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

**Corresponding author:**

Patricio Ricardo Muñoz (p.munoz@ufl.edu)

# Pedigree-based relationship matrices

### **Model A2**

It is a diploid pedigree-based relationship computed as presented in Mrode (2014) as described by Henderson (1976). The algorithm computes the matrix $A$ with elements $a\_{ij}$ considering the proportion of parental gametes that are identity by descent. Thus, for every individual $a\_{i}$ with parents $m$ (mom) and $d$ (dad):

If $m$ and $d$ of $a\_{i}$ are unknown, they are assumed unrelated, and

$a\_{ij}=a\_{ji}=0$; for every $j$ from 1 to $i-1$

$$a\_{ii}=1$$

 If only $m$ is known, $d$ is assumed nonrelated, and

$a\_{ij}=a\_{ji}=0.5(k\_{jm})$; for every $j$ from 1 to $i-1$

$$a\_{ii}=1$$

If $m$ and $d$ are known, then:

$a\_{ij}=a\_{ji}=0.5(a\_{jm}+a\_{jd})$; for every $j$ from 1 to $i-1$

$$a\_{ii}=1+0.5(a\_{md})$$

### **Model A4**

It is an autotetraploid pedigree-based relationship matrix computed as presented in Kerr et al. (2012). The algorithm first computes the matrix $K$ with elements $k\_{ij}$ considering the proportion of parental gametes that are identity by descent. As we do not have evidence of double-reduction phenomenon in blueberries, we did not consider it proportion in the matrices, i.e., $w=0$ in the original Kerr et al. (2012). Thus, for every individual $k\_{i}$ with parents $m$ (mom) and $d$ (dad):

If $m$ and $d$ of $k\_{i}$ are unknown, they are assumed unrelated, and

$k\_{ij}=k\_{ji}=0$; for every $j$ from 1 to $i-1$

$$k\_{ii}=1/4$$

 If only $m$ is known, $d$ is assumed nonrelated, and

$k\_{ij}=k\_{ji}=0.5(k\_{jm})$; for every $j$ from 1 to $i-1$

$$k\_{ii}=(5+4k\_{mm})/24$$

If $m$ and $d$ are known, then:

$k\_{ij}=k\_{ji}=0.5(k\_{jm}+k\_{jd})$; for every $j$ from 1 to $i-1$

$$k\_{ii}=(1+k\_{mm}+k\_{dd}+3k\_{md})/6$$

 Finally, the $A$matrix is given by $A=4K$**.**

# Marker-base matrices

They are marker-base relationship matrices derived from VanRaden (2008) and Aguilar et al. (2011). It is noteworthy Model G4 is also used in Ashraf et al. 2016).

### **Model G2**

The marker matrix $M$with $m$ markers and $n$ genotypes has values corresponding to gene content of the second allele where the elements are set to -1, 0, and 1, for the homozygote, heterozygote, and other homozygote, respectively. A centered matrix $Z=M-P$, where $p\_{i}=2(p\_{i}-0.5)$, is built. Then, the relationship matrix $A$is given by:

$$A=\frac{ZZ^{'}}{h}$$

Being $h$ is a scale factor equal and $h=2\sum\_{i=1}^{m}p\_{i}(1-p\_{i})$.

### **Model G4**

The marker matrix $M$ has values corresponding to gene content of the second allele (0, 1, 2, 3, and 4). $Z$is the mean-centered $M$. Then, the relationship matrix $A$is given by:

$$A=\frac{ZZ^{'}}{h}$$

Being $h$ is a scale factor equal where $h=\sum\_{i=0}^{m}s\_{i}^{2}$ and $s\_{i}^{2}$ is the variance of the vector $z\_{i}$ (centered marker vector).

### Model Gr

The marker matrix $M$ has values corresponding to the genotypic ratio $\#A/(\#A+\#a)$, where$ \#A$ is the allele count of the alternative allele and $\#a$ is the allele count of reference allele. No dosage calling was performed in this moded. $Z$is the mean-centered $M$. Then, the relationship matrix $A$is given by:

$$A=\frac{ZZ^{'}}{h}$$

Being $h$ is a scale factor equal where $h=\sum\_{i=0}^{m}s\_{i}^{2}$ and $s\_{i}^{2}$ is the variance of the vector $z\_{i}$ (centered marker vector).

# Pseudo-r code to generate the matrices

**----------------------------------------------------------------------**

## Installing and loading AGHmatrix package ##

install.packages("AGHmatrix"); library(AGHmatrix)

## Loading the data

pedigree <- read.table("pedigree.csv")

markers\_diploid <- read.table("SNPs\_diploid.csv")

markers\_tetraploid <- read.table("SNPs\_tetraploid.csv")

markers\_ratio <- read.table("SNPs\_ratio.csv")

## Building A2 matrix

A2 <- Amatrix(ped, ploidy=2)

## Building A4 matrix

A4 <- Amatrix(ped, ploidy=4)

## Building G2 matrix

G2 <- Gmatrix(markers\_diploid, ploidy=2)

## Building G4 matrix

G4 <- Gmatrix(markers\_tetraploid, ploidy=4)

## Building Gr matrix

Gr <- Gmatrix(markers\_ratio, ratio=TRUE)

**----------------------------------------------------------------------**

# References

Aguilar I, Misztal I, Legarra A, Tsuruta S (2011) Efficient computation of the genomic relationship matrix and other matrices used in single‐step evaluation. J. Anim. Breed. Genet. 128: 422-428

Amadeu RR, Cellon C, Olmstead JW, Garcia AA, Resende MF et al. (2016) AGHmatrix: R Package to construct relationship matrices for autotetraploid and diploid species: a blueberry example. Plant Genome 9: 3

Ashraf BH, Byrne S, Fé D, Czaban A, Asp T et al. (2016) Estimating genomic heritabilities at the level of family-pool samples of perennial ryegrass using genotyping-by-sequencing. Theor. Appl. Genet. 129: 45-52

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

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

Mrode RA (2014) Linear models for the prediction of animal breeding values. Cabi, Boston, Massachusetts

VanRaden PM (2008) Efficient methods to compute genomic predictions. J. Dairy Sci. 91: 4414-4423