Genetic architecture and genomic prediction accuracy of apple quantitative traits across environments

Authors

Jung, Michaela[1,2]*; Keller, Beat[1,2]; Roth, Morgane[1,3]; Aranzana, Maria José[4,5]; Auwerkerken, Annemarie[6]; Guerra, Walter[7]; Al-Rifaï, Mehdi[8]; Lewandowski, Mariusz[9]; Sanin, Nadia[7]; Rymenants, Marijn[6,10]; Didelot, Frédérique[11]; Dujak, Christian[5]; Font i Forcada, Carolina[4]; Knauf, Andrea[1,2]; Laurens, François[8]; Studer, Bruno[2]; Muranty, Hélène[8]; Patocchi, Andrea[1]

Author affiliations

[1]Breeding Research group, Agroscope, 8820 Wädenswil, Switzerland

[2]Molecular Plant Breeding, Institute of Agricultural Sciences, ETH Zurich, 8092 Zurich, Switzerland

[3]GAFL, INRAE, 84140 Montfavet, France

[4]IRTA (Institut de Recerca i Tecnologia Agroalimentàries), 08140 Caldes de Montbui, Barcelona, Spain

[5]Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Campus UAB, 08193 Bellaterra, Barcelona, Spain

[6]Better3fruit N.V., 3202 Rillaar, Belgium

[7]Research Centre Laimburg, 39040 Auer, Italy

[8]Univ Angers, Institut Agro, INRAE, IRHS, SFR QuaSaV, F-49000 Angers, France

[9]The National Institute of Horticultural Research, Konstytucji 3 Maja 1/3, 96-100 Skierniewice, Poland

[10]Laboratory for Plant Genetics and Crop Improvement, KU Leuven, B-3001, Leuven, Belgium

[11]Unité expérimentale Horticole, INRAE, F-49000 Angers, France

*corresponding author

Email addresses (in order of authors)

'michaela.jung@usys.ethz.ch'; 'beat.keller@usys.ethz.ch'; 'morgane.roth@inrae.fr'; 'mariajose.aranzana@irta.cat'; 'annemarie@better3fruit.com'; 'Walter.Guerra@laimburg.it'; 'mehdi.al-rifai@inrae.fr'; 'mariusz.lewandowski@inhort.pl'; 'Nadia.Sanin@laimburg.it'; 'marijn@better3fruit.com'; 'frederique.didelot@inrae.fr'; 'christian.dujak@cragenomica.es'; 'carolina.font@irta.cat'; 'andrea.knauf@agroscope.admin.ch'; 'francois.laurens@inrae.fr'; 'bruno.studer@usys.ethz.ch'; 'helene.muranty@inrae.fr'; 'andrea.patocchi@agroscope.admin.ch'

Abstract

Implementation of genomic tools is desirable to increase the efficiency of apple breeding. The apple reference population (apple REFPOP) proved useful for rediscovering loci, estimating genomic prediction accuracy, and studying genotype by environment interactions (G×E). Here we show contrasting genetic architecture and genomic prediction accuracies for 30 quantitative traits across up to six European locations using the apple REFPOP. A total of 59 stable and 277 location-specific associations were found using GWAS, 69.2% of which are novel when compared with 41 reviewed publications. Average genomic prediction accuracies of 0.18–0.88 were estimated using single-environment univariate, single-environment multivariate, multi-environment univariate, and multi-environment multivariate models. The G×E accounted for up to 24% of the phenotypic variability. This most comprehensive genomic study in apple in terms of trait-environment combinations provided knowledge of trait biology and prediction models that can be readily applied for marker-assisted or genomic selection, thus facilitating increased breeding efficiency.

Introduction

Apple (*Malus domestica* Borkh.) is the third most produced fruit crop worldwide[1]. Since its domestication in the Tian Shan mountains of Central Asia, the cultivated apple developed into a separated near-panmictic species[2]. Over the centuries, thousands of apple cultivars have been raised and conserved thanks to grafting[3]. Extensive relatedness among cultivars with a strong influence of a

few founders through the history of apple breeding has been reported despite their high genetic diversity[4-6]. Only a fraction of the existing cultivars are grown commercialy[3] and they require an intensive use of pesticides for crop protection. To diversify apple production, it is desirable to produce new cultivars for sustainable intensive agriculture and adapted to future climate, while remaining attractive to consumers.

Apple breeding is labor- and time-intensive, but selection efficiency can be improved by integrating DNA-informed techniques into the breeding process[7]. Marker-assisted selection allows breeders to predict the value of a target trait based on its association with a genetic marker. The method leads to removal of inferior seedlings without phenotyping, thus reducing the labor costs when decreasing the number of individuals passing to the next selection step[7]. Quantitative trait locus (QTL) mapping has been traditionally used to investigate the genetic basis of variation in traits such as pathogen resistance, phenology, and some fruit quality traits[8-11]. To bridge the gap between the discovery of marker-trait associations and their application in breeding, protocols that transfer the knowledge obtained by QTL analyses into DNA tests were established[12,13]. However, marker-assisted selection in apple remains restricted to a limited number of traits associated with single genes or a handful of large-effect QTL, such as pathogen resistance and fruit firmness, acidity, or color[14]. DNA-informed selection is rarely deployed in apple when breeding for quantitative traits with complex genetic architecture, though this task became feasible with the recent technological developments in apple genomics.

In the genomics era, advancements in genotyping and sequencing technologies led to a broad range of new tools for genetic analyses. In the case of apple, several reference genomes have been produced[15-19], single nucleotide polymorphism (SNP) genotyping arrays of different densities such as 20K or 480K SNPs have been developed[20,21], and genotyping-by-sequencing methods have been adopted[22,23]. Genome-wide association study (GWAS) emerged as a method for exploring the genetic basis of quantitative traits[24]. GWAS in apple have been used to identify associations between markers and various traits such as fruit quality and phenology traits[22,23,25-29]. The associations found with GWAS can be translated into DNA tests for marker-assisted selection. Besides GWAS, genomic selection was

developed to exploit the effects of genome-wide variation at loci of both large and low effect on quantitative traits using a single model[30] and is sometimes called marker-assisted selection on a genome-wide scale[31]. For genomic selection, prediction models are first trained with phenotypic and genomic data of a training population. In a second step, the models predict the performance of breeding material based on the genomic data alone. These genomic estimated breeding values are then used to make selections among the breeding material, thus increasing the breeding efficiency and genetic gain. Several studies have assessed genomic prediction accuracy for apple quantitative traits related to fruit quality and phenology[22,23,29,32-36]. Genomic selection can double genetic gain, as demonstrated by yield traits in dairy cattle[37], but the accuracy of genomic prediction for yield traits in apple has not been studied. Analyses of genomic datasets beyond 100K SNPs have been limited to flowering and harvest time (GWAS and genomic prediction)[26,36], fruit firmness and skin color (GWAS)[28,38]. Marker density, trait architecture, and heritability have been shown to differentially affect prediction performance in simulated data and for apple[34,36,39] and their impact on genomic analyses should therefore be further empirically tested. Moreover, GWAS for the same traits measured at different locations, the effect of genotype by environment interaction (G×E) on genomic prediction accuracy, and predictions with multivariate genomic prediction models have not been evaluated yet in apple.

Plants are known for their strong phenotypic response to environmental factors, a phenomenon regularly tested in plant breeding using multi-environment trials. In general, when statistical models are applied to measurements from multi-environment trials, the effect of environment on individuals remains constant at single locations, but the G×E leads to changes in the ranking of genotypes across locations. With an increasing proportion of G×E effect relative to genotypic effect, both heritability and response to selection decrease[40]. A noticeable effect of contrasting European environments and G×E on two apple phenology traits – floral emergence and harvest date – has been reported, which demands testing the multi-environment modelling approaches in apple[36]. A location-specific GWAS may be used to identify loci with stable effects across environments and loci specific to individual

104 locations[41]. Multi-environment prediction models can account for G×E by explicitly modeling
105 interactions between all available markers and environments[42]. Borrowing information from other
106 genotypes across environments through markers, the G×E method can outperform more simple
107 modelling approaches that ignore G×E[42-44]. Additionally, taking advantage of information that traits
108 provide about one another, a multivariate (also called multi-trait) genomic prediction can be applied.
109 This method may be useful in case the assessment of one trait remains costly, but another correlated
110 trait with less expensive measurement is available or can be assessed more easily[45]. The multivariate
111 prediction can also be extended to a multi-environment approach when treating measurements from
112 different environments as distinct traits[46].
113 A population of 269 diverse apple accessions from across the globe and 265 progeny from 27 parental
114 combinations originating in recent European breeding programs constitutes our apple reference
115 population (apple REFPOP)[36]. The apple REFPOP has a high-density genomic dataset of 303K SNPs and
116 was deemed suitable for the application of genomics-assisted breeding[36]. Combined with extensive
117 phenotypic information, the apple REFPOP provides the groundwork for marker-assisted and genomic
118 selection across contrasting European environments. Hence, 30 traits related to productivity, tree
119 vigor, phenology, and fruit quality were measured in the apple REFPOP during up to three years and
120 at up to six locations with various climatic conditions of Europe (Belgium, France, Italy, Poland, Spain,
121 and Switzerland). First, GWAS was performed to dissect the genetic architecture of the studied traits,
122 identify associated loci stable across locations and location-specific loci, and to observe signs of
123 selection on loci of large effect. Second, this study aimed to measure prediction accuracy for these
124 traits using single-environment univariate, single-environment multivariate, multi-environment
125 univariate, and multi-environment multivariate genomic prediction models. Finally, a critical analysis
126 of our results provided recommendations for future implementation of genomic prediction tools in
127 apple breeding.
128
129 **Results**

Phenotypic data analysis

The accession and progeny groups of the apple REFPOP were evaluated for 30 quantitative traits at up to six locations. The measurements for ten traits were collected at one location, while the remaining 20 traits were available from at least two locations (three traits were measured in two locations, three traits in four locations, eleven traits in five locations and three traits in six locations, Supplementary Table 1). Most traits (25) were assessed during three seasons while five traits were measured during two seasons (Supplementary Table 1). Accounting for environmental effects in the phenotypic data, BLUPs of traits (best linear unbiased prediction of random effects of genotypes, see Equation 1) were produced across all locations and separately for each location. The traits showed unimodal as well as multimodal distributions (Supplementary Figure 1). Differences of various extent between the accession and progeny groups were observed (Supplementary Figure 2). As expected, high phenotypic and genotypic correlations (>0.7) between traits were observed within trait categories, namely the phenology, productivity, fruit size, outer fruit, inner fruit, and vigor category (Figure 1a). A few moderate positive phenotypic correlations (0.3–0.7) were found between trait categories such as harvest date and fruit firmness (0.51), yellow color and russet cover (0.55), soluble solids content and russet cover (0.36), or between yield (weight and number of fruits) and vigor trait category (0.36–0.51, Figure 1a). High average correlations were observed between the environments (combinations of location and year) for harvest date (0.82 [0.73, 0.95]) or red over color (0.80 [0.62, 0.92]) whereas low average correlations (<0.3) were present between environments for flowering intensity (0.18 [-0.49, 0.68]) and trunk increment (0.16 [-0.31, 0.55], Supplementary Table 2, Supplementary Figure 3). A shift of the progeny group compared to the accession group towards smaller, more numerous and less russeted fruits was observed (Figure 1b).

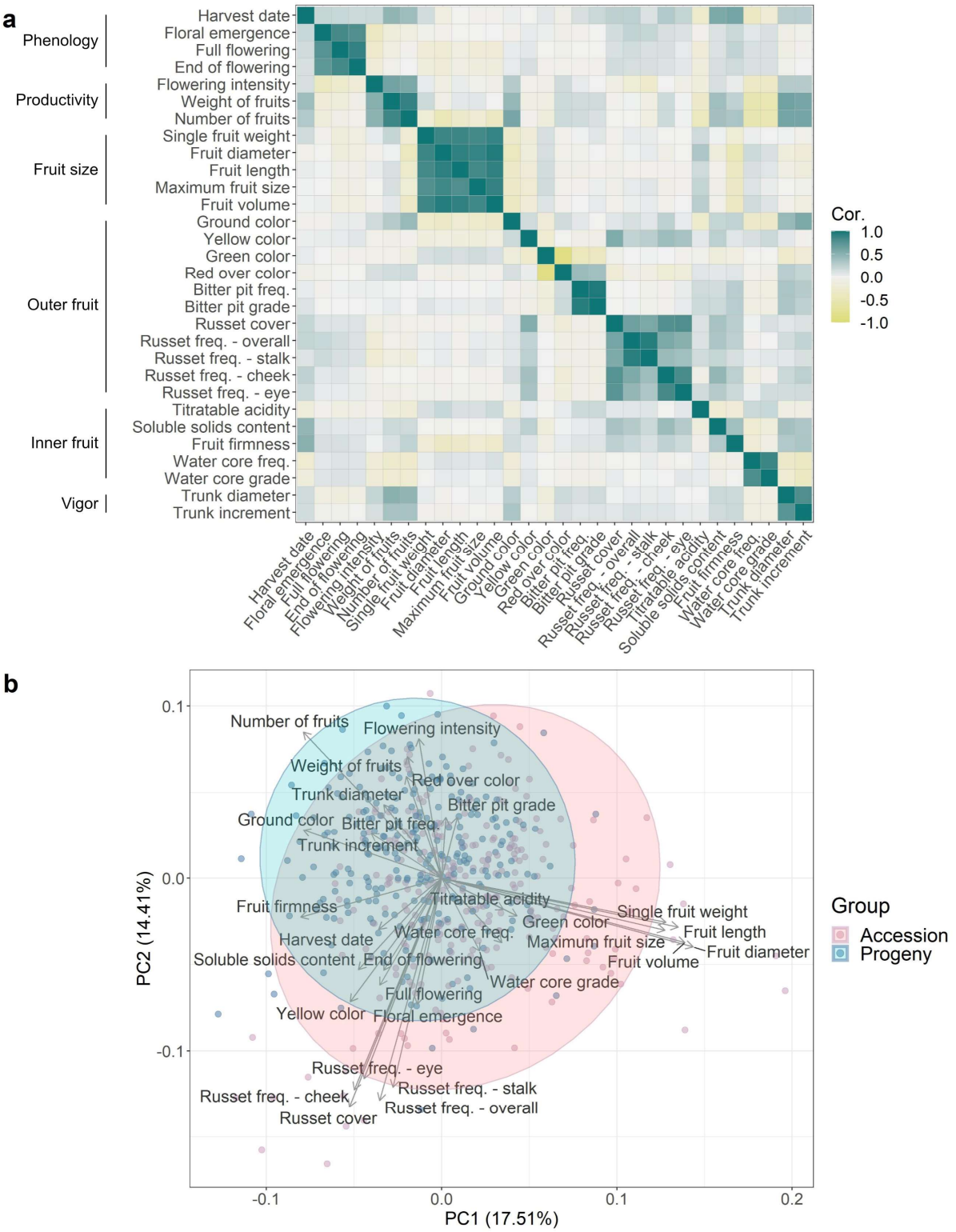


Figure 1: Exploratory phenotypic data analysis of the studied quantitative apple traits. a Pairwise correlations between traits with the phenotypic and genomic correlations in the lower and upper triangular part, respectively. Phenotypic correlation was assessed as Pearson correlation between pairs of across-location BLUPs, the genomic correlation as Pearson correlation between pairs of

