

SCIENTIFIC REPORTS



OPEN

Marker-trait associations and genomic predictions of interspecific pear (*Pyrus*) fruit characteristics

Satish Kumar¹, Chris Kirk², Cecilia Hong Deng³, Angela Shirtliff⁵, Claudia Wiedow², Mengfan Qin⁴, Jun Wu⁴ & Lester Brewer⁵

Interspecific pear (*Pyrus* spp.) hybrid populations are often used to develop novel cultivars. Pear cultivar breeding is a lengthy process because of long juvenility and the subsequent time required for reliable fruit phenotyping. Molecular techniques such as genome-wide association (GWA) and genomic selection (GS) provide an opportunity to fast-forward the development of high-value cultivars. We evaluated the genetic architecture of 10 pear fruit phenotypes (including sensory traits) and the potential of GS using genotyping-by-sequencing of 550 hybrid seedlings from nine interrelated full-sib families. Results from GWA suggested a complex polygenic nature of all 10 traits as the maximum variance explained by each marker was less than 4% of the phenotypic variance. The effect-size of SNPs for each trait suggested many genes of small effect and few of moderate effect. Some genomic regions associated with pear sensory traits were similar to those reported for apple – possibly a result of high synteny between the apple and pear genomes. The average (across nine families) GS accuracy varied from 0.32 (for crispness) to 0.62 (for sweetness), with an across-trait average of 0.42. Further efforts are needed to develop larger genotype-phenotype datasets in order to predict fruit phenotypes of untested seedlings with sufficient efficiency.

Pear is currently grown commercially in almost every continent of the world. The number of catalogued species in the genus *Pyrus* varies according to different studies, but commercial breeding has mainly focussed on three species: *P. communis* (European pear), and two Asian pears namely *P. pyrifolia* and *P. x bretschneideri*¹. High flavour, buttery and juicy texture are among the key characteristics of European pears, while Asian pears generally have crisp texture and subtle flavour. Breeding programmes in New Zealand and elsewhere have combined Asian and European pears to develop crisp, juicy and highly flavoured hybrid cultivars.

Pear breeding programmes generally involve mating of selected parents to create hybrid seedling populations, selection amongst seedling populations, and testing of best performing seedlings. Traditional pear breeding is an expensive and lengthy process primarily because seedlings grown on their own roots typically have a long juvenile period. Reduction of generation time is a focus of many breeding programmes as this has the largest influence on the time taken for new products to reach market². One way of reducing the juvenility period is by growing seedlings in the glasshouse to accelerate the growth rate before planting them into the orchard. New genomic technologies also offer the possibility of accelerating and increasing efficiencies and effectiveness of breeding programmes for new pear cultivars. Most pear fruit traits, such as texture, size, storage ability, resistance to scuffing, flavour and aroma are reported to be affected by numerous loci with small-to-moderate effects³⁻⁶.

A traditional MAS scheme ignores the contribution of all other genes, and thus could result in a lower response to selection, especially for polygenic traits. High-speed and reduced cost of genotyping technologies have facilitated the availability of large number of single nucleotide polymorphic (SNP) markers enabling researchers to study marker-trait associations across the whole genome. High-density genotyping platforms have also facilitated the implementation of genomic selection (GS). GS involves simultaneous estimation of genomewide SNPs

¹The New Zealand Institute for Plant and Food Research Limited, Hawke's Bay Research Centre, Havelock North, New Zealand. ²The New Zealand Institute for Plant and Food Research Limited, Palmerston North Research Centre, Palmerston North, New Zealand. ³The New Zealand Institute for Plant and Food Research Limited, Mount Albert Research Centre, Auckland, New Zealand. ⁴Centre of Pear Engineering Technology Research, Nanjing Agricultural University, Nanjing, 210095, China. ⁵The New Zealand Institute for Plant and Food Research Limited, Motueka Research Centre, Motueka, New Zealand. Correspondence and requests for materials should be addressed to S.K. (email: satish.kumar@plantandfood.co.nz)

	FIRM	CRIS	JUIC	SWET	SOUR	FINT	SCUF	SHAP	RUSS	AVFW
FIRM	0.47									
CRIS	0.81*	0.40								
JUIC	0.23*	0.39*	0.58							
SWET	0.09	0.21*	0.41*	0.16						
SOUR	-0.09	-0.11	-0.02	-0.31*	0.62					
FINT	0.17	0.10	0.26*	0.38*	-0.46*	0.46				
SCUF	-0.24*	-0.19*	-0.08	0.05	0.21*	-0.21*	0.51			
SHAP	-0.05	-0.05	-0.04	-0.07	0.06	-0.10	-0.12	0.57		
RUSS	-0.16	-0.15	-0.24*	-0.05	0.24*	-0.27*	-0.08	0.03	0.59	
AVFW	-0.10	0.04	0.05	0.12	0.30*	-0.32*	0.04	0.33*	0.28*	0.69

Table 1. Genetic parameters of various pear fruit quality traits (firmness: FIRM; crispness: CRIS; juiciness: JUIC; sweetness: SWET; sourness: SOUR; flavour intensity: FINT; fruit scuffing: SCUF; shape: SHAP; russet: RUSS; fruit weight: AVFW). Diagonals are estimated narrow-sense heritability and off-diagonals are genetic correlations. Significant ($p < 0.0001$) correlations are marked with*.

effects to predict genomic breeding values⁷. The traditional MAS is best suited for monogenic or oligogenic traits whereas GS is ideal for traits controlled by many loci with small-to-moderate effects. Therefore, a two-stage selection strategy combining MAS and GS has the potential to accelerate breeding cycles and improve the efficiency of fruit breeding programmes⁸. There are reports of GWAS in germplasm populations of Japanese pear *P. pyrifolia*^{6,9} and *Pyrus* spp.⁵, but the evaluation of GS has only been attempted in Japanese pear⁶.

Despite sensory traits being primary selection criteria for developing new cultivars, there appears to be no report of GWA and GS for pear sensory traits – something that hinders the acceleration of breeding cycles. The main objectives of this study were to conduct GWA to find the candidate genomic regions for pear fruit traits including sensory eating quality traits, and evaluate the potential of GS using a hybrid population derived from crosses between Asian and European pears.

