Journal of Advanced Research 42 (2022) 149–162

Contents lists available at ScienceDirect

Journal of Advanced Research

journal homepage: www.elsevier.com/locate/jare

Original Article

A bulked segregant analysis tool for out-crossing species (BSATOS) and QTL-based genomics-assisted prediction of complex traits in apple

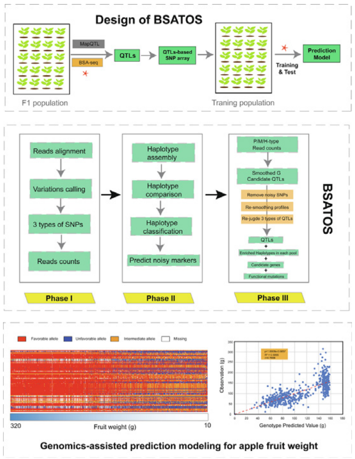
Fei Shen[a,b,c,1], Luca Bianco[b,1], Bei Wu[a], Zhendong Tian[a], Yi Wang[a], Ting Wu[a], Xuefeng Xu[a], Zhenhai Han[a,*], Riccardo Velasco[d], Paolo Fontana[b,*], Xinzhong Zhang[a,*]

[a] College of Horticulture, China Agricultural University, Beijing 100193, China
[b] Research and Innovation Center, Edmund Mach Foundation, 38010 S. Michele all'Adige, Italy
[c] Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China
[d] Research Centre for Viticulture and Enology, CREA, Conegliano, Italy

HIGHLIGHTS

- The BSATOS takes advantage of using haplotypes and markers with different segregation patterns to identify QTLs.
- A novel integrated strategy was developed to conduct genomics-assisted prediction (GAP) in out-crossing species.
- GAP models were successfully developed for apple fruit weight, ripening date, and soluble solid content.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 13 January 2022
Revised 6 March 2022
Accepted 22 March 2022
Available online 26 March 2022

Keywords:
Genomics-assisted prediction
Bulked segregant analysis
QTL
Apple

ABSTRACT

Introduction: Genomic heterozygosity, self-incompatibility, and rich-in somatic mutations hinder the molecular breeding efficiency of outcrossing plants.
Objectives: We attempted to develop an efficient integrated strategy to identify quantitative trait loci (QTLs) and trait-associated genes, to develop gene markers, and to construct genomics-assisted prediction (GAP) modes.
Methods: A novel protocol, bulked segregant analysis tool for out-crossing species (BSATOS), is presented here, which is characterized by taking full advantage of all segregation patterns (including AB × AB markers) and haplotype information. To verify the effectiveness of the protocol in dealing with the complex traits of outbreeding species, three apple cross populations with 9,654 individuals were adopted.
Results: By using BSATOS, 90, 60, and 77 significant QTLs were identified successfully and candidate genes were predicted for apple fruit weight (FW), fruit ripening date (FRD), and fruit soluble solid content

Peer review under responsibility of Cairo University.

* Corresponding authors.
E-mail addresses: shenf1028@gmail.com (F. Shen), luca.bianco@fmach.it (L. Bianco), 124055573@qq.com (B. Wu), 614966808@qq.com (Z. Tian), wangyi@cau.edu.cn (Y. Wang), wuting@cau.edu.cn (T. Wu), xuefengx@cau.edu.cn (X. Xu), rschan@cau.edu.cn (Zhenhai Han), riccardo.velasco@crea.gov.it (R. Velasco), paolo.fontana@fmach.it (P. Fontana), zhangxinzhong999@126.com (X. Zhang).

[1] Equal contribution.

https://doi.org/10.1016/j.jare.2022.03.013
2090-1232/© 2022 The Authors. Published by Elsevier B.V. on behalf of Cairo University.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

(SSC), respectively. The gene-based markers were developed and genotyped for 1,396 individuals in a training population, including 145 Malus accessions and 1,251 F1 plants of the three full-sib families. GAP models were trained using marker genotype effect estimates of the training population. The prediction accuracy was 0.7658, 0.6455, and 0.3758 for FW, FRD, and SSC, respectively.

Conclusion: The BSATOS and GAP models provided a convenient and efficient methodology for candidate gene mining and molecular breeding in out-crossing plant species. The BSATOS pipeline can be freely downloaded from: https://github.com/maypoleflyn/BSATOS.

© 2022 The Authors. Published by Elsevier B.V. on behalf of Cairo University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Introduction

Unlike inbred crops, many outcrossing plant species exhibit three prominent reproductive properties hindering their breeding efficiency and making it more difficult to explore their genetic/genomic characteristics. The first is the heterozygous genetic background, the second is the self-incompatibility barrier and the third is the large number of somatic mutations preserved and gradually accumulated generation after generation via vegetative propagation. Therefore, the breeding schemes of inbreeding and outbreeding plants apply to the general principles but differ in methodology from each other.

Bulk segregant analysis (BSA) is a cost-efficient method for quantitative trait loci (QTL) mapping and has been greatly improved in recent years by next-generation sequencing (NGS) technology [1]. For outbreeding species like apple, however, both parents of a hybrid population are heterozygous, thus the F1 hybrids segregate and sometimes their progeny exhibits ectopic segregating patterns [2,3]. These segregating patterns lead to two problems in QTL identification: (I) how to smooth the statistics of allele frequency difference (AFD) between the two extreme bulks, and (II) how to individuate the parental origins of the alleles of the QTLs. The double pseudo-testcross (DPTC) hypothesis for genetic linkage map construction enlightened the front road of BSA-seq in outbreeding species [4]. To date, unfortunately, no substantial progress was reported to link BSA-seq with DPTC and to fully utilize the huge number of markers which are heterozygous in both maternal and pollen parents.

Marker assisted selection (MAS) uses a limited number of markers to select for interesting traits. MAS is hence most effective when major QTLs explain a high proportion of genetic variance, but is less efficient in case of many genes of small effects, low heritability of the trait, and/or high QTL × environmental interaction [5,6]. On the other hand, genomic selection (GS) uses a high number of markers spread across the entire genome to define the breeding value of an individual to be tested [6,7]. Although the genotyping cost has massively dropped in this decade, screening thousands of individuals still may constitute an economical impediment to be broadly applicable [5,6].

Some evidences suggest that too many redundant markers are used in GS and that a smaller number of more significant markers led to similar or slightly higher prediction accuracy of the breeding values: for example, this was observed in several cereal, fruit, and forestry species [8–11]. In many cases, the addition of already known QTL markers may further increase the prediction accuracy using QTL-based genomics assisted prediction (GAP) approach [12–15]. Moreover, in GS models, marker effects are often arbitrarily assigned as additive so that the non-additive effects are ignored [16,17]. In some outbreeding species, GS has been performed by using a set of QTL-derived markers and obtained high accuracy for some traits, especially those with high heritability [14,18,19]. As reported in dairy cattle, a GS for bovine respiratory disease trained in one state cannot accurately predict disease risk in the other state, this issue was solved finally by using a prediction model using QTL-based markers [20]. By using four significant QTL-derived markers, the apple harvest date was predicted with high accuracy as 0.7375 [21]. The implementation of QTL-based GAP integrates pure GS and MAS. The genetic values can be accurately estimated when the genotype effects of QTL-based markers are properly estimated [22,23].

In this study, we developed an integrated strategy to deal with complex traits of outbreeding plants and test-cased it on three apple cross populations with 9,654 individuals. To make the BSA-seq strategy more effective, we developed a BSA-seq data processing software package, 'BSA tools for outbreeding species' (BSATOS). BSA-seq was used to identify genome-wide QTLs for apple fruit weight, soluble solid content, and ripening date using BSATOS. Finally, to integrate QTL-based markers with GS, QTL-based GAP models for these traits were developed and cross-validated. These protocols and the software were well-applicable and the GAP models can efficiently assist breeding programs in apple and other outbreeding plants.

Results

Development of BSATOS

BSATOS uses reads from the F1 population and their parents (pollen parent was shortened as P and maternal parent as M) in FASTQ files or pre-aligned BAM files from the two extreme bulks and the parents. The other inputs are the reference genome in FASTA format and gene models in GFF/GTF format. BSATOS provides the user with integrated information regarding QTL profiles, candidate genes, candidate functional variations, and enriched haplotype blocks (Fig. 1). The BSATOS pipeline as shown in Fig. 1 and can be freely downloaded from: https://github.com/maypoleflyn/BSATOS.

In phase I of BSATOS, reads from parents (P and M) and the two bulked extremity pools of high (H) and low (L) phenotype values were mapped to the reference genome using the Burrows-Wheeler Aligner [24]. SNPs and InDels (SNVs) were then identified and genotyped using only uniquely mapped reads by SAMtools and larger segment of genomic structure variations (SVs) were also detected with DELLY2 [24,25]. All the variants were finally annotated by ANNOVAR [26]. The high-quality SNVs that are heterozygous in at least one of the two parents were split into three subsets: AA × AB (gP, the genotype of the markers is homozygous in the maternal parent and heterozygous in the pollen parent), AB × BB (gM, the genotype of the marker is heterozygous in the maternal parent and homozygous in the pollen parent) and AB × AB (gMP, the genotype of the markers is heterozygous in both parents). Read counts with different genotypes from H and L pools were then extracted and separated based on which category they fell in (gM, gP or gMP).

In phase II, haplotype blocks were assembled from reads either using the maximum-likelihood-based tool HapCUT2 or the Hidden Markov Model-based algorithm integrated in SAMtools, depending