158 genomic BLUPs estimated from a G-BLUP model. Trait categories are outlined along the vertical axis.
159 **b** Principal component analysis biplot based on across-location BLUPs of apple traits with the addition
160 of location-specific BLUPs for traits measured at a single location.
161
162 **Genome-wide association studies**
163 Across-location GWAS for 20 traits measured at more than one location (Supplementary Table 1) and
164 location-specific GWAS for all 30 traits were used to explore the genetic basis of the assessed traits.
165 The quantile-quantile plots showed that the observed and expected distributions of p-values
166 corresponded well and no apparent inflation of p-values was found (Supplementary Figure 4 and 5).
167 Across-location GWAS revealed 59 significant ($-log_{10}(p) > 6.74$) marker-trait associations in 18 traits
168 (Figure 2a, Supplementary Table 3). No significant associations were observed for trunk diameter and
169 russet cover in the across-location GWAS. In the location-specific GWAS, 309 significant marker-trait
170 associations for all 30 traits were discovered (Figure 2b, Supplementary Table 3). Of these 309 marker-
171 trait associations, 32 associations for twelve traits were shared between the location-specific GWAS
172 and the across-location GWAS (Supplementary Table 3). The coefficient of determination (R^2) of
173 significant associations was the largest for red over color (0.71), green color (0.55) and harvest date
174 (0.42, Figure 2c, Supplementary Table 3).
175 Significant associations with different traits co-localized at identical positions or occurred very close in
176 some genomic regions (distance between marker positions below 100 kb, Figure 2c, Supplementary
177 Table 3). In the across-location GWAS, a marker significantly associated with harvest date on
178 chromosome 3 (position 30,681,581 bp) was located next to two markers associated with fruit firmness
179 (positions 30,587,378 and 30,590,166 bp). The same marker on the position 30,681,581 bp was also
180 associated with harvest date, ground color, overall russet frequency and soluble solids content
181 measured at several different locations (location-specific GWAS). Similarly, the association with
182 harvest date on chromosome 16 (position 9,023,861 bp) was closely located to a marker associated
183 with fruit firmness (position 8,985,888 bp) in the across-location GWAS. The traits related to bitter pit

184 analyzed in the across-location GWAS, i.e., bitter pit frequency and grade, showed significant
185 associations on chromosome 16, position 7,681,416 bp. Several associations with traits measuring fruit
186 skin russet in the across-location GWAS co-localized on chromosome 12 (position 23,013,281 bp,
187 russet frequency on cheek and in the eye) and 17 (position 27,249,890 bp, overall russet frequency
188 and russet frequency at stalk). A marker at position 18,679,105 bp on chromosome 1 was associated
189 with both single fruit weight from the across-location GWAS and fruit diameter from Switzerland
190 (found with the location-specific GWAS). The association with marker at position 2,005,502 bp on
191 chromosome 8 was shared between fruit diameter and fruit volume from Switzerland and single fruit
192 weight from Belgium. On chromosome 11, fruit diameter, fruit volume and single fruit weight from
193 Switzerland, as well as single fruit weight from Belgium, shared the association at position 18,521,895
194 bp. Additionally, position 3,622,193 bp on chromosome 11 was shared between the associations of
195 fruit length and single fruit weight from Switzerland. For red over color and green color, the association
196 with a marker on chromosome 9 (position 33,799,120 bp) occurred in across-location and four
197 location-specific GWAS, while a close marker (position 33,801,013 bp, less than 2kb away) was
198 associated in the two other location-specific GWAS. Additional significant marker-trait associations
199 occurred in the same genomic regions among the location-specific GWAS and between the across-
200 location and location-specific GWAS (Supplementary Table 3).
201 Previous reports on QTL mapping and GWAS in apple were extensively reviewed and 41 publications
202 reporting on traits measured similarly to our own were found and taken for comparison
203 (Supplementary Table 4). The QTL positions from literature and the marker-trait associations found in
204 this study were assigned to chromosome segments (top, center, and bottom of a chromosome).
205 Unique segment-trait combinations were discovered in the literature (166), in the across-location
206 GWAS (52) and in the location-specific GWAS (172,
207 Figure 3a). Out of all segment-trait combinations across our GWAS, 30.8% overlapped with the
208 previously published results of QTL mapping or GWAS and the rest (69.2%) were novel. All previously
209 published segment-trait combinations for the trait groups bitter pit and trunk were also detected in

our study, whereas no overlap between the former and present associations was found for ground color and sugar trait groups (Figure 3b, Supplementary Figure 6).

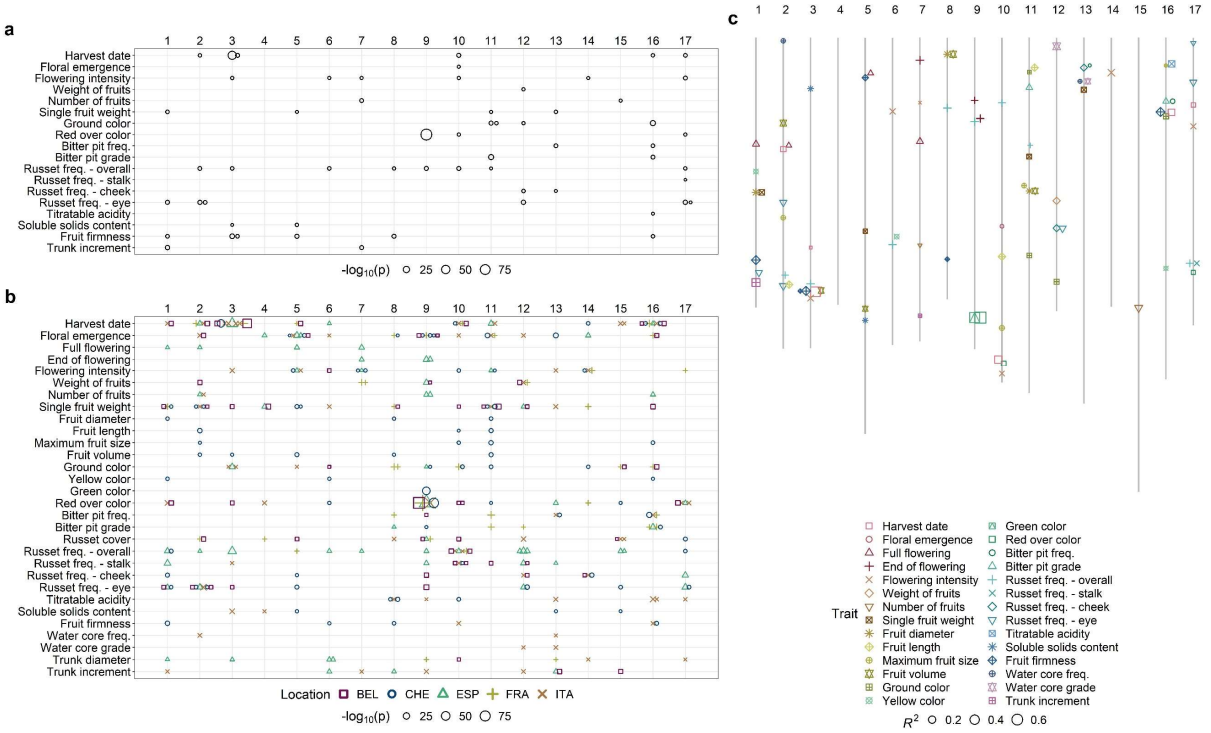


Figure 2: Significant marker-trait associations found by GWAS. a Distribution of the significant associations and corresponding p-values from across-location GWAS over the 17 apple chromosomes. **b** Distribution of the significant associations and corresponding p-values from location-specific GWAS over the 17 apple chromosomes. Locations are labeled as BEL (Belgium), CHE (Switzerland), ESP (Spain), FRA (France) and ITA (Italy). **a-b** Size of the symbols indicate the $-log_{10}(p)$. The x-axis shows chromosome numbers. **c** Physical positions (in bp) of the significant associations on chromosomes with their respective coefficients of determination (R^2) from the across-location GWAS complemented with the location-specific GWAS for traits measured at a single location. Size of the symbols indicate the R^2. The x-axis shows chromosome numbers.

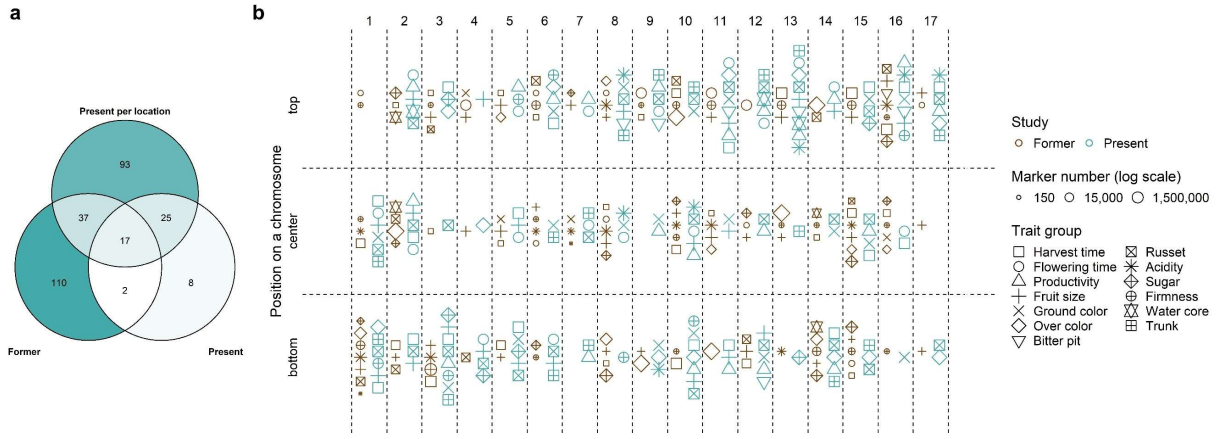


Figure 3: Comparison of the significant marker-trait associations with previously published associations. a Venn diagram comparing the unique associations, which were either previously published (former), reported in the across-location GWAS (present) or the location-specific GWAS (present per location). Color intensity and the values reflect the number of associations per diagram area. **b** Scatterplot of unique associations comparing published associations (former) with the merged across-location and location-specific GWAS (present). The traits were assembled into trait groups based on their similarity. Symbol size reflects the number of markers used in the studies. In case more than one publication reported an association in the same chromosome segment, only the report with the largest number of markers is shown (see Supplementary Table 4 for the complete list of previously published associations). **a-b** Positions of associations were assigned to three chromosome segments: top, center and bottom. Only the unique combinations of trait groups with segments and type of study (former or present) are shown.

Allele frequency dynamics over generations

Eleven major significant marker-trait associations (R^2>0.1) were identified in the global GWAS results (across-location GWAS with the addition of location-specific GWAS for traits measured at a single location only, Figure 4). Among these major associations, changes in the frequency of alleles with an increasing effect on trait phenotypes were quantified in 30 ancestral accessions (five ancestor generations of the progeny group, Supplementary Table 5) and all 265 progenies included in the apple REFPOP (Figure 4a). Compared to the ancestral accessions, the frequency of the allele with an

increasing effect on phenotype (Figure 4c) was higher in the progeny for the alleles associated with later harvest date and increased flowering intensity, titratable acidity, fruit firmness and trunk increment (Figure 4a). For the marker associated with green color and red over color, the allele frequencies were equivalent for ancestors and progeny, which reflected the minor allele frequency of nearly 0.5 for both traits (Figure 4b,d). Noticeably, at the markers closely associated with harvest date and fruit firmness on chromosome 3, the allele associated with later harvest date and firmer fruits was fixed in all progeny, while the allele with a decreasing effect on the phenotype was present with a frequency below 0.1 in the whole apple REFPOP (Figure 4a-d). The allele associated with larger trunk increment on chromosome 1 was found in progeny known to segregate for *Rvi6*, and it was present in only two accessions ('Prima' and X6398) that are also known to carry the apple scab resistance gene *Rvi6*, which is located about 1.8 Mb from the SNP associated with trunk increment (Figure 4b-c). The remaining associations (R^2≤0.1) reported by the global GWAS showed various trends in allele frequencies across generations such as increased frequency of alleles associated with increased weight of fruits in the progeny (Supplementary Figure 7). The individual parental combinations of the progeny group were often fixed for single alleles (Figure 4b, Supplementary Figure 8). Boxplots of the across-location BLUPs against the dosage of the reference allele (0, 1, 2) for the eleven major significant marker-trait associations showed additive effects of the alleles on phenotypes (Supplementary Figure 9). Squared Pearson's correlations in a window of ~3,000 markers surrounding each of the major significant marker-trait associations showed that markers in linkage disequilibrium extended over larger distances around some marker-trait associations (Supplementary Figure 10). When visually compared with other loci, the associations with harvest date and fruit firmness on chromosome 3 as well as red over color and green color on chromosome 9 were found in genomic regions of the highest linkage disequilibrium between markers (Supplementary Figure 10). The markers associated with trunk increment and *Rvi6* also showed signs of linkage disequilibrium (Supplementary Figure 10).

