Combining genetic resources and elite material populations to improve the accuracy of genomic prediction in apple

Xabi Cazenave,[1] Bernard Petit,[1] Marc Lateur,[2] Hilde Nybom (ID) ,[3] Jiri Sedlak (ID) ,[4] Stefano Tartarini (ID) ,[5] François Laurens (ID) ,[1] Charles-Eric Durel (ID) ,[1] and Hélène Muranty (ID) [1]*

[1]Univ Angers, INRAE, Institut Agro, IRHS, SFR QuaSaV, F-49000 Angers, France,
[2]Plant Breeding and Biodiversity, Centre Wallon de Recherches Agronomiques, Gembloux, Belgium,
[3]Department of Plant Breeding, Swedish University of Agricultural Sciences, Kristianstad, Sweden,
[4]Výzkumný a Šlechtitelský ústav Ovocnářský Holovousy s.r.o, Holovousy, Czech Republic, and
[5]Department of Agricultural Sciences, University of Bologna, Bologna, Italy

*Corresponding author: Email: helene.muranty@inrae.fr

Abstract

Genomic selection is an attractive strategy for apple breeding that could reduce the length of breeding cycles. A possible limitation to the practical implementation of this approach lies in the creation of a training set large and diverse enough to ensure accurate predictions. In this study, we investigated the potential of combining two available populations, *i.e.*, genetic resources and elite material, in order to obtain a large training set with a high genetic diversity. We compared the predictive ability of genomic predictions within-population, across-population or when combining both populations, and tested a model accounting for population-specific marker effects in this last case. The obtained predictive abilities were moderate to high according to the studied trait and small increases in predictive ability could be obtained for some traits when the two populations were combined into a unique training set. We also investigated the potential of such a training set to predict hybrids resulting from crosses between the two populations, with a focus on the method to design the training set and the best proportion of each population to optimize predictions. The measured predictive abilities were very similar for all the proportions, except for the extreme cases where only one of the two populations was used in the training set, in which case predictive abilities could be lower than when using both populations. Using an optimization algorithm to choose the genotypes in the training set also led to higher predictive abilities than when the genotypes were chosen at random. Our results provide guidelines to initiate breeding programs that use genomic selection when the implementation of the training set is a limitation.

Keywords: genomic selection; training set design; population combination; germplasm; *Malus domestica*; Genomic Prediction; GenPred; Shared Data Resource

Introduction

Breeding programs in outbred fruit tree crops can take many years before new varieties are released, in part because of the long juvenile phase of the trees. Shortening the breeding cycle length for these crops could help increase genetic gain (van Nocker and Gardiner 2014). The length of the breeding cycle is also a constraint when breeders aim to introgress new traits from distant relatives or genetic resources, because several generations are usually needed before genotypes with the desired traits can be released as varieties.

In both cases, the use of molecular markers is an attractive strategy for the early identification of the most promising selection candidates or genetic resources (Myles 2013). Since most apple breeding programs around the world rely on a limited number of genotypes that are frequently used as parents for variety development (Noiton and Alspach 1996), mating strategies based on molecular markers could be used to broaden the genetic base of elite material (Yu *et al.* 2016) by identifying without

phenotyping promising genetic resources that could be used as novel parents in pre-breeding programs (Crossa *et al.* 2016).

Recently, an approach referred to as genomic selection has particularly gained popularity among plant breeders (Voss-Fels *et al.* 2019), as genetic gain is expected to be higher with genomic selection compared to marker-assisted breeding, especially for complex traits (Xu *et al.* 2020). Genomic selection uses a training population of individuals that have been genotyped and phenotyped in order to estimate marker effects, then allowing the estimation of genomic breeding values of a candidate population that has only been genotyped (Meuwissen *et al.* 2001). The success of genomic selection depends on the accuracy of the predicted breeding values. The size of the training set (Zhang *et al.* 2017; Edwards *et al.* 2019), the relatedness between the training set and the candidates (Clark *et al.* 2012; Lehermeier *et al.* 2014) or the genetic architecture of the trait (Daetwyler *et al.* 2010; Wimmer *et al.* 2013) have been reported as factors affecting prediction accuracy. In fruit tree crops, the potential of genomic

selection has been outlined (Kumar *et al.* 2012a; Nsibi *et al.* 2020) and its implementation could help in efficiently using genetic resources in apple (Kumar *et al.* 2020). However, initiating pre-breeding programs based on genomic predictions can be challenging, in part because maintaining and phenotyping populations of large size is time-consuming and costly. The establishment of a large training set could thus limit the implementation of genomic selection. A way to overcome this limitation could be to combine data coming from several breeding populations, like historical data or data from other countries. This strategy has been explored in animal breeding (Hayes *et al.* 2009; Ibáñẽz-Escriche *et al.*, 2009; Wientjes *et al.* 2015) and more recently in plant breeding (Technow and Totir 2015; Sverrisdóttir *et al.* 2018; Olatoye *et al.* 2020) generally showing little to no gain in prediction accuracy. This observation may result from differences in marker effects between the combined populations, as well as from the differences in relatedness between these populations and the candidates to predict. Prediction models that allow the estimation of population-specific marker effects have been proposed (Karoui *et al.* 2012; Schulz-Streeck *et al.* 2012; Lehermeier *et al.* 2015) but few studies have used such models in plant breeding (Ramstein and Casler 2019; Rio *et al.* 2019). Differences in LD patterns between the combined populations may also explain the low observed gains from combination. In this case, it has been suggested to use high-density marker data in order to ensure consistency in SNP-QTL linkage disequilibrium between populations (de Roos *et al.* 2008).

Multi-population training sets can also be interesting when the genotypes to predict results from crosses between the combined populations. If no phenotypic records are available for the progenies of such crosses (as can be the case when initiating a new breeding program or when crossing selection material with exotic germplasm), one could consider combining populations in order to increase the diversity of the training set (Brandariz and Bernardo 2019). In this context, an optimum proportion of each combined population may exist. If this occurs and the sizes of the populations differ, it can be relevant to use a subset of the populations rather than to combine all the genotypes into a unique training set. When the size of the training set has to be fixed in advance, different algorithms exist to optimize the composition of the training set (Rincent *et al.* 2012; Akdemir *et al.* 2015; Mangin *et al.* 2019; Ou and Liao 2019) and they could be used to choose the genotypes to include in a training set in each of the combined populations (Isidro *et al.* 2015).

In this study, we considered the two abovementioned scenarios to investigate the interest of combining different populations into a unique training set for genomic prediction in apple. First, we assessed the potential increase in prediction accuracy when combining two populations, namely genetic resources and elite material, instead of using only one of the two, either using a standard GBLUP model or a model that allows the marker effects to differ between populations. We then predicted the GEBV of the progenies from crosses between elite material and genetic resources and evaluated the prediction accuracy when using different proportions of genetic resources and elite material in the training set and different strategies to choose the genotypes of this training set.

The objectives of our study were to (1) compare the predictive ability of genomic prediction models using simple or combined training sets within and across populations, (2) investigate the effect of different proportions of the two populations used in a training set and (3) assess the impact of high and medium marker density in these cases.

Materials and methods

In this study, three datasets were used for genomic predictions: the first one (hereafter referred to as the FBo-Hi dataset) regroups data coming from two past European projects while the second dataset (from here on referred to as the REFPOP dataset) contains data from an ongoing European project. The genotypes in both the FBo-Hi and REFPOP datasets are either old varieties (called genetic resources from now on) or progenies from breeding programs (called elite material in this study). Due to the different experimental design between the panels of the two datasets (see below), the FBo-Hi and REFPOP datasets provide an opportunity to investigate the effect of genomic predictions with or without the presence of genotype by environment interactions.

The third dataset contains data from crosses between old and modern varieties, which are part of a pre-breeding initiative engaged at IRHS, INRAE Angers. In this study, this dataset was only used as a validation set (VS).

Plant material
FBo-Hi panel
The panel regroups two different apple populations, from here on referred to as the genetic resources and elite material of the FBo-Hi dataset, for which phenotypic and genotypic data are available. The genetic resources represent European apple germplasm preserved in six European core-collections along with data coming from the EU-FP7 FruitBreedomics (FBo) project described in Laurens *et al.* (2018), while the elite material consists of progenies originating from biparental combinations from six European breeding programs that were gathered for pedigree-based QTL analysis during the HiDRAS (Hi) project (Kouassi *et al.* 2009). A total of 1194 unique genotypes (mainly old dessert apple cultivars) from the genetic resources were genotyped and phenotyped for at least one trait (see below). Similarly, 1018 progenies from 23 biparental combinations were genotyped and phenotyped for at least one trait. The genotypes of the FBo-Hi panel were not replicated across sites, except for some genotypes that were used to adjust phenotypic data.

REFPOP panel
The apple REFPOP is described in detail in Jung *et al.* (2020). This panel consists of 269 cultivars (hereafter referred to as genetic resources of the REFPOP) that are representative of the worldwide apple genetic diversity and of 265 progenies (hereafter referred to as elite material of the REFPOP) originating from 27 biparental combinations from various European breeding programs. Some of the genotypes of the REFPOP panel are also part of the FBo-Hi panel, as 189 accessions and 155 progenies are found in both panels. The panel is replicated across six locations in six European countries with contrasting environments and each genotype is replicated at least twice in each environment. At each site, the genetic resources and elite material are planted in the same orchard.

Panel of hybrids
The panel consists of 348 so-called "hybrids" originating from 10 biparental combinations of approximately the same size. Each combination involved a controlled cross between an old cultivar and an elite cultivar. The mating design is presented in Supplementary Table S1 and involved five old cultivars and five elite cultivars. Each hybrid genotype was represented by only one tree in the orchard in Angers, France.

Genotypic data
Genotyping of the plant material

For both the FBo-Hi and REFPOP panels, the genetic resources were genotyped using the Affymetrix Axiom Apple 480K SNP genotyping array (Bianco *et al.* 2016) and the elite material using the Illumina Infinium 20K SNP genotyping array (Bianco *et al.* 2014). The hybrids were also genotyped with the 20K genotyping array. After filtering, 303,239 SNP markers were retained from the 480K array and 10,295 SNP markers from the 20K array, of which 7060 were common to the 480K array. More information about the filtering and quality check procedure can be found in Jung *et al.* (2020).

Genotype imputation

For the elite and hybrids genotypes, the 20K SNPs were completed to reach the 480K genotyping array density by imputation with BEAGLE 4.0 software (Browning and Browning 2007), which can use a reference set of phased marker data along with pedigree information to improve the imputation quality. The reference panel proposed by Jung *et al.* (2020) was used in this study. Missing SNP marker data in the reference panel were first imputed and haplotypes were phased using the default parameters of BEAGLE 4.0. The phased marker data along with pedigree information inferred in Muranty *et al.* (2020) and updated in an ongoing apple pedigree project (Howard *et al.* 2018) were then provided to the software for the imputation from medium to high-density. Since high-density genotypes were not available for the elite material in the FBo-Hi and hybrids datasets, we could not directly assess the imputation accuracy by comparing imputed and true marker data. Imputation accuracy was estimated by computing the mean of the markers Allelic R^2 (AR^2) values provided by Beagle 4.0 after imputation, where the AR^2 value of a marker is an approximation of the correlation between true and imputed genotypic values based on the posterior probabilities of imputation (Browning and Browning 2007). We considered that a marker was well imputed if its AR^2 value was larger than 0.7.

Phenotypic data
FBo-Hi dataset

The genetic resources were phenotyped between 2012 and 2014 in six European research institutes. Several fruit quality and phenology traits were measured by assessor pairs according to the recommendations of the European Cooperative Programme for Plant Genetic Resources (Watkins and Smith 1982) and the notation for each trait was harmonized between institutes. When available, each institute also provided FBo-Hi notations for the measured traits (Supplementary Table S2).

