Genomic prediction of autotetraploids; influence of relationship matrices, allele dosage, and continuous genotyping calls in the phenotype prediction

Ivone de Bem Oliveira[*, †], Marcio F. R. Resende Jr.[‡], Luis Felipe V. Ferrão[*], Rodrigo R. Amadeu[*], Jeffrey B. Endelman[§], Matias Kirst[††], Alexandre S. G. Coelho[†], and Patricio R. Munoz[*]

[*]Blueberry Breeding and Genomics Lab, Horticultural Sciences Department, University of Florida, Gainesville, FL, USA, 32611

[†]Plant Genetics and Genomics Lab, Agronomy College, Federal University of Goiânia, GO, Brazil, 74690-900

[‡]Sweet Corn Genomics and Breeding, Horticultural Sciences Department, University of Florida, Gainesville, FL, USA, 32611

[§]Department of Horticulture, University of Wisconsin, Madison, WI, USA, 53706

[††]Forest Genomics Lab, School of Forestry Resources and Conservation, University of Florida, Gainesville, FL, USA, 32610

17 Running head: Genomic prediction in autopolyploids

18

19 **Keywords:** autopolyploid, allelic dosage, GS, relationship matrix, *Vaccinium*

20 **Corresponding author:**

21 Patricio Ricardo Muñoz

22 Blueberry Breeding & Genomics Lab

23 University of Florida

24 Address: 2550 Hull Road, P.O. Box 110690, Gainesville, FL, 32611

25 Phone: (352) 273-4837

26 Email: p.munoz@ufl.edu

ABSTRACT

Estimation of allele dosage in autopolyploids is challenging and current methods often result in the misclassification of genotypes. Here we propose and compare the use of next generation sequencing read depth as continuous parameterization for autotetraploid genomic prediction of breeding values, using blueberry (*Vaccinium corybosum* spp.) as a model. Additionally, we investigated the influence of different sources of information to build relationship matrices in phenotype prediction; no relationship, pedigree, and genomic information, considering either diploid or tetraploid parameterizations. A real breeding population composed of 1,847 individuals was phenotyped for eight yield and fruit quality traits over two years. Analyses were based on extensive pedigree (since 1908) and high-density marker data (86K markers). Our results show that marker-based matrices can yield significantly better prediction than pedigree for most of the traits, based on model fitting and expected genetic gain. Continuous genotypic based models performed as well as the current best models and presented a significantly better goodness-of-fit for all traits analyzed. This approach also reduces the computational time required for marker calling and avoids problems associated with misclassification of genotypic classes when assigning dosage in polyploid species. Accuracies are encouraging for application of genomic selection (GS) for blueberry breeding. Conservatively, GS could reduce the time for cultivar release by three years. GS could increase the genetic gain per cycle by 86% on average when compared to phenotypic selection, and 32% when compared with pedigree-based selection.

INTRODUCTION

Polyploidy events are not an exception in plants, as about 70% of Angiosperms and 95% of Pteridophytes underwent at least one polyploidization event (Soltis and Soltis 1999). Polyploids are normally grouped into two categories, autopolyploids and allopolyploids, but intermediate forms are also possible, such as segmental allopolyploids (Spoelhof *et al.* 2017). Thresholds for polyploid classification have been controversial, but following the general taxonomic definition, autopolyploids arise from within-species whole genome duplication, and allopolyploids arise from whole genome duplication prior to or after an inter-specific hybridization event (Soltis *et al.* 2007).

Because speciation via ploidy increase can generate new phenotypic variability, this phenomenon is considered a powerful evolutionary source (Hieter and Griffiths 1999; Soltis *et al.* 2016). Despite the important role of polyploidization in plant evolution, its effects on inheritance of many agronomic traits and population genetics are still poorly understood when compared with diploid species (Dufresne *et al.* 2014). This especially holds true for autopolyploids. The complex nature of autopolyploid genetics is due to the presence of genotypes with higher allele dosage than diploids, larger number of genotypic classes, possibility of multivalent pairing, and poor knowledge of chromosome behavior during meiosis (Slater *et al.* 2013; Dufresne *et al.* 2014; Mollinari *et al.* 2015).

The advent of high-throughput genotyping methods, associated with the development of genetic and statistical analysis tools, has generated significant genetic gains for diploid species (Desta and Ortiz 2014). However, the application of genomic information to polyploid crops remains a challenge (Comai *et al.* 2005; Grandke *et al.* 2016). Although methods for the analysis and interpretation of genetic data in polyploids have recently been described (see review in Bourke *et al.* 2018), much development is needed, especially for new breeding approaches, such as genomic selection.

Genomic selection (GS) is a method to increase the efficiency and accelerate the selection process in breeding programs. GS is used to capture the simultaneous effects of molecular markers distributed across the genome, based in the premise that linkage disequilibrium between causal polymorphisms and markers allow the prediction of phenotypes based on the genotypic values (Meuwissen *et al.* 2001; Zhang *et al.* 2011; de los Campos *et al.* 2013). The first GS studies addressing polyploids considered diploid genetic models to circumvent the complexity involved in accurately defining allelic dosage (i.e., the number of copies of each allele at a given polymorphic locus). Promising results have been reported for polyploids (e.g. Gouy *et al.* 2013; Annicchiarico *et al.* 2015; Ashraf *et al.* 2016), however simplified assumptions were mostly used for genetic and statistical inferences (Garcia *et al.* 2013). Only a few studies have added different factors accounting for polyploid effects (*e.g.*, Slater *et al.* 2016; Sverrisdottir *el al.* 2017). Thus, more appropriate methods for GS in polyploids could be evaluated, possibly improving trait prediction.

Polyploidy can affect phenotypes through allelic dosage (additive effect of multiple copies of the same alleles), or by creating more complex interactions between loci or alleles, such as dominance or epistasis (Osborn *et al.* 2003). Thus, the inclusion of allelic dosage information may improve GS results (*e.g.*, better fit, increase of accuracy) by creating a more realistic representation of the effects of each genotypic class. Although the evidence of dosage effects in the expression of important economic traits exists (Guo *et al.* 1996; Birchler *et al.* 2001; Adams *et al.* 2003; Osborn *et al.* 2003), few studies linking dosage effects to phenotype prediction have been reported in autopolyploid species (*e.g.*; Slater *et al.*, 2016; Sverrisdottir *el al.* 2017; Nyine *et al.* 2018; Endelman *et al.* 2018). It is interesting to note that genotype classification is one of the major challenges for polyploids. Studies about genotyping calling evaluation for autopolyploids with next generation sequencing (NGS) data

showed that none of the existing methods performs properly (Grandke *et al.* 2016), unless high sequencing coverage (60-80*x*) is used (Uitdewilligen *et al.* 2013).

Here we compare a novel approach to GS in the context of autopolyploid, using *Vaccinium corymbosum* (southern highbush blueberry, SHB) as a model. The cultivated SHB is an autotetraploid, presenting 2n = 4X = 48 chromosomes (Lyrene *et al.* 2002). Inbreeding depression is strong in SHB and population improvements have been achieved by long-term recurrent phenotypic selection alongside with long testing phase and slow genetic gain per generation (Lyrene 2008). Our goal was to investigate and compare the influence of different relationship matrices that consider different ploidy information on phenotype prediction, using novel genotyping approaches based on next-generation sequencing.

MATERIAL AND METHODS

Population and phenotyping

The population used in this study encompasses one cycle of the University of Florida blueberry breeding program's recurrent selection, comprising 1,847 SHB individuals. This population was originated from 124 biparental controlled crosses, from 146 parents that presented superior phenotypic performance (cultivars and advanced stage of breeding). Phenotypic data of eight yield and fruit quality-related traits were collected during two production seasons (2014 and 2015), when the plants were 2.5 and 3.5 years of age. Yield (rated using a 1-5 scale), weight (g), firmness (g mm^{-1} of compression force), scar diameter (mm), fruit diameter (mm), flower bud density (reported as buds per 20 cm of shoot), soluble solids content (^{o}Brix), and pH were evaluated. The last three traits were phenotyped only in one year – soluble solids content and pH were phenotyped in 2014 and flower buds in 2015.