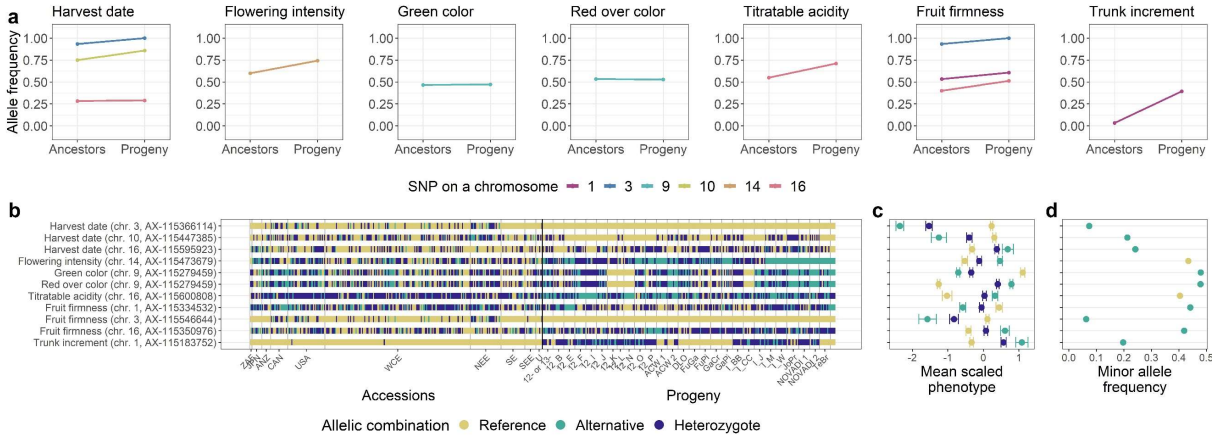


Figure 4: Allele frequency dynamics of the major significant marker-trait associations. a-d The associations were chosen based on the coefficient of determination (R^2>0.1) from the global GWAS. **a** For each association, frequency of the allele with increasing effect on trait phenotypes in the apple REFPOP is shown. For the progeny group (progeny) and its five ancestor generations (ancestors), the allele frequencies are shown as points connected with a line. Out of all known ancestors, the allele frequency was estimated for 30 accessions included in the apple REFPOP. Colors of the points and lines correspond to chromosome locations of the associated SNPs. **b** Allelic combinations carried by the apple REFPOP genotypes, sorted according to geographic origin of accessions and affiliation of progeny to parental combinations (the x-axis was labeled according to Supplementary Table 1 and 2 in Jung et al.[36]). **c** Phenotypic BLUPs of traits and their standard error for each allelic combination, centered to mean 0 and scaled to standard deviation of 1. **d** Frequency of the minor allele in the whole apple REFPOP. **b-d** The legend and y-axis are shared between plots. In d, the color of an allele corresponds to the color of the homozygous allelic combination of the same allele in b and c.

Genomic prediction

Four single-environment univariate prediction models – random forest (RF), BayesCπ, Bayesian reproducing kernel Hilbert spaces regression (RKHS) and genomic-BLUP (G-BLUP) – and a single-environment multivariate model with an unstructured covariance matrix of the random marker effect (MTM.UN) were compared using across-location BLUPs and location-specific BLUPs as phenotypes from a single environment. Among these models, the average prediction accuracies per trait ($\bar{r}_t$)

ranged between 0.18 for russet cover and 0.88 for red over color, both extreme values observed with RF (Supplementary Table 6). The prediction accuracies estimated for G-BLUP were further used as reference for model comparisons due to its widespread use in genomic prediction. When the prediction accuracy of the G-BLUP model was averaged over all traits ($\bar{r}$), the obtained $\bar{r}$ was equal to 0.50. The RF showed an $\bar{r}_t$ higher than G-BLUP for 9 out of 30 traits and an $\bar{r}$ of 0.49. BayesCπ, RKHS and MTM.UN showed an $\bar{r}$ of 0.50, 0.51 and 0.50 and exceeded $\bar{r}_t$ of G-BLUP in one, twelve and ten traits, respectively. Generally, a similar performance of all five models was observed (Figure 5a).

When compared with the baseline model G-BLUP, the single-environment multivariate model MTM.UN showed an improved prediction accuracy for several traits when they were modelled in combination with a correlated trait (genomic correlation larger than 0.3, Figure 5a, Supplementary Table 6). The inclusion of floral emergence as correlated trait improved $\bar{r}_t$ of full flowering and end of flowering. A combination with weight of fruits improved $\bar{r}_t$ of flowering intensity. Fitting the model using fruit length showed an increased $\bar{r}_t$ of single fruit weight and using single fruit weight led to an increase in $\bar{r}_t$ for fruit diameter, fruit length, maximum fruit size and fruit volume. Using soluble solids content resulted in an increase of $\bar{r}_t$ for russet cover, while using russet frequency at cheek led to an improved $\bar{r}_t$ of russet frequency at stalk. Prediction accuracies for all possible combinations of correlated traits can be found in Supplementary Table 7.

Two multi-environment univariate models – across-environment G-BLUP (G-BLUP.E) and marker by environment interaction G-BLUP (G-BLUP.E.G×E) – and the multi-environment multivariate factor-analytic model (MTM.FA) were compared using two cross-validation scenarios corresponding to different experimental scenarios. In the first cross-validation scenario (CV1), traits were predicted for 20% of genotypes in each environment (i.e., their phenotypes were masked in all environments for model training). In the second cross-validation scenario (CV2), traits were predicted for 20% of genotypes in all but the Swiss environments (i.e., for these genotypes the environments “CHE.2018”, “CHE.2019” and “CHE.2020” were retained for model training). For the models applied with CV1, the $\bar{r}_t$ ranged between 0.13 (for russet frequency in the eye obtained with MTM.FA) and 0.70 (for harvest

317 date estimated with G-BLUP.E.G×E, Supplementary Table 6). With CV2, the lowest $\bar{r}_t$ of 0.29 was
318 measured for trunk increment with G-BLUP.E.G×E and the maximum $\bar{r}_t$ of 0.86 was found for harvest
319 date with both G-BLUP.E and G-BLUP.E.G×E models (Supplementary Table 6). The prediction
320 performance of G-BLUP.E, G-BLUP.E.G×E and MTM.FA was generally lower under CV1 than under CV2
321 (Figure 5b, Supplementary Table 6). For all traits, the G-BLUP.E.CV1, G-BLUP.E.G×E.CV1 and
322 MTM.FA.CV1 showed lower $\bar{r}_t$ than the single-environment G-BLUP, the $\bar{r}$ being equal to 0.40, 0.40
323 and 0.36, respectively. The G-BLUP.E.G×E.CV1 performed better than G-BLUP.E.CV1 for 14 out of 30
324 traits. The G-BLUP.E.CV2 and G-BLUP.E.G×E.CV2 outperformed G-BLUP for 13 out of 20 traits. The G-
325 BLUP.E.CV2 and G-BLUP.E.G×E.CV2 both showed $\bar{r}$ equal to 0.57. The increase in $\bar{r}_t$ from G-BLUP to G-
326 BLUP.E.CV2 (0.35) as well as from G-BLUP to G-BLUP.E.G×E.CV2 (0.36) was the most pronounced for
327 russet cover. The performance of G-BLUP.E.CV2 and G-BLUP.E.G×E.CV2 remained below the level of
328 G-BLUP predictions for productivity traits (flowering intensity, weight and number of fruits), ground
329 color, soluble solids content, fruit firmness and trunk increment. The G-BLUP.E.G×E.CV2 performed
330 better than G-BLUP.E.CV2 for 8 out of 20 traits. The $\bar{r}$ of MTM.FA.CV2 was equal to 0.52 and therefore
331 similar to G-BLUP, however, the model outperformed G-BLUP for nine out of 20 predicted traits
332 (Supplementary Table 6). The MTM.FA showed higher prediction accuracy than both G-BLUP.E and G-
333 BLUP.E.G×E for two traits under CV1 and five traits under CV2 (Supplementary Table 6).
334 Across all model groups, the best prediction performance was found for harvest date, green color and
335 red over color (Figure 5, Supplementary Table 6). The lowest prediction accuracy was found for traits
336 related to bitter pit and russet as well as yellow color. Additionally, the prediction accuracy for
337 flowering intensity and trunk increment with the multi-environment models remained strongly below
338 the $\bar{r}_t$ of the corresponding single-environment models.
339

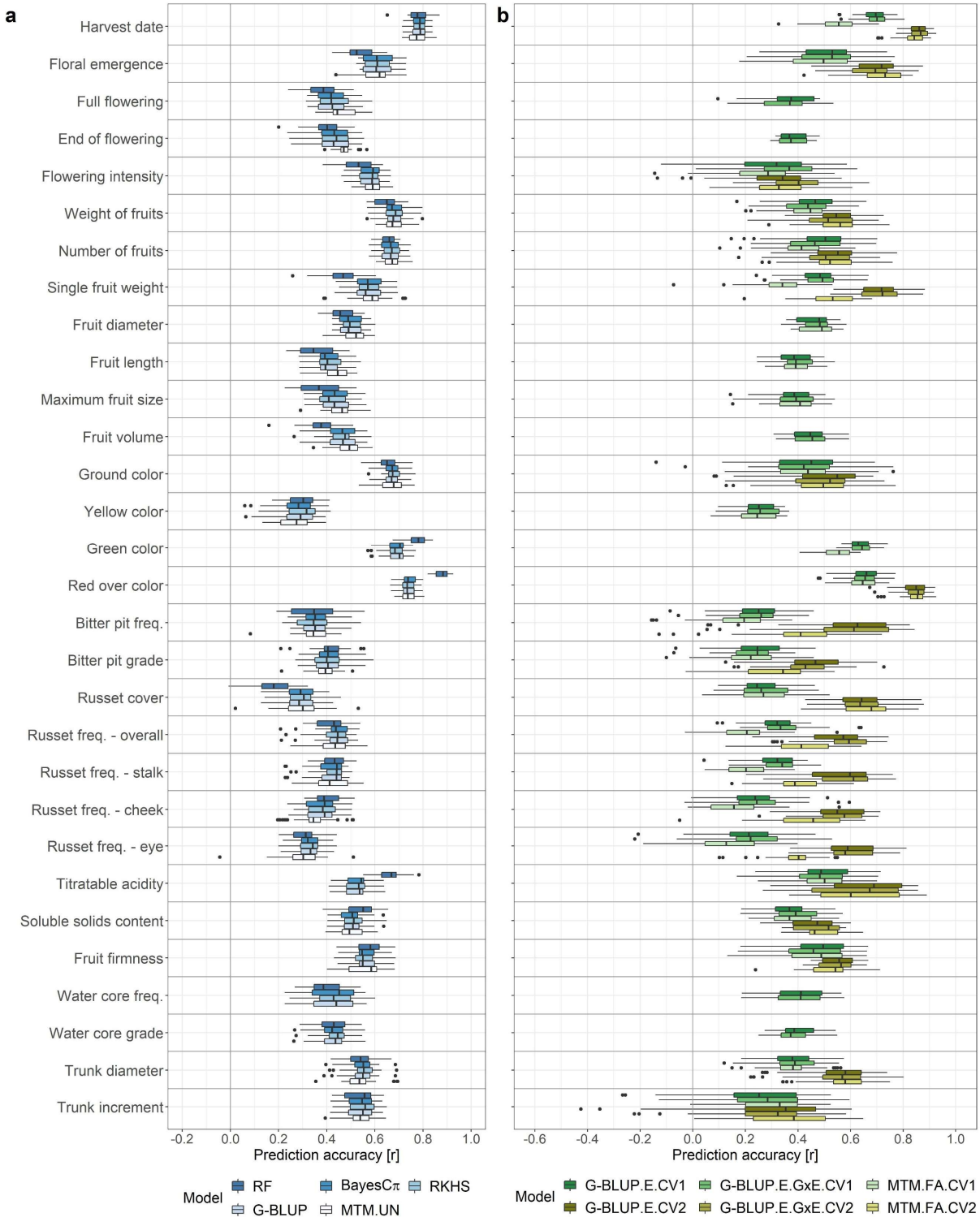
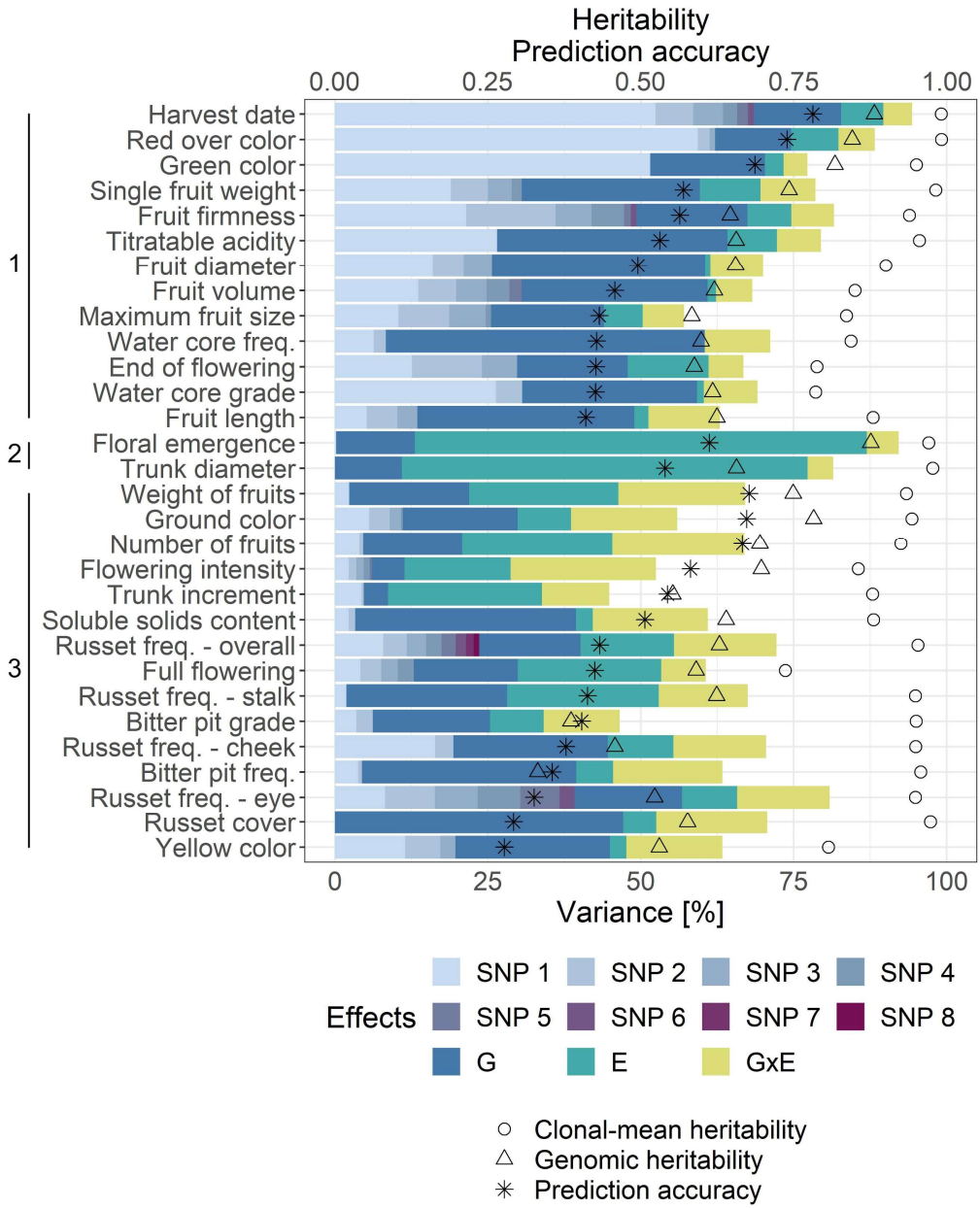


