markers sets have been developed[99] (I. Eduardo unpublished results) to genotype this locus and to identify the alleles that determine the different flesh consistency types, including freestone melting flesh, clingstone melting flesh, clingstone non-melting flesh, and clingstone non-softening flesh.

The SMF phenotype is defined by a slower process of postharvest fruit-softening than the MF types, with "Big Top" being the reference cultivar for the trait[100]. Using high-density genetic maps of two populations with "Big Top" as one of the parents, two QTLs involved in the trait were identified in G4 and G5, both affecting fruit-softening rate and maturity date[101]. The QTL on G4 co-localizes with a gene determining the slow ripening trait, a recessive mutation that determines the production of fruit that never ripe[102], and with the major QTL determining the maturity date described in section "Phenology (blooming and maturity dates) and climate change adaptation." The *Prupe.4G186800* gene, coding for a NAC transcription factor, was proposed as a strong candidate for the causal gene of these traits[48]. The second QTL, mapping on G5, was detected only in "Big Top" and also co-locates with a QTL for maturity date, overlapping with a region where other NAC transcription factors are located.

Another trait related with peach flesh texture is SH, which presents a crispy flesh when fully ripe and is known to be controlled by a single gene (*Hd*/*hd*). Based on digital gene expression analysis, a YUCCA flavin monooxygenase gene (*PpYUC11-like*, *ppa008176m*) located on G6 was proposed as a candidate gene for *Hd* (Table 1), and an SSR in the intron of this gene was identified as a possible diagnostic marker[103]. Later on, this gene and this marker were confirmed using GWAS[33] and the insertion of a transposon-like sequence was detected upstream of the PpYUC11 gene in the *hd* allele, suggesting its causal effect for the SH phenotype[104].

To have a complete picture of the genetic control of peach texture, it is necessary to: (1) identify the alleles of the genes involved in this trait to better understand how these genes interact with each other and with other important genes, such as the ones determining maturity date; and (2) understand how these gene interaction networks are influenced by the environment. Achieving all these goals relies on improved higher-throughput phenotyping methods that increase the power to discriminate the different texture types and the number of samples analyzed. This will require appropriate populations or germplasm collections to be evaluated and the development, where necessary, of plant resources specifically bred to facilitate phenotypic analysis of this important character.

Fruit allergens

Fruit allergy is becoming a significant health issue. There have been frequent reports of peach fruit allergy globally, especially in China and the Mediterranean countries where peach consumption is high and *Artemisia* pollen allergy is prevalent[105]. Peach-related allergic symptoms include the oral allergy syndrome, urticaria, gastrointestinal symptoms, and anaphylaxis. Sensitization to peach lipid transfer protein allergens also results in broad cross-reactivity to other fruits and plant foods[106]. Five allergens in peach have been identified: Pru p 1

(pathogenesis-related protein, PR-10), Pru p 2 (thaumatin-like protein, PR-5), Pru p 3 (nonspecific lipid transfer protein, PR-14), Pru p 4 (profilin) and Pru p 7 (peamaclein, Gibberellin-regulated protein)[107,108]. With the availability of the peach WGS, all putative peach allergen genes have been located precisely[107]: a cluster of six *Pru p 1* genes, *Pru p 2.04*, *Pru p 4.01*, and *Pru p 7*, are located on peach chromosome 1; *Pru p 2.01* on chromosome 3; three members of *Pru p 3* in a cluster on chromosome 6; *Pru p 4.02* and *Pru p 2.02* on chromosome 7; and *Pru p 2.03* on chromosome 8[109]. Gene expression studies of these genes in fruit, leaf, and anther tissues showed differential patterns and fruit specific genes that have been further investigated under different light-shading treatments[109,110], demonstrating their variable expression in response to light.

Pru p 1 allergen is cross reactive to birch pollen allergen Bet v 1 in northern of China and Europe, its sensitization resulting in mild oral allergy symptoms. Two isoforms (Pru p 1.01, Pru p 1.02) are expressed in peach fruit and have similar IgE-binding properties[111]. The Pru p 3 allergen can cause a severe food allergic reaction, while different peach fruit types (peach, nectarine) have a variable content of this allergen in peel and pulp[110]. There is no genetic diversity of the *Pru p 3.01* protein-coding sequence in *P. persica*. Based on the peach genome sequence, further cloning and sequencing of the upstream region in peach germplasm identified three different allele sequences, with variable frequencies in different subpopulations[110]. These differences may link to the variation of Pru p 3 content in different cultivars.

There is a need to select hypoallergenic peach cultivars that can be consumed by mild peach allergy patients. Currently about 100 core peach accessions are being evaluated for their Pru p 3 content by a newly developed sensitive method of monoclonal antibody-based enzyme-linked immunosorbent assay[112]. Preliminary results showed multiple genetic and environmental factors influencing the content. In general, nectarine and red flesh peaches have lower Pru p 3 content. Quantification of Pru p 1 content is currently in process. Peach fruit allergens accumulate more during fruit development, especially at the ripening stage. Most of these allergenic proteins belong to pathogenesis-related protein families and have biological functions, such as broad disease resistance. They may also be related to fruit softening and fruit quality traits. Combined omics' tools used in range-wide peach germplasm samples will provide new qualitative and quantitative insights on these allergens and their impact on other agronomic traits.

Disease resistance

Host natural resistance is essential for cost-effective and environmentally safe strategies to address the challenges of day-by-day biotic stress. Whereas, initially, stone fruit tree cultivation relied on seed propagation, it is now done by grafting. While vegetative propagation in perennials has the advantage of growing extensively desirable genotypes, cultivars remain unchanged for centuries while pathogens rapidly adapt to them, causing devastating diseases that are increasingly more difficult to circumvent with chemicals and/or eradication measures. Additionally, moving toward environment-friendly and healthy production systems is strongly desired by consumers, demanding transition to a chemical pesticide-free agriculture with limited impact on the environment.

In this context, the use of resistant cultivars is the foremost breeding strategy for crop protection. *Prunus* crop species are susceptible to over 70 pests and diseases, but resistant varieties are not readily available and only a few pathosystems are the subject of research programs. Based on the high-quality peach reference genome and tools developed from it, linkage maps were obtained from F_1 and F_2 populations segregating for resistance to brown rot[113,114], green peach aphid[115–118], root-knot nematodes[119–122], fungal gummosis[123], bacterial spot[124], powdery mildew[69,118,125], X-disease[126], and sharka[18,127–130]. Whereas currently none of the genes and QTLs identified are functionally validated, except for the nematode resistance *Ma* locus[131] (Table 1), these data can be used to implement MAB for resistance as demonstrated for bacterial spot disease and green peach aphid in peach[117,132] and sharka in apricot[128,133,134]. A parallel approach to these structurally based genomic analyses is the functionally based identification of differentially expressed genes, in a context of pathogen infection. This strategy relies on the alignment of the RNA-seq reads to the peach reference genome and on its annotation and gene ontology. Only one of those studies based on the transcriptome analysis generated a peach candidate gene that was confirmed in a bacteria/tobacco heterologous system[135]. In spite of this difficulty in functionally validating candidate genes associated with resistance traits, one promising strategy is to search, in the peach reference genome, for orthologous genes that were proven to control resistance/susceptibility to the same pathogen in other hosts[136]. However, this translational research strategy, leveraging knowledge from model plants to *Prunus* crop species, is limited by (1) the significant challenge of genetically transforming stone fruit trees and (2) the acceptance of genetically modified fruits by consumers. Meanwhile, benefiting from whole-genome resequencing in several peach cultivars and other *Prunus* species[119,137–140], our knowledge on resistance, pathogenesis-related, and defense elicitor gene diversity and evolution in *Prunus* species is increasing.

Breeding for resistance to pests and pathogens in stone fruit crops encounters the usual problems associated with breeding perennial plants. This process is significantly hampered by the complexities inherent to tree testing and evaluation, our lack of knowledge of the pest/pathogen

biological interaction, and the diversity and genetic determinism of the target resistance traits that is often quantitative. On top of that, screening for pathogen/pest resistance in stone fruit tree germplasm collections is a very time-consuming and effort intensive task. This likely explains the low number of pests and pathosystems under study, in the past decade. Therefore, there is a great need for academic and practical attention to this area and the development of transnational, high-throughput phenotyping platforms.

The incorporation of resistance genes in current stone fruit breeding programs from related wild species is a promising option. However, it requires significant efforts to identify the valuable germplasm that is unfortunately scarce in the stone fruit crop species. Indeed, excluding resistance mechanisms described in wild *P. kansuensis*, *P. davidiana*, and *P. armeniaca* species[115,119,141], there is very little information about resistance sources in non-crop *Prunus* relatives. This calls for a more systematic survey of the extent *Prunus* wild relatives around the world, as well as their phenotypic characterization for various resistance traits. These materials could be included in the above-mentioned platforms for high-throughput phenotyping. Once new sources of resistance are identified in wild relatives, the next step is the introgression of the genetic factors linked to resistance into elite *Prunus* cultivars through interspecific crosses. Yet, there are inherent difficulties in doing so that can be minimized by using a marker-assisted introgression (MAI) approach recently proposed and described in section "Applications of genome information in *Prunus* improvement: MAB, MAI, genomic selection, and visualization of the genome-wide genetics of elite individuals" of this paper[142]. Genetic analysis of domesticated and non-domesticated stone fruit genomes may identify the presence–absence variations contributing to trait variation. Toward this goal, a groundbreaking initiative would be the construction of a *Prunus* pan-genome, based on already-sequenced *Prunus* genomes[143,144]. This *Prunus* pan-genome would promote and accelerate the development of appropriate genomic tools for all *Prunus* species, including wild relatives.

Whereas all mentioned resistance studies deal with one single pest or pathogen at a time, fruit trees are challenged in the orchard by a cohort of pests and pathogens, in combination or successively. Moreover, plants in the field are under constant threat of multiple abiotic and biotic stresses. Very complex physiological and molecular interactions take place, resulting in unique response(s) of the plants to withstand the combined effect of these stresses. In such situations, plants can show either enhanced or reduced pest/pathogen resistance. While the combined effect of biotic and abiotic stresses has been extensively studied in annual crops, only one study dealt with a perennial host (*Populus* sp.) and none on fruit trees[145]. This establishes a need for studies combining resistance mechanisms to multiple pests and pathogens as well as resistance/tolerance to both biotic and abiotic stresses. Hence, future research programs should explore the utilization of association mapping approaches based on high-throughput genotyping of stone fruit, wild and crop, germplasm collections that will be tested for a variety of biotic and abiotic factors.

Today, epigenetics has become a crucial research field in plant responses to biotic and abiotic stresses, especially in long-lived species such as fruit and forest trees. Plants subjected to contrasting environments have evolved different epigenomes, suggesting that epigenomes are remodeled in plants under stress, such as pathogen and herbivory attacks[146,147] contributing to long-term adaptation[34]. Recently, several studies provided valuable information related to the epigenetic control in plant stress adaptation. Epigenetics refers to the variability in gene expression occurring as a result of modification of DNA and associated proteins. Hence, epimutations are often associated with specific sequence contexts, including tandem and inverted repeat or TE insertions that are thought to attract DNA methylation to nearby loci[34,148]. This feature, together with comparative epigenomic studies, may provide an interesting way to screen for new epialleles associated with responses to biotic stress[149]. However, such whole-genome surveys of epigenome variation are seriously missing in stone fruit tree species research. We need pilot experiments conducted by a coordinated action of laboratories that combine whole-genome bisulfite sequencing, ChIP-seq, and transcriptome analysis in order to successfully map epigenetic factors controlling responses to biotic/abiotic stresses and, in consequence, identify new phenotypic diversity.