The elite material was phenotyped between 2003 and 2005 in six European countries. Several traits linked to the productivity and fruit quality were measured. More details about the phenotyping procedure can be found in Muranty *et al.* (2015) and Kouassi *et al.* (2009).

Five traits were phenotyped in both populations: harvest date, fruit over-color, fruit juiciness, fruit acidity, and fruit crispness. As we were interested in combining phenotypic information from both populations, only these traits were used for the genomic predictions. When the traits in the two populations were evaluated using different ordinal scales, a correspondence table between the two scales was created in order to use phenotypic data that could be combined.

REFPOP dataset

Traits related to yield, fruit quality and phenology were measured between 2018 and 2020 using a common protocol in each location. The detailed protocol is described in Jung *et al.* (2021). Harvest date and fruit over-color were the only two traits that were measured in both the FBo-Hi and the REFPOP panels.

Dataset of hybrids

The hybrids were phenotyped in 2019 and 2020. Fruit weight, fruit number, fruit over-color, and harvest date were measured for each tree following the protocol proposed for the REFPOP dataset.

Data adjustment

The phenotypic data of the REFPOP dataset and the genetic resources of the FBo-Hi dataset were adjusted for year and site effects and those of the hybrids were adjusted for year effect, as they were evaluated in only one site. To do so, the estimated marginal means of the raw phenotypic data were computed using the emmeans function of the emmeans package (Lenth 2021). Prior to this adjustment, the phenotypic data of the REFPOP dataset were also corrected for spatial heterogeneity as in Jung *et al.* (2020) using the P-spline ANOVA approach with the PSANOVA function of the SpATS package (Rodríguez-Álvarez *et al.* 2018).

Heritability of genotypic means was estimated in both cases as $H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_E^2}{n}}$, where σ_G^2 is the variance of genotype effects, σ_E^2 is the residual variance and n is the mean number of observations per genotype.

Genomic heritability was estimated as $h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$, where σ_g^2 and σ_e^2 are the posterior means of the genomic and residual variances estimated by the MG-GBLUP model (see below). The standard deviation of these genomic heritability values was also computed from the posterior genomic and residual variances as in Lehermeier *et al.* (2015). In the case of the elite material of the FBo-Hi dataset, the best linear unbiased predictors (BLUPs) of clonal values computed in Kouassi *et al.* (2009) were used as phenotypic values.

Genetic characterization of the populations
Structure and genetic diversity across the populations

The structure of the populations was investigated through a principal component analysis (PCA) performed on the SNP marker data of the hybrids, the genetic resources, and the elite material. A PCA was thus performed for the FBo-Hi dataset and another one for the REFPOP dataset. To reduce the computational time of the analysis, the marker data were first pruned based on linkage disequilibrium using the indep-pairwise function in the PLINK 1.9 program (Purcell *et al.* 2007), by pruning the markers with a pairwise $r^2 > 0.1$ in a 50-kb window. In total, 12,363 SNP markers were retained after pruning from the FBo-Hi dataset and 12,290 markers from the REFPOP dataset. The PCA was carried out with the prcomp function from the R *stats* package. To further assess the differentiation between the elite material, the genetic resources, and the hybrids, pairwise F_{ST} values between the populations were computed using the pairwise.WCfst function from the R *hierfstat* package (Goudet 2005).

The allele frequencies along the genome of the elite material and genetic resources were compared using sliding windows for both datasets. For each chromosome, windows of 2 Mb with a shift of 400 kb were built and the average minor allele frequency of the SNPs included in the windows was computed for the genetic

resources. The frequency of the minor allele in the genetic resources was then computed in the elite material following the same procedure. When the number of SNPs of a window was less than 200, the mean minor allele frequency was set to missing.

Linkage disequilibrium

For each population of the FBo-Hi and REFPOP datasets, the linkage disequilibrium (r^2) was computed as the square correlation coefficient between pairs of markers within a 500 kb distance using the r^2 function of the PLINK 1.9 program. Marker pairs were placed into bins of 500 bp according to their pairwise distance and the LD decay was plotted as the mean r^2 of each bin.

Evaluated scenarios for genomic predictions

We evaluated the interest of population combination for two purposes: in the first case, we predicted a population (genetic resources or elite material) using either the same population (within-population prediction), the other population (across-population prediction) or a training set (TS for short) including genotypes from both populations (combination prediction). In the second case, we predicted the "genetic resources × elite" hybrids previously described with a training set combining genotypes from the genetic resources and elite material populations (Prop_hybrids scenario). In the latter case, we investigated the effect of different proportions of the two combined populations on the predictive ability given different TS sizes.

Genomic prediction models

Two different prediction models were evaluated in this study. For both models, the genomic estimated breeding values (GEBV) of the candidates were calculated using the following mixed model:

$$y = X\beta + Zu + e,$$

where y is a vector of phenotypic values, X is an incidence matrix relating fixed effects to the observations, β is the grand mean, Z is an incidence matrix linking the observations to the breeding values, u is a vector of breeding values for each individual and e is the vector of residuals.

Standard GBLUP model

We used the GBLUP model which is derived from the mixed model presented above with $u \sim N\left(0, G\sigma_g^2\right)$ and $e \sim N\left(0, \sigma_e^2\right)$. σ_g^2 and σ_e^2 are respectively the genetic and residual variances and G is the genomic relationship matrix (GRM, see below). The variance components and the breeding values were estimated with the kin.blup function of the R *rrBLUP* package (Endelman 2011).

Multigroup GBLUP model

A second model that takes the genomic correlation between the populations into account was used. In contrast to the standard GBLUP, this model allows the marker effects to be different (but correlated) between the two populations. Following Lehermeier *et al.* (2015), the model can be written as:

$$\begin{pmatrix} y_{GR} \\ y_{Elite} \end{pmatrix} = \begin{pmatrix} X_{GR} & 0 \\ 0 & X_{Elite} \end{pmatrix} \begin{pmatrix} \beta_{GR} \\ \beta_{Elite} \end{pmatrix} + \begin{pmatrix} Z_{GR} & 0 \\ 0 & Z_{Elite} \end{pmatrix} \begin{pmatrix} u_{GR} \\ u_{Elite} \end{pmatrix} + \begin{pmatrix} e_{GR} \\ e_{Elite} \end{pmatrix}$$

where:

$$\begin{pmatrix} u_{GR} \\ u_{Elite} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{g_{GR}}^2 & \sigma_{g_{GR,Elite}} \\ \sigma_{g_{Elite,GR}} & \sigma_{g_{Elite}}^2 \end{pmatrix} \otimes G \right).$$

Here, X_{GR} and X_{Elite} are incidence matrices relating fixed effects to the observations in each population, β_{GR} and β_{Elite} represent the mean of the two populations, Z_{GR} and Z_{Elite} are the incidence matrices relating the breeding values to the observations in each population, u_{GR} and u_{Elite} are the vectors of breeding values of the genetic resources and the elite material, $\sigma_g^2 GR$ and $\sigma_g^2 Elite$ are the genomic variances of the genetic resources and the elite material, $\sigma_{g_{GR,Elite}}$ is the genomic covariance between the two populations and e_{GR} and e_{Elite} are vectors of residuals for each population. The genomic correlation between the two populations, defined as the correlation of the marker effects of each population, was computed from the estimated parameters as:

$$r_{GR,\ Elite} = \frac{\sigma_{g_{GR,Elite}}}{\sqrt{\sigma_g^2 GR\ \sigma_g^2 Elite}}.$$

The genomic variances and covariances were estimated using a Gibbs sampler implemented in the *MTM* package. The Gibbs sampler was run for 20,000 iterations, the first 4000 being discarded as burn-in. One in every two samples was kept and the genetic parameters were then estimated by computing the posterior means of the remaining samples.

Genomic relationship matrix

For both models, the genomic relationship matrix (GRM) G was estimated as in VanRaden (2008):

$$G = \frac{ZZ'}{2\Sigma p_i(1 - p_i)},$$

where $Z = M - 2p_i$ is the centered matrix of the marker data, M is the matrix of marker data and p_i is the minor allele frequency at locus i. Here, Z corresponds to the marker data of both genetic resources and elite material combined into one dataset, allowing the estimation of the relationships within and across populations. The GRM was calculated with the A.*mat* function of the rrBLUP package.

The GRM used in the models was computed according to the populations used in the training and VSs as presented in Table 1. In the case of within-population prediction, the marker data of the predicted population was used to compute the GRM, while the marker data of the two populations was used to compute the GRM for the across-population prediction and combination prediction cases. To evaluate the influence of the marker density on

Table 1 Composition of the training and validation set in the different scenarios

| Method	Training set	Validation set
WP	Genetic resources	Genetic resources
	Elite	Elite
AP	Elite	Genetic resources
	Genetic resources	Elite
Comb	Genetic resources + Elite	Genetic resources
	Genetic resources + Elite	Elite
Prop_hybrids	Genetic resources + Elite	Hybrids

the predictive ability of the models, the GRM was computed using the high-density marker data and a subset corresponding to the SNP markers of the 20K genotyping array in each case.

Predictive ability

The predictive ability of the different models was evaluated as the Pearson correlation coefficient between the GEBV and the phenotypic values of the individuals in the VS. In a first step, the elite material and genetic resources were considered separately and a fivefold cross-validation scheme was used to randomly split the genotypes between training and VSs within each population, allowing within-population genomic predictions (from here on referred to as WP method). The same VSs were also predicted using all the genotypes in the complementary population (across-population prediction, from here on referred to as AP method) or by combining the genotypes of the two populations into a unique training set (combined-populations prediction, Comb and MG-Comb methods, see below). In this case, both the standard GBLUP (Comb method) and multi-group GBLUP (MG-Comb method) models described above were used. This procedure was replicated 20 times for both the FBo-Hi and REFPOP datasets. The approach is summarized in Figure 1A and the different TS and VS compositions are shown in Table 1. For the Comb method, we also studied the effect of the TS size by combining an increasingly large number of genotypes from the same population as the candidates with all the genotypes from the complementary population. This approach is described in more detail in Supplementary Materials S1.

