

## Be-Breeder - Learning: a new tool for teaching and learning plant breeding principles

Roberto Fritsche-Neto<sup>1\*</sup> and Filipe Inácio Matias<sup>1</sup>

Crop Breeding and Applied Biotechnology  
16: 240-245, 2016  
Brazilian Society of Plant Breeding.  
Printed in Brazil  
<http://dx.doi.org/10.1590/1984-70332016v16n3n36>

**Abstract:** *The Be-Breeder application is an on-line tool constructed through the R software for the purpose of assisting in some of the main genetic and statistical analyses related to the area of plant breeding. In addition, Be-Breeder provides a section called “Learning”, which in a simple click-point manner allows explanation of theories related to the effect of inbreeding, population structure, qualitative and quantitative traits, heterosis, population size, effect of selection, and composition of hybrids. Be-Breeder is available for network use on the website of the Allogamous Plant Breeding Laboratory (Laboratório de Melhoramento de Plantas Alógamas) of ESALQ-USP through the link: <http://www.genetica.esalq.usp.br/alogamas/R.html>.*

**Key words:** *Heterosis, quantitative genetics, inbreeding, population structure, selection.*

### INTRODUCTION

Most Agronomy courses have mandatory subjects related to plant breeding. This determination of mandatory subjects by CONFEA (CONFEA 1973) and the Ministry of Education (MEC 2006) in the course curriculum is based on the importance of this topic for the education/training of the agronomist and for agriculture. For example, it is estimated that half of the increase in yield of the main agricultural crops has come about through plant breeding processes (Kuhr et al. 1985, Byerlee and Moya 1993). Facts of this nature have also led to the creation of numerous graduate study programs in this line of research. The aim of these programs is to train human resources in understanding the mechanisms and concepts in regard to this theme, as well as possible applications of the principles that guide plant breeding.

In this context, seeking to assist teaching and learning, different tools, such as computer applications or software, can be used in clarification of more complex theories and equations. For this purpose, the Be-Breeder application, a compilation of on-line routines visualized through the Shiny package of the R software (Chang et al. 2015), has a section called *Learning*, with interactive learning strategies. They are an easily applied alternative tool for explaining theories related to the effect of inbreeding, population structure, qualitative and quantitative traits, heterosis, population size, effect of selection, and composition of hybrids.

### BE-BREEDER APPLICATION

The Be-Breeder application is teaching tool freely available for on-line use at

**\*Corresponding author:**  
E-mail: roberto.neto@usp.br

**Received:** 20 July 2016  
**Accepted:** 25 July 2016

<sup>1</sup> USP/ESALQ – Genetics, Av. Pádua Dias, 11, CP 83, 13418-900, Piracicaba, SP, Brazil

<http://www.genetica.esalq.usp.br/alogamas/R.html>. In the *Learning* section, the user has interactive tabs that deal with themes related to quantitative genetics, population genetics, and breeding methods. The tabs available and suggestions on how to use them are provided below.

### Inbreeding Effect

Inbreeding is the process of raising the frequency of gene loci in homozygosity by means of successive processes of crosses between individuals with kinship, with maximum expression through self-pollination in plants (Nass 2001, Bos and Caligari 2007). This allows the formation of pure lines that have various applications (Shull 1909, Johannsen 2014).

In dealing with the inbreeding process in plants and reporting reduction in the percentage of loci in heterosis by half in each self-pollination cycle, the user can use visualization of histograms for a single locus with two alleles: “A” and “a”. Upon choosing generation of self-pollination from  $F_1$  to  $F_9$  and the F.Inf ( $\infty$ ), it can be observed that at each cycle the percentage of the genotype “Aa” is cut in half, beginning at 100% in  $F_1$  to 0% in F.Inf; in contrast, the percentage of homozygous genotypes “AA” and “aa” increases from zero in generation  $F_1$  to 50% in generation F.Inf.

In this tab, the user can also observe the fluctuation in additive ( $\sigma_a^2$ ) and dominant ( $\sigma_d^2$ ) variance between and within populations as a function of the inbreeding coefficient ( $F$  of Wright). Additive variance between is estimated by the expression  $\sigma_{aA}^2 = 2F\sigma_a^2$ , additive variance within by  $\sigma_{aW}^2(1 - F)\sigma_a^2$ , dominant variance between by  $\sigma_{dA}^2(1 - F)\sigma_d^2$ , and dominant variance within by the expression  $\sigma_{dW}^2(1 - F)\sigma_d^2$  (Nass 2001).

