User Manual for

(Version 2.0)

Last updated on April 4, 2019

Disclaimer: While extensive testing has been performed by Allogamous Plant Breeding Laboratory at University of São Paulo, results are, in general, reliable, correct or appropriate. However, results are not guaranteed for any specific set of data.

Support documents: Extensive support documents, including this user manual, scripts, data, and results, are available at Allogamous Plant Breeding Laboratory website: http://vencovsky.esalq.usp.br:3838/shiny/be-breeder/.

All the source code is publicly available on GitHub at https://github.com/filipematias23/Be-Breeder/.

Question and comments: alogamas@usp.br

Citation:

Matias FI, Granato ISC, Fritsche-Neto R (2018) Be-Breeder: an R/Shiny application for phenotypic data analyses in plant breeding. **Crop Breeding and Applied Biotechnology**. 18(18): 241–243. doi: 10.1590/1984.

Matias FI, Granato ISC, Dequigiovanni G, Fritsche-Neto R (2017) Be-breeder – An application for analysis of genomic data in plant breeding. **Crop Breeding and Applied Biotechnology**. 17(1): 54–58. doi: 10.1590/1984-70332017v17n1n8.

Fritsche R, Matias FI (2016) Be-Breeder - Learning: a new tool for teaching and learning plant breeding principles. **Crop Breeding and Applied Biotechnology**. 16(3): 240–245. doi: 10.1590/1984-70332016v16n3n36.

The Be-Breeder project is supported by the Allogamous Plant Breeding Laboratory at Department of Genetics, Luiz de Queiroz College of Agriculture, University of São Paulo.

Contents

Introduction

What is Be-Breeder?

Be-Breeder 2.0 is a R/Shiny application for statistical and plant breeding analyses developed by Prof. Dr. Roberto Fritsche Neto, Dr. Filipe Inacio Matias, and PhD candidates Júlia Silva Morosini and Fernando Garcia Espolador, who integrate the Laboratory of Allogamous Plant Breeding group - Department of Genetics, "Luiz de Queiroz" College of Agriculture, University of Sao Paulo, Brazil.

Be-Breeder is freely available at http://www.genetica.esalq.usp.br/alogamas/R.html. Implemented in a web application, Be-Breeder does not require advanced mathematical, programming skills, or to be installed on a local computer. In addition, the user can save the numeric and graphical outputs. For the analyses where data input is required, example datasets are provided as text files in the Help option. Be-Breeder 2.0 comprises three main modules: Learning, Phenotypic Breeding, and Molecular Breeding. A wide range of breeding steps is considered, from field evaluation, data management, and molecular markers use to statistical analyses and the choice of breeding methods.

The intellectual property is reserved in periodicals by scientific citations of Be-Breeder papers.

Features

Be-breeder 2.0 provides a deep overview of the whole landscape of breeding scenarios, comprising three main modules:

- Learning;
- Phenotypic Breeding;
- Molecular Breeding.

In the 2.0 version, four major advancements have been made: the inclusion of statistical-genetic analyses for specific genetic designs, such as diallels, considering different proposals to obtain the specific and general combining abilities for genotypes ranking and selection; the incorporation of new features in approaches already present aiming to provide a better understanding, statistically and conceptually; the upgrade throughout all the application, both on interface and codes, regarding a better computational efficiency and clustering of subjects considering the new ones that have been added; and the open-source codes, which are now available online and for download. For the analyses where data input is required, example datasets are provided as text files in the Help option.

In the next topics, we provide a detailed description regarding the concept, the analysis and the parameters required for each feature embedded in the application.

1 Learning

1.1 Inbreeding Effect

Inbreeding is the process of raising the frequency of homozygous loci through successive of crosses between genetic related individuals (Nass, 2001, Bos and Caligari, 2007). This process enables the obtention of pure lines (Shull, 1909, Johannsen, 2014) and the maximum expression of inbreeding is via self-pollination.

In this tab, the user can visualize histograms for a single locus with two alleles: "A" and "a" (Figure 1). Upon choosing generation of self-pollination from F_1 to F_9 and the F.Inf (∞), one can observe 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.

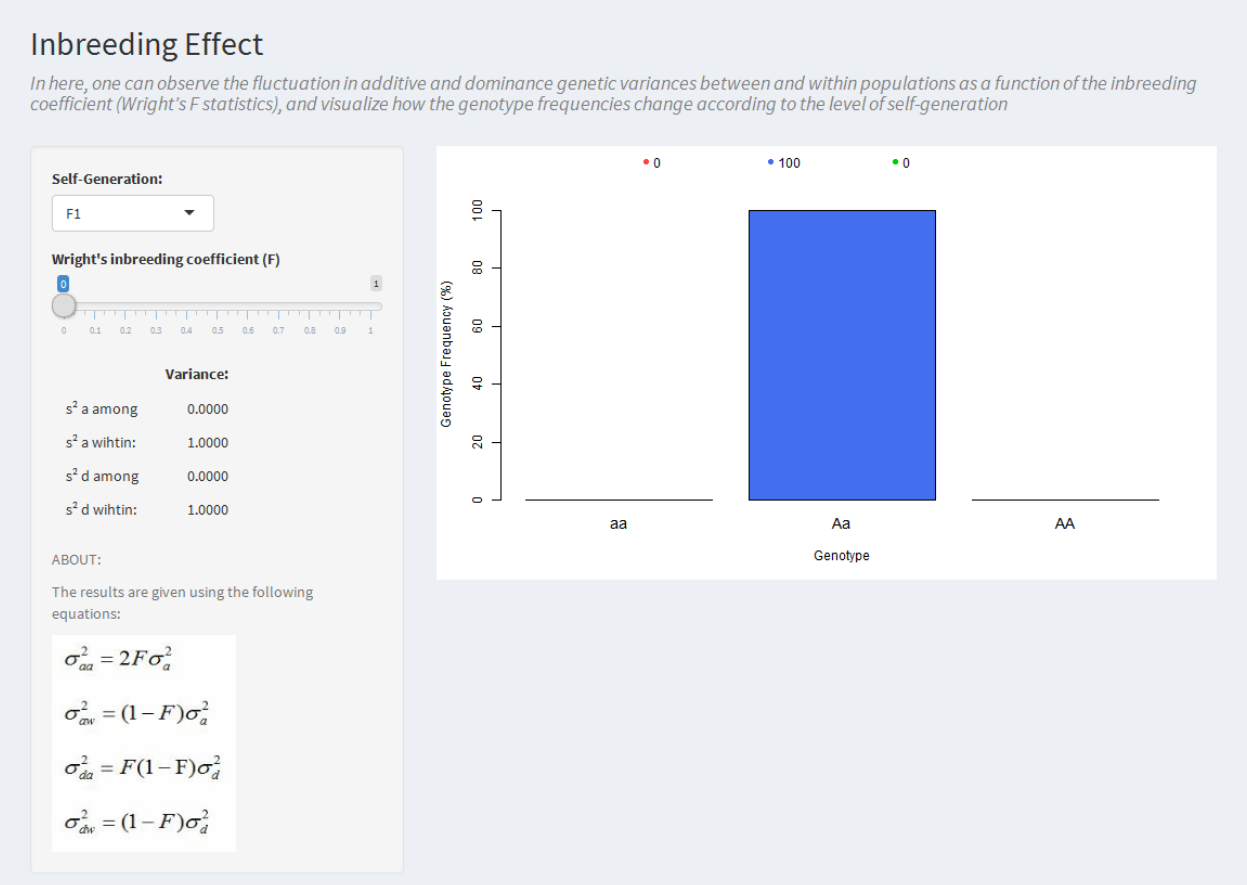


Figure 1. Inbreeding Effect tab from Be-Breeder 2.0

Furthermore, it is possible to observe the fluctuation in additive (σ_a^2) and dominant (σ_d^2) variances between and within populations as a function of the inbreeding coefficient (Wright's F). 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 , $\sigma_{dA}^2 = F(1-F)\sigma_d^2$, and dominant variance within by the expression $\sigma_{dW}^2 = (1-F)\sigma_d^2$ (Nass, 2001).

1.2 Qualitative x 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 context, an algorithm was developed 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 (Figure 2). 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.

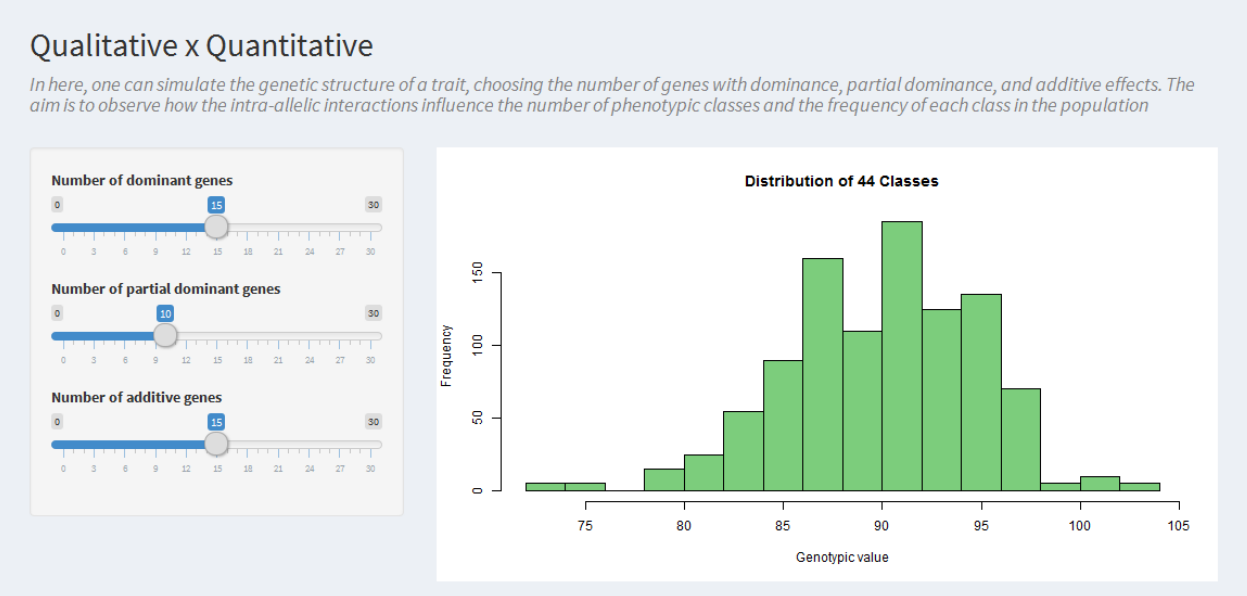


Figure 2. Qualitative x Quantitative tab from Be-Breeder 2.0

1.3 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 $n^0 = \frac{\log(1-p)}{\log(1-IH)}$, 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 (Figure 3), 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.

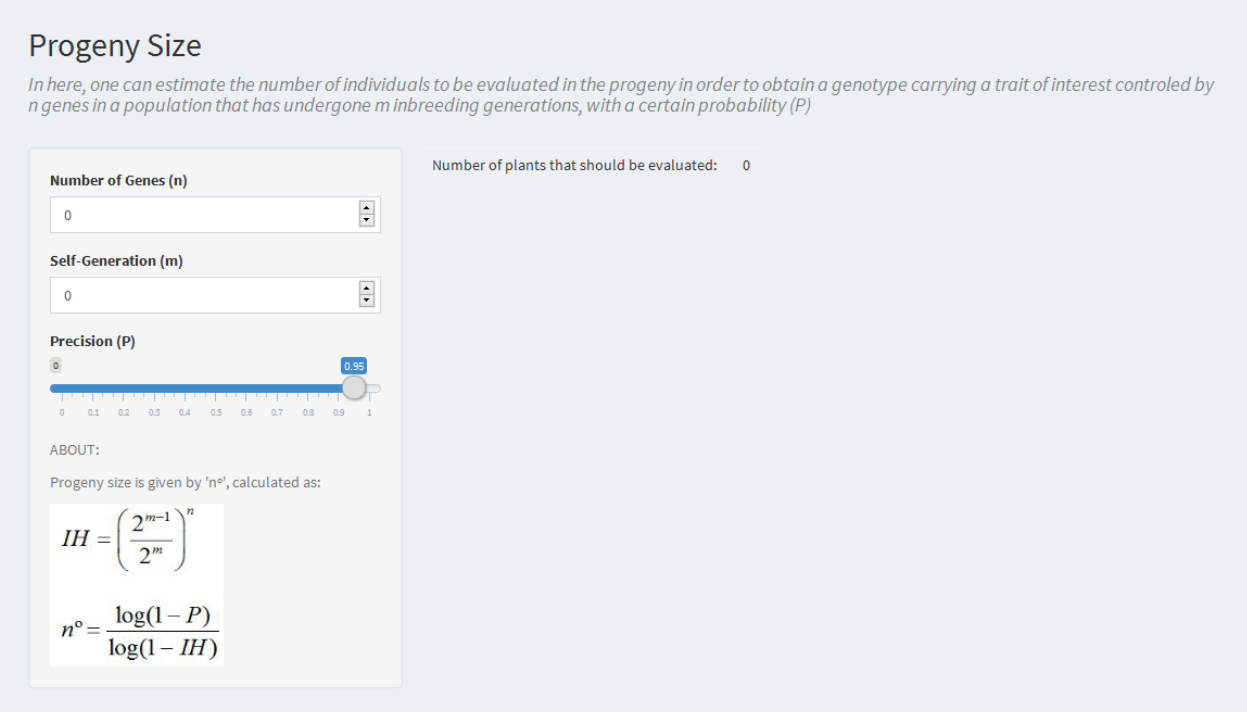


Figure 3. Progeny Size tab from Be-Breeder 2.0

1.4 Selection Effect (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 $n^0AA = (p^2 + pqF)N$, $n^0Aa = (2pq - 2pqF)N$, and $n^0aa = (q^2 + pqF)N$, 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 (χ^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, applying a selection intensity in a population with an original mean value (μ_0), a selected population is obtained with a mean value (μ_s), in which the difference between these mean values is equivalent to the differential of selection (*DS*) (Figure 4). By multiplying DS by the heritability of the trait (h^2), gain from selection is obtained ($GS = DS * h^2$), which upon being added to μ_0 will give rise to the predicted mean of the improved population in the next evaluation cycle (Falconer et al., 1996, Bernardo, 2010).

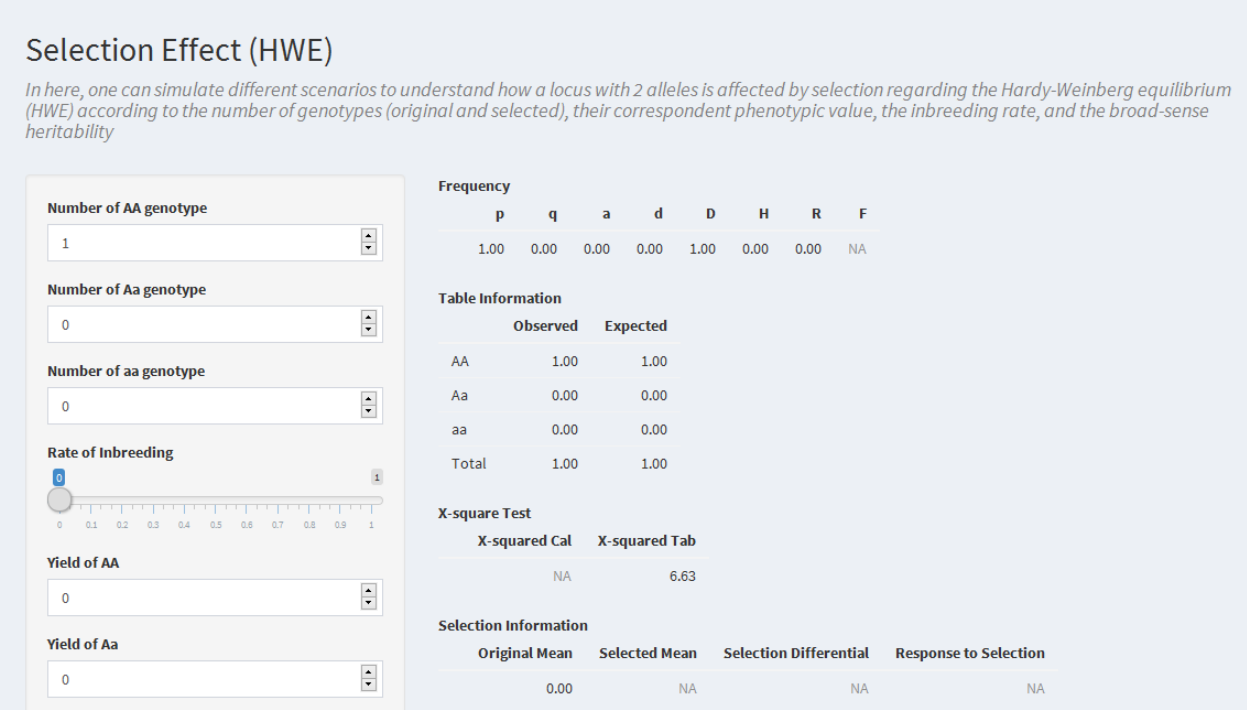


Figure 4. Selection Effect (HWE) tab from Be-Breeder 2.0

In the tab, 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 χ^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 μ_0, μ_s, DS, and GS as output.