Five berries (fully mature and presenting picking quality) were randomly sampled to compose the measurement of fruit traits for each individual. Fruit weight was measured using

126 an analytical scale (CP2202S, Sartorious Corp., Bohemia, NY). The FirmTech II firmness
127 tester (BioWorks Inc., Wamego, KS) was used to measure fruit diameter and firmness. The
128 scar diameter was obtained by image analysis of the fruits using FIJI software (Schindelin *et*
129 *al.* 2012). The number of flower buds was counted in the main cane upright shoot, in the top
130 20 cm. A digital pocket refractometer (Atago, U.S.A., Inc., Bellevue, WA) was used to obtain
131 soluble solids measures from 300µl of berry juice. The pH was measured using a glass pH
132 electrode (Mettler-Toldeo, Inc., Schwerzenbach, Switzerland). More details are provided by
133 Amadeu *et al.* (2016), Cellon *et al.* (2018), and Ferrão *et al.* (2018).
134
135 **Genotyping**
136 Genomic DNA was extracted and genotyped using sequence capture by Rapid
137 Genomics (Gainesville, FL, USA). Polymorphisms were genotyped in genomic regions
138 captured by 31,063 120-mer biotinylated probes, designed based on the 2013 blueberry draft
139 genome sequence (Bian *et al.* 2014; Gupta *et al.* 2015). Sequencing was performed in the
140 Illumina HiSeq2000 platform using 100 cycle paired-end sequencing. After trimming (quality
141 score of 20), demultiplexing, and removing barcodes, reads were aligned to the draft genome
142 using Mosaik v.2.2.3 (Lee *et al.* 2014). Genotypes were called using FreeBayes v.1.0.1
143 (Garrison and Marth 2012) considering the diploid and tetraploid options. Single-nucleotide
144 polymorphisms (SNPs) were filtered considering i) minimum sequencing depth of 40
145 (average depth for the population); ii) minimum SNP quality score (QUAL) of 10; iii) only
146 biallelic markers; iv) maximum population missing data of 0.5; and v) minor population
147 allele frequency of 0.05. After filtering a total of 85,973 SNP were used in the GS analysis.
148 Further information regarding population composition and genotyping approach were
149 described in Ferrão *et al.* (2018). The genotypes for the diploid calling were coded as *0* (*AA*),
150 *1* (*AB*), or *2* (*BB*). For the tetraploid parameterization they were coded as *0* (*AAAA*), *1*

(*AAAB*), *2* (*AABB*), *3* (*ABBB*), and *4* (*BBBB*). A third parameterization (assumption-free method) was used, which considered allele ratio $\#A/(\#A + \#a)$, where $\#A$ is the allele count (sequencing depth) of the alternative allele and $\#a$ is the allele count of the reference allele. No dosage calling was performed in this model (File S1); these data varied continuously between *0* and *1*.

Population genetics analysis

In order to compare the information captured by each genomic-based relationship matrix, we performed linkage disequilibrium (LD), and principal components (PC) analyses. Pearson correlation tests (r^2) were performed for pairwise LD estimation among SNPs within scaffolds, considering draft reference genomes (Bian *et al.* 2014; Gupta *et al*, 2015). One SNP was randomly sampled per probe interval, and a total of 22,914 SNP were used in the analysis. LD was obtained for all marker-based scenarios: i) diploid (*G2*); ii) tetraploid (*G4*) and iii) ratio (*i.e.*, continuous genotypes; *Gr*). The LD decay over physical distance was determined as the mean distance at the LD threshold of $r^2 = 0.2$. To compare the LD among scenarios, the mean distances (Kb) and their interval confidences at $r^2 = 0.2$ were compared. The diversity captured from each relationship matrix was compared by PC using the R package adegenet v. 1.3-1 (Jombart and Ahmed 2011).

In order to compare the information present in the marker matrices, we also evaluated the observed heterozygosity in the population. For this, we obtained the ratio between the number of heterozygote genotypes and the total number of individuals. To estimate the heterozygosity for the continuous genotypes, empirical limits were established based on the mean and standard deviations presented for homozygotes classes of the tetraploid parameterization.

Models

One-step single-trait Bayesian linear mixed models were used to predict breeding values for each individual in the population, as follows:

$$\bar{y} = \mu + Xb + Z_1 c + Z_2 r + Z_3 a + Z_4 bxa + e \quad (1)$$

Where $\bar{y}$ is a vector of the phenotypic values of the trait being analyzed, μ is the population's overall mean, b is the fixed effect of year, c is the random effect of ith column position in the field ~ N (0, $I\sigma_c^2$), r is the random effect of the ith row position in the field ~ N (0, $I\sigma_r^2$), a is the random effect of genotype ~ N (0,$G_a\sigma_a^2$), where G_a was replaced by the different additive relationship matrices as described in the next section. The bxa is the random effect of the year by genotype interaction ~ N (0, $I\sigma_{bxa}^2$), and e is the random residual effect ~ N (0, $I\sigma_e^2$). Row and column effects were considered nested within year only for the traits evaluated in two years. For traits measured a single year, the same equation (1) was used without the year and the year by genotype interactions. The variance components for each random variable were: additive (σ_a^2), column (σ_c^2), row (σ_r^2), year-by-genotype interaction (σ_{bxa}^2), and residual (σ_e^2). X, Z_1, Z_2, Z_3, and Z_4 were incidence matrices for year, column, row, genotype, and year by genotype interaction, respectively. The narrow-sense heritabilities were estimated considering the ratio between the additive variance component and the total phenotypic variance (sum of all variance components).

Relationship matrices

To quantify the effect of the genetic information used to build the relationship matrices on the predictive ability (PA), we performed analyses considering different approaches to modeling the genotypic values in autotetraploid species (Table 1, File S1). The factors tested were: i) the source of information used to build the relationship matrix

200 (pedigree, genomic, or no relationship information); and ii) ploidy information (diploid,
201 tetraploid, and assumption-free method).
202
203 **Table 1.** Methods and assumptions used to compare the influence of relationship matrices,
204 ploidy and continuous genotypes in the prediction of breeding values for blueberry.

Relationship matrix	Model	Ploidy assumption	Methodology
Identity	*I*	none	none
Pedigree-based	*A2*	2	Henderson (1976)
	A4	4	Kerr *et al.* (2012)
Maker-based	*G2*	2	VanRaden (2008)
	G4	4	
	Gr	none	

205
206 The methods chosen to obtain the relationship matrices are shown in the Table 1. The
207 pedigree-based relationship matrices (*A*) were built considering a diploid model (Henderson
208 1976) and autotetraploid model without double-reduction (Kerr *et al.* 2012). The marker-
209 based relationship matrices (*G*) were based on the incidence matrices of markers effects (*X*)
210 according to VanRaden (2008) and adapted by Ashraf *et al.* (2016). Different assumptions
211 can be made regarding the marker allele dosage in autotetraploids (Table 2). We built the *X*
212 matrices under three assumptions regarding the additive marker allele dosage effect: *i)* a
213 pseudo-diploid model, where all the heterozygous genotypes are assumed as one class,
214 corresponding to a unique effect (data coded as *0*, *1*, and *2*); *ii)* an additive autotetraploid
215 model, where each genotype had a specific value, and cumulative additive effect is assumed
216 (data coded as *0*, *1*, *2*, *3*, and *4*); and *iii)* an assumption-free method based on the ratio of
217 reads count for the alternative and reference alleles (continuous parameterization, assuming
218 values between *0* and *1*), where also a cumulative additive effect is assumed. For the
219 construction of the relationship matrices based on marker data, the missing genotypes were
220 substituted by the mean. The R package AGHmatrix v. 0.0.3003 (Amadeu *et al.* 2016) was
221 used to obtain all relationship matrices.

222 **Table 2.** Genotype codes for marker-allele dosage effects with different assumptions.
223 Adapted from Slater *et al.* (2016).

Genotype	Pseudo-Diploid	Autotetraploid	Continuous values*
AAAA	0	0	
AAAB	1	1	
AABB	1	2	0 - 1
ABBB	1	3	
BBBB	2	4	

* Continuous values with a ploidy assumption-free parameterization

224

225 **Model implementation**

226 The six models described above (Table 1) were fitted using the R package (R Core
227 Team 2018) BGLR v. 1.0.5. (de los Campos and Pérez-Rodríguez 2016). The predictions
228 were based on 30,000 iterations of the Gibbs sampler, in which 5,000 were taken as burn-in,
229 and a thinning of five. The number of iterations, burn-in, and thinning interval parameters
230 were evaluated to define the final values used in the analysis (Figure S1). A single step
231 regression approach was applied to perform all phenotypic BLUP (I matrix), pedigree-BLUP
232 (P-BLUP), and genomic-BLUP (G-BLUP). Default hyper-parameters were previously
233 described (Perez and de los Campos 2014).