Results

Genetic parameters. The distribution of adjusted fruit phenotypes is shown in Supplementary Fig. S1. Estimates of narrow-sense heritability were low for sweetness (SWET, 0.16), but moderate–high (0.40–0.69) for all other traits (Table 1). Fruit weight (AVFW) was the most heritable trait at 0.69, followed by sourness (SOUR). The highest genetic correlation (0.81) was observed between sensory firmness (FIRM) and crispness (CRIS). Sourness (SOUR) was adversely correlated with sweetness (SWET: -0.31) and flavour intensity (FINT: -0.46), and SWET was favourably correlated (about 0.40) with juiciness (JUIC) and FINT. Fruit with high FIRM were relatively less susceptible to scuffing (SCUF) as there was a significant correlation (-0.24) between these traits. High russet (RUSS) fruit tended to display high SOUR and less JUIC. Estimated genetic correlation of AVFW with SOUR, FINT, SHAP and RUSS was found to be significant (Table 1).

Population structure and LD decay. Overall, a product-moment correlation of 0.69 was observed between pedigree-based and SNP-based estimates of pair-wise coefficient of relationships. The average SNP-based within-family pairwise relationship ranged from 0.50 (family p490) to 0.72 (p449), and the average relationships among seedlings from different families ranged from 0.30 and 0.52 (Fig. 1). Family p449 was derived from crossing two siblings, hence showed a relatively higher relationship coefficient. A plot of the first two principal components (PCs) of the SNP-based realized relationship matrix (*G*) grouped seedlings largely according to their familial relationships (Fig. 2). The first two PCs were used to account for population structure in GWA models. Some individuals did not cluster within their pedigree-assigned full-sib family groupings. For example, two individuals from family p493 clustered with p491, which suggested some pollen contamination or mislabelling. The pattern of LD (r^2) decay in the genetically related population of 550 individuals showed a high degree of LD even at longer distances between markers. For example, the average r^2 for SNPs separated by 0.5 cM, 1.0 cM, and 5.0 cM was 0.29, 0.26, and 0.19, respectively (Fig. 3).

Genetic architecture. Using single-locus GWA, significant ($p < 0.001$) SNP-trait association signals for FIRM and CRIS were identified on LG3 and LG10; for SCUF on LG2, LG4 and LG10; for SHAP on LG11 and LG15; for SOUR on LG1, LG6 and LG13; for SWET on LG4, LG5 and LG13; and for FINT on LG1 and LG8 (Fig. 4). The majority of SNPs individually explained about 0.5% of phenotypic variance (Supplementary Fig. S2), while the maximum effect-size varied between 2% (for AVFW and RUSS) and 3.5% (SWET) (Table 2). The largest-effect SNP was common between FIRM and CRIS on LG10, while the SNPs with largest effect on JUIC and SWET were located on LG5 (Table 2). The distribution of observed ASEs (presented as phenotypic standard deviation (PSD)) for each trait were moderately leptokurtic, suggesting only few SNPs with moderate effect, and the highest ASE varied between 0.41 and 0.68 (Fig. 5). The Kolmogorov-Smirnov test showed that these observed distributions were significantly ($p < 0.05$) different from normal and exponential density functions, but fitted best to a gamma distribution (i.e. the majority of the SNPs having a small effect and a few a moderate effect). The estimated shape and rate parameters of gamma distribution were about 1.25 and 15.0, respectively, for all traits.

The number of significant ($p < 0.001$) SNPs identified using multi-locus methods MLM and MRMLM were 79 and 77, respectively, compared to 67 identified from single-locus GWA (Supplementary Table S1). Majority

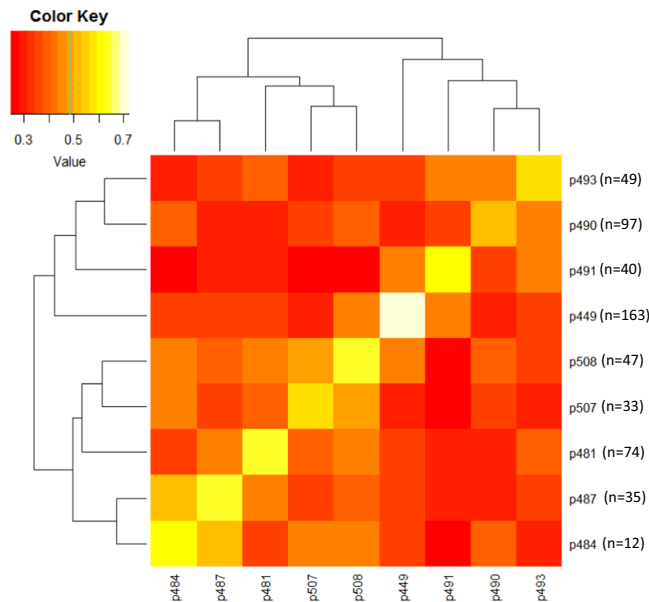


Figure 1. The average within- and between-family pairwise coefficient of relationships for various pear families. The number of offspring in each family (n) is also shown.

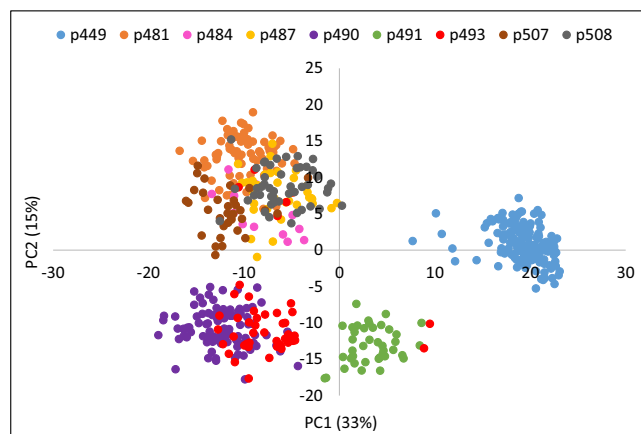


Figure 2. Principal component (PC) analysis plot of the first two components of 550 seedlings derived from their marker genotypes. Pedigree-based grouping (i.e. full-sib families) is also depicted in different colours.

