Genome-enabled estimates of additive and non-additive genetic variances and prediction of apple phenotypes across environments

Satish Kumar*, Claire Molloy*, Patricio Muñoz§, Hans Daetwyler†, David Chagné‡, Richard Volz*

*The New Zealand Institute for Plant & Food Research Limited (PFR), Hawke’s Bay Research Centre, Havelock North, New Zealand
§Agronomy Department, University of Florida, Gainesville, FL 32611, USA
†AgriBio Centre, La Trobe University, Bundoora 3083, Victoria, Australia
‡PFR, Palmerston North Research Centre, Palmerston North, New Zealand
Short running title: Genomic prediction of apple phenotypes

Key words: GBLUP, genomic prediction, apple breeding, heritability, Malus × domestica

Corresponding author:
Dr Satish Kumar
The New Zealand Institute for Plant & Food Research Limited
Hawke’s Bay Research Centre, Crosses Road
Private Bag 1401
Havelock North 4130
New Zealand
Tel.: +64 69758893; Fax: +64 69758881;
Email: Satish.Kumar@plantandfood.co.nz
ABSTRACT

The non-additive genetic effects may have an important contribution to total genetic variation of phenotypes, so estimates of both the additive and non-additive effects are desirable for breeding and selection purposes. Our main objectives were to: estimate additive, dominance and epistatic variances of apple (*Malus × domestica* Borkh.) phenotypes using relationship matrices constructed from genome-wide dense single nucleotide polymorphism (SNP) markers; and compare the accuracy of genomic predictions using genomic best linear unbiased prediction models with or without including non-additive genetic effects. A set of 247 clonally replicated individuals was assessed for six fruit quality traits at two sites, and also genotyped using Illumina 8K SNP array. Across several fruit quality traits, the additive, dominance and epistatic effects contributed about 30%, 16% and 19%, respectively, to the total phenotypic variance. Models ignoring non-additive components yielded upwardly biased estimates of additive variance (heritability) for all traits in this study. The accuracy of genomic predicted genetic values (GEGV) varied from about 0.15 to 0.35 for various traits, and these were almost identical for models with- or without including non-additive effects. However, models including non-additive genetic effects further reduced the bias of GEGV. Between-site genotypic correlations were high (>0.85) for all traits, and genotype-site interaction accounted for <10% of the phenotypic variability. The accuracy of prediction, when the validation set was present only at one site, was generally similar for both sites, and varied from about 0.50 to 0.85. The prediction accuracies were strongly influenced by trait heritability, and genetic relatedness between the training and validation families.
INTRODUCTION

The main goal of most fruit breeding programs is to identify the best performing individuals (or genotypes) as potential commercial cultivars, and the secondary goal is to identify genotypes that can be used as parents in future crosses. The genetic value (GV), which includes both additive and non-additive genetic effects, predicts a genotype’s performance for a trait when asexually propagated plants are used for production (White et al. 2007). Non-additive genetic variation results from interactions between alleles, whereby interactions between alleles at the same locus are called dominance, and interactions between alleles at different loci are called epistasis. When non-additive genetic effects are an important source of variation, the use of additive genetic value or breeding value (BV) to determine potential commercial cultivars may lead to the selection of genotypes that do not have the highest genetic value (GV).

In best linear unbiased prediction (BLUP) of BVs, information from performance of relatives (e.g. half-sibs, full-sibs) is incorporated through the use of the additive relationship matrix derived from pedigree records (Henderson 1975). This matrix represents the expected parental contributions (i.e. genome sharing) between individuals in the pedigree. Non-additive performance of relatives could also be included in BLUP analysis through non-additive relationship matrices derived from the pedigree. However, partitioning of total genetic components into additive and non-additive parts, using standard pedigree-based models, requires specific family structures involving mating designs (e.g. factorial, diallel). Furthermore, clonally replicated progeny trials enable the estimation of individual-level BV and GV, as well as the magnitude of genotype-by-environment interaction (G×E) (Burdon and Shelbourne 1974). BLUP analysis of such trials can provide ranking for the selection of
potential parents using additive effects (BV), and potential commercial cultivars from combining predicted additive and non-additive effects (i.e. GV).

DNA marker-based relationship coefficients are more accurate than pedigree-based relationships because they account for deviations from expected parental contributions due to Mendelian sampling or segregation distortion. Furthermore, genomic relationship matrices (GRM) are particularly advantageous when pedigree relationships are not available for some individuals and/or when the recorded pedigree information may contain errors (Muñoz et al. 2014a; Speed and Balding 2015). Studies on animal and plant species have demonstrated that the accuracy of artificial selection can be increased by using GRM rather than pedigree-based relationship matrices (Hayes et al. 2009; Lee et al. 2010; Kumar et al. 2013). Numerous approaches have been developed to derive GRM based on genome-wide single nucleotide polymorphic (SNP) data (Van Raden 2008; Legarra et al. 2009). Similarly to the additive GRM, approaches to derive SNP-based estimates of non-additive relationships (e.g. dominance, epistasic) have also been developed (Su et al. 2012; Sun et al. 2014).