Proportion of GR/elite in the training set

When predicting the hybrids, the influence of the elite/GR proportion in the TS was also investigated (Prop_hybrids scenario in Table 1). To do so, the total size of the TS was first fixed and the TS was then built with increasing proportions of elite individuals, from 0% to 100% by steps of 20%, and correspondingly decreasing proportions of genetic resources individuals (Figure 1B). From here on, the tested proportions will be noted as "$prop_{elite}$/$prop_{GR}$ proportion," where $prop_{elite}$ and $prop_{GR}$ are values between 0 and 1 that refer to the tested proportions (for example the 0.2/0.8 proportion means that the TS contains 20% of elite genotypes and 80% of genotypes from the genetic resources). For all the proportions, different TS sizes were evaluated: the TS size ranged from 50 to 500 by steps of 50 individuals when the genotypes in the TS were chosen from the FBo-Hi dataset and from 50 to 250 by steps of 50 individuals when using the REFPOP dataset. In this last case, the maximum TS size was smaller than for the FBo-Hi datasets because all the proportions could not be tested otherwise given the smaller number of genotypes in each population of the REFPOP. The individuals in the TS were chosen with 3 different methods, as described below. These methods were applied separately to the elite and GR populations by sampling n_1 genotypes from the elite population and n_2 genotypes from the genetic resources, where n_1 and n_2 depend on the studied proportion and $n_1 + n_2 = n$ is a predefined TS size as described above. The chosen individuals in the two populations were then combined into a unique TS that was used for the prediction of the hybrids The GEBV of the hybrids were predicted with the standard GBLUP
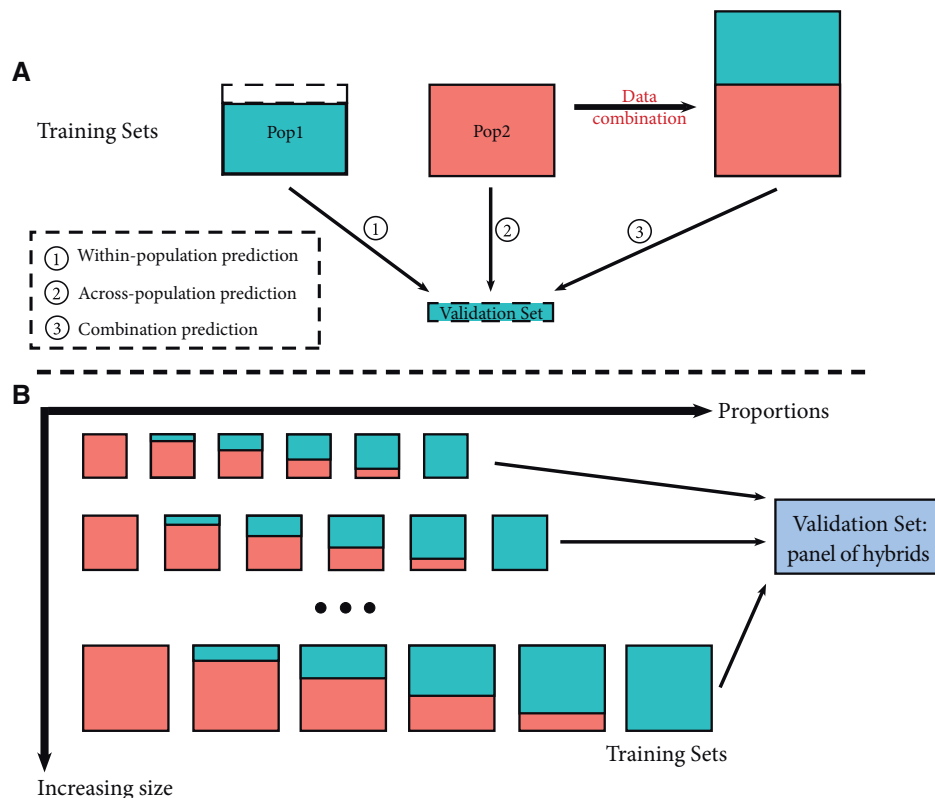
Figure 1 Schematic representation of the tested scenarios in this study. (A) The VS is composed of 20% of one of the two populations (elite material or genetic resources) and is predicted either by the remaining 80% of the population (within-population prediction), by the other population (across-population prediction) or by combining both populations (combination prediction). (B) The panel of hybrids is used as VS and is predicted by combining elite material and genetic resources into a unique population. Several training sets with increasing sizes and different proportions of the two populations are tested.

model described above. In order to compare the predictive abilities obtained with defined training set sizes or with larger training set sizes, the hybrids were also predicted using all the genetic resources and elite material genotypes in the FBo-Hi dataset or the REFPOP dataset.

Method 1: choice of the TS based on the CDpop criterion

The CDpop criterion proposed by Rincent *et al.* (2017) was used to choose the individuals for the TS. CDpop is derived from the coefficient of determination or CD (Rincent *et al.* 2012), which is a measure of the expected reliability of the predictions. The optimization based on CDpop aims to maximize the mean CD of the contrast matrix of the VS to predict, as defined in Rincent *et al.* (2017). For the two populations, an initial number n_1 (respectively n_2) of genotypes was first sampled. An exchange algorithm was then used to replace an individual within each group by an individual from the same population that had not been included yet, and this latter was kept in the group if the mean CD of the group increased after its inclusion. With our data, we found that the mean CD reached a plateau after 1000 to 5000 iterations. Since the choice of the TS based on CDpop is based on the initial sampling, the final TS composition can vary. The method was thus repeated 20 times with 3000 iterations.

Method 2: choice of the TS based on relatedness

For each population, the genotypes were chosen based on their relationship with the VS. To do so, we calculated the mean relationship coefficient between an individual and the candidates from the GRM and chose the genotypes with the highest mean relationship to be part of the TS.

Method 3: choice of the TS by random sampling

In the "Random_strat" method, the TS was constituted by combining n_1 genotypes sampled from the elite population and n_2 genotypes sampled from the genetic resources. The "Random" method is similar except that n genotypes were sampled to constitute the TS, regardless of the population to which they belonged. The two methods were applied 100 times.

Results

Imputation accuracy

We used the AR^2 values provided by Beagle 4.0 as a proxy of imputation accuracy. The distribution of the AR^2 values for the imputed datasets is shown in Supplementary Figure S1. Respectively 81% and 79% of the markers had an AR^2 value higher than 0.7 in the elite material of the FBo-Hi dataset and hybrids dataset. In the case of the REFPOP dataset, Jung *et al.* (2020) reported an imputation accuracy of 0.95 using Beagle 4.0 and the same reference panel as in this study.

Characterization of the populations

In order to investigate the potential of combining genetic resources and elite material into a training set, we first investigated the genetic and phenotypic differences between the two populations. Since the results were very similar between the two datasets, we only present the results obtained with the FBo-Hi dataset. See Supplementary Data for the same analyses with the REFPOP dataset.

Figure 2 shows the first two principal components from the PCA obtained using the pruned SNP marker data. A clear distinction between the two populations can be observed, although some genetic resources (for example Golden Delicious, Jonathan or Cox's Orange Pippin) appear to be closer from the elite material than from the other genetic resources. As expected, the hybrids

were plotted between the two populations. The F_{ST} value computed for the marker data indicated a low differentiation between the genetic resources and elite material (F_{ST} = 0.023, see Supplementary Table S3). The linkage disequilibrium decay was rapid in both genetic resources and elite material (Supplementary Figure S2), the decay being faster in the genetic resources with an average r^2 calculated at 1, 5, 100, and 250 kb of 0.24, 0.2, 0.18, and 0.15 in the genetic resources and 0.29, 0.26, 0.22, and 0.2 in the elite material in the FBo-Hi dataset. Very similar values were observed in the REFPOP dataset. In the two panels, allelic frequencies were similar between elite material and genetic resources for the largest part of the genome (Supplementary Figures S3 and S4), although some major differences could be observed for some genomic regions (for example in chromosomes 3 and 15).

Supplementary Figures S5–S7 show that the phenotypic distribution is different for some traits between the two populations. Fruits phenotyped in the elite material were juicier and crispier than in the genetic resources, and there were more fruits per tree in the elite material. The mean fruit over-color was also slightly higher in the elite material, although the same bimodal distribution could be observed in the two populations. The mean fruit acidity, fruit weight, and harvest date were similar for both populations, but a wider range in acidity score was measured in the elite material compared to genetic resources, while the range in harvest date was larger in the genetic resources. The fruit weight distribution was extremely similar in both populations. In both datasets, the heritability values were generally high (>0.7) for all the traits (Supplementary Figure S8 and Table S2).
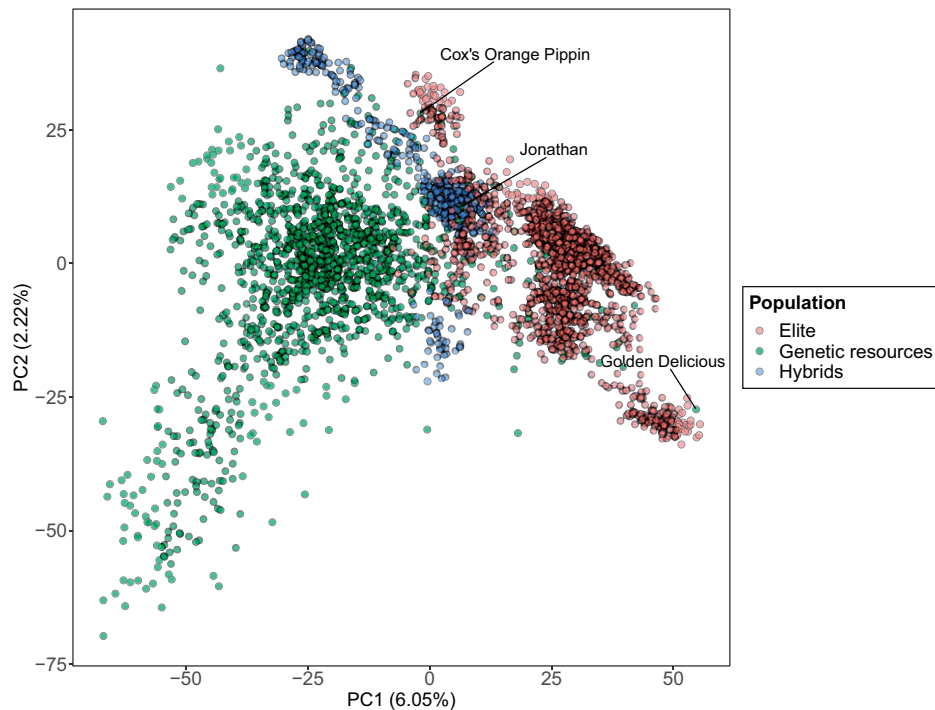
The genomic correlations between the two populations were moderate to high, ranging from 0.42 to 1 (Table 2). The lowest genomic correlations were measured for fruit weight, fruit crispness, and fruit juiciness. The correlation was high for fruit over-color and acidity (around 0.7 in both cases) and was highest for harvest date (in both datasets) and fruit number, with a correlation of almost 1.

Predictive abilities when combining populations

Table 3 presents the predictive abilities for within and across-population predictions, as well as for predictions obtained when combining elite material and genetic resources into a unique training set for the FBo-Hi and REFPOP datasets.

For the within-populations predictions, the predictive abilities varied between 0.33 and 0.82 for the elite material of the FBo-Hi dataset (respectively between 0.46 and 0.79 for the REFPOP dataset), and between 0.34 and 0.83 for the genetic resources of the FBo-Hi dataset (respectively between 0.5 and 0.79 for the REFPOP dataset). In both datasets, predictive abilities were high for harvest date (between 0.79 and 0.83 for FBo-Hi and between 0.6 and 0.79 for REFPOP, see Figure 3A) and fruit over-color (between 0.59 and 0.72 for FBo-Hi and between 0.64 and 0.79 for REFPOP, see Figure 3B), while they were moderately high for the other traits (0.33 to 0.58, Supplementary Figures S9–S13). The predictive abilities were higher when predicting harvest date (with a difference up to 0.19 and smaller standard errors in the REFPOP dataset), fruit weight (difference up to 0.06) and fruit number (difference up to 0.05) in the genetic resources than in the elite material, while fruit over-color (difference up to 0.14) and juiciness (difference up to 0.09) were better predicted when using elite material in the training and VS.

The across-population predictions always resulted in lower predictive abilities than the within-population predictions, with values ranging from 0.13 (juiciness) to 0.64 (harvest date) for the FBo-Hi dataset and from 0.19 (fruit weight) to 0.73 (fruit over-color) for the REFPOP dataset. The standard errors of predictive abilities were also generally higher than for within-population predictions. Across-

Figure 2 PCA performed using pruned marker data of the FBo-Hi and hybrids panel.

