The multi-site PeachRefPop collection: a true cultural heritage and an invaluable international scientific tool for fruit trees

One sentence summary: design and realization of the PeachRefPop, the first international multi-site reference collection in peach (P. persica): an invaluable tool for scientific studies in perennial species

List of Authors
Marco Cirilli, Sabrina Micali, Maria José Aranzana, Pere Arus, Annarosa Babini, Teresa Barreneche, Marco Bink, Celia M. Cantin, Angelo Ciacciulli, José Enrique Cos-Terrer, Pavlina Drogoudi, Iban Eduardo, Stefano Foschi, Daniela Giovannini, Walter Guerra, Alessandro Liverani, Igor Pacheco, Thierry Pascal, Benedicte Quilot-Turion, Ignazio Verde, Laura Rossini and Daniele Bassi
*equal contribution

Author’s contributions
Plant material: DB, SF, BQ, TB, TP, PA, MJA, IV, AL, DG; host institutions: DB, MC, SF, LR, PA, IE, JT, PD, IV; phytosanitary controls: AB; site characterization: WG; experimental design: MCAMB; phenotypic data analysis: AL, DG, CC; genotypic data analysis: IP, AC, MT, MJA, SM; core collection design and evaluation: SM, IV, MC; manuscript writing: MC, SM; manuscript revision: DB, LR, IV, SM and all authors for respective parts. Conception and coordination: LR and DB.

Affiliation
1Department of Agricultural and Environmental Sciences (DISAA), University of Milan, Via Celoria 2, Milan, Italy;
2CREA, Research Centre for Olive, Citrus and Tree Fruit, Via di Fioranello, 52, Rome, Italy;
3IRTA, Centre de Recerca en Agrigenòmica CSIC-IRTA-UAB-UB (CRAG), Campus UAB, Cerdanyola del Vallès (Bellaterra), 08193 Barcelona, Spain;
4Phytosanitary Service, Regione Emilia-Romagna, Bologna, Italy;
5UMR 1332 BFP, INRA Université de Bordeaux, F-33140 Villenave d’Ornon;
6Hendrix Genetics Research, Technology & Services B.V., PO Box 114, 5830 AC Boxmeer, Netherlands;
7Departamento de Horticultura Instituto Murciano de Investigación y Desarrollo Agrario y Alimentario (IMIDA), La Alberca, Murcia, Spain;
8Hellenic Agricultural Organization ‘Demeter’, Department of Deciduous Fruit Trees, Institute of Plant Breeding and Genetic Resources, Naoussa, Greece;
9Centro Ricerche Produzioni Vegetali, Cesena (FC), Italy;
10CREA, Research Centre for Olive, Citrus and Tree Fruit, Via la Canapona 1 bis, Forli, Italy;
Corresponding authors and contacts
Laura Rossini, e-mail: laura.rossini@unimi.it
Daniele, e-mail: daniele.bassi@unimi.it

Abstract
As sessile organisms, plants evolved a range of adaptive mechanisms to adjust their development and physiology to variable external conditions, particularly in perennial species subjected to long-term interplay with the environment. Exploiting the allelic diversity within available germplasm and leveraging the knowledge of the mechanisms regulating genotype interaction with the environment are crucial to address climatic challenges and assist the breeding of novel cultivars with improved resilience. The development of multisite collections is of utmost importance for the conservation and utilization of genetic materials and will greatly facilitate the dissection of genotype-by-environment interaction. Such resources are still lacking for perennial trees, facing with the intrinsic difficulties of successful propagations, materials exchange and living collection maintenance. This work describes the concept, design and realization of the first multi-site peach reference collection (PeachRelPop: PRP) located across different European countries and sharing the same experimental design. Other than an invaluable tool for scientific studies in perennial species, the PRP is configuring as the first milestone of an international collaborative project for the conservation and exploitation of European peach germplasm resources and, ultimately, as a true heritage for future generations.

Introduction
Since the Roman garden ‘hortus’, fruit tree orchards have represented distinctive features of the Mediterranean rural landscape, a synthesis of the interaction among genotype, environment and human customs (Biasi et al., 2009). The diversity of pedo-climatic conditions and production systems, along with plasticity of the genotype and human traditions has shaped the selection of a multitude of local cultivars. These materials represent a cultural and genetic heritage of generations of farmers and a ‘common good’ to preserve for present and future generations.

As sessile organisms, plants evolved a range of adaptive mechanisms to adjust their development and physiology to variable external conditions, particularly in perennial species, subjected to a long-term environmental exposure and interaction. Climate changes are impacting cultivation environments, raising the need for more resilient cultivars able to maintain performances across variable (and often unpredictable) weather conditions (Varshney et al., 2011; Luedeling,
2012; Ramirez and Kallarackal, 2015). Also, increasing the sustainability of fruit production (particularly in terms of resource demands and disease management) requires leveraging knowledge of the interactions between plants, soil, and environmental factors and how they affect productivity and end-product quality (Coakley et al., 1999; Singh et al., 2013; Parajuli et al., 2018). Peach ([Prunus persica] L. (Batsch)) originated in China (Li et al., 2019), later reaching Persia, the Mediterranean Basin, Europe and the Americas, is now the third most cultivated fruit tree species in temperate regions. Beside its importance as a crop, peach is a recognized model for genetic and genomic studies in fruit trees, representing the ideal system for addressing two main challenges in fruit tree breeding:

1) understanding and harnessing the allelic diversity within available gene pools; noteworthy for peach, the intercompatibility with related species of the Amygdalus subgenus (almond, P. davidiana, P. kansuensis, etc.) has been long considered a source of natural variability, particularly for the introgression of disease resistances (Gradziel, 2002; Foulonget al., 2003). However, interspecific hybrids have had poor applicability in current breeding programs (Cirilli et al., 2017), although new genomic based strategies could change this trend (Serra et al., 2016). Conversely, landraces and local ecotypes could be a source of resilience traits more straightforward to introgress, making their preservation and exploitation a suitable strategy for dealing with the changing climatic conditions.