234

235 **Validation and model comparison**

236 For each trait, models were compared based on their PA, stability (mean square
237 errors), goodness-of-fit, and expected genetic gain. A 10-fold cross validation scheme was
238 applied to compute model PA. Because each validation group might have a different mean
239 (Resende *et al.* 2012b), the phenotypic PA were obtained as the Pearson correlation
240 coefficient between the empirical best linear unbiased estimation values (eBLUEs) obtained
241 by considering all the variables in the equations 1 as fixed (*i.e.*, Least Square means
242 estimations; LSMeans) and the cross-validated breeding values (BV) predicted by the models
243 for each validation fold. The goodness-of-fit for the different models was evaluated with

244 measures of the posterior mean of the log likelihood obtained from the full data set. The
245 model with the lowest value for this parameter defined the best fit for the data. For the
246 expected genetic gain estimation we used the following formula: $\Delta G=(PA \cdot \sigma_a \cdot i)/L$, where
247 *PA* is the phenotypic predictive ability, σ_a is the square root of additive genetic variance in
248 the population, *i* is the selection intensity, and *L* is the breeding cycle length. The selection
249 intensity (*i*) was considered constant for all methods.
250 Phenotypic and genotypic data used for diploid and tetraploid parameterizations are
251 available from Dyrad Digital Repository (accession number doi: 10.5061/dryad.kd4jq6h).
252 Data for continuous parameterization is available for review upon request. Data will be
253 available at Dyrad Digital Repository. The authors affirm that all data necessary for
254 confirming the conclusions of the article are present within the article, figures, and tables.
255
256 **RESULTS**
257 **Population genetics analysis**
258 Linkage disequilibrium decayed below $r^2 = 0.2$ at distances of 88.3 Kb, 92.6 Kb, and
259 98.2 Kb for the diploid, tetraploid and continuous models, respectively (Figure 1A-C). No
260 significant difference was observed considering the confidence interval for the mean distance
261 (Kb) at $r^2 = 0.2$ among different ploidies and continuous genotyping scenarios (Figure S2).
262 Similarly, no major differences were found between parameterizations within
263 methodology (*i.e.*, pedigree-based or marker-based methods) in the PC analysis (Figure S3).
264 The first two PC components of the marker-based (*G*) matrices were consistent across all
265 matrices, explaining approximately 20% of the variation. For example, *G2* matrix captured
266 20.60% of the variation, while *G4* captured 21.71%, and *Gr* captured 23.36% (Figure S3 A-
267 C). The PC results were consistent between pedigree methodologies as well. Approximately
268 38% of the variation was explained (*i.e.*, 37.74% of the variability was explained for the *A2*

matrix and 37.86% was explained for the *A4* matrix, Figure S3 D-E). The results obtained in the PC analysis did not justify a stratified sampling of cross-validation populations, since no evidence of sub-population structure was detected for any of the relationship matrices.

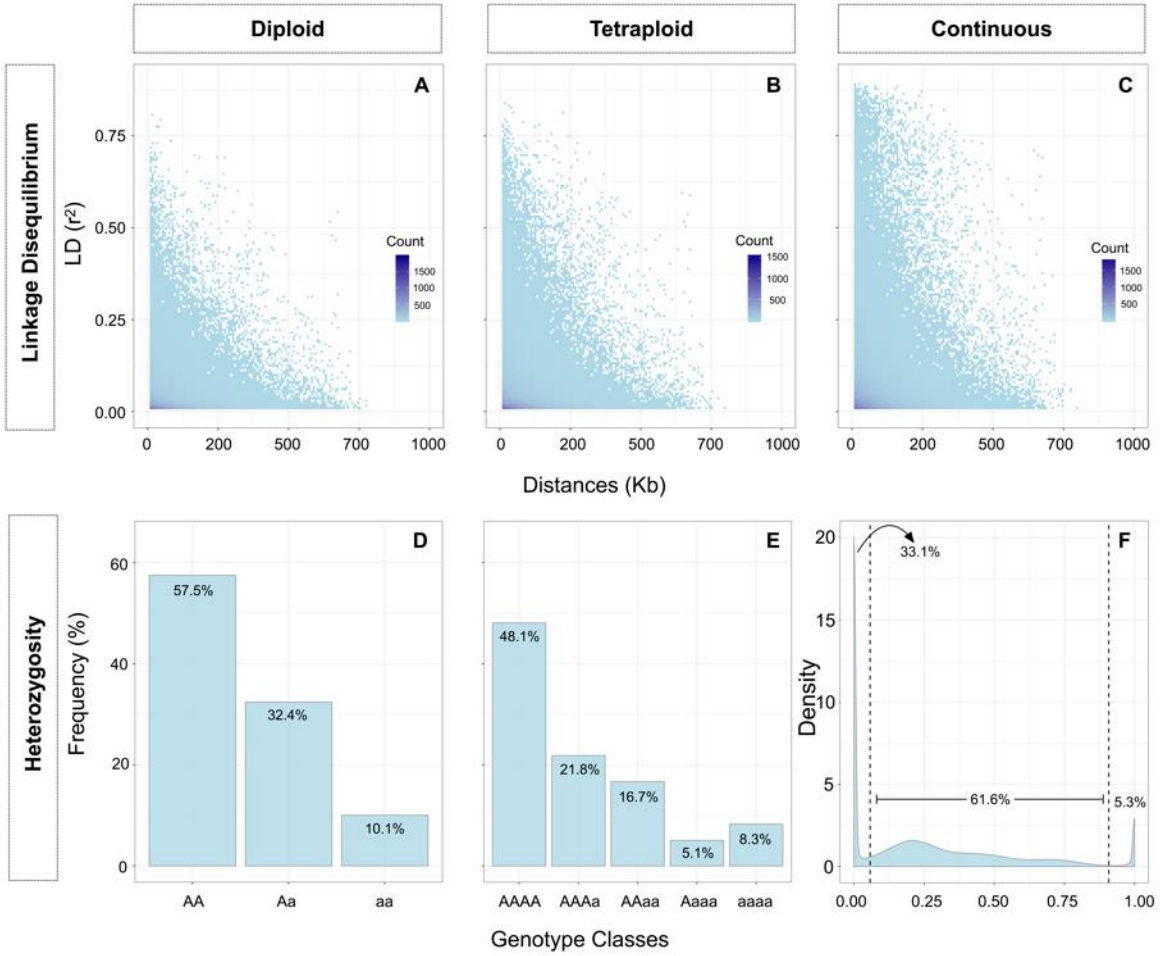

Figure 1. Linkage disequilibrium decay and heterozygocity for blueberry. Linkage disequilibrium decay estimation using one marker per probe, within scaffolds for (**A**) diploid, (**B**) tetraploid and (**C**) continuous genotype parameterizations. Heterozygosity observed in (**D**) diploid, (**E**) tetraploid, and (**F**) heterozygosity empirically established for the continuous genotypes scenario, assuming the limits of $0.058 \leq X \leq 0.908$.

Considering the heterozygosity observed in each scenario, genotypes assumed as homozygotes in the diploid parameterization were classified as one of the possible heterozygote classes in the tetraploid and in the assumption-free parameterizations (Figure 1D-F). As a result of this process, the tetraploid parameterization presented 37.50% more heterozygotes than the diploid parameterization. Considering the empirical thresholds established to compare the proportion of "heterozygotes" in the continuous genotypes with

the ploidy parameterizations, values equal to or below 0.058 and equal to or above 0.908 were considered as "homozygotes" classes (dashed lines, Figure 1F). With this, 61.59% of the genotypes were considered "heterozygotes", thus the continuous method would have presented 89.92% and 41.23% more heterozygotes than the diploid and the tetraploid parameterization, respectively. Nevertheless, some misclassification of data into classes in the diploid and tetraploid parameterization might have occurred (Figure 2A-B).

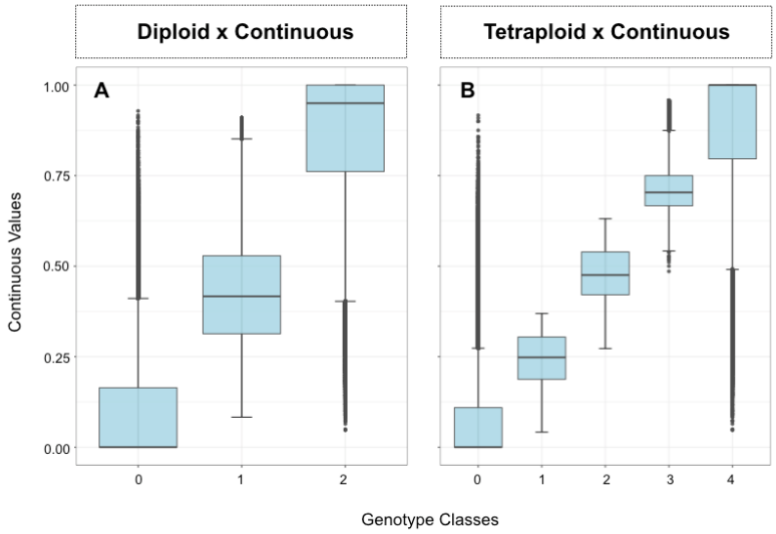

Figure 2. Relationship of the continuous values considering the classes assumed in the (**A**) diploid and (**B**) tetraploid parameterizations.

Variance estimates

The posterior means of the genetic parameters are summarized in Table 3. All the traits presented additive genetic variance significantly higher than zero. A wide range of variance was observed within a given parameter for the different methodologies, and most of the values were significantly different from each other (considering Tukey test results; Table 3, Table S1). Marker-based methodologies generated significantly smaller estimations for variance components when compared with pedigree-based estimations. Within marker-based methodologies, the assumption-free parameterization generated significantly smaller estimations. The effects of the difference in the estimation of variance components are

reflected in the estimated heritabilities – smaller values were estimated for marker-based methodologies. The lowest heritability was obtained for soluble solids, flower buds, and pH. Considering all methods, narrow-sense heritability values varied between 0.152 and 0.574, for flower buds and fruit weight, respectively.