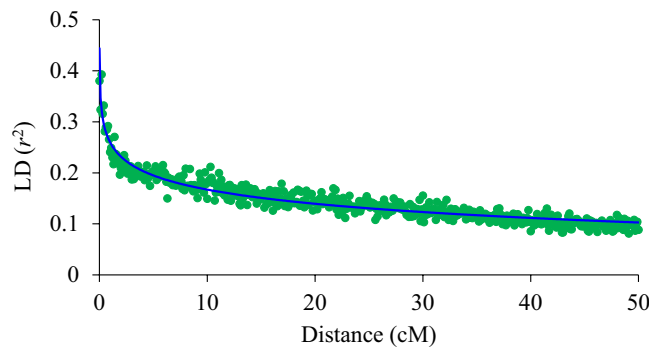


Figure 3. Average linkage disequilibrium (LD) measured as r^2 , for pairs of single nucleotide polymorphisms (SNPs) in increments of 0.1 cM, according to the distance between SNPs in the population of 550 seedlings.

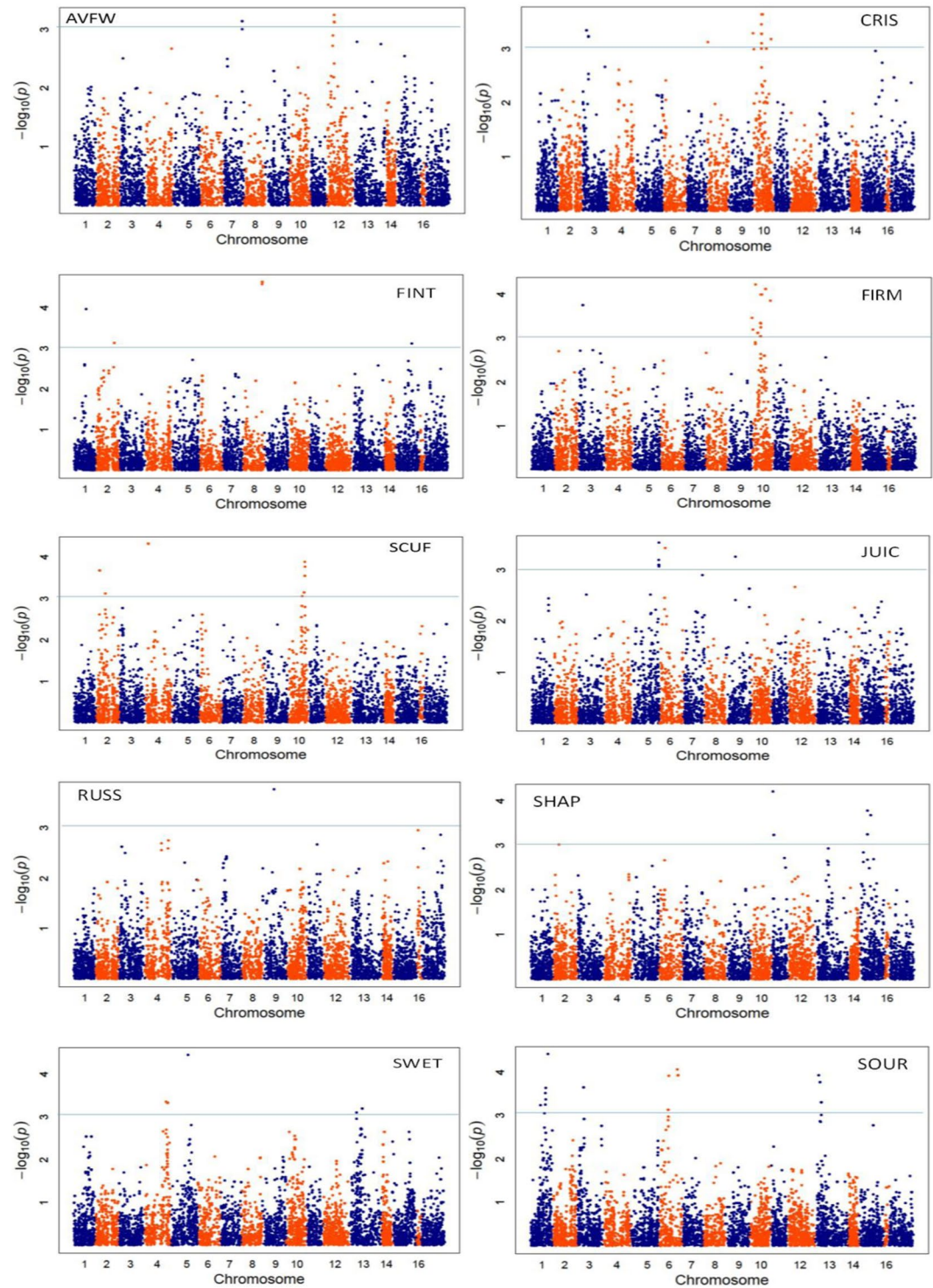


Figure 4. Manhattan plots of the $-\log_{10}(p)$ values for various pear fruit traits (firmness: FIRM; crispness: CRIS; juiciness: JUIC; sweetness: SWET; sourness: SOUR; flavour intensity: FINT; fruit scuffing: SCUF; shape: SHAP; russet: RUSS; fruit weight: AVFW) from a genome-wide scan are plotted against position on each of 17 linkage groups. Blue horizontal line indicates the significance threshold $p < 0.001$.

of the significant SNPs identified using the two multi-locus methods were on the same genomic locations as those from single-locus method (GAPIT) for all traits (Supplementary Figs S3 and S4, Supplementary Table S1). MLM and MRMLM identified SNPs significantly associated with RUSS on LG 4, 16 and 17, which were insignificant in single-locus GWA (Supplementary Table S1).

