AGHmatrix: R Package to Construct Relationship Matrices for Autotetraploid and Diploid Species: A Blueberry Example

Rodrigo R. Amadeu, Catherine Cellon, James W. Olmstead, Antonio A. F. Garcia, Marcio F. R. Resende Jr., and Patricio R. Muñoz*

Abstract
Progress in the rate of improvement in autopolyploid species has been limited compared with diploids, mainly because software and methods to apply advanced prediction and selection methodologies in autopolyploids are lacking. The objectives of this research were to (i) develop an R package for autopolyploids to construct the relationship matrix derived from pedigree information that accounts for autopolyploidy and double reduction and (ii) use the package to estimate the level and effect of double reduction in an autotetraploid blueberry breeding population with extensive pedigree information. The package is unique, as it can create A-matrices for different levels of ploidy and double reduction, which can then be used by breeders to fit mixed models in the context of predicting breeding values (BVs). Using the data from this blueberry population, we found for all the traits that tetrasomic inheritance created a better fit than disomic inheritance. In one of the five traits studied, the level of double reduction was different from zero, decreasing the estimated heritability, but it did not affect the prediction of BVs. We also discovered that different depths of pedigree would have significant implications on the estimation of double reduction using this approach. This freely available R package is available for autopolyploid breeders to estimate the level of double reduction present in their populations and the impact in the estimation of genetic parameters as well as to use advanced methods of prediction and selection.

Core Ideas
- We developed an R package for autopolyploids to construct the relationship matrix.
- We estimated the level and effect of double reduction in blueberry.
- The package is unique as it can create A-matrices for different levels of ploidy.
- The package can create A-matrices for different levels of double reduction.

Several important agronomic and horticultural crops such as alfalfa (Medicago sativa L.), potato (Solanum tuberosum L.), and highbush blueberry (Vaccinium corymbosum L.) are autotetraploids, with four homologous chromosomes present in each linkage group. As more than two homologous chromosomes exist in autotetraploids, during meiosis chromosome pairing can occur between a pair of randomly chosen chromosomes (bivalents) or between more than two homologous chromosomes (multivalents) (Fisher, 1947). This polysomic...
inheritance results in different segregation ratios for a given locus compared with diploid species, as up to four copies of an allele lead to the possibility of higher orders of allele interaction (see details in Gallais, 2003). Thus, considering a biallelic marker, there are five possible states for a given A allele (referred hereafter as dosage): nulliplex, simplex, duplex, triplex, and quadruplex representing aaaa, Aaaa, AAAa, and AAAA, respectively. Moreover, the phenomenon known as double reduction, where sister chromatids segregate into the same gamete, creates difficulties when studying quantitative or population genetics in autotetraploid species. Despite the fact that double reduction is not well studied, its occurrence has practical implications in plant breeding because it can result in increased inbreeding even when two unrelated individuals are crossed (Ma, 1995; Ma, 1993; Gallais, 2003; Kerr et al., 2012). These challenges, and the fact that these differences are often ignored in breeding strategies, have been identified as the main reasons why the rate of genetic improvement has been limited and sometimes null in autotetraploid species as compared with their diploid counterparts (Brummer, 1999; Katapa-Mupondwa et al., 2002; Jansky, 2009). As an example, the genetic gain for yield in alfalfa since 1940 has been between 0 and 0.30% yr⁻¹ on average, while 2% has been estimated in annual diploid crops (Brummer, 1999).

Breeding autotetraploid species also typically take longer than diploid species, and the development of an improved cultivar can take between 10 and 20 yr (Lyrene, 2008; Bradshaw and Bonierbale, 2010). One reason is that it takes longer for autotetraploids to eliminate deleterious alleles because of the greater number of possible states (i.e., Aaaa, AAAa, and AAAA) (Gallais, 2003). Even though progress has been made in these species, materialized in the many cultivars released, current improvement strategies (e.g., phenotypic recurrent selection) and analysis methods are rudimentary and known to be inefficient (Slater et al., 2013).