2) systematic dissection of genotype-by-environment (G × E) and/or by-management (G × E × M) interactions as primary sources of variability for several important quantitative traits (Bassi et al., 2006; Myles, 2013; Chagné et al., 2014). This is a critical point for genetic analyses of complex traits, such as genome-wide association studies (GWAS) where germplasm collections are characterized to identify quantitative trait loci (QTLs) across different environments, or genome-wide selection (GS), used to predict genomic estimated breeding values.

The comprehension of genetic, epigenetic and physiological mechanisms as well as the estimation of G × E and/or G × E × M effects requires the development of multisite replicated collections and ad hoc experimental designs. The availability of such type of resources is rapidly growing in annual species, while it has not yet been implemented in perennial trees.

During the past century, peach orchard systems have changed dramatically following innovations in orchard design, training systems and agronomic management (Fideghelli et al., 1988; Corelli-Grappadelli and Marini, 2008), other than cultivar evolution. Noteworthy, the first reported 'modern' orchard was a peach plantation established in Massa Lombarda (Ravenna, Italy) at the end of the 19th century using the white fleshe local cultivar-population ‘Buco Incavato’ (Bellucci, 1908). In the last decades, considerable breeding efforts have assisted the intensification of cultivation techniques and the development of horticultural quality concepts with the introduction of novel, fit-for-purpose cultivars (Byrne et al., 2009). In Europe, peach has a long cultivation history, tracing back to the Ancient and Middle ages, and characterized by the isolation and propagation of
chance seedlings operated by farmers, through which each country has set its own pool of locally adapted cultivars (Bassi and Layne, 2009). The paradigm shift to the modern controlled-crosses approach in early US breeding programs has been the foundation of the dramatic varietal improvement of the last century, beginning with the introduction of seedling materials from China in the mid-19th century (e.g. 'Chinese Cling', progenitor of most modern cultivars) (Faust and Timon, 1995; Byrne et al., 2009). The worldwide spread of improved US materials, favored also by the limited activities in other countries, has resulted in a rapid replacement of landraces and local accessions, particularly in Europe. From the second half of the 20th century, however, novel programs started in several European countries, although they were mostly based on US breeding stocks with a marginal role for local cultivated germplasm. This led to a consequent loss of many local cultivars, in parallel with a progressive narrowing of the genetic bases in modern cultivars (Aranzana et al., 2010; Verde et al., 2013).

As awareness of genetic erosion in modern plant breeding increased (Fu and Dong, 2015) the conservation and exploitation of genetic resources has become a fundamental aspect in crop breeding (Ford-Lloyd and Jackson, 1986). Considerable efforts have been made in the collection and characterization of many plant germplasms (including fruit tree species), along with the development of approaches for their effective management and utilization (Gepts, 2006). The concept of 'core collection', a subset of a germplasm collection of a species that captures most of the genetic diversity while reducing redundancy, has represented an ideal solution for reducing costs and increasing the efficiency of conservation programs (Frankel and Brown, 1984). Several allocation methods have been developed for selecting core collections, attempting to maximize allelic richness or allele coverage (MSTRAT, PowerCore, GenoCore), minimize or maximize genetic distance (GDOpt, SimEli) or simultaneously accommodating for multiple criteria (Core Hunter) (Gouesnard et al., 2001, Kim et al., 2003; Thachuk et al., 2009; Odong et al., 2011; Krishnan et al., 2014). However, the effectiveness of the sampling strategies varied depending on the objective of the core collection, the statistical approach for its definition and the measures for evaluating its quality (Odong et al., 2013). Furthermore, beyond statistical considerations, other aspects are often considered by the institutions hosting the collection, such as historical and socio-economic importance, relevance for breeding activities, popularity among growers and consumers, or distinctive phenotypic characteristics.

In *P. persica*, the absence of wild or feral populations makes ex situ collections the main valuable reservoirs of allelic variability for many traits not yet exploited in current breeding programs. Remarkable progress has been achieved in the phenotypic and genotypic characterization of peach genetic resources (Badenes et al., 2015; Cirilli et al., 2018; Yu et al., 2018), taking advantage of genome sequencing and the development of cutting-edge genotyping tools (Verde et al., 2012; Verde et al., 2017, Aranzana et al., 2019). In the framework of the European collaborative project FruitBreedomics (Laurens et al., 2018), a coordinated characterization of peach collections was
accomplished across relevant European repositories (Micheletti et al., 2015; Hernandez-Mora et al., 2017), promoting increased utilization of resources and encouraging the sharing of conservation responsibilities. For example, the Prunus Working Group within the Fruit Network in the European Cooperative Programme on Plant Genetic Resources (ECPGR) is dealing with Prunus, including peach (Benediková and Giovannini, 2013). Nevertheless, long-term maintenance of collections remains particularly challenging due to intrinsic vulnerabilities (e.g. direct exposure to environmental variables and pathogens) and costs for in vivo maintenance through vegetative propagation to preserve the original genotypes. Moreover, compliance to phytosanitary requirements hampers the sharing of resources among institutions, each having its own stock of materials, resulting in redundancies or risk of loss for unique accessions.

This article describes the concept, design and realization of the first multi-site peach reference collection (named as PeachRefPop: PRP) across five locations in three European countries (Italy, Spain and Greece). Other than an invaluable tool for scientific studies, the PRP is configuring as the first milestone of an international collaborative project for the conservation and exploitation of European peach germplasms resources and, ultimately, as a true heritage for future generations.