Table 2 Genomic and environmental variances and genomic correlations estimated from the Gibbs sampler for each trait in the two population of the FBo-Hi and REFPOP datasets

Trait	Elite			Genetic Resources			
	σ^2_g	σ^2_e	h^2	σ^2_g	σ^2_e	h^2	$r_{\text{Elite, GR}}$
FBo-Hi dataset							
Harvest date	194.2 (20.58)	85.81 (5.92)	0.69 (0.03)	314.5 (25.47)	108.91 (10.37)	0.74 (0.03)	0.99 (0.02)
Fruit over-color	0.6 (0.08)	0.53 (0.03)	0.52 (0.04)	0.8 (0.09)	0.43 (0.04)	0.65 (0.04)	0.65 (0.07)
Acidity	1.5 (0.24)	1.63 (0.1)	0.48 (0.05)	1.1 (0.15)	1.01 (0.08)	0.53 (0.05)	0.74 (0.07)
Juiciness	0.8 (0.15)	1.09 (0.07)	0.42 (0.05)	0.5 (0.07)	0.7 (0.05)	0.43 (0.04)	0.52 (0.12)
Crispness	0.4 (0.1)	1.16 (0.07)	0.27 (0.05)	0.5 (0.08)	0.85 (0.06)	0.37 (0.05)	0.45 (0.13)
REFPOP dataset							
Harvest date	245.8 (46.47)	52.34 (14.25)	0.82 (0.06)	341.2 (38.37)	36.33 (13.84)	0.9 (0.04)	1 (0)
Fruit over-color	1 (0.16)	0.34 (0.06)	0.75 (0.05)	1.3 (0.15)	0.25 (0.06)	0.83 (0.05)	0.72 (0.08)
Fruit number	69.6 (23.01)	112.77 (15.64)	0.38 (0.1)	70.6 (16.03)	47.21 (10.4)	0.59 (0.1)	0.98 (0.03)
Fruit weight	0.4 (0.09)	0.21 (0.04)	0.66 (0.08)	0.6 (0.1)	0.23 (0.05)	0.74 (0.06)	0.42 (0.15)

The standard deviation of the genomic and environmental variances, of the heritability values, and of the genomic correlations is shown between brackets.

population predictive abilities could be moderately high for some traits, such as harvest date (between 0.47 and 0.64 for the FBo-Hi dataset and between 0.41 and 0.47 for the REFPOP dataset) or fruit over-color (between 0.35 and 0.47 for the FBo-Hi dataset and between 0.52 and 0.73 for the REFPOP dataset). All the traits were better predicted when using genetic resources in the training set to predict elite material than when using elite material to predict genetic resources, with the exception of fruit number and crispness that were slightly better predicted in the latter case. The decrease in predictive ability between within and across-population prediction was also more important when genetic resources constituted the VS, especially for fruit over-color (for instance, with a decrease from 0.71 to 0.45 for the FBo-Hi dataset at high marker density), fruit weight (from 0.57 to 0.19) and juiciness (from 0.49 to 0.18).

Combining populations into a unique training set allowed to obtain predictive abilities that were slightly higher than with the corresponding within-population prediction in most cases

(exceptions are some predictions for harvest date, fruit over-color, and juiciness, for which the combination led to the same mean predictive abilitiy). The highest increases in predictive ability were observed for the fruit over-color in the genetic resources in the REFPOP (from 0.69 to 0.73 with medium density). Allowing marker effects to differ between populations (MG-GBLUP model) rarely led to improvements of the predictive ability compared to a prediction using a standard GBLUP model, and when it did, the increase in predictive ability was limited, the highest gain being for juiciness when predicting genetic resources with an increase of 0.02 at both medium and high marker density.

Using a high marker density allowed higher predictive abilities compared to medium density only for harvest date and fruit weight, especially when predicting across populations (for instance, an increase from 0.60 to 0.64 for the FBo-Hi was observed for harvest date when predicting elite material with a training set of genetic resources). For some traits, measured predictive

Table 3 Predictive abilities of the measured traits in the within-population (WP), across-population (AP), combined populations (Comb), and MG-GBLUP method

Trait	Method	Elite material		Genetic resources	
		Medium density	High density	Medium density	High density
FBo-Hi dataset					
Harvest date	WPP	**0.79 (0.03)**	0.79 (0.03)	**0.82 (0.02)**	**0.83 (0.02)**
	APP	0.6 (0.05)	0.64 (0.04)	0.47 (0.05)	0.51 (0.05)
	Comb	0.78 (0.03)	0.79 (0.03)	**0.82 (0.02)**	**0.83 (0.02)**
	MG-GBLUP	**0.79 (0.03)**	**0.8 (0.03)**	**0.82 (0.02)**	**0.83 (0.02)**
Fruit over-color	WPP	**0.72 (0.03)**	0.71 (0.03)	0.59 (0.04)	0.57 (0.05)
	APP	0.47 (0.05)	0.45 (0.05)	0.39 (0.06)	0.35 (0.06)
	Comb	**0.72 (0.03)**	**0.72 (0.03)**	**0.61 (0.05)**	0.58 (0.05)
	MG-GBLUP	**0.72 (0.03)**	**0.72 (0.03)**	**0.61 (0.04)**	**0.59 (0.04)**
Crispness	WPP	0.34 (0.06)	0.33 (0.06)	0.34 (0.05)	0.34 (0.05)
	APP	0.15 (0.07)	0.17 (0.06)	0.21 (0.05)	0.22 (0.05)
	Comb	**0.36 (0.06)**	**0.36 (0.06)**	0.35 (0.05)	**0.36 (0.05)**
	MG-GBLUP	0.35 (0.06)	0.35 (0.06)	**0.35 (0.04)**	0.35 (0.05)
Juiciness	WPP	**0.49 (0.05)**	0.48 (0.05)	**0.4 (0.05)**	0.41 (0.05)
	APP	0.18 (0.07)	0.16 (0.07)	0.13 (0.07)	0.14 (0.07)
	Comb	**0.49 (0.05)**	**0.49 (0.05)**	0.38 (0.05)	0.4 (0.05)
	MG-GBLUP	**0.49 (0.05)**	**0.49 (0.05)**	**0.4 (0.05)**	**0.42 (0.05)**
Acidity	WPP	0.48 (0.05)	0.47 (0.05)	0.45 (0.05)	0.43 (0.05)
	APP	0.39 (0.05)	0.39 (0.05)	0.28 (0.06)	0.23 (0.06)
	Comb	**0.51 (0.04)**	**0.5 (0.04)**	**0.47 (0.06)**	**0.46 (0.06)**
	MG-GBLUP	**0.51 (0.04)**	0.49 (0.04)	**0.47 (0.05)**	0.45 (0.05)
REFPOP dataset					
Harvest date	WPP	**0.6 (0.09)**	0.6 (0.09)	0.79 (0.04)	0.78 (0.04)
	APP	0.44 (0.12)	0.45 (0.12)	0.41 (0.08)	0.47 (0.08)
	Comb	**0.6 (0.10)**	**0.63 (0.10)**	**0.8 (0.04)**	**0.8 (0.04)**
	MG-GBLUP	**0.6 (0.10)**	0.62 (0.10)	0.79 (0.04)	0.79 (0.04)
Fruit over-color	WPP	0.79 (0.05)	0.78 (0.05)	0.69 (0.06)	0.64 (0.06)
	APP	0.73 (0.05)	0.69 (0.06)	0.56 (0.09)	0.52 (0.09)
	Comb	0.81 (0.05)	0.80 (0.05)	**0.73 (0.05)**	**0.68 (0.06)**
	MG-GBLUP	**0.82 (0.05)**	**0.81 (0.05)**	**0.73 (0.05)**	**0.68 (0.06)**
Fruit weight	WPP	0.52 (0.09)	0.52 (0.09)	0.57 (0.08)	0.58 (0.08)
	APP	0.29 (0.11)	0.3 (0.11)	0.19 (0.10)	0.2 (0.10)
	Comb	**0.54 (0.10)**	**0.55 (0.09)**	**0.59 (0.08)**	**0.61 (0.07)**
	MG-GBLUP	**0.54 (0.09)**	0.54 (0.09)	**0.59 (0.08)**	0.6 (0.08)
Fruit number	WPP	0.48 (0.08)	0.46 (0.08)	0.5 (0.11)	0.51 (0.11)
	APP	0.27 (0.12)	0.24 (0.03)	0.27 (0.11)	0.29 (0.11)
	Comb	**0.49 (0.08)**	**0.47 (0.09)**	**0.52 (0.12)**	**0.54 (0.11)**
	MG-GBLUP	0.47 (0.08)	0.45 (0.09)	0.51 (0.11)	0.52 (0.11)

The standard deviation of the predictive abilities is shown between brackets. Bold values represent the highest predictive ability value for a given trait at a given marker density.

abilities were higher with medium marker density than with high marker density, as can be observed for over-color in both datasets and for fruit acidity. The decrease in predictive ability was also more important for across-population predictions for these two traits, especially when predicting genetic resources (decrease of 0.04 for fruit over-color in both datasets and of 0.05 for acidity). For the remaining traits, the influence of the marker density depended on the population used in the VS: predictive abilities for fruit number, juiciness, and crispness decreased when predicting elite material using a high marker density but increased in the genetic resources.

Proportion of the combined populations

When predicting the hybrids, we also investigated the influence of the proportion of elite material and genetic resources used in the training set. Four methods were compared to define the training set: the CDpop algorithm (CDstrat method), a choice based on kinship between the training set and VS (Kinship method), or a random choice of genotypes either within each population (Random_strat method) or regardless of the populations (Random method). The results for harvest date and fruit over-color are presented in Figure 4 and the predictive abilities for fruit

weight and fruit number are presented in Supplementary Figures S14 and S15. The predictive ability values obtained with the different methods for all the traits are also included in the Supplementary File "Prop_predAbi".