### Qualitative vs. Quantitative

The number of genes and their different forms of intra-allelic interaction influence the number of phenotypic classes and also the frequency of each class in the population. Consequently, in the total number of genes that control the trait, as well as the types of interactions, concepts arise and transition of qualitative traits to quantitative ones (Ramalho et al. 2008, Borém and Miranda 2013). In this respect, a simple algorithm was developed in Be-Breeder in which the user can construct a genetic structure of a trait, choosing the number of genes with dominance effect, the number of genes with partial dominance effect, and the number of genes with additive effect. As a response, a histogram relating frequency and number of classes can be visualized in the output box. It should be emphasized that, just as in natural biological systems, the randomness factor was embedded in the algorithm in such a way that, although the number of genes chosen is the same, it will not indicate that the number of classes will necessarily be the same for polygenic effects, allowing inferences to be made regarding segregation.

### Progeny Size

Observation of a determined genotype is dependent on the number of genes acting on a trait, such that the greater the number of genes, the more individuals are necessary in population sampling to verify all the possible genotypes or the genotype desired. Inbreeding, promoted by self-pollination, has a direct influence on this observation through the fact of increasing loci in homozygosity in the population, reducing genotypic variability. The size of the population to be evaluated is given by the expression  $n^\circ = \frac{\log(1-p)}{\log(1-IH)}$ , in which  $p$  represents the probability of observing a determined genotype, and  $IH$  is estimated by  $IH = \left(\frac{2^{m-1}}{2^m}\right)^n$ , in which  $m$  is the number of generations of self-pollination and  $n$  is the number of genes that control the trait. This number of individuals can easily be obtained in this tab of Be-Breeder, in which the user can simulate numerous scenarios and observe fluctuation in the population size as a function of generation of self-pollination, the number of genes, and the probability of observation.

### Effect of Selection (HWE)

In population genetics, it is relevant to check if a determined gene locus is in Hardy-Weinberg equilibrium (HWE) as a function of the frequency of the alleles  $A(p)$  and  $a(q)$ . From this information, it is possible to determine the number of individuals expected for each genotype using the expressions  $n^\circ AA = (p^2 + pqF)N$ ,  $n^\circ Aa = (2pq - 2pqF)N$ , and  $n^\circ aa = (q^2 + pqF)N$ , in which  $N$  is the total number of individuals of the population and  $F$  is the inbreeding coefficient of Wright. Upon comparing the number expected with the number observed for each genotype, it is possible to perform

a chi-square test ( $\chi^2$ ) to verify significance through the expression  $\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$  for  $i = (AA, Aa, aa)$ , in which  $O_i$  is the number of individuals observed for genotype  $i$  and  $E_i$  is the number of individuals expected for genotype  $i$ .

Nevertheless, the effect of selection is the main piece of information the breeder uses to conduct a breeding program. Thus, upon applying a selection intensity in a population with an original mean value ( $\mu_o$ ), a selected population is obtained with a mean value ( $\mu_s$ ), in which the difference between these mean values is equivalent to the differential of selection ( $DS$ ). By multiplying  $DS$  by the heritability of the trait ( $h^2$ ), gain from selection is obtained ( $GS = DS \cdot h^2$ ), which upon being added to  $\mu_o$  will give rise to the predicted mean of the improved population in the next evaluation cycle (Falconer et al. 1996, Bernardo 2010).

In the application, the user can simulate numerous scenarios, modifying the allelic frequencies as a function of the phenotypic observations, thus allowing speculations in regard to the effect of the composition of the population, heritability of the trait, and the intensity of selection in a breeding population. The output is a table with the following information: allelic frequency, genotypic frequencies, the number of individuals observed, the number of individuals expected, and the  $\chi^2$  test to verify Hardy-Weinberg equilibrium. In addition, the user can simulate selection by identifying the number of individuals selected in each genotypic class and the respective genetic values, obtaining the parameters  $\mu_o$ ,  $\mu_s$ ,  $DS$ , and  $GS$  as output.

### Components of Genetic Variance

Genetic variance is composed of variance of additive and non-additive effects that fluctuate as a function of allelic frequencies (Bernardo 2010). So as to deal with these concepts, a function was developed that allows the user to indicate the frequency of the allele "A" ( $p$ ), which is obtained from the difference with allele "a" ( $q$ ). The user can also modify the additive effect ( $a$ ) and dominance effect ( $d$ ) and observe the effect of these factors ( $p$ ,  $a$ , and  $d$ ) on the magnitude and relationship between total genetic variance and its additive and dominance components. The expressions used to estimate the variances are  $\sigma_a^2 = 2pq[a + d(q - p)]^2$  for the additive,  $\sigma_d^2 = (2pqd)^2$  for the dominance, and  $\sigma_G^2 = \sigma_a^2 + \sigma_d^2$  for the total of a determined gene locus (Falconer et al. 1996).