One of such autotetraploid species is blueberry, including highbush, southern highbush (V. corymbosum interspecific hybrids), and lowbush (V. angustifolium L.). Blueberries are an economically important small fruit crop for the United States. In 2014, 0.25 Tg (543.5 million pounds) of blueberries were harvested for fresh production and valued at $824.9 million (USDA, 2015). United States production has increased an average of 20% every 2 yr since 2008, while acreage has increased by nearly 75% between 2005 and 2012. This increase is based in part on the reputation of blueberry fruit and products as health foods because of the high amounts of antioxidants such as anthocyanins. Antioxidants have been correlated with beneficial health effects in numerous biological systems including immunology, ophthalmology, cardiology, neurology, metabolism, and inflammation (Cassidy et al., 2013; Fan et al., 2012; Heim et al., 2012; Krikorian et al., 2010; Li et al., 2013; Liu et al., 2012; McAnulty et al., 2011; Shukitt-Hale, 2012; Tipton et al., 2013; Yousef et al., 2013).

The blueberry breeding program at the University of Florida develops southern highbush blueberry cultivars (2x = 4n = 48). These hybrid cultivars are derived mainly from crosses between northern highbush tetraploid cultivars (V. corymbosum) and diploid species (V. darrowii), with lesser contributions from various other Vaccinium species such as V. arboreum, V. elliottii, and V. virgatum (Brevis et al., 2008; Lyrene 1997, 2008; Olmstead et al., 2013). Many traits under selection in this species are polygenic (Lyrene 1993) and the few inheritance studies of tetraploid V. corymbosum and V. corymbosum × V. darrowii crosses have shown inheritance patterns that follow expected autotetraploid segregation ratios (Draper and Scott, 1971; Krebs and Hancock, 1989; Qu and Hancock, 1995). Additionally, the degree of double reduction present in blueberry is not yet known, and previous reports were not conclusive (Krebs and Hancock, 1989).

The use of quantitative genetic methods for breeding of blueberries or any other autotetraploid plant is not as developed as diploid species. In conventional diploid breeding, the use of linear mixed models, popularly known as best linear unbiased prediction (BLUP), is a method used to estimate variance components through restricted maximum likelihood (REML) and predict BVs. This methodology was developed by Henderson in the 1950s (Van Vleck, 1998) and was first applied to livestock breeding (VanRaden, 2008). In the past two decades, BLUP has gained popularity in plant breeding, first with perennial and forestry crops and more recently with annual crops (Piepho et al., 2008). In this period, the applicability of BLUP has been tested and adopted in breeding diploid crops (Bernardo, 1994, 1996; Cossu et al., 2010; Resende et al., 2012; Massman et al., 2013). However, its use in many polyploidy species, such as blueberries, has not been explored yet.

One of the main characteristics in the use of REML and BLUP in breeding is the use of a numerator matrix (A) containing an estimate of the pairwise relationship among individuals. This information is used to model the variance covariance between the populations. Kerr et al. (2012) developed an extension of the numerator relationship matrix based on pedigree values (A-matrix) for autopolyploids. This calculation includes the ploidy level and w coefficient (proportion of the alleles which are identical by descent, double reduction). Using this new tetraploid A-matrix, Slater et al. (2013) found that higher genetic gains could be obtained for quantitative traits in autotetraploid potato when compared with the traditional phenotypic selection. However, there is currently no software available to construct such matrices so that autotetraploid breeders can test and implement these new and more efficient selection methods.

The objective of this study was two-fold: (i) to develop an R package for construction of an autotetraploid relationship matrix that considers autopolyploidy and double reduction derived from pedigree information and (ii) to estimate the level of double reduction present in the University of Florida blueberry breeding germplasm and
to study its effects on model fitting and genetic parameter estimations. This information will allow the future implementation of BLUP and REML methods to estimate BVs in the breeding program, study of genetic diversity, inbreeding, definition of heterotic groups, etc.

Materials and Methods

Blueberry Breeding Population and Pedigree
The southern highbush blueberry population used in this study was generated from the breeding program at the University of Florida. This population was developed through 124 controlled crosses of 148 selected parents made in February 2011. Seeds were extracted from the resulting mature fruit, cold stratified for 5 mo, and planted in November 2011 as a family in 2-L pots in a greenhouse. One hundred seedlings from each family were transplanted to trays in January 2012 and planted in May 2012 in a high-density (~20,000 plants per 0.2 ha) nursery at the Plant Science Research and Education Unit in Citra, Florida. In May 2013, five to 32 plants were selected from each family, and the remaining plants were removed from the planting, leaving 1996 plants.