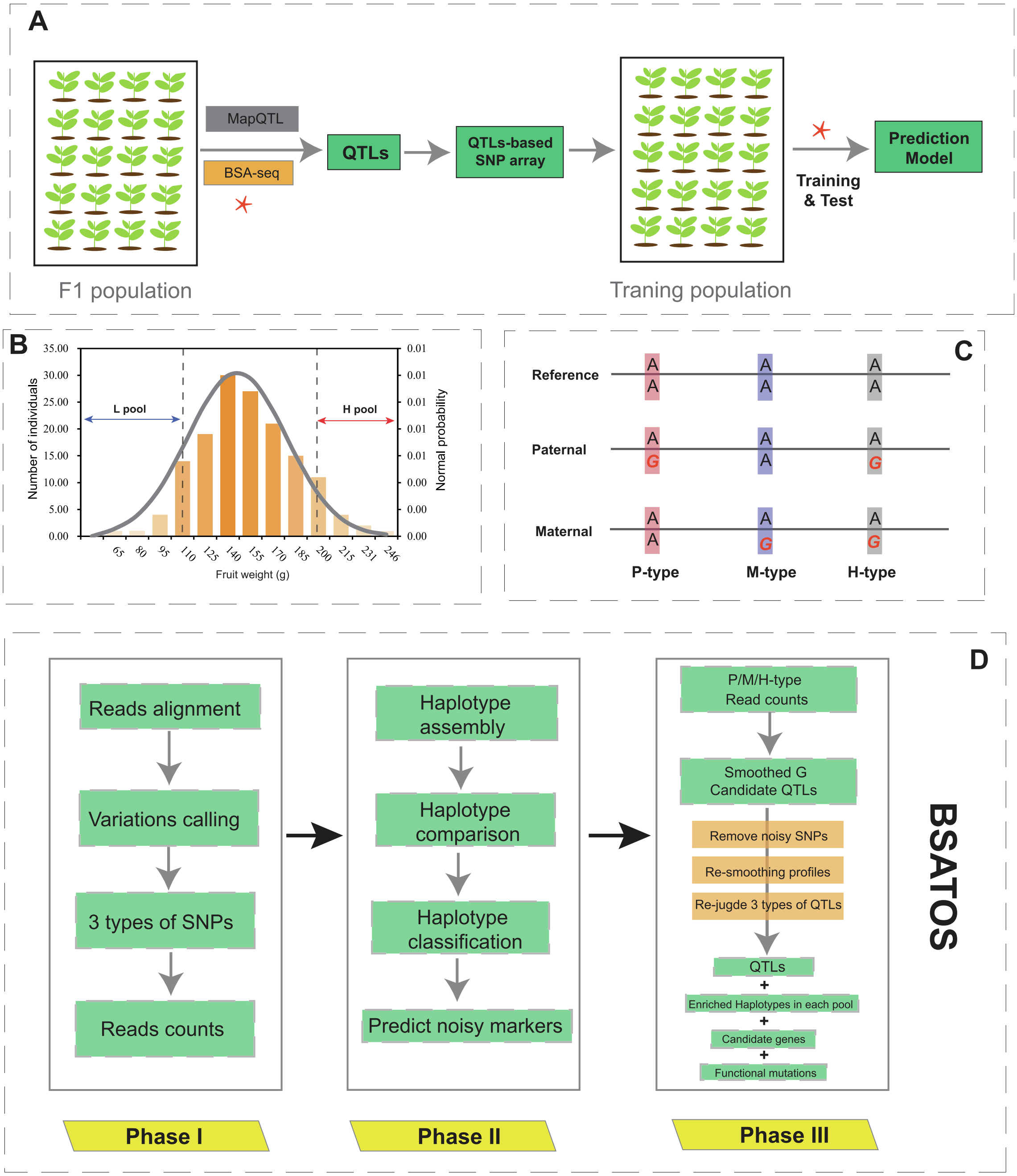


Fig. 1. The schematic diagram of Bulked Segregant Analysis Tool for Out-crossing Species (BSATOS) A. The outlined application of BSATOS and pipeline of constructing genomics-assisted prediction modles. B. The selection of extreme individuals for pooled segregant bulks with the trait apple fruit weight as an example. C. The illustration of the three types of markers considered by BSATOS. D. The detailed implementation of BSATOS pipeline.

on the user's choice (more information provided in the online documentation) [27,28]. The haplotype blocks of each sample were compared with each other to impute missing sites and merge adjacent blocks. Finally, SNP haplotype blocks from the two parents were compared and SNPs located in the haplotypes were classified based on the phase information between the two parents.

In phase III, the three categories of markers (gM, gP and gMP) were statistically analyzed respectively [29]. G values were calculated for each site from the read counts and Nadaraya-Watson kernel regression was used as a smoothing function to compute G′ values within a user-defined sliding window [29]. A log-normal distribution statistics was computed on G′ values for each category of markers as well as the false discovery rate (FDR) [29,30]. Regions featuring a high G′ and FDR lower than a user-defined threshold (default 0.01) were chosen as candidate QTLs. Considering the segregation patterns and allele frequencies observed in H and L pools, the markers with AFD not in agreement with their corresponding haplotype were considered as noise and removed. G′ and FDR were recalculated as above using different sizes for the sliding window and QTL regions were refined using overlapping peaks. The QTL analysis was performed three times using each of the three subsets of data, gM, gP and gMP to map QTLs to the maternal, pollen or both parents, respectively. The physical distance of the genes from the QTL peak was calculated to propose candidate genes responsible for the interesting phenotype. Functional mutations/alleles underlying QTLs were screened based on their genetic segregation pattern (AA × AB, AB × BB or AB × AB) and haplotype information. Finally, integrated information including candidate genes, functional alleles, functional annotation, and allele frequency in each pool were produced as outputs.

Deciphering complex traits using BSATOS

To test the efficiency of BSATOS, we used the data of three quantitative traits complexly controlled by multiple genes in apple: fruit weight, soluble solid content, and fruit ripening date. F1 plants from three hybrid populations were phenotyped over at least three years (2014 ~ 2017). All these traits are segregating continuously, exhibiting an approximately Gaussian distribution in every hybrid population. The broad sense heritability of fruit weight, soluble solid content and fruit ripening date for the three years were averaged as high as 0.89, 0.73, and 0.81, respectively. The segregating patterns implied that all of these traits were controlled by multiple genes.

Based on the phenotypes measured over the years, six bulks were defined for each trait, including 23~45 hybrids with extreme phenotypes for each bulk. A total of 508.3 G bps re-sequencing data of the pooled DNA from the bulks were obtained and processed by BSATOS. An average of 96.53% reads of each pool could be mapped to the GDDH13 apple reference genome and yielded high density (32 ~ 1,287 SNP per million bps) distributed on the entire genome. The density of paternal SNPs (AA × AB) was relatively lower than that of maternal SNPs (AB × BB) or double heterozygous SNPs (AB × AB).

In total, 25, 48, and 17 significant QTLs for fruit weight were identified scattered on 10, 8, and 3 chromosomes in the families of 'Jonathan' (J) × 'Golden Delicious' (G), 'Zisai Pearl' (Z) × 'Red Fuji' (F), and Z × G, respectively (Fig. 2). All chromosomes except chr03, chr04, and chr08 were occupied by at least one QTL. Of these QTLs, one locus coincided in all the three populations (JG-H16.1/ZF-Z16.8/ZG-G16.1) and 11 coincided or overlapped in two populations. The G′ values of two QTLs, JG-J15.6 and ZG-G16.1, were higher than 20 (Fig. 2). Several QTLs overlapped with the previously reported QTLs associated with fruit size/weight, e.g., the major QTLs (JG-J15.1, JG-H15.1, and JG-J15.2) identified in J × G families covered fs15.1 and fs15.2 reported by Liao et al., 2021 (Fig. 2). However, no significant signal at the rare locus fs4.1 has been detected in all the three families, which was consistent with the human selection on the locus during domestication (Fig. 2) [31]. Several vital candidate genes associated closely with regulation of fruit development, gibberellin synthesis, auxin synthesis/-transport were located on the peaks of the QTLs, e.g., *MdKAN2* was located on chr16, which is homolog of *AtKAN1* in *Arabidopsis* and regulates lateral organ polarity and organ morphogenesis [32]; *MdOFPs* on chr15 are homologous to tomato *SlOVATE* and *SUN*, respectively, which are involved in the regulation of fruit size or shape [30,33]; *MdGAIs*, *MdGAOX2*, and *MdG2OX8* on chr01, chr05, chr09, and chr15 are the homolog of essential genes participating in gibberellins biosynthesis [34,35].

For fruit ripening date, 26, 16, and 18 significant QTLs were mapped on the 7, 2, and 4 chromosomes in the progenies of the J × G, Z × F, and Z × G populations, respectively (Fig. 3). All these QTLs are located on 10 chromosomes. Seven of the eight QTLs on chr03 detected in the Z × F population were also identified in the Z × G populations. Five QTLs exhibited G′ values higher than 20 (Fig. 3). We identified several essential genes involved ethylene response, regulation of cell division, and auxin synthesis, notably, there was a cluster of ethylene-responsive factors (*ERF*s) and ethylene receptor (*ETR*s) at the peak of the QTL with the highest G′ value on the chr16 (Fig. 3).

For fruit soluble solid content, 48, 7 and 22 QTLs were identified from J × G, Z × F, and Z × G populations, respectively (Fig. 4). These QTLs were mapped on 13 chromosomes except chr05, chr06, and chr10-chr12. The G′ values of 26 QTLs were higher than 10.0, most of which were mapped on chr00, chr09, chr15, and chr16 (Fig. 4). Several key regulators of fruit development (e.g., *MdMYB*s, *MdbHLH*s, and *MdZFP*s) as well as some essential genes participating in sugar metabolism were identified at the peaks of some QTLs, e.g., genes encoding sorbitol dehydrogenase, *MdSDH*s, were located on the most significant peak regions (Fig. 4).

Gene marker development and estimation of marker genotype effect

Ultimately, a total of 71, 54 and 52 genes containing SNVs/SVs were selected as the candidate genes associated with fruit weight, soluble solid content, and fruit ripening date, respectively. GenoPlexs® primers were designed flanking a SNV/SV within the coding region or 2 kb sequence upstream the ATG codon of the candidate genes.

Using 1,396 individuals, including 1,251 hybrids and 145 *Malus* accessions, the genotype effects of the 71 markers associated with fruit weight were estimated. Of these markers, the largest genotype effect was detected for SIZE9471 (68.9 g) and SIZE906 (43.89 g), whereas the smallest genotype effect (−109.93 g) was detected for SIZE2805, SIZE6268, and SIZE9195. The markers SIZE1413 and SIZE2888 exhibited approximately dominant allelic relationships, SIZE832 and SIZE906 showed additive allelic interaction, while most of the other markers showed over or partial dominance among alleles (Fig. 5A).

The genotype effects of the 54 makers associated soluble solid content were estimated using 1,362 individuals, including 1,217 hybrids and 145 *Malus* accessions. The largest reliable positive genotype effects for soluble solid content were detected for the markers TSS234 (0.56), TSS263 (0.56), and TSS255 (0.49), while the largest confident negative genotype effects were estimated for the markers TSS219 (−2.49), TSS203 (−1.76), and TSS261 (−1.47) (Fig. 6A). The marker TSS249 and TSS255 exhibited approximately additive allelic interaction whereas apparent dominant allelic effect was detected for TSS228, TSS229, TSS237, TSS253, TSS258-4, TSS258-9 and TSS268 (Fig. 6A).

By using 1,033 individuals including hybrids and *Malus* accessions, the marker genotype effects for fruit ripening date were estimated. The confident genotype effects varied from −25.07 days after full bloom (DAFB) (LY284 GG) to 17.29 DAFB (new278 CT) among genotypes of the 52 markers (Fig. 7A). Most of the markers from H-type QTLs exhibited partial dominant allelic interaction among genotypes within the marker, e.g. S349, S477, and LL288 (Fig. 7A). Complete dominant allelic interaction was exerted on fruit ripening date by some of the markers such as CYYL1399, MY154, and XL13, while only markers neww441 and LY064 showed additive allelic interaction (Fig. 7A).