Trait	SNP	Linkage group	Position (cM)	Significance ($-\log_{10}p$)	ASE	R ² (%)
FIRM	S764_78213	10	95.60	4.00	0.52	3.0
CRISP	S764_78213	10	95.60	3.64	0.48	2.6
JUIC	S210_30053	5	298.97	3.54	0.41	2.2
SWET	S182_250115	5	189.89	4.45	0.33	3.5
SOUR	S465_110990	1	187.40	4.42	0.44	3.0
FINT	S203_236962	8	214.20	4.63	0.41	3.3
SCUF	S29076_1553	10	184.92	3.88	0.44	2.4
SHAP	S4855_850	11	16.65	4.21	0.65	3.0
RUSS	S150_272459	9	123.17	3.77	0.48	2.0
AVFW	S340_202551	12	79.42	3.25	0.53	2.0

Table 2. Single nucleotide polymorphism (SNP) with the largest effect (phenotypic variance explained by the SNP; R²) on various pear fruit quality traits (firmness: FIRM; crispness: CRIS; juiciness: JUIC; sweetness: SWET; sourness: SOUR; flavour intensity: FINT; fruit scuffing: SCUF; shape: SHAP; russet: RUSS; fruit weight: AVFW). The allele substitution effect (ASE), measured in phenotypic standard deviation units, of the largest-effect SNP is also presented. Statistical significance of SNP effect is shown using $-\log_{10}p$ values.

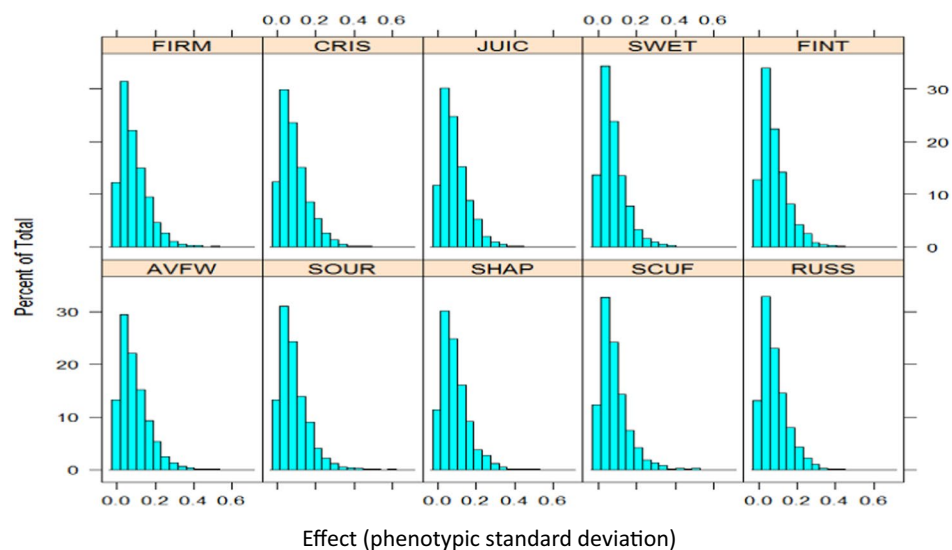


Figure 5. Histogram of allele substitution effects (in phenotypic standard deviation units) of SNPs for pear fruit quality traits (firmness: FIRM; crispness: CRIS; juiciness: JUIC; sweetness: SWET; sourness: SOUR; flavour intensity: FINT; fruit scuffing: SCUF; shape: SHAP; russet: RUSS; fruit weight: AVFW).

Genomic prediction accuracies. We applied a cross-validation scheme by using each full-sib family in turn as a validation population (VP), resulting in a nine-fold cross validation. The results are displayed using a boxplot graph (Fig. 6). The average (across nine families) accuracy varied from 0.32 (CRIS) to 0.62 (SWET), and the range of predicted accuracy was lowest (0.24) for SWET and highest (0.77) for SCUF. The higher prediction accuracy for SWET was partly due to low genetic and phenotypic variability (Table 1, Supplementary Fig. 1). Across all 10 traits, the prediction accuracy was lowest (0.32) for family p449 and the highest (0.51) for family p487.

Discussion

Understanding of the heritability (h^2) of a selection trait is critical in designing molecular breeding strategies such as MAS and GS^{7,8}. Using genomic relationship matrix (GRM) in mixed-model equations would provide improved estimation of h^2 compared to the pedigree-based relationships¹⁰. For a quantitative trait controlled by many genes, the proportion of h^2 explained by a SNP would indicate its worthiness for use in traditional MAS. Most traits in this study were under moderate-to-high genetic control, with GRM-based estimates of h^2 being comparable with earlier published pedigree-based estimates. For example, h^2 of fruit weight (0.68) and acid taste (0.62) were almost identical to those reported by Minamikawa *et al.*⁶. The h^2 for fruit firmness in interspecific populations has been reported to be around 0.60–0.70^{11,12}, slightly higher than that found in this study (0.47). Heritability of FINT in interspecific populations was reported to be 0.54¹³, which is similar to that observed in this study (0.46). However, h^2 of fruit sweetness as reported here (0.16) was half of that observed by Abe *et al.*¹⁴ in hybrid seedlings. In our study, the average (over traits) GRM-based h^2 was 0.50, which was slightly lower than the pedigree-based estimate

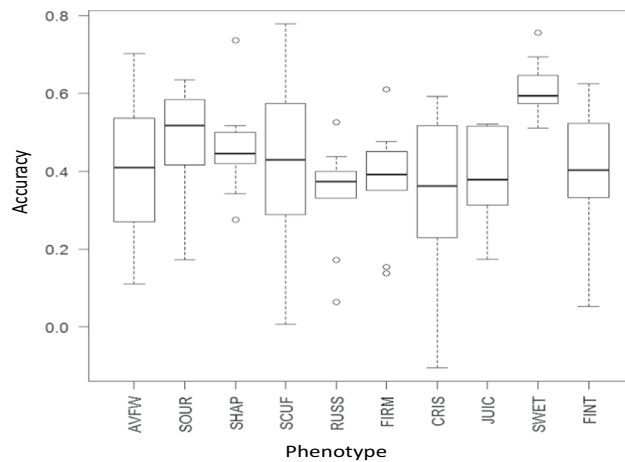


Figure 6. Accuracy of genomic selection for pear fruit quality traits (firmness: FIRM; crispness: CRIS; juiciness: JUIC; sweetness: SWET; sourness: SOUR; flavour intensity: FINT; fruit scuffing: SCUF; shape: SHAP; russet: RUSS; fruit weight: AVFW).

(0.60). In general, GRM-based h^2 were shown to be in the similar range to pedigree-based estimates in various species^{15–17}, which is consistent with results from this study.