We observed that the predictions based on the CDstrat method generally outperformed the Kinship, Random_strat and Random methods, with the exception of the prediction of the fruit number where the Random_strat method allowed to reach slightly higher predictive abilities. The standard errors of the predictive abilities were also lower when using the CDstrat method than when using the Random_strat or Random method. This observation holds true for all the tested training population sizes and at medium and high marker density in the case of harvest date predicted with the FBo-Hi dataset and for fruit weight (except for one tested proportion, see below). When predicting harvest date with the REFPOP dataset, the CDstrat and Kinship methods gave very similar results for all the tested proportions and training set sizes, but CDstrat allowed to reach higher predictive abilities for the 1/0 proportion, especially for small training set sizes (with a maximum predictive ability difference of 0.46 between the two methods for a TS of 50 genotypes). For fruit over-color, the best method depended on the tested proportion: with
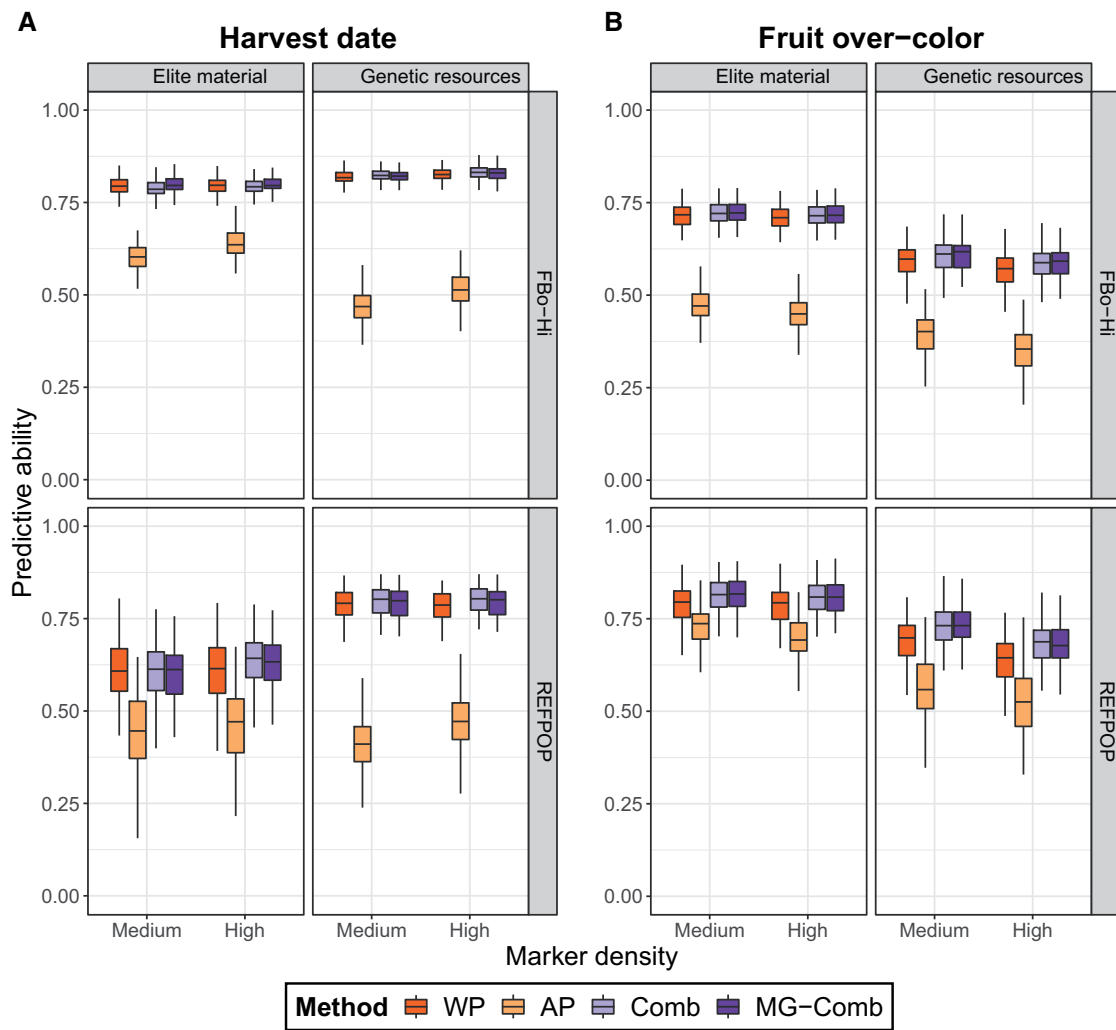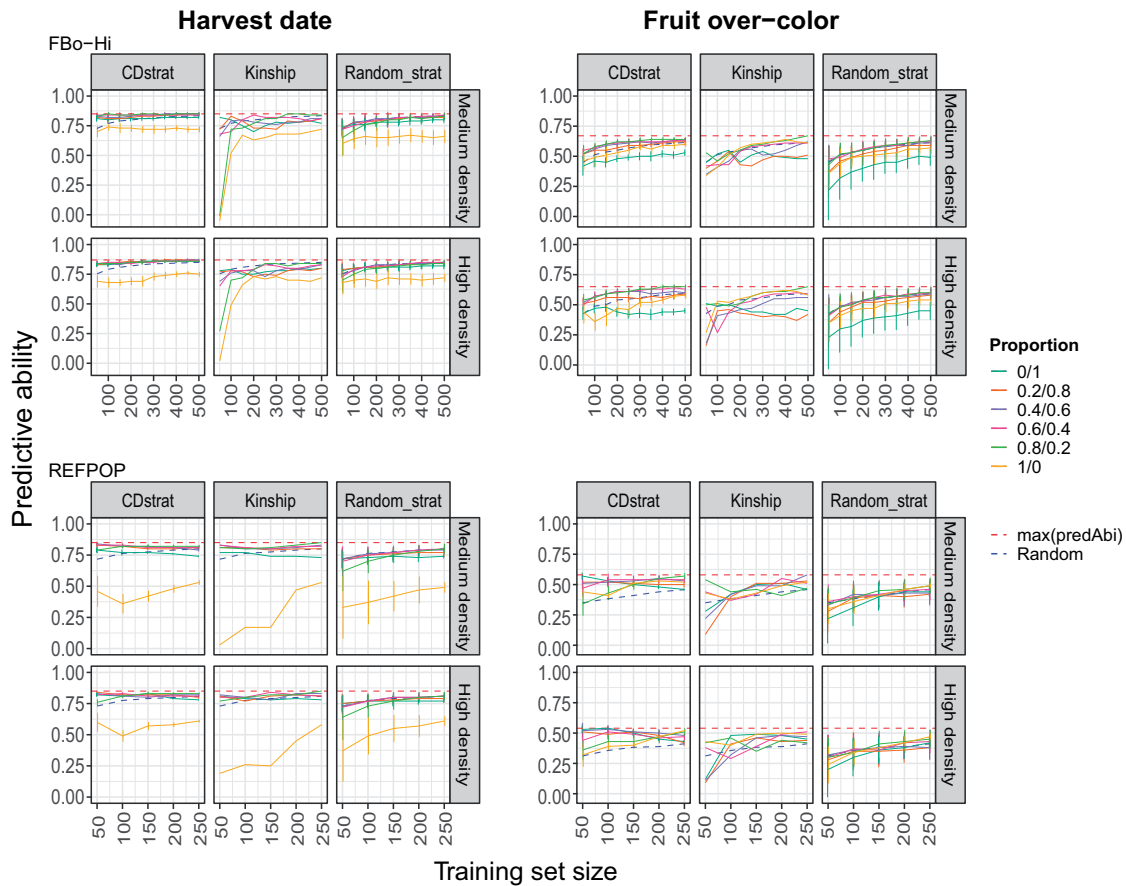
Figure 3 Predictive abilities for (A) harvest date and (B) fruit over-color in the FBo-Hi or REFPOP dataset with medium and high marker density. WP: within-population prediction AP: across-population prediction Comb: combination prediction MG-Comb: combination prediction with the MG-GBLUP method.

the FBo-Hi dataset, CDstrat outperformed Kinship except for the extreme proportions (*i.e.*, TS constituted of only one of the two populations), the highest increases in predictive ability being again measured for small training set sizes. With the REFPOP dataset, CDstrat outperformed Kinship only until a TS size of 100 genotypes, and the two methods gave similar results for larger TS sizes. For all the tested TS sizes, most genotypes included in the TS with the CDstrat method were selected at least 15 times out of the 20 repetitions (data not shown).

The predictive abilities based on the CDstrat method rapidly reached a plateau for harvest date. The maximum predictive ability of 0.85 could be reached with 100 genotypes from the FBo-Hi dataset at medium density, and the maximum predictive ability was reached with 50 genotypes from the REFPOP dataset for both densities. For fruit over-color, a bigger TS size resulted in higher predictive abilities with the FBo-Hi dataset for all but the 0/1 and 1/0 proportions. The increase in TS size also led to an increase in predictive ability only for some proportions for fruit over-color with the REFPOP dataset: increasing the TS from 50 to 250 genotypes allowed an increase in predictive ability for proportions 0.4/0.6 to 1/0 (with increase between 0.02 for the 0.4/0.6 proportion and 0.43 for the 0.8/0.2 proportion) but the predictive ability decreased for the 0/1 and 0.2/0.8 proportions. A similar

pattern was observed for fruit weight: when increasing the TS size, predictive abilities also increased for proportions 0.6/0.4 to 1/0 but decreased for proportion 0.2/0.8 and 0/1 beyond respectively 150 and 100 genotypes in the TS. The predictive abilities for fruit number increased when the TS size increased for all the tested proportions, but the gains were smaller than for the other traits (maximum increase from 0.16 to 0.26 when increasing the TS size from 50 to 250 genotypes). In all the studied traits, the standard errors of the predictive abilities also decreased when the TS size increased.

With the Kinship method, the largest increases in predictive ability were measured for harvest date predicted with the FBo-Hi dataset, especially for the 0.8/0.2 and 1/0 proportions at medium marker density (respectively from −0.01 and −0.05 with a TS size of 50 genotypes to 0.84 and 0.72 with a TS size of 500 genotypes) but not in the REFPOP dataset: the predictions with 50 genotypes in the TS led to better predictive abilities than a TS with 250 genotypes, except for the 0.8/0.2 and 0/1 proportions. The increase in TS size also led to higher predictive abilities for all the tested proportions for fruit number, with larger gains than for the CDstrat method (with a maximum increase from 0.13 to 0.29 when increasing the TS size from 50 to 250 genotypes). The predictive ability also increased for fruit weight and fruit over-color for all

Figure 4 Predictive abilities for harvest date and fruit over-color in the dataset of hybrids with medium and high marker density when the training set is composed of varying proportions of elite material and genetic resources of the FbBo-Hi or REFPOP dataset. max(predAbi): maximum predictive ability obtained for a given marker density with one of the three tested methods regardless of the training set size Random: predictive ability obtained when randomly choosing the genotypes included in the training set.

the tested proportions, except for the 0/1 proportion for fruit weight and the 0.8/0.2 proportion for fruit over-color when predicted with the REFPOP dataset, the maximum predictive ability being observed with a TS size of 50 genotypes in both cases.

For each trait, the proportion allowing the highest predictive ability depended on the TS size and the method used to select the genotypes of the TS. For the three methods, the worst predictive abilities were obtained when only one population was present in the TS, with predictive abilities lower than the ones obtained from random sampling in the two populations: the 0/1 proportion led to the lowest predictive abilities for fruit weight, fruit number and fruit over-color when predicted with the FBo-Hi dataset while for harvest date, the predictions with the 1/0 proportion gave the lowest predictive abilities. Note that for fruit weight and fruit over-color, the 1/0 proportion also led to predictive abilities lower than the predictive abilities obtained with the Random method.

As was observed in the within and across-population predictions, the effect of the marker density on predictive ability depended on the trait: harvest date generally benefited from the higher marker density in the two datasets, especially when the TS size increased. Interestingly, predictive abilities at high marker density with the 1/0 proportion were lower than with medium marker density for small TS sizes when predicting with the FBo-Hi dataset, while the increase in predictive ability due to higher marker density was the highest for the 1/0 proportion when predicting with the REFPOP dataset, with increases ranging from 0.08 to 0.15 with the CDstrat method. Predictive abilities for

fruit weight were higher at high density only for TS sizes up to 100 genotypes and medium density led to higher predictive abilities for larger TS sizes with the CDstrat method, except for the 0/1 proportion, in which case high density always gave better results (with increases in predictive ability ranging from 0.04 and 0.13). Predictive abilities measured from high marker density data for fruit over-color and fruit number were slightly lower than when using medium marker density for every method and every tested proportion, except for the 1/0 proportion for fruit over-color predicted with the REFPOP dataset with the Kinship method which allowed slightly higher predictive abilities until a TS size of 200 genotypes.

Discussion

One possible limitation to the implementation of genomic selection in fruit tree crops is the establishment of a training set large enough to allow accurate prediction of selection candidates because of the possible costs of maintaining and phenotyping trees. In this study, we first evaluated the potential of combining two different apple populations to increase the size of the training set by estimating the predictive ability in this case. We then investigated the effect of using subsets of the two populations to predict hybrids resulting from crosses between genetic resource and elite material. In each case, we compared the predictive abilities obtained using either medium or high SNP marker density.

Predictive abilities within and across populations

When predicting within-population, the measured predictive abilities were moderate to high, ranging from 0.33 for fruit crispness to 0.83 for harvest date. These values were overall in line with previous studies in apple that predicted the same traits. For example, harvest date was well predicted in other studies performed on apple (Migicovsky *et al.* 2016; McClure *et al.* 2018; Jung *et al.* 2020; Minamikawa *et al.* 2021) with predictive ability values between 0.6 and 0.8, while prediction accuracies for fruit weight and fruit quality traits were generally low to moderate (Kumar *et al.* 2015; Migicovsky *et al.* 2016; McClure *et al.* 2018; Minamikawa *et al.* 2021), with predictive abilities not exceeding 0.5. An exception was the study from Kumar *et al.* (2012b) that reported predictive ability values over 0.7 for traits related to fruit quality, in a case where genotypes from seven full-sib families derived from a 4 × 2 factorial design were randomly used for cross-validation, leading to a higher relatedness between training and population set than in our study.