Figure 5: Genomic prediction accuracy in apple quantitative traits using eight genomic prediction models and two cross-validation scenarios. **a** Prediction accuracy of four single-environment univariate models, i.e., random forest (RF), BayesCπ, Bayesian reproducing kernel Hilbert spaces regression (RKHS) and genomic-BLUP (G-BLUP), and one single-environment multivariate model with an unstructured covariance matrix of the random marker effect (MTM.UN). The models were applied

346 with a five-fold cross-validation where 20% of the genotypes were masked in each of the five runs. The
347 MTM.UN was used in case a trait showed genomic correlation larger than 0.3 with at least one other
348 trait. **b** Prediction accuracy of two multi-environment univariate models, i.e., across-environment G-
349 BLUP (G-BLUP.E) and marker by environment interaction G-BLUP (G-BLUP.E.G×E), and the multi-
350 environment multivariate factor-analytic model (MTM.FA). The models were applied under two five-
351 fold cross-validation scenarios CV1 and CV2. The CV1 was applied for all traits using G-BLUP.E and G-
352 BLUP.E.G×E and for traits measured in at least three environments using MTM.FA. The CV2 was applied
353 for traits measured in Switzerland and in at least a one other location. **a-b** Prediction accuracy was
354 estimated as a Pearson correlation coefficient between the observed and the predicted values of
355 genotypes whose phenotypes were masked in a five-fold cross-validation. For the multi-environment
356 models, the correlation coefficients were estimated for each environment separately. In the box plot,
357 the bottom and top line of the boxes indicate the 25th percentile and 75th percentile quartiles (the
358 interquartile range), the center line indicates the median (50th percentile). The whiskers extend from
359 the bottom and top line up to 1.5-times the interquartile range. The points beyond the 1.5-times the
360 interquartile range from the bottom and top line are labeled as dots.
361
362 **Synthesis of phenotypic and genomic analyses**
363 The across-environment clonal mean heritability was generally very high in the evaluated traits, the
364 value being close to one for harvest date and red over color and not lower than 0.80 for all the other
365 traits with the exception of full flowering (0.74), end of flowering (0.79) and water core grade (0.79,
366 Figure 6, Supplementary Table 6). The genomic heritability, which is the proportion of phenotypic
367 variance explained by the markers, was larger than 0.80 for harvest date, floral emergence, green color
368 and red over color, the value was not lower than 0.40 for all the other traits with the exception of
369 bitter bit frequency (0.33) and grade (0.39, Figure 6, Supplementary Table 6).
370 The effects of genotype and significantly associated markers together explained a substantial part of
371 the phenotypic variance of traits, the largest sums of these genotypic effects were observed for harvest

372 date (82.8%) and red over color (74.6%, Figure 6, Supplementary Table 6). Altogether, the sum of the
373 genotypic effects explained a very low proportion of the total variance for floral emergence (13.1%),
374 flowering intensity (11.4%), trunk diameter (10.9%) and trunk increment (8.7%). The major proportion
375 of the phenotypic variance was explained by the effect of environment for floral emergence (73.9%)
376 and trunk diameter (66.3%). The lowest impact of environment was found for traits measured at only
377 one location over two or three years such as fruit diameter or water core frequency, both showing an
378 effect of environment (i.e., year) below 1%. The effect of G×E was the most pronounced for
379 productivity traits, i.e., flowering intensity (23.7%), weight of fruits (20.8%) and number of fruits
380 (21.6%).. The proportion of the G×E effect was the lowest for harvest date (4.7%), floral emergence
381 (5.2%), red over color (5.9%) and trunk diameter (4.2%) among the traits measured at more than one
382 location and for end of flowering (5.7%), fruit volume (5.9%) and green color (3.9%) among the traits
383 measured at one location. A high proportion of the phenotypic variance remained unexplained by the
384 model parameters for flowering intensity (47.5%), bitter pit grade (53.4%) and trunk increment
385 (55.1%).
386 Hierarchical clustering of the phenotypic variance components revealed three clusters of traits (Figure
387 6). A strong genotypic effect and a comparably low effect of environment and G×E was observed for
388 13 traits assigned to the cluster one. Most of the phenotypic variance was explained by the effect of
389 environment in floral emergence and trunk diameter, which were grouped in cluster two. Finally, 15
390 traits with a pronounced effect of environment and/or G×E were grouped in cluster three.



391

392 **Figure 6: Synthesis of phenotypic and genomic analyses.** Across-environment clonal mean heritability,
393 genomic heritability, average prediction accuracy ($\bar{r}_t$) for the single-environment G-BLUP and the
394 proportion of phenotypic variance explained by the effect of each significantly associated marker (SNP
395 1–8), genotype (G), environment (E) and genotype by environment interaction (G×E). The significantly
396 associated markers corresponded to results of the global GWAS. Phenotypic variance components
397 were used to estimate clusters of traits outlined along the vertical axis. Within each cluster, the traits
398 were sorted according to $\bar{r}_t$.

399

Discussion

Discovered loci overlap between association studies and traits. Our GWAS permitted to enlighten the architecture of analyzed traits as well as the identification of numerous marker-trait associations stable across, and specific to, the locations of the apple REFPOP. The particular design of the experiment, including the diversity of the plant material used (accessions and small progeny groups), multiple locations, and multiple years of evaluation, resulted in about two thirds of the discovered associations being novel when compared with the loci published in studies spanning more than two decades. Our study design also allowed us to replicate the identification of many previously known loci associated with the studied traits.

The association of one locus with two or more seemingly independent traits (i.e., caused by pleiotropy) and linkage disequilibrium between loci associated with different traits are frequent for complex traits[47]. The GWAS performed in this study showed several marker-trait associations at identical or close positions for different traits. The interdependency between harvest date and fruit firmness, which can be also observed empirically for early cultivars that soften more, may be an example of pleiotropy or linkage disequilibrium between loci. Harvest date and fruit firmness are known to be regulated by ethylene production[48] and associated with loci present on chromosomes 3 (*NAC18.1*), 10 (*Md-ACO1*, *Md-PG1*), 15 (*Md-ACS1*) and 16[22,49-52].

In this work, closely located (distance <100 kb) associations with both harvest date and fruit firmness were found on chromosome 3. Migicovsky et al.[22] reported an overlap between associations with harvest time and fruit firmness on chromosome 3 falling within the coding region of *NAC18.1*. The authors hypothesized that the lack of associations on other chromosomes was likely due to low SNP density around the causal loci (the study used a GBS-derived 8K SNP dataset). The larger number of associations reported here might be a result of the high SNP density (303K SNPs) deployed in GWAS, however, not all previously reported loci were re-discovered.

The SNPs associated with harvest date and fruit firmness on chromosome 10 were further apart (~6 Mb). For harvest date, one of the associations on chromosome 10 was stable across locations and

426 several associations were location specific. However, the association on chromosome 10 with fruit
427 firmness was found for the Italian location only. It has been shown that chromosome 10 contains more
428 than one QTL controlling fruit firmness[49-51], but stable across-location association with fruit firmness
429 on chromosome 10 was missing in our study. One of the known loci on chromosome 10, the *Md-PG1*
430 gene, is responsible for the loss of fruit firmness after storage[51,53]. In apple REFPOP, fruit firmness was
431 measured within one week after the harvest date and this very short storage period might have
432 contributed to the less pronounced effect of the locus *Md-PG1* in our GWAS.
433 Two associations with harvest date measured in Italy but no association with fruit firmness were found
434 on chromosome 15. Although a marker for *Md-ACS1* related to ethylene production was previously
435 mapped on chromosome 15[50], and QTL for fruit firmness was discovered on the same chromosome[49],
436 these markers did not co-locate, but rather, mapped at the opposite extremes of chromosome 15[49,50].
437 Likewise, the connection between harvest date and fruit firmness on chromosome 15 could not be
438 confirmed here.
439 Our GWAS showed associations with harvest date and fruit firmness on chromosome 16, which were
440 located 38 kb apart. In the past, loci associated with harvest date and fruit firmness have been reported
441 in the same region on chromosome 16[26,49]. The role of this locus in the regulation of harvest date and
442 fruit firmness remains unknown and requires further research.
443 In practice, ripeness of fruit (harvest date) is decided based on ground color and starch content. The
444 GWAS results showed that the association on chromosome 3 was not only found for harvest date and
445 nearby markers associated with fruit firmness, but also corresponded to associations with ground color
446 and soluble solids content. This might be explained by the fact that these traits are used to define
447 ripeness and thus harvest date. Further, the association of the *NAC18.*1 locus on chromosome 3 with
448 overall russet frequency would support the known enhanced expression of *NAC* transcription factors
449 in russet skin[54].
450 Co-localizations between associations found for different measures of bitter pit on chromosome 16,
451 russet on chromosomes 12 and 17, fruit size on chromosomes 1, 8 and 11, and skin color on

chromosome 9 are likely the result of relatedness among trait measurements. The measures that are easiest to score can be used in future to phenotype these traits.

Signs of selection in marker-trait associations of large effect. The design of apple REFPOP allowed for the discovery of major marker-trait associations and for the analysis of changes in allele frequency between 30 ancestral accessions and 265 progeny included in the apple REFPOP. Comparing ancestors with the progeny, higher frequencies of the alleles associated with later harvest date and increased flowering intensity, titratable acidity, fruit firmness and trunk increment were found for the progeny. Of these traits, harvest date and fruit firmness are correlated, probably due to pleiotropy or linkage disequilibrium of causal loci, as it was shown in this and previous studies[22]. Consequently, the consistently higher frequency of alleles contributing to later harvest and firmer apples in the progeny is because the softening of harvested apples is undesirable and likely selected against[55]. Signs of selection for increased firmness were also recently found in USDA germplasm collection[5]. Our study also showed fixation of the late-harvest and high-firmness alleles on chromosome 3 in the whole progeny group, which suggests a loss of genetic diversity in the modern breeding material at this locus. For flowering intensity, a trait positively correlated with apple yield, a new locus was discovered on chromosome 14. The increased frequency of the allele contributing to higher flowering intensity in the progeny, its presence in all parental genotypes, and fixation in some parental combinations may be the result of breeding for high yield. The major locus found for acidity on chromosome 16 was consistent with the *Ma* locus frequently detected in various germplasm[8,11]. The total number of the high-acidity alleles for *Ma* and *Ma3*, which is another regularly detected acidity locus, was shown to be higher in parents of a European breeding program (Better3fruit, Belgium) than in parents used in the USDA breeding program[11,56]. The desired acidity level might depend on local climate of the breeding program and market preferences[56]. The increase in frequency of the allele contributing to higher acidity in the progeny may indicate a current preference towards more acidic apples in European breeding, but further investigation is needed to clarify the trend. The last locus of large effect showing allele frequency dynamics between generations was found for trunk increment. The allele

associated with an increase in trunk increment may have been selected in the progeny due to its potential impact on productivity suggested by moderate positive correlations between tree vigor (trunk diameter and increment) and yield-related traits. Additionally, the marker associated with trunk increment was 1.8 Mb apart from a SNP marker associated with *Rvi6* gene responsible for resistance against apple scab[10]. These two markers (AX-115183752 for trunk increment and AX-115182989 (also called Rvi6_42M10SP6_R193) for apple scab) showed a correlation of 0.15 and occurred within a region of increased linkage disequilibrium between markers (Supplementary Figure 10). All accessions were homozygous for the reference allele of AX-115183752 associated with decreased trunk increment (Figure 6c) except for 'Prima' and X6398, which were heterozygous. The scab-resistant accessions 'Prima' and X6398 (which is a second-generation offspring of 'Prima'[57]) but also 'Priscilla-NL' (known to be heterozygous for *Rvi6*[58]), were also heterozygous for AX-115182989. All other accessions were homozygous for the reference allele not associated with *Rvi6*. The allele on chromosome 1 associated with increased trunk increment may have been co-selected with the *Rvi6* locus responsible for resistance against apple scab.

Signs of intense selection for red skin were recently detected in the USDA germplasm collection when compared with progenitor species of the cultivated apple[5]. Our results show that the associations with red over color and green color, which phenotypically mirrored red over color and was associated with the same marker, did not show changes in allele frequency between ancestors and progeny included in the apple REFPOP. Some parental combinations showed almost exclusively the allele increasing red skin color, other parental combinations exhibited a lack of the allele. This uneven distribution of the alleles in the progeny group pointed to different directions of selection for fruit skin color in the European breeding programs (Figure 4b).