An approach termed genomic BLUP (or GBLUP), which directly estimates genomic breeding values (GEBV) using additive GRM, is now commonly used in genomic selection (GS) studies of animal and plant species (Hayes et al. 2009; Daetwyler et al. 2010; Heslot et al. 2012; Lin et al. 2014). GBLUP has been shown to be equivalent to ridge regression BLUP (Goddard 2009; Stranden and Garrick 2009), and one of the advantages of GBLUP is that the model can use the well-known mixed models framework similar to the traditional pedigree-based BLUP models. Also, GBLUP can be extended to include additive and non-additive genetic effects to potentially increase the accuracy of estimates when non-additive genetic
effects make a substantial contribution to the genetic variation (Su et al. 2012; Muñoz et al. 2014b; Sun et al. 2014).

Although numerous studies on estimation of genetic parameters of apple fruit phenotypes have been conducted (Durel et al. 1998; Alspach and Oraguzie 2002; Kouassi et al. 2009; Kumar et al. 2011), there has not yet been a systematic dissection of genetic variance of apple phenotypes into additive and non-additive variances. The objectives of this study were threefold. The first was to estimate additive, dominance and epistatic variances using SNP-based relationship matrices. The second was to compare the accuracy of genomic predictions using GBLUP models with or without including non-additive genetic effects. The third was to evaluate genotype-by-environment interaction (G×E) and its impact on prediction accuracy.

MATERIALS AND METHODS

Genetic Material

A field trial consisting of 17 full-sib apple (Malus × domestica Borkh.) families, involving 25 advanced selections as parents, was used as the source of budwood for this study. Based on pedigree records, the average coefficient of relationship between the 25 parents was 0.15 (File S1), resulting in varying degree of relationships between the 17 full-sib families. Budwood from each of the 255 randomly selected genotypes (15 from each of the 17 families) were collected in the winter (June-July) of 2010. This budwood was held in cold storage before chip-budding eight lateral buds from each of the 255 genotypes onto nursery-grown one-year-old ‘M9’ rootstocks in spring (September 2010), making eight clonal replicates of each of the 255 genotypes. In the following winter (July 2011), all successfully propagated seedlings were planted (3.0 x 0.5 m spacing) at two orchard sites, one in each of
the key apple production regions in New Zealand, namely Hawkes Bay (HB, lat. 39°40´S, long. 176°53´E) and Motueka (MOT, lat. 41°6´S, long. 172°58´E). Generally, HB is warmer during winter, spring and autumn than MOT. Rainfall and humidity are higher in MOT than in HB, particularly during spring. Comparing the soil types of the two sites, MOT has a more fertile soil than the HB orchard – thus trees are more vigorous at the former site. The initial goal was to plant four replicates of each genotype at each of the two sites; however, because of problems in the propagation survival, the number of plants available for each genotype varied. Ninety percent of the genotypes were represented by at least three clonal replicates at each site. There were only five individuals with only one replicate at each site. All trees received standard commercial management for nutrition, pesticide, fruit hand-thinning, and irrigation.

Phenotypes

Fruit harvesting, which was carried out in the fruiting season (February – May) during 2013 and 2014, began when fruit from each tree were judged to be mature, based on a change in skin background colour from green to yellow, and when the starch pattern index was between 2 and 3 using a scale from 0 (full starch) to 7 (no starch). Phenotypes of each individual were collected in both years. One sample of six fruit from each clonal replicate of each genotype were stored for 70 days at 0.5°C, followed by a further 7 days at 20°C before evaluation. Fruit firmness (FF) was determined on opposite sides of each fruit after peel removal using a Fruit Texture Analyzer (GÜSS) fitted with an 11-mm diameter probe tip. A list of visual, sensory, and instrumental traits assessed following Kumar et al. (2011) is shown in Table 1.