The pedigree of the University of Florida blueberry breeding program was assembled using internal pedigree records, the NCGR–Corvallis Vaccinium Catalog (National Center for Genome Resources–Corvallis, 2012), and the Brooks and Olmo Register of Fruit and Nut Varieties (Brooks and Olmo, 1952, 1997). The file was ordered chronologically beginning with the founding parents in 1908. The pedigree file contains 7755 lines with annotated pedigree since 1908.

Phenotyping
Phenotypic data collected for selected seedlings in 2014 and 2015 included plant yield, fruit weight, fruit diameter, fruit firmness, and fruit stem scar diameter. Yield was measured using a 1-to-5 rating scale, where 1 equaled none to very few berries on the plant and 5 equaled a yield comparable with commercial cultivars of the same age plant. Yield ratings were recorded over a 2-wk period before the blueberries were fully mature.

Fruit traits were determined using five randomly sampled representative berries from each genotype. Because of the diversity of harvest timing among the individual selections, the berries were collected over a 6-wk period from the beginning of April to the middle of May in each year. The total sample of berries was hand harvested from each genotype when they were fully mature and had no insect or visual damage. After harvest, the berries were kept in a cooler and stored at 4°C overnight prior to analysis. Weight (g) was measured on each individual berry using an analytical scale (CP2202S, Sartorius Corp.). The same five berries were oriented equatorially for fruit size diameter (mm) and firmness (g mm⁻¹ compression force) measurement using a FirmTech II (BioWorks Inc.). The minimum and maximum force thresholds were set at 50 g and 350 g, respectively.

Subsequently, each of the berries was placed with the stem scar positioned upward on a tray in a light box with a digital SLR (Pentax K-x, Ricoh Imaging) camera positioned 50 cm above the berries. The camera was set at a shutter speed of 1/50, aperture F 5.6, and ISO 200. A ruler was placed in each image as a size reference. Subsequently, the images were uploaded into FIJI (Schindelin et al., 2012) and the scale was set using the ruler. Scar diameter (mm) was measured and recorded for each fruit. The average weight (g), diameter (mm), firmness (g mm⁻¹), and scar diameter (mm) of the five berries was calculated and used for subsequent analyses.

Analysis and Model Comparison
Variance components were estimated with REML and BVs with BLUP using ASReml version 3.0 (Gilmour et al., 2009) with the following univariate linear mixed models:

\[
y = \mathbf{1_n} \mu + \mathbf{X}_b \mathbf{b} + \mathbf{Z}_c \mathbf{c} + \mathbf{Z}_r \mathbf{r} + \mathbf{Z}_a \mathbf{a} + \mathbf{Z}_e \mathbf{e} + \epsilon
\]

where \(y\) is a vector of the trait being analyzed (i.e., yield, weight, firmness, diameter, or scar), \(\mathbf{1}\) is a vector of ones, \(\mu\) is the overall mean, \(b\) is the fixed effect of year, \(c\) is the random effect of column nested within year \(\sim \mathcal{N}(0, \sigma^2_c)\), \(r\) is the random effect of row nested within year \(\sim \mathcal{N}(0, \sigma^2_r)\), \(a\) is the random effect of genotype \(\sim \mathcal{N}(0, \sigma^2_a)\), \(b \times a\) is the random effect of the interaction of year \(\times\) genotype \(\sim \mathcal{N}(0, \sigma^2_{ba})\), \(c\) is the random residual effect \(\sim \mathcal{N}(0, \sigma^2_e)\), \(\mathbf{X}, \mathbf{Z}_c, \mathbf{Z}_r, \mathbf{Z}_a\) are incidence matrices for year, column, row, genotype and year \(\times\) genotype interaction, respectively. The unbiased heritability was calculated as \(h^2 = \sigma^2_g/(\sigma^2_g + \sigma^2_e + \sigma^2_r + \sigma^2_{ba} + \sigma^2_{bg})\). The \(A\) matrix was constructed with the AGHMatrix R package as described below.