Across-population predictions always resulted in predictive abilities lower than within-population predictions. This is an expected result as linkage disequilibrium, allele effects, and allele frequencies are generally different between populations. In our case, the LD decay was indeed more rapid in the genetic resources than in the elite material. The allele frequencies were similar along the 17 chromosomes but differed for some genomic regions, which could be due to selection, thus leading to different allele effects estimates as discussed in further. For some traits like fruit crispness or juiciness, the different phenotypic distribution in the two populations could also explain the poor predictive abilities in across-population prediction (Supplementary Figures S4–S6). Roth *et al.* (2020) reported a similar result when predicting apple fruit texture of full-sib families using a germplasm collection. The high genomic correlations between the two populations could have suggested that across-population predictions could perform as well as within-population prediction, which was not the case in our study. Similar results were observed by Lyra *et al.* (2018) and Rio *et al.* (2019), who hypothesized that this apparently contradictory observation could be explained by population-specific allele frequencies at the causal QTL. Such differences are expected in our case because of selection (see below). The moderate to high predictive abilities observed for some traits like harvest date and fruit over-color and the high genomic correlations between the two populations nevertheless suggest that marker effects are conserved to a certain extent between genetic resources and elite material. In the case of the FBo-Hi dataset, the lower predictive abilities observed in the across-population predictions could also be due to G×E interactions since the trees of the two populations were phenotyped by different assessors in different sites and years. Combining genotypes that were phenotyped in environments that differ from the environments of the VS is not always detrimental for genomic predictions (Jarquin *et al.* 2016) but when updating the training set, such genotypes should be discarded and replaced by genotypes from the same environment when it becomes possible to do so.

Potential of combining populations

Combining datasets can help increase predictive abilities in genomic prediction by allowing larger training populations, which is a key parameter for accurate predictions. However, the marker effects may be different between the combined populations, in which case the addition of new genotypes to the training set is not expected to improve predictive abilities (Lund *et al.* 2016). For instance, several studies in livestock found no increase in predictive ability when combining breeds (Hayes *et al.* 2009; Erbe *et al.* 2012) and adding genetically distinct individuals to the training population can even result in lower predictive abilities (Lorenz and Smith 2015). Multi-population genomic prediction models combining populations that have diverged for a few generations are then expected to lead to higher predictive abilities (de Roos *et al.* 2009; Wientjes *et al.* 2016) and the choice of the populations to combine should be based on knowledge about the genetic distance between the populations.

Several studies showed evidence of selection during apple domestication, *e.g.*, for fruit size (Yao *et al.* 2015; Duan *et al.* 2017), fruit quality traits (Wedger *et al.* 2021), metabolite content (Khan *et al.* 2014) or disease-related genes (Singh *et al.* 2019). However, little information is available about the genomic consequences of the more recent apple improvement and the resulting genetic differences between genetic resources and modern cultivars. Supplementary Figure S4 shows that the genotypes from the elite material are less acid, juicier and crispier than genetic resources, which can be a consequence of breeders' choices since these fruit quality traits have been targeted for decades in apple breeding programs (Janick and Moore 1975; Liao *et al.* 2021). On average, elite material is also harvested later than genetic resources, most probably because fruits with a later ripening are more firm (Migicovsky *et al.* 2016) and can be stored longer (Nybom *et al.* 2008). Modern varieties also have a lower phenolic content than genetic resources in general (Ceci *et al.* 2021; Watts *et al.* 2021). A recent pedigree reconstruction study showed that the two populations are separated by approximately 5–10 generations (Muranty *et al.* 2020), which could result in the persistence of marker phase and effects across populations.

In this study, combining elite material and genetic resources generally led to small increases in predictive ability compared to within-population predictions. For some traits, the combination allowed to obtain slightly higher predictive abilities in the FBo-Hi dataset, in spite of the genotypes from the two populations being evaluated at different sites and years. This suggests that the potential adverse effect of GxE interactions on predictive ability was counterbalanced by the large increase in population size. The addition of genotypes from the complementary population did not decrease the predictive abilities when combining data, but we can hypothesize that selecting subsets of the two populations to combine closer genotypes into a training set could be advantageous. The approaches proposed in the Prop_hybrids scenario could thus be used for the Comb scenario. Supplementary Figures S16–20 show that the small increase in predictive ability when combining populations could be due to the initial large population sizes for some of the studied traits, in which case the marker effects would already be accurately estimated. Combining populations may thus be of particular interest when one of the two populations is small and the complementary population is large. To further increase the size of the training set in such a case, historical data from trees that only have pedigree information could also be used in a single-step GBLUP model that uses both genotyped and nongenotyped individuals at the same time (Sood *et al.* 2020).

Toward improvements of the predictive abilities

For all the studied traits, the MG-GBLUP model, which allows different but correlated effects between populations, did not perform better than the standard GBLUP model. This result can be explained in the light of the genomic correlations between the two populations: when the genomic correlation between the

combined populations is equal to one, the MG-GBLUP is equivalent to the standard GBLUP. Therefore, the MG-GBLUP and standard GBLUP were expected to yield similar results since the genomic correlation between the genetic resources and elite material was high for most traits. The GEBV obtained from the GBLUP and MG-GBLUP model were almost the same even when the correlation between elite material and genetic resources was moderate (Supplementary Figures S21 and S22). Lehermeier *et al.* (2015) point out that MG-GBLUP is expected to perform better than GBLUP when population used for within-population prediction is small because borrowing information from the combined population could improve marker effect estimation in this case. In this study, the populations used for genomic prediction were large, which could explain why the GBLUP and MG-GBLUP performed similarly even when the genomic correlation between the combined populations was moderate. In addition, the MG-GBLUP shows limitations when dealing with genotypes that cannot clearly be assigned to a given population (Lehermeier *et al.* 2015), which is sometimes the case with the elite material and genetic resources (Figure 2 and Supplementary Figure S23). In this case, genomic prediction models that account for admixture could be better adapted (Rio *et al.* 2020).

Ibáñez-Escriche *et al.* (2009) suggested that models that fit population-specific marker effects may not be necessary at high marker density because a density high enough could allow the marker-QTL association to be the same in the combined populations. In this study, we observed limited increases in predictive abilities when using high-density marker data. Other studies also reported that predictive abilities could reach a plateau after a given number of markers (Hickey *et al.* 2014; Jung *et al.* 2020) and that high marker density was not always needed to obtain high predictive abilities. Considering the rapid linkage disequilibrium decay in the genetic resources, we could have expected the predictive abilities to be lower at medium than at high density when the marker effects are estimated from such a population. However, the increases in predictive ability when using high-density marker data were generally small for all the datasets. This could imply that even at medium density, some markers are in high linkage disequilibrium with the most important causal QTLs and that increasing marker density does not better capture their variation in this case. In the REFPOP dataset, Jung *et al.* found that predictive ability was the same for harvest date and floral emergence at medium or high density and that as few as 5000 markers could be sufficient to reach a similar predictive ability. In the case of the elite material, imputation errors could also explain the absence of predictive ability improvement when using high-density marker data. However, we obtained very similar predictive abilities for all the traits when we removed markers with an AR^2 value under 0.7 (data not shown).

For some traits, the predictive abilities were even lower with high than with medium marker density. Two of these traits, namely fruit over-color and acidity, are controlled by an oligogenic determinism with a few known major genes (Chagné *et al.* 2016; Verma *et al.* 2019). As the GBLUP and MG-GBLUP models make the hypothesis that all the marker effects are drawn from the same normal distribution and do not allow marker effects to be null, a high marker density could overshrink the marker effects and poorly capture the effect of the large QTLs, as suggested by Daetwyler *et al.* (2010). Similar to our case, Erbe *et al.*(2012) found that predictive abilities were lower with a GBLUP model and 800K markers than with the same model with 50K markers and suggested that the number of effects to estimate was too important compared to the number of phenotypic

records. In their case, models that allow marker effects to be set to zero or models resulting in a selection of a subset of markers led to higher predictive abilities. Bayesian models could thus result in higher predictive abilities for traits with an oligogenic genetic architecture (Hayes *et al.* 2009). If major QTLs are known, they could also be considered as fixed effects (Bernardo, 2014; Sarinelli *et al.* 2019) or be weighted accordingly in the GRM (Tiezzi and Maltecca 2015; Raymond *et al.* 2018). Since genomic predictions are generally used for traits with a polygenic determinism, high marker density should give better results than lower marker densities in general. Moreover, DoVale *et al.* (2021) showed that, in outbred crops, genomic prediction models should be updated more regularly when high marker densities are not used. Using high marker density for genomic selection could thus still prove useful when combining populations, especially if a high-quality imputation step can be implemented to reduce the costs of genotyping (Song *et al.* 2019).

Optimization of the training set composition

When predicting the hybrids, we never observed a particular proportion of the two combined populations that would allow higher predictive abilities. However, using only one of the two populations was detrimental for each studied trait, regardless of the training set size or the method used to choose the genotypes. This point highlights the need to use at least some genotypes from both populations in the training set in order to achieve high enough predictive abilities. This observation can have at least two explanations. First, we observed in our data that the phenotypic variation in the hybrids is larger than the variation of the two populations taken separately. When using only one of the two populations in the training set, the range of phenotypic values of the chosen population can by consequence not be large enough to accurately estimate the marker effects. Second, if some alleles linked to the desired trait are in low frequency in one population and segregate in the other population, using only the population with the low-frequency alleles as the training set will lead to incorrect marker effect estimations. For example, Migicovsky *et al.* (2021) showed that the *NAC18.1* gene marker, which is linked to harvest date and fruit firmness, is homozygous for the favorable allele in the nine most marketed varieties in the United States, and several marker alleles detected by GWAS in the REFPOP panel are fixed in the elite population whereas they segregate in the genetic resources (Jung *et al.*, 2021). As expected in this case, the prediction for harvest date in the hybrids when using only the elite material in the training set led to predictive abilities lower than when using both populations. Another such example would be the *Rvi6* gene that confers resistance to apple scab and all QTLs in the associated introgressed segment: the favorable allele at *Rvi6* segregates in elite material (Jung *et al.* 2021) but is absent in old varieties, because the gene introgression from the wild relative *M. floribunda* is recent (Gessler and Pertot 2012).

When the training set size is a limitation for the implementation of genomic selection, methods to optimize the choice of the genotypes in the TS when genomic data are already available can be advantageous. In this study, we observed that for three out of four traits the hybrids were better predicted when the TS was chosen based on the CDpop criterion compared to a choice based on kinship or genotypes sampled randomly, especially for small training set sizes. The CD algorithm has been implemented in order to better sample the genetic diversity than algorithms based on kinship alone (Rincent *et al.* 2012), which could explain the better performance of the CDstrat method, since it is probably

necessary to use a training set with a large diversity to predict the hybrids, as discussed above.

Note that we applied the CD algorithm to each population separately in order to simultaneously study the effect of the proportion of the two populations on predictive abilities. One way to further optimize the composition of the training set could be to use the CDpop criterion with genotypes of the elite material and genetic resources in a single step, letting the algorithm choose the proportion of the two populations to be used. However, using the optimization procedure based on the CDpop criterion is computationally demanding for large populations, as the computation of the criterion requires to calculate the inverse of the GRM of the genotypes in the training and VSs (Rincent *et al.* 2017), which could not be achieved in our case when considering both populations at the same time. It would also be interesting to evaluate the effect of the optimization methods proposed in this study in a case where the hybrids would be predicted by a training set also composed of hybrids. This situation should lead to the highest predictive abilities, and while the combination of populations proposed in this study is a suitable approach to initiate the prediction of hybrids, a training set composed of hybrids only should be envisioned as soon as enough phenotypic and genotypic data for the hybrids are available. Such a training set can be built gradually, by replacing a part of the genotypes of the two populations by newly phenotyped and genotyped hybrids (Fritsche-Neto *et al.* 2021).