Tree architecture

Tree architecture can have a substantial impact on yield and management costs and also is of interest to peach ornamental and rootstock breeding programs. Different peach tree architectures have been described including the standard spreading form, brachytic dwarf, semi-dwarf, compact, spur-type, narrow-leaf, weeping, columnar and upright, or semi-columnar[150,151].

Brachytic dwarfing in peach is controlled by a single gene, *Dw/dw*, located on chromosome 6 that was cloned using a bulk segregant analysis approach combined with genome resequencing[152]. The responsible allele is a loss-of-function mutation in a gibberellic acid receptor (*PpGID1c*) gene (Table 1). A different allele of this gene also causing brachytic dwarfism has been identified[153]. Another gene responsible for dwarfing in peach, *Tssd/tssd*, causes the temperature-sensitive semi-dwarf phenotype and has been fine mapped to a region spanning

500 kb on G3, where 69 genes are annotated in the *Prunus* reference genome[154].

Trees with weeping habit or, by contrast, with branches with vertical growth (pillar) have been identified. For both traits, the genes causing their atypical branch growth orientation have been identified and characterized. In the case of weeping habit, the trait is determined by a single gene (*Pl*/*pl*) located on G3[118]. Using a genomic sequencing approach, a SAM domain gene (*ppa013325; Prupe.3G200700*) affecting branch orientation has been recently identified and functionally validated by silencing in *P. domestica* trees[155] (Table 1). The trait characterized by vertically oriented branches called broomy or pillar is determined by a major gene (*Br/br*). The heterozygous phenotype has a distinctive phenotype called upright. A *PpeTAC1* candidate gene for *Br* has been identified on G2 coupling bulked segregant analysis with NGS[156] (Table 1). This gene is a putative ortholog of rice *TAC1* (*tiller angle control 1*). The mutation causing the pillar phenotype is an insertion located in exon 3 of *PpeTAC1* (*ppa010082; Prupe.2G194000*) causing a stop codon. RNA interference (RNAi) silencing of this gene in *P. domestica* produced trees with a more extreme pillar phenotype than the one described in peach[157].

All the available knowledge of genes controlling these traits and molecular markers to select them makes it possible to efficiently breed new cultivars with characteristics more adapted to different growing practices, including the possibility of high-density plantations and mechanization[158]. Furthermore, other traits related to root architecture need to be included in the global picture. Overexpression of one of these genes, deeper rooting 1 (*DRO1*), in *P. domestica* resulted in trees with deeper roots[159] that could produce plants more adapted to drought.

The gametophytic self-incompatibility (GSI) system of *Prunus*

Fertilization and seed formation are essential for fruit production in *Prunus* because stone fruit and almonds are unable to bear fruit parthenocarpically. Special attention, therefore, has been given to the *Prunus* GSI system and extensive studies have been conducted to elucidate physiological, genetic, and genomic aspects of GSI. During the past two decades, the specificity determinants of pistil (S-ribonuclease; S-RNase) and pollen (*S* haplotype-specific F-box protein; SFB) in *Prunus* have been identified by classical protemoics and genomics techniques based on two-dimensional gel electrophoresis and chromosome walking, respectively[160,161]. Identification of the pistil *S* and the pollen *S* determinants led to the development of PCR-based *S* genotyping and marker-assisted selection for self-compatible (SC) individuals. Furthermore, a series of molecular and genetic analyses of *Prunus* SC *S* haplotypes revealed the possible existence of a distinct recognition mechanism in the S-RNase-based GSI system in *Prunus*[8,160].

Recent studies utilizing WGS information in *Prunus* shed light on the evolution of the *S* locus and establishment of the *Prunus*-specific GSI recognition mechanism. Evolutionary paths of the establishment of *Prunus S-RNase* and *SFB* were investigated by tracking their gene duplication patterns[162]. Phylogenetic analysis and estimation of proxy ages for the establishment of *S-RNase* and its homologs in several rosaceous species showed that the divergence of *S-RNase* in the subtribe Malinae and the genus *Prunus* predated the gene in the most recent common ancestors of Rosaceae species. Furthermore, the duplicated S-RNase-like genes were accompanied by duplicated pollen S-like F-box genes, suggesting segmental duplications of the *S* locus. Analysis of the expression patterns and evolutionary speed of duplicated S-RNase-like genes in *Prunus* suggested that these genes have lost the SI recognition function, resulting in a single *S* locus. Furthermore, phylogenetic analysis with SFB and its orthologs in other angiosperm genomes indicated that *Prunus SFB* does not cluster with the pollen *S* of other plants and diverged early after the establishment of the Eudicots[29]. *Prunus SFB* likely originated from a recent *Prunus*-specific gene duplication event. Transcriptomic and evolutionary analyses of the *Prunus S* paralogs are consistent with the establishment of a *Prunus*-specific SI system and the possibility of subfunctionalization differentiating the newly generated *SFB* from the original pollen *S* determinant. The *S* loci in the current Rosaceae species might have evolved independently from the duplicated *S* loci, which could explain the presence of genus-specific SI recognition mechanisms in the Rosaceae.

Recently, WGS information has been further utilized to identify pollen part modifiers specifically present in *Prunus*[63,163]. Availability of WGS information in diverse species of *Prunus* and resequencing information of cultivars and strains in a given *Prunus* species will further contribute to elucidating the *Prunus*-specific GSI system.

Applications of genome information in *Prunus* improvement: MAB, MAI, genomic selection, and visualization of the genome-wide genetics of elite individuals

Routine application of MAB has long been the main avenue to which stone fruit breeders have expected to benefit from genomics advances. The ability to predict the phenotype before it is expressed and especially in the early stages of each plant's life is of high value in crops that can take ≥5 years before the phenotype can be assessed in the field. Several collaborative attempts to translate that vast

information of QTL and gene analysis in user-friendly informative DNA tests have been attempted[164,165] (www.fruitbreedomics.com; www.rosbreed.org). However, only a limited number of reports on development, validation, and successful implementation of MAB have been published so far[79,87,94,102,133,166–172]. To aid in publicly sharing DNA information, the promises, progress, and prospects of using DNA-based information in breeding in Rosaceae and other horticultural crops have been recently summarized[173–175] and steps to translate promising QTL information to a trait-predictive DNA tests have been proposed[176]. With additional *Prunus* genome sequences being released, an exponential increase in discovery of marker–trait associations is expected. A shift in selection approaches to molecular genotyping and DNA informed breeding efforts have proven useful and are now economically feasible on much larger scales.

The variability of wild and cultivated compatible relatives of *Prunus* species is an enormous reservoir that may contain useful alleles to be integrated into target species, particularly those that have less diversity such as the peach[20,177]. However, this variability has seldom been used in *Prunus* breeding because of: (1) poor knowledge of the inheritance of traits of interest that could be introgressed from exotic sources; (2) limited availability of interspecific crosses and incomplete knowledge of intercrossability among species, and (3) the several generations required to recover an elite genetic background to result in commercially valuable fruit tree cultivars. A method to address these drawbacks, MAI, was proposed and tested successfully in peach × almond crosses[142]. MAI consists of generating a large BC1 population ($N > 1000$) from an interspecific hybrid. A small collection of seedlings (15–25) with a few introgressions (the "prIL set") and overall containing the whole donor genome can be selected from this progeny. These individuals, when backcrossed again or selfed, will produce with reasonable frequency genotypes with the genetic background of the recurrent species, peach in this case, and a single fragment in homozygosis or heterozygosis of the donor parent (almond). In the case of the peach × almond progeny used, the first individuals with a single introgression were extracted 9 years after the beginning of the experiment[142]. Currently, a collection of introgression lines (ILs) heterozygous or homozygous for a single introgressed almond fragment is available with 85% and 45% coverage of the almond genome, respectively (P. Arús unpublished results). Phenotyping the prIL set is also an opportunity for identifying, with low resolution, the genome positions of major genes with dominant or additive almond alleles. The characters of interest can be examined in a genetic background very close to that of the recurrent parent and therefore with limited influence of the donor variability at other genetic regions. This situation makes the prIL set an interesting tool for assessment of the variability supplied by an exotic genome and is an affordable means for long-term storage of the exotic genome in an elite background, often only one generation away from a commercial cultivar. Implementing MAI approaches to other crop × wild or exotic *Prunus* species is an interesting objective for ensuring future availability of the overall diversity of this genus for breeding purposes.

Genomic selection is a molecular genotyping technique that shows promise in fruit tree crops for enhancing breeding efficiency via increased prediction accuracy and selection intensity and decreased generational interval[178–180]. Application of genomic selection in plant breeding is still fairly new and requires further investigation of the most appropriate model(s) and the optimal number of markers and plants to use[181,182]. Preliminary reports on the factors affecting accuracy of genomic prediction in fruit crops suggested importance of relatedness between the reference and validation individuals, with prediction accuracy increasing in highly related material[178,180]. In addition, multicycle analysis in strawberry[180] and peach[179] showed a steady increase of prediction accuracy for all traits as data were aggregated across cycles in the reference population.

Genetic architecture of the trait influences the size of the reference population needed to accurately estimate the SNP effects. The smaller the largest SNP effects are, the larger a reference population is needed[178]. Modeling for the optimal number of markers in apricot[61] and strawberry[180] suggested that prediction accuracy of 88–97% is achieved with ~500–1000 informative, non-redundant, randomly distributed markers. Evaluation of the feasibility of genomic selection in peach for highly polygenic traits linked to yield and fruit quality suggested similar prediction accuracy with over 60% for fruit weight, sugar content, and titratable acidity[179].

Genomic information can also be used to estimate genetic relationships among germplasm populations by modeling the performance of an individual in different environments[183,184] and to estimate an individuals' breeding value[49,179,184,185]. High correlation (0.88; $P < 0.0001$) between predicted genomic breeding value and fruit size was observed in cherry[185]. An initial study on predicting genomic breeding value for fruit weight, sugar content, and titratable acidity in peach showed promising results for application of genomic selection in peach breeding programs[179].