Development of GAP models and cross-validation

We implemented the allele-aware strategy by assigning marker genotype effect to all the three genotypes (see methods) into BSATOS, which constructed the GAP models for the three essential apple traits and assessed the prediction accuracy using different sizes of training set. Using all the 71 markers for apple fruit weight, the GAP models produced by BSATOS gave a higher prediction accuracy (an average of 0.738) than the ridge-regression best linear unbiased prediction (rr-BULP) package (an average of 0.451) (Fig. 5B and 5C). Besides, the GAP model by produced by BSATOS showed better stability than rr-BULP when using a different percentage of training set, especially for high percentages (Fig. 5B and 5C). Five-fold cross-validation confirmed that the average prediction accuracy of GAP for fruit weight was 0.737 (Fig. 5C). This prediction accuracy was higher than that of rr-BLUP (r = 0.452)

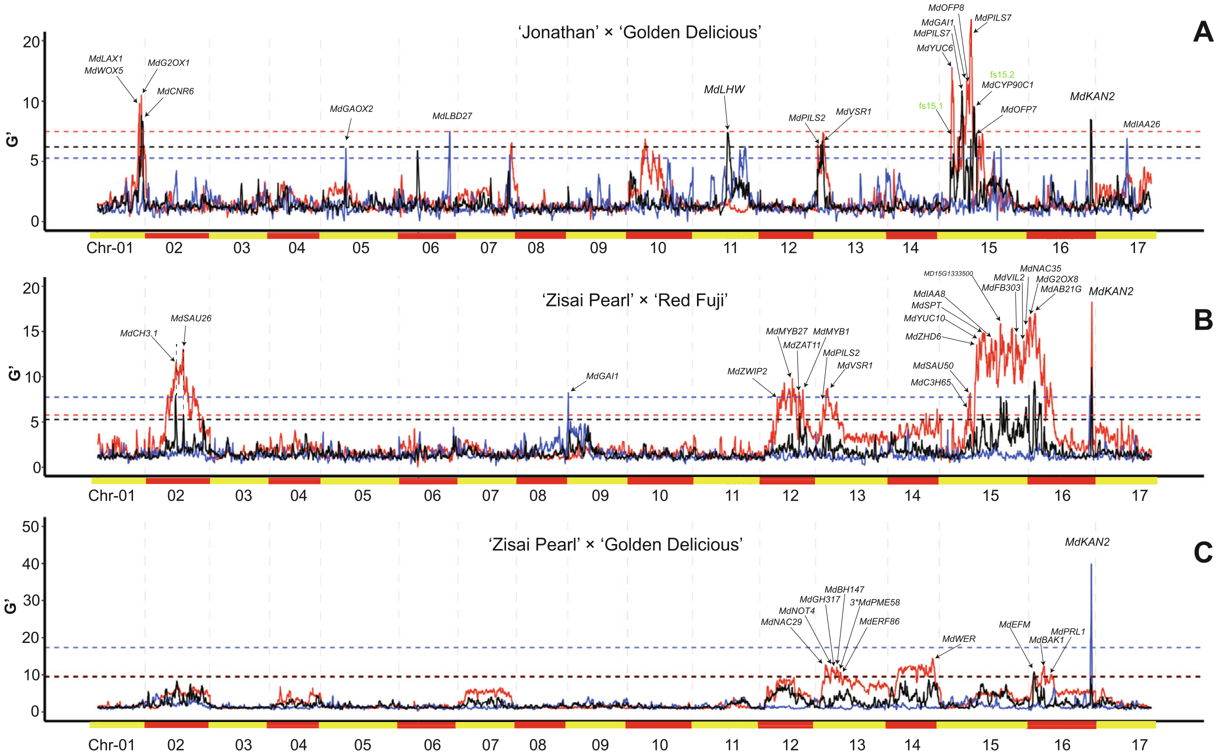


Fig. 2. Genome-wide quantitative trait loci (QTL) identification for apple fruit weight using three biparental cross populations, *Malus domestica* Borkh. 'Jonathan' × 'Golden Delicious'; *M. asiatica* Nakai 'Zisai Pearl' × *M. domestica* Borkh. 'Red Fuji', and 'Zisai Pearl' × 'Golden Delicious'. *Candidate genes at the QTL peaks were marked. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

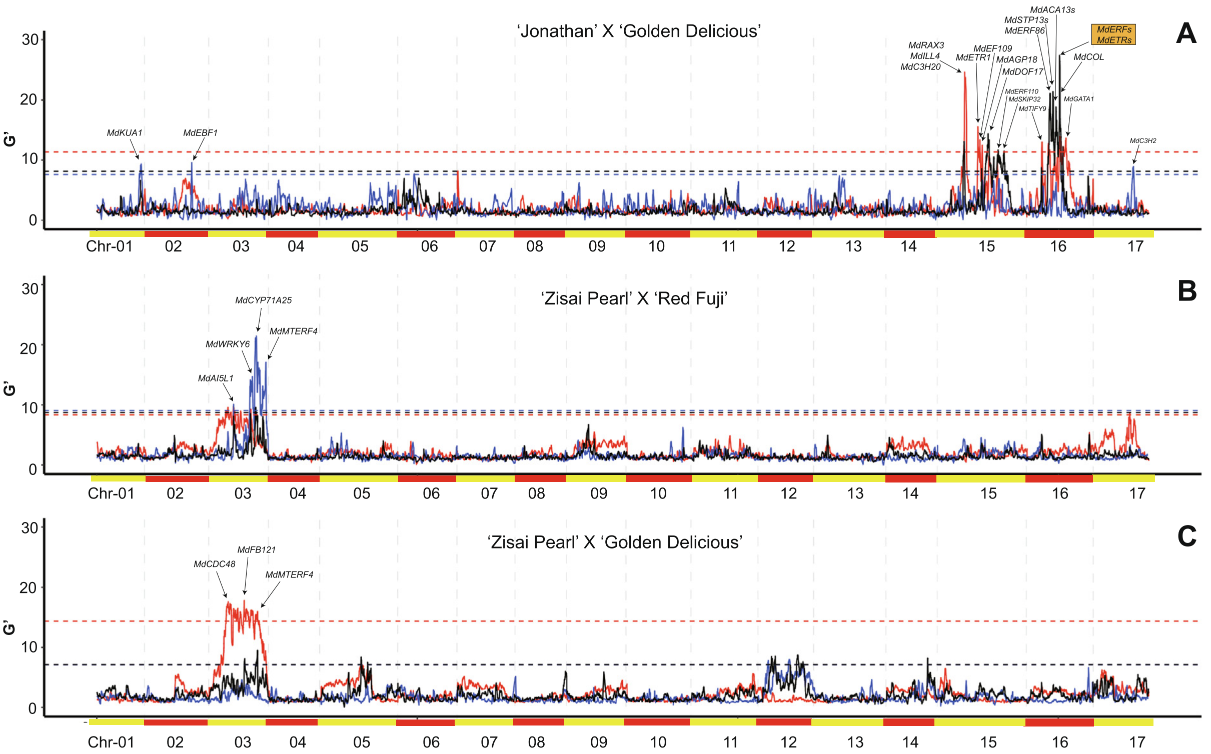


Fig. 3. Genome-wide quantitative trait loci (QTL) identification for apple fruit ripening date using three biparental cross populations, *Malus domestica* Borkh. 'Jonathan' × 'Golden Delicious'; *M. asiatica* Nakai 'Zisai Pearl' × *M. domestica* Borkh. 'Red Fuji', and 'Zisai Pearl' × 'Golden Delicious'. *Candidate genes at the QTL peaks were marked. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

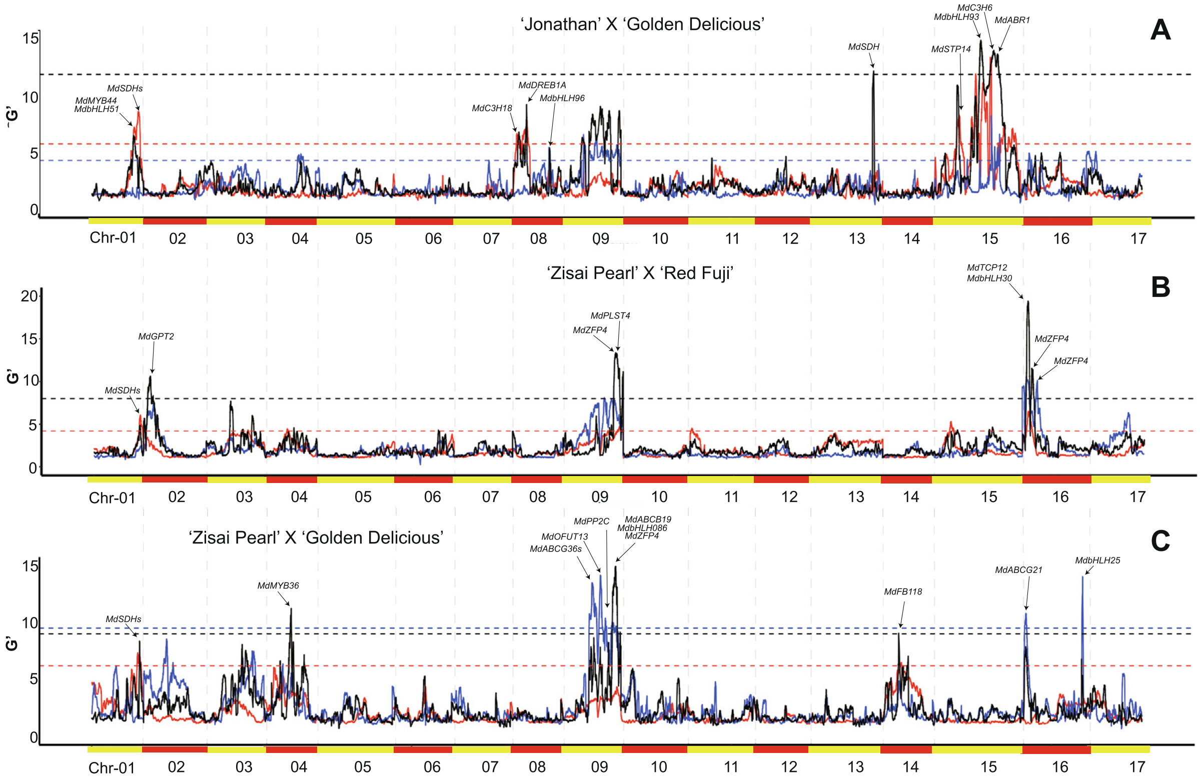


Fig. 4. Genome-wide quantitative trait loci (QTL) identification for apple fruit soluble solid content using three biparental cross populations, *Malus domestica* Borkh. 'Jonathan' × 'Golden Delicious', *M. asiatica* Nakai 'Zisai Pearl' × *M. domestica* Borkh. 'Red Fuji', and 'Zisai Pearl' × 'Golden Delicious'. * Candidate genes at the QTL peaks were marked. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(Fig. 5B and 5C). We defined the allelic genotypes conferring the highest/lowest/intermediate effect and observed that the genotypic profiles changed as the fruit weight gradually increases, indicating the contribution of both the marker effect and the allelic effect to the trait (Fig. 5D, 5E, and 5G). The GAP model for fruit weight was finally developed using all the 71 markers and an accuracy of 0.7658 was obtained when using all the individuals (Fig. 5F). When the genotype predicted value (GPV) criterion was > 140 g, the selection rate was 41.9% and the selection efficiency was 61.1%, which means the observed phenotype values of 146/239 individuals were >140 g. Of the 570 individuals used for simulative selection, there were 161 with the fruit weight phenotype > 140 g, so by filtering the GPV > 140 g, 90.7% (146/161) were selected, which were defined as the exhaustivity (90.7%).

After removing the individuals with missing marker genotype data, the GAP model for soluble solid content was created by BSATOS with 52 markers in a training population containing 1,435 individuals. By using different percentages of the training set, the average prediction accuracy of GAP models obtained by BSATOS was 0.362, which was significantly higher than that by rr-BULP (0.276) (P-value < 0.01) (Fig. 6B and 6C). Moreover, the accuracy of the GAP model produced by BSATOS was more stable than that by rr-BULP (Fig. 6B and 6C). The prediction accuracy of the ultimate GAP model was up to 0.3758 by using all the 52 markers (Fig. 6F). The genetic heatmap showed few significant differences between extreme individuals, indicating the minor marker effect or allelic effect of the genetic loci controlling sugar content (Fig. 6D, 6E, and 6G). By applying a filter of GPV > 15.0% in a simulative selection, the selection rate was 40.9%, the efficiency was 61.2%, whereas the exhaustivity was 55.7%.

