



Implementation of different relationship estimate methodologies in breeding value prediction in kiwiberry (*Actinidia arguta*)

Daniel Mertten¹ · Samantha Baldwin² · Canhong H. Cheng³ · John McCallum⁴ · Susan Thomson⁵ · David T. Ashton⁶ · Catherine M. McKenzie⁷ · Michael Lenhard⁸ · Paul M. Datson⁹

Received: 8 June 2023 / Accepted: 2 October 2023 / Published online: 18 October 2023
© The Author(s) 2023

Abstract In dioecious crops such as *Actinidia arguta* (kiwiberries), some of the main challenges when breeding for fruit characteristics are the selection of potential male parents and the long juvenile period. Currently, breeding values of male parents are estimated through progeny tests, which makes the breeding of new kiwiberry cultivars time-consuming and costly. The application of best linear unbiased prediction (BLUP) would allow direct estimation of sex-related traits and speed up kiwiberry breeding. In

this study, we used a linear mixed model approach to estimate narrow sense heritability for one vine-related trait and five fruit-related traits for two incomplete factorial crossing designs. We obtained BLUPs for all genotypes, taking into consideration whether the relationship was pedigree-based or marker-based. Owing to the high cost of genome sequencing, it is important to understand the effects of different sources of relationship matrices on estimating breeding values across a breeding population. Because of the increasing implementation of genomic selection in crop breeding, we compared the effects of incorporating different sources of information in building relationship matrices and ploidy levels on the accuracy of BLUPs' heritability and predictive ability. As kiwiberries are autotetraploids, multivalent chromosome formation and occasionally double reduction can occur during meiosis, and this can affect the accuracy of prediction. This study innovates the breeding programme of autotetraploid kiwiberries. We demonstrate that the accuracy of BLUPs of male siblings, without phenotypic observations, strongly improved when a tetraploid marker-based relationship matrix was used rather than parental BLUPs and female siblings with phenotypic observations.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11032-023-01419-8>.

D. Mertten (✉)
The New Zealand Institute for Plant and Food Research Ltd (PFR), Auckland 1142, New Zealand
e-mail: Daniel.Mertten@plantandfood.co.nz

D. Mertten · M. Lenhard
Institute for Biochemistry and Biology, University of Potsdam, 14476 Potsdam-Golm, Germany

S. Baldwin · J. McCallum · S. Thomson
Plant and Food Research, Lincoln 7608, New Zealand

C. H. Cheng · P. M. Datson
Kiwifruit Breeding Centre, Auckland 1142, New Zealand

D. T. Ashton
Plant and Food Research, Nelson 7010, New Zealand

C. M. McKenzie
Plant and Food Research, Te Puke 3182, New Zealand

Keywords Best linear unbiased prediction · Autopolyploid · *Actinidia arguta* · Genomic selection · Accuracy · Cross-validation

Introduction

Successful plant breeding is the art of identifying and selecting potential parents with desirable traits and exceptional performance for the next round of crossing from within a variable population. Modern plant breeding can utilise further information, such as genomics and new statistical analysis tools, to improve parental selection.

Variation in any trait is due to genetic, environmental and other factors (such as maternal effects and crop management tools). Selection on a trait requires that much of the variation in the trait be due to segregating heritable genetic factors rather than due to the environment. To estimate the genetic effects, statistical methodologies that estimate best linear unbiased predictions (BLUPs) have been developed to estimate variance components and predict breeding values (Patterson and Thompson 1971; Henderson 1974). With improvements in computational power and computing techniques, these approaches have been modified and improved to increase their accuracy in predicting breeding values. Incorporating pedigree information and environmental effects into these statistical methods increases the accuracy of genetic analysis of quantitative traits by eliminating some of the bias linked to the sharing of genes among related individuals and has led to a faster genetic gain in animal- as well crop-breeding programmes (Kennedy and Sorenson 1988; Kennedy et al. 1988). The covariance describing kinship among individuals is represented by the additive relationship matrix (A) or numerator relationship matrix (NRM). Rules for calculating A have been developed for diploid animal species such as livestock, where gametes carry only one of the two alleles (Henderson 1976). Algorithms have been developed to calculate an additive relationship matrix (A) and its inverse (A^{-1}) in diploid species, mainly for animal breeding. In diploid species, it is assumed that gametes cannot carry two or more alleles that are identical by descent (IBD) because of meiotic reduction division. In autopolyploids, where non-preferential pairing of chromosome occurs, such an assumption cannot be made. The A matrix is the probability that an allele is identical by descent (kinship coefficient) among individuals, multiplied by two for diploids, by four for tetraploids, by six for hexaploids and so forth (Gallais 2003; Kerr et al. 2012).

In many plant species, the estimation of breeding values is confounded by polyploidy. Whole genome

duplication is a common event in angiosperms (Soltis et al. 2004, 2015; Wood et al. 2009; Baduel et al. 2018). Different forms of polyploids are defined by the number of multiple coexisting chromosome sets and the pairing pattern of chromosome inheritance. The two extreme forms are auto- or allopolyploidy, but a mixed form of both can also be found (allo-autopolyploidy). Autopolyploids result from genome duplication or the combination of two very closely related species and show non-preferential chromosome pairing between their homologous chromosomes during meiosis. By contrast, allopolyploids result from the combination of chromosome sets from two or more distantly related species and show preferential chromosome pairing behaviour during meiosis (Sears 1976; Soltis and Soltis 1999; Comai 2005; Soltis et al. 2007). Because of non-preferential chromosome pairing, it has been thought that autopolyploids show a high frequency of multivalent chromosome formations. However, in some autopolyploid crops, including blueberry, kiwifruit and potato, almost exclusively bivalent chromosome formation with occasionally (<10%) multivalent formation has been observed (Soltis et al. 1993; Qu et al. 1998; Fjellstrom et al. 2001; Wu et al. 2014; Choudhary et al. 2020). Through multivalent chromosome formations, double reduction can occur during meiosis, resulting in sister-chromatids segregating into the same gametes (Bradshaw 2007; Bourke et al. 2015; Muthoni et al. 2015).

When dealing with autopolyploid and pedigree-based relationship information, bias in heritability estimation and breeding value prediction can occur if double reduction is ignored. Double reduction affects the inbreeding rate of a breeding population and therefore the kinship between individuals. Studies in blueberries and potatoes revealed correlations between double reductions and the genome locations of quantitative trait loci (Bourke et al. 2015).

Polyploidy is an important consideration for breeding of kiwifruit (*Actinidia*). Several species and hybrids of *Actinidia* have been introduced into cultivation, the main two being *Actinidia chinensis* (Planch.) var. *chinensis* and *Actinidia chinensis* var. *deliciosa* (A. Chev.) A. Chev. (Huang and Ferguson 2007; Datson et al. 2017). Studies on different *Actinidia* species have revealed non-preferential chromosome pairing in natural and induced polyploidy selections. Non-preferential chromosome pairing occurs during meiosis when chromosomes pair with more

than one potential homologue partner. Both natural and induced polyploids in kiwifruit can form multivalent chromosome formations (Mertten et al. 2012; Wu et al. 2014). This finding suggests an adjustment of NRM to polyploidy, and double reduction (ω) should be considered in the breeding strategy of *Actinidia* spp. and other crops that include true autopolyploids with occasional double reduction (Haynes and Douches 1993; Kerr et al. 2012; Choudhary et al. 2020).

Tetraploid *Actinidia arguta* (Sieb. et Zucc.) Planch. ex Miq. var. *arguta* ($2n=4x=116$) (kiwiberry) is one of the species recently introduced into cultivation. Kiwiberries produce small fruit with a soft, hairless, edible skin. Like other *Actinidia* species, kiwiberries are usually dioecious, with staminate flowers on male vines and pistillate flowers on female vines. Female flowers fully develop ovaries, styles, stigmata and stamens but produce non-viable, empty pollen. Male flowers produce viable pollen but rudimentary female organs lacking ovules that are able to develop into fruit (Rizet 1945; Schmid 1978; White 1990).