A new opportunity is available for breeders to effectively access and make decisions on the genetics of their germplasm. The breakthrough has four components. The first is a technological advance in genotyping technologies that efficiently provide genome-wide genetic polymorphism data; SNP arrays are particularly useful here because the same loci are assessed on all plants. The second

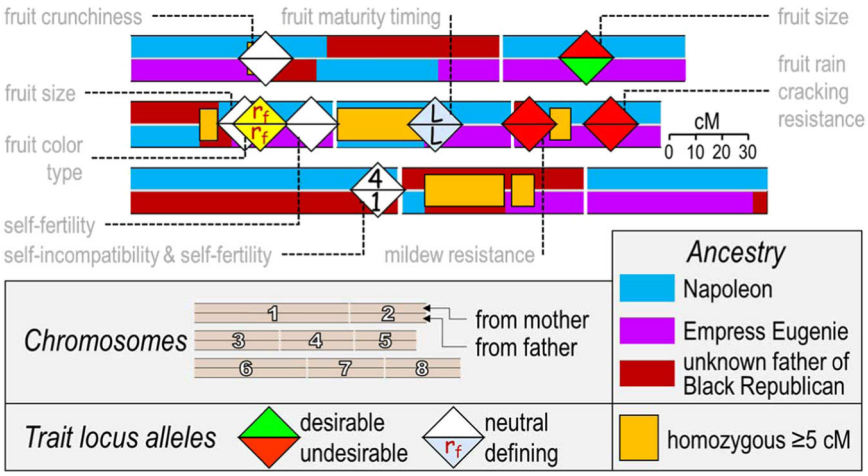

Fig. 3 Haplotype mosaic of "Rainier", a classic Washington-bred sweet cherry cultivar. Segments that "Rainier" inherited via its parents from its three specified ancestors are displayed across the eight chromosomes of sweet cherry. In some cases, these ancestral segments are homozygous, highlighting consequences of inbreeding and signifying common ancestry in generations behind known ancestors. Trait locus alleles are indicated with phenotypic effects and ancestral origins; despite the commercial success of "Rainier," it can be seen that there is still much to be improved. These results were obtained from single-nucleotide polymorphism (SNP) data curation and pedigree ascertainment by L. Cai and C. Peace using the RosBREED cherry 6K SNP array v1 on a U.S. breeding germplasm set ($n \sim 500$)[21]. Diagram is from Peace et al.[175]

component is conceptual, giving explicit attention to each elite germplasm individual (parents, selections, and cultivars) just as breeders do. The common way to deal with large genotypic datasets is a quantitative summary across the germplasm, but this approach easily overlooks each germplasm individual. The third component is an analytical synthesis of genome-wide genotypic data into three genetic vectors of breeding relevance: (1) allelic variation (patterns of similarity among alleles within each locus); (2) linkage (patterns of recombination among loci within each individual); and (3) relatedness (patterns of shared ancestry among individuals within a population). Together, the patterns of alleles over loci over generations can provide emergent genetic knowledge on elite germplasm individuals to improve the precision with which individuals are used in breeding. The key qualitative leap in information now available is the positioning of all recombination events that have led to each plant. SNP datasets of pedigree-connected material must be carefully curated by removing errors, ascertaining all pedigree connections, phasing each individual, and imputing missing data[186]. Haploblocking (section "Haploblocking *Prunus* genomes") also helps. The fourth component is enabling user experience of the holistic "genotype" of each elite individual. It is proposed that sensory experience of DNA information that focuses on elite individuals, just as for phenotypic information, would help breeders better understand what they have and then target development of what they want. While DNA-based genetic information cannot yet be heard or tasted, one way for breeders to "experience the genotype" of an elite individual is by seeing it. Visual aids are well recognized as vital tools for transmitting information for long-term memory retrieval and empowering users to leverage their creativity and understand large data patterns[187]. While improvements in graphical genotyping software are desired as the fifth component to facilitate routine use of genome-wide genetic information in breeding decisions, in the meantime ad hoc collation of the above information can be made as "haplotype mosaics" (Fig. 3). Software is in development to display a circular format for haplotype mosaic visualization. In sweet cherry, visualizing the genetics of elite genomes has recently been used to reveal new pedigree relationships among cultivars and trace origins and distributions of valuable trait locus alleles in ancestors and parents[188].

Development of biotechnology tools: genetic transformation and gene editing

The lack of efficient regeneration systems in *Prunus* is a primary limitation for stable genetic transformation procedures and generation of transformant individuals. New breeding techniques (NBTs), involving approaches such as RNAi, DNA methylation, and gene editing, could overcome the monolithic requirement of having a final T-DNA-modified individual. Novel techniques in gene transfer have a rediscovered value due to simple requirements such as an efficient transient expression[189] or a direct intake of pre-assembled[190] editing reagents into the cell. For this reason, transient expression systems, protoplast regeneration, somatic embryogenesis, and organogenesis regeneration procedures are currently being optimized for NBT application.

The availability of genome drafts in *Prunus* species enables advances toward improved and safer applications

involving NBTs. Regarding gene editing technology, prior knowledge of the genome can help to ensure system efficiency and specificity. The Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)/*Cas*9 system, discovered as an adaptive line of defense against viral infection in *Archaea*[191], is the most common gene editing technique that allows for the direct generation of sequence modifications in the genome. Targeted mutagenesis by this tool involves making guide RNAs (gRNAs) that target customized sequences in the genome to direct the *Cas*9 nuclease activity to generate double-strand breaks adjacent to the gRNA-joined location. gRNA refers to a short synthetic RNA composed of a scaffold sequence necessary for *Cas*-binding and a user-defined 20 nucleotide spacer that determines the genomic target to be modified. The target sequence recognized by the spacer will be a protospacer sequence, located contiguous to an adjacent motif recognized by *Cas*9 required for DNA cleavage, that is an NGG nucleotide arrangement called protospacer adjacent motif (PAM). Currently, new nucleases replacing Cas9 are the focus of new research that uses different PAM motifs[192].

By computing on described genomes, PAM datasets and their contiguous regions can be reviewed and summarized in datasets. In this way, several datasets for gRNA and PAM required in DNA editing have been described and used for on-target and off-target predicting activity of gRNAs inside a genome[193]. In the case of *Prunus* spp., *P. mume* datasets with information about the target sites of Cas9 (i.e., NGG) and newer nucleases such as Cpf1 (which uses TTTNs and TTNs as PAMs) are available at the CRISPR-Local[194] website (http://crispr.hzau.edu.cn/cgi-bin/CRISPR-Local/download). Similarly, based on the peach and sweet cherry genomes, a Cas9 gRNA designer and analyzer for both species is available (www.fruit-tree-genomics.com/biotools).

The system CRISPR/Cas9 has been successfully used in plants and delivery of editing components into the plant cell has been mostly achieved by their stable integration into the genome by gene transfer techniques relying on *Agrobacterium*-mediated transformation[195]. Efforts to edit individuals without foreign DNA insertion into the genome have involved the delivery of assembled ribonucleoprotein editing reagents[190]. A different approach implies the delivery of DNA-replicons using disarmed or "deconstructed" viruses[189], which allows for a high copy number in the cell without the insertion of the replicon into the plant genome.

Efficacy of the above-mentioned techniques in vivo relies on the knowledge of the specific target sequences and, ideally, of whole-genome information to prevent secondary effects on eventual off-targets. For this reason, transient expression assays have become relevant as screening systems prior to a "precise breeding" full-path experimentation. These procedures are relevant because they avoid the technical difficulty and the time needed for whole plant generation, which is a major restriction in *Prunus*[9,196]. Protocols based on the use of explants such as leaf discs, stem segments, or even whole plants have been achieved in other woody species such as grapevines[197,198]. Nevertheless, these approaches are restricted to a few *Prunus* species such as *P. domestica*[199] and *P. salicina*[200] in experiments that can be considered preliminary assays that later led to stable transformation procedures[201,202]. In *P. persica*, different tissues were successfully used as explants for biolistic-mediated β-glucuronidase gene expression[203] in the search for a stable transformation procedure. In addition, protoplast procedures in *P. avium*[204] and *P. avium* hybrids[205] can be proposed as optimal tools for gene editing capability for in vivo evaluation. This technique could help to overcome chimerism, a significant issue in woody plant genetic transformation.

A significant amount of research has been carried out regarding *Prunus* spp. regeneration using different explants and approaches, and optimal conditions for this process are determined by many factors[9]. Whereas leaf explants can lead to adequate regeneration systems in several members of the Rosaceae family[206], extremely recalcitrant species in *Prunus* limit regeneration mostly to using seed explants[9]. Recently, leaf explants of *P. domestica* have shown increased adventitious shoot production by constitutive expression of the class I *KNOX* gene from corn[207].

Improved regeneration efficiencies have been obtained in other species. In sweet cherry varieties, more efficient methods were described using leaf and nodal segments[208] and mature cotyledons[209]. Better results have been reported in other commercially important genotypes, including sour cherry[210,211], black cherry (*P. serotina*)[212], and several cherry rootstocks[211,213,214]. In another example, a complete pipeline of methods for transformation and regeneration of transgenic apricot plants from "Helena" leaves has been described[215,216]. In all cases, researchers used *A. tumefaciens* EHA 105 for gene transfer.

Experimental procedures exploiting RNAi approaches allowed the generation of *Prunus necrotic ringspot virus* (PNRSV) resistant individuals of a cherry rootstock[217]. Grafting of these individuals demonstrated small RNA transport from rootstock to scion, resulting in "Emperor Francis" scions becoming resistant to PNRSV[218]. In this way, the use of *Prunus* genotypes currently suitable for genetic transformation can be proposed as small RNA generators (donors) for systemic transmission of resistance through transgrafting to compatible scions from different genetic backgrounds, directing either systemic sRNA-directed gene silencing or sRNA-directed DNA methylation.

Existing and new bioinformatics tools and resources for *Prunus*

The availability of extensive sequence data forms an essential genomics resource in designing various research platforms to understand the biology of crops and to apply the knowledge in their improvement. Three primary sequence databases are GenBank[219], European Nucleotide Archive[220], and the DNA Data Bank of Japan[221]. As of September 12, 2018, a search of NCBI for data on *Prunus* revealed 4 genomes, 85,014 genes, 196,367 proteins, 2,994,795 nucleotide sequences, 2481 probes, and 2712 records in the short read archive (SRA). The WGS data include *P. persica* v2.0[10,16], *P. avium* v1.0[12], *P. mume* v1.0[11], and *P. yedoensis* var. *nudiflora* (assembly Pyn.v1)[13]. Users can also download the WGS data from their FTP site and perform BLAST analyses. In the GDR[222], in addition to two assemblies of peach, *P. persica* genome v1.0 (IPGI 2013) and v2.0, the whole-genome assembly of sweet cherry *P. avium* genome v1.0 and almond *P. dulcis* Texas v2.0 are available for *Prunus*. Additional data provided by GDR on these assemblies include computational annotation of predicted genes with homology to genes of closely related or model plant species and assignment of InterPro protein domains[223] and GO terms[224,225]. NCBI runs a separate gene annotation pipeline to annotate WGS data. To help researchers compare two different gene annotation sets, GDR performs BLAST analysis between the NCBI annotated genes and genes from the original genome assemblies. The WGS data in GDR can be accessed through the species page, gene/transcript search page, JBrowse[226], and BLASTX[227]. Newly added functionality in the gene/transcript search page allows users customize output to include various functional annotation data in the result table.

The *Prunus* SNP array data and the genotyping data using these SNPs are available to view, search, and download in JBrowse and other pages in GDR. In the SSR and SNP genotype search pages, data from nine SSR and four SNP genotyping projects are available for *Prunus*. The new SNP genotype search page as well as a new plugin in JBrowse in GDR allows users to choose a specific genomic region and peach accessions to view the individual genotype.

The latest and most efficient tool for transcriptome analysis, RNA-seq, allows not only the assessment of the expression level of specific genes but also the detection of less-represented transcripts, allelic-specific expression of transcripts, post-transcriptional mutations, and the expression of splice variants. GDR provides a page that links to *Prunus* RNA-Seq and DNA datasets in the NCBI SRA. The expression data from selected RNA-seq analyses will be available from GDR using Tripal Analysis Expression Module[228]. This module allows users to submit gene sets to generate a heatmap and also view expression patterns of a mRNA in a feature page. The published RNA-Seq and dbEST datasets are also analyzed by GDR to create a reference transcriptome (RefTrans) for major species and provides putative gene function identified by homology to known proteins. For *Prunus*, *P. avium* GDR RefTrans v1.0 and *Prunus persica* GDR RefTrans v1.0, are currently available. Currently, there is not much epigenomic data available for Rosaceae but more data are expected to come[229] and search and analysis tools for epigenomic data will be available in GDR.