The GAP model for fruit ripening date were created by BSATOS with all the 52 markers and the prediction accuracy was 0.6455 using all the individuals (Fig. 7F). Although there was no significant difference in prediction accuracy between the GAP produced by BSATOS and rr-BULP in any size of training sets, GAP by BSATOS exhibited better convergence and stability (P-value > 0.05) (Fig. 7B and 7C). For example, the average prediction accuracy of five-fold cross validation was 0.6354, comparable to the prediction accuracy of rr-BLUP (0.651) (Fig. 7B and 7C). Significant genetic contribution to the fruit ripening date was derived from both the marker effect and the allelic effect (Fig. 7D and 7E). The genotypic heatmap exhibited the distinct genotypes between early maturing and late maturing individuals (Fig. 7G). When the GPV criterion was ≥ 170 DAFB in a simulative selection for fruit ripening date, the selection rate was 23.6%. The selection efficiency and exhaustivity were 69.8% and 39.4%, respectively.

Discussion

BSATOS is a powerful tool for QTL identification in outcrossing plants

Fruit ripening date is a complex quantitative trait, so over 50 QTLs for harvest date were clustered into 16 unique genomic regions on more than five chromosomes [21,36,37]. By using BSATOS, we identified 60 significant QTLs for fruit ripening date on 10 chromosomes in the three hybrid populations, including 21 and nine QTLs on chr03 and chr16, respectively (Fig. 3). Of the 82 SNPs associated with ripening period, 70 and 9 were located on chr03 and chr16, respectively [37], the spanning genomic regions, 29,196,200 – 31,243,065 bp on chr03 and 9,032,064 –

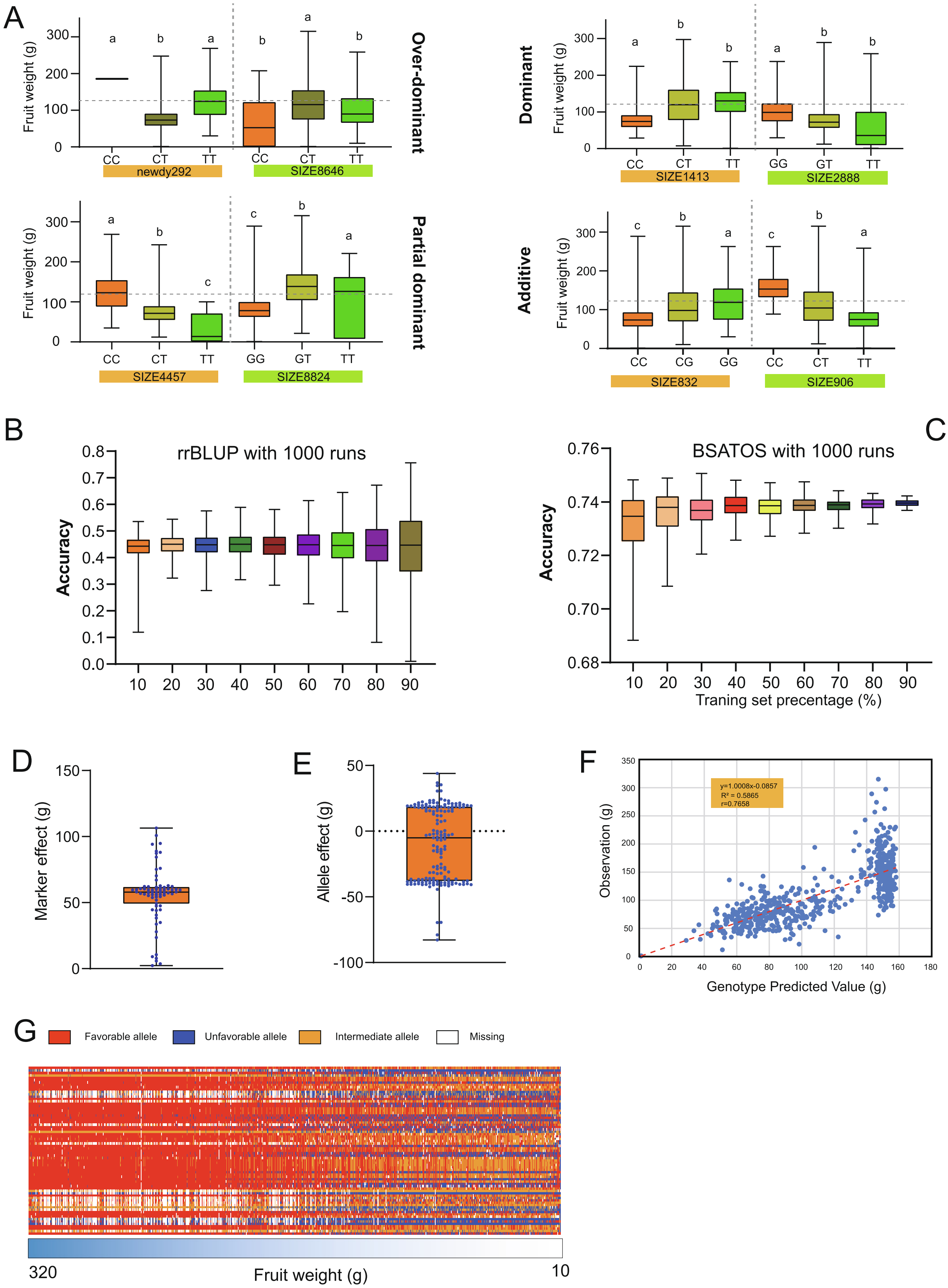


Fig. 5. Marker genotype effects, genomics-assisted prediction (GAP) modeling and cross-validation of the GAP model for apple fruit weight. A. Markers exhibiting typical allelic genetic interactions. B. The prediction accuracy of ridge-regression best linear unbiased prediction (rr-BLUP) using different percentages of training individuals. C. The accuracy of GAP using different percentage of training individuals D. the marker effect of markers for FW E. the allele effect of markers for FW F. Dot-plots showing linear regression between genotype predicted value (GPV) and observed phenotype value (OPV) for apple fruit weight. G. Heatmap showing the genotype across the 1,396 individuals; the allelic genotype conferring the highest/lowest/intermediate fruit weight effect as favorable/unfavorable/intermediate alleles.

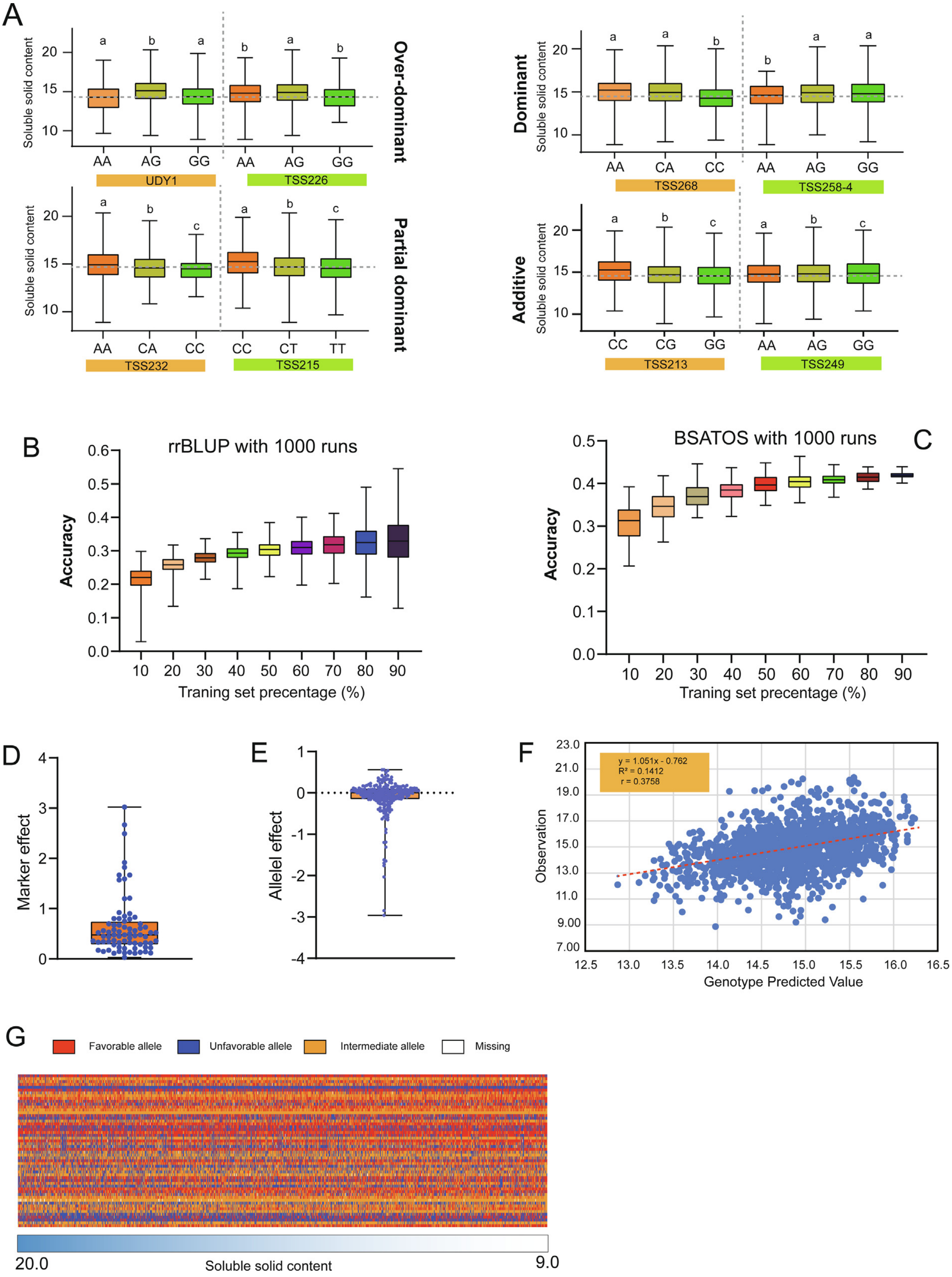


Fig. 6. Genomics-assisted prediction models for apple fruit soluble solid contents (TSSC). A. the demonstration of markers exhibiting typical genetic effects. B. the accuracy of rr-BLUP using different percentage of training individuals C. the accuracy of BSATOS using different percentage of training individuals D. the marker effect of markers for apple fruit soluble solid contents E. the allele effect of markers for apple fruit soluble solid contents F. Dot-plots showing linear regression between genotype predicted value (GPV) and observed phenotype value (OPV) for apple fruit soluble solid contents G. the heatmap of the genotype across the 1,396 individuals; We defined the allelic genotypes conferring the highest/lowest/intermediate TSSC effect as favorable/unfavorable/intermediate alleles.