Matrix Construction with AGHMatrix
To build the pedigree-based relationship matrices, we constructed an R package named AGHMatrix. The package is freely available for download at https://github.com/prmunoz/AGHMatrix. To build the relationship matrix derived from the pedigree using the AGHMatrix package, the pedigree information needs to be formatted in three columns (individual, Parent 1, and Parent 2) and read as a table in R (using read.table or read.csv functions). The pedigree file should be sorted from older to newer generations, thus the oldest generation (founders) without pedigree information should appear at the top of the file, while the current generation being phenotyped is at the bottom of the file. Before proceeding to build the relationship matrix, the AGHMatrix verifies that the above order is correct, and if not, it will flag and permute the order of misplaced individuals. More details of the package’s functions can be found in the manual at https://github.com/prmunoz/AGHMatrix/blob/master/vignettes/tutorial.pdf.

The autotetraploid pedigree-based relationship matrix \(\mathbf{A}_w\) was calculated as presented in Slater et al. (2013). The algorithm first computes the matrix \(K\)
considering the proportion of parental gametes that are identity by descent as a result of double reduction \( w \). Thus for every individual \( k \) with parents \( s \) and \( d \),

If \( s \) and \( d \) of \( a \) are unknown, they are assumed unrelated, and

\[
\begin{align*}
  k_{ij} &= k_{ji} = 0; & \text{for } j = 1:(i-1) \\
  k_{ii} &= (1+w)/4
\end{align*}
\]

If only \( s \) is known, \( d \) is assumed nonrelated, and

\[
\begin{align*}
  k_{ij} &= k_{ji} = 0.5(k_{ij}); & \text{for } j = 1:(i-1) \\
  k_{ii} &= [5 + 7w + 4k_{is}(1-w)]/24
\end{align*}
\]

If \( s \) and \( j \) are known, then

\[
\begin{align*}
  k_{ij} &= k_{ji} = 0.5(k_{ij} + k_{ji}); & \text{for } j = 1:(i-1) \\
  k_{ii} &= [1 + 2w + (1-w)k_{ss} + (1-w)k_{dd} + 3k_{sd}]/6
\end{align*}
\]

Finally, the matrix \( A_w \) is given by \( A_w = 4K \). Six \( A_w \) matrices were constructed from the pedigree assuming varying levels of double reduction \( w \): 0, 0.05, 0.10, 0.15, 0.20, and 0.25. Additionally, we also built the traditional diploid pedigree-based relationship matrix (A-matrix) using the AGHmatrix R package, which is calculated according to the recursive method presented in Mrode (2014) and described by Henderson (1976). The AGHmatrix R package can also build the genomic-base relationship matrix, G-matrix, for diploids using either the method proposed by VanRaden (2008) or from Powell et al. (2010). To calculate the G-matrix, molecular data needs to be in table format where \( n \) is the number of individuals (in rows) and \( m \) is the number of loci (in columns). The loci information is set to \(-1, 0, 1\), for the genotypes aa, Aa, and AA, respectively. Although the G-matrix capability is present in the package, the genomic matrix was not built or used in this study.

The autotetraploid matrices with different levels of double reduction were used in the linear mixed model described above. These autotetraploid models were compared against the diploid counterpart using the Akaike information criteria (AIC). As the level of double reduction \( w \) is unknown, we inferred its level based on the model that maximized the likelihood of the data given the parameters of the linear mixed model.

**Results and Discussion**

**Using AGHmatrix to Determine Trait-Specific Levels of Double Reduction**

A unique challenge to plant breeders dealing with autotetraploid species has been the determination of the degree of double reduction and its impact on heritability and BV estimates. This challenge has been exacerbated with the lack of molecular markers and genetics resources in most autotetraploid species. Therefore, the AGHmatrix package presented in this study may be the solution for many autotetraploid breeders to obtain an estimation of double reduction in their populations.

Using blueberry as an example, we demonstrated the method to achieve such goals. We first built seven relationship matrices (A-matrices) base on the pedigree: one assuming a diploid system and the other six assuming an autotetraploid system with different levels of double reduction ranging from 0 to 0.25. Each of these matrices was fit in the same mixed model equation using the same phenotypic data for each of the traits considered. We then obtained the AIC value from the REML. The AIC difference between the disomic and tetrasomic model (Fig. 1) was used to indicate the model that best fit the data.