Table 3. Genetic parameters estimated for eight yield and fruit-related traits analyzed with six linear mixed models, considering the use of ploidy information and continuous genotypes. Source of information, and dosage parameterizations for the relationship matrices indicated by the letters (*I*, *A*, or *G*), and numbers (2 or 4), respectively*

Trait	Relationship matrix	Additive Variance	Residual Variance	Heritability	Goodness-of-fit[1]	EGG[2] 2014	EGG[2] 2015
Soluble Solid (°Brix)	*I*	0.806 b	1.794 d	0.257 a	-2976.31 a	0.018 b	-
	A2	0.777 c	2.129 b	0.239 b	-3149.85 c	0.021 ab	-
	A4	0.764 c	2.125 b	0.236 b	-3148.067 c	0.021 ab	-
	G2	0.848 a	2.026 c	0.262 a	-3106.762 b	0.028 a	-
	G4	0.673 d	2.109 b	0.215 c	-3142.493 c	0.026 a	-
	Gr	0.546 e	2.241 a	0.174 d	-3195.892 d	0.022 ab	-
Flower Buds	*I*	2.133 a	4.752 d	0.270 a	-4024.785 a	-	0.018 a
	A2	1.247 cd	6.080 a	0.153 de	-4275.597 d	-	0.019 a
	A4	1.232 d	6.070 a	0.152 e	-4274.008 d	-	0.018 a
	G2	2.106 a	5.562 c	0.251 b	-4192.561 b	-	0.030 a
	G4	1.526 b	5.881 b	0.188 c	-4244.712 c	-	0.025 a
	Gr	1.315 c	6.115 a	0.161 d	-4281.097 d	-	0.023 a
Fruit Diameter	*I*	2.236 f	6.804 b	0.162 f	-8142.071 a	0.047 b	0.041 c
	A2	3.647 a	6.854 b	0.250 a	-8147.133 a	0.063 b	0.054 bc
	A4	3.581 b	6.825 b	0.247 b	-8141.58 a	0.061 b	0.054 bc
	G2	3.428 c	6.799 b	0.242 c	-8139.628 a	0.088 a	0.079 a
	G4	2.992 d	6.954 ab	0.216 d	-8178.698 ab	0.083 a	0.072 ab
	Gr	2.910 e	7.219 a	0.207 e	-8243.89 b	0.082 a	0.071 ab
Fruit Firmness	*I*	509.180 f	737.735 b	0.275 f	-16181.026 a	0.567 c	0.798 c
	A2	806.908 a	741.089 b	0.401 a	-16183.975 a	0.881 b	1.16 b
	A4	786.601 b	742.547 b	0.395 b	-16189.39 a	0.877 b	1.135 b
	G2	725.192 c	734.332 b	0.376 c	-16171.694 a	1.243 a	1.511 a
	G4	659.584 e	749.865 b	0.351 e	-16207.163 a	1.217 a	1.446 a
	Gr	687.685 d	783.729 a	0.354 d	-16283.119 b	1.257 a	1.490 a
pH	*I*	0.053 a	0.118 d	0.253 a	-592.837 a	0.005 a	-
	A2	0.052 a	0.140 c	0.241 b	-764.731 b	0.006 a	-
	A4	0.052 a	0.140 c	0.238 b	-762.886 b	0.005 a	-
	G2	0.052 a	0.141 c	0.241 b	-769.534 b	0.007 a	-
	G4	0.040 b	0.147 b	0.191 c	-805.822 c	0.006 a	-
	Gr	0.035 c	0.153 a	0.165 d	-840.501 d	0.006 a	-

Fruit Scar	*I*	0.086 f	0.073 d	0.381 f	-351.218 a	0.008 c	0.009 c
	A2	0.139 a	0.075 c	0.528 a	-395.541 ab	0.013 b	0.014 b
	A4	0.135 b	0.075 bc	0.522 b	-414.241 bc	0.013 b	0.014 b
	G2	0.123 d	0.075 cd	0.500 c	-392.265 ab	0.018 a	0.018 a
	G4	0.115 e	0.077 b	0.479 e	-451.945 c	0.018 a	0.017 a
	Gr	0.126 c	0.081 a	0.494 d	-538.444 d	0.019 a	0.018 a
Fruit Weight	*I*	0.217 f	0.214 b	0.374 f	-2215.277 b	0.013 c	0.014 c
	A2	0.403 a	0.207 c	0.574 a	-2157.103 a	0.021 b	0.021 b
	A4	0.393 b	0.205 c	0.568 b	-2136.886 a	0.021 b	0.021 b
	G2	0.344 d	0.206 c	0.535 c	-2152.251 a	0.030 a	0.029 a
	G4	0.323 e	0.215 b	0.513 e	-2220.552 b	0.029 a	0.027 a
	Gr	0.352 c	0.231 a	0.522 d	-2349.684 c	0.030 a	0.028 a
Yield	*I*	0.326 f	0.444 bc	0.310 f	-3683.364 a	0.012 b	0.015 c
	A2	0.549 a	0.442 bc	0.447 a	-3668.232 a	0.019 a	0.022 b
	A4	0.536 b	0.442 bc	0.441 b	-3667.668 a	0.020 a	0.021 b
	G2	0.470 c	0.441 c	0.407 c	-3662.062 a	0.026 a	0.030 a
	G4	0.421 d	0.458 b	0.374 d	-3730.163 a	0.024 a	0.028 a
	Gr	0.411 e	0.493 a	0.356 e	-3864.783 b	0.023 a	0.027 a

*Letters based on Tukey test performed considering estimations obtained from 10 independent runs of the full models with BGLR (equation 1). [1] Posterior Mean of the Log Likelihood. [2] Expected Genetic Gain on trait scale.

Effect of the genetic information to build the relationship matrices

The incorporation of relationship information in the analysis generated better PA results than the phenotypic-BLUP model without it. Overall, we observed that higher values for the phenotypic PA were obtained when marker-based relationship matrices were used, when compared with phenotypic and pedigree BLUP (*I* and *A* matrices, respectively). However, the marker-based and pedigree-based results were not always significantly different from each other (Figure 3, Table S1). The use of molecular data yielded phenotypic PA values ranging from 0.27 (pH) to 0.49 (fruit scar) in 2014, and from 0.15 (flower buds) to 0.51 (fruit firmness) in 2015. Lower PA values were obtained for traits with lower heritability and better results were observed for the second year of evaluation. The biggest increase in the PA values can be seen for fruit firmness – when we compared marker and pedigree results, we observed an average increase of 13.37% in 2014. Also, an increase in the PA values of 11% was observed for fruit diameter and yield in 2015 when markers were used instead of pedigree data.

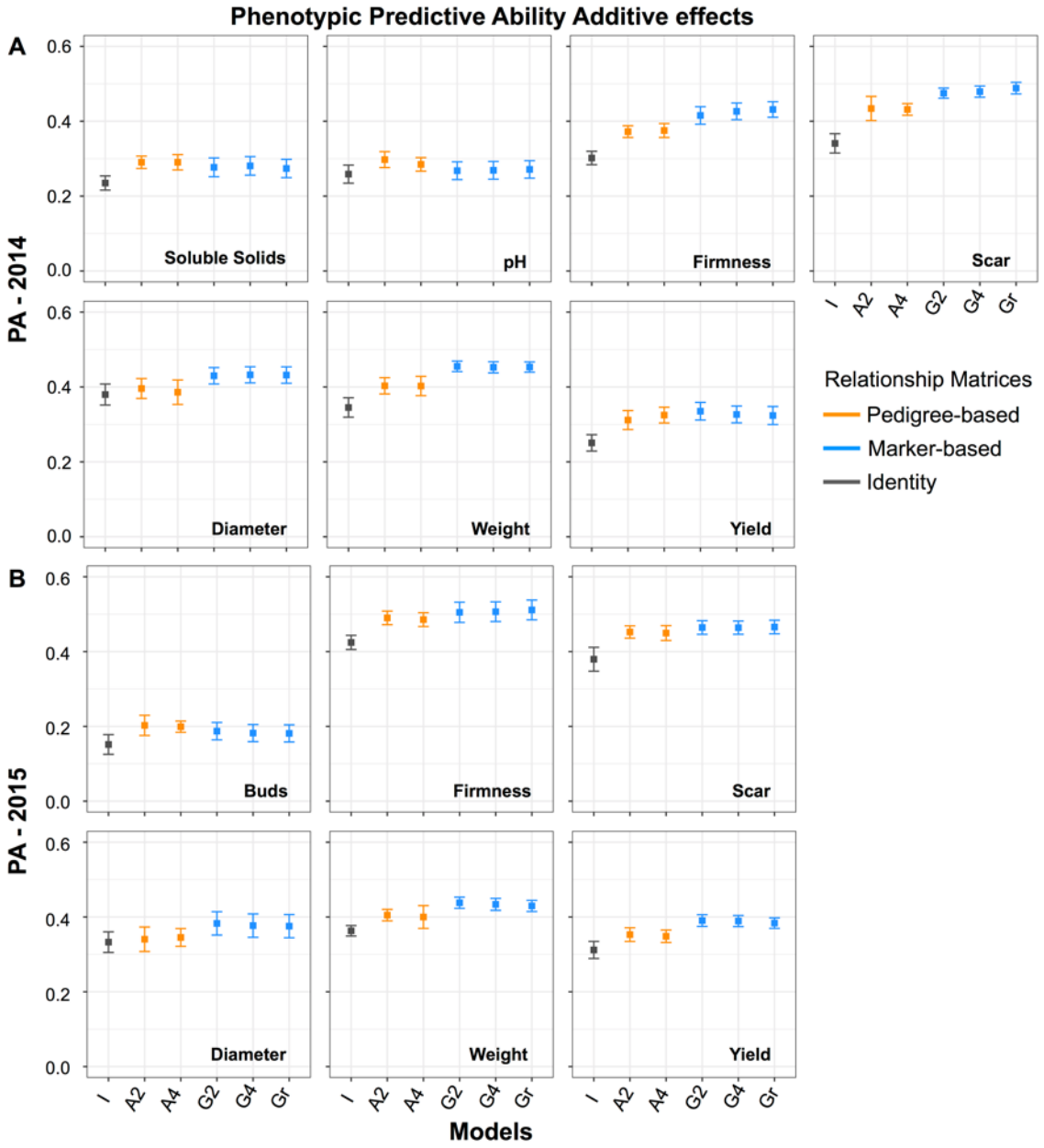

Figure 3. Phenotypic predictive abilities for (**A**) seven traits in 2014, and (**B**) for six traits in 2015 for different dosage parameterizations (indicated by the numbers 2 or 4) of the relationship matrices (indicated by the letters *I*, *A*, and *G*) in the prediction of breeding values of 1,847 SHB individuals.