Results

1. Criteria for construction of a reference panel of peach accessions and progenies

The PRP collection was built with the aim of selecting a reduced germplasm pool, reflecting the original genetic and phenotypic diversity (Figure 1) and the cultural and socio-economic value of peach cultivation, for its exploitation in future breeding programs. A four-step procedure was followed (exemplified in Figure 2):

1. Definition of the collection domain
2. Establishment of the number of entries
3. Identification of the criteria for entries selection
4. Choice and allocation of the entries

1.1 Definition of the PeachRefPop domain

In order to build a resource representing the European peach diversity and breeding history, the starting point was the genetic material characterized in the framework of the FP7 European project FruitBreedomics (http://fruitbreedomics.com/) in a coordinated effort involving different universities and research institutions across Europe and China. A total of 1,580 Prunus accessions (comprising P. persica and its hybrids with P. davidiana and almond), were phenotyped and genotyped with the IPSC 9k SNP array, as previously described (Micheletti et al., 2015). The inclusion of only peach (including P. ferganensis, Verde et al. 2012) among all the available Prunus accessions was the leading concept behind the definition of the PRP reference collection. Indeed,
as a consequence of many factors (genetic diversity, evolution history, mating system, geographical
distribution etc.), sampling strategies for the inclusion of wild relatives (e.g. species of Amygdalus
subgenus) may substantially differ from those for a cultivated species (e.g. peach) (Brown and
Marshall, 1995). Moreover, to avoid limitations on the exchange of plant material, the domain was
restricted to the European repositories. Based on these criteria, the starting panel for building the
PRP amounted to a total of 1,262 P. persica accessions. Besides accessions, progenies from
controlled crosses also represent a valuable source of informative materials for both genetic analysis
and breeding (or pre-breeding) activities. For this reason, 1,467 individuals from 18 progenies and
their parents (including an interspecific cross with a P. davidiana accession), also analyzed during
the FruitBreedomics project (Hernandez Mora et al., 2017), were considered in the construction
process.

1.2 Establishment of the PeachRefPop size

The definition of the size is one of the most critical decisions for the establishment of a reference
population. For fruit tree crops, the costs of in vivo maintenance are particularly onerous, and
together with long-term space availability in the field, the main limiting factor of running a germplasm
collection. In the perspective of analyzing the interactions between genotype and environment and/or
management practices, or performing genetic studies such as GWAS and GS, an adequate panel
size and experimental design are key factors for the power and reliability of statistical analyses. On
the other hand, for agrobiodiversity conservation purposes, the least number of accessions to include
in a core set depends on the level of genetic repetitiveness present in the original germplasm pool.
The first step towards the establishment of the PRP size was the assessment of the allelic richness
and redundancy observed at marker loci. Two series of core collections of incremental size were
generated, one based on the genetic diversity (Maximization method, OPT), the other through
random sampling (RAN). The maximization procedure (M strategy by Schoen and Brown, 1993), is
based on the sampling of the total allelic diversity observed at marker loci in the least number of
entries. By plotting the genetic diversity measured over the core size, a convex curve was obtained,
indicating the presence of redundancy across the European peach germplasm collection. The
inflection point, corresponding to a plateau in the increase of diversity, was observed at the level of
core 26. At this core size, 99.9% of the total genetic diversity was captured in the core obtained with
the M method in comparison to 93.5% with random sampling (Figure 3). The outperformance of the
optimized versus the random selection was observed across all the core sizes, indicating that the
Maximization strategy was more efficient and was to be preferred for conservation purposes in our
germplasm.

According to some recent works in peach (reviewed in Aranzana et al., 2019), a number of about
100 – 150 unrelated accessions usually provides a satisfactory resolution for identifying major loci
or developing prediction models. In light of all the above premises, an ideal target number of 400
entries was deemed adequate for allocating a minimum of 150 accessions and a maximum of 250 seedlings from progenies (including the parents) based on the outputs of selection criteria.

1.3 Identification of the selection criteria
In spite of the genetic redundancy observed and excluding the rare cases of synonymy, the vast majority of the accessions are not overlapped across the various collections, being conserved for a multitude of reasons and purposes, including scientific research, agrobiodiversity preservation or support to breeding activities. To reconcile these reasons with the aim of creating a feasible, usable and multi-purpose reference collection to be shared among European institutions, a mixed approach has been considered for selecting the accessions. A subset of entries was sampled by using an analytical strategy, based on the criteria of maximizing genetic (and phenotypic) diversity, also taking into account the availability of whole genome re-sequencing data (WGRS); the remaining entries were selected using an empirical strategy, leveraging the knowledge of an experts panel (e.g. breeders, experienced researchers and curators of each repository) and considering the traditional and historical value at national and/or regional levels, the relevance for breeders, growers and consumers, and particular agronomic or pomological characteristics. Moreover, in order to maintain a balanced representation of the genetic structure of the whole collection, the empirical selection of accessions was partially supported by information on population structure (Structure and PCA analysis available from Micheletti et al., 2015). Complementing the choice of accessions, progenies were selected based on the availability of detailed genotypic and/or phenotypic information, genetic background, scientific relevance and, above all, priority traits for breeding.

1.4 Choice, evaluation and description of the PeachRefPop accessions
Capturing the maximum amount of genetic diversity present in the entire collection while reducing redundancy was the primary driver for sampling the first PRP subset (the core set). For this purpose, the advanced M method, implemented in the software PowerCore (Kim et al., 2007) through a modified heuristic algorithm, was used to select a core from the initial panel of accessions, based on a set of 3,894 filtered SNPs previously described in Micheletti et al., (2015). After superimposing 17 accessions with available whole-genome re-sequencing data, an ideal core of 69 accessions (PwC_69) was extracted, representing a sampling size of 5.5% (Supplemental File 1). Considering the many variables that could affect the actual availability of materials for grafting, a flexible approach was further developed to rank each accession of the whole panel based on genotypic and phenotypic diversity. Four different sets made up of 100 cores of 70 entries each were constructed with MSTRAT by setting different combinations of genotypic (9 subsets of SNPs extracted approximately every 1.8 Mb in order to avoid linkage between them) and phenotypic data (7 qualitative and 10 quantitative traits, following transformation of the latter into categories) (Supplemental File 2). Accessions were ranked in groups according to the average frequency of inclusion across the four sets
(Supplemental File 3). Combining the core population extracted by PowerCore with the MSTRAT ranking list resulted in a shortlist of 69 accessions (41 and 28, respectively, indicated as Core._69), ensuring the inclusion of the maximum possible level of genetic diversity. For the completion of the final PRP panel, the remaining 100 accessions (Priority._100) were empirically selected by experts, following the above specified criteria.