Single nucleotide polymorphism (SNP) genotyping
A total of 247 (out of 255) individuals were genotyped using the IRSC apple 8K SNP array v1 (Chagné et al. 2012), based on the Infinium® II technique. Genomic DNA (gDNA) was extracted from leaves using the NucleoSpin® Plant II kit (Macherey-Nagel GmbH and Co KG, Düren, Germany), and quantified using the Quant-iT™ PicoGreen® Assay (Invitrogen). Two hundred nanograms of gDNA were used as a template for the reaction, following the manufacturer’s instructions. SNP genotypes were scored using the Genotyping Module (version 1.8.4) of the Illumina® GenomeStudio software (Illumina Inc.). The reliability of each genotype call was measured using the GenCall score set at a minimum of 0.15, which is a lower bound for calling genotypes relative to its associated cluster. SNPs were subsequently discarded using the following sequence of criteria: GenCall score at the 50% rank (50% GC) < 0.40; cluster separation (ClusterSep) <0.25; more than 5% missing calls. SNPs with minor allele frequency <0.05 were discarded, and then segregation discrepancy was checked within families. A total of 2828 SNPs were retained after various quality checks. BEAGLE 3.1 software (Browning and Browning 2007), with default settings, was then used for imputing missing SNP genotypes.

Genomic BLUP model

A two-step approach was used to estimate genetic parameters and make genomic predictions. The first step consisted of calculating best linear unbiased estimates (BLUEs) to account for fixed effects such as replicates and year, so that each individual had a single phenotypic value for each trait at each site. These estimates were then used as ‘phenotypes’ for estimation of variance components and BLUP of additive, dominance and epistatic effects using the following linear mixed model:

\[ y = Xb + Z_1a + Z_2d + Z_3i + Z_4t_1 + Z_5t_2 + Z_6t_3 + e \]  \hspace{1cm} (1)
where $y$ is a vector ($ns \times 1$) of adjusted phenotypes (BLUEs) on a trait; $n$ is the number of individuals (247) and $s$ is the number of sites (2); $b$ is a vector of fixed effect (i.e. the intercept, site); $a \sim N(0, G_{a}\sigma_{a}^{2})$, $d \sim N(0, G_{d}\sigma_{d}^{2})$, and $i \sim N(0, G_{aa}\sigma_{aa}^{2})$ are the vectors ($n \times 1$) of random additive, dominance and additive-by-additive epistatic effects respectively; $G_{a}$, $G_{d}$ and $G_{aa}$ are the additive, dominance and epistatic relationship matrices ($n \times n$) respectively; $\sigma_{a}^{2}$, $\sigma_{d}^{2}$ and $\sigma_{aa}^{2}$ represent additive, dominance and epistatic variance respectively; $\epsilon \sim N(0, I_{s} \otimes \sigma_{e}^{2})$ is a vector of random residual terms and $I$ is an identity matrix ($ns \times ns$); $t_{1} \sim N(0, I_{s} \otimes G_{a}\sigma_{as}^{2})$ and $t_{2} \sim N(0, I_{s} \otimes G_{d}\sigma_{ds}^{2})$ and $t_{3} \sim N(0, I_{s} \otimes G_{aa}\sigma_{aa}^{2})$ are the vectors ($ns \times 1$) of random interactions of $a$, $d$, and $i$ with site respectively; $I_{s}$ represents an identity matrix with order equal to the number of sites; $\otimes$ denotes the Kronecker product operation; $Z_{1} – Z_{3}$ are incidence matrices ($ns \times n$) for the main genetic effects, and $Z_{4} – Z_{6}$ are incidence matrices ($ns \times ns$) of interaction of main genetic effects with site respectively.

The genomic relationship matrices ($G_{a}$, $G_{d}$, and $G_{aa}$) were constructed using SNP marker information according to methods from previous studies (VanRaden 2008; Su et al. 2012). Briefly, $G_{a} = \frac{MM'}{\Sigma 2p_{i} q_{i}}$, where $M$ is an n×m matrix ($n =$ number of individual, $m =$ number of SNP loci) representing genotypes at each locus. The coefficient of $i^{th}$ column of the $M$ matrix are (0-2$p_{i}$), (1-2$p_{i}$) and (2-2$p_{i}$) for genotypes AA, AB and BB respectively; $p_{i}$ and $q_{i}$ are the frequencies of allele A and B at $i^{th}$ SNP locus, respectively. Similarly, $G_{d} = \frac{HH'}{\Sigma 2p_{i} q_{i} (1-2p_{i} q_{i})}$, where $H$ is a n×m matrix of heterozygosity coefficients with elements (0-2$p_{i} q_{i}$) and (1-2$p_{i} q_{i}$) for homozygous (AA, BB) and heterozygous (AB) individuals at $i^{th}$ SNP locus, respectively.

The epistatic genomic relationship matrix ($G_{aa}$) was obtained as G#G, where # denotes the Hadamard product operation (Su et al. 2012). Pedigree-based relationship matrices,
accounting for genetic relationships between the 25 parents, were also used for comparison purposes. Equation 1 was implemented in software ASReml v3.0 (Gilmour et al. 2009).