Performance of the single-environment univariate genomic prediction models. Single-environment univariate genomic prediction models were applied to individual traits after accounting for environmental effects and averaging across locations and/or years. The observed small differences between genomic prediction accuracies of various models (Figure 5a) were in accordance with

previous model comparisons where distinctions among models were negligible[39,59]. The largest extremes in prediction accuracy between traits were found with random forest, which allowed for the overall highest prediction accuracy among all compared models for red over color. The explanation for the striking performance of random forest for red over color might be found in the results of our GWAS. This trait of oligogenic architecture was associated with a few low-effect loci and one locus of large effect explaining 61% of the red over color phenotypic variance measured in the apple REFPOP. High correlations between many markers, i.e., linkage disequilibrium, were found in the vicinity of the large-effect locus (Supplementary Figure 10). Random forest is known to assign higher importance to correlated predictor variables (here the markers) in the tree building process[60], which may have contributed to the particularly high prediction accuracy found for red over color with random forest. The prediction accuracy for red over color reached ~0.4 in several former prediction studies[22,23,29,34] and was approximately doubled in our work, which demonstrated the potential of the current study design for accurate genomic predictions. For harvest date, the currently reported prediction accuracy of 0.78 was only slightly higher than the accuracy of 0.75 obtained with the initial apple REFPOP dataset measured during one year[36], but these accuracies showed a considerable improvement over other accuracies of approximately 0.5–0.6 reported elsewhere[22,23,29]. As shown before[36], these results underline the suitability of apple REFPOP design for the application of genomic prediction.

Prediction accuracy for traits such as yellow color or russet cover were on the opposite side of the spectrum when compared to harvest date and red over color. The prediction accuracy of yellow color and russet cover was low, although the genotypic effects explained 45% and 47% of the phenotypic variance, respectively. The across-environment clonal-mean heritability of russet cover was high (0.97), while the heritability for yellow color was slightly lower (0.81, Figure 6). Yellow color showed a moderate phenotypic correlation of 0.55 with russet cover, suggesting that the phenotyping device might have classified some russet skin as yellow color. Symptoms of powdery mildew could have been misinterpreted as russet skin. The decreased performance of genomic prediction models might stem

529 from inaccurate phenotyping methods, insufficient SNP density in the associated regions, or other
530 factors, all of which could not be explained in this work.

531 ***Role of genotype by environment interactions in multi-environment univariate genomic prediction.***
532 The multi-environment univariate genomic prediction models either directly estimated environmental
533 effects (across-environment G-BLUP, called here G-BLUP.E) or additionally borrowed genotypic
534 information across environments and thus considered the G×E (marker by environment interaction G-
535 BLUP, called here G-BLUP.E.G×E)[42]. The average accuracy of the G-BLUP.E.G×E model across traits was
536 only slightly higher than the accuracy of the G-BLUP.E. In contrast, the G-BLUP.E.G×E model had
537 substantially greater prediction accuracy than the G-BLUP.E model when applied in wheat[42]. In the
538 latter study, a productivity trait was measured under simulated conditions of mega-environments and
539 the effect of G×E explained up to ~60% of the phenotypic variance[42]. Our work only focused on
540 European environments and the largest proportion of phenotypic variance assigned to G×E was 24%
541 for a productivity trait (flowering intensity). Furthermore, the average proportion of G×E across traits
542 was approximately 12%, which may explain the mostly negligible differences between the G-BLUP.E
543 and G-BLUP.E.G×E models. Our results were in line with the low interaction of additive genetic effects
544 with location of up to ~6% obtained for apple fruit quality traits measured at two locations in New
545 Zealand[33], and the limited G×E reported for fruit maturity timing in sweet cherry across continents[61].
546 For approximately half of the tested traits, the G-BLUP.E.G×E did not outperform G-BLUP.E. For these
547 traits, the G-BLUP.E ignoring G×E may be sufficient to account for the environmental effects across
548 European sites because it is computationally simpler and therefore less demanding. Traits such as
549 flowering intensity, soluble solids content, trunk increment or traits related to fruit size and russet
550 showed an improved performance under G-BLUP.E.G×E when compared to G-BLUP.E. For traits
551 positively responding to G-BLUP.E.G×E, the G×E should be considered when making predictions across
552 environments. The highest improvement of prediction accuracy with G-BLUP.E.G×E when compared
553 to G-BLUP.E was found for flowering intensity, the difference between the models amounting to 0.07
554 (Figure 5b). This result might be explained by the highest contribution of G×E to the phenotypic

555 variance of flowering intensity among all traits (Figure 6). A comparably high contribution of G×E was
556 also found for weight of fruits and number of fruits, though no improvement with G-BLUP.E.G×E model
557 was observed for these traits. When comparing the relative contributions of variance components to
558 the phenotypic variance of flowering intensity, weight of fruits and number of fruits, the genotype
559 explained 11%, 22% and 21%, the environment 17%, 24% and 25%, and the G×E 24%, 21% and 22%,
560 respectively. Although the proportions of G×E were similar in the three compared traits, the effects of
561 genotype and environment explained a higher proportion of the variance for weight of fruits and
562 number of fruits than for flowering intensity. This may have contributed to the surprisingly lower
563 accuracy of the G-BLUP.E.G×E model when compared with G-BLUP.E for weight of fruits and number
564 of fruits, but additional investigations may be needed to clarify this result in the future.
565 The G-BLUP.E.G×E model assumes positive correlations between environments and is therefore mostly
566 suitable for the joint analysis of correlated environments[42,62]. As shown by Lopez-Cruz et al.[42] and in
567 our study, this assumption of G-BLUP.E.G×E resulted in the best model performance for traits showing
568 high positive correlations between environments (here harvest date and red over color) and the worst
569 performance for traits exhibiting low correlations between environments (here flowering intensity and
570 trunk increment, Figure 5b, Supplementary Table 2, Supplementary Figure 3). For flowering intensity
571 and trunk increment, bivariate prediction of the environments or prediction with a different G×E model
572 not assuming positive correlations between environments might be more appropriate than the
573 currently applied approach[42,63].
574 ***Multivariate models as a useful element in the genomic prediction toolbox.*** Multivariate (also called
575 multi-trait) models were shown to be useful for predicting traits that are costly to phenotype when a
576 correlated trait less expensive to phenotype was available[45]. In our study, when the prediction accuracy
577 of the single-environment multivariate model MTM.UN was compared with the baseline model G-
578 BLUP, several combinations of related and unrelated traits led to increased accuracy. For the related
579 traits with a high phenotypic correlation (Figure 1a), prediction of traits measured at one location were
580 often improved when a related trait measured across different locations was included. This was the

case for the combination of floral emergence with full flowering and end of flowering and for single fruit weight combined with fruit diameter, fruit length, maximum fruit size and fruit volume. Inclusion of soluble solids content in MTM.UN resulted in increased prediction accuracy for russet cover, although the traits showed only a moderate correlation and no obvious explanation for this result could be found. Our study supports the potential of multivariate models to borrow information that correlated traits provide about one another and identified trait combinations that can be successful under the multivariate setup.

In place of the correlated traits, environments of a single trait can be implemented in a multivariate model[46]. The average prediction accuracy over all traits was ~0.04 lower in the multi-environment multivariate (MTM.FA) than in the multi-environment univariate genomic prediction models (G-BLUP.E and G-BLUP.E.G×E). Compared to G-BLUP.E and G-BLUP.E.G×E, the MTM.FA showed the potential to perform equally well for six (CV1) and three traits (CV2) and was able to outperform both models for two (CV1) and five traits (CV2). In cases where MTM.FA outperformed G-BLUP.E and G-BLUP.E.G×E, a very limited increase in prediction accuracy of 0.01 was found for all traits but trunk increment, for which the increase was equal to 0.07 under the second cross-validation scenario. Except for the noticeable increase in prediction accuracy for trunk increment that could not be explained by our analyses, the performance of MTM.FA was similar to G-BLUP.E and G-BLUP.E.G×E, which establishes the multivariate model as a useful tool for multi-environment genomic prediction in apple.

Two approaches to genomic prediction addressed with cross-validation scenarios. The cross-validation scenarios CV1 and CV2 were applied with multi-environment genomic prediction models to test two genomic prediction approaches typically faced in breeding. The CV1 imitated evaluation of breeding material that was yet untested in field trials. The CV2 was implemented to simulate incomplete field trials where breeding material was evaluated in some but not all target environments. More specifically, the CV2 investigated a situation where the breeding material has been evaluated at one location (the breeding site, in this case Switzerland) and the material's potential over other European sites was predicted without its assessment in a multi-environment trial, which may increase

selection efficiency at latter stages of evaluation. As CV2 provided more phenotypic information to the models than CV1, a higher genomic prediction accuracy was found under CV2 when compared with CV1, which was anticipated[33,42]. The CV2 was tested by calibrating the model with Swiss observations only. The application of CV2 could be extended to other apple REFPOP locations to provide useful information for the breeding programs located at these sites. The choice of cross-validation scenario did not affect the general ranking of the average genomic prediction accuracies estimated for the evaluated traits.

Implications for apple breeding. Phenotypic variance decomposition into genetic, environmental, G×E and residual effects was compared with the results of GWAS and genomic prediction as well as heritability estimates. The comprehensive comparison indicated three classes of traits with contrasting genetic architecture and prediction performance. Characteristics of these trait classes and proposals for their efficient prediction strategies are described in the following paragraphs.

The first class included harvest date and red over color that showed a few loci of large effect and some additional loci of low effect, the highest prediction accuracies, and the highest across-environment clonal-mean heritability among all traits. Both traits showed a very high proportion of the genotypic effect explaining ~75% of the phenotypic variance. For harvest date and red over color, the marker with the largest effect explained 52% and 59% of the phenotypic variance and all marker effects in genomic prediction captured together 88% and 85% of the phenotypic variance (i.e., genomic heritability of 0.88 and 0.85), respectively. Selection for these traits exhibiting a strong genetic effect of one locus could be done using marker-assisted selection, although only a part of the variance would be explained by a single marker. Better results can be achieved using genomic prediction, as this was able to explain a substantially larger amount of the phenotypic variance. Other traits such as fruit firmness, titratable acidity, end of flowering or traits related to fruit size and water core were grouped in the same cluster as harvest date and red over color (Figure 6). These traits showed a strong genotypic effect and a comparably low effect of environment and G×E, suggesting that selection for

the traits would be efficient when performed using single-environment genomic prediction models rather than multi-environment prediction.

The second class of traits was represented by floral emergence and trunk diameter displaying a high proportion of the environmental effect (~70%) and a similar ratio of variance explained by genotypic effects compared to variance explained by G×E effects (~2.5). The genomic prediction accuracy did not considerably deviate from the average accuracy over all traits. Several marker associations with these traits were identified using location-specific GWAS. However, in the across-location GWAS, only one association explaining a very small part of phenotypic variance (floral emergence) or no association (trunk diameter) were discovered. Consequently, such traits predominantly driven by the effect of environment can be successfully selected based on genomic prediction, but the lack of associations stable across environments limits the applicability of marker-assisted selection to this class of traits.

In the third class, the productivity traits (flowering intensity, weight of fruits and number of fruits) showed the largest proportion of variance explained by G×E (~20%), with similar amounts of variance explained by genotypic effects for weight of fruits and number of fruits, but half as much variance explained by genotypic effects for flowering intensity (Figure 6). As a consequence, only flowering intensity showed higher prediction accuracy with G-BLUP.E.G×E than G-BLUP.E model. As shown above, the G×E should be considered when making predictions across environments for traits responding positively to the G-BLUP.E.G×E model, but G-BLUP.E may be sufficient for other traits to account for the environmental effects. To our knowledge, this is the first report of genomic prediction for apple yield components and our results can aid the establishment of productivity predictions in apple breeding. Other traits falling within the same cluster as the productivity traits, namely full flowering, ground color, yellow color, soluble solids content, trunk increment, and traits related to bitter pit and russet, showed a pronounced effect of environment and/or G×E (Figure 6). Multi-environment genomic prediction models can be efficient when applying genomic selection to these traits.

The decision to apply either marker-assisted or genomic selection can be based on genetic architecture of traits of interest and resources available in a breeding program. For breeding of yet genetically unexplored traits, variance decomposition of historical phenotypic data prior to genomic analyses may help describe trait architecture, assign traits to one of the three classes described in the previous paragraphs, and finally determine the most appropriate method of genomics-assisted breeding. From all traits explored in this study, the marker-trait associations with large and stable effects across environments found for harvest date, flowering intensity, green color, red over color, titratable acidity, fruit firmness and trunk increment could be implemented into DNA tests for marker-assisted selection. These tests would allow for a reduction of labor costs in a breeding program when removing inferior seedlings without phenotyping[7]. Although generally requiring more statistical competences than marker-assisted selection, genomic selection can make use of both large- and low-effect associations between markers and traits when accommodating thousands of marker effects in a single genomic prediction model. For all studied traits, our results showed that marker effects estimated in genomic prediction were able to capture a larger proportion of the phenotypic variance than individual markers associated with the traits. Therefore, genomic selection should become the preferred method of genomics-assisted breeding for all quantitative traits explored in this study to ultimately increase their breeding efficiency and genetic gain.