Estimates of genetic diversity were used to compare the starting panel of 1,262 European accessions (EU._1262), the core collection obtained by PowerCore (PwC._69) and the final PRP, composed by joining Priority._100 and Core._69 subsets. In addition, Core Hunter software was used to create additional core sets, either of 69 and 169 entries, based on the optimization of various criteria, including allelic coverage (CV._169) and three distance-based algorithms A-NE (AN._69 and AN._169), E-NE (EN._69 and EN._169) and E-E (EE._69 and EE._169). Concerning parameters accounting for allelic diversity, all sets showed high and similar values for the allelic coverage (CV), while the number of effective alleles (N_e) and expected heterozygosity (H_e) were slightly lower for the Priority._100 subset (Table 1). The Shannon-Weaver diversity index (SH) was comparable among the different subsets, ranging between 0.595 in EE._169 and 0.534 in Priority._100. SH generally displays higher values in the presence of a reduced redundancy (Peet, 1975). In contrast, values of observed heterozygosity (H_o) tended to be more variable, ranging from a minimum of 0.202 in PwC._69 to a maximum of 0.318 in AN._69. According to Odong et al. (2013), distance-based criteria were used for further evaluations, such as the minimization of A-NE distance, particularly indicated for generalist collections (as the PRP), and maximization of either E-E or E-NE, both suitable for core collection representing the extremes of the entire collection. A-NE distance generally tends to decrease along with the increase of core size, being minimized in the AN._169 and AN._69 core sets (0.137 and 0.172, respectively), a priori optimized using this selection criterion. Despite the relative low performance of both Priority._100 and Core._69 (0.188 and 0.195, respectively), PRP sets showed satisfactory values for this index (0.165), most probably as a consequence of the increased size. Regarding E-E and E-NE, PRP (as well as Priority._100) showed lower values, particularly for E-NE distance, indicating the presence of a certain redundancy within the panel.

The population structure of peach germplasm is well represented in the PRP, in agreement with the presence of clusters of breeding-derived accessions (further separated in peach- and nectarine-type groups), Occidental traditional and admixed entries with prevalent Oriental origins (Figure 4A). Structure is also preserved in the other core sets, except for that selected through the E-E distance algorithm, tending to oversample the admixed group (Supplemental Figure 1). PCA was also run to check the distribution of the PRP with respect to the other sets, the first two components explaining 15.9 and 8.4%, respectively, of the total variance detected. In the scatter plot, 95% confidence ellipses show almost overlapping areas (except for EE._169), confirming that the PRP panel is well distributed to represent the structure of the starting germplasm (Figure 4B).
Finally, Neighbor-joining (NJ) tree, based on the dissimilarity matrix between the whole panel of 1,262 accessions, was also built to assess the distribution of PRP accessions (Figure 4C). A number of accessions of historical and regional importance, mostly belonging to the Occidental traditional cluster were included. For example, French cultivars dating from late Middle Age (‘Grosse Mignonne’, ‘Millecoton de Septembre’, ‘Reine des Verges’, ‘Brugnon Violet’) (Okie et al., 2008), traditional non-melting Spanish cultivars (‘Amarillo de Agosto 1’, ‘Calante’, ‘Campiell’, ‘Jesca’, ‘Groc Abel’, ‘Groc Alto’) (Badenes et al., 1998; Wünsch et al., 2006) and the Italian ‘Crasiomolo Rosso’ (a white fleshed nectarine belonging to the ‘Sbergie’ type) and ‘Poppa di Venere’, firstly reported at the end of eighteenth century (Majoli, 1790 - 1810). The richness of the Italian peach germplasm is also widely represented by materials from several regions, including Sicilia (‘Imera’, ‘Tardiva di Ficarazzi’, ‘Settembrina di Bivona’, ‘Gialla di Moavero’) (Marchese et al 2005), Campania (‘Zingara Nera’), Puglia (‘Percoco di Turi’), Liguria (‘Michelini’), Emilia-Romagna (‘Buco Incavato’, ‘San Varano 2’ and ‘San Varano 3’, ‘Rosa del West’, this last used for the preparation of the famous cocktail ‘Bellini’) and Tuscany (‘Regina di Londa’) (Gallesio, 2003; Monte et al., 2006; Liverani and Giovannini, 2016). Early breeding materials, mainly from US programs and funders of most of the currently cultivated materials are also included, along with commercial cultivars of worldwide diffusion (Supplemental File 4).

Finally, PRP accessions embrace a wide range of phenotypic variability for traits related to fruit quality, resistance or tolerance against major diseases (brown rot, powdery mildew, leaf curl, aphids and Sharka disease), tree habitus and phenology (Figure 5 and Supplemental File 4).

1.5 Choice and description of the PeachRefPop progenies

Seedlings from 15 cross populations from the research and breeding activities of some European universities and institutions were also added. Most of these accessions were already described in depth (Hernandez-Mora et al., 2017). The leading criterion for the choice of breeding materials was the effective segregation of priority traits in peach, mainly related to phenotype (fruit developmental period, maturity date), fruit quality (fresh weight, soluble solid content, titratable acidity, texture and aroma) and disease resistance (brown rot, powdery mildew, green peach aphids and Plum Pox Virus) (Table 2). A range of breeding materials was considered, such as F1, F2, BC1 populations as well as hybrids with P. davidiana, particularly interesting as a source of PPV resistance (Decroocq et al., 2005). Within each selected cross-population (except for Sf x G and CREA-Forli progenies), a distance matrix estimated from IPSC 9K SNP data was used to support the choice of the seedlings (data not shown).

2. Experimental design and orchard sites description

Accessions and seedlings were propagated through grafting on a common ‘GF677’ rootstock by the same nursery. All accessions and seedlings were grafted on the same year (August 2015) to obtain
trees with the same age. In order to ensure an adequate compromise between the number of replicate trees and sustainable costs of maintenance, an augmented design with replicated checks was adopted (Figure 6A). Accessions and seedlings were assigned to 8 subgroups (i.e. ‘a1’ to ‘a4’ for accessions and ‘s1’ to ‘s4’ for seedlings) and arranged in two blocks:

- the M1 block, which includes two copies of the entire collection of 169 accessions plus 20 cross parents (the ‘a’ subgroup) and 214 seedlings (the ‘s’ subgroup). Replicates are randomly arranged in two separate sub-blocks. The composition of the M1 block is the same across all sites but the order is randomized.