In the subsequent text, Equation 1 will be termed as Model ADE, which includes three genetic components (a, d and i). BLUP-GV (= a + d + i) of all 247 individuals were obtained from Equation 1 using all available data. Estimates of variance components derived from Equation 1 were used for calculating narrow-sense (\(h^2\)) and broad-sense (\(H^2\)) heritability, as the ratio of additive (\(\sigma^2_a\)) to phenotypic variance (sum of all variance components in the model), and the ratio of total genetic variance (\(\sigma^2_G = \sigma^2_a + \sigma^2_d + \sigma^2_i\)) to phenotypic variance, respectively. The method of Yamada (1962) was followed to calculate between-site genotypic correlations using estimates of variance components from Equation 1. Reduced forms of Equation 1, by including only the additive component (Model A), and additive and dominance components (Model AD), were also tested. Goodness of fit for each model (i.e. with and without non-additive genetic effects) was evaluated by the log-likelihood value. The superiority of an alternative model over an additive model, obtained using all available data, was tested using a likelihood ratio test.

**Model validation**

To account for the family structure of our data, we applied a cross-validation scheme by using each full-sib family in-turn as a validation population (VP) and the remaining 16 families as a training population (TP), resulting in a 17-fold cross validation. Prediction accuracy of a model was estimated as the correlation between genomic GV (GEGV) and the observed BLUP-GV of individuals in the VP. GEGV was defined as \((a + d + i), (a + d), \) and \(a\) for the models ADE, AD and A, respectively. The observed BLUP-GVs were linearly
regressed on the GEGV, where the regression coefficient reflected the degree of bias of the GEGV, and a regression coefficient of one indicated no bias.

Two different validation strategies were used to address practical situations. The first strategy mimics a scenario where a set of individuals (i.e. VP) have not been field tested at any of the two sites (HB and MOT) of this study, i.e. individuals in the VP were treated as missing at both sites. Thus, the prediction accuracy of a model in the first validation strategy was based solely on the performance of individuals in the TP. The second validation strategy considers a scenario where individuals in the VP are tested at one site (say, HB) but missing at the other site (say, MOT). The prediction accuracy in this strategy relies on the performance of TP in the target environment (MOT) and the performance of all individuals (VP and TP) from the test site (HB). This process was repeated so that all 17 families at both sites were predicted.

Data availability

File S1 contains pedigree relationships between parents, and Files S2 and S3 contains phenotypic and genotypic data respectively.

RESULTS

Basic features of the population

The average fruit weights (WT) at the HB and MOT sites were 183 g and 192 g, respectively, with a coefficient of variation (CV) of 23% at both sites (Table 1). Statistically significant differences between the site averages were observed for fruit weight (WT), greasiness (GRE), fruit firmness (FF) and crispness (CRI). The coefficients of variation (CV) were high and similar at both sites suggesting that heterogeneity of trait expression was consistent across the
two key apple production sites. The average GRE score was higher but FF and CRI were significantly lower at MOT than at HB (Table 1).

A plot of the first two principal components of the SNP data matrix grouped seedlings largely according to their familial relationships (Figure 1A). In almost every family there were some individuals that did not cluster within their pedigree-assigned full-sib family groupings. Such cases were more prevalent in families such as A536, A332, A334 and A535. A distribution of the SNP-based additive coefficient of relationship between all pairs of 247 individuals is shown in Figure 1B. The SNP-based and pedigree-based average coefficients of relationship were 0.34 and 0.19 respectively. The SNP-based relatedness coefficients for pairs of supposedly half-sibs and full-sibs varied considerably (Figure 1C). The average within-family relatedness coefficient varied from 0.35 (A535) to 0.66 (A537), and the between-family relatedness coefficient varied between 0.22 and 0.50 (File S4). Results in Figures 1A and 1C suggested some pollen contamination and/or mislabelling, which can be accounted for in analyses such as those in the present study by using SNP-based relationships.

**Estimates of variance components and heritability**

The additive variance ($\sigma^2_A$) was the major source of variability for WT (52%) and FF (50%), but the non-additive sources were very prominent for other traits (Table 2; File S5). On average (across traits), the additive, dominance and epistatic effects contributed about 30%, 16% and 19% to the total phenotypic variance. The pedigree-based average estimates of additive and non-additive variance were 22% and 25% respectively (results not shown). For various traits, the magnitude of variance due to interaction of genetic effects with site varied between 0 and 3%, 0 and 4%, and 0 and 8% for additive, dominance and epistatic effects respectively. Estimates of between-site genotypic correlation varied between 0.87 (GRE) and
0.95 (WT), and the interaction of genetic effects with site collectively accounted for less than 10% of the total phenotypic variance (Table 2).