The current breeding approach for kiwifruit species, such as *A. arguta*, parallel approaches employed in animal breeding. This methodology entails the selection of genotypes using techniques like single-seed descent, accompanied by the utilisation of pedigree records to preserve critical relationship information. Consequently, genotypes displaying desirable trait performances are carefully chosen and clonally propagated for subsequent commercial cultivation.

Owing to the sex linkage of some desired quantitative traits in dioecious crop breeding, it is not possible to select the superior individuals of genotypes within a cross when phenotype observations cannot be made, e.g., the breeding values of fruit characteristics in male genotypes within the same cross are estimated as family means and cannot be distinguished on an individual level. Thus, there is a need to find methods that enable individual estimation of a trait-value for a non-expressed trait. In particular, the selection of male parents requires progeny testing, as they do not provide phenotypic information on their genetic background, for the breeder. Recently, genomic methods have been developed to enable this prediction (Testolin 2011; Datson et al. 2017; Cheng et al. 2019). In polyploids with their multiple homologous chromosome sets, allele dosage information is crucial to estimating marker-based additive variance–covariance relationships between individuals to

predict breeding values. To date, there is no publication addressing the application of genomically estimated breeding values (GEBV) to breeding of autotetraploid kiwiberries.

To explore the effects of incorporating probabilistic versus realised relationship matrices into a linear mixed model equation for commercially important fruit quality traits and vine characteristics, we modified the equation through the application of different types of relationship matrices (using pedigree or genomic information) and varying the complexity of assumptions of chromosome inheritance. The effects of these modifications on the breeding value estimates for parental generations, female progenies with trait records and male progenies with no records, are compared.

Materials and methods

Plant population and phenotyping

A seedling population of tetraploid *A. arguta*, consisting of two incomplete factorial crossing designs (Supplementary Table 1), was generated within the parental breeding programme at The New Zealand Institute for Plant and Food Research Limited (PFR). In 2014, 1791 seedlings from 50 crosses were planted at the PFR Motueka Research Centre (41°50' S; 172°58' E). A minimum of 20 randomly selected seedlings with a mix of males and females per cross was planted in groups of seven with replication at cross-level in the field trial. Plants were separated by distances of 0.5 m within a row and 3.0 m between rows. Seedlings were grown on a pergola support system, the most common production system used in New Zealand. Upon plant establishment, the observed number of seedlings per cross exhibited a range spanning from a minimum of 2 seedlings to a maximum of 80 seedlings. Notably, only 3 of the 50 crosses yielded fewer than 10 seedlings, whereas 5 crosses yielded a number of 40 seedlings. Overall, there were 36 progeny per cross on average, with a median of 39 progeny for each cross. Plants were established in the field for 2 years, after which fruiting vines were assessed. Two canes from the current growing season were trained horizontally during summer and remained after winter pruning for vine assessments. The numbers of progeny within each cross varied, and phenotype data of some

individuals were missing in some years, making the phenotypic data incomplete.

One vine characteristic (fruit load) and five fruit characteristics (fruit weight, dry matter, ripe soluble solids content, fruit circularity crosswise and lengthwise) were assessed for this study. During the 5-year trial, fruit load was recorded in 2017 and 2018. Fruit load was scored from 0 to 9 (Supplementary Table 2), but category zero individuals, with no fruit, were not included in this study. The assessment of fruit load in female vines followed a scoring system based on the number of fruits they bore. Vines with a fruit count of up to 4 received a score of 0.5. Those with up to 10 fruits were assigned a score of 1, while vines carrying up to 30 fruits garnered a score of 2. As the fruit load increased, the scoring correspondingly escalated: vines with up to 60 fruits achieved a score of 3, those with up to 100 fruits were rated at 4 and vines containing up to 200, 300, and 400 fruits received scores of 5, 6 and 7, respectively. Vines that developed up to 500 fruits were designated a score of 8, whereas vines shouldering more than 500 fruits attained the highest score of 9. Fruit assessments were performed when fruit maturity was indicated by >90% of seeds being black. Fruit weight (g), recorded from 2017 to 2019, was the mean of 30 randomly picked fruit across each vine. Dry matter percentage (DM) was recorded from 2017 to 2019. Three representative fruits were sampled randomly, and a cross-sectional slice of 2–5 mm was cut for DM calculation (Fenton and Kennedy 1998). Ten fruits were sampled from harvest and kept at 1 °C for 14 days, followed by 1 day at room temperature to ripen. Ripe soluble solids content (SSC) of three sampled ripe fruits was measured in 2018 and 2019 using a digital pocket refractometer (ATAGO®). Six fruits, when available on the vine, were taken for measuring fruit circularity in 2019 and 2020. Three fruits were cut in half lengthwise and placed flesh side up on a black background. From the remaining three fruits, an equatorial 5-mm slice (crosswise) was cut and also placed on black background. The outline of fruit was extracted from images using background thresholding from the OpenCV library. The circularity of the fruit outline was then measured as the proportion overlap between the area of the outline and the area of a circle that was the same total area as the outline and centred on the outline (1 = perfect circle). The trait properties were analysed using the R-package

“moments” v. 0.14.1 (R Core Team 2020; Komsta and Novomestky 2022).

DNA extraction and genotyping

Young leaf tissue was collected in spring, and DNA was extracted by Slipstream Automation (Slipstream Automation, Palmerston North, New Zealand). Final dsDNA concentration was standardised to a quantity of ~500 ng per sample and vacuum dried to the requirement of the high throughput targeted resequencing platform Flex-Seq® Ex-L of RAPiD Genomics (RAPiD Genomics Gainesville, FL, USA). Resulting sequence reads were mapped against the *A. chinensis* var. *chinensis* “Russell” reference genome (Tahir et al. 2022). Alignments were generated using BWA-MEM (Li 2013) and SAMtools (Danecek et al. 2021) using default parameters. SNP calling was performed in ANGSD with region selection based on target intervals (Korneliussen et al. 2014). Dosage estimation of tetraploid *A. arguta* x *A. arguta* population and SNP filtering were performed using the R-package “Updog” V2. Dosage genotypes were called for offspring and parental lines using an empirical Bayesian approach (Gerard et al. 2018). A further filtering of SNPs was performed for quality, allele bias ($0.5 < \text{bias} < 2$), over-dispersion ($\text{od} < 0.02$) and sequencing error ($\text{seq} < 0.01$) (R Core Team 2020; Tahir et al. 2020). Genotypes were called under the tetraploid ($4x$) assumption as 0 (AAAA), 1 (AAAB), 2 (AABB), 3 (ABBB) and 4(BBBB). For pseudo-diploid ($2x$) genotyping, all heterozygote genotypes were assumed to be one class and therefore recoded as 0 (AAAA = AA), 1 (AAAB, AABB, ABBB = AB) and 2 (BBBB = BB).

Linear mixed model and relationship matrices

A linear mixed model (LMM) was used to predict breeding values for a segregating population comprising two incomplete crossing designs:

$$y = \mu + Xb + Za + e$$

where y is a vector of phenotypic values of the analysed trait, μ is the overall population mean, b is a vector of fixed effect (multiple years of observation) with the incidence matrix X , a is the unobserved random effect of genotypes with $a \sim N(0, G\sigma_a^2)$ where σ_a^2

is the additive variance and Z the incidence matrix of genotypes and e is the random residual effect with $e \sim N(0, I\sigma_e^2)$.

Variance components and their standard errors were estimated using ASReml-R software in R (Gilmour et al. 2015; R Core Team 2020). ASReml-R uses restricted maximum-likelihood (REML) methodology, which can be applied to unbalanced crossing designs (Patterson and Thompson 1971). Narrow-sense heritability (h_{NS}^2) on an individual plant basis was estimated for each trait, considering the proportion of additive variance component and total variance component $\sigma_p^2 : h_{NS}^2 = \frac{\sigma_a^2}{\sigma_p^2}$ (Falconer and Mackay 1996).