- the M2 block, which includes two partial replicates of the collection (85 accessions, 10 cross-parents and 112 seedlings). Partial replicates are randomly arranged in two separate sub-blocks, as for the M1 one. According to a pair wise scheme, one ‘a’ and one ‘s’ subgroup is common to at least two sites.

- the accessions ‘Big Top’, ‘Springcrest’ and ‘Nectaross’ were set as checks, randomly distributed over M1 (10 replicates each) and M2 (7 replicate each) blocks.

A full copy of the PRP was planted in orchards of 4 institutions from 3 countries (Greece, Italy and Spain) in addition to a partial one hosted by CREA-Roma (Italy) (Figure 6B), and including only the ‘a’ subgroup without block randomization (Figure 6B).

The site of the Institute of Plant Breeding and Genetic Resources (IPB&GR) in Naoussa (Imathia region, Greece) is located at geographical coordinates 40°37’ N, 22°06’ E and 119 m altitude. The area is characterized by dry summers and enough chilling hours. The average annual temperature is 15.5° C, with an average minimum and maximum of 10.8° and 21.6° C, respectively. The average annual precipitation is 724 mm (mostly concentrated in autumn-winter period). The soil is sandy-loam, neutral reaction (pH 6.8), low carbonate content and 2.50% of organic matter (average measures from soil depth until 45 cm).

The site of the Institute of Agrifood Research and Technology (IRTA) in Gimeneells (Catalonia region, Spain) is located at geographical coordinates 41°65’ N, 0°39’ E and 259 m altitude. The average annual temperature is 14.3° C, with an average minimum and maximum of 2.3° and 27.4° C, respectively. The average annual precipitation is 349 mm. The soil is sandy-loam, sub-alkaline reaction (pH 7.7), very high carbonate content and 2.65% of organic matter.

The site of the Instituto Murciano de Investigacion y Desarrollo Agrario y Alimentario (IMIDA) in Mula (Murcia region, Spain) is located at geographical coordinates 38°3’55.595 N, 1°25’42.931 O and 278 m altitude. The average annual temperature is 17.8° C, with an average minimum and maximum temperatures of 3.2° and 29.7° C, respectively. The average annual precipitation is 308 mm. The soil is clay, sub-alkaline reaction (pH 7.8), very high carbonate content and 2.65% of organic matter.

The site of the Centro di Ricerca per le Produzioni Vegetali (CRPV) in Imola (Emilia-Romagna region, Italy) is located at geographical coordinates 44°33’ N, 12°33’ E and 53 m altitude.
The average annual temperature is 13.9° C, with an average minimum and maximum of 12.5° and 25.1° C, respectively. The average annual precipitation is 766 mm. The soil is silty-loam, neutral reaction (pH 7.2), moderate carbonate content and 1.47% of organic matter.

The site of the Research Centre for Olive, Citrus and Tree Fruit of Rome (CREA, Italy), is located at 41°47′48.0″N 12°33′58.1″E and 79 m altitude. The average annual temperature is 16.8° C, with an average minimum and maximum of 12.1° and 21.8° C, respectively. The average annual precipitation is 792 mm. The soil is sandy-loam, sub-alkaline reaction (pH 7.7), no carbonate content and 1.9% of organic matter.

Discussion and Perspectives

The concept of the PRP arises from the growing awareness about current and common issues on ex situ peach conservation across European institutions. Fluctuations in funds availability and intrinsic constraints of living orchard collections threaten the long-term preservation of diversity resources, causing a progressive loss of valuable materials. Reference or core collections have been designed for several fruit tree species, for example olive (Khadari et al., 2003; El Bakkali et al., 2012; Belaj et al., 2012), grape (Laucou, et al., 2011), cherry (Campoy et al., 2016), apple (Gross et al., 2013; Lassois et al., 2016), apricot (Krichen et al., 2012). Nevertheless, they have mainly been created for improving resource allocation in the context of a single institution or repository. The development of a trans-national and shared strategy is configuring as the most promising opportunity in the conservation approach. Actual establishment of the PRP has required huge coordination efforts, facing with the effective availability of materials, the difficulties of their exchange and the success of clonal propagations (particularly for old, often unique, accessions). The sampling strategy for PRP has been defined to accommodate multiple purposes while maintaining the maximum possible diversity compared to the starting panel. The final panel was assembled by the combination of two different subsets: the first (Core_69), ensuring the preservation of the total allele number with the minimum number of accessions, was extracted by widely adopted maximization strategies, either using a class coverage criterion (PowerCore) or Shannon-Wheaver Index (MSTRAT), this last penalizing redundancy. The second subset accommodating for other scopes (Priority_100) was chosen by experts, with a robust knowledge on the genetic structure in peach providing a reliable criterion for assisting selection. As a whole, genetic analysis supports that PRP composition is highly representative of the diversity of peach germplasms present in European collections, as it retains all the allelic variability present within the starting panel, specifically targets defined genetic clusters according to the genetic structure and includes most relevant phenotypic traits. Indeed, differences among the various sampling strategies were negligible for allelic coverage (CV), expected heterozygosity and SH index, revealing a buffer effect towards optimization criteria. Such effect could be expected, since peach has experienced a severe domestication bottleneck with a reduction of genetic diversity, followed by a strong artificial selection during domestication and modern
improvement (Verde et al., 2013; Yu et al., 2018; Li et al., 2019). This is also reflected in the narrow
genetic bases of peach germplasm available across main European repositories. Thus, the high
level of allelic redundancy allows selecting many different subpopulations able to retain the same
amount of genetic variation. In spite of this, a preliminary validation using distance-based criterion
not used in the selection stage provides satisfactory results for the A-NE index, the most indicative
for evaluating the quality of multipurpose collection (Odong et al., 2013). Conversely, E-E and,
particularly, E-NE indices resulted less optimized, due to a certain redundancy on the Priority_100
subset (i.e. a higher number of genotypes providing unique alleles). This was mainly due to the
inclusion of accessions of traditional and breeding values, respectively belonging to the Occidental
Traditional and Occidental breeding clusters, characterized by a very narrow genetic background.
Clearly, the inclusion of these materials is crucial in the overall perspective of balancing diversity and
usefulness, as they integrated various fundamental qualities such as popularity, prestige, tradition
and breeding. A similar mixed strategy was also recently optimized for creating a core collection for
Swiss pear germplasm (Urrestarazu et al., 2019).