Conclusion

We showed in this study that combining genetic resources and elite material into a unique training set could be beneficial for genomic predictions. First, larger training populations can be obtained with this approach, leading to slightly higher predictive abilities in return. Second, using both populations in the training set appeared necessary to predict "genetic resources × elite" hybrids. Combining populations could thus be an effective way to initiate pre-breeding programs that incorporate genomic prediction when a large training population is too costly. Genotypic and phenotypic data from ongoing breeding programs can be used to generate the training set, or historical data can be used to further reduce the costs of phenotyping. The training set composition can be further optimized to reduce the number of genotypes to include, which can be of particular interest for traits that are hard or long to phenotype, like biennial fruit bearing or abiotic stress tolerance. The proposed approach can be used in other fruit tree species provided that the genetic differences of the combined populations are taken into account when necessary.

Data availability

All SNP genotypic data generated with the 480K array used in this study have been deposited in the INRAe dataset archive (https://data.inrae.fr/) at https://doi.org/10.15454/IOPGYF. The SNP genotypic data of the REFPOP elite material generated with the 20K array and used in this study have been deposited in the INRAe dataset archive at https://doi.org/10.15454/1ERHGX. The SNP genotypic data of the elite material of the FBo-Hi panel generated using the 20K array and used in this study have been deposited at https://doi.org/10.15454/PMVCFI. The SNP genotypic data of the hybrids generated using the 20K array in this study have been deposited at https://doi.org/10.15454/SB2JSB. The raw

phenotypic data of the REFPOP panel have been deposited at https://doi.org/10.15454/VARJYJ. The raw phenotypic data of the genetic resources of the FBo-Hi dataset have been deposited at https://doi.org/10.15454/KNECMS, the BLUPs of clonal values of the elite material of the FBo-Hi dataset have been deposited at https://doi.org/10.15454/UJPCOV, the raw phenotypic data of the hybrids have been deposited at https://doi.org/10.15454/H6PXRX. The scripts used to obtain the results presented in this manuscript are available at https://sourcesup.renater.fr/projects/apple-gensel.

Supplementary material is available at *G3* online.

Conflicts of interest

The authors declare that there is no conflict of interest.

Literature cited

Akdemir D, Sanchez JI, Jannink J-L. 2015. Optimization of genomic selection training populations with a genetic algorithm. Genet Sel Evol. 47:10.1186/s12711-015-0116-6

Bernardo R. 2014. Genomewide Selection when Major Genes Are Known. Crop Sci. 54:68–75. doi: 10.2135/cropsci2013.05.0315.

Bianco L, Cestaro A, Sargent DJ, Banchi E, Derdak S, Di Guardo M, Salvi S, Jansen J, Viola R, Gut I, et al. 2014. Development and Validation of a 20K Single Nucleotide Polymorphism (SNP) Whole Genome Genotyping Array for Apple (Malus × domestica Borkh). PLoS ONE. 9:e110377 10.1371/journal.pone.0110377

Bianco L, Cestaro A, Linsmith G, Muranty H, Denancé C, et al. 2016. Development and validation of the Axiom ® Apple480K SNP genotyping array. Plant J. 86:62–74. doi: 10.1111/tpj.13145.

Brandariz SP, Bernardo R. 2019. Small ad hoc versus large general training populations for genomewide selection in maize biparental crosses. Theor Appl Genet. 132:347–353. 10.1007/s00122-018-3222-3

Browning SR, Browning BL. 2007. Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. The American Journal of Human Genetics. 81:1084–1097. 10.1086/521987

Ceci AT, Bassi M, Guerra W, Oberhuber M, Robatscher P, et al. 2021. Metabolomic characterization of commercial, old, and Red-Fleshed apple varieties. Metabolites. 11:378.doi:10.3390/metabo11060378.

Chagné D, Kirk C, How N, Whitworth C, Fontic C, et al. 2016. A functional genetic marker for apple red skin coloration across different environments. Tree Genetics & Genomes. 12:Doi:10.1007/s11295-016-1025-8.

Clark SA, Hickey JM, Daetwyler HD, van der Werf JHJ. 2012. The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. Genet Sel Evol. 44:4.doi:10.1186/1297-9686-44-4.

Crossa J, Jarquín D, Franco J, Pérez-Rodríguez P, Burgueño J, et al. 2016. Genomic prediction of gene bank wheat landraces. G3 (Bethesda). 6:1819–1834. doi:10.1534/g3.116.029637.

Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA. 2010. The impact of genetic architecture on genome-wide evaluation methods. Genetics. 185:1021–1031. doi:10.1534/genetics.110.116855.

DoVale JC, Carvalho HF, Sabadin F, Fritsche-Neto R. 2021. Reduction of genotyping marker density for genomic selection is not an affordable approach to long-term breeding in cross-pollinated crops. biorXiv.doi: 10.1101/2021.03.05.434084

Duan N, Bai Y, Sun H, Wang N, Ma Y, et al. 2017. Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. Nat Commun. 8:249. doi: 10.1038/s41467-017-00336-7.

Edwards SM, Buntjer JB, Jackson R, Bentley AR, Lage J, et al. 2019. The effects of training population design on genomic prediction accuracy in wheat. Theor Appl Genet. 132:1943–1952. doi: 10.1007/s00122-019-03327-y.

Endelman JB. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome. 4:250–255. doi: 10.3835/plantgenome2011.08.0024.

Erbe M, Hayes BJ, Matukumalli LK, Goswami S, Bowman PJ, et al. 2012. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J Dairy Sci. 95:4114–4129. doi: 10.3168/jds.2011–5019.

Fritsche-Neto R, Galli G, Borges KLR, Costa-Neto G, Alves FC, et al. 2021. Optimizing genomic-enabled prediction in small-scale maize hybrid breeding programs: a roadmap review. Front Plant Sci. 12:658267.doi:10.3389/fpls.2021.658267.

Gessler C, Pertot I. 2012. Vf scab resistance of Malus. Trees. 26: 95–108. doi:10.1007/s00468-011-0618-y.

Goudet J. 2005. hierfstat, a package for r to compute and test hierarchical F-statistics. Mol Ecol Notes. 5:184–186. doi: 10.1111/j.1471-8286.2004.00828.x.

Hayes BJ, Bowman PJ, Chamberlain AC, Verbyla K, Goddard ME. 2009. Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genet Sel. Evol. 41:51.

Hickey JM, Dreisigacker S, Crossa J, Hearne S, Babu R, et al. 2014. Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. Crop Sci. 54:1476–1488. doi:10.2135/cropsci2013.03.0195.

Howard NP, Albach DC, Lubby JJ. 2018. The identification of apple pedigree information on a large diverse set of apple germplasm and its application in apple breeding using new genetic tools. In: Foerdergemeinschaft Oekologischer Obstbau e. V. (FOEKO). Hohenheim: Germany.

Ibánez-Escriche N, Fernando RL, Toosi A, Dekkers JCM. 2009. Genomic selection of purebreds for crossbred performance. Genet Sel Evol. 41:12.doi:10.1186/1297–9686-41-12.

Ibánẽz-Escriche N, Fernando RL, Toosi A, Dekkers JC.Genomic selection of purebreds for crossbred performance. Genet Sel Evol. 2009;41(1):10.1186/1297-9686-41-12

Isidro J, Jannink J-L, Akdemir D, Poland J, Heslot N, et al. 2015. Training set optimization under population structure in genomic selection. Theor Appl Genet. 128:145–158. doi:10.1007/s00122-014-2418-4.

Janick J, Moore JN, editors. 1975. Advances in Fruit Breeding. West Lafayette, IN: Purdue University Press.

Jarquin D, Specht J, Lorenz A. 2016. Prospects of genomic prediction in the USDA soybean germplasm collection: historical data creates robust models for enhancing selection of accessions. G3 (Bethesda). 6:2329–2341. doi:10.1534/g3.116.031443.

Jung M, Keller B, Roth M, Aranzana MJ, Auwerkerken A, et al. 2021. Genetic architecture and genomic prediction accuracy of apple quantitative traits across environments. biorXiv. doi: 10.1101/2021.11.29.470309.

Jung M, Roth M, Aranzana MJ, Auwerkerken A, Bink M, et al. 2020. The apple REFPOP-a reference population for genomics-assisted breeding in apple. Hortic Res. 7:189. doi:10.1038/s41438-020–00408-8.

Karoui S, Carabaño MJ, Díaz C, Legarra A. 2012. Joint genomic evaluation of French dairy cattle breeds using multiple-trait models. Genet Sel Evol. 44:39.doi:10.1186/1297-9686-44-39.

Khan SA, Tikunov Y, Chibon P-Y, Maliepaard C, Beekwilder J, et al. 2014. Metabolic diversity in apple germplasm. Plant Breed. 133: 281–290. doi:10.1111/pbr.12134.

Kouassi AB, Durel C-E, Costa F, Tartarini S, van de Weg E, et al. 2009. Estimation of genetic parameters and prediction of breeding values for apple fruit-quality traits using pedigreed plant material in Europe. Tree Genet Genomes. 5:659–672. doi: 10.1007/s11295-009-0217-x.

Kumar S, Bink MCAM, Volz RK, Bus VGM, Chagné D. 2012a. Towards genomic selection in apple (Malus × domestica Borkh.) breeding programmes: prospects, challenges and strategies. Tree Genet Genomes. 8:1–14. doi:10.1007/s11295-011-0425-z.

Kumar S, Chagné D, Bink MCAM, Volz RK, Whitworth C, *et al.* 2012b. Genomic selection for fruit quality traits in apple (Malus×domestica Borkh.), (T. Zhang, Ed.). PLoS One. 7: e36674.doi:10.1371/journal.pone.0036674.

Kumar S, Hilario E, Deng CH, Molloy C. 2020. Turbocharging introgression breeding of perennial fruit crops: a case study on apple. Hortic Res. 7:47.doi:10.1038/s41438-020-0270-z.

Kumar S, Molloy C, Muñoz P, Daetwyler H, Chagné D, *et al.* 2015. Genome-enabled estimates of additive and nonadditive genetic variances and prediction of apple phenotypes across environments. G3 (Bethesda). 5:2711–2718. doi:10.1534/g3.115. 021105.

Laurens F, Aranzana MJ, Arus P, Bassi D, Bink M, *et al.* 2018. An integrated approach for increasing breeding efficiency in apple and peach in Europe. Hortic Res. 5:11. doi:10.1038/s41438-018-0016-3.

Lehermeier C, Krämer N, Bauer E, Bauland C, Camisan C, *et al.* 2014. Usefulness of multiparental populations of maize (*Zea mays* L.) for genome-based prediction. Genetics. 198:3–16. doi:10.1534/genetics.114.161943.

Lehermeier C, Schön C-C, de los Campos G. 2015. Assessment of genetic heterogeneity in structured plant populations using multivariate whole-genome regression models. Genetics. 201:323–337. doi:10.1534/genetics.115.177394.

Lenth RV. 2021. emmeans: estimated Marginal Means, aka Least-Squares Means. R package version 1.6.0.

Liao L, Zhang W, Zhang B, Fang T, Wang X-F, *et al.* 2021. Unraveling a genetic roadmap for improved taste in the domesticated apple. Mol Plant. 14:1454–1471. doi:10.1016/j.molp.2021.05.018.

Lorenz AJ, Smith KP. 2015. Adding genetically distant individuals to training populations reduces genomic prediction accuracy in barley. Crop Sci. 55:2657–2667. doi:10.2135/cropsci2014.12.0827.