With the increasing number of species with WGS, several web-based databases are available for comparative genomics. A few contain the WGS of Rosaceae species. CoGe (The Place to Compare Genomes)[230] contains data of 48,063 genomes including most of the Rosaceae genomes. Plaza[231] and the plant genome duplication database (PGDD)[232] contain some of the Rosaceae WGS, including *P. persica* v1.0. Phytozome[233] currently contains *P. persica* v2.0. In GDR, *P. persica* v2.0 is used in a synteny analysis with eight other Rosaceae whole-genome assemblies using MCScanX[234]. The synteny results are available through the new Synteny Viewer. The gene/mRNA page also lists the orthologs and paralogs identified in this analysis and provides a link to the Synteny Viewer.

One of the most important roles of GDR as the community database for the Rosaceae family is to integrate all the different types of data to maximize the value. In addition to the data mentioned above, GDR includes genetic data, such as map; marker; and QTL, genotype, and phenotype data. GDR currently contains 168 genetic maps for *Prunus*, which can be viewed and compared through a new graphic interface, MapViewer. Trait locus data for *Prunus* in GDR includes 1491 QTLs and 39 major genes for 148 horticultural traits. GDR contains 471,854 *Prunus* genetic markers including 151,479 SNPs. The SNP data are available as JBrowse tracks, downloadable files, and to search and download from the SNP marker search page as well as the all marker search page. The marker search page has a new feature, filtering by trait name, allowing users to search for markers that are associated with QTLs. In addition to the genotype data, phenotypic data from projects such as RosBREED, are available through the "Search Trait Evaluation" page. The public breeding data can also be accessed using the Breeding Information Management System (BIMS). BIMS allows Rosaceae breeders to store, manage, archive, and analyze their private or public breeding data. Future efforts in GDR include more data curation, integration, standardization, and further tool development to utilize the integrated data across crops, data types, and discipline.

Conclusions

Prunus encompasses a group of economically important and closely related crops, sharing an essentially common

genome. Each of them followed a distinct pathway of evolution pre- and post-domestication leading to various biological and reproductive outcomes, including that certain species remain intercrossable producing fertile progeny, while others are reproductively isolated. New sequence-based approaches provide an opportunity for the in-depth study of the genetics and evolution of this unique group of crops. Evolutionary studies of other perennial crops based on whole-genome sequence data are already underway, such as in apple[235] and grape[236]. These studies have uncovered interesting patterns of evolution during and after domestication, including identification of selective sweeps, evolution of population sizes, and introgression from other species as part of the domestication process. Overall, perennial crop evolution appears to follow specific trends as compared with the more extensively studied annual crops[237]. Similar studies for *Prunus* are starting to emerge, particularly in peach[20,25,27,29] and almond[26]. These results, including the comparison of evolution patterns in a group of closely related crop species, are likely to provide fascinating stories that will help to understand the nature of available variability and its potential for developing fruit crops needed for the future.

Even with synteny between *Prunus* species so high, having a single high-quality peach sequence is insufficient. The genomes of some other *Prunus* species have been recently released or will be in the near future, but there is still a need for extending the de novo sequence to all crop species of the genus, a set of key wild relatives and several representatives for each crop selected to cover their main variability/population structure transects. The inter-specific and intraspecific variability that constitutes the pan-genome of *Prunus* has to be characterized. This is crucial for understanding the causal mechanisms explaining the existing genetic diversity among cultivated species and for the analysis of chromosomal regions containing the genes responsible for the variability of important traits within each species.

One of the features of *Prunus* variability is the existence of a large set of genes with major effects on the phenotype (major genes or major QTLs)[4,238]. This is particularly important in the case of peach and may have been favored by its self-compatibility system, enabling for the easy generation of selfed progenies with low or no inbreeding depression, and by artificial selection in the cultivated species for obvious mutants in the fruit, flower, plant habit, and many other characters. Identifying the genes controlling these traits and characterizing their molecular variability has important basic and applied consequences. The progress realized since the release of the first peach genome on gene cloning has been enormous (Table 1) despite inherent difficulties of long intergeneration time, low seed production and germination, and the recalcitrant in vitro behavior of most *Prunus* species. This trend will likely continue in the next decades, further enabled by the exponential increase of DNA sequence information availability and progress of gene editing approaches. These technological advances may help to circumvent some of the current limitations for functional validation of candidate genes and incorporating new valuable alleles into new cultivars of these species.

Establishing routine MAS for major genes in breeding *Prunus* and other temperate fruit crops[117,174,175] is a first step toward employing whole-genome approaches in selecting better cultivars. Knowledge of the genotype across the genome in large sets of historical and modern accessions enables the establishment of highly reliable pedigrees and ways of visualizing, and thus better understanding and using, the available variability in the set of parents used in each breeding program. In addition, this approach allows the complexity of the genome to be simplified by using concepts like haploblocking. Whole-genome selection with a set of markers loosely covering the genome greatly reduces the number of generations needed for recovery of elite germplasm after an elite × wild or exotic cross as demonstrated by MAI[142]. This opens the door for understanding and capturing the genetic variability present in the enormous gene reservoir of the *Prunus* wild and cultivated species. Similar approaches can be used to identify individuals highly similar to a given high-value heterozygous genotype in its selfed progeny or to select two complementary homozygotes that, when crossed, produce a genotype similar to one of contrasted value as described in "resynthesis"[239]. Combined with MAI, this strategy provides a way of introgressing new alleles of interest into an established heterozygous genotype. Finally, genomic selection is a powerful approach for *Prunus* crops with their long intergeneration periods, and research so far shows that it is feasible and potentially useful[61,179]. It is now crucial to design and test genomic selection schemes specifically adapted to the needs of perennial crops. Overall, the increasing amount of DNA information on genes and QTLs and the development of new breeding strategies and methods such as gene editing are providing a significantly enhanced toolbox that can provide a strong push to the development of superior new cultivars. The use of this enhanced toolbox and its integration into the public and private breeding programs will require additional investments in training and translational research[175], following in the steps and leveraging the achievements of successful large-scale European[164] and US[165] initiatives.

Acknowledgements

This research was supported in part by grants from the Ministry of Economy and Competitiveness (MINECO/FEDER projects AGL2015-68329-R and RTA2015-00050-00-00, Severo Ochoa Program for Centres of Excellence in R&D

201-2019 SEV-2015-0533 and CERCA Programme-Generalitat de Catalunya) from Spain; USDA-NIFA-Specialty Crop Research Initiative project, RosBREED: "Enabling marker-assisted breeding in Rosaceae" (2009-51181-05808) and RosBREED 2: "Combining disease resistance with horticultural quality in new rosaceous cultivars" (2014-51181-22378) and USDA NIFA Hatch project 1014919 from USA; and Key Project for New Agricultural Cultivar Breeding in Zhejiang Province (2016C02052-5) from China.

Author details

[1]IRTA, Centre de Recerca en Agrigenòmica CSIC-IRTA-UAB-UB, Campus UAB, Edifici CRAG, Cerdanyola del Vallès (Bellaterra), 08193 Barcelona, Spain. [2]UMR 1332 BFP, INRA, University of Bordeaux, A3C and Virology Teams, 33882 Villenave-d'Ornon Cedex, France. [3]Allergy Research Center, Zhejiang University, 310058 Hangzhou, China. [4]Clemson University, Clemson, SC 29634, USA. [5]Department of Horticulture, Michigan State University, 1066 Bogue Street, East Lansing, MI 48824-1325, USA. [6]Department of Horticulture, Washington State University, Pullman, WA 99164-6414, USA. [7]Biotechnology Laboratory, La Platina Research Station, Instituto de Investigaciones Agropecuarias, Santa Rosa, 11610 La Pintana, Santiago, Chile. [8]Laboratory of Pomology, Graduate School of Agriculture, Kyoto University, Kyoto 606-8502, Japan. [9]Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria (CREA) – Centro di ricerca Olivicoltura, Frutticoltura e Agrumicoltura (CREA-OFA), Rome, Italy. [10]University of Kentucky, 106 T. P. Cooper Hall, Lexington, KY 40546-0073, USA

Conflict of interest

The authors declare that they have no conflict of interest.

Received: 31 January 2019 Revised: 10 March 2019 Accepted: 13 March 2019
Published online: 05 April 2019