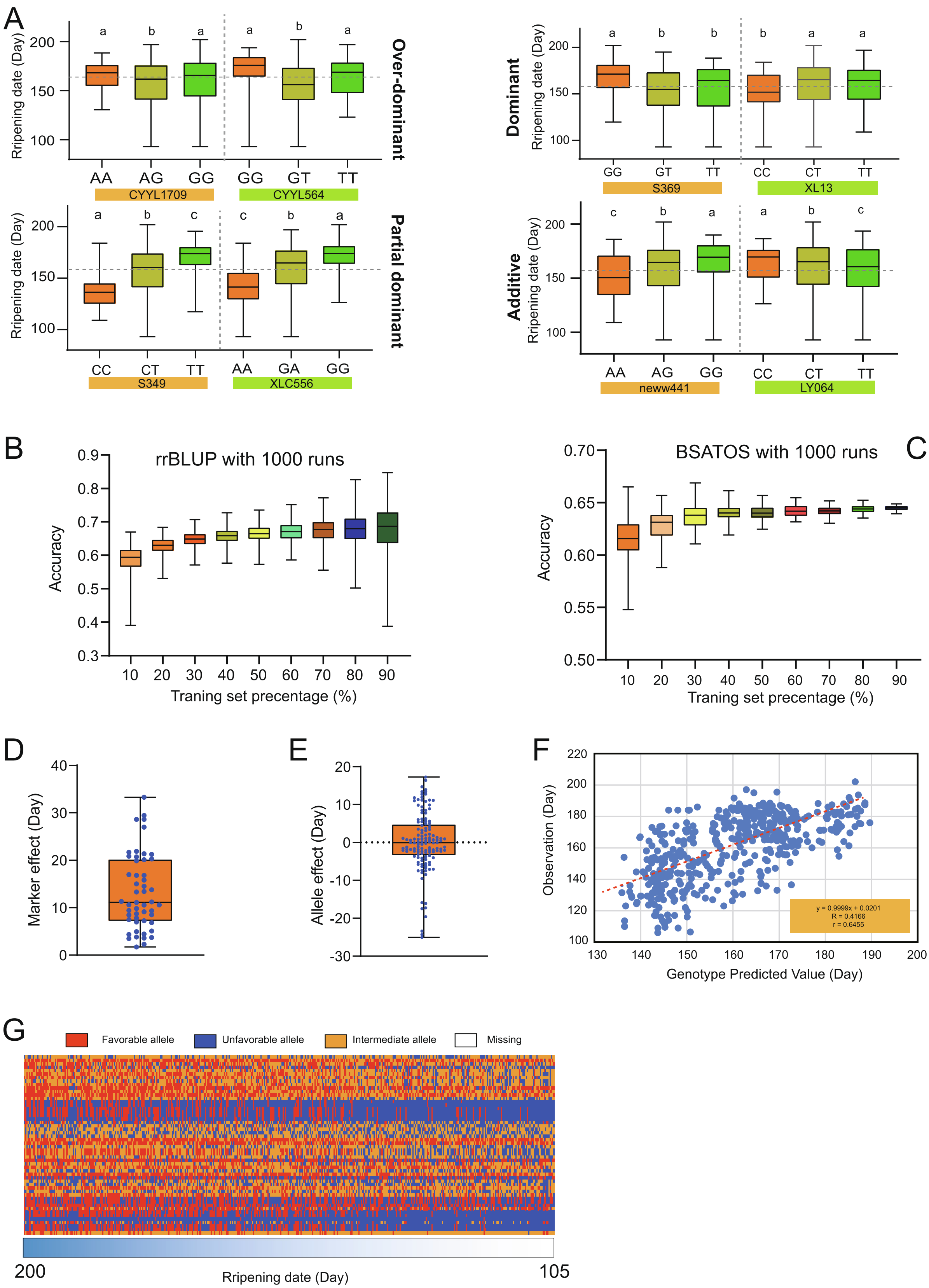


Fig. 7. Genomics-assisted prediction models for apple fruit ripening date. A. the demonstration of markers exhibiting typical genetic effects. B. the accuracy of rr-BLUP using different percentage of training individuals C. the accuracy of BSATOS using different percentage of training individuals D. the marker effect of markers for apple fruit ripening date E. the allele effect of markers for apple fruit ripening date F. Dot-plots showing linear regression between genotype predicted value (GPV) and observed phenotype value (OPV) for apple fruit ripening date G. the heatmap of the genotype across the 1,396 individuals; We defined the allelic genotypes conferring the highest/lowest/intermediate ripening date effect as favorable/unfavorable/intermediate alleles.

9,306,332 bp on chr16 were almost exactly covered by QTLs individuated in this study. Besides those hot spots on chr 03 and chr16, more QTLs with relatively low significance were detected on chr01, chr02, chr05, chr06, chr12, chr14, and chr17. These results were not only highly consistent with the previous reports [37–40], but also indicated that QTLs can be more effectively mapped by using BSATOS.

Another complex quantitative trait is fruit soluble solid contents in apples. QTLs or MetaQTLs for fruit soluble sugar/solids, and/or individual sugar compositions were mapped and some were validated on all the chromosomes except chr10 and chr14 [41–43]. In the present study, we identified 77 QTLs on 13 chromosomes, which were more than any of the previous reports.

Fruit weight is under a much complex genetic control. A set of SNPs, QTLs or MetaQTLs for fruit weight/size were mapped on chr04, chr05, chr06, chr07, chr08, chr11, chr12, chr13, chr15, chr16, and chr17 [31,41,43–45]. Consistently, we mapped 90 confident QTLs for fruit weight on10 chromosomes, and those on chr01, chr02 and chr14 were newly discovered.

The good performance of BSATOS in QTL identification is attributed to the large population size, high marker density, and inclusions of double heterozygous markers (AB × AB) to the marker set, because AB × AB type of markers are often conveying large effects on phenotype variations (Figs. 2-4). BSATOS can also precisely individuate the parental origin of QTL-associated alleles, screen functional variations through haplotype information and assign a statistical significance to the identified QTLs. For outbreeding species like apple, the genetic scenario is quite different from that of conventional inbred crops (e.g., rice and maize). In inbred species, three statistical methods, namely SNP index, Euclidean distance, and G′ statistics, are commonly used in BSA-seq to detect the QTL signals [1,3,30]. To determine the best statistical method for the special scenario in out-crossing species like apple, the efficiency of the three statistical methods for QTL identification was compared, and the results indicated that G′ statistics was the best with distinct unbiasedness and low false positives in QTL detection [3]. Then G′ statistics embedded in BSATOS has been successfully applied to identify QTLs or either the key genes associated with several complex traits, e.g. apple fruit acidity [2], apple cover color degree [46], fruit cold storability [47], apple root growth angle [48], salt-alkaline tolerance [49], and fruit ring rot disease resistance [3]. Moreover, candidate genes and genetic variations on these genes were efficiently predicted and several variations were confirmed to be functional [2,47,48].

In addition, the use of distantly related parents for creating multiple segregating populations is also critical for maximizing the efficiency of QTL identification, which has been confirmed by several previous reports on FlexQTL and MetaQTL [16,17,41,50–54]. GWAS using large scale accessions and pedigree-based QTL mapping using multiple unrelated families were more effective on saturating genome wide QTL-based markers [37,39,52,55]. This study makes use of ‘Jonathan’ and ‘Golden Delicious’, which are the founders of modern apple cultivars, ‘Red Fuji’, which is a direct descendant of other two founders: ‘Red Delicious’ and ‘Ralls Janet’ [43,56,57]. ‘Zisai Pearl’ is a Chinese domesticated cultivar that taxonomically belongs to *M. asiatica* Nakai. These parental cultivars have covered nearly half of the genetic composition of *M. domestica* ancestors. Thus, the identified QTLs are more saturated than those that would have been identified by using a single bi-parental population.

The number of QTLs varied between hybrid populations, because a marker segregating in one population may not segregate in another. For example, in this study, H15.5 (S349) for ripening date was identified in the J × G but was not detectable in hybrids of neither Z × G nor Z × F. The effect of a certain marker may be detected in a population where the marker does not segregate. The genotype effect of some non-segregating markers altered the average phenotype performance of the whole population. e.g., for fruit weight, the genotype of the marker SIZE6268, SIZE2805, DDY6, and XDY160 did not segregate in the J × G population, however, the effects (21.65, 18.76, 18.42, and 19.24, respectively) were present and therefore increased the overall phenotype value of the population.

The accurate phenotyping is also important for QTL mapping and genotype effect estimation. In annual crops, phenotypic data are often collected from multiple sites and multiple seasons [13,67,75]. In perennial woody plants, however, multi-site trials cost much and the authors prefer to collect phenotype data of multi-year on a single site like that in this study [19,39,58]. Kumar and colleagues assessed six apple fruit quality traits at two sites in New Zealand, the between-site genotypic correlations were higher than 0.85 for all traits, and genotype-site interaction accounted for less than 10% of the phenotypic variance, the prediction accuracy was similar when the validation set was used for one site or for both sites [61].

QTL-based GAP is capable to integrate MAS and GS

For complex quantitative traits, QTL-based GAP may obtain comparable or better accuracy than that of GS but with remarkably reduced cost. GS uses genome-wide markers and makes the genotyping cost for thousands of individuals unaffordable. The accuracy of GS was remarkably high and ranged from 0.68 to 0.89 for apple soluble solid content, astringency and titratable acidity in 1,120 seedlings of seven full-sib families generated from six parents [12]. However, by using pedigreed full-sib families, the GS accuracy for fruit size ranged from 0.08 to 0.33, averaging to 0.23 [16]. Similarly, the predictability was ~0.5 for fruit harvest date, ~0.2 for fruit weight and soluble solid contents in a natural population comprised of 172 *Malus* accessions in a two years trial [39]. In an independent study using founder haplotypes, the GS predictability was around 0.6, 0.3, and 0.2 for apple pickday, fruit weight, and Brix, respectively [43]. In other outcrossing perennials like strawberry (*Fragaria ananassa* Duch.) and Japanese pear (*Pyrus pyrifolia* Nakai), different GS models for fruit weight, yield, or soluble solid content showed a relatively low predictive ability of 0.18 ~ 0.70 [19,58,59]. QTL-based GAP has been successfully established for apple cover color degree [46], fruit cold storability [60], apple root growth angle [48], and apple rootstock salt-alkaline tolerance [49]. The GAP prediction accuracy in this study was 0.7658, 0.6455, and 0.3758 for apple fruit weight, fruit ripening date, and soluble solid content, respectively. These accuracies were relatively higher than that of most reported pure GS with high density SNP array [16,39,61].

The prediction accuracy for soluble solid content (0.3758) was however relatively lower than that for fruit weight and fruit ripening date. The relatively lower broad sense heritability for fruit soluble solid content (0.73) might be one of the causes of this, because the prediction accuracy was strongly influenced by trait heritability [16,61]. The low prediction accuracy for apple soluble solids was also reported in GS by using data of germplasm accessions (around 0.35) or using historical phenotypic data (about 0.25) [38,39]. The prediction accuracy of GS/GAP depends also on the genetic structure of the population and the relatedness between training and validation populations [61]. When a collection of diverse accessions or several populations covered most variations in a few founder cultivars, the prediction accuracies would not be as high as that obtained by using a well-designed training population (0.89 for soluble solids) reported by Kumar and colleagues [38,39,61].

The size and composition of the training population are important factors affecting the prediction accuracy. In alfalfa (*Medicago* *sativa* L.), when the training population was obtained by a mixture of commercial cultivars, the genomic prediction accuracies went from 0.34 to 0.51 for one cycle selection of total biomass yield, but by using two single populations, the GS accuracy was in the range 0.32–0.35 [62,63], revealing large impacts of population composition on prediction accuracy [7,64]. In this study three full-sib families derived from four unrelated parents were explored, and 145 *Malus* accessions with broad ancestral background were also included in the training population, which ensured good prediction accuracy and increased the practical applicability of the GAP models.