LD is a measure of associations between SNP alleles and the alleles at QTL¹⁸. In practice, LD between pairs of SNP markers are obtained because the genomic coordinates of QTLs are generally unknown. The extent of LD obtained in the hybrid seedling populations in this study is higher than that in a population of Asian and European pear germplasm accessions⁵. For markers separated by 10 cM, the r^2 (0.17) in our study is almost identical to that reported (0.18) by Minamikawa *et al.*⁶ in full-sib progenies of Japanese pear. The average r^2 between the adjacent SNPs was 0.47, which is higher than that reported for pear (0.33)⁶ and apple (0.32)¹⁷.

The higher magnitude of short-range and long-range LD in our study could be a result of the genetic structure of the seedling population. Clustering patterns of families (Fig. 2) reflected sharing of parents between the various families. Strong relatedness between different families, and a bottleneck in the breeding history of parents of these families, could be among the factors underpinning the high observed LD^{9,18}. High LD between the adjacent SNPs also plays a key role in improved accuracy of GS¹⁹.

Peak association signals for various traits were located close to genomic regions that have been previously identified. For example, putative QTLs identified for SCUF on LG2, 4 and 10 are in agreement with earlier reports^{5,20}. Similar to our study, there are reports of QTLs for pear fruit shape index on LG2 and LG11^{3,21}. Cao *et al.*²² reported a large effect QTL for apple fruit shape index on LG11, suggesting an orthologous region between apple and pear genomes. A QTL for AVFW was mapped on LG7²¹, which agrees with our results, but some studies^{3,4} mapped QTLs for AVFW on different linkage groups – suggesting this trait has a complex polygenic nature. As might have been expected given the high genetic correlation between FIRM and CRIS, these traits shared the largest-effect SNPs located on LG10 and LG3. Genomic locations on LG10 and LG3 have previously been shown to be associated with fruit firmness in apple^{8,17,23}. QTLs influencing apple and pear fruit traits (e.g. fruit softening; harvest maturity) have been mapped on the same LGs of apple and pear genomes^{6,24} – further evidence for high synteny between pear and apple genomes^{25,26}.

The same SNP marker was found to be associated with SWET and JUIC on LG5. A significant genetic correlation (0.41) between these traits (Table 1) would suggest some genes with pleiotropic effects. There are no reports on QTLs for pear SWET, but QTLs for soluble solids concentration (SSC) have been investigated as SSC is the best objective predictor of sensory sweetness²⁷. In agreement with results of this study, QTLs for SSC have been mapped on LG4⁴ and LG5^{3,21}. A high correlation was reported between malic acid content and SOUR taste in apple fruit²⁷, hence common QTLs for these fruit traits would be expected. Minamikawa *et al.*⁶ mapped a large-effect QTL for acid content in pear fruit on LG6, which supports our GWA results. Unlike our study, there are no QTLs reported for SOUR taste on LG1 and LG13 – suggesting a possibility of population-specific QTL²⁸. Similar to apple FINT QTL²⁹, significant marker-trait associations were observed for pear FINT on LG1 and LG8. SNPs within the apple genes *MdCXE4* (LG1) and *MdMYB44* (LG8) were shown to be associated with apple FINT and/or fruit acid content^{29,30}. Further work is needed to identify and test the functionality of putative candidate genes underpinning marker-trait associations reported in this study, and evaluate their synteny with apple genes.

Multi-locus GWA could be more powerful than single-locus GWA because they account for LD between SNPs, and a small-effect locus may be more apparent when other large-effect loci are already fitted in the model^{31,32}. Similar to a previous study⁶ on pear GWA, a higher number of significant SNPs were detected in the multi-locus GWA compared to the single-locus analysis. However, the majority (nearly 90%) of the significant SNPs, especially the large-effect SNPs, were common between the two approaches in our study (Supplementary Table S1). As some small-effect SNPs (e.g. for RUSS) were not detected in the single-locus model, multi-locus GWA could be more powerful in such cases. However, it's important to note that the threshold ($p < 0.001$) used to identify significant SNPs in our study is low and the false-discovery rate adjusted p -values were > 0.05 for most traits. These results emphasise that an independent validation of the significant SNPs would be required irrespective of the GWA method used.

Understanding the genetic architecture of pear fruit phenotypes is facilitated by having information regarding the distribution of QTL effects. The variance explained by a SNP is mainly a function of the size of the QTL associated with the significant marker and LD between the marker and QTL⁷. Consequently, the distribution of estimated SNP effects should resemble the distribution of the underlying QTL effects. In this study, the maximum variance explained by a SNP for any trait was very small (<4%) and there were very few SNPs with ASE higher than 0.50 PSD, so gamma distributions best fitted to the observed SNP effects for all fruit traits. These results indicated that many QTLs of small effect and only a few moderate effect QTLs control the fruit traits investigated in this study. Hayes *et al.*³³ also reported a gamma distribution of QTL effect sizes in dairy and pig breeding programmes; this is supported by similar observations in other species^{34,35}. Our results suggest that the response from MAS in most pear fruit phenotypes would be small – hence GS could be a better selection tool. MAS would still be useful for traits controlled by major genes, such as pear red skin colour³⁶.

The fundamental difference between MAS and GS is that the former only utilizes the SNPs that are significant in a GWAS, whereas the latter uses high-density genome-wide SNPs so that all QTLs are expected to be in LD with one or more SNPs. In the case of GS, potentially all the genetic variance for a trait can be tracked because the marker effect does not need to exceed a pre-determined significance threshold to be used to predict breeding value⁷. In fruit breeding programmes, the traditional MAS for major gene traits (e.g. disease resistance, skin/flesh colour) is followed by orchard testing for polygenic traits – hence the time required for developing new cultivars is not shortened to any great extent. A two-step approach, i.e. MAS for monogenic traits followed by GS for polygenic traits, obviates the need for Stage-1 seedling testing, hence fast-forwarding the development of new cultivars⁸. This strategy has been evaluated in the New Zealand-based PFR apple breeding programme, and it could be adapted to pear breeding programmes as well.