We considered several different approaches for building the relationship matrices to estimate BLUPs. The effect of pedigree-based and marker-based relationship matrices and the effect of including ploidy-levels and double-reduction coefficients to build the variance–covariance matrices were compared. The R package “AGHmatrix” v. 2.0.4 (Amadeu et al. 2016; R Core Team 2020) was used to build all relationship matrices. The methodologies used in this study are summarised in Table 1

Model comparison and cross-validation

The plant population in this study can be divided into different levels of sub-populations. The core element is a total number of 842 female progeny with phenotype and genotype information used to estimate BLUPs, while 910 male progeny, 31 parents and 11 distantly related ancestors contribute only genotype information but no phenotype information. Because of the lack of developing fruit,

39 seedlings did not contribute any phenotypic information but were included in the genotyping process. Owing to the sex-linkage of fruit traits, only female progeny contributed phenotypic information to the BLUP estimation. Most of the parents used to develop the two factorials had been developed from previous controlled crosses, and pedigree information for each of these was available. The 13 females in the first factorial were selected for their own performance and crossed with two male selections from the germplasm at PFR Motueka. The second factorial comprised 13 male parents, previously selected from their seedling populations based on their overall family means and crossed with two commercial female cultivars (Supplementary Table 1).

An overview of breeding value prediction is provided in Fig. 1. A total of 1752 progeny of the two factorial crossing schemes was used. Phenotypic information of female progeny was used to predict breeding values (BLUPs) for both the parental generation and progeny generation under the assumption of a pedigree-based or marker-based relationship matrix. The LMM was validated by applying a tenfold cross-validation scheme to compute different validation criteria. For cross-validation, female progeny with phenotypic observations were assigned randomly into 10 groups. At each validation step, information from one group (validation set) was masked and predicted by the remaining groups (training set). The randomised grouping was repeated 10 times to eliminate structural occurrences in datasets and the population. Individuals without phenotypic observation records (parental and male individuals) were not included in the model validation method (Supplementary Fig. 1). Each group was used only once as a validation set, and the correlation of observation to prediction (predictive ability, PA), mean squared error (MSE), regression coefficient of observed phenotypes to their breeding value prediction (bias), variance components and expected genetic gain (EGG) were calculated. The genetic gain was estimated using the following equation: $\Delta G = \frac{1}{2}(PA * \sigma_a * i)/L$, where PA is the correlation between observed phenotype and prediction, σ_a is the square root of additive variance, i is the selection intensity and L is the length of breeding cycle. We set i and L equal to 1 to be consistent for all models. Because of dioecy, only female progeny were considered.

A LMM including all female progenies (full model) was used to compute the accuracy of breeding value prediction for all tested variance–covariance matrices. Female progeny with phenotypic observations were

Table 1 Model used to estimate breeding values of dioecious *Actinidia arguta* kiwiberry crop. (*) A double reduction coefficient (ω) of 0.01 was also chosen to implement a multivalent chromosome behaviour during meiosis. (**) A minor-allele frequency (maf) of 0.05 was chosen by the visual validation of normality of the residual

Relationship matrix	Model	Ploidy	Reference
Pedigree-based	A2	2x	(Henderson 1976)
	A4	4x	(Kerr et al. 2012)
	A4 ^o	4x*	(Kerr et al. 2012)
Marker-based	G2	2x**	(Yang et al. 2010)
	G4	4x	(VanRaden 2008) adapted by (Ashraf et al. 2016)

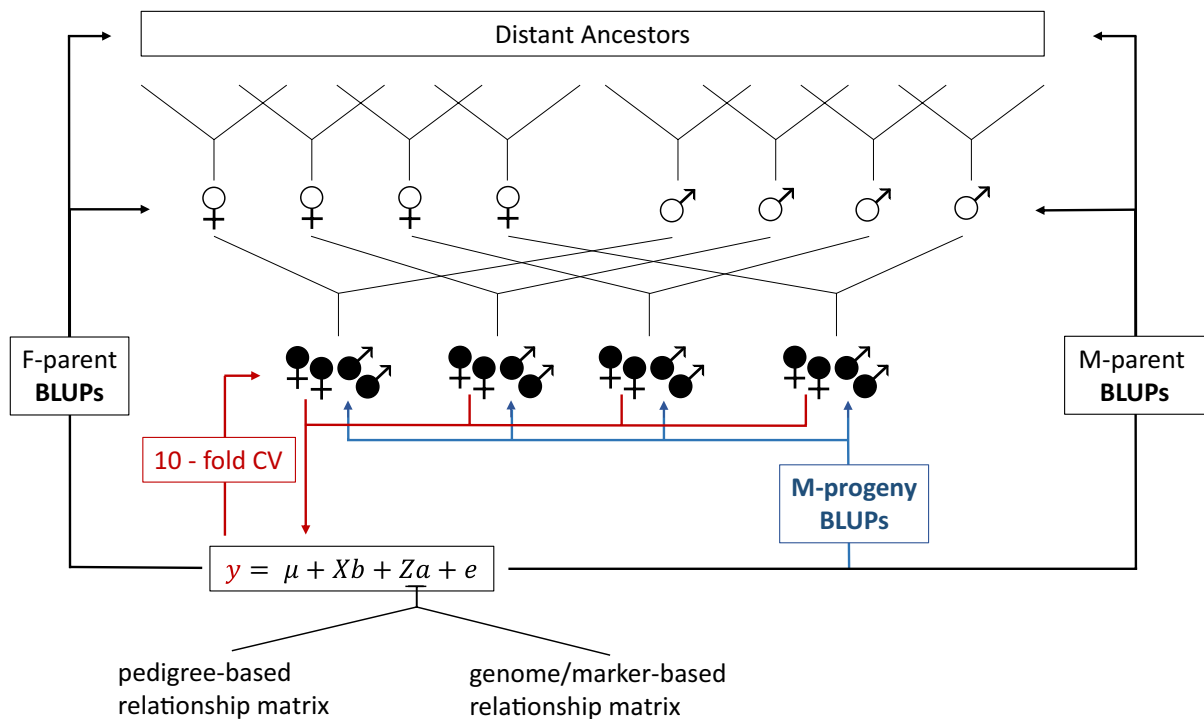


Fig. 1 Overview of breeding value predictions for female (F-) and male (M-) *Actinidia arguta* genotypes. From a previous population of 12 crosses (distant ancestors), 13 females, one unrelated additional female and 13 males were selected and used as parents within two incomplete factorial designs, represented by four females ♀ and four males ♂. From the two factorials, 842 female (♀) and 910 male (♂) progeny were

selected. Only female progeny were phenotyped and used as response variable (y) within the linear mixed model (LMM). The full model was used to predict best linear unbiased predictions (BLUPs) for female and male parents as well as male and female progeny. A tenfold cross-validation (tenfold CV) was set and repeated 10 times to validate the predictive ability for phenotyped female progenies

used to train the model, and BLUPs were estimated for all individuals included in the crossing scheme. Results were grouped into parental generation (with distant ancestors), individuals with observation (female progenies) and no-observations (male progenies).