Climate challenges in peach growing areas increase the need for resilient cultivars, able to
maintain productivity while showing an increased capacity for adaptation to sub-optimal conditions.
Nevertheless, resilience and adaptive traits often have a complex inheritance and a strong
interaction with the environment or cultivation practice. The partitioning of phenotypic variation into
genotypic, environmental and their interaction components involves ad hoc experimental designs
and integration of field data on a common set of genetic materials under a range of different
environmental/management conditions. Multi-environment trials (METs) have been extensively used
to study GxExM interactions, carry out GWAS and develop GS models for complex traits in annual
crops (Malosetti et al., 2013; Gutierrez et al., 2015; Zhu et al., 2018; Bustos-Korts et al., 2019).
In contrast, such experimental designs are lagging in fruit trees, largely because of the need for large
and diverse germplasm sets for quantitative genetics analyses and the above-mentioned difficulties
in material propagation and exchange. The PRP aims to fill this gap, as the replicated design and
the different pedo-climatic conditions across sites are particularly indicated for the dissection of
interactions between genotype and environment and/or management practice. Also, the inclusion of
both accessions and progenies allows development and testing of novel statistical approaches for
genomics-assisted breeding, such as joint linkage-association analysis (Yu et al., 2008; Lu et al.,
2010) and genome-wide selection (Resende et al., 2012; van Nocker and Gardiner, 2014).

In perspective, the PRP should fulfill several purposes, from scientific research to education
and traineeship of young breeders. A better understanding of diversity is expected to encourage the
use of broad-ranging germplasm (maybe also in other existing ex situ collections) in breeding
programs. In the last decades, the mission of many agriculture-oriented institutions has shifted from
the traditional focus of establishing horticultural collections to a wider target of preserving germplasm
resources and agricultural heritage (Hammer et al., 2003; Havens et al., 2006). This objective is of
 utmost importance for fruit tree species of ancient cultivation history, such as peach. For these reasons, a number of traditional and local cultivars (either old or relatively modern) has been included in the PRP, as a safeguard of an integral part of the rural landscape and collective memory. Since information and descriptions about local germplasms are scarce and often restricted to cultivation areas, their choice has been directly handled by curators of each repository, with the aid of experienced breeders.

**Materials and Methods**

**Dataset**

A set of 1,262 accessions was selected as representative of the *Prunus* germplasm maintained in collections of four different European countries. The complete list of institutions providing plant materials, SNP genotyping and phenotypic data for seven monogenic traits have been previously described (Micheletti et al., 2015). All data, including phenotyping of 10 quantitative traits, were retrieved from FruitBreedomics database available at http://bioinformatics.tecnoparco.org/fruitbreedomics/ website.

**Construction of core subsets**

The advanced M (maximization) strategy implemented in PowerCore v. 1.0 (Kim et al., 2007) using 3,894 SNP markers, was carried out to extract a core subset able to capture all the alleles observed in the entire collection. The size of the final core collection depends on the level of variability and redundancy present in the whole panel and cannot be set *a priori*. Seventeen kernel accessions with available whole-genome re-sequencing data were superimposed through the 'preferential selection' tool, which retains the accessions defined by the user without validation. The standard M strategy implemented in MSTRAT (Gouesnard et al., 2001) was also applied. MSTRAT algorithm selects a subset of n accessions from the N accessions of the entire collection by maximizing the number of alleles (and/or trait classes) at each locus. The sampling size estimated with PowerCore was set as default parameter and four sets of 100 core collections were constructed by using different combinations of genotypic and phenotypic data. Due to the restraints in the number of variables MSTRAT is able to manage, different subsets of approximately 100 SNPs each were obtained through an *ad hoc* developed Perl script program, by extracting 1 SNP every 1,800 Kbp, corresponding to the max boundary for LD found in some subpopulations of the original plant material (Micheletti et al. 2015). Seven qualitative and 10 quantitative traits (these last transformed into qualitative categories) were used as phenotypic data. For each run, the core size was set to 70 and 100 independent replicates with 100 iterations were generated. The Shannon-Weaver diversity index was used as a second criterion to classify core subsets. Redundancy was assayed through the 'Redundancy' tool implemented in MSTRAT, which samples two different sets of core collections of increasing size, as defined by the user, through the application of the maximization strategy or random sampling. For this analysis a subset of 445 SNP markers was pruned from the whole set of
4271 using Plink v1.07 with a window size of 50, a shift of 7 and a variance inflation factor (VIF) of 
2. Redundancy was assayed in the whole panel of accessions with a step of 5 in the first 100, 5 
repetitions and 50 iterations. The Mixed Replica search algorithm implemented in the Core Hunter II 
software (De Beukelaer et al., 2012) was used to generate a core collection of fixed size (either of 
69 and 169 entries) based on the optimization of the Modified Rogers’ (MR) distance measure 
(Wright, 1978), with a weight of 1.0. For the evaluation of the quality of the different core subsets, 
genetic distance-based criteria were considered: the average genetic distance between all the 
entries of each core collection (E-E); the average distance between each entry and the nearest 
neighboring entry for each core collection (N-E); the average distance between each genotype of 
the entire collection and the nearest entry in each core collection (A-NE). The quality of each 
collection increase for lower value of A-NE (the maximum representation is obtained for AN = 0, 
when each accession is represented by itself or by an identical duplicate), and higher value both for 
E-NE (maximizes the average distance between each selected individual and the closest other 
selected item in the core) and E-E (maximizes the average distance between each pair of selected 
individuals in the core.).