Various studies on apple have shown that the correlation between sensory and instrumental measures of firmness, acidity, and sweetness were about 0.75, 0.80 and 0.50, respectively^{27,37}, suggesting that genomic prediction accuracies of sensory and instrumental measures could be similar. Minamikawa *et al.*⁶ used instrumental measures of some of the sensory traits used in our study. The average prediction accuracy of traits (fruit weight, firmness, sweetness, and sourness), that were common with Minamikawa *et al.*⁶, was almost identical (cf. 0.46). For the across-family validation scheme implemented in this study, the genetic relationships between the training and validation families and population-level LD are among the key drivers of genomic prediction accuracy. Genomic predictions were reported to be most accurate when models were trained with some individuals from the validation families because of the close relation between training and validation sets^{8,38}. Prediction accuracies of pear fruit phenotypes could be improved further by increasing the training population size, and also by combining parental and breeding populations⁶.

When the training and validation samples are observed independently over different sites/environments, prediction accuracies can be lower depending on the magnitude of genotype-by-site interaction³⁹. Training data would need to be obtained from different sites and years in order to develop robust genomic prediction models for pear fruit traits. A common approach for the evaluation of GS in fruit breeding programmes is to cross-validate using validation samples from the same generation^{6,8,39,40}. The accuracy of predicting phenotypes of successive generations would be lower than within-generation accuracy due to marker-QTL LD decay¹⁹. Meuwissen *et al.*⁴¹ suggested that Bayesian GS models could outperform GBLUP because they capture marker-trait LD that persists in the successive breeding cycles. However, various studies^{42,43} based on empirical data suggested very little or no advantage of Bayesian GS models for most traits, so GBLUP method is widely adopted for GS in commercial breeding programmes^{44,45}.

Similar to animal and plant species^{7,46,47}, fruit breeding programmes need to develop unique multi-generation genotype-phenotype datasets to evaluate persistence of accuracy of genomic predictions over several generations under different environmental conditions. Based on the accuracy of GS in our study, we conclude that it shows strong potential to accelerate the pear breeding cycle by making selections prior to extensive fruit-quality phenotyping. Thus, a GWAS-GS combination could be an effective tool for increasing the efficiency of pear breeding programmes.

Method

Plant material. An interspecific pear breeding programme at Plant & Food Research Limited (PFR) New Zealand was initiated in 1986 using commercial cultivars of European, Chinese and Japanese pear as parents. Second-generation populations were created in 1996 using the best selections produced from the first-generation hybrid families as parents³⁶. A subset of the third-generation families created during 2007–08 using best seedlings from the second-generation were used for this study. A total of 12 second-generation selections were used as parents to create nine families and details of mating design and relatedness among the 12 parents were reported earlier³⁶.

As described earlier by Kumar *et al.*³⁶, a random subset of seedlings that reached a minimum height after growing in a field nursery for 1 year were propagated and planted in PFR's orchard in Motueka during 2011. Fruit were harvested during the fruiting season in 2015 and 2016 and a random sample of six fruit from each seedling was stored for 28 days at 3 °C, then a further 1 day at 20 °C before evaluation³⁶. Sensory traits including firmness (FIRM), crispness (CRISP), juiciness (JUIC), sweetness (SWET), sourness (SOUR) and flavour intensity (FINT) were evaluated on a scale from 0 (=lowest) to 9 (=highest) by two trained assessors, and one overall score for each trait was given to each seedling. Russet (RUSS) was scored visually on a scale from 0 (=lowest) to 9 (=highest). Scuffing (SCUF) was rated on a 0–9 scale (0 = no darkening; 9 = solid brown or black colouration) following the method described by Brewer *et al.*⁴⁸. Fruit shape index (SHAP) was measured using a two dimensional shape chart⁵ and fruit weight (AVFW) was measured as the average weight of the six fruits.

DNA extraction, variants discovery, and linkage map construction. Protocols for DNA extraction, genotyping-by-sequencing (GBS) library preparation, and variant calling were reported earlier as the SNP data used in this study are those used for fine-mapping of pear red skin colour gene in the same nine families³⁶. Resulting SNPs were used to construct a consensus linkage map as reported earlier by Kumar *et al.*³⁶. Briefly, a total of 16 paternal and maternal maps were constructed using *Joinmap* v4.1 software, and then common SNPs shared by at least three maps were selected as a bridge to merge the maps using R package *MergeMap*. Finally, 7,509 high quality SNPs were mapped and used for GWA and GS in this study.

Genomic BLUP model. Phenotypes adjusted for fixed effects (e.g. year and assessor effect) were used for estimation of variance components and best linear unbiased prediction (BLUP) of additive effects using the following model¹⁷:

$$y = \mu \mathbf{1}_n + \mathbf{Za} + e \quad (1)$$

where y is a vector of adjusted phenotypes; μ is an intercept, $\mathbf{1}_n$ is a vector of 1 s; \mathbf{Z} is the known design matrix relating to \mathbf{a} , the unknown vector of random additive genetic effects with $\mathbf{a} \sim N(0, \mathbf{G} \sigma_a^2)$. The scalar σ_a^2 is the additive variance and e is a vector of independent random deviates with variance σ_e^2 . A realized (or genomic) relationship matrix (\mathbf{G})⁴⁹ was obtained using all available SNPs, and Eq. (1) was implemented in software ASReml v3.0⁵⁰. Estimates of variance components derived from Eq. (1) were used for calculating narrow-sense heritability (h^2) as the ratio of additive (σ_a^2) to phenotypic variance ($=\sigma_a^2 + \sigma_e^2$). Product-moment correlations between breeding values were used as estimates of genetic correlation among various traits.

Model validation. The dataset in this study was composed of nine families, so each family in-turn was used as a validation population (VP) and the remaining eight families were used as a training population (TP). GBLUP of seedlings in a VP were predicted by fitting Eq. (1) where their phenotypes were considered as missing values. Prediction accuracy was obtained as the correlation between observed and predicted BLUP of individuals in the VP. This process was repeated so that all nine families were predicted. The mean (averaged over nine validation sets) accuracy and its standard error were presented for each trait.