1.5 Genetic Variance Components

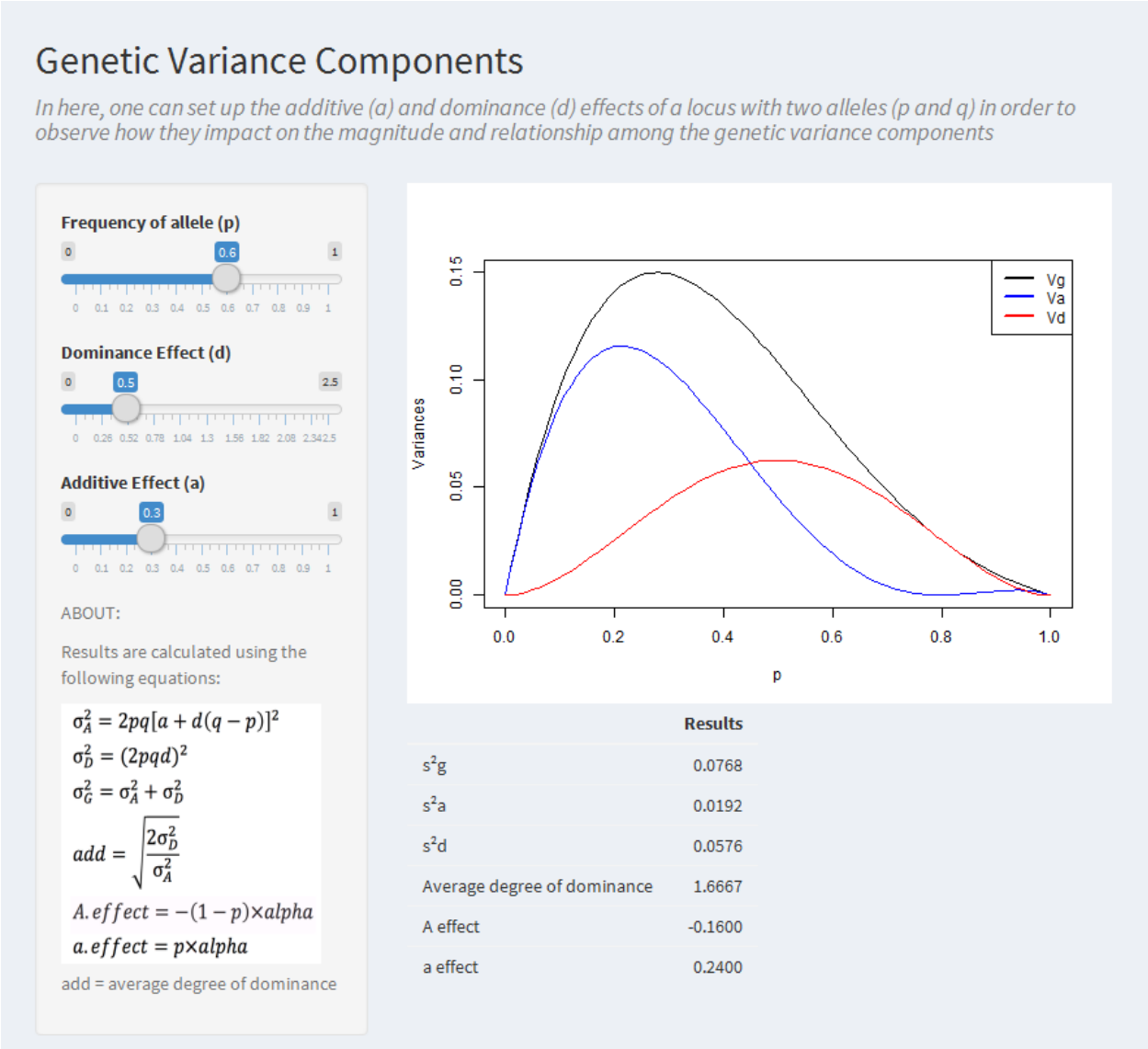


Figure 5. Genetic Variance Components tab from Be-Breeder 2.0

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 (Figure 5). 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).

1.6 Synthetic Pop. Construction

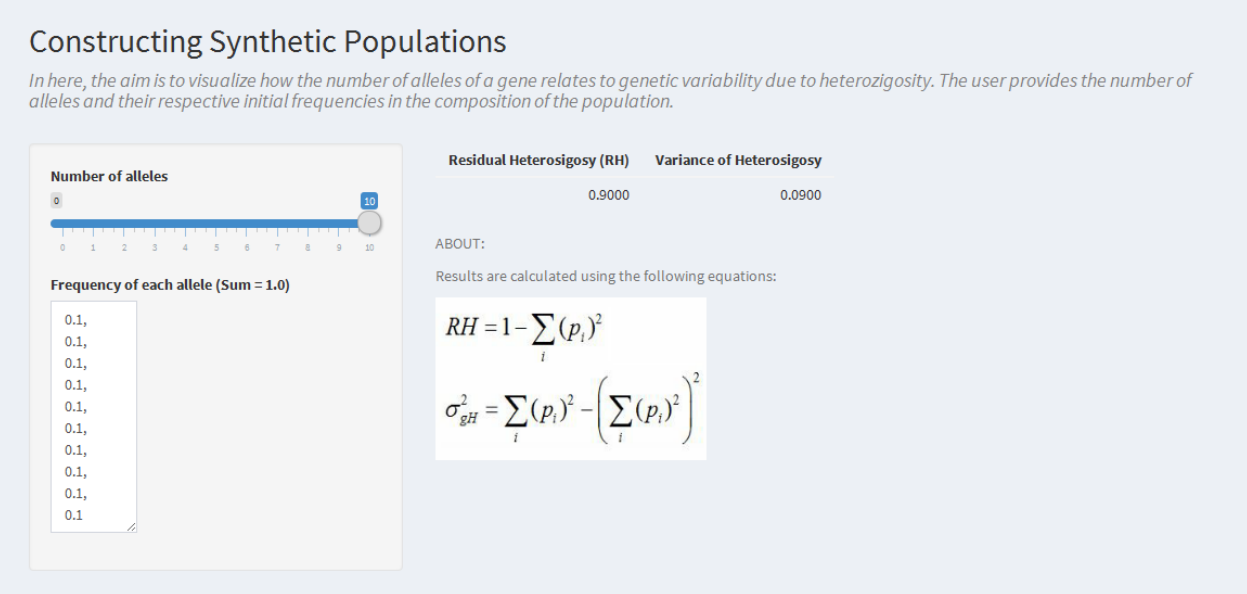


Figure 6. Synthetic Population Construction tab from Be-Breeder 2.0

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 $\sigma_{gH}^2 = \sum_i (p_i)^2 - (\sum_i (p_i)^2)^2$. 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, the aim here is that 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 (Figure 6). Thus, it is possible to identify the ideal composition of parents for formation of a population for numerous situations.

1.7 Recurrent Selection

1.7.1 Intrapopulation

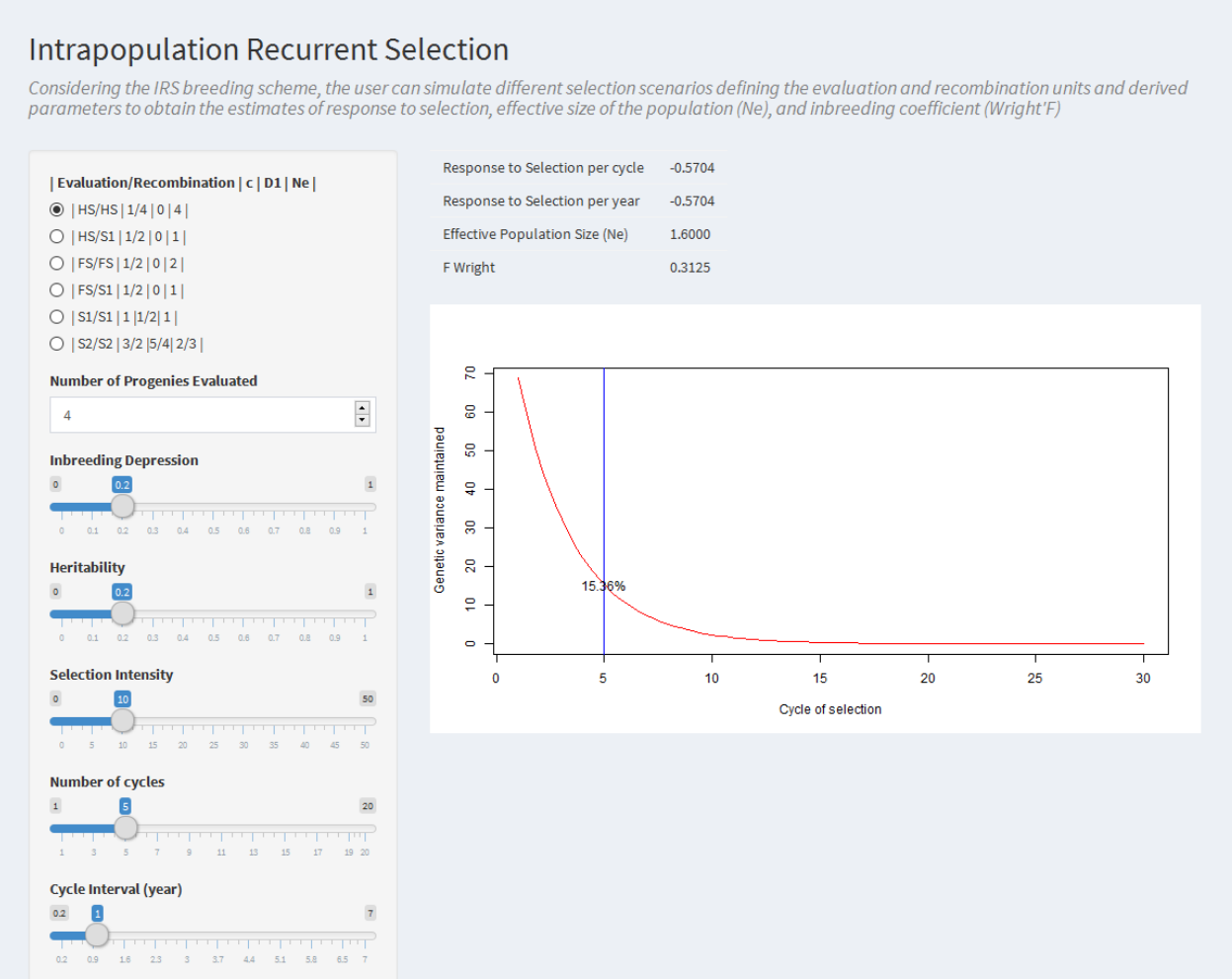


Figure 7. Intrapopulational Recurrent Selection tab from Be-Breeder 2.0

Intrapopulational recurrent selection (IRS) 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 (Figure 7).

The expression of response to intrapopulational recurrent selection, according to Falconer et al. (1996), is given by $GS = [\, i * c * \frac{(\sigma_a^2 + F * D_1)}{\sigma_p} - \frac{ID}{2Ne}]$, in which GS is gain from selection, i is the standardized selection differential, c and D_1 are values that depend on the selection arrangement by parental control (Table 1), σ_a^2 is additive genetic variance, σ_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 $Ne = Ne_{tab} * N * i$, in which Ne_{tab} is the value dependent on the selection arrangement, 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 intrapopulational 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_{tab}), and coefficient between additive and dominance effects of the homozygotes (D_1)

Evaluation	Recombination	c	Ne_{tab}	D_1
HS	HS	1/4	4	0
HS	S_1	1/2	1	0
FS	FS	1/2	2	0
FS	S_1	1/2	1	0
S_1	S_1	1	1	1/2
S_2	S_2	3/2	2/3	5/4

1.7.2 Reciprocal

Reciprocal recurrent selection (RRS) 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, in Be-Breeder it is also possible to simulate different selection arrangements, number of progenies evaluated, selection intensity, and heritability for recurrent reciprocal selection (Figure 9).

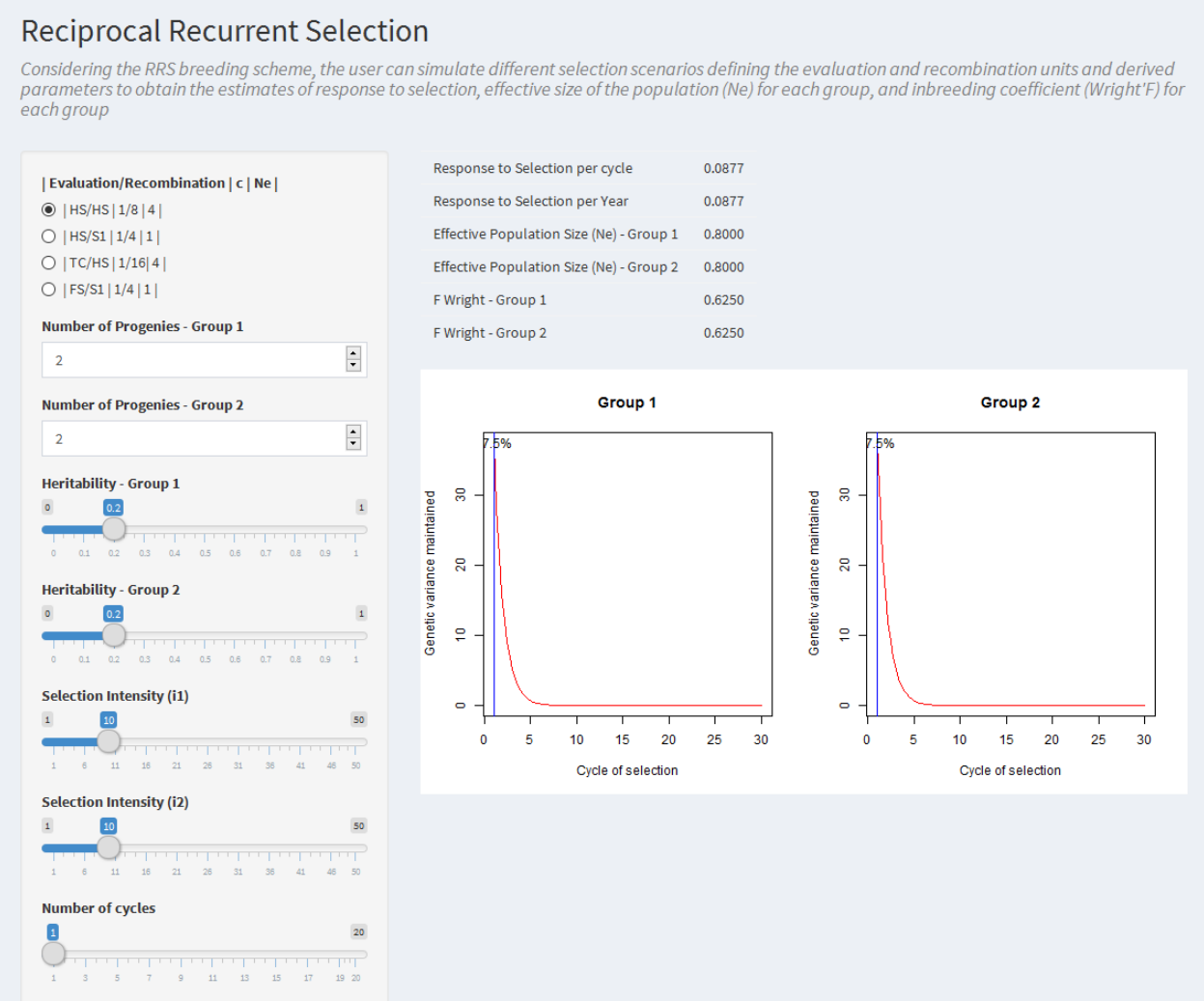


Figure 8. Reciprocal Recurrent Selection tab from Be-Breeder 2.0

For that purpose, the response of reciprocal recurrent selection was estimated by the expression $GS = i_1 * c * \frac{\sigma^2_{a1}}{\sigma^2_{p1}} + i_2 * c * \frac{\sigma^2_{a2}}{\sigma^2_{p2}}$, 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), σ^2_a is additive genetic variance, and σ_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 (Figure 8); they are estimated in a manner similar to that described in the item IRS.