Genetic diversity and population analyses
Genetic diversity measures were performed using GenAlex 6.41 (Peakall et al., 2006) and include: 
number of effective alleles (Ne), the number of equally frequent alleles required to give the observed 
level of heterozygosity), levels of observed (Ho) and expected (He) heterozygosity, and the Shannon-
Weaver index (I_S). Allelic coverage was calculated by the function CV implemented in Core Hunter 
II software. Population structure was inferred using a model-based clustering algorithm ADMIXTURE 
v1.22 (Alexander et al., 2009). From SNP data, the software identifies K a priori genetic clusters 
provided by the user and for each individual it estimates the probability of membership to each 
cluster. A preliminary analysis was performed by inputting successive values of K from 2 to 6. The 
value of K that maximized the predictive accuracy was chosen based on a 10-fold cross-validation 
procedure with 10 different fixed initial seeds (Supplemental Figure 2). Data of Principal 
Component Analysis (PCA) were retrieved from a previous work (Micheletti et al., 2015). The 95% 
confidence ellipses in the scatter plot were estimated using PAST software (Hammer et al., 2001). 
Phylogenetic tree was built from a pairwise genetic distance matrix between individuals clustered 
with NJ method in TASSEL (Bradbury et al., 2007). Bootstrap replicate and tree reconstruction were 
performed in MEGA7 software (Kumar et al., 2016).

Supplemental Materials
Supplemental File 1. PowerCore output.
Supplemental File 2. MSTRAT outputs for the four settings.
Supplemental File 3. Accession ranking by MSTRAT frequencies.
Supplemental File 4. PeachReffPop accessions description.
**Supplemental Figure 1.** Population structure estimated in the core sets AN_169, EE_169, EN_169 and CV_169.

**Supplemental Figure 2.** Predictive accuracy (cross-validation error) of population stratification in both Refpop_169 and EU_1262 as determined by Admixture software.

**Acknowledgments**

We wish to thank Claudio Buscaroli and Martina Lama for field assistance, Remo Chiozzotto for lab assistance and Michela Troggio (Fondazione Edmund Mach) for genotypic analyses. We thank the INRA’s ‘Prunus Genetic Resources Center’ of INRA-Nouvelle Aquitaine-Bordeaux for preserving and managing the peach collections.
**Table 1.** Genetic analysis and parameters for the different core subsets. N<sub>e</sub>: number of effective alleles; SH: Shannon-Weaver diversity index; H<sub>o</sub>: observed Heterozygosity; H<sub>e</sub>: expected Heterozygosity; CV: percentage allelic coverage; MR distance: average Modified Rogers genetic distance; E-E: average entry to entry distance; A-NE: average distance between each genotype of the collection and the nearest entry, E-NE: average distance between each entry and the nearest entry.

<table>
<thead>
<tr>
<th>Set name</th>
<th>N&lt;sub&gt;e&lt;/sub&gt;</th>
<th>SH</th>
<th>H&lt;sub&gt;o&lt;/sub&gt;</th>
<th>H&lt;sub&gt;e&lt;/sub&gt;</th>
<th>CV</th>
<th>EE</th>
<th>EE</th>
<th>EE</th>
</tr>
</thead>
<tbody>
<tr>
<td>EU_1262</td>
<td>1.621</td>
<td>0.547</td>
<td>0.292</td>
<td>0.367</td>
<td>0.995</td>
<td>0.285</td>
<td>-</td>
<td>0.131</td>
</tr>
<tr>
<td>PwC_69</td>
<td>1.675</td>
<td>0.574</td>
<td>0.202</td>
<td>0.39</td>
<td>0.987</td>
<td>0.318</td>
<td>0.203</td>
<td>0.237</td>
</tr>
<tr>
<td>EE_69</td>
<td>1.705</td>
<td>0.587</td>
<td>0.234</td>
<td>0.401</td>
<td>0.992</td>
<td>0.347</td>
<td>0.209</td>
<td>0.229</td>
</tr>
<tr>
<td>AN_69</td>
<td>1.645</td>
<td>0.560</td>
<td>0.318</td>
<td>0.378</td>
<td>0.977</td>
<td>0.286</td>
<td>0.172</td>
<td>0.210</td>
</tr>
<tr>
<td>CV_169</td>
<td>1.638</td>
<td>0.556</td>
<td>0.285</td>
<td>0.375</td>
<td>0.995</td>
<td>0.302</td>
<td>0.163</td>
<td>0.212</td>
</tr>
<tr>
<td>EE_169</td>
<td>1.721</td>
<td>0.595</td>
<td>0.224</td>
<td>0.408</td>
<td>0.994</td>
<td>0.330</td>
<td>0.183</td>
<td>0.191</td>
</tr>
<tr>
<td>AN_169</td>
<td>1.643</td>
<td>0.559</td>
<td>0.300</td>
<td>0.377</td>
<td>0.987</td>
<td>0.290</td>
<td>0.137</td>
<td>0.203</td>
</tr>
<tr>
<td>EN_169</td>
<td>1.683</td>
<td>0.578</td>
<td>0.277</td>
<td>0.394</td>
<td>0.993</td>
<td>0.315</td>
<td>0.175</td>
<td>0.256</td>
</tr>
<tr>
<td>Core_69</td>
<td>1.713</td>
<td>0.593</td>
<td>0.247</td>
<td>0.406</td>
<td>0.988</td>
<td>0.303</td>
<td>0.195</td>
<td>0.212</td>
</tr>
<tr>
<td>Priority_100</td>
<td>1.597</td>
<td>0.534</td>
<td>0.283</td>
<td>0.356</td>
<td>0.979</td>
<td>0.277</td>
<td>0.188</td>
<td>0.179</td>
</tr>
</tbody>
</table>

*PeachRefPop* 1.647 0.563 0.270 0.379 0.988 0.290 0.165 0.180
Table 2. Description of the progenies used for establishing the PeachRefPop collection. Traits abbreviation: FD, flowering date; MD, maturity date; SSC, soluble solid content; FW, fruit weight; BR, brown rot; TA, titratable acidity; SwS, slow-softening texture; PM, powdery mildew; PPV, Plum Pox Virus; GPA, green peach aphid; SH, stony hard texture.