The narrow-sense heritability ($h^2$) estimate was low (< 0.20) for JUI and FIN, moderate (0.20 – 0.40) for GRE and CRI, and high (>0.40) for WT and FF (Table 2). Estimated broad-sense heritability ($H^2$) was 0.34 for FIN and >0.60 for other traits, reflecting a substantial amount of non-additive genetic variance. Figure 2 shows that the estimated $h^2$ from different models varied between 0.39 and 0.70 when the epistatic effect was dropped (i.e. model AD), and between 0.47 and 0.82 when dominance and epistatic effects were excluded (i.e. model A) from Equation 1 (see File S6 for details). These are much higher than $h^2$ estimates (0.11 – 0.52) of various traits obtained using the model ADE. The likelihood ratio test showed that model ADE had relatively low -2 log likelihood compared with model A (File S7), but the goodness of fit was not improved significantly ($p > 0.05$; degrees of freedom = 4).

**Accuracy of prediction**

Prediction accuracies for the first validation strategy (no field presence of validation set) are shown in Figure 3. Correlation between the observed BLUP-GV and genomic predicted GV (i.e. GEGV) was essentially the same for the additive model (model A) and for the model including additive and non-additive effects (model ADE), and varied from about 0.15 (for JUI) to 0.35 (for FF) (Figure 3). The coefficient of regression of BLUP-GV on GEGV for various traits varied between 0.75 and 1.0 for model ADE, and between 0.55 and 0.75 for model A (File S8). These results suggested that the degree of bias is higher for the additive model than for the model ADE. For the second validation strategy (validation set present only at one site), the accuracy of predicted GEGV was generally similar for both sites, and varied from about 0.50 (for FIN) to about 0.83 (for WT and FF) (Figure 4). Prediction accuracy of
validation families was moderately correlated \( r = 0.44 \) with the degree of genetic relatedness to the training families (File S9). Prediction accuracy also showed strong relationship \( r = 0.81 \) with trait heritability (File S10).

**DISCUSSION**

**Additive and non-additive genetic variances**

Genetic relationship matrices, constructed using genome-wide SNP markers, were fitted using a GBLUP approach to estimate non-additive genetic variation. The difference between the pedigree-based and SNP-based additive coefficient of relationship \( (0.19 \text{ v. } 0.34) \) reflected that there were perhaps many more ancestral relationships not captured by the pedigree records. Different kinds of relationships (e.g. half-sib, full-sib) along with permanent environmental or maternal effects shared by siblings make it challenging to partition non-additive variance from the additive variance using traditional methods (Hill *et al.* 2008; Lee *et al.* 2010). Nevertheless, ours is the first formal attempt to estimate additive, dominance and epistatic variances of apple phenotypes. The genome-enabled estimate of non-additive variance was similar to the additive variance (35% versus 30%), suggesting that non-additive effects are as important as additive effects for various apple phenotypes. Mäki-Tanila and Hill (2014) showed that when heterozygosity at individual loci is high, multilocus epistasis could lead to large increases in the non-additive part of the genotypic variance. The use of F1 families, where heterozygosity is expected to be high, could have contributed to the large non-additive variance observed in this study. Similarly to the results of a study on loblolly pine (Muñoz *et al.* 2014b), the genome-enabled estimates of genetic variance components in this study were somewhat larger than their pedigree-based counterparts. This could be due to better capture of relatedness using genome-wide markers.
The additive variance decreased when dominance effects were included in the model, and it decreased further when epistatic effects were considered. The average $h^2$ was 0.69, 0.52 and 0.30 for the models A, AD and ADE respectively (Figure 2). These results highlight the confounding nature of the additive and non-additive effects. In other words, effects which the model ADE allocated to dominance and epistatic effects, were now being absorbed by the inferred additive genetic component in model A. This means that models ignoring non-additive components would yield upwardly biased estimates of additive variance if the true variance components of ignored effects were not zero (Lu et al. 1999). The covariance between full-sibs and half-sibs includes $\frac{1}{4}\sigma_{aa}^2$ and $\frac{1}{16}\sigma_{aa}^2$ respectively, so there is a ‘hidden’ non-additive component even if an additive model is fitted. Moreover, depending on the distribution of allele frequencies and linkage disequilibrium (LD) patterns, some of the variance due to interaction of alleles can manifest as additive variance (Hill et al. 2008; Zuk et al. 2012). Similarly to our results, decrease in $h^2$ by including non-additive effects in the model have been reported by various studies on forest trees (Waldmann et al. 2008; Araújo et al. 2012; Muñoz et al. 2014b).