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	C	Ne_{tab}
HS	HS	1/8	4
HS	S_1	1/4	1
TC	HS	1/16	4
FS	S_1	1/4	1

1.8 Hybrids (Jenkins)

1.8.1 Hybrid Prediction

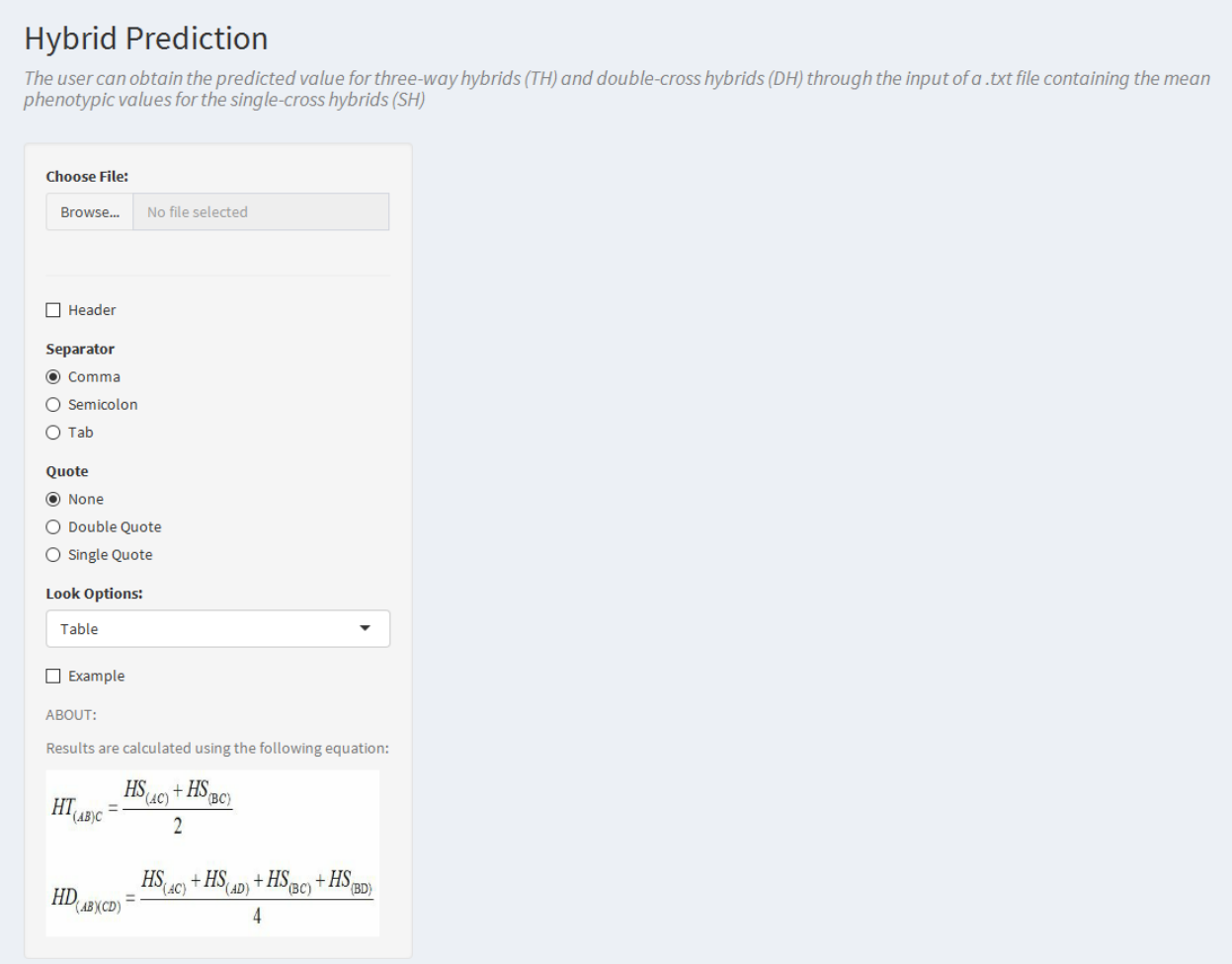


Figure 9. Hybrid Prediction tab from Be-Breeder 2.0

Among the main products coming from plant breeding, hybrids stand out for the important role they exercise in the world economy (USDA 2019). 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 (Figure 9). For this purpose, the expressions of Jenkins (1934) are used as a basis, in which $HT_{(AB)C} = \frac{HS_{(AC)}+HS_{(BC)}}{2}$ and $HD_{(AB)(CD)} = \frac{HS_{(AC)}+HS_{(AD)}+HS_{(BC)}+HS_{(BD)}}{4}$. The sequence of the columns in the *.txt* file must follow the example indicated in Table 3.

Table 3. Example of *.txt* file for input in the Hybrid Effect tab 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

1.8.2 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 (Figure 10). This information can be obtained for a single population through the expressions (Vencovsky and Barriga, 1992): $n^0HS = \frac{n(n-1)}{2}$, $n^0HT = \frac{n(n-1)(n-2)}{2}$, and $n^0HD = \frac{n(n-1)(n-2)(n-3)}{8}$. 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: $n^0HS = a * b$, $n^0HT = a * (a-1) * b + b * (b-1) * a$, and $n^0HD = a * (a-1) * b * (b-1)$.

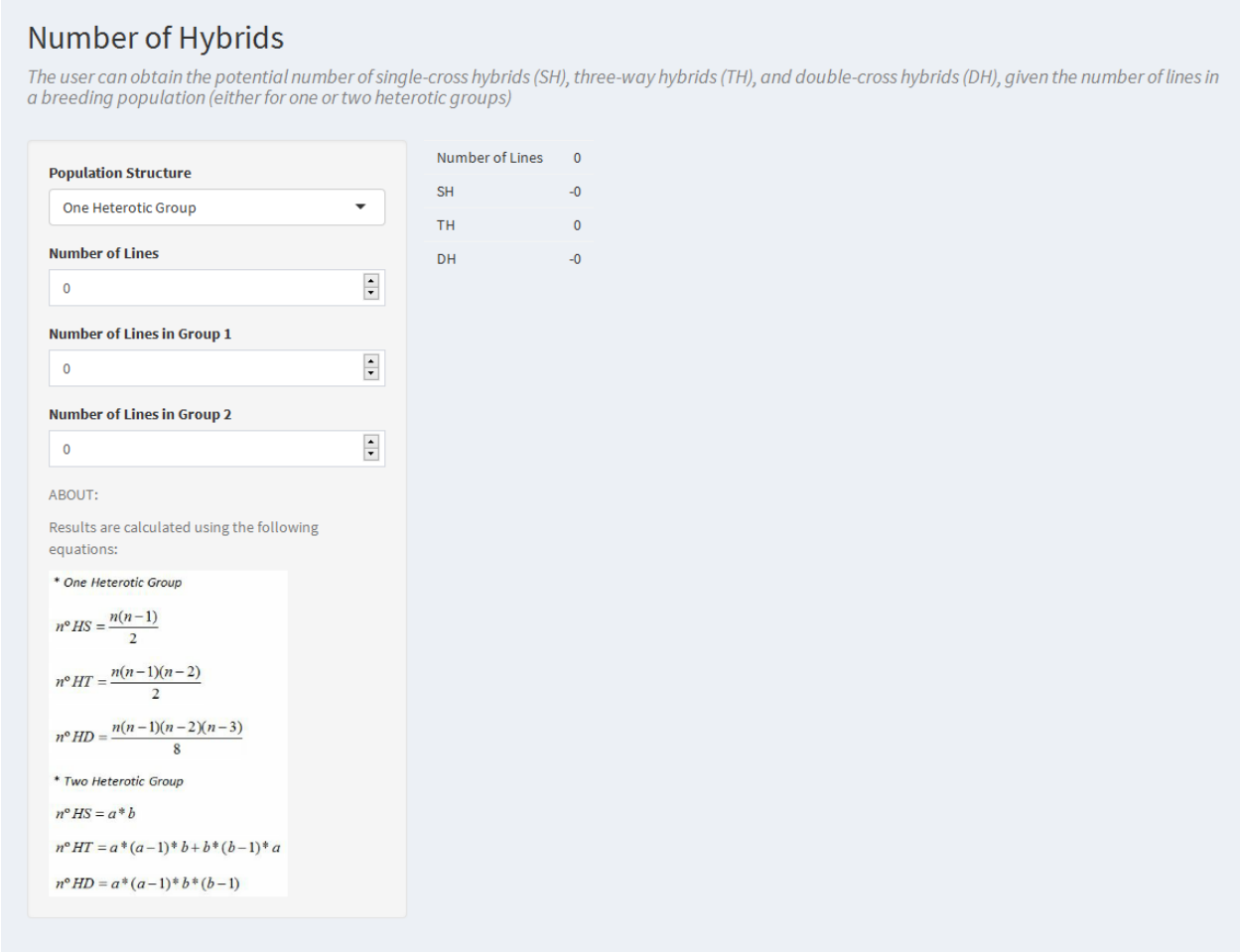


Figure 10. Number of Hybrids tab from Be-Breeder 2.0

1.9 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, this tab provides 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 (Figure 11). 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.

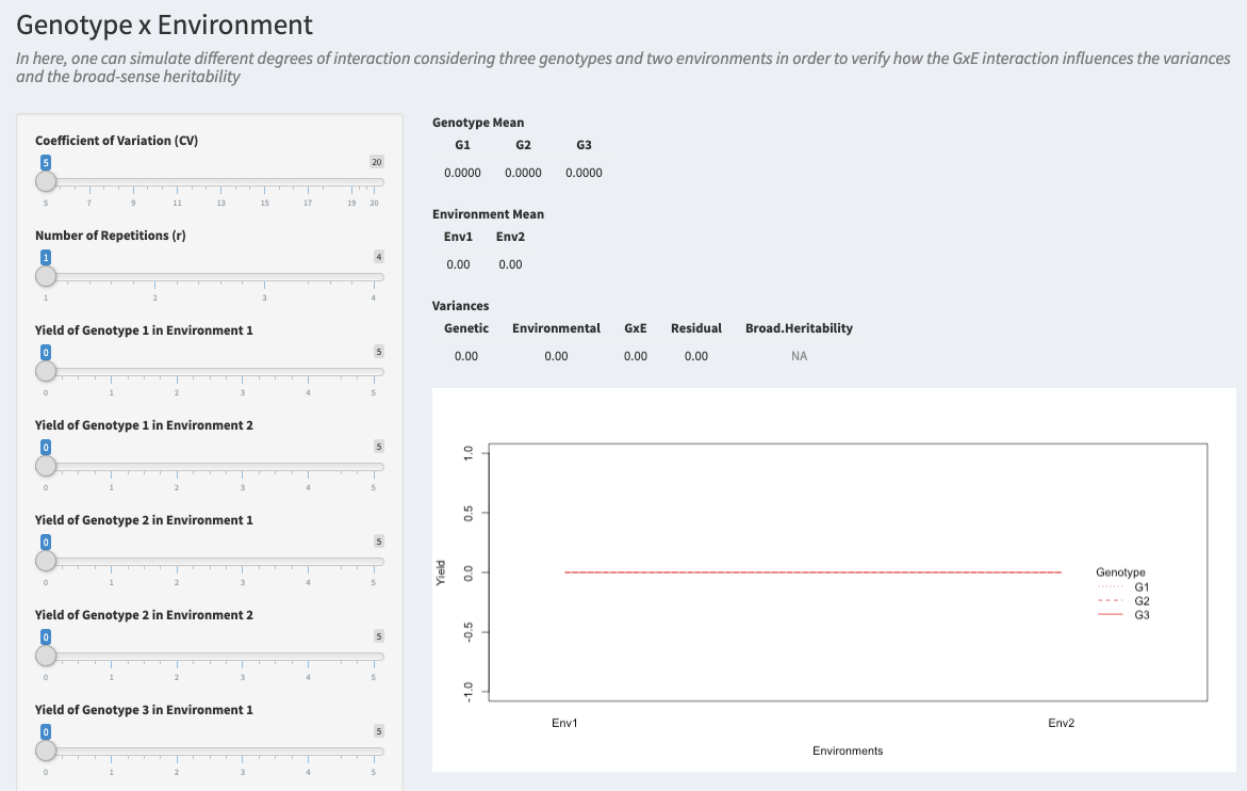


Figure 11. Genotype x Environment tab from Be-Breeder 2.0

1.10 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 $H = F1 - \left(\frac{P1+P2}{2}\right)$ or $H = (p-r)^2 d$ (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 $Hb = F1 - BP$. 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 (Figure 12).

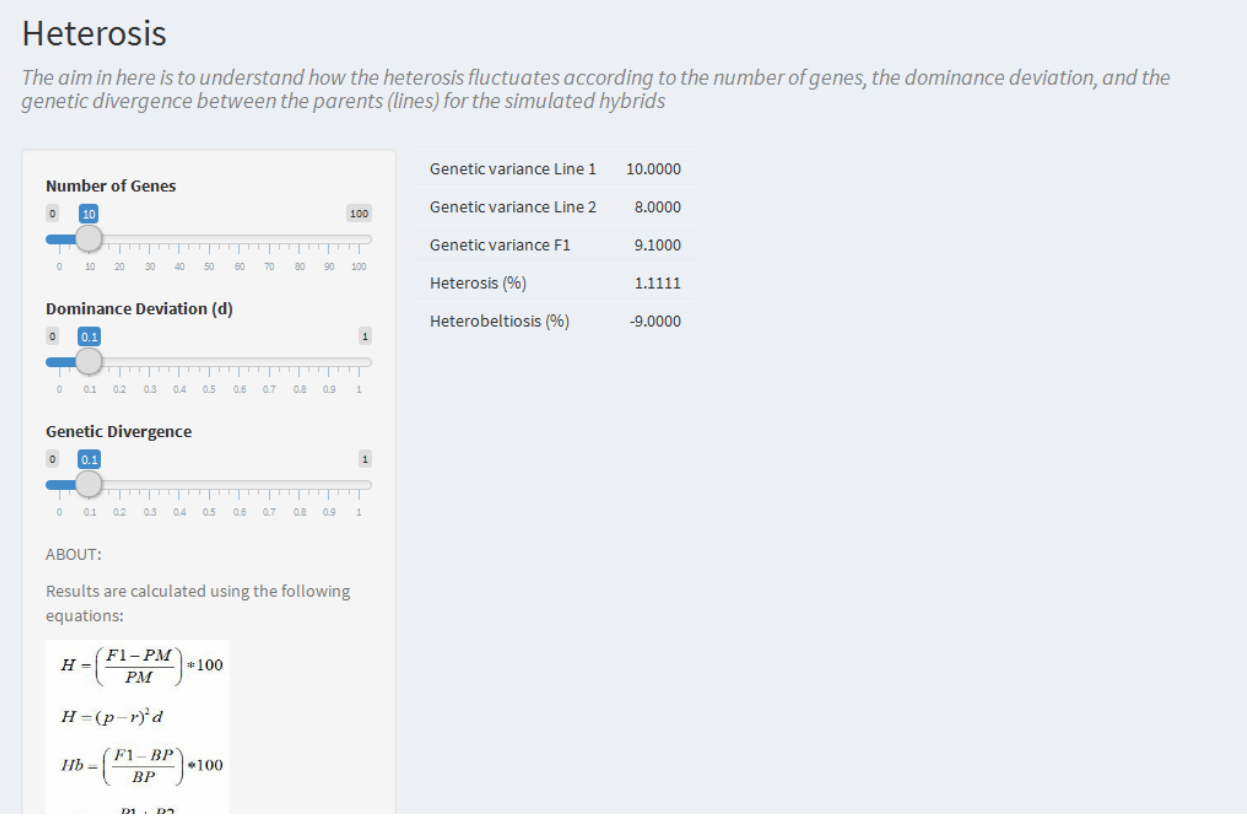


Figure 12. Heterosis tab from Be-Breeder 2.0

1.11 Tester Effect

This topic aims to determine the effect of testers in an experimental condition in order to verify their impacts on genetic variability (GV), where $G = \frac{p(1-p)}{2} * (1 - \mathrm{F}) * (\mathrm{a} + (1 - 2\mathrm{r})\mathrm{d})^2$ (Bos and Caligari, 2008).