The accuracy of BLUP estimation was calculated using the definition described by Henderson in 1975: $accuracy = \sqrt{1 - \frac{PEV}{\sigma_a^2 K_{ii}}}$, where PEV is the predicted error variance of the predicted error of breeding values of each individual and σ_a^2 is the additive variance and K_{ii} the diagonal element of variance co-variance matrix with $K_{ii} = 1 + F$, where F is the inbred coefficient of individual i . The calculation of accuracy requires the diagonal elements of the mixed model equation (LHS=left-hand-side), when calculating standard error (SEP):

$$\begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z + K^{-1}\lambda \end{pmatrix} \begin{pmatrix} \hat{b} \\ \hat{u} \end{pmatrix} = \begin{pmatrix} X'y \\ Z'y \end{pmatrix},$$

where K^{-1} is equal to the inverse of A^{-1} (pedigree-based) or G^{-1} (marker-based) relationship matrix and its inverse therefore $\lambda = \frac{\sigma_e^2}{\sigma_u^2}$ (shrinkage factor) and the coefficient matrix, (Henderson 1975 ; Mrode and Thompson 2014)

$$\begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z + K^{-1}\lambda \end{pmatrix} = \begin{pmatrix} C_{ii} & C_{ij} \\ C_{ji} & C_{jj} \end{pmatrix}.$$

Calculating PEV, the diagonal elements of inverse of the coefficient matrix are required, as shown by Henderson (1975):

$$\begin{pmatrix} C_{ii} & C_{ij} \\ C_{ji} & C_{jj} \end{pmatrix}^{-1} = \begin{pmatrix} C_{ii}^{ii} & C_{ij}^{ii} \\ C_{ji}^{ii} & C_{jj}^{ii} \end{pmatrix}$$

$$PEV = C^{dd} \sigma_e^2,$$

with diagonal element C^{dd} of the inverse coefficient matrix, or.

$$PEV_i = (d_i \sigma_e^2),$$

where d_i is the diagonal element of the inverse of LHS, and σ_e^2 is the residual variance. For every individual included in relationship matrix, a standard error is calculated (SEP) with the following:

$$SEP = \sqrt{\text{var}(a - \hat{a})} = \sqrt{(d_i \sigma_e^2)}$$

(Henderson 1975; Mrode and Thompson 2014; Gilmore et al. 2015; Isik et al. 2017). All models and scenarios were compared using Tukey's honestly significant difference (HSD) multiple comparison, considering independent runs of each "fold" as well as each iteration, implemented in the R-packages "stats" and "multcompView" v. 0.1–8 (Hothorn et al. 2008; R Core Team 2020). Visualisation of data analysis was performed using "ggplot2" V3.3.5, "ggbreak" v. 0.1.1 and "patchwork" v. 1.1.1 in R (Wickham 2016; Pedersen 2020; R Core Team 2020; Xu et al. 2021).

Results

We assessed five methods for calculating the relationship matrix and breeding values accuracy across different generations. All traits showed continuous distributions with a moderate skewness except for fruit load, which had a skewness that was very close to zero and therefore symmetric. Fruit circularity traits were moderately left-skewed (i.e. skewness values were negative), and fruit traits were fairly to moderately right-skewed (i.e. skewness values were positive) (Table 2).

A total of 1752 *A. arguta* progeny of 50 crosses were planted and managed under commercial breeding programme conditions for 5 years. Pedigree information across the population and 7259 (G4) or 2660 (G2) genome-wide distributed bi-allelic markers were available for analysing the effects of incorporating different relationship matrices.

For the G2 model, genotypes were classified in two homozygote classes and one heterozygote class under the assumption of re-calling genotypes from tetraploid dosage call to pseudo-diploid. Distribution of allele dosage classes under the assumption of $2x$ and $4x$ is shown in Supplementary Fig. 2.

The effect of relationship matrix on variance component estimation and estimated genetic gain

The genetic parameters of the full model, which includes all progeny with phenotypic information, and the mean over 10 iterations of the tenfold cross-validation model, are summarised in Supplementary Table 3. The impact of the relationship matrix on estimated variance components when employing the full model for all traits is shown in Fig. 2a–b. In the pedigree-based model, the additive variance (Fig. 2a) was consistently higher than that observed under the assumption of marker-based models, across all traits except for fruit weight. There was no difference in residual variance using the full model among the three pedigree-based models (Supplementary Table 3). No significant difference in additive variance between the diploid and tetraploid (pedigree-based) models was observed, except when 10% double reduction was included (Supplementary Table 3). Consequently, narrow-sense heritability, as the ratio of

Table 2 Overall trait properties for one *Actinidia arguta* kiwiberry vine trait and five fruit traits of female progeny. Trait properties were averaged over multiple years

Trait	Min	1 st Qu	Median	Mean	3 rd Qu	Max	Skewness	No. of years
Fruit load	0.50	3.00	5.00	4.46	6.00	9.00	0.08	2
Fruit weight (g)	1.00	6.30	7.73	7.98	9.34	19.73	0.60	3
Dry matter (%)	12.44	18.65	20.59	20.62	22.50	29.39	0.14	3
Ripe soluble solids content (Brix)	10.10	14.50	15.91	16.09	17.70	24.50	0.23	2
Fruit circularity (crosswise)	0.92	0.95	0.96	0.96	0.97	0.99	−0.42	2
Fruit circularity (lengthwise)	0.92	0.94	0.95	0.94	0.96	0.97	−0.40	2

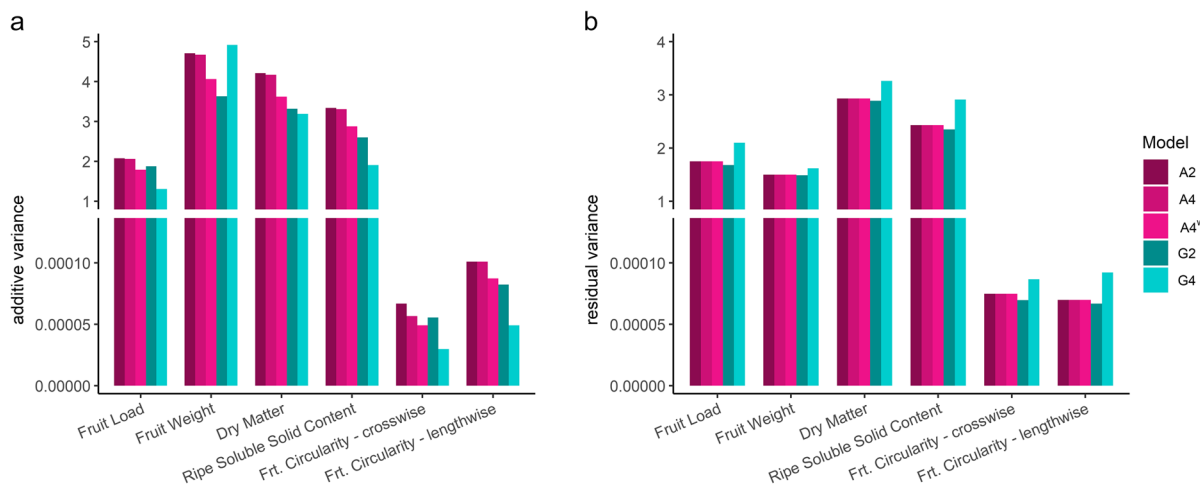


Fig. 2 Comparison of variance components. Variance components are compared for a single-vine trait and five fruit traits analysed with a linear mixed model (LMM), considering different relationship matrices, indicated by the letter (A and G),

ploidy (2 and 4). In **a**, the comparison of calculated additive variance is displayed, while in **b**, the comparison of residual variance is shown

additive to phenotypic variance, showed no significant difference between pedigree-based models under the assumption of disomic (A2) and tetrasomic (A4) inheritance for all traits compared to models including double reduction of marker effects (Supplementary Table 3). Between diploid and tetraploid marker-based methodologies, a significant difference in additive and residual variance was observed (Supplementary Table 3). When G2 was taken into account, the residual variance was lower for all traits, while it was higher considering G4, across all traits (Fig. 2b). In both models (pedigree-based and marker-based), additive variance was very low for both fruit-shape traits compared to fruit-load and fruit-quality traits.

The expected genetic gain (EGG) for each trait and year and overall for the average of multiple years is shown in Supplementary Table 3. Across all traits, no significant difference was observed for diploid and tetraploid probabilistic parametric models. Between the G2 and G4 models, only fruit dry matter content showed no significant differences of overall EGG but showed a significant difference of EGG in 2019. All other traits showed a difference between G2 and G4. The estimate of overall EGG tended to be lower when G4 was used, compared with G2 (Supplementary Table 3).

The effect of relationship matrix and ploidy level on the accuracy of BLUPs

We investigated the accuracy of predicted BLUPs for all six traits. BLUPs and the corresponding accuracy were estimated for all individuals with or without observations, incorporating different relationship matrix approaches into the LMM equation. The standard error of BLUP estimation, and therefore the accuracy of predicted breeding values of an individual, relies on the available information. Parental BLUPs, and therefore the accuracy of breeding value prediction, depend heavily on phenotypic records of progeny and relatives as well as the number of relatives. However, the accuracy of individuals within the progeny generation depends on the individual performance of those with phenotypic records, or on the family mean for individuals without phenotypic observations.