References

1. Wen, J. et al. Phylogenetic inferences in Prunus (Rosaceae) using chloroplast ndhF and nuclear ribosomal ITS sequences. *J. Syst. Evol.* **46**, 322–332 (2008).
2. Dirlewanger, E. et al. Comparative mapping and marker-assisted selection in Rosaceae fruit crops. *Proc. Natl. Acad. Sci. USA* **101**, 9891–9896 (2004).
3. Arús, P., Yamamoto, T., Dirlewanger, E. & Abbott, A. G. in *Plant Breeding Reviews* (ed. Janick, J.) Ch. 4 (John Wiley & Sons Inc., Hoboken, NJ, 2005).
4. Arús, P., Verde, I., Sosinski, B., Zhebentyayeva, T. & Abbott, A. G. The peach genome. *Tree Genet. Genomes* **8**, 531–547 (2012).
5. Lester, D. R., Sherman, W. B. & Atwell, B. J. Endopolygalacturonase and the melting flesh (M) locus in peach. *J. Am. Soc. Hortic. Sci.* **121**, 231–235 (1996).
6. Peace, C. P., Crisosto, C. H. & Gradziel, T. M. Endopolygalacturonase: a candidate gene for freestone and melting fleshin peach. *Mol. Breeding* **16**, 21–31 (2005).
7. Bielenberg, D. G. et al. Sequencing and annotation of the evergrowing locus in peach [Prunus persica (L.) Batsch] reveals a cluster of six MADS-box transcription factors as candidate genes for regulation of terminal bud formation. *Tree Genet. Genomes* **4**, 495–507 (2008).
8. Matsumoto, D. & Tao, R. Distinct self-recognition in the Prunus S-RNase-based gametophytic self-incompatibility system. *Hort. J.* **85**, 289–305 (2016).
9. Prieto, H. in *Genetic Transformation* (ed. Alvarez, M.) Ch. 4 (InTech, Rijeka, 2011).
10. Verde, I. et al. The high-quality draft genome of peach (Prunus persica) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nat. Genet.* **45**, 487–U47 (2013).
11. Zhang, Q. et al. The genome of *Prunus mume*. *Nat. Commun.* **3**, 1318 (2012).
12. Shirasawa, K. et al. The genome sequence of sweet cherry (*Prunus avium*) for use in genomics-assisted breeding. *DNA Res.* **24**, 499–508 (2017).
13. Baek, S. et al. Draft genome sequence of wild *Prunus yedoensis* reveals massive inter-specific hybridization between sympatric flowering cherries. *Genome Biol.* **19**, 127 (2018).
14. Arumuganathan, K. & Earle, E. Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* **9**, 208–218 (1991).
15. Howad, W. et al. Mapping with a few plants: using selective mapping for microsatellite saturation of the Prunus reference map. *Genetics* **171**, 1305–1309 (2005).
16. Verde, I. et al. The Peach v2.0 release: high-resolution linkage mapping and deep resequencing improve chromosome-scale assembly and contiguity. *BMC Genomics* **18**, 225 (2017).
17. Verde, I. et al. Development and evaluation of a 9K SNP array for peach by internationally coordinated SNP detection and validation in breeding germplasm. *PLoS ONE* **7**, e35668 (2012).
18. Mariette, S. et al. Genome-wide association links candidate genes to resistance to Plum Pox Virus in apricot (*Prunus armeniaca*). *New Phytol.* **209**, 773–784 (2016).
19. Peace, C. et al. Development and evaluation of a genome-wide 6K SNP array for diploid sweet cherry and tetraploid sour cherry. *PLoS ONE* **7**, e48305 (2012).
20. Micheletti, D. et al. Whole-genome analysis of diversity and SNP-major gene association in peach germplasm. *PLoS ONE* **10**, e0136803 (2015).
21. Li, X.-w et al. Peach genetic resources: diversity, population structure and linkage disequilibrium. *BMC Genet.* **14**, 84 (2013).
22. Xie, R. et al. Evaluation of the genetic diversity of Asian peach accessions using a selected set of SSR markers. *Sci. Hortic.* **125**, 622–629 (2010).
23. Campoy, J. et al. Genetic diversity, linkage disequilibrium, population structure and construction of a core collection of *Prunus avium* L. landraces and bred cultivars. *BMC Plant Biol.* **16**, 49 (2016).
24. Delplancke, M. et al. Combining conservative and variable markers to infer the evolutionary history of *Prunus* subgen. *Amygdalus* s.l. under domestication. *Genet. Resour. Crop Evol.* **63**, 221–234 (2016).
25. Yu, Y. et al. Genome re-sequencing reveals the evolutionary history of peach fruit edibility. *Nat. Commun.* **9**, 5404 (2018).
26. Velasco, D., Hough, J., Aradhya, M. & Ross-Ibarra, J. Evolutionary genomics of peach and almond domestication. *G3 Genes Genomics Genet.* **g3**, 116.032672 (2016).
27. Cao, K. et al. Comparative population genomics reveals the domestication history of the peach, Prunus persica, and human influences on perennial fruit crops. *Genome Biol.* **15**, 415 (2014).
28. Cao, K. et al. Genome-wide association study of 12 agronomic traits in peach. *Nat. Commun.* **7**, 10 (2016).
29. Akagi, T., Hanada, T., Yaegaki, H., Gradziel, T. M. & Tao, R. Genome-wide view of genetic diversity reveals paths of selection and cultivar differentiation in peach domestication. *DNA Res.* **23**, 271–282 (2016).
30. Okie, W. & Weinberger, J. H. in *Fruit Breeding: Tree and Tropical Fruits* (ed. Janick, J. M. J.) Ch. 10 (Wiley, New York, 1996).
31. Xie, Z. et al. Mutation rate analysis via parent-progeny sequencing of the perennial peach. I. A low rate in woody perennials and a higher mutagenicity in hybrids. *Proc. Biol. Sci.* **283**, 20161016 (2016).
32. Zhang, Q. et al. The genetic architecture of floral traits in the woody plant *Prunus mume*. *Nat. Commun.* **9**, 1702 (2018).
33. Cirilli, M. et al. Integrative genomics approaches validate PpYUC11-like as candidate gene for the stony hard trait in peach (*P. persica* L. *Batsch*). *BMC Plant Biol.* **18**, 88 (2018).
34. Dubin, M. J. et al. DNA methylation in Arabidopsis has a genetic basis and shows evidence of local adaptation. *Elife* **4**, e05255 (2015).
35. Kawakatsu, T. et al. Epigenomic diversity in a global collection of *Arabidopsis thaliana* accessions. *Cell* **166**, 492–505 (2016).
36. He, J. Q., Harrison, R. J. & Li, B. A novel 3D imaging system for strawberry phenotyping. *Plant Methods* **13**, 93–93 (2017).
37. Yandún Narváez, F. J., Salvo del Pedregal, J., Prieto, P. A., Torres-Torriti, M. et al. & Auat Cheein, F. A. LiDAR and thermal images fusion for ground-based 3D characterisation of fruit trees. *Biosyst. Eng.* **151**, 479–494 (2016).
38. Rouphael, Y., Spíchal, L., Panzarová, K., Casa, R. & Colla, G. High-Throughput Plant Phenotyping for Developing Novel Biostimulants: From Lab to Field or From Field to Lab? Front. Plant. Sci. 9 (2018).
39. Voorrips, R. E., Bink, M. C. A. M., Kruisselbrink, J. W., Koehorst-van Putten, H. J. J. & van de Weg, W. E. PediHaplotyper: software for consistent assignment of marker haplotypes in pedigrees. *Mol. Breeding* **36**, 1–10 (2016).
40. Bink, M. C. A. M. et al. Bayesian QTL analyses using pedigreed families of an outcrossing species, with application to fruit firmness in apple. *Theor. Appl. Genet.* **127**, 1073–1090 (2014).
41. Cai, L., Voorrips, R. E., van de Weg, E., Peace, C. & Iezzoni, A. Genetic structure of a QTL hotspot on chromosome 2 in sweet cherry indicates positive selection for favorable haplotypes. *Mol. Breeding* **37**, 85 (2017).
42. Bianco, L. et al. Development and validation of a 20K single nucleotide polymorphism (SNP) whole genome genotyping array for apple (Malus x domestica Borkh). *PLoS ONE* **9**, 9 (2014).

43. Peace, C. & Norelli, J. Experiencing "the genotype" with haplotype mosaics - an Events article. *Community Breeders' Page. RosBREED Newsletter* **7**, 6–7 (2017).
44. Olsen, J. E. Light and temperature sensing and signaling in induction of bud dormancy in woody plants. *Plant Mol. Biol.* **73**, 37–47 (2010).
45. Perry, T. O. Dormancy of trees in winter. *Science* **171**, 29–36 (1971).
46. Lang, G. A., Early, J. D., Martin, G. C. & Darnell, R. L. Endo-, para-and eco-dormancy: physiological terminology and classification for dormancy research. *Hortic. Sci.* **22**, 271–277 (1987).
47. Dirlewanger, E. 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* **109**, 280–292 (2012).
48. Pirona, R. et al. Fine mapping and identification of a candidate gene for a major locus controlling maturity date in peach. *BMC Plant Biol.* **13**, 1–13 (2013).
49. Hernández Mora, J. R. et al. Integrated QTL detection for key breeding traits in multiple peach progenies. *BMC Genomics* **18**, 404 (2017).
50. Castède, S. et al. Genetic determinism of phenological traits highly affected by climate change in Prunus avium: Flowering date dissected into chilling and heat requirements. *New Phytol.* **202**, 703–715 (2014).
51. Castède, S. et al. Mapping of candidate genes involved in bud dormancy and flowering time in sweet cherry (*Prunus avium*). *PLoS ONE* **10**, e0143250 (2015).
52. Kitamura, Y., Takeuchi, T., Yamane, H. & Tao, R. Simultaneous down-regulation of DORMANCY-ASSOCIATED MADS-box6 and SOC1during dormancy release in Japanese apricot (Prunus mume) flower buds. *J. Hortic. Sci. Biotech.* **91**, 476–482 (2016).
53. Jiménez, S., Reighard, G. L. & Bielenberg, D. G. Gene expression of DAM5 and DAM6 is suppressed by chilling temperatures and inversely correlated with bud break rate. *Plant Mol. Biol.* **73**, 157–167 (2010).
54. Zhu, Y. et al. RNA-Seq-based transcriptome analysis of dormant flower buds of Chinese cherry (*Prunus pseudocerasus*). *Gene* **555**, 362–376 (2015).
55. Vimont, N. et al. The presence of H3K4me3 histone mark is positively correlated with expression at the DAM loci in sweet cherry during dormancy. Acta Hortic. 1235 (2019). (in press). https://doi.org/10.17660/ActaHortic.2019.1235.57.
56. Sasaki, R. et al. Functional and expressional analyses of PmDAM genes associated with endodormancy in Japanese apricot. *Plant Physiol.* **157**, 485–497 (2011).
57. Zhao, K. et al. Comprehensive cloning of *Prunus mume* dormancy associated MADS-Box genes and their response in flower bud development and dormancy. *Front. Plant Sci.* **9**, 17 (2018).
58. Leida, C., Conesa, A., Llácer, G., Badenes, M. L. & Ríos, G. Histone modifications and expression of DAM6 gene in peach are modulated during bud dormancy release in a cultivar-dependent manner. *New Phytol.* **193**, 67–80 (2012).
59. Rothkegel, K. et al. DNA methylation and small interference RNAs participate in the regulation of MADS-box genes involved in dormancy in sweet cherry (*Prunus avium* L.). *Tree Physiol.* **37**, 1739–1751 (2017).
60. Kronenberg, Z. N. et al. FALCON-Phase: integrating PacBio and Hi-C data for phased diploid genomes. *bioRxiv* 327064 (2018).
61. Nsibi, M. et al. Genomic selection - which prospects in *Prunus armeniaca*? Preliminary results issued from fruit quality traits. In *9th International Rosaceae Genomics Conference Program and Abstracts*; 26–30 June; Nanjing, China. 2018; p 61.
62. Dirlewanger, E. et al. Sweet cherry: new genomic tools for the creation of varieties adapted to future conditions. In *9th International Rosaceae Genomics Conference Program and Abstracts*; 26–30 June; Nanjing, China. 2018; p 31.
63. Ono, K., Akagi, T., Morimoto, T., Wünsch, A. & Tao, R. Genome re-sequencing of diverse sweet cherry (Prunus avium) individuals reveals a modifier gene mutation conferring pollen-part self-compatibility. *Plant Cell Physiol.* **59**, 1265–1275 (2018).
64. Herrero, J. & Ibarz, P. (eds) Cartografía de Frutales de Hueso y Pepita (CSIC-Estación experimental de Aula Dei, 1972).
65. Egea, J., Dicenta, F. & Burgos, L. 'Rojo Pasion' apricot. *HortScience* **39**, 1490–1491 (2004).
66. Olmstead, J. W., Iezzoni, A. F. & Whiting, M. D. Genotypic differences in sweet cherry fruit size are primarily a function of cell number. *J. Am. Soc. Hortic. Sci.* **132**, 697–703 (2007).
67. Quilot, B. et al. QTL analysis of quality traits in an advanced backcross between *Prunus persica* cultivars and the wild relative species *P. davidiana*. *Theor. Appl. Genet.* **109**, 884–897 (2004).
68. da Silva Linge, Cs et al. Genetic dissection of fruit weight and size in an F2 peach (*Prunus persica* (L.) *Batsch*) progeny. *Mol. Breeding* **35**, 1–19 (2015).
69. Donoso, J. M. et al. Exploring almond genetic variability useful for peach improvement: mapping major genes and QTLs in two interspecific almond x peach populations. *Mol. Breeding* **36**, 1–17 (2016).
70. Salazar, J. A. et al. Genotyping by sequencing for SNP-based linkage analysis and identification of QTLs linked to fruit quality traits in Japanese plum (*Prunus salicina* Lindl.). *Front. Plant Sci.* **8**, 476 (2017).
71. Campoy, J. A., Le Dantec, L., Barreneche, T., Dirlewanger, E. & Quero-García, J. New insights into fruit firmness and weight control in sweet cherry. *Plant Mol. Biol. Rep.* **33**, 783–796 (2015).
72. Frary, A. fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. *Science* **289**, 85–88 (2000).
73. De Franceschi, P. et al. Cell number regulator genes in Prunus provide candidate genes for the control of fruit size in sweet and sour cherry. *Mol. Breeding* **32**, 311–326 (2013).
74. Fresnedo-Ramírez, J. et al. QTL mapping and breeding value estimation through pedigree-based analysis of fruit size and weight in four diverse peach breeding programs. *Tree Genet. Genomes* **12**, 25 (2016).
75. Chakrabarti, M. et al. A cytochrome P450 regulates a domestication trait in cultivated tomato. *Proc. Natl. Acad. Sci. USA* **110**, 17125–17130 (2013).
76. Qi, X., Liu, C., Song, L., Li, Y. & Li, M. PaCYP78A9, a cytochrome P450, regulates fruit size in sweet cherry (*Prunus avium* L.). *Front. Plant Sci.* **8**, 2076 (2017).
77. Guo, J. et al. Comparative transcriptome and microscopy analyses provide insights into flat shape formation in peach (*Prunus persica*). *Front. Plant Sci.* **8**, 2215 (2018).
78. Dirlewanger, E. et al. Development of a second-generation genetic linkage map for peach [Prunus persica (L.) Batsch] and characterization of morphological traits affecting flower and fruit. *Tree Genet. Genomes* **3**, 1–13 (2006).
79. López-Girona, E. et al. A deletion affecting an LRR-RLK gene co-segregates with the fruit flat shape trait in peach. *Sci. Rep.* **7**, 6714 (2017).
80. Bliss, F. A. et al. An expanded genetic linkage map of Prunus based on an interspecific cross between almond and peach. *Genome* **45**, 520–529 (2002).
81. Brandi, F. et al. Study of 'Redhaven' peach and its white-fleshed mutant suggests a key role of CCD4 carotenoid dioxygenase in carotenoid and norisoprenoid volatile metabolism. *BMC Plant Biol.* **11**, 24 (2011).
82. Falchi, R. et al. Three distinct mutational mechanisms acting on a single gene underpin the origin of yellow flesh in peach. *Plant J.* **76**, 175–187 (2013).
83. Espley, R. V. et al. Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. *Plant J.* **49**, 414–427 (2007).
84. Chagné, D. et al. Mapping a candidate gene (MdMYB10) for red flesh and foliage colour in apple. *BMC Genomics* **8**, 212 (2007).
85. Sooriyapathirana, S. S. et al. QTL analysis and candidate gene mapping for skin and flesh color in sweet cherry fruit (*Prunus avium* L.). *Tree Genet. Genomes* **6**, 821–832 (2010).
86. Bretó, M. P., Cantín, C. M., Iglesias, I., Arús, P. & Eduardo, I. Mapping a major gene for red skin color suppression (highlighter) in peach. *Euphytica* **213**, 14 (2017).
87. Frett, T. J., Reighard, G. L., Okie, W. R. & Gasic, K. Mapping quantitative trait loci associated with blush in peach [*Prunus persica* (L.) Batsch]. *Tree Genet. Genomes* **10**, 367–381 (2014).
88. Zhou, H. et al. Molecular genetics of blood-fleshed peach reveals activation of anthocyanin biosynthesis by NAC transcription factors. *Plant J.* **82**, 105–121 (2015).
89. Fang, Z.-Z., Zhou, D.-R., Ye, X.-F., Jiang, C.-C. & Pan, S.-L. Identification of candidate anthocyanin-related genes by transcriptomic analysis of 'Furongli' plum (*Prunus salicina* Lindl.) during fruit ripening using RNA-seq. Front. Plant Sci. **7**, 1338 (2016).
90. Eduardo, I. et al. QTL analysis of fruit quality traits in two peach intraspecific populations and importance of maturity date pleiotropic effect. *Tree Genet. Genomes* **7**, 323–335 (2011).
91. Salazar, J. A., Ruiz, D., Egea, J. & Martínez-Gómez, P. Transmission of fruit quality traits in apricot (Prunus armeniaca L.) and analysis of linked quantitative trait loci (QTLs) using simple sequence repeat (SSR) markers. *Plant Mol. Biol. Rep.* **31**, 1506–1517 (2013).
92. Dirlewanger, E. et al. Mapping QTLs controlling fruit quality in peach (*Prunus persica* (L.) Batsch. *Theor. Appl. Genet.* **98**, 18–31 (1999).