In accordance, it is possible to handle the following parameters: the frequency of an allele (p), the accuracy (r), the additive effect (a), the dominance effect (d), and the inbreeding coefficient (F). The results displayed concern the amount of GV and four graphs, each one related to the parameters p, r, d, and F (Figure 13).

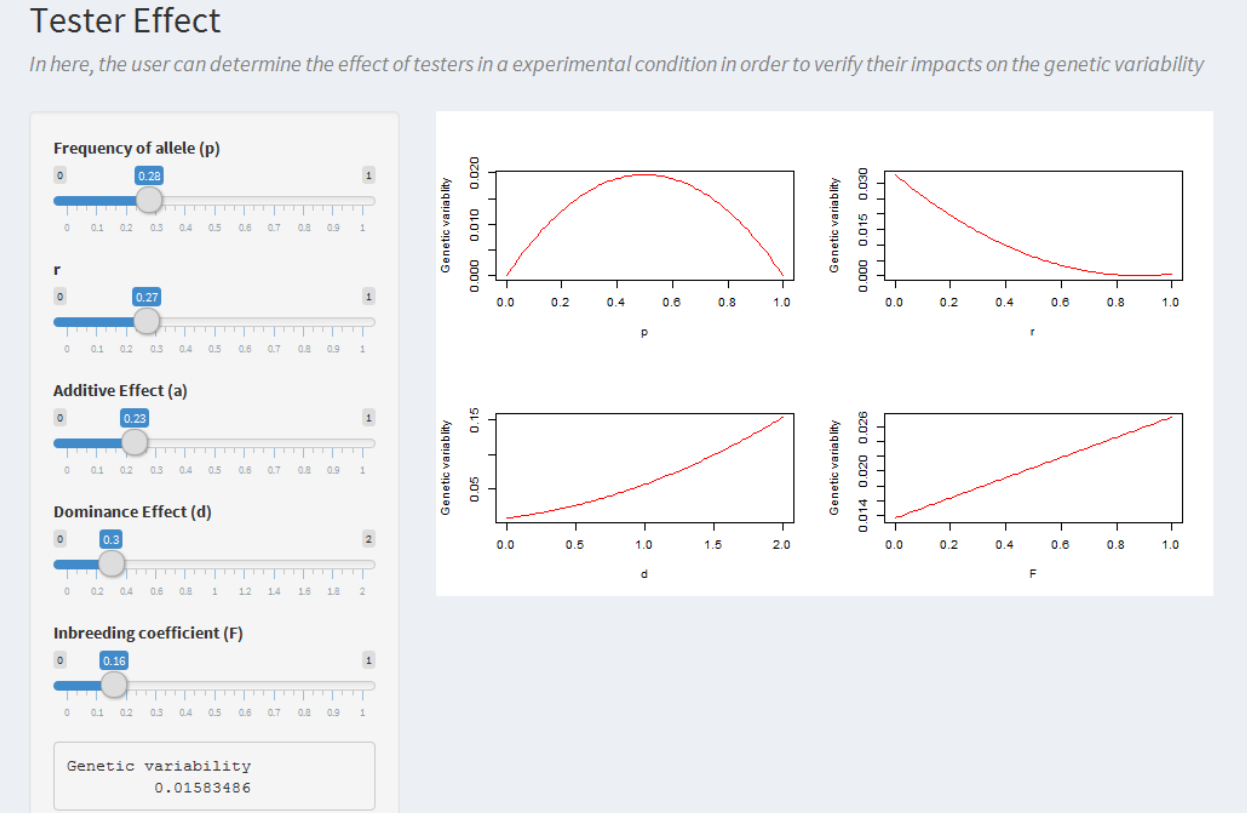


Figure 13. Tester Effect tab from Be-Breeder 2.0

1.12 Genetic Drift

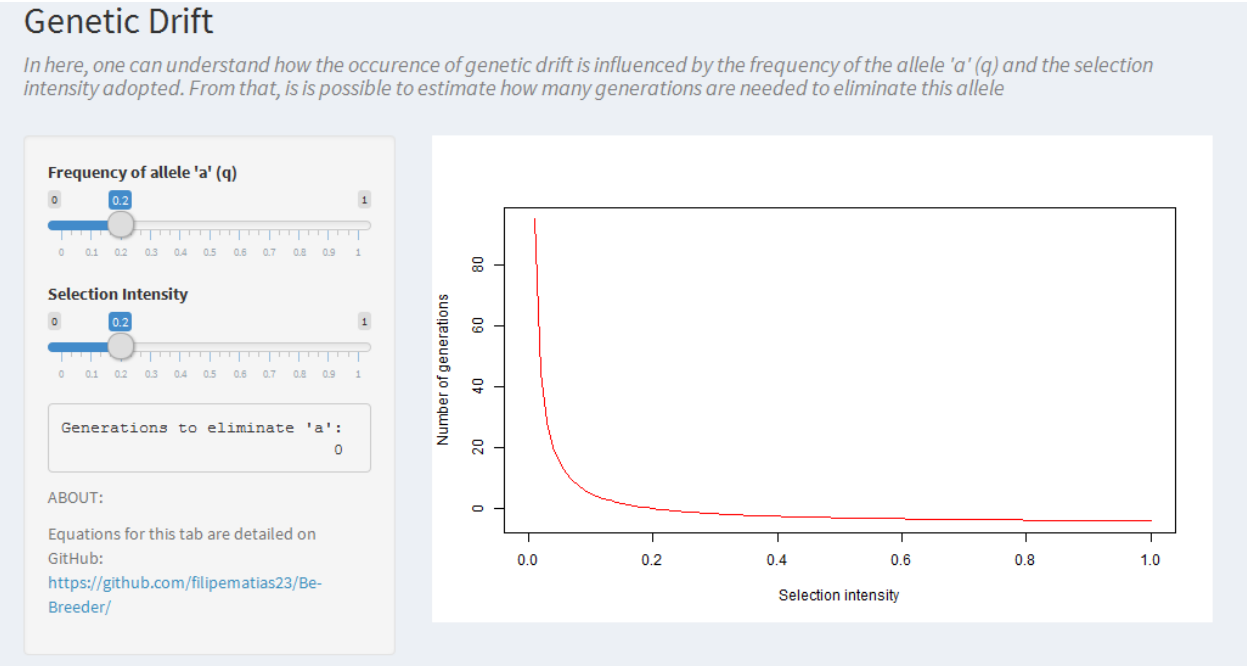


Figure 14. Genetic Drift tab from Be-Breeder 2.0

In here, we analyze the mechanism of evolution related to the change in the frequency of alleles in a population due to a random sampling of individuals, called Genetic Drift. This event has great influence on breeding programs especially when few genotypes are selected for a next step, i.e., a strong selection intensity is adopted. For analyzing this effect regarding the number of generations to eliminate an allele (z), the user may inform the allele frequency (q), and the selection intensity (s) desired, according the equation $z = \frac{q-s}{q*s}$ (Lynch and Walsh, 1998). As output, the app releases the number of generations to eliminate this allele from a given population, and a graph showing the trend of this selection intensity effect (Figure 14).

1.13 Indirect Selection

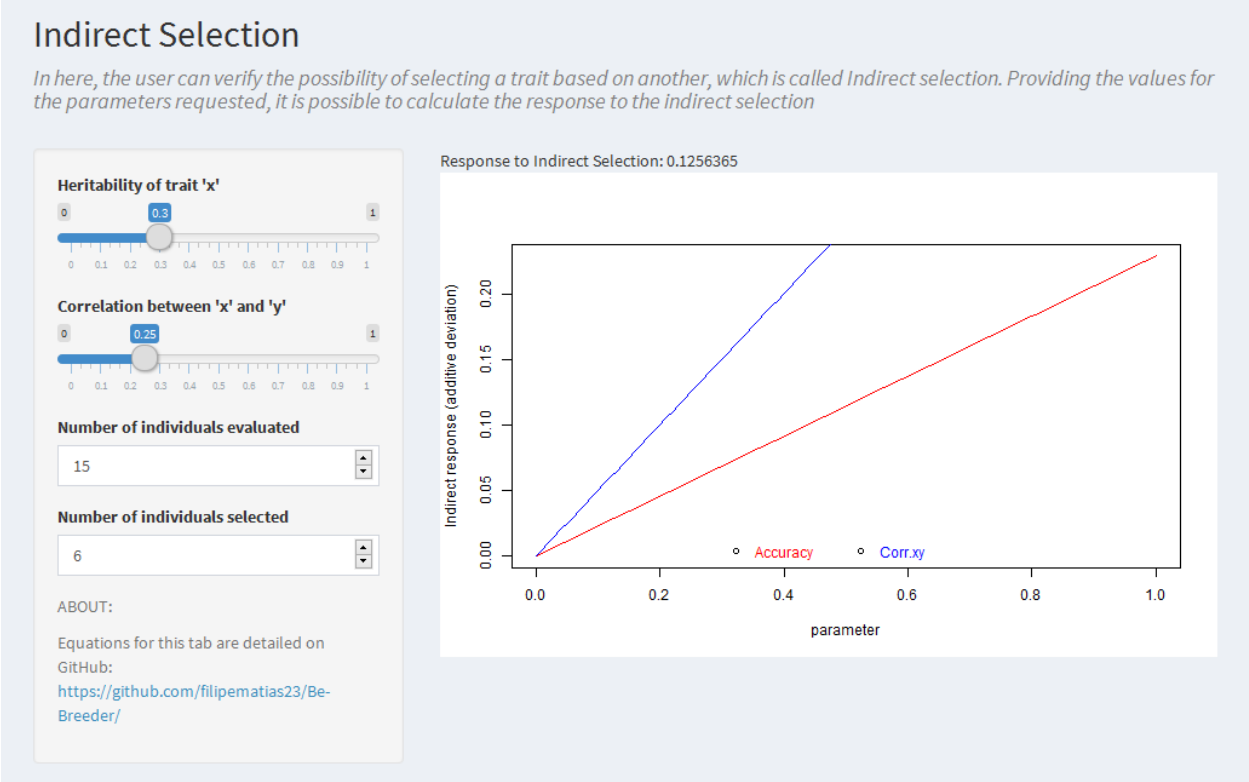


Figure 15. Indirect Selection tab from Be-Breeder 2.0

The objective in this topic is to verify the possibility of selecting a trait based on another, a procedure called Indirect selection. For that, the user must input the heritability of a character of interest x, the correlation between this trait (x) and another (y), the number of individuals evaluated, and the number of individuals selected (Figure 15). Based on that, the application computes the value of response to indirect selection and informs graphically how feasible this response is. The function constructed for this analysis is entirely available on GitHub (https://github.com/filipematias23/Be-Breeder/).

1.14 Residual Effect on Selection

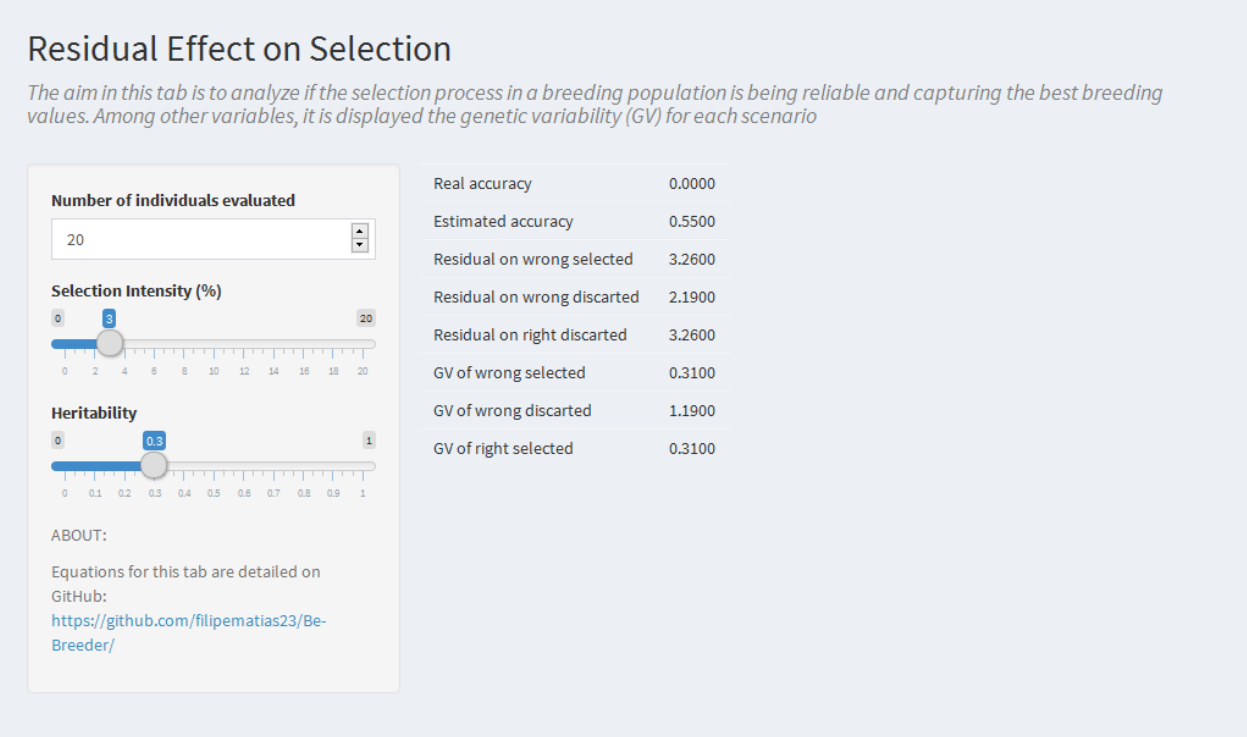


Figure 16. Residual Effect on Selection tab from Be-Breeder 2.0

In order to analyze if the selection is being reliable and capturing the best breeding values, Be-Breeder 2.0 calculates the residual effect on selection. The parameters needed are the number of individuals to be evaluated, selection intensity (%), and heritability (Figure 16). Therefore, the following outputs are calculated: real accuracy, estimated accuracy, residual on wrong selected, residual on the wrong discarded, residual on the right discarded, breeding value of the wrong selected, breeding value of wrong discarded, and the breeding value of right selected individuals.

1.15 Replicates vs Population size

This topic allows the user to play with two central concerns in plant breeding: the number of replicates and the population size. Usually, the capacity of genotypes evaluation is limited in breeding programs, being important to evaluate compensatory aspects regarding cost and field plot utilization. Based on that, the user here may infer what the impact of an increase in the number of replicates on the population size is, and especially, what is the increment of that on the response to selection (RS), simulating different scenarios in function of the number

of replicates (r), number of plots available (p), number of genotypes selected (s), and trait heritability (h^2), according the equation $RS = \left(\frac{e^{-0,5\left[\left(\frac{s}{p}\right)^2 - \left(\frac{s \times r}{p}\right)^2\right]}}{r} \times \sqrt{\frac{r}{1 + h^2 * (r-1)}} - 1 \right) 100$ (Figure17).