In this study, female progeny with observations and parents without phenotypic observations showed similar high accuracy of breeding value predictions. Within the parental generation, no significant differences were observed when using different pedigree-based relationship matrices. For all traits, the accuracy of prediction was significantly lower under the assumption of pseudo-diploidy of the marker-based relationship, whereas tetraploid

genetic marker methodology showed no differences from pedigree-based relationship methodologies (Fig. 3a). Including own phenotypic performance for all female progeny, marker-based relationship matrices significantly improved the estimation of accuracy, compared with pedigree-based methodologies (Fig. 3b). No difference was observed between A2 and A4, but including a double reduction coefficient in the LMM reduced the accuracy (Fig. 3b). The highest effect of realised relationship matrix (G4) on the accuracy of BLUP estimation was observed when individuals had no trait records (Fig. 3c–d). All relationship methodologies were compared for male progenies, which do not have trait records (Fig. 3c). The tetraploid G-matrix significantly improved the accuracy of breeding values. The results of male progeny population were compared with the results of the tenfold cross-validation approach, where observations were masked for females in the validation set (Fig. 3d). The sets performed almost identically.

Model validation and the effect of relationship methodology and ploidy

We investigated the correlation coefficient between mean observations over multiple years and predicted breeding values when observations were masked (validation set). An indicator of inflation/deflation of predicted breeding value variance was explored. A regression coefficient (β) of 1.0 (threshold line) indicates no differences in variance between observed phenotypes and predicted breeding values. In comparison with the pedigree-based LMM approach, A2, A4 and A4^w showed a mean regression coefficient close to $\beta=1.0$, indicating a similar variance among predicted breeding values and mean phenotypic observation (Supplementary Fig. 3a). No significant difference between pedigree-based models was observed. Under the assumption of pseudo-diploidy, a bias > 1 for all traits and a significant difference between G2 and other models were observed. Whereas the tetraploid model (G4) showed a bias less than 1.0 (threshold) and a significant difference from

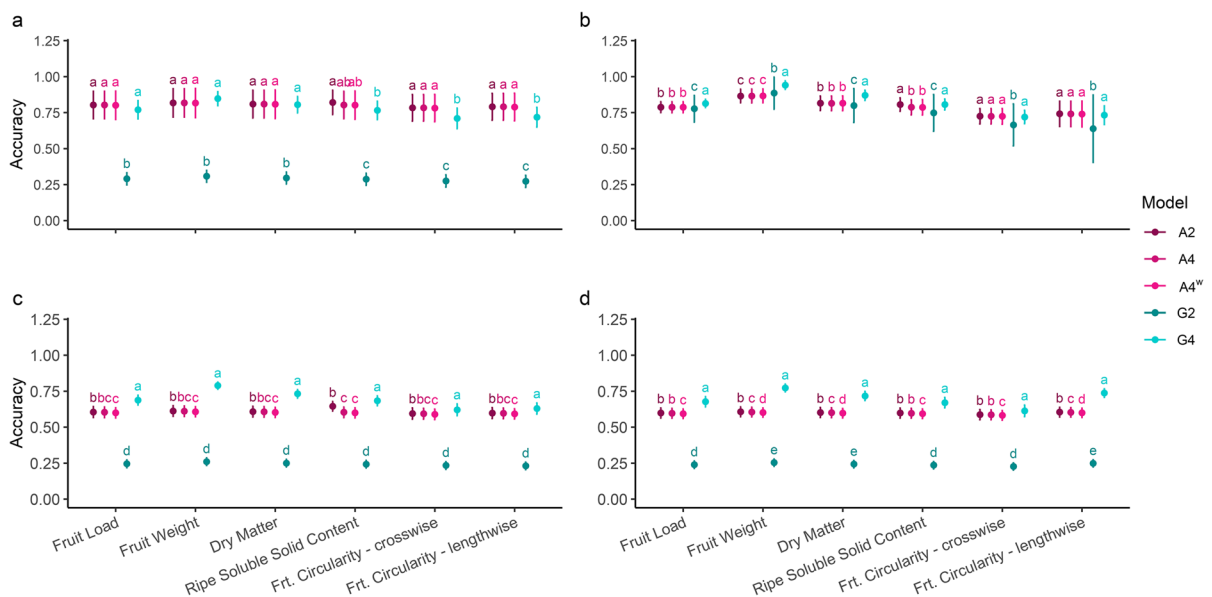


Fig. 3 The effects of different relationship matrices (A=pedigree-based and double reduction coefficient $w=0.1$, G=marker-based) and the effect of considering ploidy level (2 or 4) on the accuracy of best linear unbiased prediction (BLUP) estimation were compared for each *Actinidia arguta* kiwiberry trait and within different sub-populations. Mean accuracy is shown for **a** parental generation, **b** female progeny

and **c** male progeny. Plot groups (**a–c**) result from a full-model set; all females with observations are included. Plot group (**d**) is the result of validation sets from a 10×tenfold cross-validation methodology. Letters are from Tukey's HSD test; models with the same letter within a trait are not significantly different at a significance level of 0.05

other tested models; a higher variance of predicted breeding values was observed compared with the phenotypic observation (Supplementary Fig. 3a).

The correlation between observation and predicted breeding values (predictive ability) for all tested methodologies of calculating relationship matrices was obtained by computing the overall Pearson's correlation of each validation set. No difference in predictive ability was observed between pedigree-based and pseudo-diploid realised relationships based on the predicted abilities (*PA*) for all studied traits. *PA* of the tetraploid-realised relationship methodology (G4) varied depending on the trait. A significant difference between G4 and the pedigree-based approach was observed for fruit load, whereas no difference between G2 and G4 was observed for dry matter content, ripe soluble solids content, or fruit circularity (Supplementary Fig. 3b).

The quality of model prediction was measured by the mean-squared error (MSE) for each model approach and trait. In general, the MSE was higher under the G4 model assumption compared with other models. The only significant difference between G4 and the other studied models was observed for fruit weight. No significant difference was observed between all three pedigree-based models and between pedigree-based and G2 models (Supplementary Fig. 3c).

Discussion

LMMs to estimate best linear unbiased predictions (BLUPs) were first developed in animal breeding to estimate additive random individual effects and are now used in plant breeding. An improvement for predicted breeding values and accelerating genetic gain of economically important quantitative traits can be achieved by including the genetic information of individuals (Meuwissen et al. 2001). Genomic selection uses markers distributed across the whole genome to construct an additive relationship matrix directly from the genotypic information and the covariance relationship of breeding values between individuals (Calus 2010). Genomic-based relationships exploit not just genetic information between families but also differentiate the relationship between individuals within a family, whereas a pedigree-based relationship assumes an equal probabilistic relationship within a family through common ancestors. In this study, five

approaches to generating relationship matrices were applied to predict breeding values for six economically important, sex-linked traits in kiwiberry.

Validation variables (accuracy, bias, predictive ability and mean squared error) were estimated under the different relationship matrices (A=pedigree-based, G=marker-based) and accounted for ploidy (2=diploid and 4=tetraploid). Studies of double reduction during meiosis in natural and induced tetraploid *A. chinensis* populations showed a 10% multivalent chromosome formation during meiosis in induced tetraploids (Wu et al. 2014).

Evidence of marker segregation in autotetraploids showed that the rate of double reduction increases towards the telomere because multivalent and chromosome formations cross-over events are more likely. Nevertheless, double reduction is often ignored and this, therefore, may be one reason for the low rate of genetic improvement in polyploids compared with their diploid counterparts (Bourke et al. 2015; Amadeu et al. 2016). Testing the effect of double reduction in *A. arguta*, a second A4 model with a double reduction coefficient of 10% ($w=0.1$) was proposed in this study.