Yield, weight, and diameter had similar negative slopes, each decreasing more dramatically with increasing double reduction. The pattern of scar was similar but with a marginal decrease from double reduction varying from 0 to 0.25. Firmness was the only trait that had the best fit when double reduction was larger than zero; the fitness of the model was maximized when double reduction ranged from 0.15 to 0.20. The tetraploid model assuming \( w = 0 \) was the best fit for four of the five traits. The absence of double reduction for the majority of traits is similar to the findings of Krebs and Hancock (1989) when studying blueberry based on segregation of a limited number of isoenzyme loci in *V. corymbosum*. The most plausible explanation for the null double reduction comes from the physical barrier imposed by the size of the chromosomes. Blueberries have small chromosomes (1.5–2.5 \( \mu m \)), which could limit multivalent formation, an essential step for double reduction to occur (Krebs and Hancock, 1989; Hall and Galletta, 1971). Also, cytogenetic studies of meiotic pairing in *V. corymbosum* found the majority of pairing to be bivalent with few quadrivalent formations (Jelenkovic and Harrington, 1971; Jelenkovic and Hough, 1970; Vorsa and Novy, 1995; Qu et al., 1998). Qu et al. (1998) found that >90% of the chromosomes formed bivalents in a highbush cultivar and a wild tetraploid selection. Jelenkovic and Hough (1970) found similar results when studying three *V. corymbosum* cultivars, where up to four quadrivalents were found in a pollen cell, but most cells had only one. Additionally, preferential pairing lowers the expected frequency of multivalent formation between homologous chromosomes (Sybenga, 1996). Preferential pairing has been reported in *V. corymbosum* and *V. corymbosum × V. darrowii* hybrids (Draper and Scott, 1971; Vorsa and Novy, 1995).

While the above explanations are feasible for the zero double reduction found for four of the five traits, they do not explain the high double reduction estimated for firmness. However, a similar study in autotetraploid potato found that the model best estimating the proportion of double reduction varied between 0 and 0.25 depending on the trait considered (Slater et al., 2013). Double reduction is expected to differ from trait to trait because it is a position-dependent phenomenon.
fluctuating between chromosomes based on the frequency of multivalent formation and within chromosomes depending on where the measured loci reside (Wu et al., 2001). Bourke et al. (2015) found a parallel increase between the rate of double reduction and the distance from the centromere using a high-density set of markers in autotetraploid potato. The rate of double reduction increases toward telomeres because there is a greater probability of a crossover to occur between the centromere and the loci (Bourke et al., 2015; Luo et al., 2006; Welch, 1962; Wu et al., 2001). The high proportion of double reduction for firmness suggests that genes controlling this trait may reside toward the distal end of a chromosome. Thus, in the same line, the genes controlling yield, weight, diameter, and scar may be located closer to the centromere, be quantitative in nature and dispersed over the whole genome, or located on chromosomes that do not exhibit much multivalent formation.

**Impact of Increasing Double Reduction Levels on Heritability and Relationship Estimates**

Heritability estimations were not significantly different when assuming tetrasomic inheritance with zero double reduction ($w = 0$) instead of disomic inheritance for all traits (Table 1). However, firmness was the only trait that showed a better fit with the tetrasomic model considering higher levels of double reduction, thus the heritability for the tetrasomic model was estimated using $w = 0.15$, which was the level of double reduction that fit the data best (Table 1). The heritabilities estimates for firmness had the largest difference between the two models, decreasing from 0.40 to 0.35 as a result of the larger proportion of double reduction used in the estimation. As seen in Fig. 2, there

---

**Table 1. The heritability of five breeding traits calculated assuming disomic and tetrasomic inheritance. Tetrasomic inheritance was estimated assuming no double reduction for all traits except firmness where 0.15 was used.**

<table>
<thead>
<tr>
<th>Trait</th>
<th>Disomic inheritance</th>
<th>Tetrasomic inheritance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Heritability</td>
<td>Standard error</td>
</tr>
<tr>
<td>Yield</td>
<td>0.46</td>
<td>0.02</td>
</tr>
<tr>
<td>Weight</td>
<td>0.58</td>
<td>0.02</td>
</tr>
<tr>
<td>Firmness†</td>
<td>0.40</td>
<td>0.03</td>
</tr>
<tr>
<td>Fruit diameter</td>
<td>0.26</td>
<td>0.03</td>
</tr>
<tr>
<td>Scar diameter</td>
<td>0.54</td>
<td>0.02</td>
</tr>
</tbody>
</table>