93. Boudehri, K. et al. Phenotypic and fine genetic characterization of the D locus controlling fruit acidity in peach. *BMC Plant Biol.* **9**, 59 (2009).
94. Eduardo, I. et al. Development of diagnostic markers for selection of the subacid trait in peach. *Tree Genet. Genomes* **10**, 1695–1709 (2014).
95. Desnoues, E. et al. Dynamic QTLs for sugars and enzyme activities provide an overview of genetic control of sugar metabolism during peach fruit development. *J. Exp. Bot.* **67**, 3419–3431 (2016).
96. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
97. Farcuh, M. et al. Sugar metabolism reprogramming in a non-climacteric bud mutant of a climacteric plum fruit during development on the tree. *J. Exp. Bot.* **68**, 5813–5828 (2017).
98. Gu, C. et al. Copy number variation of a gene cluster encoding endopolygalacturonase mediates flesh texture and stone adhesion in peach. *J. Exp. Bot.* **67**, 1993–2005 (2016).
99. Sandefur, P., Clark, J. R. & Peace, C. in *Horticultural Reviews* (ed. Janick, J.). Ch. 6 (John Wiley & Sons, Inc., New York, 2013).
100. Ghiani, A. et al. Melting of 'Big Top' nectarine fruit: some physiological, biochemical, and molecular aspects. *J. Am. Soc. Hortic. Sci.* **136**, 61–68 (2011).
101. Serra, O. et al. Genetic analysis of the slow-melting flesh character in peach. *Tree Genet. Genomes* **13**, 77 (2017).
102. Meneses, C. et al. A codominant diagnostic marker for the slow ripening trait in peach. *Mol. Breeding* **36**, 77 (2016).
103. Pan, L. et al. PpYUC11, a strong candidate gene for the stony hard phenotype in peach (Prunus persica L. Batsch), participates in IAA biosynthesis during fruit ripening. *J. Exp. Bot.* **66**, 7031–7044 (2015).
104. Tatsuki, M. et al. Insertion of a transposon-like sequence in the 5 '-flanking region of the YUCCA gene causes the stony hard phenotype. *Plant J.* **96**, 815–827 (2018).
105. Gao, Z.-S. et al. Peach allergy in China: a dominant role for mugwort pollen lipid transfer protein as a primary sensitizer. *J. Allergy Clin. Immun.* **131**, 224–226.e3 (2013).
106. Matricardi, P. et al. EAACI molecular allergology user's guide. *Pediatr. Allergy Immun.* **27**, 1–250 (2016).
107. Chen, L. et al. Genomic characterization of putative allergen genes in peach/almond and their synteny with apple. *BMC Genomics* **9**, 543 (2008).
108. Tuppo, L. et al. Peamaclein - a new peach allergenic protein: similarities, differences and misleading features compared to Pru p 3. *Clin. Exp. Allergy* **43**, 128–140 (2013).
109. Yang, Z. et al. Differential transcript abundance and genotypic variation of four putative allergen-encoding gene families in melting peach. *Tree Genet. Genomes* **7**, 903–916 (2011).
110. Ma, Y.-t. et al. The PpLTP1 primary allergen gene is highly conserved in peach and has small variations in other Prunus species. *Plant Mol. Biol. Rep.* **32**, 652–663 (2014).
111. Gao, Z.-S. et al. IgE-binding potencies of three peach Pru p 1 isoforms. *Mol. Nutr. Food Res.* **60**, 2457–2466 (2016).
112. Gao, Z.-S. et al. Quantification of peach fruit allergen lipid transfer protein by a double monoclonal antibody-based sandwich ELISA. *Food Anal. Method* **9**, 823–830 (2016).
113. Pacheco, I. et al. QTL mapping for brown rot (*Monilinia fructigena*) resistance in an intraspecific peach (*Prunus persica* L. Batsch) F1 progeny. *Tree Genet. Genomes* **10**, 1223–1242 (2014).
114. Martínez-García, P. J. et al. Application of genomic and quantitative genetic tools to identify candidate resistance genes for brown rot resistance in peach. *PLoS ONE* **8**, e78634 (2013).
115. Sauge, M. H., Lambert, P. & Pascal, T. Co-localisation of host plant resistance QTLs affecting the performance and feeding behaviour of the aphid Myzus persicae in the peach tree. *Heredity* **108**, 292–301 (2012).
116. Lambert, P. & Pascal, T. Mapping Rm2 gene conferring resistance to the green peach aphid (Myzus persicae Sulzer) in the peach cultivar 'Rubira®'. *Tree Genet. Genomes* **7**, 1057–1068 (2011).
117. Lambert, P. et al. Identifying SNP markers tightly associated with six major genes in peach [Prunus persica (L.) Batsch] using a high-density SNP array with an objective of marker-assisted selection (MAS). *Tree Genet. Genomes* **12**, 121 (2016).
118. Pascal, T. et al. Mapping of new resistance (Vr2, Rm1) and ornamental (Di2, pl) Mendelian trait loci in peach. *Euphytica* **213**, 132 (2017).
119. Cao, K., Wang, L. R., Zhu, G. R., Fang, C. H. & Chen, C. W. Isolation, characterisation and phylogenetic analysis of resistance gene analogues in a wild species of peach (*Prunus kansuensis*). *Can. J. Plant Sci.* **91**, 961–970 (2011).
120. Duval, H. et al. High-resolution mapping of the RMia gene for resistance to root-knot nematodes in peach. *Tree Genet. Genomes* **10**, 297–306 (2014).
121. Cao, K. et al. Identification of a candidate gene for resistance to root-knot nematode in a wild peach and screening of its polymorphisms. *Plant Breeding* **133**, 530–535 (2014).
122. Maquilan, M. A. D., Olmstead, M. A., Olmstead, J. W., Dickson, D. W. & Chaparro, J. X. Genetic analyses of resistance to the peach root-knot nematode (Meloidogyne floridensis) using microsatellite markers. *Tree Genet. Genomes* **14**, 47 (2018).
123. Mancero-Castillo, D., Beckman, T. G., Harmon, P. F. & Chaparro, J. X. A major locus for resistance to Botryosphaeria dothidea in Prunus. *Tree Genet. Genomes* **14**, 26 (2018).
124. Yang, N., Reighard, G., Ritchie, D., Okie, W. & Gasic, K. Mapping quantitative trait loci associated with resistance to bacterial spot (*Xanthomonas arboricola* pv. pruni) in peach. *Tree Genet. Genomes* **9**, 573–586 (2013).
125. Salazar, J. A. et al. Quantitative trait loci (QTLs) identification and the transmission of resistance to powdery mildew in apricot. *Euphytica* **211**, 245–254 (2016).
126. Lenz, R. R. & Dai, W. Mapping X-disease phytoplasma resistance in *Prunus virginiana*. *Front. Plant. Sci.* **8**, 2057 (2017).
127. Vera Ruiz, E. M. A. et al. Narrowing down the apricot Plum pox virus resistance locus and comparative analysis with the peach genome syntenic region. *Mol. Plant Pathol.* **12**, 535–547 (2011).
128. Soriano, J. M. et al. Identification of simple sequence repeat markers tightly linked to plum pox virus resistance in apricot. *Mol. Breeding* **30**, 1017–1026 (2012).
129. Zuriaga, E., Romero, C., Blanca, J. M. & Badenes, M. L. Resistance to Plum Pox Virus (PPV) in apricot (Prunus armeniaca L.) is associated with down-regulation of two MATHd genes. *BMC Plant Biol.* **18**, 25 (2018).
130. Cirilli, M. et al. Genetic dissection of Sharka disease tolerance in peach (P. persica L. Batsch). *BMC Plant Biol.* **17**, 192 (2017).
131. Claverie, M. et al. The Ma gene for complete-spectrum resistance to Meloidogyne species in Prunus is a TNL with a huge repeated C-terminal post-LRR region. *Plant Physiol.* **156**, 779–792 (2011).
132. Gasic, K. et al. Bacterial spot resistance in peach: functional allele distribution in breeding germplasm. *Acta Hortic.* **1084**, 69–74 (2015).
133. Decroocq, S. et al. Selecting with markers linked to the PPVres major QTL is not sufficient to predict resistance to Plum pox virus (PPV) in apricot. *Tree Genet. Genomes* **10**, 1161–1170 (2014).
134. Passaro, M., Geuna, F., Bassi, D. & Cirilli, M. Development of a high-resolution melting approach for reliable and cost-effective genotyping of PPVres locus in apricot (*P. armeniaca*). *Mol. Breeding* **37**, 1–7 (2017).
135. Sherif, S., El-Sharkawy, I., Paliyath, G. & Jayasankar, S. PpERF3b, a transcriptional repressor from peach, contributes to disease susceptibility and side branching in EAR-dependent and -independent fashions. *Plant Cell Rep.* **32**, 1111–1124 (2013).
136. Wang, X. H. et al. Silencing of the host factor eIF(iso)4E gene confers Plum pox virus resistance in plum. *PLoS ONE* **8**, e50627 (2013).
137. Sherif, S., Paliyath, G. & Jayasankar, S. Molecular characterization of peach PR genes and their induction kinetics in response to bacterial infection and signaling molecules. *Plant Cell Rep.* **31**, 697–711 (2012).
138. Terefe-Ayana, D., Kaufmann, H., Linde, M. & Debener, T. Evolution of the Rdr1 TNL-cluster in roses and other rosaceous species. *BMC Genomics* **13**, 409 (2012).
139. Ruiz, C., Nadal, A., Montesinos, E. & Pla, M. Novel Rosaceae plant elicitor peptides as sustainable tools to control *Xanthomonas arboricola* pv. pruni in Prunus spp. *Mol. Plant Pathol.* **19**, 418–431 (2018).
140. Ruiz, C. et al. Diversity of plant defense elicitor peptides within the Rosaceae. *BMC Genet.* **19**, 11 (2018).
141. Decroocq, S. et al. New insights into the history of domesticated and wild apricots and its contribution to Plum pox virus resistance. *Mol. Ecol.* **25**, 4712–4729 (2016).
142. Serra, O. et al. Marker-assisted introgression (MAI) of almond genes into the peach background: a fast method to mine and integrate novel variation from exotic sources in long intergeneration species. *Tree Genet. Genomes* **12**, 96 (2016).
143. Sun, C. et al. RPAN: rice pan-genome browser for similar to 3000 rice genomes. *Nucleic Acids Res.* **45**, 597–605 (2017).
144. Ou, L. et al. Pan-genome of cultivated pepper (Capsicum) and its use in gene presence–absence variation analyses. *New Phytol.* **220**, 360–363 (2018).