The estimated $h^2$ of WT (0.52) in this study was very similar to those in some other reports (Alspach and Oraguzie 2002; Kouassi et al. 2009). Although the $h^2$ of FIN (0.11) was similar to that reported by Kouassi et al. (2009), it was lower than that found by Durel et al. (1998) (0.39). Heritability ($h^2$) of CRI and JUI has been reported to vary between 0.15 and 0.40 (Durel et al. 1998; Alspach and Oraguzie 2002; Kouassi et al. 2009), but higher heritabilities (c. 0.60) have been reported (Kumar et al. 2011). The estimated $H^2$ in our study varied between 0.34 (FIN) and 0.79 (FF), with an average of about 0.65 (Table 2). King et al. (2000), using clonally replicated seedlings from a single full-sib family, reported estimates of $H^2$ of FF, CRI, and JUI to be 0.52, 0.57, and 0.46, respectively. McKay et al. (2011) reported...
estimated $H^2$ values of JUI and CRI as 0.72 and 0.76 respectively, which were higher than those observed in our study (0.60 and 0.66, respectively). The ratio of non-additive ($\sigma^2_d+\sigma^2_{de}$) to additive ($\sigma^2_a$) variance varied between 0.48 (WT) and 2.35 (JUI), with an average of 1.53, suggesting that non-additive variance contributed significantly to the expression of apple fruit phenotypes.

**Influence of explicit fitting of non-additive components on prediction accuracy**

BLUP-BV or de-regressed EBVs have commonly been used as phenotypes in most applications of genomic selection in animal and plant species (Garrick *et al*. 2009; Hayes *et al*. 2009; Kumar *et al*. 2012; Resende *et al*. 2012). Such data allow only the estimation of allele substitution effects, so distinguishing between additive and non-additive effects is not possible. The use of BLUEs as phenotypes in genomic predictions is desirable to distinguish between additive and non-additive effects, and to reduce the computation time compared with using multiple records per individual (Su *et al*. 2012; Daetwyler *et al*. 2014; Xu *et al*. 2014).

Using the phenotypes adjusted for non-genetic effects (replication, year), we obtained BLUP-GV (sum of additive, dominance and epistatic effects) of all individuals for cross-validation of genomic predictions. There was a near-perfect correlation between BLUP-GV and BLUEs for all traits, reflecting that pedigree/family relationships had little effect on the performance ranking (i.e. BLUP-GV), in contrast to unbalanced datasets with little or no clonal replication of individuals (e.g. Garrick *et al*. 2009). Similar observations were reported in studies on forest tress species (Zapata-Valenzuela *et al*. 2012).

The accuracy of predicting unobserved BLUP-GV varied from about 0.15 to 0.35, and these were almost identical between the models with- or without including non-additive effects. As shown by Xu *et al*. (2014), the lack of improvement in accuracy could be attributed to the
high correlation between different estimated variance components. For all traits in our study, vectors of estimated additive effects were highly correlated with non-additive effects for both the marker-based (c. 0.80) and pedigree-based (c. 0.60) models. If genes are not independently distributed (i.e. LD) in the parents of a population, a covariance between the additive and non-additive effects would be introduced (Melchinger 1988), and what otherwise would be epistatic variance becomes additive or dominance variance (Hill et al. 2008). Studies on animal species showed that prediction accuracy increased when both (additive and non-additive) effects rather than just additive effects were included (Su et al. 2012; Sun et al. 2014). Although the prediction accuracy of models A and ADE was similar for the apple phenotypes, the prediction bias was lower for the latter. The regression coefficients for the predictions using the model ADE were closer to 1, so including non-additive effects did not improve the accuracy but helped to reduce the bias of predicting GV.

Relatively low prediction accuracy (Figure 3) in this study compared with that in the study by Kumar et al. (2012) was mainly due to the sample size (c. 250 cf. c. 1000). Moreover, the large family sizes and a random cross-validation scheme in Kumar et al. (2012) meant that genomic predictions were essentially made within the same families that were part of the training population. For across family validation scheme as in this study, genetic relationship between the training and validation families, and population-level LD, are the key drivers of prediction accuracy (Legarra et al., 2008). Relationships at the level of parents and grandparents resulted in high relatedness among the 17 families in this study (File S4). Genetic relatedness between the training and validation families, and trait heritability were positively correlated with the prediction accuracy (Files S9, S10), which is consistent with previous studies (Habier et al., 2007; Daetwyler et al., 2008).
Genotype-environment interaction (GxE)

It is common practice to include some common ‘checks’ or ‘controls’ across various test environments, to extrapolate performance of an accession at an untested site. Pedigree information could also be used for predicting performance if some known relatives were tested at the target site. Genome-wide SNPs, which provide precise estimates of genetic relationships, will provide better links between accessions at different sites for the understanding of GxE and predicting performance in different environments. The accuracy of predicting unobserved BLUP-GV at HB using performance from MOT, or vice versa, was about 0.75 for all traits except JUI (0.62) and FIN (0.52). Although the GxE variance components for JUI and FIN were similar to those for other traits (Table 2), the lower heritability of these two traits might have contributed to these results.