† Tetrasomic inheritance was calculated assuming double reduction of 0.15.
is a small difference between the disomic-estimated heritability for firmness and the tetrasomic-estimated heritability with no double reduction. The difference becomes more pronounced, and the narrow-sense heritability for the trait decreases as the proportion of double reduction increases. This trend was seen in all the traits measured (data not shown). Comparing disomic and tetrasomic inheritance with \( w = 0.10 \) in autotetraploid potatoes, Slater et al. (2013) observed no consistent trend and found the difference in heritability ranged from 0.01 to 0.13 depending on the trait being analyzed.

The downward trend in heritability observed for blueberry fruit firmness can be explained by the increased inbreeding caused by increasing the levels of double reduction. As double reduction increased, the kinship coefficient also increased (Table 2). The maximum coefficient was determined by the amount of double reduction, whereas the minimum kinship coefficient was unaltered by the level of double reduction. A large number of the individuals in the pedigree had unknown ancestry, which negatively skewed the distribution. Unknown ancestry was a result of missing information, unspecified clones, and wild selections. These individuals were assumed to be unrelated and thus unaffected by double reduction. As double reduction increased the kinship coefficient increased as well. This can be seen by comparing the kinship coefficients obtained for \( w = 0 \) and those obtained for \( w = 0.25 \) (Table 2). The effects of double reduction are cumulative with larger kinship coefficients more susceptible to its effects. In the first quartile, the difference between the kinship coefficients estimated assuming \( w = 0 \) and the coefficient estimated assuming \( w = 0.25 \) was 0.021, while this difference almost doubles (0.046) in the third quartile.

The BLUP analysis uses these kinship coefficients and estimations of heritability to predict BVs. Thus, inaccurate estimations of double reduction could lead to bias predictions. Higher relationships should suffer a larger impact as these are more affected by double reduction.

![Fig. 2. The estimated narrow-sense heritability for firmness assuming disomic inheritance and tetrasomic inheritance with increasing coefficients of double reduction.](image)

### Table 2. Average of the relationship coefficient between two individuals assuming various levels of double reduction.

<table>
<thead>
<tr>
<th>Double Reduction</th>
<th>Min.</th>
<th>First quartile</th>
<th>Median</th>
<th>Mean</th>
<th>Third quartile</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.067</td>
<td>0.106</td>
<td>0.115</td>
<td>0.148</td>
<td>1.000</td>
</tr>
<tr>
<td>0.05</td>
<td>0</td>
<td>0.071</td>
<td>0.113</td>
<td>0.123</td>
<td>0.158</td>
<td>1.050</td>
</tr>
<tr>
<td>0.10</td>
<td>0</td>
<td>0.076</td>
<td>0.120</td>
<td>0.130</td>
<td>0.168</td>
<td>1.100</td>
</tr>
<tr>
<td>0.15</td>
<td>0</td>
<td>0.080</td>
<td>0.126</td>
<td>0.138</td>
<td>0.177</td>
<td>1.150</td>
</tr>
<tr>
<td>0.20</td>
<td>0</td>
<td>0.084</td>
<td>0.133</td>
<td>0.145</td>
<td>0.186</td>
<td>1.200</td>
</tr>
<tr>
<td>0.25</td>
<td>0</td>
<td>0.088</td>
<td>0.139</td>
<td>0.151</td>
<td>0.194</td>
<td>1.250</td>
</tr>
</tbody>
</table>
Nevertheless, there was a strong correlation ($r = 0.999$) between the diploid and tetraploid ($w = 0.15$) estimated BVs for firmness (Fig. 3). This strong correlation was seen for all traits when analyzed with tetraploid models assuming $w = 0$ (data not shown). Similarly, Slater et al. (2013) found a strong correlation (>0.95) for eight of the nine traits when comparing the diploid estimated BVs with the autotetraploid estimated BVs. The depth of the pedigree could help balance and compensate erroneous estimations, resulting in small changes to the BVs over various levels of double reduction. In this case, the pedigree covered the whole breeding program beginning with its origin in 1908.