Conclusion

This study laid the groundwork for marker-assisted and genomic selection across European environments for 30 quantitative apple traits. The apple REFPOP experimental design facilitated identification of a multitude of novel and known marker-trait associations. Our multi-environment trial provided accurate genomics-estimated breeding values for apple genotypes under various environmental conditions. Limited G×E detected in this work suggested consistent performance of genotypes across different European environments for most studied traits. Utilizing our dataset, more efficient selection of traits related to yield may lead to higher productivity and increased genetic gain

683 in the future[37]. Improved fruit quality would appeal to consumers and tree phenology could be
684 synchronized with current and future climates to secure production. The genomic prediction models
685 developed here can be readily used for selecting germplasm in breeding programs, thus providing
686 breeders with tools increasing selection efficiency. Beside the apple REFPOP, one other large multi-
687 environment reference population for fruit trees, the PeachRefPop[64], was designed in Europe.
688 Application of our study design to other horticultural crops such as peach can promote broader use of
689 genomics-assisted breeding in the future.
690

691 ## Methods

692 ### Plant material

693 The apple REFPOP was designed and established by the collaborators of the FruitBreedomics project[65]
694 as described by Jung et al.[36]. The apple REFPOP consists of 534 genotypes from two groups of diploid
695 germplasm. The accession group consists of 269 accessions of European and non-European origin
696 representing the diversity in cultivated apple. The progeny group of 265 genotypes stemmed from 27
697 parental combinations produced in the current European breeding programs. In 2016, the apple
698 REFPOP was planted in six locations representing several biogeographical regions in Europe, in (i)
699 Rillaar, Belgium, (ii) Angers, France, (iii) Laimburg, Italy, (iv) Skierniewice, Poland, (v) Lleida, Spain and
700 (vi) Wädenswil, Switzerland (one location per country). Every genotype was replicated at least twice
701 per location. All plants included in this study were treated with agricultural practice common to each
702 location. Calcium spraying was avoided due to its influence on bitter pit. Flowers were not thinned,
703 but the fruits were hand-thinned after the June fruit drop and up to two apples per fruit cluster were
704 retained.
705

706 ### Genotyping

707 A high-density genome-wide SNP marker dataset was produced as reported by Jung et al.[36]. Briefly,
708 SNPs from two overlapping SNP arrays of different resolution, (i) the Illumina Infinium® 20K SNP

709 genotyping array[20] and (ii) the Affymetrix Axiom® Apple 480K SNP genotyping array[21], were curated
710 and then joined applying imputation with Beagle 4.0[66] using the recently inferred pedigrees[4]. Non-
711 polymorphic markers were removed to obtain a set of 303,148 biallelic SNPs. Positions of SNPs were
712 based on the apple reference genome obtained from the doubled haploid GDDH13 (v1.1)[16].
713
714 **Phenotyping**
715 Thirty phenotypic traits related to phenology, productivity, fruit size, outer fruit, inner fruit, and vigor
716 were evaluated at up to six locations of the apple REFPOP during up to three seasons (2018–2020).
717 Trunk diameter was measured in 2017 in some locations, enabling for a trunk increment calculation
718 for 2018. The traits were recorded as described in the Supplementary Methods. Two phenology traits
719 measured in 2018, i.e., floral emergence and harvest date, were previously analyzed by Jung et al.[36].
720
721 **Phenotypic data analyses**
722 Spatial heterogeneity was modeled separately for each trait and environment (nested factor of
723 location and year) using the spatial analysis of field trials with splines (SpATS) to account for the
724 replicate effects and differences due to soil characteristics[67]. Phenotypic values of traits adjusted for
725 spatial heterogeneity within each environment were estimated at the level of trees (adjusted
726 phenotypic values of each tree) and genotypes (adjusted phenotypic values of each genotype)[36].
727 The general statistical model for the following phenotypic data analyses fitted via restricted maximum
728 likelihood (R package lme4[68]) was:
729 $$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zb} + \boldsymbol{\varepsilon}$$ (Equation 1)
730 where $\boldsymbol{y}$ was a vector of trait phenotypes, $\boldsymbol{X}$ the design matrix for the fixed effects, $\boldsymbol{\beta}$ the vector of
731 fixed effects, $\boldsymbol{Z}$ the design matrix for the random effects, $\boldsymbol{b}$ the vector of random effects and $\boldsymbol{\varepsilon}$ the
732 vector of random errors. The $\boldsymbol{b}$ was a $q \times 1$ vector assuming $\boldsymbol{b} \sim N(0, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma}$ was a variance-
733 covariance matrix of the random effects. The assumptions for the $N \times 1$ vector of random errors were
734 $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{I}\sigma_{\varepsilon}^2)$ with $N \times N$ identity matrix $\boldsymbol{I}$ and the variance σ_{ε}^2, the N being the number of trees.

735 To assess the reliability of environment-specific data, a random-effects model was first fitted
736 separately for each trait and environment to estimate an environment-specific clonal mean
737 heritability. Applying the Equation 1, the response $\boldsymbol{y}$ was a vector of the raw (non-adjusted) phenotypic
738 values of each tree. On the place of $\boldsymbol{X}$, a vector of ones was used to model the intercept β. The vector
739 of genotypes acted as a random effect in $\boldsymbol{Z}$**.** The environment-specific clonal mean heritability was
740 calculated from the variance components of the random-effects model as:

741 $$H^2 = \frac{\sigma_g^2}{\sigma_p^2} \text{ (Equation 2)}$$

742 where the phenotypic variance $\sigma_p^2 = \sigma_g^2 + \sigma_\varepsilon^2/\bar{n}_r$ was obtained from the genotypic variance σ_g^2, error
743 variance σ_ε^2 and the mean number of genotype replications $\bar{n}_r$. The environment-specific clonal mean
744 heritability was used to eliminate location-year-trait combinations with a heritability value below 0.1.
745 For the remaining location-year combinations, a mixed-effects model following the Equation 1 was
746 fitted to the vector of the adjusted phenotypic values of each tree as response ($\boldsymbol{y}$). The effects of
747 environments, i.e., combination of location and years, were used as fixed effects and the effects of
748 genotypes and genotype by environment interactions as random effects. Estimated variances of the
749 model components were used to evaluate the across-environment clonal mean heritability calculated
750 using the Equation 2 with the phenotypic variance estimated as:

751 $$\sigma_p^2 = \sigma_g^2 + \frac{\sigma_{ge}^2}{n_e} + \frac{\sigma_\varepsilon^2}{n_e \bar{n}_r} \text{ (Equation 3)}$$

752 where σ_{ge}^2 was the genotype by environment interaction variance and n_e represented the number of
753 environments.
754 An additional mixed-effects model following the Equation 1 was fitted to the adjusted phenotypic
755 values of each tree ($\boldsymbol{y}$) using the effects of location, year and their interaction as fixed effects and the
756 effects of genotypes as random effects. Due to the skewness of their distributions, $\boldsymbol{y}$-values of the
757 traits weight of fruits, number of fruits and trunk diameter were log-transformed. BLUPs ($\widehat{\boldsymbol{b}}$) extracted
758 from the model were further denoted as across-location BLUPs. To estimate the location-specific
759 BLUPs, a model according to the Equation 1 was fitted with a subset of the adjusted phenotypic values
760 of each tree from single locations ($\boldsymbol{y}$) using the effects of years as fixed effects and the effects of

761 genotypes as random effects. The across-location BLUPs and the adjusted phenotypic values of each
762 genotype were used to assess phenotypic correlation as the Pearson correlation between pairs of traits
763 and between pairs of environments within traits, respectively. The across-location BLUPs with the
764 addition of location-specific BLUPs for traits measured at a single location were further denoted as the
765 main BLUPs. In the main BLUPs, the missing values were replaced with the mean of the BLUPs of the
766 same trait and the data were scaled and centered to finally estimate a principal component analysis
767 biplot[69], where multivariate normal distribution was assumed for the ellipses.
768
769 **Genome-wide association studies**
770 The Bayesian-information and linkage-disequilibrium iteratively nested keyway (BLINK)[70] implemented
771 in the R package GAPIT 3.0[71] was applied using the genomic matrix $\boldsymbol{M}$, an $n \times m$ matrix for a population
772 of size $n = 534$ genotypes (i.e., accessions and progeny) with $m = 303{,}148$ markers, with across-
773 location BLUPs (across-location GWAS) or location-specific BLUPs (location-specific GWAS) as the
774 response. BLINK was used with two principal components and the minor allele frequency threshold
775 was set to 0.05. Marker-trait associations were identified as significant for p-values falling below a
776 Bonferroni-corrected significance threshold $\alpha^* = \alpha/m$ with $\alpha = 0.05$ $(-log_{10}(p) > 6.74)$. The
777 proportion of phenotypic variance explained by each significantly associated SNP was assessed with a
778 coefficient of determination (R^2). The R^2 was estimated from a linear regression model, which was
779 fitted with a vector of SNP marker values (coded as 1, 2, 3) as predictor and either the across-location
780 BLUPs or location-specific BLUPs as response. GWAS based on the across-location BLUPs with the
781 addition of location-specific BLUPs, in cases where traits were measured at a single location only, was
782 further denoted as the global GWAS. The position of the last SNP on a chromosome was used to
783 estimate chromosome length, which was used to divide each chromosome into three equal segments,
784 i.e., top, center and bottom. The marker-trait associations were assigned to these chromosome
785 segments based on their positions.

Previous reports on QTL mapping and GWAS in apple were reviewed to perform an extensive comparison with our GWAS results (Supplementary Table 4). Published results for traits measured similarly to the traits studied in the present work were considered, with the traits being assembled into trait groups: harvest time (harvest date and similar), flowering time (floral emergence, full flowering, end of flowering and similar), productivity (flowering intensity, weight of fruits, number of fruits and similar), fruit size (single fruit weight, fruit diameter, fruit length, maximum fruit size, fruit volume and similar), ground color (ground color, yellow color and similar), over color (red over color, green color and similar), bitter pit (bitter pit frequency, bitter pit grade and similar), russet (russet cover, russet frequency overall, at stalk, on cheek and in the eye and similar), acidity (titratable acidity and similar), sugar (soluble solids content and similar), firmness (fruit firmness and similar), water core (water core frequency, water core grade and similar) and trunk (trunk diameter, trunk increment and similar). The positions of published associations within respective chromosomes were visually assigned to the three chromosome segments, i.e., top, center and bottom. The total number of markers used was recorded (Supplementary Table 4). Where the number of overlapping markers between the maternal and paternal linkage maps was not provided in a publication, the marker numbers for both maps were summed.

In the global GWAS results, the allele frequency was studied over generations. The ancestors of genotypes were identified making use of the apple pedigrees of Muranty et al.[4]. For all significant marker-trait associations from the global GWAS, frequency of the allele associated with increased phenotypic value was estimated for the progeny group and for its five ancestor generations. To represent the ancestors, the allele frequency was estimated for the 30 accessions of them included in the apple REFPOP. For major significant marker-trait associations with $R^2 > 0.1$ reported in the global GWAS, linkage disequilibrium was estimated as squared Pearson's correlations in a window of 3,000 markers surrounding each of the associations. A smaller window size was used for associations located towards the end of a chromosome.

A mixed-effects model following the Equation 1 was fitted to the vector of the adjusted phenotypic values of each tree as response ($\boldsymbol{y}$) using the effects of environments as fixed effects and the effects of genotypes, genotype by environment interactions, and additional effects for each SNP significantly associated with the trait (a factor of the respective SNP values in $\boldsymbol{M}$) as random effects. In cases where traits with no marker-trait associations were found in the global GWAS, the additional random effects of significantly associated SNPs were omitted from the model. The mixed-effects model for every trait was used to estimate proportions of phenotypic variance explained by the model components as described in Jung et al.[36]. The proportions of phenotypic variance explained by the random effects of genotypes and significantly associated SNPs were summed to obtain the proportion of variance explained by a genotypic effect. The proportions of phenotypic variance explained by genotypic, environmental, genotype by environment interaction, and residual effects were scaled and centered to be finally used for discovering similarities between the traits. For this purpose, a hierarchical clustering following Ward[72] was applied to the distance matrix of the set of effects. The number of clusters was estimated from a dendrogram, which was cut where the distance between splits was the largest.

Genomic prediction

The general statistical model for genomic prediction was

$$\boldsymbol{y} = 1\boldsymbol{\mu} + \boldsymbol{u} + \boldsymbol{\varepsilon} \quad \text{(Equation 4)}$$

where $\boldsymbol{y}$ was a vector of trait phenotypes, $\boldsymbol{\mu}$ was an intercept, $\boldsymbol{u}$ represented a vector of random effects and $\boldsymbol{\varepsilon}$ was a vector of residuals. Different vectors of $\boldsymbol{y}$ and assumptions for $\boldsymbol{u}$ and $\boldsymbol{\varepsilon}$ were used across eight single- and multi-environment genomic prediction models.

Single-environment genomic prediction. The single-environment genomic prediction models were fitted after the environmental effects were accounted for during the phenotypic data analysis, a process also called two-step genomic prediction. Therefore, the across-location BLUPs and location-specific BLUPs acted here as phenotypes from a single environment. Four univariate prediction models

and one multivariate model were implemented. First, regression with random forest (RF) was performed[73]. In this and the following three univariate models, the response $\boldsymbol{y}$ was defined as a $n \times 1$ vector of the main BLUPs. The centered and scaled additive genomic matrix $\boldsymbol{M}$, an $n \times m$ matrix for a population of size $n = 534$ with $m = 303{,}148$ markers, was used as further input. The number of trees to grow in the RF was 500 and the number of variables randomly sampled as candidates at each split was (rounded down) $mtry = m/3$. Second, BayesCπ was applied[74], where the random marker effects $\boldsymbol{u} = \sum_{k=1}^{m} z_k a_k$ with z_k an $n \times 1$ vector of the number of copies of one allele at the marker k and a_k being the additive effect of the marker k. The prior for a_k depended on the variance $\sigma^2_{a_k}$ and the prior probability π that a marker k had zero effect, the priors of all marker effects having a common variance $\sigma^2_{a_k} = \sigma^2_a$. The π parameter was treated as an unknown with uniform(0,1) prior. The random vector of residual effects followed a normal distribution $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{I}\sigma^2_\varepsilon)$ with $n \times n$ identity matrix $\boldsymbol{I}$ and the variance σ^2_ε. Third, the Bayesian reproducing kernel Hilbert spaces regression (RKHS) was implemented using a multi-kernel approach[75]. The multi-kernel model was fitted with $L = 3$ random marker effects $\boldsymbol{u} = \sum_{l=1}^{L} \boldsymbol{u_l}$ following a distribution $\boldsymbol{u} \sim N(0, \boldsymbol{K_l}\sigma^2_{ul})$, with $\boldsymbol{K_l}$ being the reproducing kernel evaluated at the lth value of the bandwidth parameter $h = \{h_1, \ldots, h_L\} = \{0.1, 0.5, 2.5\}$ and the variance σ^2_{ul}. For each random effect, the kernel matrix $\boldsymbol{K} = \{K(x_i, x_{i'})\}$ was an $n \times n$ matrix $K(x_i, x_{i'}) = \exp\{-h \times D_{ii'}\}$, where $\boldsymbol{D} = \left\{D_{ii'} = \frac{\sum_{k=1}^{m}(x_{ik} - x_{i'k})^2}{m}\right\}$ was the average squared-Euclidean distance matrix between genotypes, and x_{ik} the element on line i (genotype i) and column k (kth marker) of the centered and scaled additive genomic matrix $\boldsymbol{M}$. The residual effect assumed $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{I}\sigma^2_\varepsilon)$. Fourth, from the centered and scaled additive genomic matrix $\boldsymbol{M}$, the genomic relationship matrix $\boldsymbol{G}$ was computed as $\boldsymbol{G} = \boldsymbol{M}\boldsymbol{M}'/m$ and used to fit the genomic-BLUP (G-BLUP) model applying a semi-parametric RKHS algorithm, with the random marker effects following $\boldsymbol{u} \sim N(0, \boldsymbol{G}\sigma^2_u)$ with variance σ^2_u and the model residuals assuming $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{I}\sigma^2_\varepsilon)$[76]. Fifth, a multivariate model with an unstructured covariance matrix of the random marker effect (here abbreviated as MTM.UN) was fitted for chosen pairs of traits using the Bayesian multivariate Gaussian model environment MTM (http://quantgen.github.io/MTM/vignette.html). The main BLUPs acted as