The prediction accuracy of the second validation strategy (Figure 4) was at least twice that of the first validation strategy (Figure 3). These results supported earlier findings that predicting the performance of untested individuals is more challenging than predicting the performance of individuals that have been evaluated at some sites (Burgueño et al. 2012). When the training and validation samples are observed independently over different sites and ages/years, prediction accuracies can be affected, depending on the magnitude of genotype-by-site and genotype-by-age interactions (Resende et al. 2012).

This study showed that non-additive genetic effects are important sources of genetic variation for apple phenotypes, so there is a good case to account for these effects when selecting individuals as potential cultivars. For selection of individuals as parents for new crosses the additive effects is probably more relevant as the non-additive effects do not transfer very effectively to the next generation. The GBLUP approach is relatively simple to implement to
estimate additive and non-additive variances and to predict GEBV and GEGV. GBLUP models fitting only the additive effects can predict phenotype with similar accuracy to models fitting both the additive and non-additive effects, but the degree of bias will be lower for the latter. Also, GBLUP models ignoring non-additive effects will provide an overestimated additive variance that will exaggerate the expected genetic gains in breeding. The high correlation between the additive and non-additive effects may be due to the genetic relatedness between samples, so including non-additive genetic effects in GBLUP models may still improve prediction accuracy in populations of less-related individuals – this hypothesis needs to be tested.

ACKNOWLEDGEMENTS

The PFR Markers & Mapping Team is thanked for help in collecting leaf samples. Slipstream Automation Ltd, New Zealand is thanked for providing high-throughput DNA extraction services. We also thank AgResearch Invermay, New Zealand, for providing Illumina® SNP Array genotyping. This research was partly supported by Prevar (Contract no. 26015) and the New Zealand Ministry of Business, Innovation and Employment (MBIE) through a Core Funding programme. PFR colleagues Dr Luis Gea and Dr Jo Stephens are thanked for their helpful comments. We thank reviewers and the editor for constructive feedback on the manuscript.

LITERATURE CITED


Melchinger, A. E., 1988 Means, variance, and covariances between relatives in hybrid populations with disequilibrium in the parent populations, pp. 400-415 in *Proceedings*

Muñoz, P. R., M. F. Resende, D. A. Huber, T. Quezada, M. D. Resende et al., 2014a

Muñoz, P. R., M. F. Resende, S. A. Gezan, M. D. V. Resende, G. de los Campos et al., 2014b


Table 1 Apple (*Malus × domestica* Borkh.) fruit mean trait values at Hawke’s Bay (HB) and Motueka (MOT) sites, New Zealand.

<table>
<thead>
<tr>
<th>Trait (abbreviation)</th>
<th>Description (unit)</th>
<th>HB</th>
<th></th>
<th>MOT</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight (WT)</td>
<td>Average individual fruit weight (g)</td>
<td>183*</td>
<td>23</td>
<td>192*</td>
<td>23</td>
</tr>
<tr>
<td>Greasinessa (GRE)</td>
<td>Greasy sensation when finger touches skin</td>
<td>2.3*</td>
<td>66</td>
<td>3.3*</td>
<td>62</td>
</tr>
<tr>
<td>Firmness (FF)</td>
<td>Force required to puncture skinless tissue (kg/cm$^2$)</td>
<td>8.7*</td>
<td>21</td>
<td>8.1*</td>
<td>25</td>
</tr>
<tr>
<td>Crispnessa (CRI)</td>
<td>Amount and pitch of sound generated when a 1 cm$^3$ segment of flesh is first bitten with the front teeth</td>
<td>4.8*</td>
<td>21</td>
<td>4.5*</td>
<td>25</td>
</tr>
<tr>
<td>Juicinessa (JUI)</td>
<td>Amount of free fluid released on chewing</td>
<td>4.3</td>
<td>19</td>
<td>4.1</td>
<td>22</td>
</tr>
<tr>
<td>Flavour intensity² (FIN)</td>
<td>Sensory perception of ‘overall’ flavour intensity</td>
<td>3.7</td>
<td>9</td>
<td>3.6</td>
<td>11</td>
</tr>
</tbody>
</table>

a Scoring from 0 (lowest) to 9 (highest)