Inclusion of significant QTL in the marker panel may largely improve the prediction accuracy. In animals, the addition of several QTL-derived markers to the GS marker set led to an increase in prediction reliability from 0.585 to 0.606 and from 0.488 to 0.519 for milk fat contents of cattle one and three generations away, respectively [65], which was consistent with the results obtained not only in animals like pigs but also plants like wheat [66,67]. As a similar result was observed in the prediction accuracy using 1.0 ~ 2.5% selected SNP markers and 13 ~ 18 tagged QTLs. The accuracy obtained was better than that using a whole set of 200 K SNPs including 95 QTLs [42]. Moreover, the accuracy of GS for sheep parasite resistance using carefully selected sequence variants from the QTL regions can be improved by 9% [68]. These data demonstrated that the use of a small number of loci with large effects on a trait may result in the best accuracy [69,70]. Interestingly, high accuracy can also be obtained by using the QTL-derived markers only, using 125–200 SSR markers with the highest heterozygosity would have marginally improved accuracy to 0.56 for rubber production in *Hevea brasiliensis* [71]. In this study, genome-wide QTL-based markers were used to generate GAP models with relatively high predictability for apple fruit ripening date, fruit weight, and soluble solid content, indicating the potential use of QTL data to reduce marker density and therefore costs.

In inbreeding plants, GEBV is estimated in many GS models such as rr-BLUP, Bayesian least absolute shrinkage and selection operator (LASSO) etc., which emphasizes additive effects [11]. In outbreeding species, however, non-additive effects contribute to a large proportion of the genetic variations due to the high level of heterozygosity and the clonal propagation of many outcrossing plants [2,72]. In the GAP models, the genetic effect includes both allelic additive and non-additive effects, the prediction accuracy was therefore higher than that obtained by rr-BLUP.

In the GAP model for fruit weight, we found that the observed phenotype value (OPV) of some triploid *M. domestica* cultivars were 250–350 g, such as 'HAC-9' (276.0 g), 'Jonagold' (250.4 g), 'Shizuka' (258.1gg), 'Mutsu' (314.1 g), 'Crispin' (298.1 g), etc. These OPVs were much larger than their GPVs (176.8 g – 188.8 g). Because the genotypes of these triploids were output with the format of diploids by the genotyping by sequencing (GBS) protocol, one of the three alleles in a triploid accession was routinely omitted, therefore the GPV was under-estimated by neglecting one-third allelic effects. Additionally, the genotype of some somatic mutant cultivars was nearly the same and thus they have the same GPV, but the OPV varied remarkably, for examples, 'Golden B', 'Smoothee', 'Spur Golden Delicious', and 'Golden Spur' were russet-less, or spur-type mutant cultivars derived from 'Golden Delicious', their GPV for fruit weight was 187.5 g, but the OPV varied between 110.0 g and 236.4 g. These data indicated that the present GAP model for fruit weight failed to accurately predict the effect of polyploidy or somatic mutations.

Further dissection of the molecular control and regulation network of the traits is much eagerly desired for precise genomic selection [73], so most markers in this study were designed on the coding regions or on the 2 kb upstream sequence of the candidate genes. These gene markers were beneficial for eliminating the effect of the rapid linkage disequilibrium decay as much as possible, or otherwise extremely high density of markers would have been required [38]. The markers designed on gene regions could in addition be helpful for the development of new diagnostic markers and to improve the prediction accuracy [74,75].

Conclusion

A novel protocol, BSATOS, for BSA-seq and QTL identification for out-crossing plant species was developed in this study. We tested BSATOS by challenging it with the identification of QTLs for apple fruit weight, fruit ripening date and solid content in three cross populations. BSATOS identified 90, 60, and 77 QTLs respectively for apple fruit weight, fruit ripening date, and fruit soluble solid content. Markers were designed on the candidate genes of each QTL region, and the marker genotype effects were estimated. Finally, the accuracy of GAP models was 0.6455, 0.7658, and 0.3758 for fruit ripening date, fruit weight, and soluble solid content, respectively. The results presented in this paper showed that BSATOS can be effectively used and that GAP models may assist highly efficient molecular breeding in out-crossing plants.

Materials and methods

Plant materials

Malus germplasm accessions (145) and hybrids of three bi-parental cross populations, *M. domestica* Borkh. 'Jonathan' × 'Golden Delicious' (J × G) (1,773 hybrids); *M. asiatica* Nakai 'Zisai Pearl' × *M. domestica* Borkh. 'Red Fuji' (Z × F) (3,627 hybrids); and 'Zisai Pearl' × 'Golden Delicious' (Z × G) (3,492 hybrids), were used as segregating populations. The hybrid cross was performed in 2002 (J × G) and 2007 (Z × F and Z × G). All plant materials were subjected to conventional cultivation management and pest control. The phenotype values of fruit weight of the parental cultivars, J, G, Z, and F were 220.7 g, 236.4 g, 44.0 g, and 241.8 g, respectively. The ripening date phenotype of the parents, J, G, Z, and F was 153 DAFB, 151 DAFB, 179 DAFB, and 177 DAFB, respectively. The fruit soluble solid content of J, G, Z, and F was 15.35 Brix, 16.31 Brix, 16.16 Brix, and 14.60 Brix, respectively.

Phenotyping

Apple fruit of all hybrids and accessions were phenotyped in the years period 2014 ~ 2017. Fruit maturity was determined by fruit skin ground color de-greening and starch index [76,77]. The ripening date was recorded as days after full bloom (DAFB) to avoid the phenological variation among years. Fruit weight was measured as the average of ten randomly picked apples. The soluble solid content was the average value of measurements for each of three apples by a Brix meter (PAL-1, Atago, Japan) after calibrating with distilled water. The phenotype segregation and the broad-sense heritability for each trait were then analyzed using the methods described previously [78].

Re-sequencing of parental cultivars, bulk construction, and BSA-seq

To acquire polymorphic variations among the parental cultivars, the four parents were sequenced with Illumina short reads at a 50x genome coverage. Genomic DNA was extracted from the leaves of 'Golden Delicious', 'Jonathan', 'Red Fuji', and 'Zisai Pearl' using a Genomic DNA Isolation Kit (TianGen, Beijing, China). Then, the Illumina sequencing libraries were constructed using NEBNext DNA Library Prep Master Mix (NEB), pair-end (PE-150) sequencing was performed using the Illumina HiSeq2500 sequencer (Illumina, San Diego, CA), and the re-sequencing data were processed using the protocol described previously [3]. Individuals with extreme phenotypes of fruit weight, soluble solid content, and ripening date for each hybrid population were selected to construct nine pairs of DNA pools. Illumina sequencing libraries were constructed using the pooled DNA samples and pair-end (PE-150) sequencing was performed as described above. Finally, the sequencing data was processed using BSATOS.

Candidate gene prediction, GenoPlexs® marker design, and genotyping

Candidate genes were predicted from the QTL intervals following the previously described protocols [3]. Genes that do not contain functional single nucleotide variants (SNVs) and structural variations (SVs) between the two parental cultivars, on which the QTL was mapped, were removed from the list. Based on the Gene Ontology (GO) and detailed functional annotations, genes with trait-inconsistent organ/tissue/sub-cellular localization, developmental dynamics, and physiological pathway annotations were excluded. Then, genes with unexpected AFD values between the two bulk pools were also excluded from the candidate gene list. GenoPlexs® primers were designed based on the 200-bp sequence flanking the SNP, InDel or SV markers on the candidate genes. Marker genotyping was performed following the instruction of the GenoPlexs® (https://www.molbreeding.com/index.php/Technology/GenoPlexs.html).

Estimation of marker genotype effects and non-allelic interactions

To estimate the genotype effects of the markers, the 145 accessions and over 400 hybrids randomly chosen from each population were genotyped for all markers. Marker genotype effect was estimated using the complete data set of whole population by the following equation (1):

$$GE = \sum_{i}^{m} P/m - \sum_{k}^{n} OPV/n \tag{1}$$

GE: the allelic genotype effect of a marker.
P: the phenotype value of hybrids or accessions with a certain genotype of a certain marker.
m and i: m is the number of hybrids or accessions with a certain genotype of a certain marker, i = 1.
n and k: n is the number of all the hybrids from the three biparental populations and also the accessions, k = 1.
OPV: the observed phenotype value of an individual, which was the average phenotypic observations of a certain trait over multiple years.

Genomics assisted prediction model

The estimated effects of marker genotypes or pairwise genotype combinations were assigned to the test population and genotype predicted values (GPV) of the traits were calculated by the following equation (2):

$$GPV = \sum_{k}^{n} OPV/n + \alpha \times \sum_{e}^{f} GE + \beta \tag{2}$$

e and f: f is the number of markers, e = 1.
α is a vector of adjustive index.
β is the residual effect.

The accuracy of GAP was evaluated as Pearson's correlation between the GPV and the OPV of individuals in the training population. The selection efficiency was measured as the ratio of OPV-dependent selects against the GPV-depending selects out of the training population (3).

$$\text{Selection efficiency} = \frac{\text{Number of OPV selects in the GPV selected subset}}{\text{Number of GPV selects}} \times 100\% \tag{3}$$

Finally, the exhaustivity of the GAP model was evaluated by using equation (4).

$$\text{Selection exhaustivity} = \frac{\text{Number of OPV selects in the GPV selected subset}}{\text{Number of OPV selects}} \times 100\% \tag{4}$$

Cross-validation of GAP models

Cross-validation was used to evaluate the accuracy of the GAP models [11–13,59]. The marker genotype effects were re-estimated using a sub-sampled data-set of training population and the genotype effect values were input to marker genotypes of hybrids in the test population to predict GPV by the above GAP modeling. These pipelines were run 1,000 times and the average accuracy was compared with that obtained using the complete data set. To compare the reliability of GAP with the conventional GS, the rr-BLUP GS strategy was performed as a control method [12,13,79].

Compliance with Ethics Requirements

All experiments were conducted according to the ethical policies and procedures approved by the ethics committee of China Agricultural University.

Data availability

The pool sequencing data are publicly available in the National Center for Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov/) under project number PRJNA782369. The whole genome sequencing data of the four parental cultivars are publicly available in Sequence Read Archive (SRA) database (https://www.ncbi.nlm.nih.gov/sra/) with the accessions: SRR7510377, SRR7510378, SRR7510381 and SRR7510382.

CRediT authorship contribution statement

Fei Shen: Data curation, Formal analysis, Software, Investigation, Writing - original draft. **Luca Bianco:** Data curation, Formal analysis, Software, Investigation, Writing - original draft, Data curation, Resources. **Bei Wu:** Data curation, Resources. **Zhendong Tian:** Data curation, Resources. **Yi Wang:** Project Administration, Validation and Visualization. **Ting Wu:** Project Administration, Validation and Visualization. **Xuefeng Xu:** Project Administration, Validation and Visualization. **Zhenhai Han:** Conceptualization, Supervision, Writing - review & editing, Funding acquisition. **Riccardo Velasco:** Project Administration, Validation and Visualization. **Paolo Fontana:** Conceptualization, Supervision, Writing - review & editing, Funding acquisition, Data curation, Formal analysis, Software, Investigation, Writing - original draft. **Xinzhong Zhang:** Conceptualization, Supervision, Writing - review & editing, Funding acquisition, Data curation, Resources.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the earmarked fund for China Agriculture Research System (CARS-27) and ZhongBaoLvDu Agriculture Research Center. We thank Dr. Michela Troggio for the helpful discussions.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jare.2022.03.013.