145. Dangi, A. K., Sharma, B., Khangwal, I. & Shukla, P. Combinatorial interactions of biotic and abiotic stresses in plants and their molecular mechanisms: systems biology approach. *Mol. Biotechnol.* **60**, 636–650 (2018).
146. Herrera, C. M. & Bazaga, P. Untangling individual variation in natural populations: ecological, genetic and epigenetic correlates of long-term inequality in herbivory. *Mol. Ecol.* **20**, 1675–1688 (2011).
147. Dowen, R. H. et al. Widespread dynamic DNA methylation in response to biotic stress. *Proc. Natl Acad. Sci. USA* **109**, E2183–E2191 (2012).
148. Song, X. & Cao, X. Transposon-mediated epigenetic regulation contributes to phenotypic diversity and environmental adaptation in rice. *Curr. Opin. Plant Biol.* **36**, 111–118 (2017).
149. Fortes, A. M. & Gallusci, P. Plant stress responses and phenotypic plasticity in the epigenomics era: perspectives on the grapevine scenario, a model for perennial crop plants. *Front. Plant. Sci.* **8**, 82 (2017).
150. Scorza, R. Theory and practice of genetically manipulating peach tree architecture. *New York Fruit Quart.* **13**, 27–31 (2005).
151. Scorza, R., Miller, S., Glenn, D. M., Okie, W. R. & Tworkoski, T. Developing peach cultivars with novel tree growth habits. Acta Hortic. **713**, 61–64 (2006).
152. Hollender, C. A., Hadiarto, T., Srinivasan, C., Scorza, R. & Dardick, C. A brachytic dwarfism trait (dw) in peach trees is caused by a nonsense mutation within the gibberellic acid receptor PpeGID1c. *New Phytol.* **210**, 227–239 (2015).
153. Cantín, C. M., Arús, P. & Eduardo, I. Identification of a new allele of the Dw gene causing brachytic dwarfing in peach. *BMC Res. Notes* **11**, 386 (2018).
154. Lu, Z. et al. Fine mapping of the temperature-sensitive semi-dwarf (Tssd) locus regulating the internode length in peach (*Prunus persica*). *Mol. Breeding* **36**, 20 (2016).
155. Hollender, C. A. et al. Loss of a highly conserved sterile alpha motif domain gene (WEEP) results in pendulous branch growth in peach trees. *Proc. Natl Acad. Sci.* **115**, E4690–E4699 (2018).
156. Dardick, C. et al. PpeTAC1 promotes the horizontal growth of branches in peach trees and is a member of a functionally conserved gene family found in diverse plants species. *Plant J.* **75**, 618–630 (2013).
157. Hollender, C. A. et al. Alteration of TAC1 expression in Prunus species leads to pleiotropic shoot phenotypes. *Hortic. Res.* **5**, 26 (2018).
158. Miller, S. S. & Scorza, R. Response of two novel peach tree growth habits to in-row tree spacing, training system, and pruning: effect on growth and pruning. *J. Am. Pomol. Soc.* **64**, 199 (2010).
159. Guseman, J. M., Webb, K., Srinivasan, C. & Dardick, C. DRO1 influences root system architecture in Arabidopsis and Prunus species. *Plant J.* **89**, 1093–1105 (2017).
160. Tao, R. & Iezzoni, A. F. The S-RNase-based gametophytic self-incompatibility system in Prunus exhibits distinct genetic and molecular features. *Sci. Hortic.* **124**, 423–433 (2010).
161. Yamane, H. & Tao, R. Molecular basis of self-(in)compatibility and current status of S-genotyping in rosaceous fruit trees. *J. Jpn. Soc. Hortic. Sci.* **78**, 137–157 (2009).
162. Morimoto, T., Akagi, T. & Tao, R. Evolutionary analysis of genes for S-RNase-based self-incompatibility reveals S locus duplications in the ancestral Rosaceae. *Horticult. J.* **84**, 233–242 (2015).
163. Muñoz-Sanz, J. V., Zuriaga, E., Badenes, M. L. & Romero, C. A disulfide bond A-like oxidoreductase is a strong candidate gene for self-incompatibility in apricot (*Prunus armeniaca*) pollen. *J. Exp. Bot.* **68**, 5069–5078 (2017).
164. Laurens, F. et al. An integrated approach for increasing breeding efficiency in apple and peach in Europe. *Hortic. Res.* **5**, 11 (2018).
165. Iezzoni, A. et al. RosBREED: enabling marker-assisted breeding in Rosaceae. *Acta Hortic.* **859**, 389–394 (2010).
166. Picañol, R. et al. Combining linkage and association mapping to search for markers linked to the flat fruit character in peach. *Euphytica* **190**, 279–288 (2013).
167. Vendramin, E. et al. A unique mutation in a MYB gene cosegregates with the nectarine phenotype in peach. *PLoS ONE* **9**, e90574 (2014).
168. Stegmeir, T., Cai, L., Basundari, F. R. A., Sebolt, A. M. & Iezzoni, A. F. A DNA test for fruit flesh color in tetraploid sour cherry (*Prunus cerasus* L.). *Mol. Breeding* **35**, 149 (2015).
169. Sandefur, P., Oraguzie, N. & Peace, C. A DNA test for routine prediction in breeding of sweet cherry fruit color, Pav-Rf-SSR. *Mol. Breeding* **36**, 33 (2016).
170. Wang, Q. et al. DNA marker-assisted evaluation of fruit acidity in diverse peach (Prunus persica) germplasm. *Euphytica* **210**, 413–426 (2016).
171. Morgutti, S. et al. Endopolygalacturonase gene polymorphisms: asset of the locus in different peach accessions. *Am. J. Plant Sci.* **08**, 941–957 (2017).
172. Sandefur, P., Frett, T., Clark, J., Gasic, K. & Peace, C. A DNA test for routine prediction in breeding of peach blush, Ppe-R-f-SSR. *Mol. Breeding* **37**, 15 (2017).
173. Edge-Garza, D. A., Luby, J. J. & Peace, C. Decision support for cost-efficient and logistically feasible marker-assisted seedling selection in fruit breeding. *Mol. Breeding* **35**, 223 (2015).
174. Ru, S., Main, D., Evans, K. & Peace, C. Current applications, challenges, and perspectives of marker-assisted seedling selection in Rosaceae tree fruit breeding. *Tree Genet. Genomes* **11**, 1–12 (2015).
175. Peace, C. P. DNA-informed breeding of rosaceous crops: promises, progress and prospects. *Hortic. Res.* **4**, 17006 (2017).
176. Vanderzande, S. et al. Crossing the finish line: how to develop diagnostic DNA tests as breeding tools after QTL discovery. *J. Hortic.* **5**, 228 (2018).
177. Mnejja, M., Garcia-Mas, J., Audergon, J.-M. & Arús, P. Prunus microsatellite marker transferability across rosaceous crops. *Tree Genet. Genomes* **6**, 689–700 (2010).
178. Muranty, H. et al. Accuracy and responses of genomic selection on key traits in apple breeding. *Hortic. Res.* **2**, 15060 (2015).
179. Biscarini, F. et al. Genome-enabled predictions for fruit weight and quality from repeated records in European peach progenies. *BMC Genomics* **18**, 432 (2017).
180. Gezan, S. A., Osorio, L. F., Verma, S. & Whitaker, V. M. An experimental validation of genomic selection in octoploid strawberry. *Hortic. Res.* **4**, 16070 (2017).
181. Daetwyler, H. D., Calus, M. P. L., Pong-Wong, R., de los Campos, G. & Hickey, J. M. Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. *Genetics* **193**, 347–365 (2012).
182. Desta, Z. A. & Ortiz, R. Genomic selection: genome-wide prediction in plant improvement. *Trends Plant. Sci.* **19**, 592–601 (2014).
183. Piaskowski, J. et al. Genomic heritability estimates in sweet cherry reveal non-additive genetic variance is relevant for industry-prioritized traits. *BMC Genet.* **19**, 23 (2018).
184. Hardner, C. M. et al. Prediction of genetic value for sweet cherry fruit maturity among environments using a 6K SNP array. *Hortic. Res.* **6**, 6 (2019).
185. Rosyara, U. R. et al. Fruit size QTL identification and the prediction of parental QTL genotypes and breeding values in multiple pedigreed populations of sweet cherry. *Mol. Breeding* **32**, 875–887 (2013).
186. Vanderzande, S. et al. Providing access to high-quality, genome-wide SNP genotypic data for pedigreed germplasm of outbreeding species demonstrated in apple, peach and sweet cherry. PLoS ONE (2019) (Accepted). Preprint at http://biorxiv.org/cgi/content/short/514281v1. https://doi.org/10.1101/514281.
187. Lankow, J., Ritcher, J. & Crooks, R. *Infographics: The Power of Visual Storytelling* (John Wiley & Sons, New York, 2012).
188. Peace, C., Piaskowski, J. & Vanderzande, S. Visualizing the genetics of elite genomes. In *9th International Rosaceae Genomics Conference Program and Abstracts*; 26–30 June; Nanjing, China. 2018; p 32.
189. Baltes, N. J., Gil-Humanes, J., Cermak, T., Atkins, P. A. & Voytas, D. F. DNA replicons for plant genome engineering. *Plant Cell* **26**, 151–163 (2014).
190. Malnoy, M. et al. DNA-free genetically edited grapevine and apple protoplast using CRISPR/Cas9 ribonucleoproteins. *Front. Plant. Sci.* **7**, 1904 (2016).
191. Mojica, F. J. M., Díez-Villaseñor, C., Soria, E. & Juez, G. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol. Microbiol.* **36**, 244–246 (2000).
192. Gao, L. et al. Engineered Cpf1 variants with altered PAM specificities. *Nat. Biotechnol.* **35**, 789–792 (2017).
193. Pulido-Quetglas, C. et al. Scalable design of paired CRISPR guide RNAs for genomic deletion. *PLoS Comput. Biol.* **13**, e1005341 (2017).
194. Liu, H. et al. CRISPR-P 2.0: an improved CRISPR-Cas9 tool for genome editing in plants. *Mol. Plant* **10**, 530–532 (2017).
195. Demirci, Y., Zhang, B. & Unver, T. CRISPR/Cas9: an RNA-guided highly precise synthetic tool for plant genome editing. *J. Cell. Physiol.* **233**, 1844–1859 (2017).
196. Maghuly, F. et al. Long-term stability of marker gene expression in *Prunus subhirtella*: a model fruit tree species. *J. Biotechnol.* **127**, 310–321 (2007).
197. Santos-Rosa, M., Poutaraud, A., Merdinoglu, D. & Mestre, P. Development of a transient expression system in grapevine via agro-infiltration. *Plant Cell Rep.* **27**, 1053–1063 (2008).
198. Chialva, C. et al. Differential expression patterns within the grapevine stilbene synthase gene family revealed through their regulatory regions. *Plant Mol. Biol. Rep.* **36**, 225–238 (2018).