Estimation of linkage disequilibrium (LD) and genetic architecture. Estimates of LD (r^2), derived from allele frequencies at a pair of loci, were calculated between SNPs located on the same linkage group³⁶. The r^2 estimates were corrected for population structure and cryptic relatedness, and then the average across LGs were plotted against pairwise genetic distance using R package *LdcorSV*⁵¹. The LD decay curve was fitted using a standard logarithmic function. Genetic architecture, in terms of allele substitution effects (ASE) of genome-wide SNPs, of each trait was investigated using the single-locus unified mixed linear model (MLM) approach that accounts for family structure and cryptic relationships as implemented in R package *GAPIT*⁵². The following MLM was implemented for single-locus GWA⁵²:

$$y = X\beta + \mathbf{Za} + \varepsilon \quad (2)$$

where y is a vector of adjusted phenotypes; β is an unknown vector containing estimates of fixed effects (overall mean, ASE of the SNP, and population structure); X and Z are the known design matrices relating to β and \mathbf{a} (the unknown vector of random additive genetic effects with variance $\mathbf{G} \sigma_a^2$), respectively. The scalar σ_a^2 is the additive variance, \mathbf{G} is the realised or genomic relationship matrix (GRM)⁴⁹, and ε is a vector of independent random deviates with variance σ_e^2 . The estimated ASEs of all SNPs were expressed in phenotypic standard deviation units and the observed distribution of SNP effects was compared with theoretical density functions (Normal, Gamma, Exponential) using Kolmogorov-Smirnov test implemented in R package 'fitdistrplus'⁵³. The proportion of phenotypic variance explained by each SNP was also calculated using R package *GAPIT*.

Multi-locus GWA methods could potentially be more powerful than single-locus GWA especially for complex traits. Two multi-locus methods were evaluated in this study; first, a multi-locus mixed model (MLMM³¹) which accounts for population structure and cryptic relatedness and uses a stepwise regression with forward inclusion and backward elimination of SNPs as fixed cofactors; second, a multi-locus random-SNP-effect mixed linear model (MRMLM³²) which fits all SNPs simultaneously as random effects and also accounts for population structure and cryptic relatedness (<https://cran.r-project.org/web/packages/mrMLM.GUI/index.html>).

References

- da Silva, G. J., Souza, T. M., Barbieri, R. L. & Oliveira, A. C. D. Origin, domestication, and dispersing of pear (*Pyrus* spp.). *Adv. Agr.* e541097 (2014).
- Brewer, L. R. & Palmer, J. W. Global pear breeding programmes: goals, trends and progress for new cultivars and new rootstocks. *Acta Hort.* **909**, 105–120 (2011).
- Wu, J. *et al.* High-density genetic linkage map construction and identification of fruit-related QTLs in pear using SNP and SSR markers. *J. Exp. Bot.* **65**, 5771–5781 (2014).
- Yamamoto, T. *et al.* Identification of QTLs controlling harvest time and fruit skin colour in Japanese pear (*Pyrus pyrifolia* Nakai). *Breed. Sci.* **64**, 351–361 (2014).
- Kumar, S. *et al.* Genotyping-by-sequencing of pear (*Pyrus* spp.) accessions unravels novel patterns of genetic diversity and selection footprints. *Hortic. Res.* **4**, 17015 (2017).
- Minamikawa, M. F. *et al.* Genome-wide association study and genomic prediction using parental and breeding populations of Japanese pear (*Pyrus pyrifolia* Nakai). *Sci. Rep.* **8**, 11994 (2018).
- Hayes, B. J., Bowman, P. J., Chamberlain, A. J. & Goddard, M. E. Invited review: Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci.* **92**, 433–443 (2009).
- Kumar, S. *et al.* Genomic selection for fruit quality traits in apple (*Malus × domestica* Borkh.). *Plos One* **7**, e36674 (2012).
- Iwata, H. *et al.* Potential assessment of genome-wide association study and genomic selection in Japanese pear *Pyrus pyrifolia*. *Breed. Sci.* **63**, 125–140 (2013).