### Constructing Synthetic Populations

The number of alleles of a gene in a breeding population indicates variability, such that the greater the number of alleles present in the population, the greater the variability of the heterozygotes will be, according to the expression  $\sigma_{gH}^2 = \sum_i (p_i)^2 - \left(\sum_i (p_i)\right)^2$ . However, this does not indicate that the number of heterozygotes in the population will increase indefinitely; that is, retention of heterozygosity (RH) reaches a plateau with a determined number of alleles. Although this value increases toward 1.00, the number of heterozygotes in the population will remain constant, just as indicated by the expression  $RH = 1 - \sum_i (p_i)^2$  (Nietlisbach et al. 2016). It is known that the objective is to maintain heterozygosity at high levels in an allogamous population, so as to exploit heterosis and avoid inbreeding depression (Falconer et al. 1996). Nevertheless, an excessive increase in the number of alleles can increase genetic variability to critical levels, which can impede selection and standardization of the population for important agronomic traits or descriptors.

In this regard, in this function the user provides the number of alleles of a gene (maximum of 10 alleles) and their respective initial frequencies in the composition of the population. The sum total should be equal to one. Retention of heterozygosity and the genetic variance of the heterozygotes can be observed in the output window as a result, representing that, although genetic variance increases the number of heterozygotes in the population, this reaches a plateau and stabilizes. Thus, it is possible to identify the ideal composition of parents for formation of a population for numerous situations.

### Recurrent Selection

#### Intrapopulation (IRS)

Intrapopulation recurrent selection is a breeding procedure that leads to an increase in the frequencies of alleles of interest in the population without, however, drastically reducing its variability, improving the performance per se of

the population in each selective and recombination cycle (Bernardo 2010). In this context, Be-Breeder allows estimation of gain from selection (*GS*), effective size of the population (*Ne*), and inbreeding coefficient (*F* of Wright) for the *IRS* breeding arrangement in different selection scenarios, number of progenies evaluated, selection intensity, and heritability.

The expression of response to intrapopulation recurrent selection, according to Falconer et al. (1996), is given by  $GS = \left[ i \cdot c \cdot \frac{(\sigma_a^2 + F \cdot D_1)}{\sigma_p} - \frac{ID}{2Ne} \right]$ , in which *GS* is gain from selection, *i* is the standardized selection differential, *c* and *D*<sub>1</sub> are values that depend on the selection arrangement by parental control (Table 1),  $\sigma_a^2$  is additive genetic variance,  $\sigma_p$  is the phenotypic standard deviation from the unit of selection, and *ID* refers to inbreeding depression given in percentage. The *Ne* parameter refers to the effective size of the population given the expression  $Ne = Ne_{tab} \cdot N^* \cdot i$ , in which *Ne*<sub>tab</sub> is the value dependent on the selection arrangement (Table 1), *N* is the total size of the population, and *i* is selection intensity. The inbreeding coefficient *F* of Wright is estimated by  $F = \frac{1}{2Ne}$ .

**Table 1.** Selection arrangement in regard to intrapopulation recurrent selection for the population of evaluation, population of recombination (HS – half sibs, FS – full sibs, and S1 – self-pollination), *c* index, effective size (*Ne*<sub>tab</sub>), and coefficient between additive and dominance effects of the homozygotes (*D*<sub>1</sub>)

Evaluation	Recombination	<i>c</i>	<i>Ne</i> <sub>tab</sub>	<i>D</i> <sub>1</sub>
HS	HS	¼	4	0
HS	S <sub>1</sub>	½	1	0
FS	FS	½	2	0
FS	S <sub>1</sub>	½	1	0
S <sub>1</sub>	S <sub>1</sub>	1	1	0.5
S <sub>2</sub>	S <sub>2</sub>	3/2	2/3	5/4

### b) Reciprocal (RRS)

Reciprocal or intrapopulation recurrent selection is a breeding arrangement that leads to an increase in complementarity between two heterotic groups or populations from crosses. The *RRS* brings about superior hybrids by crossing these groups in each selection cycle, in which intragroup selection and recombination of the most complementary parents increases the frequency of favorable alleles within each group (Bernardo 2010). Thus, as for *IRS*, in Be-Breeder it is also possible to simulate different selection arrangements, number of progenies evaluated, selection intensity, and heritability for *RRS*.