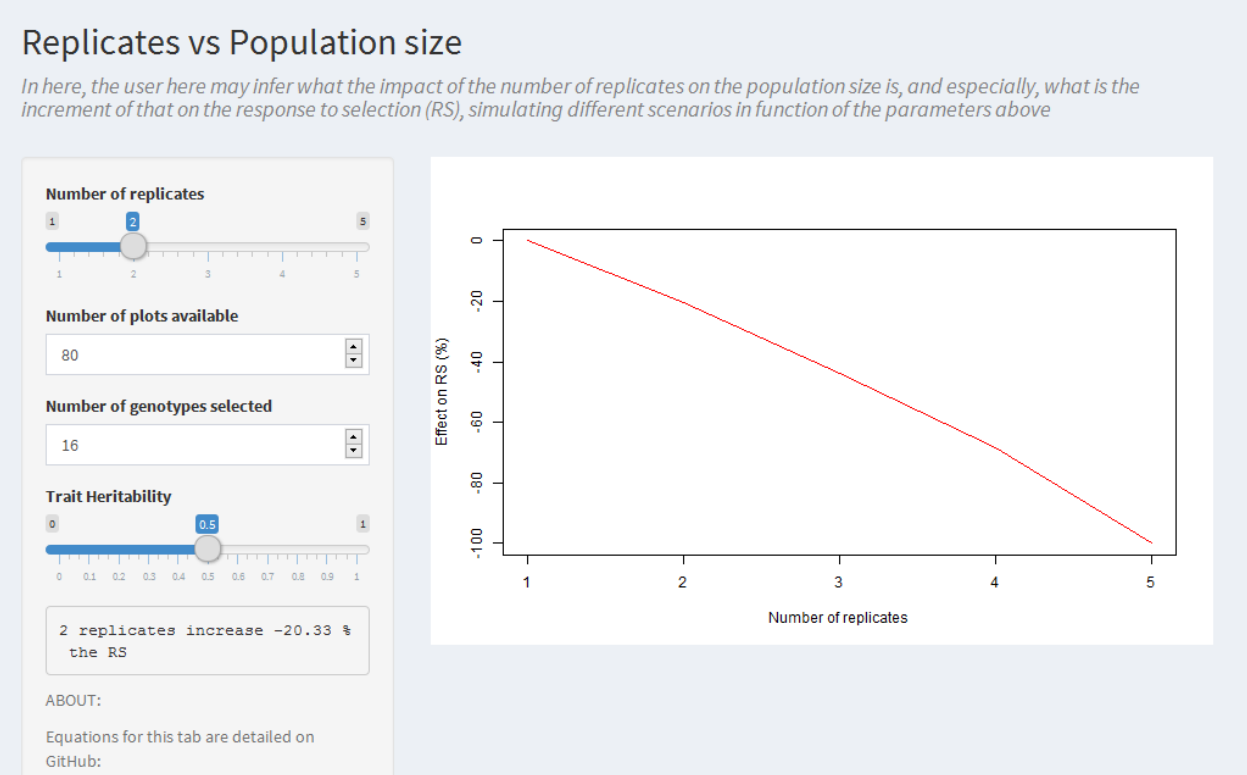


Figure 17. Replicates vs Population size tab from Be-Breeder 2.0

1.16 Economics in Plant Breeding

This analysis provides a broader framework of a breeding process, considering the cost spent per plot, the viability, and the expected genetic gain. Hence, the number of genotypes evaluated and selected, the trait heritability, the number of plot replicates, the time spent per breeding cycle (years), and the cost to evaluate each plot (U$) must be informed (Figure 18). Modifying the value of heritability and the number of plot replicates, for example, the user may observe the influence of the aspects (*i*) nature of the trait and (*ii*) experimental size on the total costs. Thus, the ratio genetic gain by total cost is provided, as well as a graphic illustration.

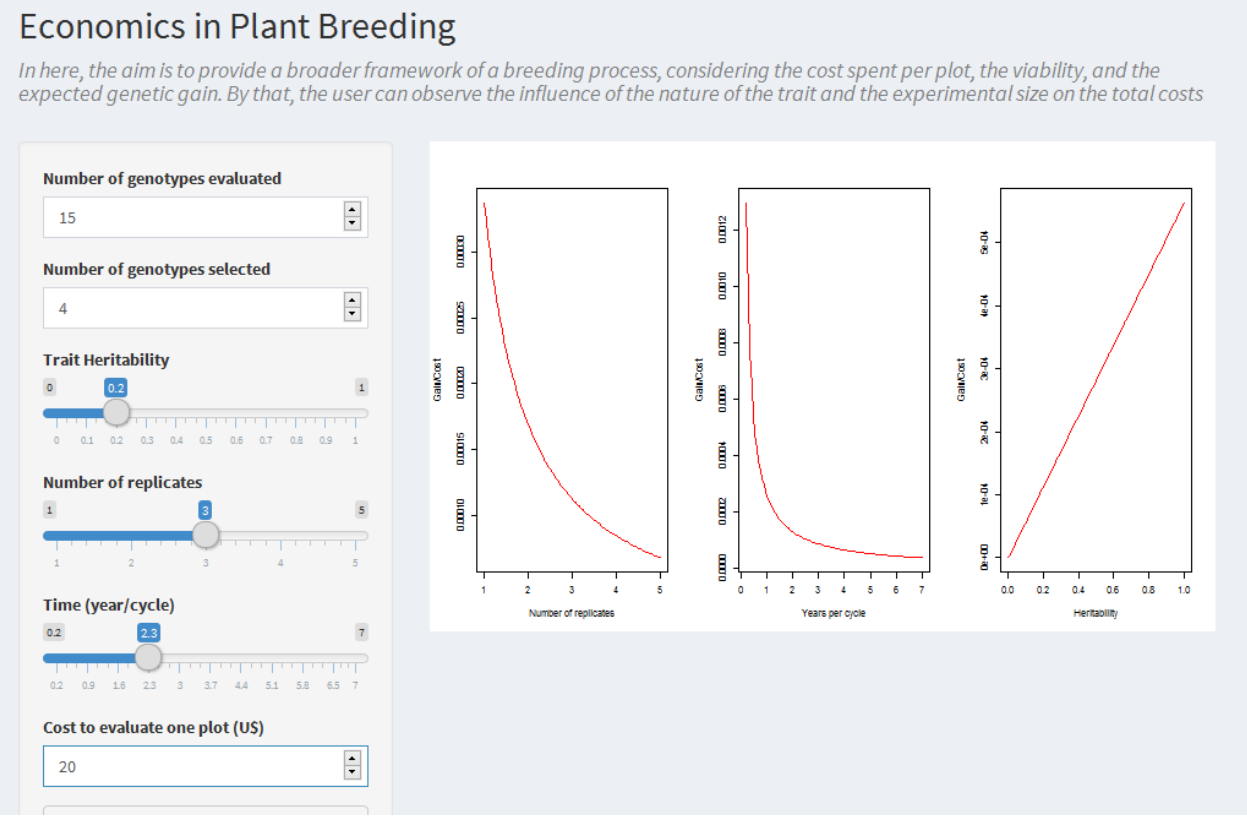


Figure 18. Economics in Plant Breeding tab from Be-Breeder 2.0

1.17 Linkage Disequilibrium (LD)

In order to estimate this phenomenon, we use the *D* coefficient for each genotype, which is a quantitative measure of allelic association (Hartl and Clark, 2007). For genotype *AB*, for example, LD is estimated as follows: $D_{AB} = p_{AB} - (p_A)(p_B)$, where p_{AB} is the observed frequency of genotype *AB*, p_A is the frequency of allele *A*, and p_B is the frequency of allele *B*. Frequencies of alleles *a* and *b* are estimated as $p_a = (1 - p_A)$ and $p_b = (1 - p_B)$, respectively. Loci can be linked either in coupling (positive *D*) or in repulsion phase (negative *D*), also called *cis* and *trans* configurations. Based on these concepts, the user may simulate different conditions informing the frequencies of the haplotypes *AB*, *Ab*, *aB*, and *ab*, and the rates of the alleles *A*, *a*, *B*, and *b* (Figure 19). The application will then estimate whether the loci are independent, using the *D* value and the linkage phase, and how strong the LD is through another metric commonly used, the square allele frequency correlation coefficient (r^2), which corresponds to $r^2 = \frac{D^2}{p_A \times p_B \times p_a \times p_b}$ (Hill and Robertson, 1968).

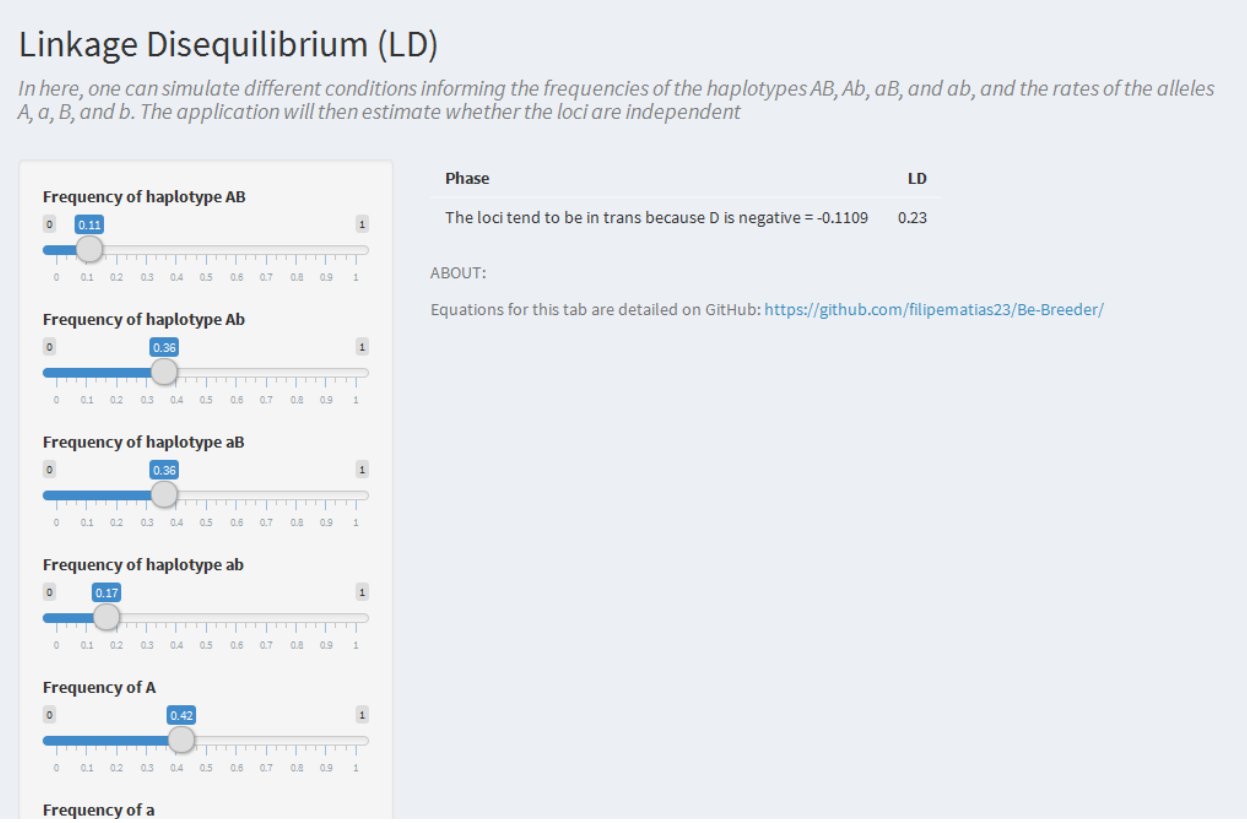


Figure 19. Linkage Disequilibrium (LD) tab from Be-Breeder 2.0

1.18 Cycles to Reduce LD

Combining the number of generations to the recombination frequency in a population, it is possible to compute how many cycles are needed to dissipate the LD using $p_{AB} = p_A p_B + D_{AB} = p_A p_B + (1-c)D_{AB}$, where c refers to the rate of recombination and the other parameters are the same as described previously (Hartl and Clark, 2007). When no recombination occurs ($c = 0$), the frequency p_{AB} from one generation to the next is maintained, but when $c > 0$, LD decay is given exponentially according to the number of generations (t), following the equations:

- Generation 0: $D_{AB}^0 = p_{AB} - p_A p_B$
- Generation 1: $D_{AB}^1 = p_{AB}^1 - p_A^1 p_B^1 = (1-c)^1 D_{AB}^0$
- Generation t: $D_{AB}^t = (1-c)^t D_{AB}^0$

Thus, to observe these events, the user must inform the loss of LD (%) aimed and the recombination ratio between loci (Figure 20). In addition to the numeric information, Be-Breeder 2.0 displays graphic responses.

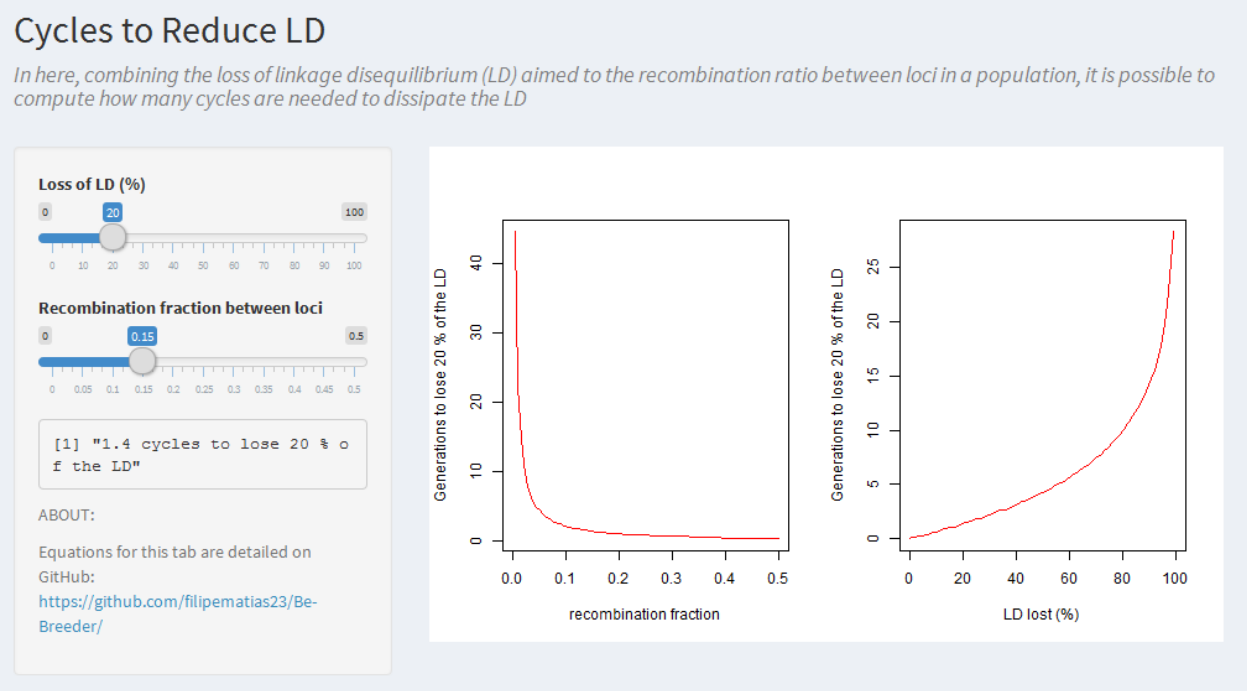


Figure 20. Cycles to reduce LD tab from Be-Breeder 2.0

2 Phenotypic Breeding

In this module and the next, Molecular Breeding, the online version of the application provides a "HELP" checkbox with a short tutorial about data format for each topic. The aim is that the user loads their own dataset in each of them. Nevertheless, it is possible to run the analyses using the data provided by the application by marking the Example checkbox. In this case, the user must check "Example" for all subtopics connected to that topic.

2.1 Experimental Analysis

In this topic the dataset is analyzed using mixed models by the *lme4* package in R (Bates et al., 2014). This package requires at least one random effect parameter to estimate fixed effects and predict the random effects. For illustration, a mixed linear model can be written as $y = \boldsymbol{X}\beta + \boldsymbol{Z}u + \epsilon$, where y is a vector of phenotypic values, β is the fixed effects vector, u is the random effects vector, and ε is the experimental error. $\boldsymbol{X}$ and $\boldsymbol{Z}$ are the incidence matrices for the vectors β and u. Analyses are performed using mixed model equations of Restricted Maximum Likelihood / Best Linear Unbiased Predictor (REML/BLUP) method, in which the variance-covariance matrix of genotype effect is an identity matrix since they are considered as non-related (Resende 2002).

2.1.1 Dataset

The user can upload his own dataset or use the example data to proceed with the analysis (Figure 21). More details are present in the Help option.