References

[1] Takagi H, Abe A, Yoshida K, Kosugi S, Natsume S, Mitsuoka C, et al. QTL-seq: Rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J 2013;74(1):174–83. doi: https://doi.org/10.1111/tpj.12105.

[2] Jia D, Shen F, Wang Yi, Wu T, Xu X, Zhang X, et al. Apple fruit acidity is genetically diversified by natural variations in three hierarchical epistatic genes: Mdsaur37, mdpp2ch and mdalmtii. Plant J 2018;95(3):427–43. doi: https://doi.org/10.1111/tpj.13957.

[3] Shen F, Huang Z, Zhang B, Wang Y, Zhang X, Wu T, et al. Mapping gene markers for apple fruit ring rot disease resistance using a multi-omics approach. G3 Genes, Genomes, Genet 2019;9:1663–78. doi: https://doi.org/10.1534/g3.119.400167.

[4] Grattapaglia D, Sederoff R. Genetic linkage maps of Eucalyptus grandis and Eucalyptus urophylla using a pseudo-testcross: Mapping strategy and RAPD markers. Genetics 1994;137:1121–37. doi: https://doi.org/10.1093/genetics/137.4.1121.

[5] Kainer D, Lanfear R, Foley WJ, Külheim C. Genomic approaches to selection in outcrossing perennials: focus on essential oil crops. Theor Appl Genet 2015;128(12):2351–65. doi: https://doi.org/10.1007/s00122-015-2591-0.

[6] Werner CR, Voss-Fels KP, Miller CN, Qian W, Hua W, Guan C-Y, et al. Effective Genomic Selection in a Narrow-Genepool Crop with Low-Density Markers: Asian Rapeseed as an Example. Plant Genome 2018;11(2):170084. doi: https://doi.org/10.3835/plantgenome2017.09.0084.

[7] Jung M, Roth M, Aranzana MJ, Auwerkerken A, Bink M, Denancé C, et al. The apple REFPOP—a reference population for genomics-assisted breeding in apple. Hortic Res 2020;7(1). doi: https://doi.org/10.1038/s41438-020-00408-8.

[8] Asoro FG, Newell MA, Beavis WD, Scott MP, Jannink J-L. Accuracy and Training Population Design for Genomic Selection on Quantitative Traits in Elite North American Oats. Plant. Genome 2011;4(2). doi: https://doi.org/10.3835/plantgenome2011.02.0007.

[9] Lenz PRN, Beaulieu J, Mansfield SD, Clément S, Desponts M, Bousquet J. Factors affecting the accuracy of genomic selection for growth and wood quality traits in an advanced-breeding population of black spruce (Picea mariana). BMC Genomics 2017;18(1). doi: https://doi.org/10.1186/s12864-017-3715-5.

[10] Chen Z-Q, Baison J, Pan J, Karlsson Bo, Andersson B, Westin J, et al. Accuracy of genomic selection for growth and wood quality traits in two control-pollinated progeny trials using exome capture as the genotyping platform in Norway spruce. BMC Genomics 2018;19(1). doi: https://doi.org/10.1186/s12864-018-5256-y.

[11] Abed A, Pérez-Rodríguez P, Crossa J, Belzile F. When less can be better: How can we make genomic selection more cost-effective and accurate in barley? Theor Appl Genet 2018;131(9):1873–90. doi: https://doi.org/10.1007/s00122-018-3120-8.

[12] Kumar S, Chagné D, Bink MCAM, Volz RK, Whitworth C, Carlisle C. Genomic selection for fruit quality traits in apple (Malus×domestica Borkh.). PLoS One 2012;7. doi: https://doi.org/10.1371/journal.pone.0036674.

[13] Jiang Y, Schulthess AW, Rodemann B, Ling J, Plieske J, Kollers S, et al. Validating the prediction accuracies of marker-assisted and genomic selection of Fusarium head blight resistance in wheat using an independent sample. Theor Appl Genet 2017;130(3):471–82. doi: https://doi.org/10.1007/s00122-016-2827-7.

[14] Minamikawa MF, Nonaka K, Kaminuma E, Kajiya-Kanegae H, Onogi A, Goto S, et al. Genome-wide association study and genomic prediction in citrus: Potential of genomics-assisted breeding for fruit quality traits. Sci Rep 2017;7(1). doi: https://doi.org/10.1038/s41598-017-05100-x.

[15] Liabeuf D, Sim S-C, Francis DM. Comparison of marker-based genomic estimated breeding values and phenotypic evaluation for selection of bacterial spot resistance in tomato. Phytopathology 2018;108(3):392–401. doi: https://doi.org/10.1094/PHYTO-12-16-0431-R.

[16] Muranty H, Troggio M, Sadok IB, Rifaï MA, Auwerkerken A, Banchi E, et al. Accuracy and responses of genomic selection on key traits in apple breeding. Hortic Res 2015;2(1). doi: https://doi.org/10.1038/hortres.2015.60.

[17] Di Guardo M, Bink MCAM, Guerra W, Letschka T, Lozano L, Busatto N, et al. Deciphering the genetic control of fruit texture in apple by multiple family-based analysis and genome-wide association. J Exp Bot 2017;68:1451–66. doi: https://doi.org/10.1093/jxb/erx017.

[18] Fikere M, Barbulescu DM, Malmberg MM, Shi F, Koh JCO, Slater AT, et al. Genomic Prediction Using Prior Quantitative Trait Loci Information Reveals a Large Reservoir of Underutilised Blackleg Resistance in Diverse Canola (Brassica napus L.) Lines. Plant. Genome 2018;11(2):170100. doi: https://doi.org/10.3835/plantgenome2017.11.0100.

[19] Minamikawa MF, Takada N, Terakami S, Saito T, Onogi A, Kajiya-Kanegae H, et al. Genome-wide association study and genomic prediction using parental and breeding populations of Japanese pear (Pyrus pyrifolia Nakai). Sci Rep 2018;8(1). doi: https://doi.org/10.1038/s41598-018-30154-w.

[20] Hoff JL, Decker JE, Schnabel RD, Seabury CM, Neibergs HL, Taylor JF. QTL-mapping and genomic prediction for bovine respiratory disease in U.S. Holsteins using sequence imputation and feature selection. BMC Genomics 2019;20(1). doi: https://doi.org/10.1186/s12864-019-5941-5.

[21] Kunihisa M, Moriya S, Abe K, Okada K, Haji T, Hayashi T, et al. Identification of QTLs for fruit quality traits in Japanese apples QTLs for early ripening are tightly related to preharvest fruit drop. Breed Sci 2014;64(3):240–51. doi: https://doi.org/10.1270/jsbbs.64.240.

[22] Bai B, Wang L, Lee M, Zhang Y, Alfiko Y, Ye BQ, et al. Genome-wide identification of markers for selecting higher oil content in oil palm. BMC Plant Biol 2017;17. doi: https://doi.org/10.1186/s12870-017-1045-z.

[23] Nishio S, Hayashi T, Yamamoto T, Terakami S, Iwata H, Imai A, et al. Bayesian genome-wide association study of nut traits in Japanese chestnut. Mol Breed 2018;38(8). doi: https://doi.org/10.1007/s11032-018-0857-3.

[24] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25(14):1754–60. doi: https://doi.org/10.1093/bioinformatics/btp324.

[25] Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. Structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics 2012;28(18):i333–9. doi: https://doi.org/10.1093/bioinformatics/bts378.

[26] Wang K, Li M, Hakonarson H. Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38. doi: https://doi.org/10.1093/nar/gkq603.

[27] Edge P, Bafna V, Bansal V. HapCUT2: Robust and accurate haplotype assembly for diverse sequencing technologies. Genome Res 2017;27(5):801–12. doi: https://doi.org/10.1101/gr.213462.116.

[28] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25(16):2078–9. doi: https://doi.org/10.1093/bioinformatics/btp352.

[29] Magwene PM, Willis JH, Kelly JK. The statistics of bulk segregant analysis using next generation sequencing. PLoS Comput Biol 2011;7. doi: https://doi.org/10.1371/journal.pcbi.1002255.

[30] Yang Z, Huang D, Tang W, Zheng Y, Liang K, Cutler AJ, et al. Mapping of Quantitative Trait Loci Underlying Cold Tolerance in Rice Seedlings via High-Throughput Sequencing of Pooled Extremes. PLoS One 2013;8. doi: https://doi.org/10.1371/journal.pone.0068433.

[31] Liao L, Zhang W, Zhang Bo, Fang T, Wang X-F, Cai Y, et al. Unraveling a genetic roadmap for improved taste in the domesticated apple. Mol Plant 2021;14(9):1454–71. doi: https://doi.org/10.1016/j.molp.2021.05.018.

[32] Kerstetter RA, Bollman K, Taylor RA, Bomblies K, Poethig RS. KANADI regulates organ polarity in Arabidopsis. Nature 2001;411(6838):706–9. doi: https://doi.org/10.1038/35079629.

[33] Xiao H, Jiang N, Schaffner E, Stockinger EJ, van der Knaap E. Van Der Knaap E. A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. Science (80-) 2008;319(5869):1527–30. doi: https://doi.org/10.1126/science.1153040.

[34] Rieu I, Ruiz-Rivero O, Fernandez-Garcia N, Griffiths J, Powers SJ, Gong F, et al. The gibberellin biosynthetic genes AtGA20ox1 and AtGA20ox2 act, partially redundantly, to promote growth and development throughout the Arabidopsis life cycle. Plant J 2008;53(3):488–504. doi: https://doi.org/10.1111/j.1365-313X.2007.03356.x.

[35] Schomburg FM, Bizzell CM, Lee DJ, Zeevaart JAD, Amasino RM. Overexpression of a novel class of gibberellin 2-oxidases decreases gibberellin levels and creates dwarf plants. Plant Cell 2003;15(1):151–63. doi: https://doi.org/10.1105/tpc.005975.

[36] Chagné D, Dayatilake D, Diack R, Oliver M, Ireland H, Watson A, et al. Genetic and environmental control of fruit maturation, dry matter and firmness in apple (Malus × domestica Borkh.). Hortic Res 2014;1(1). doi: https://doi.org/10.1038/hortres.2014.46.

[37] Urrestarazu J, Muranty H, Denancé C, Leforestier D, Ravon E, Guyader A, et al. Genome-wide association mapping of flowering and ripening periods in apple. Front Plant Sci 2017;8. doi: https://doi.org/10.3389/fpls.2017.01923.

[38] Migicovsky Z, Gardner KM, Money D, Sawler J, Bloom JS, Moffett P, et al. Genome to Phenome Mapping in Apple Using Historical Data. Plant. Genome 2016;9(2). doi: https://doi.org/10.3835/plantgenome2015.11.0113.

[39] McClure KA, Gardner KM, Douglas GM, Song J, Forney CF, DeLong J, et al. A Genome-Wide Association Study of Apple Quality and Scab Resistance. Plant Genome 2018;11(1):170075. doi: https://doi.org/10.3835/plantgenome2017.08.0075.

[40] Larsen B, Migicovsky Z, Jeppesen AA, Gardner KM, Toldam-Andersen TB, Myles S, et al. Genome-Wide Association Studies in Apple Reveal Loci for Aroma Volatiles, Sugar Composition, and Harvest Date. Plant Genome 2019;12(2):180104. doi: https://doi.org/10.3835/plantgenome2018.12.0104.

[41] Costa F. MetaQTL analysis provides a compendium of genomic loci controlling fruit quality traits in apple. Tree Genet Genomes 2015;11(1). doi: https://doi.org/10.1007/s11295-014-0819-9.