For that purpose, the response of reciprocal recurrent selection was estimated by the expression  $GS = i_1 \cdot c \cdot \frac{\sigma_{a1}^2}{\sigma_{p1}^2} + i_2 \cdot c \cdot \frac{\sigma_{a2}^2}{\sigma_{p2}^2}$ , in which *GS* is gain from selection, *i* is the standardized selection differential for each group (1 and 2), *c* is a value that depends on the selective arrangement of parental control (Table 2),  $\sigma_a^2$  is additive genetic variance, and  $\sigma_p$  is the phenotypic standard deviation for each group (1 and 2) (Falconer et al. 1996). The effective size of the population (*Ne*) and inbreeding coefficient *F* of Wright are also provided for each heterotic group (1 and 2) in the output window; they are estimated in a manner similar to that described in the item *IRS*.

### Hybrids (Jenkins)

#### Prediction of Hybrids

Among the main products coming from plant breeding, hybrids stand out for the important role they exercise in the world economy (USDA 2016). Among them, single-cross hybrids (*SH*), three-way hybrids (*TH*), and double-cross hybrids (*DH*) (Shull 1910, Jones 1918) are the most representative products on the market. In this section, the user of Be-Breeder can find the predicted genotypic value of three-way hybrids (*TH*) and double-cross hybrids (*DH*) through the input of a .txt document containing the mean phenotypic dataset of the single-cross hybrids (*SH*) coming from the lines of interest.

For this purpose, the expressions of Jenkins (1934) are used as a basis, in which  $HT_{(ABC)} = \frac{HS_{(Ac)} + HS_{(Bc)}}{2}$  and  $HD_{(AB)(CD)} = \frac{HS_{(Ac)} + HS_{(Ac)} + HS_{(Bc)} + HS_{(Bc)}}{4}$ . The sequence of the columns in the txt file must follow the example indicated in Table 3.

**Table 2.** Selective arrangement in reference to reciprocal recurrent selection for the population of evaluation, the population of recombination (TC – Testcross, HS – half sibs, FS – full sibs, and S1 – self-pollination), *c* index, tabulated effective size ( $Ne_{tab}$ )

Evaluation	Recombination	<i>c</i>	$Ne_{tab}$
HS	S	¼	1
FS	S <sup>1</sup>	¼	1
HS	HS	1/8	4
TC	HS	1/16	4

**Table 3.** Example of .txt file for input in the Be-Breeder application in the Learning section, “Hybrid Effect” tab so as to predict the phenotypic value expected of three-way and double-cross hybrids from the mean phenotypic value observed from the single-cross hybrids between the lines of interest

Lines	A	B	C	D
A	0	15	14	11
B	15	0	18	12
C	14	18	0	10
D	11	12	10	0

### b) Number of Hybrids

In this tab of the application, the user can obtain the possible number of single-cross hybrids (*SH*), three-way hybrids (*TH*), and double-cross hybrids (*DH*), given the number of lines (*n*) of a breeding population. This information can be obtained for a single population through the expressions (Vencovsky and Barriga 1992):  $n^{\circ} HS = \frac{n(n-1)}{2}$ ,  $n^{\circ} HT = \frac{n(n-1)(n-2)}{2}$ , and  $n^{\circ} HD = \frac{n(n-1)(n-2)(n-3)}{8}$ . In dealing with two heterotic groups (1 and 2) or two populations (1 and 2), the number of *SH*, *TH*, and *DH* depends on the number of lines belonging to group 1 (*a*) and the number of lines belonging to group 2 (*b*), according to the expressions:  $n^{\circ} HS = a^*b$ ,  $n^{\circ} HT = a^*(a-1)^*b + b^*(b-1)^*a$ , and  $n^{\circ} HD = a^*(a-1)^*b^*(b-1)$ .

### Genotype x Environment

In plant breeding, the statistical significance of the component of the genotype × environment interaction defines the selection strategy and commercial recommendation. The absence of interaction indicates that the environments of evaluation do not have a different influence on the behavior and on the ordering of the genotypes, and so it is sufficient to choose a single environment because the crop recommendation will be the same. Simple interaction indicates that the genotypes respond differently to environmental influences, but not enough to for there to be change in ordering, maintaining the same commercial recommendation among them. In contrast, in complex interaction, there are changes in ordering of genotypes among the environments, and a crop recommendation per location is necessary (Borém and Miranda 2013). In this context, Be-Breeder provides a tab for simulation of interactions among three genotypes in two environments, and it is possible to construct various scenarios, simulate recommendations, and make inferences regarding the implications of the G × A effect in the selection process and in data analysis. The user has columns that range from zero to five for each genotype/environment combination (genotype value of individual *i* in environment *j*), obtaining the mean values of genotypes and of environments separately as output, as well as graph visualization of each scenario.