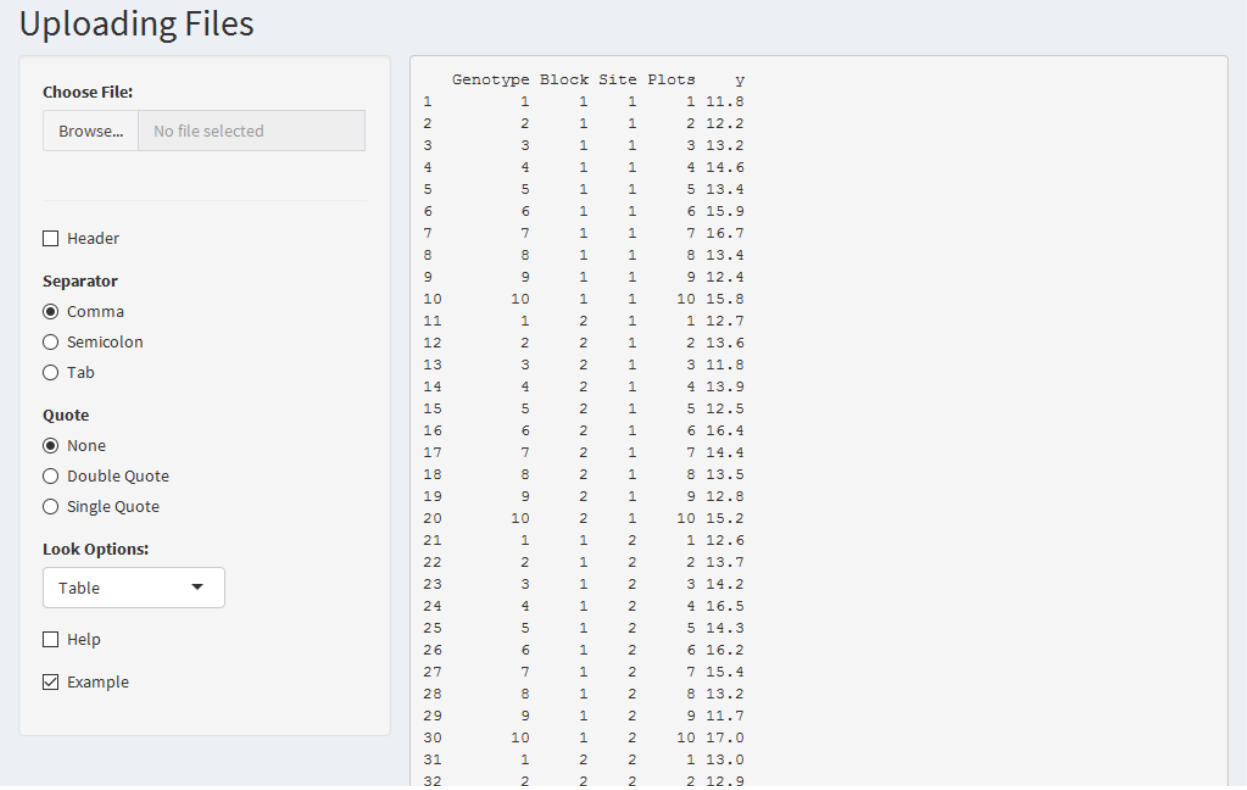


Figure 21. Dataset - Experimental Analysis tab from Be-Breeder 2.0

2.1.2 Statistical Model

The user must define the effect of each factor in the model as fixed or random. Fixed effects should be declared using the name of their respective columns in the dataset. Random effects must be declared between parentheses, with "1" followed by the name of the effect, separated by a vertical bar. For example, considering an experiment with two factors, in which factor A is declared as fixed and factor *B* as random, the model should be typed as $y \sim A + (1|B)$. Interaction between factors are declared with their names separated by a colon, i.e. $(1|A:B)$, and the nature of its effect should be typed as described for individual factors.

In this section, it is possible to obtain outputs such as the overall phenotypic mean, estimation of fixed effects, prediction of random effects, test of statistical significance for both of them, and predicted means for each genotype and trait. These latter are considering the Selection Intensity (scrollbar) set by the user (Figure 22).

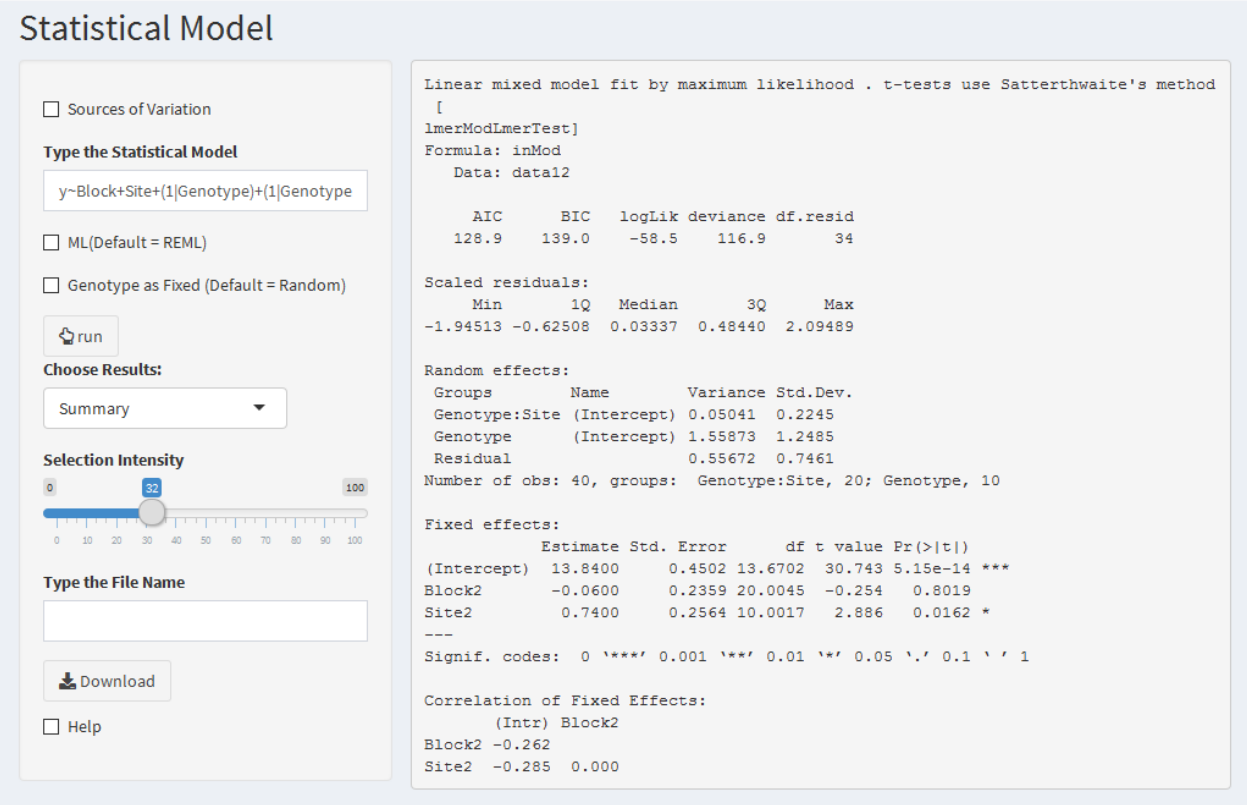


Figure 22. Statistical Model – Experimental Analysis tab from Be-Breeder 2.0

A distinctive feature of this section is the flexibility to define a statistical model by the experimental conditions from users. For more detailed descriptions and options, we recommend checking the Help option.

2.2 Diallel Analysis

The central methods for diallel procedures are considered in this topic: Griffing (1956), Gardner and Eberhart (1966), both for balanced diallels, and North Caroline II design, for unbalanced diallels (Comstock et al., 1949).

Several important parameters are released for: variance for random effects; BLUP (Best Linear Unbiased Predictor) values, which are predictors for random effects in linear mixed models; BLUE (Best Linear Unbiased Estimator) values (concerning fixed effects); *p*-values for fixed effects ; chi-square testing whether the fixed effect coefficients are equal to zero; variance of BLUPs and fixed effect coefficients; the prediction error variance estimates for the BLUPs; and the loglikelihood for the model. In this topic, we employed two packages, *EMMREML* (Akdemir et al., 2015) and *breedR* (Muñoz and Sanchez, 2017).

To run the analysis, the data is set in Dataset tab, according to the file format. Also, it is possible to run the analysis with the dataset provided in the Example button (Figure 23). Help icon displays more information about the analysis.

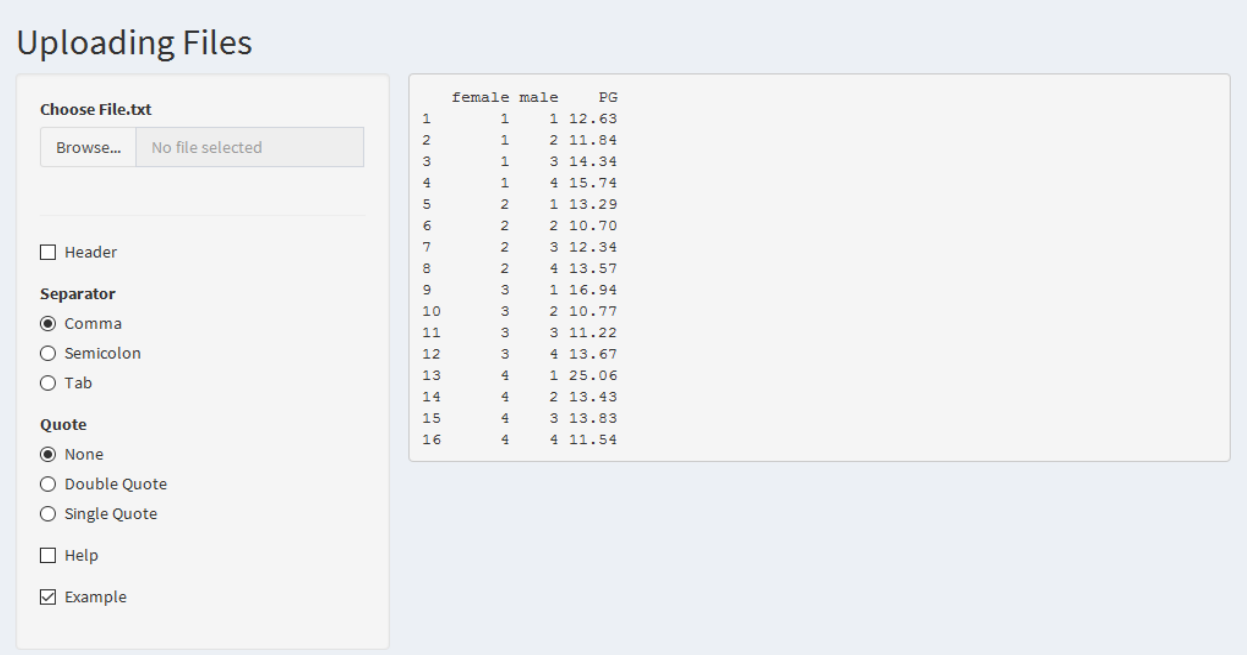


Figure 23. Dataset – Griffin design - Diallel Analysis tab from Be-Breeder 2.0

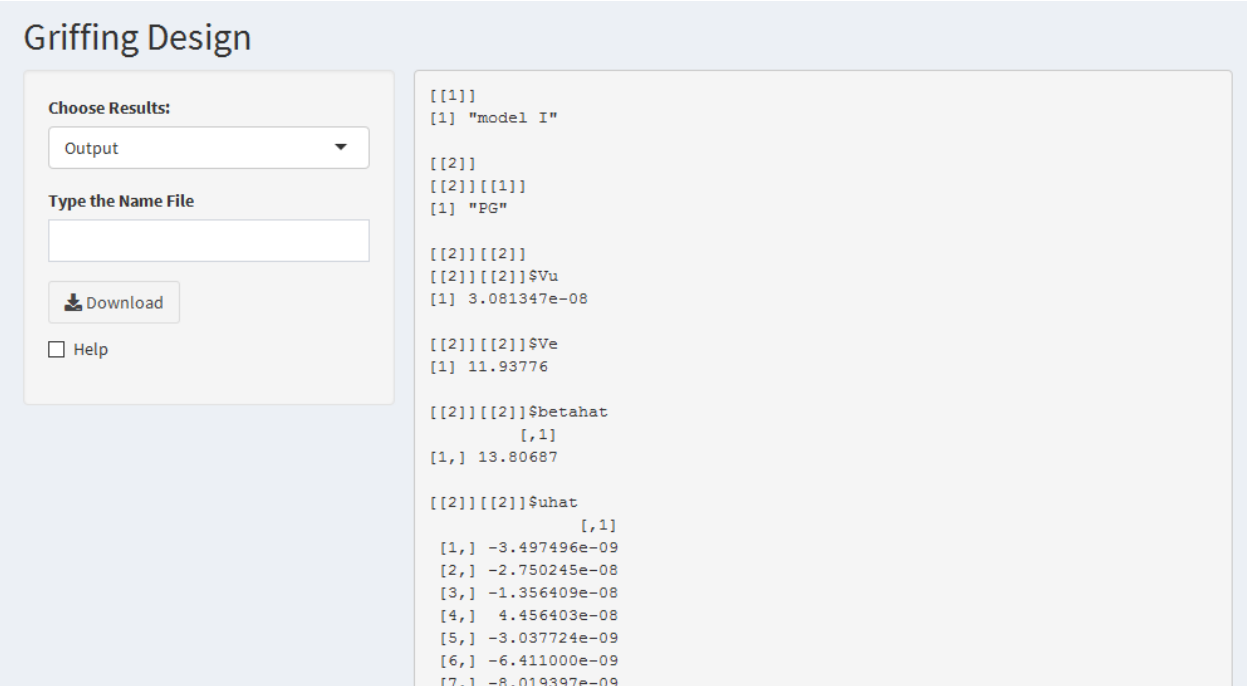


Figure 24. Analysis – Griffin design - Diallel Analysis tab from Be-Breeder 2.0

2.2.1 Griffing Design

When Griffing's methodology is chosen, Be-Breeder 2.0 intuitively points out which one of the four possible methods the dataset refers to. They are 1) Method 1 (full diallel): parents, F1, and reciprocals, 2) Method 2 (half diallel): parents and F1's, 3) Method 3: F1's and reciprocals, 4) Method 4: F1's. Then, the analysis is performed according to it.

2.2.2 Gardner and Eberhart Design

Similar to Griffing's methodology presented before, it is possible to set one of four possible methods in Gardner and Eberhart's method: (1) full diallel; (2) half diallel; (3) F1's and reciprocals, 4) only F1's. Consequently, the analysis is performed according to the method chosen.

2.2.3 Factorial Design

Several important parameters are released: variance for random effects; BLUP (Best Linear Unbiased Predictor) values, which are predictors for random effects in linear mixed models; BLUE (Best Linear Unbiased Estimator) values (concerning fixed effects); p-values for fixed effects ; chi-square testing whether the fixed effect coefficients are equal to zero; variance of BLUPs and fixed effect coefficients; the prediction error variance estimates for the BLUPs; and the loglikelihood for the model.

2.3 Index Selection

In this tab from the Phenotypic Breeding section, it is possible to perform the selection of genotypes using additive selection index (SI), which combine several quantitative traits simultaneously. The general model for SI considering *n* traits is $\mathrm{SI} = \mathrm{X}_1\,\mathrm{b}_1 + \mathrm{X}_2\mathrm{b}_2 + \ldots + \mathrm{X}_\mathrm{n}\mathrm{b}_\mathrm{n}$, where ***SI*** is the value of the index, X_i is the mean of trait *i*, and b_i is the weight attributed to each trait *i*. There is no restriction regarding the number of traits considered.

2.3.1 Index File

User's dataset must contain the genotype identification in the first column, named Genotype, followed by the information of traits, each one in a different column (Figure 25). Here, we recommend the use of predicted means or BLUP values, which can be obtained from the topic Statistical Model from Experimental Design Analysis tab, as previously described.

2.3.2 Index Analysis

The weights for each trait and their respective definition of favorable (+) or unfavorable (-) effect must be written appropriately in the Index Analysis window, according to the order the traits are mentioned in the data file. The value zero should be attributed as the weight to a trait for removing its effect from the index.As output, genotypes are sorted by the index value, from highest to lowest (Figure 26). It is also possible to simulate different levels of selection intensity and verify their effect on the outcome.