Source of information and relationship matrix methodology

Our study compared the effects of own performance and observation records of relatives on the breeding value predictions using different methodologies to build a relationship matrix. Breeding value is the estimated merit of genotypes, as parental breeding values are estimated by progeny performance; consequently, the accuracy of BLUP estimation is high (Fig. 3a). Since phenotyped individuals and pedigree relationships were the only sources of information to build this model, the accuracy of progeny with observations was high, regardless of which of the three matrices were used (Fig. 3b). Any progeny which lacked phenotypic observations did not contribute information to the prediction model. Breeding values of progeny without phenotypic records were estimated incorporating phenotypic information of siblings and relatives, which is obviously a less accurate BLUP estimation. The same pattern was also observed in the female progeny population when observations were masked (Fig. 3c–d). This finding suggests that the accuracy of estimation heavily relies on own observation records, and there is no sex-linked effect.

Through the use of markers across the whole genome, the genomic-based relationship distinguished the relationship of individuals within families by marker inheritance. Based on marker inheritance, breeding values can be estimated for individuals where no phenotypic information can be made. When genetic markers were used to build a realised relationship matrix, each individual's own phenotypic performance became less exclusive to predicting breeding values, compared with the linkage between markers and phenotypic observations. In individuals with just genotypic information, the marker-based relationship matrix allows individual breeding values to be estimated. Female progeny were used to train the model regardless of which model was used to predict breeding values; therefore, the accuracy of predicted breeding values for female progeny across all models equates to the accuracy of pedigree-based models (Fig. 3b). The accuracy of breeding value prediction of male progeny (Fig. 3c) and female progeny with masked phenotypic observations (Fig. 3d) is both highly dependent on their relationship to the training population. Between marker-based models, the G4 model significantly improved the accuracy because of the representation of five genotype classes. In contrast, when G2 was used, the heterozygous classes combined into one, resulting in a masked additive genetic effect and therefore less precise breeding value prediction.

Bias is a sufficient indicator of the shrinkage factor (λ), the proportion of residual variance to additive genetic variance. The factor lambda (λ) is shrinking the distribution of phenotypic observations towards the population mean, which results in a reduced variation in breeding values. A low shrinkage factor results in high variance of predicted breeding values compared with observed variance. Probabilistic relationship matrix-based models tend to have a bias value of around one, indicating similar variance of predicted breeding values and observation. Therefore, the model prediction is more robust for pedigree-based models. Our marker-based relationship matrix model showed significant differences from the pedigree-based models as well as between allele dosages (Supplementary Fig. 3a). This leads us to conclude that there was under- and overestimation of BLUPs compared with pedigree-based models.

There were limited differences in the predictive ability of probabilistic and realised relationship-based models. The correlation of predicted breeding values

and phenotype observations was positive for all traits in this research. Individuals (female progeny) within the validation set did not contribute phenotypic information for model development because their observations were masked. Therefore, BLUPs of individuals within the validation set were the predicted family mean, accounted for by phenotyped family members. This can lead to overestimation of BLUPs, whereas models considering realised relationships lead to more precise prediction. Our tested models showed a slightly lower predictive ability for marker-based models, suggesting improvement of the genotyping approach will improve the predictive ability and the mean squared error (Supplementary Fig. 3b–c).

Ploidy and double reduction coefficient

The effects of ploidy/allele dosage considering tetrasomic inheritance were studied for pedigree-based and marker-based model approaches. For all sub-populations, no significant differences in the accuracy of model breeding value accuracy, bias, predicted ability and mean squared error were observed between the various ploidy levels under the assumption of probabilistic relationship methodology. Using the kinship matrix to estimate the *A-matrix* was originally developed for population studies with varied ploidy levels (Kerr et al. 2012). With uniform ploidy levels, this study showed no significant differences when comparing the model criteria. It is only in polyploid populations where mixed ploidy occurs that consideration of ploidy in kinship estimation between individuals is necessary when a probabilistic-relationship is considered. Only a slightly significant difference of accuracy of prediction was observed here, including the complexity of double reduction, depending on the trait analysed.

Amadeu et al. (2016) showed that the effect of double reduction is cumulative for breeding populations with long histories and therefore more amenable to breeding value prediction. In populations with shallow pedigree histories like for *A. arguta*, the double reduction is less effective for the BLUP estimation, leading to overestimation of variance components.

The accuracy of parental BLUPs and those of other relatives, when no observations are made, depends on the relationship to individuals in the training set (Henderson 1975; Mrode and Thompson 2014). When heterozygote classes of G4 were re-scored to

G2, masked additive allele effects resulted, and therefore a significant difference in the prediction accuracy was observed (Fig. 3a, c–d). Within the training set, the prediction accuracy was reduced when observations were recorded, and a diploid marker-based relationship matrix was considered. This suggested a reduction of additive allele effect linked to phenotypic observations (Fig. 3b).

In our study, we observed no effect of ploidy or double reduction coefficient on the validation criteria (bias, predictive ability, mean squared error) for pedigree-based models, which suggests no significant difference in variance estimation. Considering allele dosage in the marker-based relationship, the variance comparison between phenotypic observations and predicted breeding values was significant. The G2 model showed a higher variability in observations than the predicted. On the other hand, G4 predicted a higher variability in BLUPs than observations (Supplementary Fig. 3a). This leads us to conclude that there was underestimated BLUP prediction using G2 and an overestimation when the G4 model was tested.

Limited differences were observed of predictive ability and mean squared error when ploidy or allele dosage were considered (Supplementary Fig. 3b–c). Gemenet et al. (2020) studied the effect of diploid, pseudo-diploid, tetraploid and hexaploid variant calling in potatoes and sweet potatoes. The authors showed that when diploidized genotype data are considered, it is more adequate to call genotype classes directly as diploid rather than re-diploidizing from high ploidy calls. We can confirm that pseudo-diploidizing, already called genotypes, is less reliable. In autopolyploids, estimating heterozygote genotype classes can be challenging. de Bem Oliveira et al. (2019) compared the influence of different relationship matrices originating from various genotype call data. The authors showed similar results in predictive ability when considering pseudo-diploid and tetraploid marker-based relationship matrices, with only minor differences observed. Due to the challenge of estimating heterozygote classes in autopolyploids, it can lead to misclassification and interfere with genomic selection, as shown in different studies (Grandke et al. 2016; Schmitz Carley et al. 2017; Bourke et al. 2018). An alternative genotyping approach was recommended, using continuous genotyping (de Bem Oliveira et al. 2019).

Our results of predictive ability contrast with the accuracy of breeding value prediction, which improved significantly when the tetrasomic inheritance of the marker-based relationship matrix was considered using the predicted error variance. All female progeny with phenotypic observations were grouped into a training and validation set (tenfold cross-validation). A consequence of grouping these small populations made the comparison of model validation less reliable, as suggested by Gurka and Edwards (2011). Further investigation using large breeding populations in dioecious crops is required.

In this study, we have shown the potential of using different variance co-variance relationship methodologies in *A. arguta* breeding programmes. Overall, the results of six traits considered in a marker-based relationship matrix showed a positive correlation of predictions to mean observations, indicating a better representative genetic architecture of genome-wide marker coverage using a multiplex PCR and new generation sequencing combination approach, compared with previous studies of other *Actinidia* species (Datson et al. 2017; Cheng et al. 2019). We were able to differentiate the effects of different relationship methodologies and ploidy to the best linear unbiased prediction in the parental generation and the progeny population, progeny both with and without phenotypic observation. Including the uncertainty of double reduction to the pedigree-based methodology had less effect on the accuracy of prediction. In the context of selecting genotypes within crosses when no phenotypic observations can be made, pedigree-based models have no power to distinguish variation. Marker-based models allow capturing variation between individuals within the same cross (Daetwyler et al. 2013; de Bem Oliveira et al. 2019). In our study, tetraploid marker-based models incorporating allele dosage significantly affected the predicted accuracy, especially in the progeny generation when no phenotypic observations were available, and these improvements were significant. This will reduce the breeding cycle by at least 3 years because no progeny testing of selected males is needed. The estimated 3 years mainly represent the time required for cross establishment before the first observations can be made. Further work including genotype by environmental interactions and non-additive effects could improve the genomic selection models (Endelman et al. 2018; Matias et al. 2019).