The use of pedigree-based relationship matrices generated higher phenotypic PA values for all the traits, when compared with the assumption of unrelated individuals (*i.e.*, identity matrix). Unlike the identity matrix, the use of pedigree-based matrix assumes that there is relationship (expected values) among individuals. The phenotypic PA obtained for the pedigree methods in 2014 yielded values from 0.20 (flower bud) to 0.49 (fruit firmness).

As with marker-based methods, smaller values were observed for traits with lower heritability (*i.e.*, pH, brix, and flower bud). For 2015, the PA results for the phenotypic-BLUP were 0.36, 0.38, and 0.42, for fruit weight, fruit scar, and fruit firmness, respectively. The PA values obtained for the same traits with pedigree-BLUP were 0.40, 0.45, and 0.49, respectively. No significant differences between the models' stability were observed (Table S1).

Use of dosage information and continuous genotypes

Our results indicate that the importance of dosage in GS will vary depending on the trait being analyzed. For example, in 2014 the PA for fruit firmness, fruit scar, and fruit diameter showed modestly better phenotypic PA when the tetraploid and continuous parameterizations were applied, as opposed to the diploid parameterization (Figure 3, Table S1). Although no significant difference was observed between marker-based models, the use of relationship matrices derived from continuous genotype data (ploidy-free parameterization) performed equally well as the best models (Figure 3, Table S1). However, the goodness-of-fit statistics show that the use of a relationship matrix obtained from the continuous genotype data significantly improved model fit for all traits (Table 3). This was followed by the tetraploid parameterization using marker-based data.

Expected genetic gain in a perennial fruit tree, blueberry

The results obtained for the expected genetic gain (EGG) are summarized in Table 3. GS offers the possibility to accelerate genetic improvement by decreasing the breeding cycle and selecting superior individuals earlier in the breeding program. Considering a breeding cycle (L) of 12 years (Cellon *et al.* 2018) we propose that routine genomic selection could be implemented in the second stage of the blueberry breeding program, which would allow the

omission of a whole stage (stage III), and a three-year reduction for cultivar release (Figure 4).

Higher EGG was obtained for all traits when marker-based matrices (*i.e.*, genomic selection) were applied (Table 3), which was mainly related to the reduction in cycle time. The implementation of GS in the second stage population would lead to an increase in the EGG varying from 27% (pH) to 119% (scar) when compared with the application of phenotypic BLUP. Considering the comparison of marker-based and pedigree-based models, an increase of 15% (pH) to 41% (fruit weight, fruit scar, and flower buds) in the EGG was observed (Table 3). In addition, the use of continuous data generated EGG values that were not significantly different of the best models for all traits (Table 3).

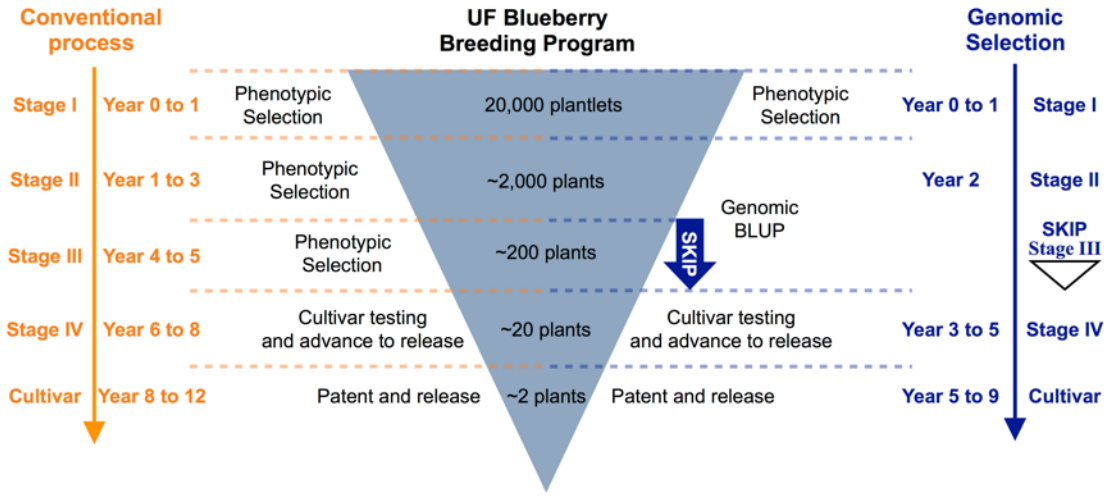

Figure 4. Example of the University of Florida blueberry breeding program stages and times of selection. Conventional process (left) compared with the proposed process implementing genomic selection (right).

DISCUSSION

In this study, six linear mixed models were applied to predict breeding values for eight yield and fruit-quality traits measured in a real blueberry breeding population as model. Analyses were based on phenotypic, pedigree, and high-density marker data from 1,847 individuals. We compared the expected genetic gain, the stability, and the PA of models

386 considering different sources to build the relationship matrices (only phenotype=BLUP,
387 phenotypes + pedigree=P-BLUP, phenotypes + genomic=G-BLUP). Our results also
388 explored models accounting for ploidy information and proposed the use of genotypic data
389 that is independent of assumptions regarding ploidy levels (continuous) to perform GS,
390 avoiding the need for *a priori* parameterization for a given ploidy level.
391
392 **Continuous data**
393 Our research showed empirical evidences that the use of continuous genotypic data
394 from NGS can be effectively applied in GS models for autotetraploid species. This method
395 was tested and compared with marker calling methodologies at the individual level in
396 genome wide association studies (Grandke *et al*. 2016). It was also tested in family pool data
397 for GS (Ashraf *et al.* 2014; Guo *et al*. 2018), as well as used at the individual level in
398 tetraploid potato for GS by Sverrisdottir *el al.* (2017). However, to our knowledge the
399 comparison of continuous genotypes with ploidy parameterizations for genomic selection has
400 not yet been reported. Here we empirically compare diploid, tetraploid, and continuous data
401 at the individual level for the application of genomic selection in an autotetraploid species.
402 In polyploids, the assignment of genotypic classes based on NGS data is a major
403 challenge, with high risk of misclassification (Grandke *et al.* 2016, Bourke *et al.* 2018). The
404 problem is further exacerbated as the ploidy increases – for a given level of ploidy, *n*, the
405 expected number of genotypic classes is *2n+1*. As a consequence, the signal distribution
406 derived from each genotypic class increasingly approximates a continuous distribution where
407 no clear separation is observed (Grandke *et al.* 2016). Despite extensive research to address
408 these challenges (Serang *et al.* 2012), advances have been mostly limited to SNP arrays in
409 tetraploid data (Carley *et al.* 2017). Studies that evaluated genotype calling with NGS data
410 obtained from polyploids show that no method works properly, and that misclassification of

411 genotypes can significantly interfere in the results of genetic studies (Grandke *et al.* 2016).
412 This misclassification can be observed in our results when a diploid, or tetraploid
413 parameterization is used in the genomic data (Figure 2A-B) with standard parameters of
414 filtering. The use of the continuous genotyping approach provides a relevant alternative to
415 overcome this issue that is independent of assumptions regarding ploidy level. Models that
416 used continuous genotypic data performed as well as the best models and resulted in
417 modestly better predictive abilities for some of the traits (*i.e.*, fruit firmness, fruit scar, and
418 fruit diameter; Table 3), but better data fit, which could indicate better prediction of future
419 populations. The use of continuous genotypes also simplifies the analysis complexity and
420 time by eliminating the genotype calling and parameterization for a give ploidy, because
421 instead, the ratio of reads assigned to each allele are used. Finally, our results showed that the
422 addition of noise associated with the continuous distribution in the genotypes significantly
423 improved model fitting for all analyzed traits (Table 3), instead of increasing the complexity
424 of the models. The benefits of continuous genotyping could easily be extended to more
425 complex polyploids (higher ploidies), where the genotype attribution is even more difficult,
426 however higher sequencing depth would probably be required. Meanwhile, for more complex
427 models, such as those that consider dominance effects, dosage calling is still necessary.
428
429 **Relationship matrices**
430 Our results also showed that including information based on the genetic merit of the
431 individuals yielded better results when compared with the phenotypic-BLUP analysis (based
432 on the identity matrix; Table 3), corroborating previous studies in the literature (e.g., Muir
433 2005; Resende *et al.* 2012a; Muñoz *et al.* 2014a). In addition, the use of marker-based
434 methodologies generated better predictions than pedigree for most of the traits. Marker-based
435 methods allow the capture of Mendelian segregation. This is especially important in our