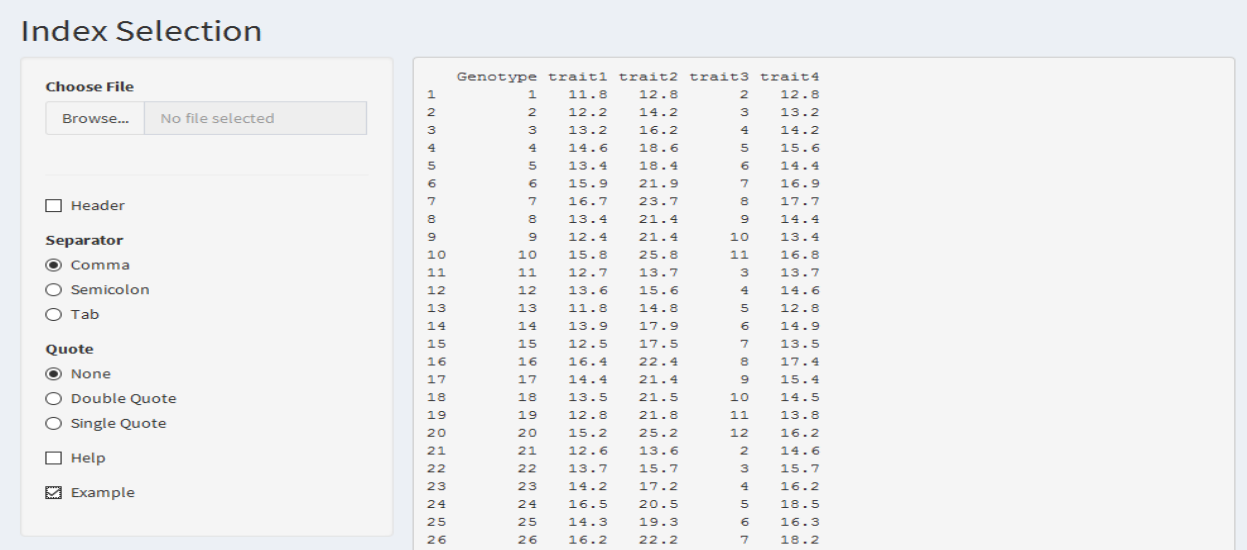


Figure 25. Index file – Index Selection tab from Be-Breeder 2.0

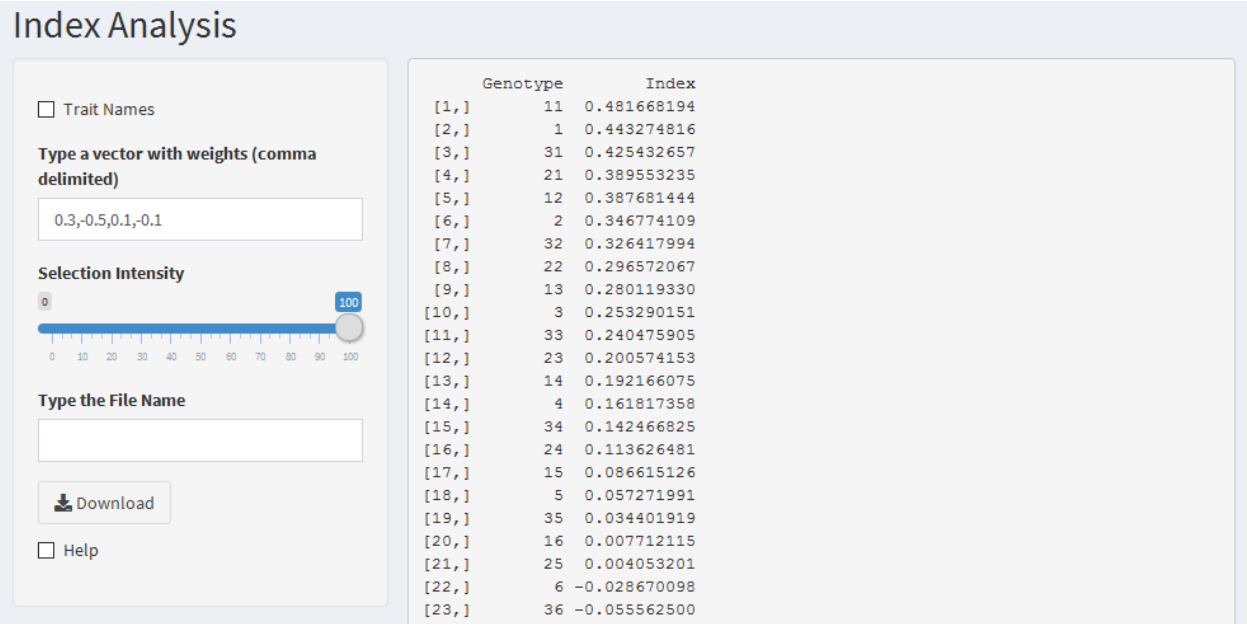


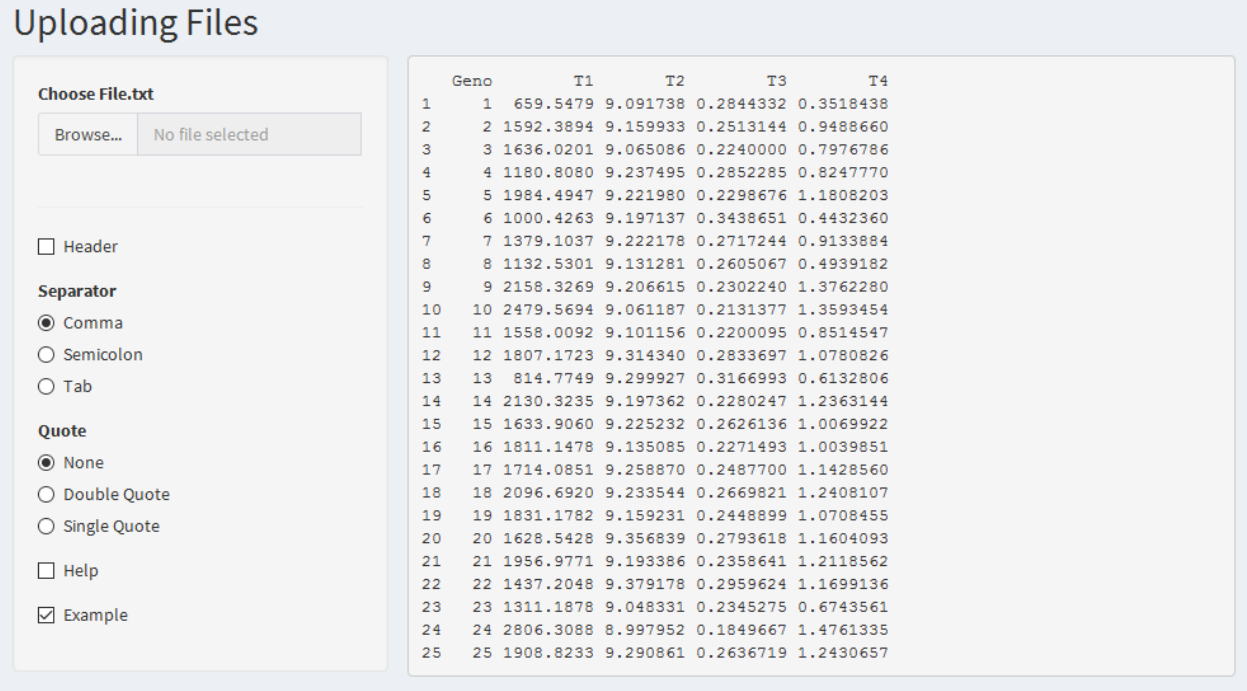
Figure 26. Index analysis – Index Selection tab from Be-Breeder 2.0

2.4 Correlation Analysis

This analysis is carried out to obtain the Pearson's method coefficients of correlation and the p-value between all the variables of a dataset (Figure 27), determining the strength of pairwise statistical association between them. Two types of graphs are also displayed to provide a visual effect (Figure 28).

2.4.1 File input

All genotypes included in the data file will be analyzed.



	Geno	T1	T2	T3	T4
1	1	659.5479	9.091738	0.2844332	0.3518438
2	2	1592.3894	9.159933	0.2513144	0.9488660
3	3	1636.0201	9.065086	0.2240000	0.7976786
4	4	1180.8080	9.237495	0.2852285	0.8247770
5	5	1984.4947	9.221980	0.2298676	1.1808203
6	6	1000.4263	9.197137	0.3438651	0.4432360
7	7	1379.1037	9.222178	0.2717244	0.9133884
8	8	1132.5301	9.131281	0.2605067	0.4939182
9	9	2158.3269	9.206615	0.2302240	1.3762280
10	10	2479.5694	9.061187	0.2131377	1.3593454
11	11	1558.0092	9.101156	0.2200095	0.8514547
12	12	1807.1723	9.314340	0.2833697	1.0780826
13	13	814.7749	9.299927	0.3166993	0.6132806
14	14	2130.3235	9.197362	0.2280247	1.2363144
15	15	1633.9060	9.225232	0.2626136	1.0069922
16	16	1811.1478	9.135085	0.2271493	1.0039851
17	17	1714.0851	9.258870	0.2487700	1.1428560
18	18	2096.6920	9.233544	0.2669821	1.2408107
19	19	1831.1782	9.159231	0.2448899	1.0708455
20	20	1628.5428	9.356839	0.2793618	1.1604093
21	21	1956.9771	9.193386	0.2358641	1.2118562
22	22	1437.2048	9.379178	0.2959624	1.1699136
23	23	1311.1878	9.048331	0.2345275	0.6743561
24	24	2806.3088	8.997952	0.1849667	1.4761335
25	25	1908.8233	9.290861	0.2636719	1.2430657

Figure 27. File input – Correlation Analysis tab from Be-Breeder 2.0

2.4.2 Coefficients and graphs

The packages *agricolae* (Mendiburu, 2017) and *corrgram* (Wright, 2017) are used to construct the graphics displayed (Figure 28).

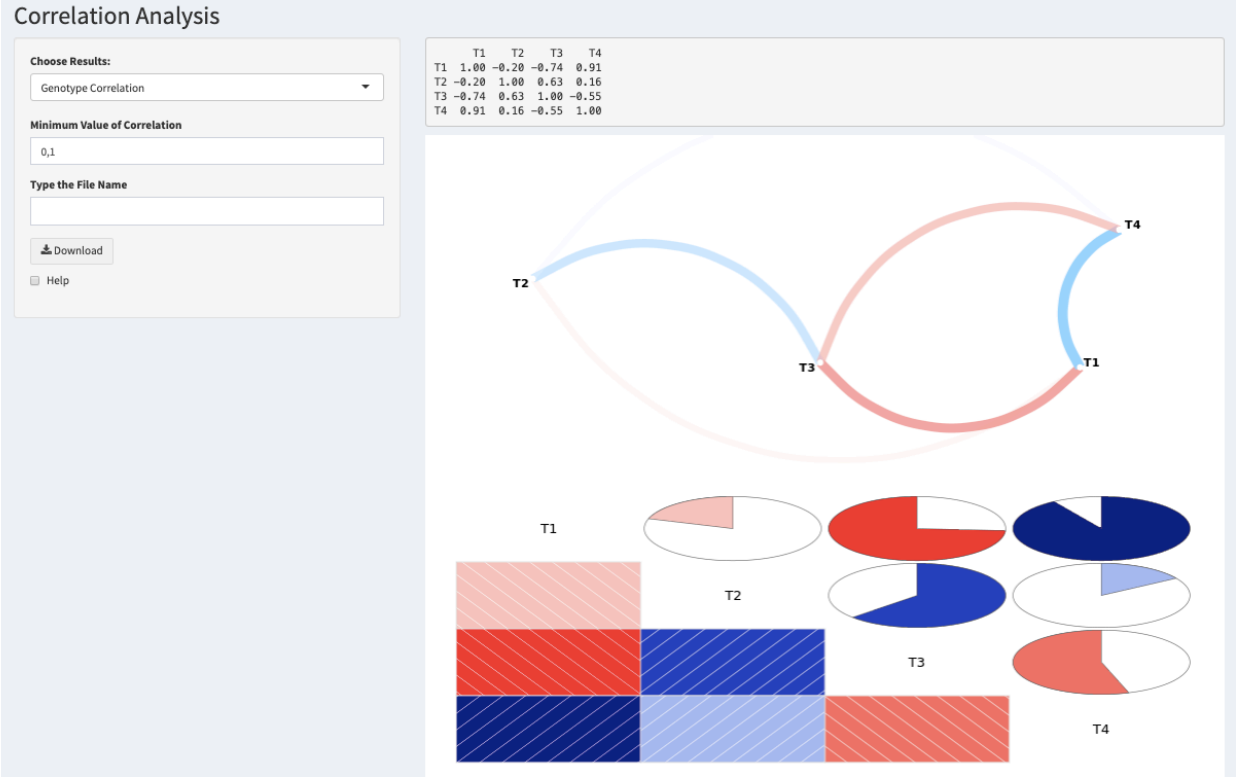


Figure 28. Coefficients and graphs – Correlation Analysis tab from Be-Breeder 2.0

2.5 Path Analysis

Path Analysis tab carry out path analysis among traits (Wright, 1923). This analysis is performed using the *agricolae* R package (Mendiburu, 2016). Just as it occurs for Selection Index, input dataset must contain the genotype identification in the first column, named Genotype, followed by information of as many traits as desired, each one in a different column.

2.5.1 Trait File

Data with phenotypic mean must be considered as trait information (Figure 29). In addition, the user must indicate the main trait, so the estimates of direct and indirect effects between that and the other ones are supplied.

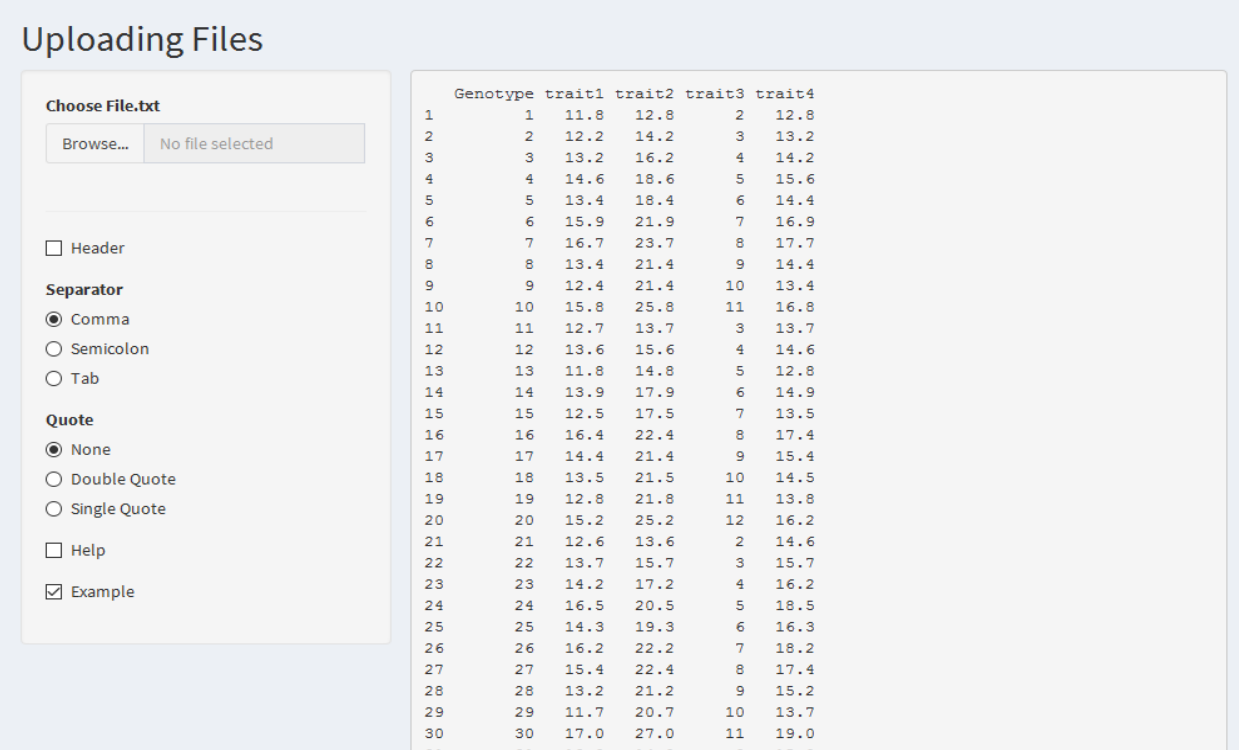


Figure 29. Trait file – Path Analysis tab from Be-Breeder 2.0

2.5.1 Path Analysis

Finally, the user can choose whether perform path analysis or traits correlation (Pearson's) in the option Choose Results.