### Heterosis

According to the Dominance and Repulsion Hypothesis (Borém and Miranda 2013), the mean performance of hybrids (F1) in relation to the mean of the parents is related to allelic complementarity between the parents (*P1* and *P2*), the genetic divergence between them, and the magnitude of the dominance deviations (Nass 2001). By being observed mainly in quantitative traits, the number of genes they control also has a certain influence on the magnitude of the heterosis observed. In this context, hybrid vigor in F1, also called biological heterosis (*H*), is estimated by the expression  $H = F1 - \frac{P1 + P2}{2}$  or  $H = (p - r)^2d$  (Falconer et al. 1996). Hybrid performance in relation to superior performance (Best Parent - *B.P*), for its part, receives the name heterobeltiosis (*Hb*) or agronomic heterosis, and is estimated by  $Hb = F1 -$

*B.P.* In light of the foregoing, in this tab of the application, it is possible to simulate different scenarios as a function of the number of genes, of deviation of dominance, and of divergence among the parents, making it possible to observe the fluctuation of heterosis and of heterobeltiosis for the simulated hybrids.

## FINAL CONSIDERATION

Through use of the Be-Breeder - Learning application, it can be observed that it provides a teaching interface of the R software, which can be easily handled and which offers an alternative for understanding concepts and expressions of quantitative and population genetics applied in plant breeding. This tool can assist researchers, professors, and students in experiments, predictions, and academic studies, among other applications, both in learning and in teaching the principles that guide plant breeding.

## REFERENCES

- Bernardo R (2010) **Breeding for quantitative traits in plants**. Stemma Press, Woodbury, 400p.
- Borém A and Miranda GV (2013) **Melhoramento de plantas**. UFV, Viçosa, 523p.
- Bos I and Caligari P (2007) **Selection methods in plant breeding**. Springer Science & Business Media, Dordrecht, 461p.
- Byerlee D and Moya P (1993) **Impacts of international wheat breeding research in the developing world, 1966-1990**. CIMMYT, Texcoco, 87p.
- Chang W, Cheng J, Allaire J, Xie Y and McPherson J (2015) Shiny: web application framework for R. **R package version 0.11.1**.
- CONFEA (1973) Resolução Nº 218, de 29 de Junho de 1973. **Diário Oficial da União (DOU)**: 0218-73.
- Falconer DS, Mackay TF and Frankham R (1996) Introduction to quantitative genetics (4th edn). **Trends in Genetics** **12**: 280.
- Jenkins MT (1934) Methods of estimating the performance of double crosses in corn. **Agronomy Journal** **26**: 199-204.
- Johannsen W (2014) The genotype conception of heredity. **International Journal of Epidemiology** **43**: 989-1000.
- Jones DF (1918) The effect of inbreeding and crossbreeding upon development. **Proceedings of the National Academy of Sciences** **4**: 246-250.
- Kuhr S, Johnson V, Peterson C and Mattern P (1985) Trends in winter wheat performance as measured in international trials. **Crop Science** **25**: 1045-1049.
- MEC - Ministério da Educação e Cultura (2006) Resolução Nº 1, de 2 de fevereiro de 2006. **Diário Oficial da União (DOU)** de 03/02/2006 Seção I: 31-32.
- Nass L (2001) **Recursos genéticos e melhoramento-Plantas**. Fundação MT, Rondonópolis, 1183p.
- Nietlisbach P, Keller L and Postma E (2016) Genetic variance components and heritability of multiallelic heterozygosity under inbreeding. **Heredity** **116**: 1-11.
- Ramalho M, Santos JB and Pinto CB (2008) **Genética na agropecuária**. UFLA, Lavras, 463p.
- Shull GH (1909) A pure-line method in corn breeding. **Journal of Heredity** **5**: 51-58.
- Shull GH (1910) Hybridization methods in corn breeding. **Journal of Heredity** **6**: 98-107.
- USDA - United States Department of Agriculture (2016) Gain reports. Available at <<http://gain.fas.usda.gov/Pages/Default.aspx>>. Access on July 28, 2016.
- Vencovsky R and BARRIGA P (1992) **Genética biométrica no fitomelhoramento**. Revista Brasileira de Genética, Ribeirão Preto, 486p.