10. Speed, D., Hemani, G., Johnson, M. R. & Balding, D. J. Improved heritability estimation from genome-wide SNPs. *Am. J. Hum. Genet.* **91**, 1011–1021 (2012).
11. Shin, I. S., Shin, Y. U. & Hwang, H. S. Heritability of fruit characters of interspecific hybrids between *Pyrus pyrifolia* and *P. ussuriensis* or *P. breschneideri*. *Acta Hort.* **800**, 535–540 (2008).
12. Saito, T. Advances in Japanese pear breeding in Japan. *Breed. Sci.* **66**, 46–59 (2016).
13. Bell, R. L. & Janick, J. Quantitative Genetic Analysis of Fruit Quality in Pear. *J. Am. Soc. Hortic. Sci.* **115**, 829–834 (1990).
14. Abe, K., Saito, Y., Kurihara, A. & Kotobuki, K. Narrow-sense heritability of fruit characters in Japanese pear (*Pyrus pyrifolia* Nakai). *Breed. Sci.* **45**, 1–5 (1995).
15. Yang, J. *et al.* Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* **43**, 519–525 (2011).
16. Fan, B. *et al.* Genome-wide association study identifies loci for body composition and structural soundness traits in pigs. *Plos One* **6**, e14726 (2011).
17. Kumar, S. *et al.* Novel genomic approaches unravel genetic architecture of complex traits in apple. *BMC Genomic* **14**, 93 (2013).
18. Flint-Garcia, S. A., Thornsberry, J. M. & Buckler, E. S. Structure of linkage disequilibrium in plants. *Annu. Rev. Phys. Chem.* **53**, 319–348 (2002).
19. Habier, D., Fernando, R. L. & Dekkers, J. C. The impact of genetic relationship information on genome-assisted breeding values. *Genetics* **177**, 2389–2397 (2007).
20. Saeed, M. *et al.* Genetic, metabolite and developmental determinism of fruit friction discolouration in pear. *BMC Plant Biol.* **14**, 1 (2014).
21. Zhang, R. P. *et al.* An AFLP, SRAP, and SSR genetic linkage map and identification of QTLs for fruit traits in pear (*Pyrus L.*). *Plant Mol. Biol. Rep.* **31**, 678–687 (2013).
22. Cao, K. *et al.* Candidate gene prediction via quantitative trait locus analysis of fruit shape index traits in apple. *Euphytica* **206**, 381–391 (2015).
23. McClure, K. A. *et al.* A genome-wide association study of apple quality and scab resistance. *Plant Genome* **11**, 170075 (2018).
24. Costa, F. A. *et al.* Map position and functional allelic diversity of Md-Exp7, a new putative expansin gene associated with fruit softening in apple (*Malus × domestica* Borkh.) and pear (*Pyrus communis*). *Tree Genet. Genomes* **4**, 575–586 (2008).
25. Yamamoto, T. *et al.* Genetic linkage maps of Japanese and European pears aligned to the apple consensus map. *Acta Hort.* **663**, 51–66 (2004).
26. Wu, J. *et al.* The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Res.* **23**, 396–408 (2013).
27. Harker, F. R. *et al.* Sensory interpretation of instrumental measurements 2: sweet and acid taste of apple fruit. *Postharvest Biol. Tech.* **24**, 241–250 (2002).
28. Würschum, T. Mapping QTL for agronomic traits in breeding populations. *Theor. Appl. Genet.* **125**, 201–10 (2012).
29. Kumar, S. *et al.* Genome-wide scans reveal genetic architecture of apple flavour volatiles. *Mol. Breed.* **35**, 118 (2015).
30. Jia, D. *et al.* Apple fruit acidity is genetically diversified by natural variations in three hierarchical epistatic genes Md SAUR 37, Md PP 2 CH and Md ALMTII. *Plant J.* **95**, 427–443 (2018).
31. Segura, V. *et al.* An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat. Genet.* **44**, 825–830 (2012).
32. Wang, S. B. *et al.* Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Sci. Rep.* **6**, 19444 (2016).
33. Hayes, B. E. & Goddard, M. E. The distribution of the effects of genes affecting quantitative traits in livestock. *Genet. Sel. Evol.* **33**, 209 (2001).
34. Albert, A. Y. *et al.* The genetics of adaptive shape shift in stickleback: pleiotropy and effect size. *Evolution* **62**, 76–85 (2008).
35. Caballero, A., Tenesa, A. & Keightley, P. D. The nature of genetic variation for complex traits revealed by GWAS and regional heritability mapping analyses. *Genetics* **201**, 1601–1613 (2015).
36. Kumar, S. *et al.* Fine-mapping and validation of the genomic region underpinning pear red skin colour. *Hortic. Res.* **6**, 29 (2019).
37. Oraguzie, N. *et al.* Postharvest assessment of fruit quality parameters in apple using both instruments and an expert panel. *Postharvest Biol. Tech.* **52**, 279–87 (2009).
38. Legarra, A., Robert-Granié, C., Manfredi, E. & Elsen, J. M. Performance of genomic selection in mice. *Genetics* **180**, 611–618 (2008).
39. Burgueño, J., de los Campos, G., Weigel, K. & Crossa, J. Genomic prediction of breeding values when modeling genotype-environment interaction using pedigree and dense molecular markers. *Crop Sci.* **52**, 707–719 (2012).
40. Biscarini, F. *et al.* Genome-enabled predictions for fruit weight and quality from repeated records in European peach progenies. *BMC Genomics* **18**, 432 (2017).
41. Meuwissen, T. H., Hayes, B. J. & Goddard, M. E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–1829 (2001).
42. Zhang, X. *et al.* Rapid cycling genomic selection in a multiparental tropical maize population. *G3-Genes Genom. Genet.* **7**, 2315–2326 (2017).
43. Hofheinz, N., Borchardt, D., Weissleder, K. & Frisch, M. Genome-based prediction of test cross performance in two subsequent breeding cycles. *Theor. Appl. Genet.* **125**, 1639–1645 (2012).
44. Hickey, J. M. *et al.* Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nat. Genet.* **49**, 1297–1302 (2017).
45. Wiggans, G. R., Cole, J. B., Hubbard, S. M. & Sonstegard, T. S. Genomic selection in dairy cattle: the USDA experience. *Annu. Rev. Anim. Biosci.* **5**, 309–327 (2017).
46. Wolc, A. *et al.* Persistence of accuracy of genomic estimated breeding values over generations in layer chickens. *Genet. Sel. Evol.* **43**, 23 (2011).
47. Zhang, X. *et al.* Rapid cycling genomic selection in a multi-parental tropical maize population. *G3-Genes Genom. Genet.* **7**, 2315–2326 (2017).
48. Brewer, L. R., Morgan, C., Alspach, P. A. & Volz, R. K. Heritability and parental breeding value estimates of abrasion-induced skin discolouration on pear fruit. *Acta Hort.* **909**, 127–135 (2011).
49. VanRaden, P. M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **91**, 4414–4423 (2008).
50. Gilmour, A. R., Gogel, B. J., Cullis, B. R. & Thompson, R. *ASReml User Guide Release 3.0*. VSN International Ltd, Hemel Hempstead, United Kingdom. (2009).
51. Mangin, B. *et al.* Novel measures of linkage disequilibrium that corrects the bias due to population structure and relatedness. *Heredity* **108**, 285–291 (2012).
52. Lipka, A. E. *et al.* GAPIT: Genome association and prediction integrated tool. *Bioinformatics* **28**, 2397–2399 (2012).
53. Delignette-Muller, M. L. & Dutang, C. Fitdistrplus: An R package for fitting distributions. *J. Stat. Software* **64**, 1–34 (2015).

Acknowledgements

This research was partly funded by the New Zealand Ministry of Business, Innovation and Employment (MBIE). Funding from the National Science Foundation of China (31820103012) is thankfully acknowledged. The PFR Markers & Mapping Team is thanked for their help in collecting leaf samples. Marlene Aldworth helped in fruit harvest and phenotyping. Constructive feedback from Margaret Carpenter and Luis Gea is gratefully acknowledged.

Author Contributions

S.K. and L.B. conceived and designed the study. C.W. and C.K. performed leaf collection and GBS library preparation. C.H.D. conducted bioinformatics analysis, and M.Q. and J.W. constructed linkage map. A.S. and L.B. led the phenotyping. S.K. performed GWA and genomic predictions and wrote the first draft. All authors helped to edit the manuscript, and read and approved the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-45618-w>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



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

© The Author(s) 2019