436 population, since it was composed of 117 full-sib families. In this context, pedigree-based
437 methods have no power to distinguish variance within families. Another advantage is that
438 marker-based methods allows the computation of genetic similarity among unidentified
439 individuals in the pedigree, and corrections of errors in the pedigree, which can affect
440 parameter estimation causing reduction in the genetic gain (Muñoz *et al.* 2014b).
441 In our results, some non-significant differences between pedigree and marker-based
442 methods were identified, which could be an effect of the extensive pedigree data used, as well
443 as bias in pedigree-based estimations. Pedigree-based methods can overestimate the
444 reliability of selection and consequently, the accuracy (Bulmer 1971; Gorjanc *et al.* 2015).
445 Furthermore, it also presents low efficiency to capture and estimate genetic relationships
446 among individuals (Resende *et al.* 2017).
447 It is interesting to notice that we used extensive pedigree information that dates back
448 to 1907 for our predictions, which may not be common in other autopolyploid breeding. This
449 extensive information can have significant implications on the estimation of relationship
450 coefficients (Amadeu *et al.* 2016) and consequently, in breeding value predictions. For
451 breeding programs with smaller pedigree depth information, the comparison between
452 accuracies of prediction from marker and pedigree-based methodologies could be even bigger
453 than what was found in our study.
454
455 **Allele dosage**
456 The results obtained for both models that assumed more than three genotypic classes
457 (*G4* and *Gr*) demonstrate the importance of considering dosage in the prediction of breeding
458 values. However, this will depend on the trait analyzed, as previously reported by Nyine *et al.*
459 (2018) and Endelman *et al.* (2018). For example, modest improvement was verified in the PA
460 for fruit firmness, fruit scar, and fruit diameter when this factor was considered in the models.

461 In addition, model fitting was significantly better for methods that accounted for dosage
462 information (Figure 3, Table 3, Table S1). The inclusion of nonadditive effects into the
463 models could also improve model accuracy. Endelman *et al.* (2018) demonstrated that the
464 inclusion of digenic effects, as well as accounting for ploidy information, presented a higher
465 accuracy over diploid models when using a SNP array.
466
467 **Genomic selection for perennial autopolyploids**
468 We also demonstrate the value of applying GS in a perennial fruit tree, blueberry. One
469 cycle of blueberry breeding takes from 12 to 15 years until the release of a new cultivar
470 (Lyrene 2008; Cellon *et al.* 2018). By applying selection based on high-density markers at
471 early stages of the program, the time to cultivar release could decrease by three years (Figure
472 4), significantly improving the expected genetic gain per unit of time. More specifically, the
473 use of GS would lead to an average increase of 86% in the EGG when compared with
474 phenotypic BLUP, and an average increase of 32% over the application of pedigree-based
475 models (Table 3). Implementing GS in this form could eliminate one stage in the breeding
476 and selection process toward cultivar development, which will reduce costs associated with
477 field trials and phenotyping. The implementation of GS would require extra financial outlay
478 when genotyping and accurately phenotyping the training population. However, the savings
479 on phenotyping and field trials of future generations (selection populations) could results in a
480 break-even financial exercise, and as a result could be a cost-effective application of GS.
481 However, this financial analysis needs to be performed for each crop in a case-by-case basis.
482
483
484
485

Funding

USDA - Agriculture and Food Research Initiative Grant no. 2014-67013-22418 to Patricio R. Munoz, James W. Olmstead and Jeffrey B. Endelman from the USDA National Institute of Food and Agriculture.

Ivone de Bem Oliveira was funded by the CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) [PSDE scholarship: 88881.131685/2016-01].

Acknowledgments

The authors thank the University of Florida blueberry breeding program technical support, especially Dr. Paul M. Lyrene, David Norden, and Werner Collante. Special thanks to James Olmstead and Catherine Cellon, who coordinate the phenotyping and genotyping of the population as part of Catherine Cellon MS degree.

Supplemental Material

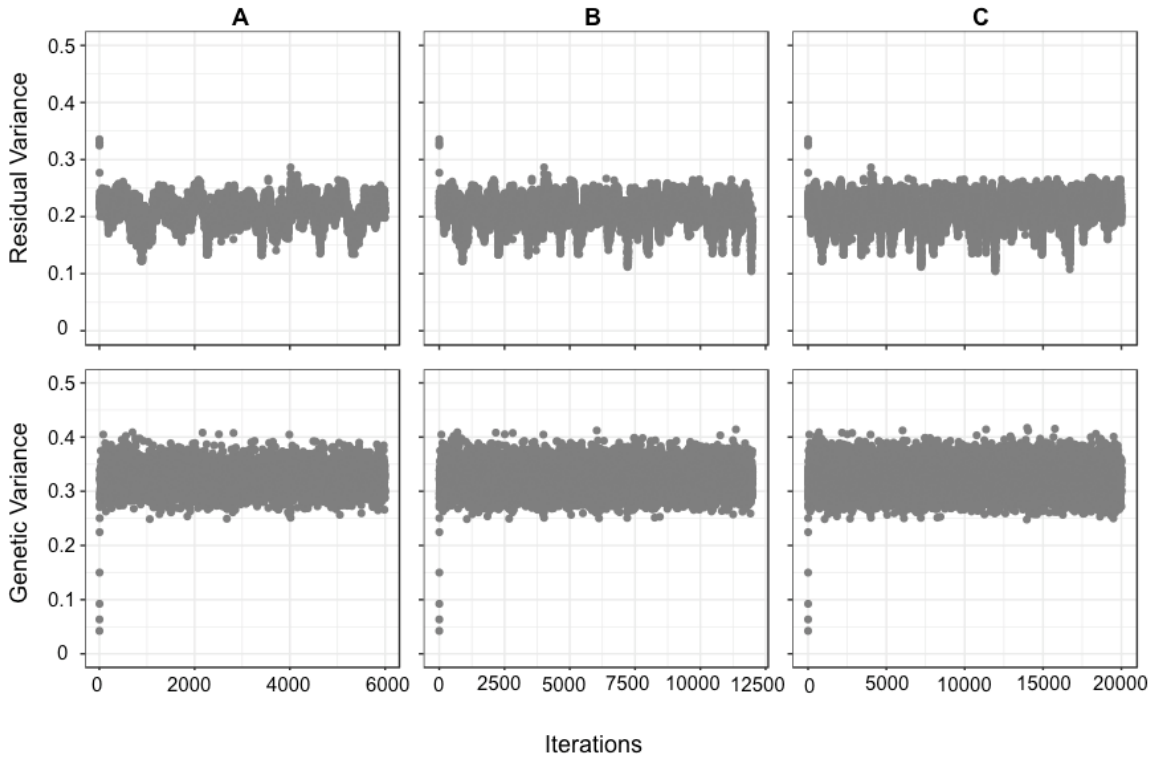

Figure S1. Model convergence obtained using different values for iterations (I), burn-in (B) and thinning (T) to perform the GS in BGLR for 1847 SHB genotypes. **A)** I=30K B=5K and T =5; **B)** I=60K B=10K and T =5; **C**) I=100K B=10K and T =5.

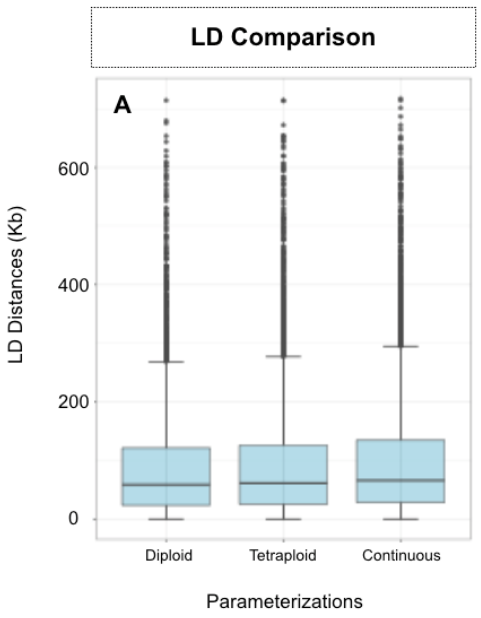

Figure S2. Distribution of linkage disequilibrium distances at the empirical threshold r^2=0.2 for the diploid, tetraploid and continuous parameterizations.