the response $\boldsymbol{y}$, which was a vector of length $n \cdot t$ with $t = 2$ being the number of traits used in the model. The vector of the random marker effects followed $\boldsymbol{u} \sim N(0, \boldsymbol{U} \otimes \boldsymbol{G})$ where $\boldsymbol{U}$ was an unstructured covariance matrix of the random marker effect with dimension $t \times t$. Model residuals assumed $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{R} \otimes \boldsymbol{I})$ with $\boldsymbol{R}$ being an unstructured covariance matrix of the residual effect. To choose the pairs of traits for MTM.UN, a G-BLUP model was applied using all genotypes to estimate genomic BLUPs, which were then used to obtain pairwise genomic correlations between traits. The pairs with the genomic correlations larger than 0.3 were retained for the MTM.UN analysis. In case a trait was included in more than one pair of traits, the result for the pair with the highest average predictive ability for this trait was reported.

BayesCπ, RKHS, G-BLUP and MTM.UN were applied with 12,000 iterations of the Gibbs sampler, a thinning of 5, and a burn-in of 2,000 discarded samples. With all models, a five-fold cross-validation repeated five times was performed, generating 25 estimates of prediction accuracy. The folds were chosen randomly without replacement to mask phenotypes of 20% of the genotypes in each run. Prediction accuracy was estimated as a Pearson correlation coefficient between phenotypes of the masked genotypes and the predicted values for the same genotypes. The RF model was implemented in the R package ranger[77], the models BayesCπ, RKHS and G-BLUP in the R package BGLR[78] and the MTM.UN model in the R package MTM (http://quantgen.github.io/MTM/vignette.html).

Multi-environment genomic prediction. Two univariate multi-environment genomic prediction algorithms and one multivariate multi-environment algorithm were implemented, the response $\boldsymbol{y}$ being a vector of the adjusted phenotypic values of each genotype of length $n \times r$ with r equal to the number of environments (nested factor of location and year). The two univariate multi-environment models reported by Lopez-Cruz et al.[42] and implemented in the R package BGLR[78] were applied to explore the effects of genotypes, environments and their interaction in genomic prediction. Of the two models, the across-environment G-BLUP model (G-BLUP.E) assumed that marker effects were constant across environments. The random marker effects followed $\boldsymbol{u} \sim N(0, \boldsymbol{G_0}\sigma_u^2)$ where $\boldsymbol{G_0} = \boldsymbol{J} \otimes \boldsymbol{G}$, the $\boldsymbol{J}$ being an $r \times r$ matrix of ones. The model residuals assumed $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{I}\sigma_\varepsilon^2)$. Additionally to the

889 constant effects of markers across environments as assumed in the previous model, the marker by
890 environment interaction G-BLUP model (G-BLUP.E.G×E) allowed the marker effects to change across
891 environments, i.e., to borrow information across environments. The random marker effects were
892 defined as $\boldsymbol{u} = \boldsymbol{u_0} + \boldsymbol{u_1}$ where $\boldsymbol{u_0} \sim N(0, \boldsymbol{G_0}\sigma^2_{u0})$ and $\boldsymbol{u_1} \sim N(0, \boldsymbol{G_1})$ with

893 $$\boldsymbol{G_1} = \begin{bmatrix} \sigma^2_{u1}\boldsymbol{G} & 0 & 0 \\ 0 & \sigma^2_{u2}\boldsymbol{G} & 0 \\ 0 & 0 & \sigma^2_{u3}\boldsymbol{G} \end{bmatrix}$$

894 assuming $r = 3$ here for easier notation. The model residuals assumed $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{I}\sigma^2_{\varepsilon})$. Finally, a
895 multivariate multi-environment factor-analytic model (here abbreviated as MTM.FA) using the
896 Bayesian multivariate Gaussian model environment implemented in the R package MTM
897 (http://quantgen.github.io/MTM/vignette.html) was fitted to the data. As in the previous two models,
898 phenotypes of the same trait from multiple environments acted as response, although this model was
899 originally designed to analyze multiple traits. The traits measured at only one location during two
900 seasons (full flowering, end of flowering, fruit volume, water core frequency and water core grade)
901 were not modeled using MTM.FA because the analysis required at least three environments. The
902 vector of the random marker effects assumed $\boldsymbol{u} \sim N(0, \boldsymbol{C} \otimes \boldsymbol{G})$ where $\boldsymbol{C}$ was an $r \times r$ covariance
903 matrix. For the factor analysis, the $\boldsymbol{C} = \boldsymbol{B}\boldsymbol{B}' + \boldsymbol{\Psi}$ where $\boldsymbol{B}$ was a matrix of loadings (regressions of the
904 original random effects into common factors) and $\boldsymbol{\Psi}$ was a diagonal matrix whose entries gave the
905 variances of environment-specific factors. The loadings were estimated for all environments and the
906 variance of the Gaussian prior assigned to the unknown loadings was set to 100. The model residuals
907 assumed $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{R} \otimes \boldsymbol{I})$ with $\boldsymbol{R}$ being an unstructured covariance matrix of the residual effect.
908 All three multi-environment genomic prediction models were applied with 12,000 iterations of the
909 Gibbs sampler, a thinning of 5 and a burn-in of 2,000 discarded samples. The folds of a five-fold cross-
910 validation were chosen randomly without replacement. The cross-validation was repeated under two
911 scenarios. In the first cross-validation scenario (CV1), the phenotypes of 20% of the genotypes were
912 masked across all environments. For the second cross-validation scenario (CV2), the phenotypes of
913 20% of the genotypes were masked across all environments except for three Swiss environments, i.e.,

914 phenotypes of all genotypes from the environments “CHE.2018”, “CHE.2019” and “CHE.2020” were
915 used for model training. Ten traits were measured in only one location and therefore excluded from
916 CV2 (i.e., full flowering, end of flowering, fruit diameter, fruit length, maximum fruit size, fruit volume,
917 yellow color, green color, water core frequency and water core grade). Prediction accuracy was
918 estimated as a Pearson correlation coefficient between the phenotypes of the masked genotypes and
919 the predicted values for these genotypes. The correlations were estimated for each predicted
920 environment separately.
921
922 **Genomic heritability**
923 The BayesCπ model was applied for each trait as described before but trained with a full set of the
924 main BLUPs as response. The genomic heritability $h^2 = V_g/(V_g + V_e)$ was estimated as the proportion
925 of phenotypic variance explained by the markers, where V_g and V_e represented the amount of
926 phenotypic variance explained and unexplained by the markers, respectively[79,80]. The genomic
927 heritability was calculated from the marker effects saved in each iteration and averaged over iterations
928 to obtain the mean genomic heritability per trait.
929
930 ## Data availability
931 All SNP genotypic data generated with the 480K array used in this study have been deposited in the
932 INRAe dataset archive (https://data.inrae.fr/) at https://doi.org/10.15454/IOPGYF. All SNP genotypic
933 data generated using the 20K array used in this study have been deposited in the INRAe dataset archive
934 at https://doi.org/10.15454/1ERHGX. The raw phenotypic data generated during the study are
935 available in the INRAe dataset archive at (TBA upon acceptance).
936
937 ## Acknowledgements
938 The authors thank the field technicians and staff, especially Sylvain Hanteville, at INRAe IRHS and
939 Experimental Unit (UE Horti), Angers, France, and technical staff at other apple REFPOP sites for the

940 maintenance of the orchards and phenotypic data collection. We thank Dr. Graham Dow for English
941 language editing. Phenotypic data collection was partially supported by the Horizon 2020 Framework
942 Program of the European Union under grant agreement No 817970 (project INVITE: "Innovations in
943 plant variety testing in Europe to foster the introduction of new varieties better adapted to varying
944 biotic and abiotic conditions and to more sustainable crop management practices"). This work was
945 partially supported by the project RIS3CAT (COTPA-FRUIT3CAT) financed by the European Regional
946 Development Fund through the FEDER frame of Catalonia 2014-2020 and by the CERCA Program from
947 Generalitat de Catalunya. We acknowledge financial support from the Spanish Ministry of Economy
948 and Competitiveness through the “Severo Ochoa Programme for Centres of Excellence in R&D” 2016-
949 2019 (SEV-20150533) and 2020-2023 (CEX2019-000902-S). C.D. was supported by "DON CARLOS
950 ANTONIO LOPEZ" Abroad Postgraduate Scholarship Program, BECAL-Paraguay. We dedicate this paper
951 to Prof. Edward Zurawicz of the National Institute of Horticultural Research in Skierniewice, Poland
952 who co-promoted this study, but sadly recently passed away.

953

954 **Competing interests**

955 The authors declare no competing interests.

956

957 **References**

958 1 FAOSTAT (Food and Agriculture Organization of the United Nations, 2019).
959 2 Cornille, A., Giraud, T., Smulders, M. J. M., Roldán-Ruiz, I. & Gladieux, P. The domestication
960 and evolutionary ecology of apples. *Trends in Genetics* **30**, 57-65,
961 doi:10.1016/j.tig.2013.10.002 (2014).
962 3 Way, R. D. *et al.* Apples (*Malus*). *Acta Horticulturae*, 3-46, doi:10.17660/ActaHortic.1991.290.1
963 (1991).

964 4 Muranty, H. *et al.* Using whole-genome SNP data to reconstruct a large multi-generation
965 pedigree in apple germplasm. *BMC Plant Biology* **20**, 2, doi:10.1186/s12870-019-2171-6
966 (2020).
967 5 Migicovsky, Z. *et al.* Genomic consequences of apple improvement. *Horticulture Research* **8**, 9,
968 doi:10.1038/s41438-020-00441-7 (2021).
969 6 Urrestarazu, J. *et al.* Analysis of the genetic diversity and structure across a wide range of
970 germplasm reveals prominent gene flow in apple at the European level. *BMC Plant Biology* **16**,
971 130, doi:10.1186/s12870-016-0818-0 (2016).
972 7 Wannemuehler, S. D. *et al.* A cost–benefit analysis of DNA informed apple breeding.
973 *HortScience horts* **54**, 1998, doi:10.21273/hortsci14173-19 (2019).
974 8 Maliepaard, C. *et al.* Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using
975 multi-allelic markers. *Theoretical and Applied Genetics* **97**, 60-73, doi:10.1007/s001220050867
976 (1998).
977 9 Kenis, K., Keulemans, J. & Davey, M. W. Identification and stability of QTLs for fruit quality
978 traits in apple. *Tree Genetics & Genomes* **4**, 647-661, doi:10.1007/s11295-008-0140-6 (2008).
979 10 Jänsch, M. *et al.* Identification of SNPs linked to eight apple disease resistance loci. *Molecular*
980 *Breeding* **35**, 45, doi:10.1007/s11032-015-0242-4 (2015).
981 11 Verma, S. *et al.* Two large-effect QTLs, *Ma* and *Ma3*, determine genetic potential for acidity in
982 apple fruit: breeding insights from a multi-family study. *Tree Genetics & Genomes* **15**, 18,
983 doi:10.1007/s11295-019-1324-y (2019).
984 12 Baumgartner, I. O. *et al.* Development of SNP-based assays for disease resistance and fruit
985 quality traits in apple (*Malus* × *domestica* Borkh.) and validation in breeding pilot studies. *Tree*
986 *Genetics & Genomes* **12**, 35, doi:10.1007/s11295-016-0994-y (2016).
987 13 Iezzoni, A. F. *et al.* RosBREED: Bridging the chasm between discovery and application to enable
988 DNA-informed breeding in rosaceous crops. *Horticulture Research* **7**, 177, doi:10.1038/s41438-
989 020-00398-7 (2020).

14 Chagné, D. *et al.* Validation of SNP markers for fruit quality and disease resistance loci in apple (*Malus* × *domestica* Borkh.) using the OpenArray® platform. *Horticulture Research* **6**, 30, doi:10.1038/s41438-018-0114-2 (2019).

15 Velasco, R. *et al.* The genome of the domesticated apple (*Malus* × *domestica* Borkh.). *Nature Genetics* **42**, 833-839, doi:10.1038/ng.654 (2010).

16 Daccord, N. *et al.* High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nature Genetics* **49**, 1099-1106, doi:10.1038/ng.3886 (2017).

17 Zhang, L. *et al.* A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. *Nature Communications* **10**, 1494, doi:10.1038/s41467-019-09518-x (2019).

18 Sun, X. *et al.* Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nature Genetics* **52**, 1423-1432, doi:10.1038/s41588-020-00723-9 (2020).

19 Broggini, G. A. L. *et al.* Chromosome-scale de novo diploid assembly of the apple cultivar 'Gala Galaxy'. *bioRxiv*, 2020.2004.2025.058891, doi:10.1101/2020.04.25.058891 (2020).

20 Bianco, L. *et al.* Development and validation of a 20K single nucleotide polymorphism (SNP) whole genome genotyping array for apple (*Malus* × *domestica* Borkh). *PLOS ONE* **9**, e110377, doi:10.1371/journal.pone.0110377 (2014).