<table>
<thead>
<tr>
<th>Cross (parents)</th>
<th>Acronym</th>
<th>Institution</th>
<th>Type of Progeny</th>
<th>Seedlings #</th>
<th>Trait(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Bolero’ × ‘Oro A’</td>
<td>B × O</td>
<td>UMI - Milan</td>
<td>F1</td>
<td>9</td>
<td>MD, SSC, FW, skin overcolor, aroma</td>
</tr>
<tr>
<td>‘Contender’ × ‘Elegant Lady’</td>
<td>C × EL</td>
<td>UMI - Milan</td>
<td>F1</td>
<td>14</td>
<td>BR, MD</td>
</tr>
<tr>
<td>‘Max 10’ × ‘Rebus 028’</td>
<td>M × R</td>
<td>UMI - Milan</td>
<td>F1</td>
<td>9</td>
<td>MD, TA, SSC, FW, SwS</td>
</tr>
<tr>
<td>‘Sweetfire’ x ‘Garcia’</td>
<td>Sf × G</td>
<td>UMI - Milan</td>
<td>F1</td>
<td>15</td>
<td>MD, TA, SSC, FW, SwS</td>
</tr>
<tr>
<td>‘Belbinette’ × ‘Nectalady’</td>
<td>Bb × Ni</td>
<td>IRTA - Lleida</td>
<td>F1</td>
<td>20</td>
<td>FD, MD, TA, SSC, FW</td>
</tr>
<tr>
<td>‘Big Top’ × ‘Nectarross’</td>
<td>Bt × Nr</td>
<td>IRTA - Lleida</td>
<td>F1</td>
<td>19</td>
<td>FD, MD, TA, SSC, FW</td>
</tr>
<tr>
<td>‘Big Top’ × ‘Armking’</td>
<td>Bt × Ak</td>
<td>IRTA - Lleida</td>
<td>F1</td>
<td>18</td>
<td>FD, MD, TA, SSC, FW</td>
</tr>
<tr>
<td>‘Subirana’ x ‘Feraude’</td>
<td>PN643</td>
<td>IRTA - Lleida</td>
<td>F1</td>
<td>7</td>
<td>Fruit shape</td>
</tr>
<tr>
<td>‘Summergrand’ x ‘P. davidiana P1908’</td>
<td>SD</td>
<td>INRA - Avignon</td>
<td>F1</td>
<td>6</td>
<td>PM, PPV</td>
</tr>
<tr>
<td>‘Zephyr’ × ([‘Summergrand’ (S) × ‘P. davidiana P1908’] × S)</td>
<td>BC2</td>
<td>INRA - Avignon</td>
<td>BC2</td>
<td>13</td>
<td>FD, PM, PPV, TA, SSC, FW</td>
</tr>
<tr>
<td>‘Pamirskij 5’ × ‘Rubira’</td>
<td>P × R</td>
<td>INRA - Avignon</td>
<td>F2</td>
<td>13</td>
<td>PM, GPA, foliage colour</td>
</tr>
<tr>
<td>FRF 1495 x FRF 1148 (Ma 16-03-059)</td>
<td>POP1376</td>
<td>CREA - Forli</td>
<td>F1</td>
<td>17</td>
<td>PM; fruit pubescence</td>
</tr>
<tr>
<td>IFF 983 x Ma 25-01-042</td>
<td>POP1115</td>
<td>CREA - Forli</td>
<td>F1</td>
<td>17</td>
<td>TA, SwS and aroma</td>
</tr>
<tr>
<td>FRF 1695 x FRF 1681</td>
<td>POP1095</td>
<td>CREA - Forli</td>
<td>F1</td>
<td>19</td>
<td>SH</td>
</tr>
<tr>
<td>FRF 813 × FRF 691</td>
<td>POP1039</td>
<td>CREA - Forli</td>
<td>F1</td>
<td>18</td>
<td>skin overcolor</td>
</tr>
</tbody>
</table>

467 Figure legends
468 Figure 1. Overview of the range of phenotypic diversity in the PeachRefPop.
469 Figure 2. Graphical summary of the overall scheme followed for selecting the PeachRefPop collection.
470 Figure 3. Redundancy in starting panel (EU_1262) estimated with MSTRAT software using a set of 445 SNPs. Datapoints represent averaged values over 5 independent repetitions for each size.
471 Figure 4. A) Population structure estimated in the whole panel (EU_1262) and PeachRefPop, as estimated for K (number of a priori cluster) equal to 4; B) PCA analysis of the subsets with a core size of 169 entries. Scores for each accession were obtained from the work of Micheletti et al., 2015. The 95% confidence ellipses in the scatter plot were estimated using PAST software. C) NJ phylogenetic tree. Blue navy squared indicated accessions with traditional and historical value, violet circle indicates the other PeachRefPop accessions, while colors reflect the population structure.
Figure 5. Distribution of main phenotypic traits in the *PeachRefPop* collection.

Figure 6. A) Experimental design and B) Google maps satellite imageries of the established *PeachRefPop* orchards across the different European sites.

References


Belaj, A., del Carmen Dominguez-García, M., Atienza, S. G., Urdíroz, N. M., De la Rosa, R., Satovic, Z., ... & Del Río, C. (2012). Developing a core collection of olive (Olea europaea L.) based on molecular markers (DArTs, SSRs, SNPs) and agronomic traits. Tree Genetics & Genomes, 8(2), 365-378.


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).

Figure 1. Overview of the range of phenotypic diversity in the PeachRefPop.
Figure 2. Graphical summary of the overall scheme followed for selecting the *PeachRefPop* collection.
**Figure 3.** Redundancy in starting panel (EU_1262) estimated with MSTRAT software using a set of 445 SNPs. Datapoints represent averaged values over 5 independent repetitions for each size.
Figure 4. A) Population structure estimated in the whole panel (EU_1262) and PeachRefPop, as estimated for K (number of a priori cluster) equal to 4; B) PCA analysis of the subsets with a core size of 169 entries. Scores for each accession were obtained from the work of Micheletti et al., 2015. The 95% confidence ellipses in the scatter plot were estimated using PAST software. C) NJ phylogenetic tree. Blue navy squared indicated accessions with traditional and historical value, violet circle indicates the other PeachRefPop accessions, while colors reflect the population structure.
Figure 5. Distribution of main phenotypic traits in the *PeachRefPop* collection.
Figure 6. A) Experimental design and B) Google maps satellite imageries of the established *PeachRefPop* orchards across the different European sites.