199. Yancheva, S. D., Vlahova, M., Gercheva, P. & Atanassov, A. Agrobacterium mediated transient expression of B-glucuronidase GUS gene in plum (*Prunus domestica* L.). *Biotechnol. Biotecnological Equip.* **8**, 12–13 (1994).
200. Canli, F. & Tian, L. Assessment of regeneration and transient expression factors for Agrobacterium-mediated transformation of *Prunus salicina* Lindl. *Eur. J. Hortic. Sci.* **74**, 66 (2009).
201. Mante, S., Morgens, P. H., Scorza, R., Cordts, J. M. & Callahan, A. M. Agrobacterium-mediated transformation of plum (*Prunus domestica* L.) hypocotyl slices and regeneration of transgenic plants. *Nat. Biotechnol.* **9**, 853–857 (1991).
202. Urtubia, C. et al. Agrobacterium-mediated genetic transformation of *Prunus salicina*. *Plant Cell Rep.* **27**, 1333–1340 (2008).
203. Ye, X., Brown, S. K., Scorza, R., Cordts, J. & Sanford, J. C. Genetic transformation of peach tissues by particle bombardment. *J. Am. Soc. Hortic. Sci.* **119**, 367–373 (1994).
204. Yao, L. P. et al. Protoplast isolation and development of a transient expression system for sweet cherry (Prunus avium L.). *Sci. Hortic.* **209**, 14–21 (2016).
205. Ochatt, S. J., Cocking, E. C. & Power, J. B. Isolation, culture and plant regeneration of colt cherry (*Prunus avium* x *P. seudocerasus*) protoplasts. *Plant Sci.* **50**, 139–143 (1987).
206. Aldwinckle, H. & Malnoy, M. Plant regeneration and transformation in the Rosaceae. *Transgenic Plant J.* **3**, 1–38 (2009).
207. Srinivasan, C., Liu, Z. & Scorza, R. Ectopic expression of class 1 KNOX genes induce adventitious shoot regeneration and alter growth and development of tobacco (*Nicotiana tabacum* L) and European plum (*Prunus domestica* L). *Plant Cell Rep.* **30**, 655–664 (2011).
208. Matt, A. & Jehle, J. A. In vitro plant regeneration from leaves and internode sections of sweet cherry cultivars (*Prunus avium* L.). *Plant Cell Rep.* **24**, 468–476 (2005).
209. Canli, F. A. & Tian, L. In vitro shoot regeneration from stored mature cotyledons of sweet cherry (Prunus avium L.) cultivars. *Sci. Hortic.* **116**, 34–40 (2008).
210. Dolgov, S. V. & Firsov, A. P. Regeneration and Agrobacterium transformation of sour cherry leaf discs. *Acta Hortic.* **484**, 577–580 (1998).
211. Song, G.-Q. & Sink, K. C. Transformation of Montmorency sour cherry (*Prunus cerasus* L.) and Gisela 6 (*P. cerasus* x *P. canescens*) cherry rootstock mediated by *Agrobacterium tumefaciens*. *Plant Cell Rep.* **25**, 117–123 (2006).
212. Liu, X. & Pijut, P. M. Agrobacterium-mediated transformation of mature *Prunus serotina* (black cherry) and regeneration of transgenic shoots. *Plant Cell Tissue Organ Cult.* **101**, 49–57 (2010).
213. Da Camara Machado, A. et al. Somatic embryogenesis of Prunus subhirtella Autumno Rosa and regeneration of transgenic plants after Agrobacterium-mediated transformation. *Plant Cell Rep.* **14**, 335–340 (1995).
214. Gutiérrez-Pesce, P. & Rugini, E. Influence of plant growth regulators, carbon sources and iron on the cyclic secondary somatic embryogenesis and plant regeneration of transgenic cherry rootstock `Colt' (*Prunus avium* x *P. pseudocerasus*). *Plant Cell Tissue Organ Cult.* **79**, 223–232 (2004).
215. Petri, C., Wang, H., Alburquerque, N., Faize, M. & Burgos, L. Agrobacterium-mediated transformation of apricot (*Prunus armeniaca* L.) leaf explants. *Plant Cell Rep.* **27**, 1317–1324 (2008).
216. Petri, C., Wang, H., Burgos, L., Sánchez-Navarro, J. & Alburquerque, N. Production of transgenic apricot plants from hypocotyl segments of mature seeds. *Sci. Hortic.* **197**, 144–149 (2015).
217. Song, G.-q et al. Engineering cherry rootstocks with resistance to Prunus necrotic ring spot virus through RNAi-mediated silencing. *Plant Biotechnol. J.* **11**, 702–708 (2013).
218. Zhao, D. & Song, G.-q Rootstock-to-scion transfer of transgene-derived small interfering RNAs and their effect on virus resistance in nontransgenic sweet cherry. *Plant Biotechnol. J.* **12**, 1319–1328 (2014).
219. Benson, D. A. et al. GenBank. *Nucleic Acids Res.* **46**, D41–D47 (2018).
220. Silvester, N. et al. The European Nucleotide Archive in 2017. *Nucleic Acids Res.* **46**, D36–D40 (2018).
221. Kodama, Y. et al. DNA Data Bank of Japan: 30th anniversary. *Nucleic Acids Res.* **46**, D30–D35 (2018).
222. Jung, S. et al. 15 years of GDR: new data and functionality in the Genome Database for Rosaceae. *Nucleic Acids Res.* **47**, D1137–D1145 (2018).
223. Finn, R. D. et al. InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.* **45**, D190–D199 (2017).
224. Ashburner, M. et al. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
225. Carbon, S. et al. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* **45**, D331–D338 (2017).
226. Buels, R. et al. JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.* **17**, 66 (2016).
227. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
228. Chen, M. et al. New extension software modules to enhance searching and display of transcriptome data in Tripal databases. *Database (Oxf.)*. https://doi.org/10.1093/database/bax052 (2017).
229. Farinati, S., Rasori, A., Varotto, S. & Bonghi, C. Rosaceae fruit development, ripening and post-harvest: an epigenetic perspective. *Front. Plant Sci.* **8**, 1247 (2017).
230. Lyons, E. & Freeling, M. How to usefully compare homologous plant genes and chromosomes as DNA sequences. *Plant J.* **53**, 661–673 (2008).
231. Van Bel, M. et al. PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Res.* **46**, D1190–D1196 (2018).
232. Lee, T.-H., Tang, H., Wang, X. & Paterson, A. H. PGDD: a database of gene and genome duplication in plants. *Nucleic Acids Res.* **41**, D1152–D1158 (2012).
233. Goodstein, D. M. et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178–D1186 (2012).
234. Wang, Y. et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).
235. Duan, N. et al. Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. *Nat. Commun.* 8, 249 (2017).
236. Zhou, Y., Massonnet, M., Sanjak, J. S., Cantu, D. & Gaut, B. S. Evolutionary genomics of grape (Vitis vinifera ssp.vinifera) domestication. *Proc. Natl Acad. Sci. USA* **114**, 11715–11720 (2017).
237. Gaut, B. S., Díez, C. M. & Morrell, P. L. Genomics and the contrasting dynamics of annual and perennial domestication. *Trends Genet.* **31**, 709–719 (2015).
238. Salazar, J. A. et al. Quantitative trait loci (QTL) and Mendelian trait loci (MTL) analysis in Prunus: a breeding perspective and beyond. *Plant Mol. Biol. Rep.* **32**, 1–18 (2014).
239. Arús, P. et al. New Marker-Based Breeding Strategies for Peach and Other Perennial Crops. in *9th International Rosaceae Genomics Conference Program and Abstracts*; 26–30 June; Nanjing, China. 2018; p 30.
240. Liang, W. et al. Genetic diversity, population structure and construction of a core collection of apple cultivars from italian germplasm. *Plant Mol. Biol. Rep.* **33**, 458–473 (2015).
241. Ying, Z. et al. Transcriptome analysis and transient transformation suggest an ancient duplicated MYB transcription factor as a candidate gene for leaf red coloration in peach. *BMC Plant Biol.* **14**, 13 (2014).
242. Gattolin, S. et al. Deletion of the miR172 target site in a TOE-type gene is a strong candidate variant for dominant double-flower trait in Rosaceae. *Plant J.* **96**, 358–371 (2018).