21 Bianco, L. *et al.* Development and validation of the Axiom®Apple480K SNP genotyping array. *The Plant Journal* **86**, 62-74, doi:10.1111/tpj.13145 (2016).

22 Migicovsky, Z. *et al.* Genome to phenome mapping in apple using historical data. *The Plant Genome* **9**, doi:10.3835/plantgenome2015.11.0113 (2016).

23 McClure, K. A. *et al.* A genome-wide association study of apple quality and scab resistance. *The Plant Genome* **11**, 170075, doi:10.3835/plantgenome2017.08.0075 (2018).

1015 24 Hirschhorn, J. N. & Daly, M. J. Genome-wide association studies for common diseases and
1016 complex traits. *Nature Reviews Genetics* **6**, 95-108, doi:10.1038/nrg1521 (2005).
1017 25 Kumar, S. *et al.* Novel genomic approaches unravel genetic architecture of complex traits in
1018 apple. *BMC Genomics* **14**, 393, doi:10.1186/1471-2164-14-393 (2013).
1019 26 Urrestarazu, J. *et al.* Genome-wide association mapping of flowering and ripening periods in
1020 apple. *Frontiers in Plant Science* **8**, 1923, doi:10.3389/fpls.2017.01923 (2017).
1021 27 Larsen, B. *et al.* Genome-wide association studies in apple reveal loci for aroma volatiles, sugar
1022 composition, and harvest date. *The Plant Genome* **12**, 180104,
1023 doi:10.3835/plantgenome2018.12.0104 (2019).
1024 28 Hu, Y. *et al.* ERF4 affects fruit firmness through TPL4 by reducing ethylene production. *The*
1025 *Plant Journal* **103**, 937-950, doi:10.1111/tpj.14884 (2020).
1026 29 Minamikawa, M. F. *et al.* Tracing founder haplotypes of Japanese apple varieties: application
1027 in genomic prediction and genome-wide association study. *Horticulture Research* **8**, 49,
1028 doi:10.1038/s41438-021-00485-3 (2021).
1029 30 Meuwissen, T. H. E., Hayes, B. J. & Goddard, M. E. Prediction of total genetic value using
1030 genome-wide dense marker maps. *Genetics* **157**, 1819 (2001).
1031 31 Meuwissen, T. Genomic selection: marker assisted selection on a genome wide scale. *Journal*
1032 *of Animal Breeding and Genetics* **124**, 321-322, doi:10.1111/j.1439-0388.2007.00708.x (2007).
1033 32 Kumar, S. *et al.* Genomic selection for fruit quality traits in apple (*Malus* × *domestica* Borkh.).
1034 *PLOS ONE* **7**, e36674, doi:10.1371/journal.pone.0036674 (2012).
1035 33 Kumar, S. *et al.* Genome-enabled estimates of additive and nonadditive genetic variances and
1036 prediction of apple phenotypes across environments. *G3 Genes|Genomes|Genetics* **5**, 2711-
1037 2718, doi:10.1534/g3.115.021105 (2015).
1038 34 Muranty, H. *et al.* Accuracy and responses of genomic selection on key traits in apple breeding.
1039 *Horticulture Research* **2**, 15060, doi:10.1038/hortres.2015.60 (2015).

1040 35 Roth, M. *et al.* Genomic prediction of fruit texture and training population optimization
1041 towards the application of genomic selection in apple. *Horticulture Research* **7**, 148,
1042 doi:10.1038/s41438-020-00370-5 (2020).
1043 36 Jung, M. *et al.* The apple REFPOP—a reference population for genomics-assisted breeding in
1044 apple. *Horticulture Research* **7**, 189, doi:10.1038/s41438-020-00408-8 (2020).
1045 37 García-Ruiz, A. *et al.* Changes in genetic selection differentials and generation intervals in US
1046 Holstein dairy cattle as a result of genomic selection. *Proceedings of the National Academy of*
1047 *Sciences* **113**, E3995-E4004, doi:10.1073/pnas.1519061113 (2016).
1048 38 Duan, N. *et al.* Genome re-sequencing reveals the history of apple and supports a two-stage
1049 model for fruit enlargement. *Nature Communications* **8**, 249, doi:10.1038/s41467-017-00336-
1050 7 (2017).
1051 39 Howard, R., Carriquiry, A. L. & Beavis, W. D. Parametric and nonparametric statistical methods
1052 for genomic selection of traits with additive and epistatic genetic architectures. *G3:*
1053 *Genes|Genomes|Genetics* **4**, 1027-1046, doi:10.1534/g3.114.010298 (2014).
1054 40 Cooper, M. & DeLacy, I. H. Relationships among analytical methods used to study genotypic
1055 variation and genotype-by-environment interaction in plant breeding multi-environment
1056 experiments. *Theoretical and Applied Genetics* **88**, 561-572, doi:10.1007/BF01240919 (1994).
1057 41 Snape, J. W. *et al.* Dissecting gene × environmental effects on wheat yields via QTL and
1058 physiological analysis. *Euphytica* **154**, 401-408, doi:10.1007/s10681-006-9208-2 (2007).
1059 42 Lopez-Cruz, M. *et al.* Increased prediction accuracy in wheat breeding trials using a marker ×
1060 environment interaction genomic selection model. *G3: Genes|Genomes|Genetics* **5**, 569-582,
1061 doi:10.1534/g3.114.016097 (2015).
1062 43 Jarquín, D. *et al.* A reaction norm model for genomic selection using high-dimensional genomic
1063 and environmental data. *Theoretical and Applied Genetics* **127**, 595-607, doi:10.1007/s00122-
1064 013-2243-1 (2014).

44 Tsai, H.-Y. *et al.* Use of multiple traits genomic prediction, genotype by environment interactions and spatial effect to improve prediction accuracy in yield data. *PLOS ONE* **15**, e0232665, doi:10.1371/journal.pone.0232665 (2020).

45 Lado, B. *et al.* Resource allocation optimization with multi-trait genomic prediction for bread wheat (Triticum aestivum L.) baking quality. *Theoretical and Applied Genetics* **131**, 2719-2731, doi:10.1007/s00122-018-3186-3 (2018).

46 Gianola, D. & Fernando, R. L. A multiple-trait Bayesian LASSO for genome-enabled analysis and prediction of complex traits. *Genetics* **214**, 305-331, doi:10.1534/genetics.119.302934 (2020).

47 Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nature Genetics* **51**, 1339-1348, doi:10.1038/s41588-019-0481-0 (2019).

48 Johnston, J. W., Gunaseelan, K., Pidakala, P., Wang, M. & Schaffer, R. J. Co-ordination of early and late ripening events in apples is regulated through differential sensitivities to ethylene. *Journal of Experimental Botany* **60**, 2689-2699, doi:10.1093/jxb/erp122 (2009).

49 Chagné, D. *et al.* Genetic and environmental control of fruit maturation, dry matter and firmness in apple (Malus × domestica Borkh.). *Horticulture Research* **1**, 14046, doi:10.1038/hortres.2014.46 (2014).

50 Costa, F. *et al.* Role of the genes Md-ACO1 and Md-ACS1 in ethylene production and shelf life of apple (Malus domestica Borkh). *Euphytica* **141**, 181-190, doi:10.1007/s10681-005-6805-4 (2005).

51 Costa, F. *et al.* QTL dynamics for fruit firmness and softening around an ethylene-dependent polygalacturonase gene in apple (Malus x domestica Borkh.). *Journal of experimental botany* **61**, 3029-3039, doi:10.1093/jxb/erq130 (2010).

52 Longhi, S., Moretto, M., Viola, R., Velasco, R. & Costa, F. Comprehensive QTL mapping survey dissects the complex fruit texture physiology in apple (*Malus* x *domestica* Borkh.). *Journal of Experimental Botany* **63**, 1107-1121, doi:10.1093/jxb/err326 (2012).

1090 53 Longhi, S. *et al.* A candidate gene based approach validates Md-PG1 as the main responsible
1091 for a QTL impacting fruit texture in apple (*Malus* x *domestica* Borkh.). *BMC Plant Biology* **13**,
1092 37, doi:10.1186/1471-2229-13-37 (2013).
1093 54 Legay, S. *et al.* Apple russeting as seen through the RNA-seq lens: Strong alterations in the
1094 exocarp cell wall. *Plant Molecular Biology* **88**, 21-40, doi:10.1007/s11103-015-0303-4 (2015).
1095 55 Johnston, J. W., Hewett, E. W. & Hertog, M. L. A. T. M. Postharvest softening of apple (Malus
1096 domestica) fruit: A review. *New Zealand Journal of Crop and Horticultural Science* **30**, 145-160,
1097 doi:10.1080/01140671.2002.9514210 (2002).
1098 56 Rymenants, M. *et al.* Detection of QTL for apple fruit acidity and sweetness using sensorial
1099 evaluation in multiple pedigreed full-sib families. *Tree Genetics & Genomes* **16**, 71,
1100 doi:10.1007/s11295-020-01466-8 (2020).
1101 57 van de Weg, E. *et al.* Epistatic fire blight resistance QTL alleles in the apple cultivar 'Enterprise'
1102 and selection X-6398 discovered and characterized through pedigree-informed analysis.
1103 *Molecular Breeding* **38**, 5, doi:10.1007/s11032-017-0755-0 (2017).
1104 58 Evans, K. M. *et al.* Genotyping of pedigreed apple breeding material with a genome-covering
1105 set of SSRs: Trueness-to-type of cultivars and their parentages. *Molecular Breeding* **28**, 535-
1106 547, doi:10.1007/s11032-010-9502-5 (2011).
1107 59 Heslot, N., Yang, H.-P., Sorrells, M. E. & Jannink, J.-L. Genomic selection in plant breeding: A
1108 comparison of models. *Crop Science* **52**, 146-160, doi:10.2135/cropsci2011.06.0297 (2012).
1109 60 Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T. & Zeileis, A. Conditional variable importance
1110 for random forests. *BMC Bioinformatics* **9**, 307, doi:10.1186/1471-2105-9-307 (2008).
1111 61 Hardner, C. M. *et al.* Prediction of genetic value for sweet cherry fruit maturity among
1112 environments using a 6K SNP array. *Horticulture Research* **6**, 6, doi:10.1038/s41438-018-0081-
1113 7 (2019).
1114 62 Crossa, J. *et al.* Genomic selection in plant breeding: Methods, models, and perspectives.
1115 *Trends in Plant Science* **22**, 961-975, doi:10.1016/j.tplants.2017.08.011 (2017).

1116 63 Cuevas, J. *et al.* Bayesian genomic prediction with genotype × environment interaction kernel
1117 models. *G3 Genes|Genomes|Genetics* **7**, 41-53, doi:10.1534/g3.116.035584 (2017).
1118 64 Cirilli, M. *et al.* The multisite PeachRefPop collection: A true cultural heritage and international
1119 scientific tool for fruit trees *Plant Physiology* **184**, 632-646, doi:10.1104/pp.19.01412 (2020).
1120 65 Laurens, F. *et al.* An integrated approach for increasing breeding efficiency in apple and peach
1121 in Europe. *Horticulture Research* **5**, 11, doi:10.1038/s41438-018-0016-3 (2018).
1122 66 Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data
1123 inference for whole-genome association studies by use of localized haplotype clustering. *The*
1124 *American Journal of Human Genetics* **81**, 1084-1097, doi:10.1086/521987 (2007).
1125 67 Rodríguez-Álvarez, M. X., Boer, M. P., van Eeuwijk, F. A. & Eilers, P. H. C. Correcting for spatial
1126 heterogeneity in plant breeding experiments with P-splines. *Spatial Statistics* **23**, 52-71,
1127 doi:10.1016/j.spasta.2017.10.003 (2018).
1128 68 Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4.
1129 *Journal of Statistical Software* **67** (2015).
1130 69 Gabriel, K. R. The biplot graphic display of matrices with application to principal component
1131 analysis. *Biometrika* **58**, 453-467, doi:10.1093/biomet/58.3.453 (1971).
1132 70 Huang, M., Liu, X., Zhou, Y., Summers, R. M. & Zhang, Z. BLINK: A package for the next level of
1133 genome-wide association studies with both individuals and markers in the millions.
1134 *GigaScience* **8**, doi:10.1093/gigascience/giy154 (2018).
1135 71 Tang, Y. *et al.* GAPIT version 2: An enhanced integrated tool for genomic association and
1136 prediction. *The Plant Genome* **9**, doi:10.3835/plantgenome2015.11.0120 (2016).
1137 72 Ward, J. H. Hierarchical grouping to optimize an objective function. *Journal of the American*
1138 *Statistical Association* **58**, 236-244, doi:10.1080/01621459.1963.10500845 (1963).
1139 73 Breiman, L. Random forests. *Machine Learning* **45**, 5-32, doi:10.1023/A:1010933404324
1140 (2001).

74 Habier, D., Fernando, R. L., Kizilkaya, K. & Garrick, D. J. Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* **12**, 186, doi:10.1186/1471-2105-12-186 (2011).

75 de los Campos, G., Gianola, D., Rosa, G. J. M., Weigel, K. A. & Crossa, J. Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. *Genetics Research* **92**, 295-308, doi:10.1017/S0016672310000285 (2010).

76 VanRaden, P. M. Efficient methods to compute genomic predictions. *Journal of Dairy Science* **91**, 4414-4423, doi:10.3168/jds.2007-0980 (2008).

77 Wright, M. N. & Ziegler, A. ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software* **77**, 1-17, doi:10.18637/jss.v077.i01 (2017).

78 Pérez, P. & de los Campos, G. Genome-wide regression and prediction with the BGLR statistical package. *Genetics* **198**, 483-495, doi:10.1534/genetics.114.164442 (2014).

79 de los Campos, G., Sorensen, D. & Gianola, D. Genomic heritability: What is it? *PLOS Genetics* **11**, e1005048, doi:10.1371/journal.pgen.1005048 (2015).

80 Lehermeier, C., de Los Campos, G., Wimmer, V. & Schön, C. C. Genomic variance estimates: With or without disequilibrium covariances? *J Anim Breed Genet* **134**, 232-241, doi:10.1111/jbg.12268 (2017).