[42] Chagné D, Vanderzande S, Kirk C, Profitt N, Weskett R, Gardiner SE, et al. Validation of SNP markers for fruit quality and disease resistance loci in apple (Malus × domestica Borkh.) using the OpenArray® platform. Hortic Res 2019;6(1). doi: https://doi.org/10.1038/s41438-018-0114-2.

[43] Minamikawa MF, Kunihisa M, Noshita K, Moriya S, Abe K, Hayashi T, et al. Tracing founder haplotypes of Japanese apple varieties: application in genomic prediction and genome-wide association study. Hortic Res 2021;8(1). doi: https://doi.org/10.1038/s41438-021-00485-3.

[44] Devoghalaere F, Doucen T, Guitton B, Keeling J, Payne W, Ling TJ, et al. A genomics approach to understanding the role of auxin in apple (Malus x domestica) fruit size control. BMC Plant Biol 2012;12(1). doi: https://doi.org/10.1186/1471-2229-12-7.

[45] Khan MA, Olsen KM, Sovero V, Kushad MM, Korban SS. Fruit Quality Traits Have Played Critical Roles in Domestication of the Apple. Plant Genome 2014;7(3). doi: https://doi.org/10.3835/plantgenome2014.04.0018.

[46] Zheng W, Shen F, Wang W, Wu B, Wang X, Xiao C, et al. Quantitative trait loci-based genomics-assisted prediction for the degree of apple fruit cover color. Plant Genome 2020;13(3). doi: https://doi.org/10.1002/tpg2.v13.310.1002/tpg2.20047.

[47] Wu B, Shen F, Wang X, Zheng WY, Xiao C, Deng Y, et al. Role of MdERF3 and MdERF118 natural variations in apple flesh firmness/crispness retainability and development of QTL-based genomics-assisted prediction. Plant Biotechnol J 2021;19(5):1022–37. doi: https://doi.org/10.1111/pbi.13527.

[48] Zheng C, Shen F, Wang Y, Wu T, Xu X, Zhang X, et al. Intricate genetic variation networks control the adventitious root growth angle in apple. BMC Genomics 2020;21:1–18. doi: https://doi.org/10.1186/s12864-020-07257-8.

[49] Liu J, Shen F, Xiao Y, Fang H, Qiu C, Li W, et al. Genomics-assisted prediction of salt and alkali tolerances and functional marker development in apple rootstocks. BMC Genomics 2020;21(1). doi: https://doi.org/10.1186/s12864-020-06961-9.

[50] Bink MCAM, Jansen J, Madduri M, Voorrips RE, Durel C-E, Kouassi AB, et al. Bayesian QTL analyses using pedigreed families of an outcrossing species, with application to fruit firmness in apple. Theor Appl Genet 2014;127(5):1073–90. doi: https://doi.org/10.1007/s00122-014-2281-3.

[51] Guan Y, Peace C, Rudell D, Verma S, Evans K. QTLs detected for individual sugars and soluble solids content in apple. Mol Breed 2015;35(6). doi: https://doi.org/10.1007/s11032-015-0334-1.

[52] Laurens F, Aranzana MJ, Arus P, Bassi D, Bink M, Bonany J, et al. An integrated approach for increasing breeding efficiency in apple and peach in Europe. Hortic Res 2018;5(1). doi: https://doi.org/10.1038/s41438-018-0016-3.

[53] Howard NP, van de Weg E, Tillman J, Tong CBS, Silverstein KAT, Luby JJ. Two QTL characterized for soft scald and soggy breakdown in apple (Malus × domestica) through pedigree-based analysis of a large population of interconnected families. Tree Genet Genomes 2018;14(1). doi: https://doi.org/10.1007/s11295-017-1216-y.

[54] Verma S, Evans K, Guan Y, Luby JJ, Rosyara UR, Howard NP, et al. Two large-effect QTLs, Ma and Ma3, determine genetic potential for acidity in apple fruit: breeding insights from a multi-family study. Tree Genet Genomes 2019;15(2). doi: https://doi.org/10.1007/s11295-019-1324-y.

[55] Di Pierro EA, Gianfranceschi L, Di Guardo M, Koehorst-van Putten HJJ, Kruisselbrink JW, Longhi S, et al. A high-density, multi-parental SNP genetic map on apple validates a new mapping approach for outcrossing species. Hortic Res 2016;3(1). doi: https://doi.org/10.1038/hortres.2016.57.

[56] Noiton DAM, Alspach PA. Founding clones, inbreeding, coancestry, and status number of modern apple cultivars. J Am Soc Hortic Sci 1996;121:773–82. doi: https://doi.org/10.21273/jashs.121.5.773.

[57] Ordidge M, Kirdwichai P, Fazil Baksh M, Venison EP, George Gibbings J, Dunwell JM. Genetic analysis of a major international collection of cultivated apple varieties reveals previously unknown historic heteroploid and inbred relationships. PLoS One 2018;13. doi: https://doi.org/10.1371/journal.pone.0202405.

[58] Iwata H, Hayashi T, Terakami S, Takada N, Saito T, Yamamoto T. Genomic prediction of trait segregation in a progeny population: A case study of Japanese pear (Pyrus pyrifolia). BMC Genet 2013;14(1):81. doi: https://doi.org/10.1186/1471-2156-14-81.

[59] Gezan SA, Osorio LF, Verma S, Whitaker VM. An experimental validation of genomic selection in octoploid strawberry. Hortic Res 2017;4(1). doi: https://doi.org/10.1038/hortres.2016.70.

[60] Wu B, Shen F, Chen CJ, Liu Li, Wang X, Zheng WY, et al. Natural variations in a pectin acetylesterase gene, MdPAE10, contribute to prolonged apple fruit shelf life. Plant Genome 2021;14(1). doi: https://doi.org/10.1002/tpg2.v14.110.1002/tpg2.20084.

[61] Kumar S, Molloy C, Muñoz P, Daetwyler H, Chagné D, Volz R. Genome-enabled estimates of additive and nonadditive genetic variances and prediction of apple phenotypes across environments. G3 Genes, Genomes, Genet 2015;5:2711–8. doi: https://doi.org/10.1534/g3.115.021105.

[62] Li X, Wei Y, Acharya A, Hansen JL, Crawford JL, Viands DR, et al. Genomic Prediction of Biomass Yield in Two Selection Cycles of a Tetraploid Alfalfa Breeding Population. Plant. Genome 2015;8(2). doi: https://doi.org/10.3835/plantgenome2014.12.0090.

[63] Annicchiarico P, Nazzicari N, Li X, Wei Y, Pecetti L, Brummer EC. Accuracy of genomic selection for alfalfa biomass yield in different reference populations. BMC Genomics 2015;16(1). doi: https://doi.org/10.1186/s12864-015-2212-y.

[64] Kumar S, Hilario E, Deng CH, Molloy C. Turbocharging introgression breeding of perennial fruit crops: a case study on apple. Hortic Res 2020;7(1). doi: https://doi.org/10.1038/s41438-020-0270-z.

[65] Ma P, Lund MS, Aamand GP, Su G. Use of a Bayesian model including QTL markers increases prediction reliability when test animals are distant from the reference population. J Dairy Sci 2019;102(8):7237–47. doi: https://doi.org/10.3168/jds.2018-15815.

[66] Sarup P, Jensen J, Ostersen T, Henryon M, Sørensen P. Increased prediction accuracy using a genomic feature model including prior information on quantitative trait locus regions in purebred Danish Duroc pigs. BMC Genet 2016;17(1). doi: https://doi.org/10.1186/s12863-015-0322-9.

[67] Hassan MA, Yang M, Fu L, Rasheed A, Zheng B, Xia X, et al. Accuracy assessment of plant height using an unmanned aerial vehicle for quantitative genomic analysis in bread wheat. Plant Methods 2019;15(1). doi: https://doi.org/10.1186/s13007-019-0419-7.

[68] Al Kalaldeh M, Gibson J, Duijvesteijn N, Daetwyler HD, MacLeod I, Moghaddar N, et al. Using imputed whole-genome sequence data to improve the accuracy of genomic prediction for parasite resistance in Australian sheep. Genet Sel Evol 2019;51(1). doi: https://doi.org/10.1186/s12711-019-0476-4.

[69] Wientjes YCJ, Calus MPL, Goddard ME, Hayes BJ. Impact of QTL properties on the accuracy of multi-breed genomic prediction. Genet Sel Evol 2015;47(1). doi: https://doi.org/10.1186/s12711-015-0124-6.

[70] Kemper KE, Bowman PJ, Hayes BJ, Visscher PM, Goddard ME. A multi-trait Bayesian method for mapping QTL and genomic prediction. Genet Sel Evol 2018;50(1). doi: https://doi.org/10.1186/s12711-018-0377-y.

[71] Cros D, Mbo-Nkoulou L, Bell JM, Oum J, Masson A, Soumahoro M, et al. Within-family genomic selection in rubber tree (Hevea brasiliensis) increases genetic gain for rubber production. Ind Crops. Prod 2019;138:111464. doi: https://doi.org/10.1016/j.indcrop.2019.111464.

[72] Li X, Xu X, Shen F, Li W, Qiu C, Wu T, et al. γ-Aminobutyric Acid Participates in the Adult-Phase Adventitious Rooting Recalcitrance. J Plant Growth Regul 2021;40(5):1981–91. doi: https://doi.org/10.1007/s00344-020-10251-9.

[73] Cheng HH, Perumbakkam S, Pyrkosz AB, Dunn JR, Legarra A, Muir WM. Fine mapping of QTL and genomic prediction using allele-specific expression SNPs demonstrates that the complex trait of genetic resistance to Marek's disease is predominantly determined by transcriptional regulation. BMC Genomics 2015;16(1). doi: https://doi.org/10.1186/s12864-015-2016-0.

[74] Leng P-F, Lübberstedt T, Xu M-L. Genomics-assisted breeding – A revolutionary strategy for crop improvement. J Integr Agric 2017;16(12):2674–85. doi: https://doi.org/10.1016/S2095-3119(17)61813-6.

[75] Mason RE, Addison CK, Babar A, Acuna A, Lozada D, Subramanian N, et al. Diagnostic markers for vernalization and photoperiod loci improve genomic selection for grain yield and spectral reflectance in wheat. Crop Sci 2018;58(1):242–52. doi: https://doi.org/10.2135/cropsci2017.06.0348.

[76] Blanpied GD, Silsby KJ. Predicting Harvest Date Window For Apples. A Cornell Coop Ext 1992.

[77] Deell JR, Lum GB, Ehsani-Moghaddam B. Effects of delayed controlled atmosphere storage on disorder development in 'Honeycrisp' apples. Can J Plant Sci 2016;96:621–9. doi: https://doi.org/10.1139/cjps-2016-0031.

[78] Sun R, Chang Y, Yang F, Wang Yi, Li H, Zhao Y, et al. A dense SNP genetic map constructed using restriction site-associated DNA sequencing enables detection of QTLs controlling apple fruit quality. BMC Genomics 2015;16(1). doi: https://doi.org/10.1186/s12864-015-1946-x.

[79] Michel S, Ametz C, Gungor H, Akgöl B, Epure D, Grausgruber H, et al. Genomic assisted selection for enhancing line breeding: merging genomic and phenotypic selection in winter wheat breeding programs with preliminary yield trials. Theor Appl Genet 2017;130(2):363–76. doi: https://doi.org/10.1007/s00122-016-2818-8.