**Effect of Pedigree Depth on Double Reduction Estimate**

In the example data set we have used, there is an extensive pedigree available beginning with known founder clones dating back to 1908 (Coville, 1937). However, the detail of pedigree information available for blueberry may not be available for other autotetraploid plant breeders. To study the consequences of using truncated or incomplete pedigree information while studying double reduction using these methods, we truncated the pedigree to 5, 10, 20, and 30 yr (back to 2009, 2004, 1994, and 1984, respectively).

Different levels of double reduction maximize the fitting of the data when using different pedigree depths (Fig. 4). By using only 5 yr of data (back to 2009) the level of double reduction is estimated to be the maximum coefficient tested in this study, 0.25, for four out of five traits, and 0.20 for the last trait. In this case, all the traits had the worst fit assuming no double reduction. When 10 yr of pedigree information was used, three traits followed the same pattern as when using 5 yr of pedigree information, no double reduction was estimated for one trait, while the fit for the final trait was maximized at a double reduction level of 0.05. As more pedigree information was added, the slopes continued to change, but even when 30 yr of pedigree information were used, it was not enough to observe the estimation of double reduction seen when all the pedigree information was used. Thus, shallow pedigrees underestimate the actual kinship coefficient between two individuals. For example Boyce (1983) found that shallow pedigrees miscalculate the level of inbreeding when analyzing 10 Standardbred stallions with 30 generations worth of pedigree information. In blueberries, there has been a steady increase in inbreeding with each consecutive generation as a result of recurrent selection (Brevis et al., 2008; Hancock and Siefker, 1982; Ehlenfeldt, 1994). Relationships are further intensified by the narrow gene pool created in the founding event. Based on pedigree information, Ehlenfeldt (1994) concluded that three cultivars, Brooks, Sooy, and Rubel, have contributed the largest portion of genes in
79 highbush cultivars (23.5, 12.5, and 28.6%, respectively) and V. darrowi contributions can be traced back to germplasm sources Florida 4B, Florida 4A, Florida 9A, and clone D (Brevis et al., 2008). The limited pedigree information cannot accurately account for these intertwined relationships. Thus, the higher proportions of double reduction compensated for the lack of information by increasing the assumed relationships between the individuals and thus better fitted the model. Atkin et al. (2009) and Mehrabani-Yeganeh et al. (1999) found the accuracy of estimating additive variance components and BVs improved when using more pedigree information. Atkin et al. (2009) discovered that the variances were estimated with the same precision once five generations were used, suggesting older generations do not contribute much information to the model. This is mainly because the coefficients of relationships with the current population become smaller as generation distance increases. However, using shallow pedigrees can lead the erroneous conclusions in the estimation of relationship coefficients and thus incorrect determination of the level of double reduction for each trait. This will also impact the estimated heritability, inbreeding, and BVs.

**Conclusion**

We have developed an R package, AGHmatrix, that can be used to quickly develop relationship matrices for autotetraploid crop species. To illustrate the utility of this package, we used the pedigree and phenotypic data from a blueberry breeding program. The majority of traits (yield, weight, diameter, and scar) had the best fit assuming tetrasomic inheritance with no double reduction. The only exception was firmness, which had the best fit when double reduction was 0.15. Changes in the estimation of heritability occurred with larger proportions of double reduction. However, there was no significant improvement in BV estimates for the blueberry traits assessed when comparing tetrasomic inheritance accounting for the correct levels of double reduction with
a simple disomic inheritance model. The effects of double reduction are cumulative, and shallow pedigree depths can underestimate the true kinship coefficient and, as a result, report inflated levels of double reduction, which can affect the estimation of the genetic parameters. As double reduction is trait dependent, this package should allow autotetraploid breeders to estimate the levels of double reduction and how it affects genetic parameters and ultimately breeding progress.

Acknowledgments

This material is based on work that is supported by the National Institute of Food and Agriculture, USDA under award no. 2014-67013-22418 to PRM, JWO, and JBE and by the Science without Borders/CAPES scholarship (proc. No. 88888.020298/2013-00) program from Brazil awarded to RRA.

References