Lund MS, van den Berg I, Ma P, Brøndum RF, Su G. 2016. Review: how to improve genomic predictions in small dairy cattle populations. Animal. 10:1042–1049. doi:10.1017/S1751731115003031.

Lyra DH, Granato ÍSC, Morais PPP, Alves FC, dos Santos ARM, *et al.* 2018. Controlling population structure in the genomic prediction of tropical maize hybrids. Mol Breed. 38. doi:10.1007/s11032-018–0882-2.

Mangin B, Rincent R, Rabier C-E, Moreau L, Goudemand-Dugue E. 2019. Training set optimization of genomic prediction by means of EthAcc. PLoS One. 14:e0205629.doi:10.1371/journal.pone. 0205629.

McClure KA, Gardner KM, Douglas GM, Song J, Forney CF, *et al.* 2018. A genome-wide association study of apple quality and scab resistance. Plant Genome. 11:170075.doi:10.3835/plantgenome2017.08.0075.

Meuwissen TH, Hayes BJ, Goddard ME. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 157:1819–1829.

Migicovsky Z, Gardner KM, Money D, Sawler J, Bloom JS, *et al.* 2016. Genome to phenome mapping in apple using historical data. Plant Genome. 9. doi:10.3835/plantgenome2015.11.0113.

Migicovsky Z, Yeats TH, Watts S, Song J, Forney CF, *et al.* 2021. Apple ripening is controlled by a NAC transcription factor. Front Genet. 12:671300.doi:10.3389/fgene.2021.671300.

Minamikawa MF, Kunihisa M, Noshita K, Moriya S, Abe K, *et al.* 2021. Tracing founder haplotypes of Japanese apple varieties: application in genomic prediction and genome-wide association study. Hortic Res. 8:49. doi:10.1038/s41438-021-00485-3.

Muranty H, Denancé C, Feugey L, Crépin J-L, Barbier Y, *et al.* 2020. Using whole-genome SNP data to reconstruct a large multi-generation pedigree in apple germplasm. BMC Plant Biol. 20:2.doi:10.1186/s12870-019-2171-6.

Muranty H, Troggio M, Sadok IB, Rifaï MA, Auwerkerken A, *et al.* 2015. Accuracy and responses of genomic selection on key traits in apple breeding. Hortic Res. 2:15060. doi:10.1038/hortres.2015.60.

Myles S. 2013. Improving fruit and wine: what does genomics have to offer? Trends Genet. 29:190–196. doi:10.1016/j.tig.2013.01.006.

Noiton DAM, Alspach PA. 1996. Founding clones, inbreeding, coancestry, and status number of modern apple cultivars. J Am Soc Hortic Sci. 121:773–782. doi:10.21273/JASHS.121.5.773.

Nsibi M, Gouble B, Bureau S, Flutre T, Sauvage C, *et al.* 2020. Adoption and optimization of genomic selection to sustain breeding for apricot fruit quality. G3 (Bethesda). 10:4513–4529. doi:10.1534/g3.120.401452.

Nybom H, Sehic J, Garkava-Gustavsson L. 2008. Modern apple breeding is associated with a significant change in the allelic ratio of the ethylene production gene *Md-ACS1*. J Hortic Sci Biotechnol. 83:673–677. doi:10.1080/14620316.2008.11512442.

Olatoye MO, Clark LV, Labonte NR, Dong H, Dwiyanti MS, *et al.* 2020. Training population optimization for genomic selection in *Miscanthus*. G3 (Bethesda). 10:2465–2476. doi:10.1534/g3.120.401402.

Ou J-H, Liao C-T. 2019. Training set determination for genomic selection. Theor Appl Genet. 132:2781–2792. doi:10.1007/s00122-019-03387-0.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, *et al.* 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 81:559–575. doi: 10.1086/519795.

Ramstein GP, Casler MD. 2019. Extensions of BLUP models for genomic prediction in heterogeneous populations: application in a diverse switchgrass sample. G3 (Bethesda). 9:789–805. doi: 10.1534/g3.118.200969.

Raymond B, Bouwman AC, Wientjes YCJ, Schrooten C, Houwing-Duistermaat J, *et al.* 2018. Genomic prediction for numerically small breeds, using models with pre-selected and differentially weighted markers. Genet Sel Evol. 50:49.doi:10.1186/s12711-018-0419-5.

Rincent R, Laloë D, Nicolas S, Altmann T, Brunel D, *et al.* 2012. Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: comparison of methods in two diverse groups of maize inbreds (*Zea mays* L.). Genetics. 192:715–728. doi:10.1534/genetics.112.141473.

Rincent R, Charcosset A, Moreau L. 2017. Predicting genomic selection efficiency to optimize calibration set and to assess prediction accuracy in highly structured populations. Theor Appl Genet. 130:2231–2247. doi:10.1007/s00122-017-2956-7.

Rio S, Moreau L, Charcosset A, Mary-Huard T. 2020. Accounting for Group-Specific Allele Effects and Admixture in Genomic Predictions: Theory and Experimental Evaluation in Maize. Genetics. 216:27–41. doi: 10.1534/genetics.120.303278.

Rio S, Mary-Huard T, Moreau L, Charcosset A. 2019. Genomic selection efficiency and a priori estimation of accuracy in a structured dent maize panel. Theor Appl Genet. 132:81–96. doi: 10.1007/s00122-018-3196-1.

Rodríguez-Álvarez MX, Boer MP, van Eeuwijk FA, Eilers PHC. 2018. Correcting for spatial heterogeneity in plant breeding experiments with P-splines. Spat Stat. 23:52–71. doi:10.1016/j.spasta.2017.10.003.

de Roos APW, Hayes, BJ, de Goddard ME. 2009. Reliability of genomic predictions across multiple populations. Genetics. 183:1545–1553. doi:10.1534/genetics.109.104935.

de Roos APW, de Hayes BJ, Spelman RJ, Goddard ME. 2008. Linkage disequilibrium and persistence of phase in Holstein–Friesian,

Jersey and Angus cattle. Genetics. 179:1503–1512. doi:10.1534/genetics.107.084301.

Roth M, Muranty H, Guardo MD, Guerra W, Patocchi A, *et al.* 2020. Genomic prediction of fruit texture and training population optimization towards the application of genomic selection in apple. Hortic Res. 7:148. doi:10.1038/s41438-020-00370-5.

Sarinelli JM, Murphy JP, Tyagi P, Holland JB, Johnson JW, *et al.* 2019. Training population selection and use of fixed effects to optimize genomic predictions in a historical USA winter wheat panel. Theor Appl Genet. 132:1247–1261. doi: 10.1007/s00122-019-03276-6.

Schulz-Streeck T, Ogutu JO, Karaman Z, Knaak C, Piepho HP. 2012. Genomic Selection using Multiple Populations. Crop Science. 52:2453–2461. doi: 10.2135/cropsci2012.03.0160.

Singh A, Dubey PK, Chaurasia R, Dubey RK, Pandey KK, *et al.* 2019. Domesticating the undomesticated for global food and nutritional security: four steps. Agronomy. 9:491.doi:10.3390/agronomy9090491.

Song H, Ye S, Jiang Y, Zhang Z, Zhang Q, *et al.* 2019. Using imputation-based whole-genome sequencing data to improve the accuracy of genomic prediction for combined populations in pigs. Genet Sel Evol. 51:58.doi:10.1186/s12711-019-0500-8.

Sood S, Lin Z, Caruana B, Slater AT, Daetwyler HD. 2020. Making the most of all data: combining non-genotyped and genotyped potato individuals with HBLUP. Plant Genome. 13:e20056.doi: 10.1002/tpg2.20056.

Sverrisdóttir E, Sundmark EHR, Johnsen HØ, Kirk HG, Asp T, *et al.* 2018. The value of expanding the training population to improve genomic selection models in tetraploid potato. Front Plant Sci. 9: 1118.doi:10.3389/fpls.2018.01118.

Technow F, Totir LR. 2015. Using Bayesian multilevel whole genome regression models for partial pooling of training sets in genomic prediction. G3 (Bethesda). 5:1603–1612. doi:10.1534/g3.115.019299.

Tiezzi F, Maltecca C. 2015. Accounting for trait architecture in genomic predictions of US Holstein cattle using a weighted realized relationship matrix. Genet Sel Evol. 47:24.doi: 10.1186/s12711-015-0100-1.

van Nocker S, Gardiner SE. 2014. Breeding better cultivars, faster: applications of new technologies for the rapid deployment of superior horticultural tree crops. Hortic Res. 1:14022.doi: 10.1038/hortres.2014.22.

VanRaden PM. 2008. Efficient methods to compute genomic predictions. J Dairy Sci. 91:4414–4423. doi:10.3168/jds.2007-0980.

Verma S, Evans K, Guan Y, Luby JJ, Rosyara UR, *et al.* 2019. Two large-effect QTLs, Ma and Ma3, determine genetic potential for acidity in apple fruit: breeding insights from a multi-family study. Tree Genetics & Genomes. 15:Doi:10.1007/s11295-019-1324-y.

Voss-Fels KP, Cooper M, Hayes BJ. 2019. Accelerating crop genetic gains with genomic selection. Theor Appl Genet. 132:669–686. doi:10.1007/s00122-018-3270-8.

Watkins R, Smith RA. 1982. International board for plant genetic resources, commission of the European communities, and committee on disease resistance breeding and use of Genebanks. In: Descriptor List for Apple (Malus.). Brussels; Rome: CEC Secretariat; IBPGR Secretariat.

Watts S, Migicovsky Z, McClure KA, Amyotte B, *et al.* 2021. Quantifying apple diversity: a phenomic characterization of Canada's apple biodiversity collection. Plants People Planet. 10211. doi:10.1002/ppp3.10211.

Wedger MJ, Schumann AC, Gross BL. 2021. Candidate genes and signatures of directional selection on fruit quality traits during apple domestication. Am J Bot. 108:616–627. doi:10.1002/ajb2. 1636.

Wientjes Y, Veerkamp RF, Bijma P, Bovenhuis H, Schrooten C, *et al.* 2015. Empirical and deterministic accuracies of across-population genomic prediction. Genet Sel Evol. 47:5.doi:10.1186/s12711 -014-0086-0.

Wientjes YCJ, Bijma P, Veerkamp RF, Calus MPL. 2016. An equation to predict the accuracy of genomic values by combining data from multiple traits, populations, or environments. Genetics. 202:799–823. doi:10.1534/genetics.115.183269.

Wimmer V, Lehermeier C, Albrecht T, Auinger H-J, Wang Y, *et al.* 2013. Genome-wide prediction of traits with different genetic architecture through efficient variable selection. Genetics. 195: 573–587. doi:10.1534/genetics.113.150078.

Xu Y, Liu X, Fu J, Wang H, Wang J, *et al.* 2020. Enhancing genetic gain through genomic selection: from livestock to plants. Plant Commun. 1:100005.doi:10.1016/j.xplc.2019.100005.

Yao J-L, Xu J, Cornille A, Tomes S, Karunairetnam S, *et al.* 2015. A *microRNA* allele that emerged prior to apple domestication may underlie fruit size evolution. Plant J. 84:417–427. doi: 10.1111/tpj.13021.

Yu X, Li X, Guo T, Zhu C, Wu Y, *et al.* 2016. Genomic prediction contributing to a promising global strategy to turbocharge gene banks. Nat Plants. 2:16150. doi:10.1038/nplants.2016.150.

Zhang A, Wang H, Beyene Y, Semagn K, Liu Y, *et al.* 2017. Effect of trait heritability, training population size and marker density on genomic prediction accuracy estimation in 22 bi-parental tropical maize populations. Front Plant Sci. 8:1916. doi:10.3389/fpls.2017.01916.

Communicating editor: P. Brown