Acknowledgements We would like to thank A. Ross Ferguson, Margaret Carpenter, Edwige J. F. Souleyre, Linley K. Jesson and Sara Montanari for critical reading of the manuscript. In this study, ChatGPT, an AI language model developed by OpenAI, was used to refine the written content of this publication.

Author contribution All authors contributed to the study conception and design. PMD designed the experimental crossing design. Genotyping was performed by ST and JMcC. DTA and CHC contributed to data collection and analysis. The first draft of the manuscript was written by DM, ML, CMcK and SB, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions Funded through the Kiwifruit Royalty Investment Programme (KRIP) by the New Zealand Institute for Plant and Food Research Limited.

Materials availability The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Amadeu RR, Cellon C, Olmstead JW, Garcia AA, Resende MF, Muñoz PR (2016) AGHmatrix: R Package to construct relationship matrices for autotetraploid and diploid species: A blueberry example. *Plant Genome* 9(3). <https://doi.org/10.3835/plantgenome2016.01.0009>
- Ashraf BH, Byrne S, Fé D, Czaban A, Asp T, Pedersen MG, Lenk I, Roulund N, Didion T, Jensen CS, Jensen J, Janss LL (2016) Estimating genomic heritabilities at the level of family-pool samples of perennial ryegrass using genotyping-by-sequencing. *Theor Appl Genet* 129:45–52. <https://doi.org/10.1007/s00122-015-2607-9>
- Baduel P, Bray S, Vallejo-Marin M, Kolář F, Yant L (2018) The “polyploid hop”: shifting challenges and opportunities over the evolutionary lifespan of genome duplications. *Front Eco Evo* 6:117. <https://doi.org/10.3389/fevo.2018.00117>
- Bourke PM, Voorrips RE, Visser RGF, Maliepaard C (2015) The double-reduction landscape in tetraploid potato as revealed by a high-density linkage map. *Genetics* 201:853–863. <https://doi.org/10.1534/genetics.115.181008>
- Bourke PM, Voorrips RE, Visser RG, Maliepaard C (2018) Tools for genetic studies in experimental populations of polyploids. *Front Plant Sci* 9:513. <https://doi.org/10.3389/fpls.2018.00513>
- Bradshaw JE (2007) The canon of potato science: 4. Tetrasomic Inheritance Potato Res 50:219–222. <https://doi.org/10.1007/s11540-008-9041-1>
- Calus MPL (2010) Genomic breeding value prediction: methods and procedures. *Animal* 4:157–164. <https://doi.org/10.1017/S1751731109991352>
- Cheng C-H, Datson PM, Hilario E, Deng CH, Manako KI, McNeilage M, Bomert M, Hoeata K (2019) Genomic predictions in diploid *Actinidia chinensis* (kiwifruit). *Eur J Hort Sci.* 84(4):213–217. <https://doi.org/10.17660/eJHS.2019/84.4.3>
- Choudhary A, Wright L, Ponce O, Chen J, Prashar A, Sanchez-Moran E, Luo Z, Compton L (2020) Varietal variation and chromosome behaviour during meiosis in *Solanum tuberosum*. *Heredity* 125:212–226. <https://doi.org/10.1038/s41437-020-0328-6>
- Comai L (2005) The advantages and disadvantages of being polyploid. *Nat Rev Genet* 6:836–846. <https://doi.org/10.1038/nrg1711>
- Daetwyler HD, Calus MPL, Pong-Wong R, de los Campos G, Hickey JM, (2013) Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. *Genetics* 193(2):347–365. <https://doi.org/10.1534/genetics.112.143313>
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H (2021) Twelve years of SAMtools and BCFtools. *GigaScience* 10(2):giab008. <https://doi.org/10.1093/gigascience/giab008>
- Datson PM, Barron L, Manako KI, Deng CH, De Silva N, Bomert M, Cheng C-H, Crowhurst R, Hilario E (2017) The application of genome selection to kiwifruit breeding. *Acta Hort* 1172:273–278. <https://doi.org/10.17660/ActaHortic.2017.1172.52>
- de Bem Oliveira I, Resende MFR Jr, Ferrão LFF, Amadeu RR, Endelman JB, Kirst M, Coelho ASG, Munoz PR (2019) Genomic prediction of autotetraploids; influence of relationship matrices, allele dosage, and continuous genotyping calls in phenotype prediction. *G3 (bethesda)* 9(4):1189–1198
- Endelman JB, Carley CAS, Bethke PC, Coombs JJ, Clough ME, da Silva WL, De Jong WS, Douches DS, Frederick CM, Haynes KG, Holm DG, Miller JC, Muñoz PR, Navarro FM, Novy RG, Palta JP, Porter GA, Rak KT, Sathuvalli VR, Thompson AL, Yencho GC (2018) Genetic variance