* Site means were significantly ($p < 0.01$) different
Table 2 Estimates of additive ($\sigma_a^2$), dominance ($\sigma_d^2$), epistatic ($\sigma_{as}^2$) genetic variance and their interaction variance in apple (Malus × domestica Borkh.) families with site ($\sigma_{as}^2$, $\sigma_{ds}^2$, $\sigma_{aas}^2$, respectively), expressed as the percentage of phenotypic variance (defined as the sum of variance components in the model), obtained using the Equation 1 (Model ADE). Estimates of narrow-sense heritability ($h^2$), broad-sense heritability ($H^2$), and between-site genotypic correlation ($r_B$) are also shown for various traits (WT: fruit weight; GRE: greasiness; FF: fruit firmness; CRI: crispness; JUI: juiciness; FIN: flavour intensity).

<table>
<thead>
<tr>
<th>Source</th>
<th>WT</th>
<th>GRE</th>
<th>FF</th>
<th>CRI</th>
<th>JUI</th>
<th>FIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_a^2$</td>
<td>51.79</td>
<td>26.52</td>
<td>49.63</td>
<td>23.39</td>
<td>18.05</td>
<td>10.67</td>
</tr>
<tr>
<td>$\sigma_d^2$</td>
<td>6.70</td>
<td>22.86</td>
<td>19.99</td>
<td>23.37</td>
<td>23.78</td>
<td>0.00</td>
</tr>
<tr>
<td>$\sigma_{as}^2$</td>
<td>18.38</td>
<td>23.64</td>
<td>9.15</td>
<td>19.17</td>
<td>18.58</td>
<td>23.41</td>
</tr>
<tr>
<td>$\sigma_{ds}^2$</td>
<td>2.31</td>
<td>1.08</td>
<td>1.32</td>
<td>3.38</td>
<td>0.00</td>
<td>0.80</td>
</tr>
<tr>
<td>$\sigma_{aas}^2$</td>
<td>2.02</td>
<td>1.72</td>
<td>4.36</td>
<td>0.00</td>
<td>0.00</td>
<td>0.95</td>
</tr>
<tr>
<td>$\sigma_e^2$</td>
<td>18.80</td>
<td>16.17</td>
<td>15.56</td>
<td>30.70</td>
<td>35.02</td>
<td>61.16</td>
</tr>
<tr>
<td>$h^2$</td>
<td>0.52</td>
<td>0.27</td>
<td>0.50</td>
<td>0.23</td>
<td>0.18</td>
<td>0.11</td>
</tr>
<tr>
<td>$H^2$</td>
<td>0.77</td>
<td>0.73</td>
<td>0.79</td>
<td>0.66</td>
<td>0.60</td>
<td>0.34</td>
</tr>
<tr>
<td>$r_B$</td>
<td>0.95</td>
<td>0.87</td>
<td>0.93</td>
<td>0.95</td>
<td>0.93</td>
<td>0.88</td>
</tr>
</tbody>
</table>
Figure 1 Principal coordinates (PC) plot of all 247 individuals in the apple (*Malus × domestica* Borkh.) families tested, derived from their marker genotypes (A). Pedigree-based grouping (i.e. full-sib families) is also depicted in different marker colours/shape (A). Distribution of the SNP-based coefficients of relationships between all pairs of 247 individuals (B), and for pairs of half-sibs and full-sibs is also shown (C).
Figure 2 Changes in estimated heritability of various apple (Malus × domestica Borkh.) traits (WT: fruit weight; GRE: greasiness; FF: fruit firmness; CRI: crispness; JUI: juiciness; FIN: flavour intensity) by fitting various genetic effects (A: additive; AD: additive and dominance; ADE: additive, dominance and epistatic) in the model.
Figure 3 Mean (across families) prediction accuracy of various apple \( (Malus \times domestica) \) traits (WT: fruit weight; GRE: greasiness; FF: fruit firmness; CRI: crispness; JUI: juiciness; FIN: flavour intensity) using the additive component (Model A) and both additive and non-additive components (Model ADE). Validation samples were assumed untested at both sites. The error bars represent standard error of the mean.
**Figure 4** Mean (across families) prediction accuracy of various apple (*Malus × domestica* Borkh.) traits (WT: fruit weight; GRE: greasiness; FF: fruit firmness; CRI: crispness; JUI: juiciness; FIN: flavour intensity) for the cross-validation strategy where validation samples are assumed untested at a particular site. The prediction model included both additive and non-additive components (i.e. model ADE). The error bars represent standard error of the mean. Hawke’s Bay=HB; Motueka= MOT; New Zealand.