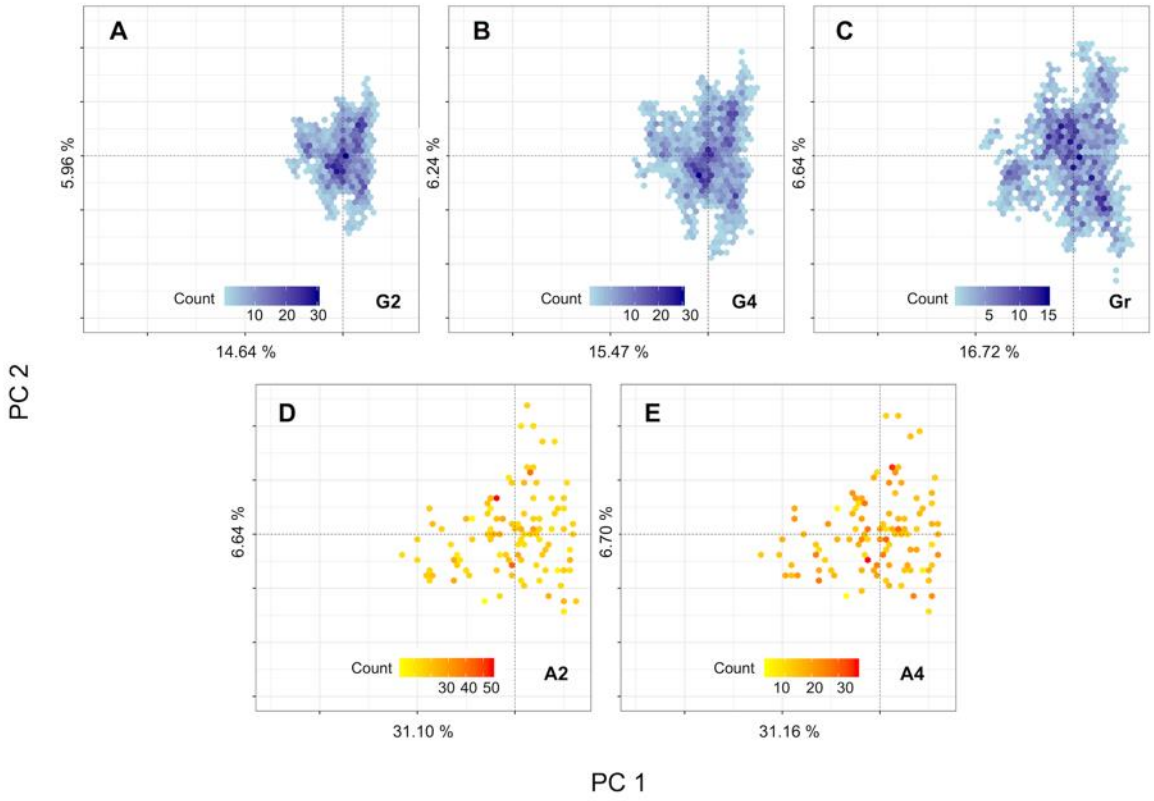

Figure S3. Principal components heat plots for the relationship matrices used in this study. In blue are the marker-based, and in orange are the pedigree-based matrices. Genomic matrices were computed using VanRaden's (2008) methodology (**A**) Diploid, (**B**) tetraploid, and (**C**) continuous relationship matrix. Pedigree-based matrices were computed for the additive effects using (**D**) Henderson (1976) methodology for diploid, and (**E**) Kerr *et al.* (2012) methodology for tetraploid.

Table S1. Accuracy and model stability values for eight yield and fruit-related traits analyzed with six linear mixed models with different dosage parameterizations of the relationship matrices. Source of information, and dosage parameterizations for the relationship matrices indicated by the letters (*I*, *A*, or *G*), and numbers (2 or 4), respectively*

Trait	Relationship matrix	Predictive ability 2014	Predictive ability 2015	MSE (2014)	MSE (2015)
Soluble Solid (°Brix)	*I*	0.235 a	-	5.795 a	-
	A2	0.290 a	-	4.926 a	-
	A4	0.290 a	-	4.531 a	-
	G2	0.277 a	-	2.238 a	-
	G4	0.281 a	-	2.197 a	-
	Gr	0.274 a	-	2.216 a	-
Flower Buds	*I*	-	0.152 a	-	9.261 a
	A2	-	0.203 a	-	8.280 a
	A4	-	0.199 a	-	8.039 a
	G2	-	0.187 a	-	11.469 a

	G4	-	0.182 a	-	11.718 a
	Gr	-	0.181 a	-	11.860 a
Fruit Diameter	*I*	0.38 a	0.333 a	8.827 a	8.918 a
	A2	0.396 a	0.340 a	8.072 a	8.638 a
	A4	0.386 a	0.345 a	9.134 a	3.409 a
	G2	0.403 a	0.383 a	19.796 a	6.339 a
	G4	0.432 a	0.377 a	19.423 a	6.484 a
	Gr	0.432 a	0.376 a	19.718 a	7.241 a
Fruit Firmness	*I*	0.302 b	0.425 a	2725.511 a	1109.026 a
	A2	0.372 ab	0.490 a	2146.811 a	734.108 a
	A4	0.375 ab	0.486 a	778.212 a	1490.294 a
	G2	0.415 a	0.505 a	1240.152 a	2083.417 a
	G4	0.426 a	0.507 a	1245.192 a	2085.359 a
	Gr	0.431 a	0.511 a	1250.673 a	2124.308 a
pH	*I*	0.259 a	-	0.244 a	-
	A2	0.297 a	-	0.257 a	-
	A4	0.285 a	-	0.126 a	-
	G2	0.268 a	-	0.253 a	-
	G4	0.269 a	-	0.259 a	-
	Gr	0.271 a	-	0.256 a	-
Fruit Scar	*I*	0.341 b	0.380 a	0.236 a	0.424 a
	A2	0.434 a	0.452 a	0.113 a	0.416 a
	A4	0.432 a	0.45 a	0.112 a	0.181 a
	G2	0.475 a	0.464 a	0.160 a	0.197 a
	G4	0.479 a	0.464 a	0.160 a	0.199 a
	Gr	0.488 a	0.466 a	0.171 a	0.184 a
Fruit Weight	*I*	0.345 b	0.363 a	0.502 a	0.337 a
	A2	0.403 ab	0.405 a	0.529 a	0.391 a
	A4	0.403 ab	0.400 a	0.404 a	1.100 a
	G2	0.455 a	0.438 a	1.030 a	0.615 a
	G4	0.453 a	0.434 a	1.059 a	0.630 a
	Gr	0.453 a	0.430 a	1.027 a	0.702 a
Yield	*I*	0.251 a	0.312 b	0.881 a	1.221 a
	A2	0.312 a	0.353 ab	0.661 a	1.214 a
	A4	0.325 a	0.348 ab	0.638 a	1.730 a
	G2	0.335 a	0.390 a	1.191 a	2.343 a
	G4	0.327 a	0.389 a	1.263 a	2.308 a
	Gr	0.324 a	0.384 ab	1.297 a	2.294 a

*Letters based on Tukey test performed considering estimations obtained from 10 independent runs of the full models with BGLR (equation 1).

Literature Cited

Adams, K. L., R. Cronn, R. Percifield, and J. F. Wendel, 2003 Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc. Natl. Acad. Sci. 100: 4649-4654.

Amadeu, R. R., C. Cellon, J. W. Olmstead, A. A. Garcia, M. F. Resende *et al.*, 2016 AGHmatrix: R Package to construct relationship matrices for autotetraploid and diploid species: a blueberry example. Plant Genome 9: 3.

Annicchiarico, P., N. Nazzicari, X. Li, Y. Wei, L. Pecetti *et al.*, 2015 Accuracy of genomic selection for alfalfa biomass yield in different reference populations. BMC genomics, 16: 1020.

Ashraf, B., V. Edriss, D. Akdemir, E. Autrique, D. Bonnett *et al.*, 2016. Genomic prediction using phenotypes from pedigreed lines with no marker data. Crop Sci. 56: 957-964.

Ashraf, B. H., J. Jensen, T. Asp, and L. L. Janss, 2014 Association studies using family pools of outcrossing crops based on allele-frequency estimates from DNA sequencing. Theor. Appl. Genet. 127:1331-1341.

Bian, Y., J. Ballington, A. Raja, C. Brouwer, R. Reid, M. Burke *et al.*, 2014 Patterns of simple sequence repeats in cultivated blueberries (*Vaccinium* section *Cyanococcus* spp.) and their use in revealing genetic diversity and population structure. Mol. Breed. 34: 675-689.

Birchler, J. A., U. Bhadra, M. P. Bhadra, and D. L. Auger, 2001 Dosage-dependent gene regulation in multicellular eukaryotes: implications for dosage compensation, aneuploid syndromes, and quantitative traits. Dev. Biol. 234: 275-288.

Bourke, P. M., R. E. Voorrips, R. G. Visser, and C. Maliepaard, 2018 Tools for genetic studies in experimental populations of polyploids. Front. Plant Sci. 9: 513.

Bulmer, M., 1971 The effect of selection on genetic variability. Am. Nat. 105: 201-211.

Carley, C. A. S., J. J. Coombs, D. S. Douches, P. C. Bethke, J. P. Palta *et al.*, 2017. Automated tetraploid genotype calling by hierarchical clustering. Theor. Appl. Genet. 130: 717-726.