- partitioning and genome-wide prediction with allele dosage information in autotetraploid potato. *Genetics* 209(1):77–87. <https://doi.org/10.1534/genetics.118.300685>
- Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Longman, Essex, UK
- Fenton GA, Kennedy MJ (1998) Rapid dry weight determination of kiwifruit pomace and apple pomace using an infrared drying technique. *N Z J Crop Hort Sci* 26(1):35–38. <https://doi.org/10.1080/01140671.1998.9514037>
- Fjellstrom RG, Beuselinck PR, Steiner JJ (2001) RFLP marker analysis supports tetrasomic inheritance in *Lotus corniculatus* L. *Theor Appl Genet* 102:718–725. <https://doi.org/10.1007/s001220051702>
- Gallais A (2003) Quantitative genetics and breeding methods in autopolyploid plants. INRA, Paris
- Grandke F, Singh P, Heuven HCM, de Haan JR, Metzler D (2016) Advantages of continuous genotype values over genotype classes for GWAS in higher polyploids: a comparative study in hexaploid chrysanthemum. *BMC Genomics* 17:672. <https://doi.org/10.1186/s12864-016-2926-5>
- Gemenet DC, Lindqvist-Kreuzer H, De Boeck B, da Silva PG, Mollinari M, Zeng Z-B, Craig Yench G, Campos H (2020) Sequencing depth and genotype quality: accuracy and breeding operation considerations for genomic selection applications in autopolyploid crops. *Theor Appl Genet* 133:3345–3363. <https://doi.org/10.1007/s00122-020-03673-2>
- Gerard D, Ferrão LFV, Garcia AAF, Stephens M (2018) Genotyping polyploids from messy sequencing data. *Genetics* 210(3):789–807. <https://doi.org/10.1534/genetics.118.301468>
- Gilmour AR, Gogel BJ, Cullis BR, Welham SJ, Thompson R (2015) ASReml user guide. Release 4.1. Structural specification. VSN international Ltd, Hemel Hempstead, HP1 1ES, UK. <https://www.vsnl.co.uk>. Accessed 8 June 2023
- Gurka MJ, Edwards LJ (2011) Mixed models. In: Rao CR, Miller JP, Rao DC (eds) Essential Statistical Methods for Medical Statistics. Amsterdam: Elsevier, pp. 146–173
- Haynes KG, Douches DS (1993) Estimation of the coefficient of double reduction in the cultivated tetraploid potato. *Theor Appl Genet* 85:857–862. <https://doi.org/10.1007/BF00225029>
- Henderson CR (1974) General flexibility of linear model techniques for sire evaluation. *J Dairy Sci* 57:963–972. [https://doi.org/10.3168/jds.S0022-0302\(74\)84993-3](https://doi.org/10.3168/jds.S0022-0302(74)84993-3)
- Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423–447. <https://doi.org/10.2307/2529430>
- Henderson CR (1976) A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* 32:69–83. <https://doi.org/10.2307/2529339>
- Hothorn T, Bretz F, Westfall P (2008) Simultaneous inference in general parametric models. *Biom J* 50:346–363. <https://doi.org/10.1002/bimj.200810425>
- Huang H, Ferguson AR (2007) Genetic resources of kiwifruit: domestication and breeding. *Hortic Rev* 33:1–121. <https://doi.org/10.1002/9780470168011.ch1>
- Isik F, Holland J, Maltecca C (2017) Genetic data analysis for plant and animal breeding. Springer International Publishing, Cham
- Kennedy BW, Schaeffer LR, Sorensen DA (1988) Genetic properties of animal models. *J Dairy Sci* 71:17–26. [https://doi.org/10.1016/S0022-0302\(88\)79975-0](https://doi.org/10.1016/S0022-0302(88)79975-0)
- Kennedy BW and Sorensen DA (1988) Properties of mixed-model methods for prediction of genetic merit. In: Weir BS, Eisen EJ, Goodman MM, Namkoog G (eds) Proceedings of the second international conference on quantitative genetics. Sinauer Associates, Inc., Sunderland, pp 91–103. <https://eurekamag.com/research/001/921/001921703.php>. Accessed 8 June 2023
- Kerr RJ, Li L, Tier B, Dutkowski GW, McRae TA (2012) Use of the numerator relationship matrix in genetic analysis of autopolyploid species. *Theor Appl Genet* 124:1271–1282. <https://doi.org/10.1007/s00122-012-1785-y>
- Komsta L and Novomestky F (2022) moments: moments, cumulants, skewness, kurtosis and related tests. R package version 0.14.1. <https://CRAN.R-project.org/package=moments>. Accessed 8 June 2023
- Korneliusson TS, Albrechtsen A, Nielsen R (2014) ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* 15:356. <https://doi.org/10.1186/s12859-014-0356-4>
- Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv: Genomics 1303.3997v2 [q-bio.GN]. <https://doi.org/10.48550/arXiv.1303.3997>
- Matias FI, Alves FC, Meireles KGX, Barrios SCL, do Valle CB, Endelman JB, Fritsche-Neto R (2019) On the accuracy of genomic prediction models considering multi-trait and allele dosage in *Urochloa* spp interspecific tetraploid hybrids. *Mol Breeding* 39:100. <https://doi.org/10.1007/s11032-019-1002-7>
- Mertten D, Tsang GK, Manako KI, McNeilage MA, Datson PM (2012) Meiotic chromosome pairing in *Actinidia chinensis* var. *deliciosa*. *Genetica* 140:455–462. <https://doi.org/10.1007/s10709-012-9693-2>
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157(4):1819–1829. <https://doi.org/10.1093/genetics/157.4.1819>
- Mrode RA and Thompson R (2014) Linear models for the prediction of animal breeding values. 3rd edn. CABI, Wallingford. <https://doi.org/10.1079/9781780643915.0000>
- Muthoni J, Kabira J, Shimelis H, Melis R (2015) Tetrasomic inheritance in cultivated potato and implications in conventional breeding. *Aust J of Crop Sci* 9(3):185–190.
- Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika* 58(3):545–554. <https://doi.org/10.2307/2334389>
- Pedersen TL (2020) patchwork: the composer of plots. R package version 1.1.1. <https://CRAN.R-project.org/package=patchwork>. Accessed 8 June 2023
- Qu L, Hancock JF, Whallon JH (1998) Evolution in an autopolyploid group displaying predominantly bivalent pairing at meiosis: genomic similarity of diploid *Vaccinium darrowii* and autotetraploid *V. corymbosum* (Ericaceae). *Am J Bot* 85:698–703. <https://doi.org/10.2307/2446540>
- R Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing,

- Vienna, Austria. <https://www.R-project.org>. Accessed 8 June 2023
- Rizet G (1945) Contribution a l' étude biologique et cytologique de l'*Actinidia chinensis*. C R Séances Soc Biol Paris 139:140–142
- Schmid R (1978) Reproductive anatomy of *Actinidia chinensis* (Actinidiaceae). Botanischer Jahrbücher Für Systematik, Pflanzengeschichte Und Pflanzengeographie 100:149–195
- Schmitz Carley CA, Coombs JJ, Douches DS, Bethke PC, Palta JP, Novy RG, Endelman JB (2017) Automated tetraploid genotype calling by hierarchical clustering. Theor Appl Genet 130:717–726. <https://doi.org/10.1007/s00122-016-2845-5>
- Sears ER (1976) Genetic control of chromosome pairing in wheat. Annu Rev Genet 10(1):31–51. <https://doi.org/10.1146/annurev.ge.10.120176.000335>
- Soltis DE, Soltis PS (1999) Polyploidy: recurrent formation and genome evolution. Trends Eco Evol 14(9):348–352. [https://doi.org/10.1016/S0169-5347\(99\)01638-9](https://doi.org/10.1016/S0169-5347(99)01638-9)
- Soltis DE, Soltis PS, Rieseberg LH (1993) Molecular data and the dynamic nature of polyploidy. Crit Rev Plant Sci 12(3):243–273. <https://doi.org/10.1080/07352689309701903>
- Soltis DE, Soltis PS, Tate JA (2004) Advances in the study of polyploidy since *Plant speciation*. New Phytol 161:173–191. <https://doi.org/10.1046/j.1469-8137.2003.00948.x>
- Soltis DE, Soltis PS, Schemske DW, Hancock JF, Thompson JN, Husband BC, Judd WS (2007) Autopolyploidy in angiosperms: have we grossly underestimated the number of species? Taxon 56(1):13–30. <https://doi.org/10.2307/25065732>
- Soltis PS, Marchant DB, Van de Peer Y, Soltis DE (2015) Polyploidy and genome evolution in plants. Curr Opin Genet Dev 35:119–125. <https://doi.org/10.1016/j.gde.2015.11.003>
- Tahir J, Brendolise C, Hoyte S, Lucas M, Thomson S, Hoeata K, McKenzie C, Wotton A, Funnell K, Morgan E, Hedderley D, Chagné D, Bourke PM, McCallum J, Gardiner SE, Gea L (2020) QTL mapping for resistance to cankers induced by *Pseudomonas syringae* pv. *actinidiae* (Psa) in a tetraploid *Actinidia chinensis* kiwifruit population. Pathogens 9(11):967
- Tahir J, Crowhurst R, Deroles S, Hilario E, Deng C, Schaffer R, Le Lievre L, Brendolise C, Chagné D, Gardiner SE, Knaebel M, Catanach A, McCallum J, Datson PM, Thomson S, Brownfield LR, Nardoza S, Pilkington SM (2022) First chromosome-scale assembly and deep floral-bud transcriptome of a male kiwifruit. Front Genet 13:852161. <https://doi.org/10.3389/fgene.2022.852161>
- Testolin R (2011) Kiwifruit breeding: from the phenotypic analysis of parents to the genomic estimation of their breeding value (GEBV). Acta Hort 913:123–130. <https://doi.org/10.17660/ActaHortic.2011.913.14>
- VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91(11):4414–4423. <https://doi.org/10.3168/jds.2007-0980>
- White J (1990) Pollen development in *Actinidia deliciosa* var. *deliciosa*: histochemistry of the microspore mother cell walls. Ann Bot 65(3):231–239. <https://doi.org/10.1093/oxfordjournals.aob.a087929>
- Wickham H (2016) ggplot2: elegant graphics for data analysis. Springer-Verlag, New York. <https://doi.org/10.1007/978-0-387-98141-3>
- Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH (2009) The frequency of polyploid speciation in vascular plants. Proc Natl Acad Sci USA 106(33):13875–13879. <https://doi.org/10.1073/pnas.081157510>
- Wu J-H, Datson PM, Manako KI, Murray BG (2014) Meiotic chromosome pairing behaviour of natural tetraploids and induced autotetraploids of *Actinidia chinensis*. Theor Appl Genet 127:549–557. <https://doi.org/10.1007/s00122-013-2238-y>
- Xu S, Chen M, Feng T, Zhan L, Zhou L, Yu G (2021) Use ggbreak to effectively utilize plotting space to deal with large datasets and outliers. Front Genet 12:774846. <https://doi.org/10.3389/fgene.2021.774846>
- Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42:565–569. <https://doi.org/10.1038/ng.608>

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