Cellon, C., R. R. Amadeu, J. W. Olmstead, M. R. Mattia, L. F. V. Ferrao *et al.*, 2018 Estimation of genetic parameters and prediction of breeding values in an autotetraploid blueberry breeding population with extensive pedigree data. Euphytica, 214: 1-13.

Comai, L., 2005 The advantages and disadvantages of being polyploid. Nat. Rev. Genet. 6: 836.

de Los Campos, G., and P. Perez-Rodriguez, 2016 BGLR: Bayesian Generalized Linear Regression. R package version 1.0.5. https://cran.r-project.org/package=BGLR

de los Campos, G., J. M. Hickey, R. Pong-Wong, H. D. Daetwyler, and M. P. Calus, 2012 Whole genome regression and prediction methods Appl.ied to plant and animal breeding. Genetics 112.

Desta, Z. A., and R. Ortiz, 2014 Genomic selection: genome-wide prediction in plant improvement. Trends Plant Sci. 19: 592-601.

Dufresne, F., M. Stift, R. Vergilino, R., and B. K. Mable, 2014 Recent progress and challenges in population genetics of polyploid organisms: an overview of current state-of-the-art molecular and statistical tools. Mol. Ecol. 23: 40-69.

Endelman, J. B., C. A. S. Carley, P. C. Bethke, J. J. Coombs, M. E. Clough *et al.*, 2018 Genetic variance partitioning and genome-wide prediction with allele dosage information in autotetraploid potato. Genetics, 300685.

Ferrão, L. F. V., J. Benevenuto, I. de Bem Oliveira, C. Cellon, J. Olmstead *et al.*, 2018 Insights into the genetic basis of blueberry fruit-related traits using diploid and polyploid models in a GWAS context. Front. Ecol. Evol. 6: 107.

Garcia, A. A., M. Mollinari, T. G. Marconi, O. Serang, R. R. Silva *et al.*, 2013 SNP genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autopolyploids. Sci. Rep. 3: 3399.

Garrison, E., and G. Marth, 2012 Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907.

Gorjanc, G., P. Bijma, and J. M. Hickey, 2015 Reliability of pedigree-based and genomic evaluations in selected populations. Genet. Sel. Evol. 47: 65.

Gouy, M., Y. Rousselle, D. Bastianelli, P. Lecomte, L. Bonnal *et al.*, 2013 Experimental assessment of the accuracy of genomic selection in sugarcane. Theor. Appl. Genet. 126: 2575-2586.

Grandke, F., P. Singh, H. C. Heuven, J. R. De Haan, and D. Metzler, 2016 Advantages of continuous genotype values over genotype classes for GWAS in higher polyploids: a comparative study in hexaploid chrysanthemum. BMC Genomics 17: 672.

Guo, M., D. Davis, and J. A. Birchler, 1996 Dosage effects on gene expression in a maize ploidy series. Genetics 142: 1349-1355.

Guo, X., F. Cericola, D. Fè, M. G. Pedersen, I. Lenk, C. S. Jensen *et al.*, 2018 Genomic Prediction in Tetraploid Ryegrass Using Allele Frequencies Based on Genotyping by Sequencing. Front. Plant. Sci. 9.

Gupta, V., A. D. Estrada, I. Blakley, R. Reid, K. Patel, M. D. Meyer *et al.*, 2015 RNA-Seq analysis and annotation of a draft blueberry genome assembly identifies candidate genes involved in fruit ripening, biosynthesis of bioactive compounds, and stage-specific alternative splicing. Gigascience 4: 5.

Henderson, C. R., 1976 A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. Biometrics 32: 69-83.

Hieter, P., and T. Griffiths, 1999 Polyploidy--more is more or less. Science 285: 210-211.

Kerr, R. J., L. Li, B. Tier, G. W. Dutkowski, and T. A. McRae, 2012 Use of the numerator relationship matrix in genetic analysis of autopolyploid species. Theor. Appl. Genet. 124: 1271-1282.

Lee, W. P., M. P. Stromberg, A. Ward, C. Stewart, E. P. Garrison *et al.*, 2014 MOSAIK: a hash-based algorithm for accurate next-generation sequencing short-read mapping. PloS one 9: e90581.

Lyrene, P. 2008 Breeding southern highbush blueberries. Plant Breed. Rev. 30: 353-414.

Lyrene, P., 2002 Development of highbush blueberry cultivars adapted to Florida. J. Am. Pomol. Soc. 56: 79.

Hayes, B. J., and M. E. Goddard, 2001 Prediction of total genetic value using genome-wide dense marker maps. Genetics, 157: 1819-1829.

Mollinari, M., and O. Serang, 2015 Quantitative SNP genotyping of polyploids with MassARRAY and other platforms, pp. 215-241 in *Plant Genotyping* edited by Humana Press, New York, NY.

Muir, W. M., 2007 Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. J. Anim. Breed. Genet. 124: 342-355.

Munoz, P. R., M. F. Resende, D. A. Huber, T. Quesada, M. D. Resende, D. B. Neal *et al.* 2014a. Genomic relationship matrix for correcting pedigree errors in breeding populations: impact on genetic parameters and genomic selection accuracy. Crop Sci. 54: 1115-1123.

Muñoz, P. R., M. F. Resende, S. A. Gezan, M.D. V. Resende, G. de los Campos *et al.*, 2014b Unraveling additive from non-additive effects using genomic relationship matrices. Genetics 114.

Nyine, M., B. Uwimana, N. Blavet, E. Hřibová, H. Vanrespaille *et al.*, 2018. Genomic prediction in a multiploid crop: genotype by environment interaction and allele dosage effects on predictive ability in banana. Plant Genome 11:170090.

Osborn, T. C., J. C. Pires, J. A. Birchler, D. L. Auger, Z. J. Chen *et al.*, 2003 Understanding mechanisms of novel gene expression in polyploids. Trends Genet. 19: 141-147.

Pérez, P., and G. de Los Campos, 2014 Genome-wide regression & prediction with the BGLR statistical package. Genetics 114.

R Core Team, 2018 R: A Language and Environment for Statistical Computing.

Resende, M. F. R., P. Munoz, J. J. Acosta, G. F. Peter, J. M. Davis *et al.*, 2012 Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. New Phytol. 193: 617-624.

Resende, M. F., P. Muñoz, M. D. Resende, D. J. Garrick, L. R. Fernando *et al.*, 2012 Accuracy of genomic selection methods in a standard dataset of loblolly pine (*Pinus taeda* L.). Genetics 111.

Resende, R. T., M. D. V. Resende, F. F. Silva, C. F. Azevedo, E. K. Takahashi *et al.*, 2017 Assessing the expected response to genomic selection of individuals and families in Eucalyptus breeding with an additive-dominant model. Heredity 119: 245.

Schindelin, J., I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair *et al.*, 2012 Fiji: an open-source platform for biological-image analysis. Nat. methods 9: 676.

Serang, O., M. Mollinari, and A. A. F. Garcia, 2012 Efficient exact maximum a posteriori computation for bayesian SNP genotyping in polyploids. PLoS One, 7: e30906.

Slater, A. T., N. O. Cogan, J. W. Forster, B. J. Hayes, and H. D. Daetwyler, 2016 Improving genetic gain with genomic selection in autotetraploid potato. Plant Genome, 9: 3.

Slater, A. T., G. M. Wilson, N. O. Cogan, J. W. Forster, and B. J. Hayes, 2014 Improving the analysis of low heritability complex traits for enhanced genetic gain in potato. Theor. Appl. Genet. 127: 809-820.

Soltis, D. E., and P. S. Soltis, 1999 Polyploidy: recurrent formation and genome evolution. Trends Ecol. Evol. 14: 348-352.

Soltis, D. E., P. S. Soltis, D. W. Schemske, J. F. Hancock, J. N. Thompson *et al.*, 2007 Autopolyploidy in angiosperms: have we grossly underestimated the number of species?. Taxon, 56: 13-30.

Soltis, D. E., C. J. Visger, D. B. Marchant, and P. S. Soltis, 2016 Polyploidy: pitfalls and paths to a paradigm. Am. J. Bot. 103: 1146-1166.

Spoelhof, J. P., P. S. Soltis, and D. E. Soltis, 2017 Pure polyploidy: closing the gaps in autopolyploid research. J. Syst. Evol. 55: 340-352.

Sverrisdóttir, E., S. Byrne, E. H. R. Sundmark, H. Ø. Johnsen, H. G. Kirk *et al.*, 2017 Genomic prediction of starch content and chipping quality in tetraploid potato using genotyping-by-sequencing. Theor. Appl. Genet. 130: 2091-2108.

Uitdewilligen, J. G., A. M. A. Wolters, B. Bjorn, T. J. Borm, R. G. Visser *et al.*, 2013 A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. PloS one 8: e62355.

VanRaden, P. M., 2008 Efficient methods to compute genomic predictions. J. Dairy Sci. 91: 4414-4423.

Zhang, Z., X. Ding, J. Liu, D. J. de Koning, and Q. Zhang, 2011 Genomic selection for QTL-MAS data using a trait-specific relationship matrix. BMC Proc. 5: S15.