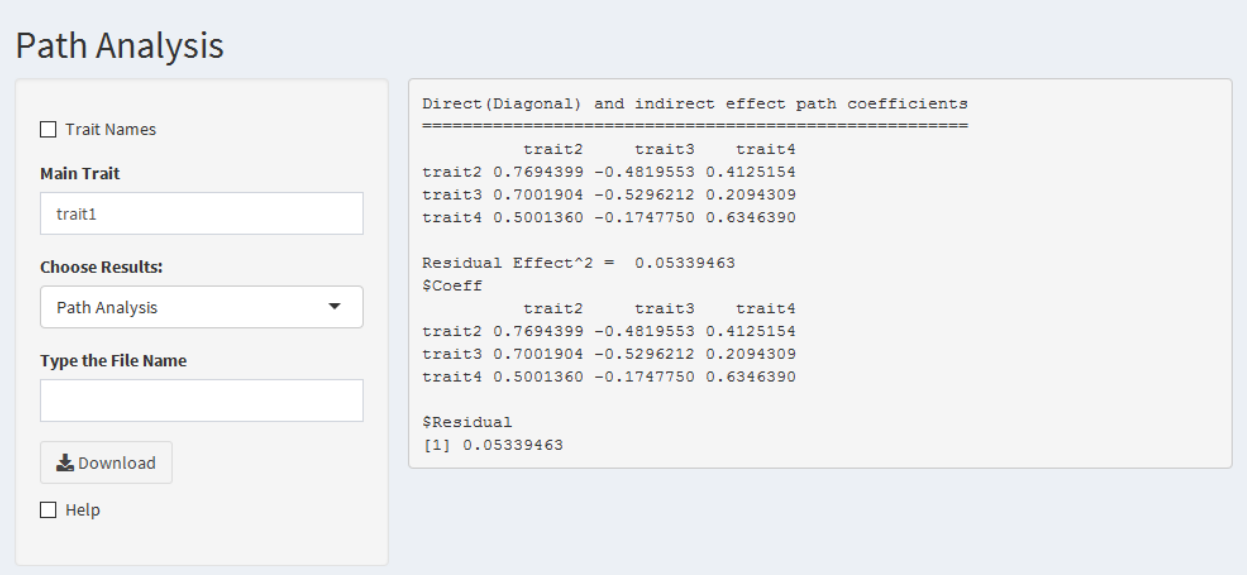


Figure 30. Path Analysis tab from Be-Breeder 2.0

2.6 Biplot Analysis

Evaluation of the genotype × environment interaction is contemplated in the tab Biplot Analysis, carried out using the built-in R functions *princomp* and *biplot*.

2.6.1 Dataset

In the entry dataset, genotypes should be arranged in rows and environments in columns (Figure 31). Two types of analysis are available, as described in the two following topics.

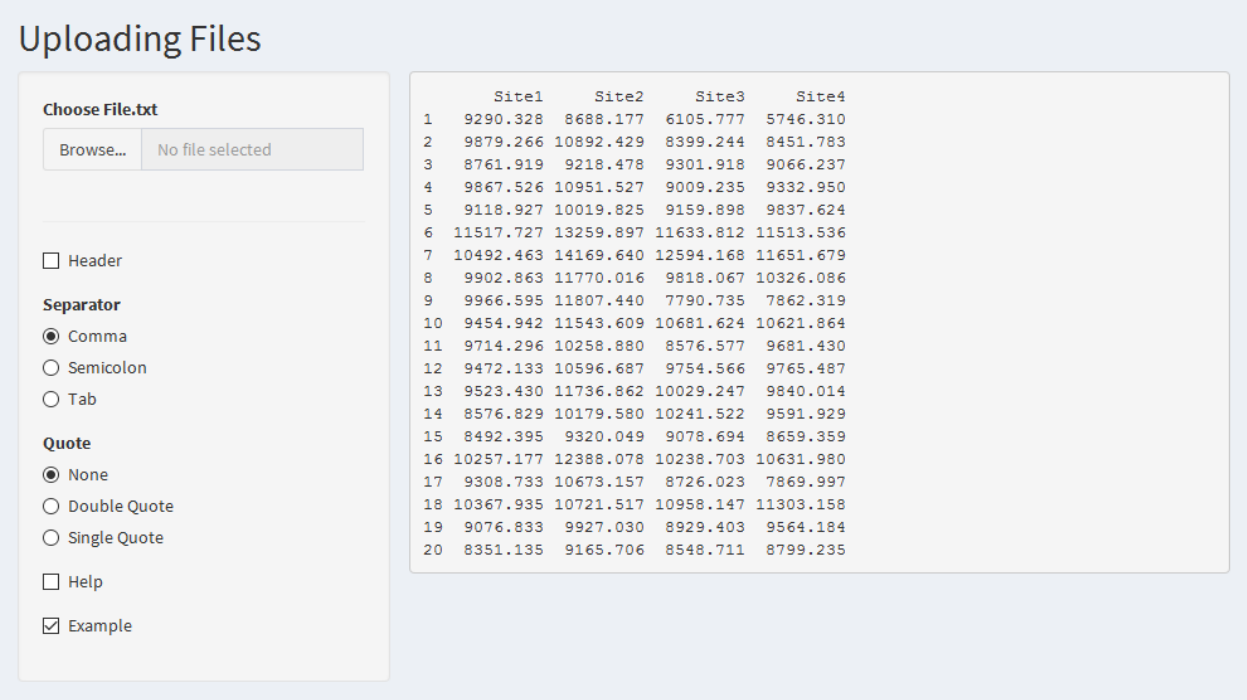


Figure 31. Dataset – Biplot Analysis tab from Be-Breeder 2.0

2.6.1 GE Biplot

The former is GE Biplot, which performs the principal components analysis, releasing the summary, the scores for each component calculated either for genotypes (G) or for environments (E), and the biplot graph, where the first two principal components are considered, being the genotypes represented by points and the environments by arrows (Figure 32).

2.6.2 GE Cluster

The latter is GE Cluster, that executes cluster analysis in the same context (in here, the user may inform the desired number of clusters), issuing cluster information, the group genotypes belong to, mean of clusters, and a dendrogram built using the R function *hclust* (Figure 33).

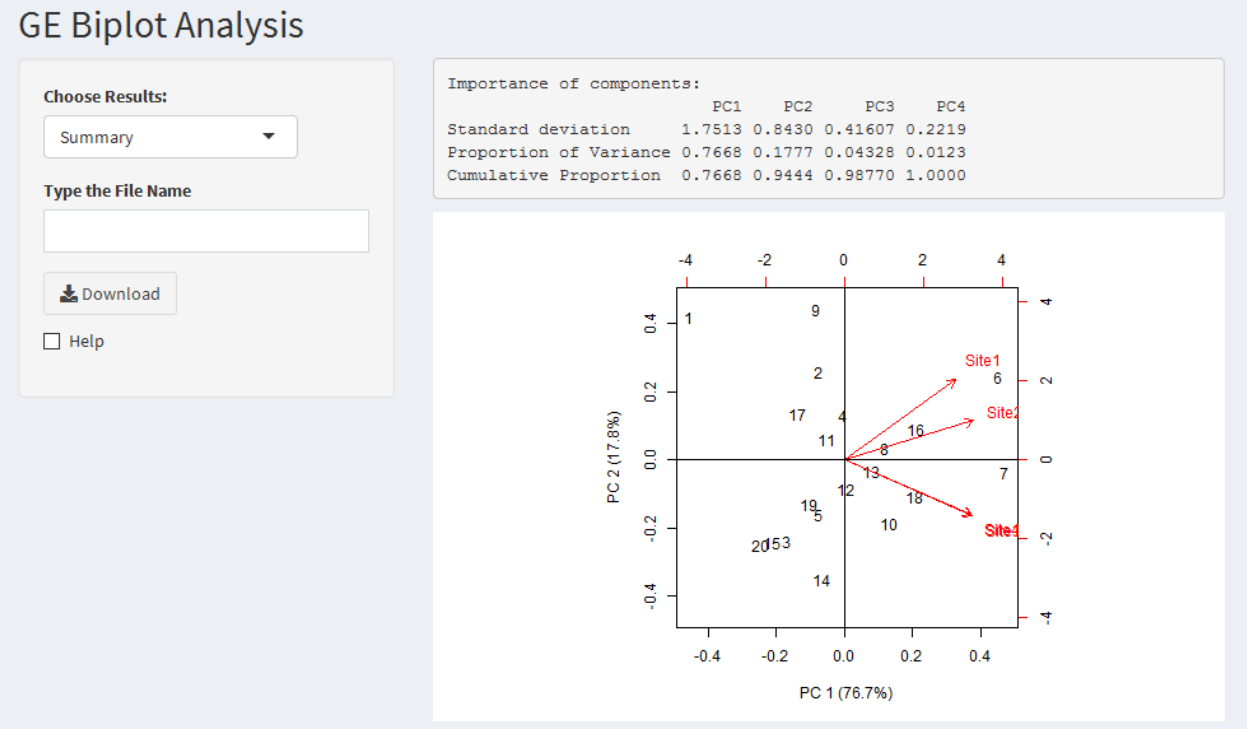


Figure 32. GE Biplot – Biplot Analysis tab from Be-Breeder 2.0

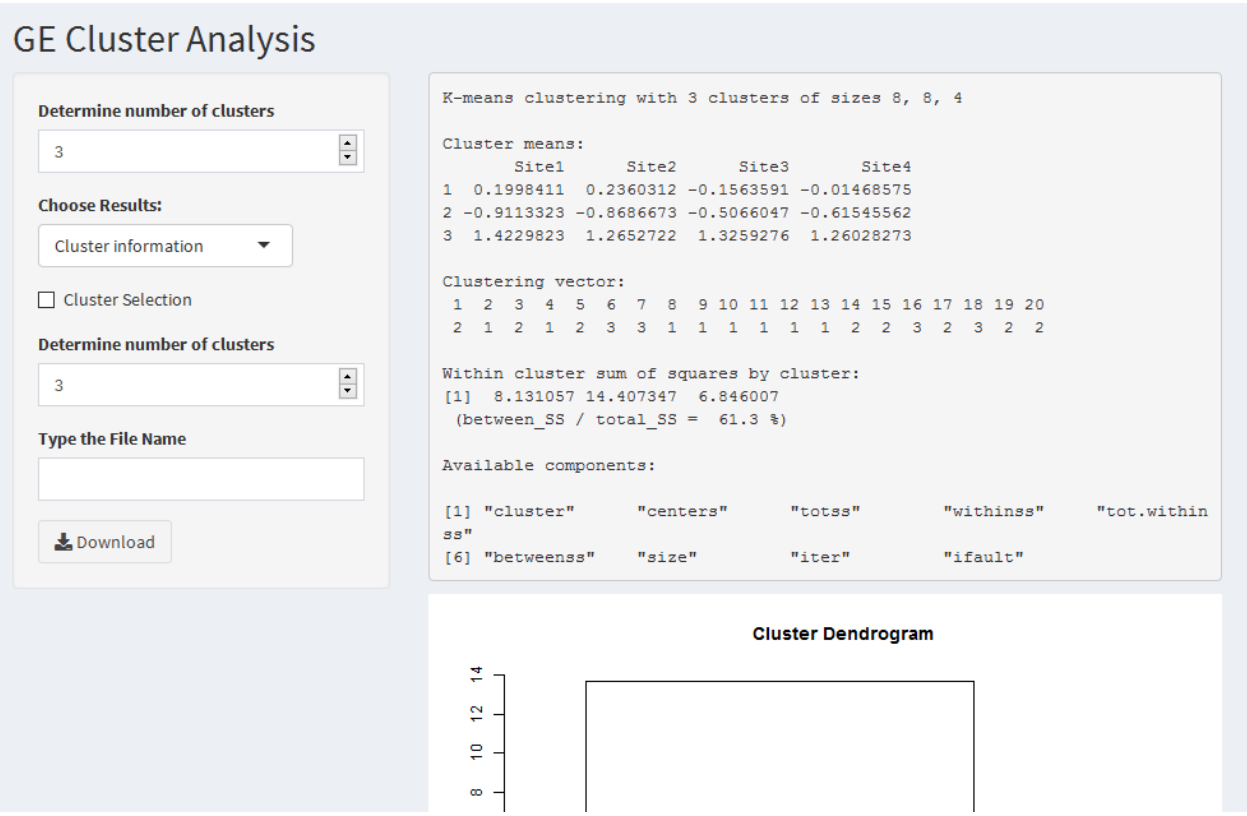


Figure 33. GE Cluster – Biplot Analysis tab from Be-Breeder 2.0

2.7 Experiment Designs

Since many possible designs can be adopted, it is crucial that the organization of the experiment and its installation in the field be in accordance to the one chosen, as well as to the species and the objective of the breeding program. In this topic, it is possible to obtain how the treatments must be installed in the field simulating the following experiment designs: Completely randomized design, Randomized complete block design, Latin square design, Balanced Incomplete Block Designs, Lattice designs, Alpha designs, Augmented block designs, Split-plot designs, and Factorial. The user may set up the number of treatments, number of blocks and replicates, depending on the parameters demanded by design (Figure 34).

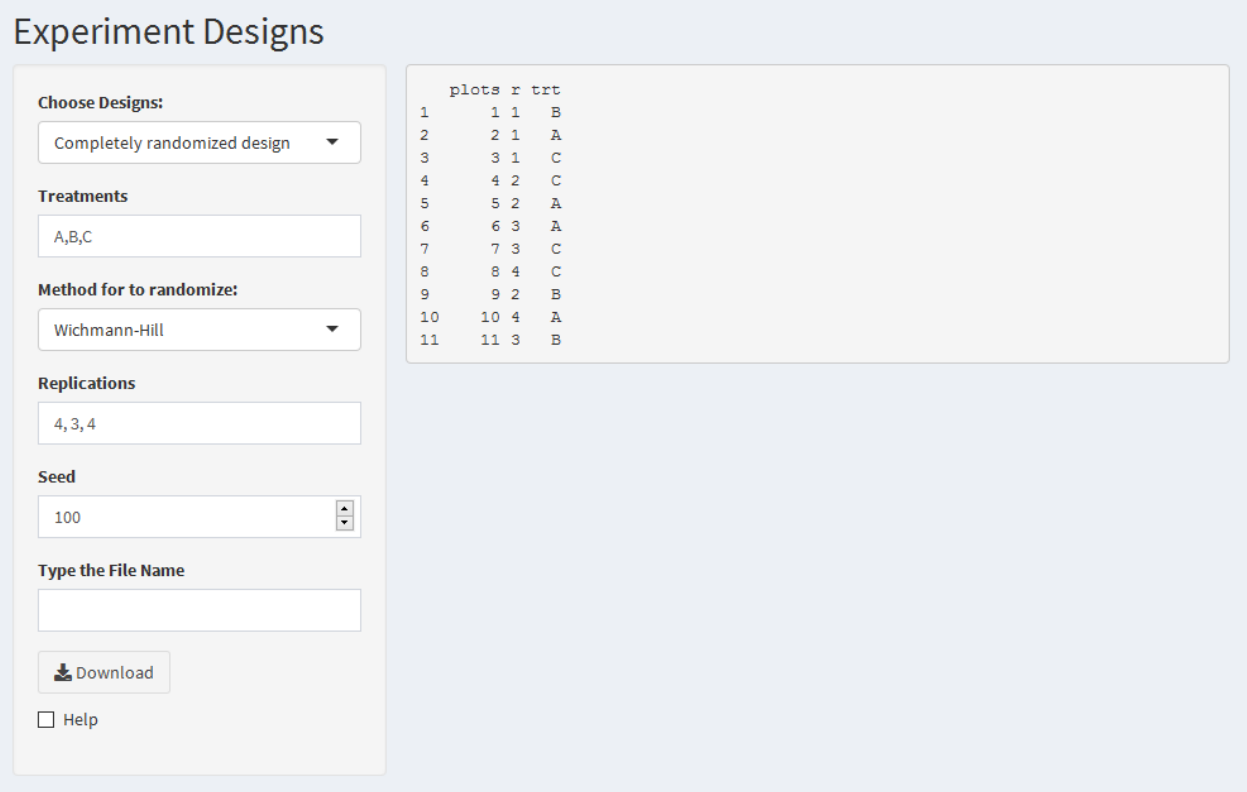


Figure 34. Designs – Experimental Designs tab from Be-Breeder 2.0

3 Molecular Breeding

3.1 Genotyping Data

3.1.1 Quality Control

This tab uses the *raw.data* function of the *snpReady* R package (Granato et al. 2018). In here, the user can control the quality of the genomic data. Data can be uploaded as matrix or data frame filled with nitrogen bases, following the structure of the initial dataset received from the genotyping process (Figure 35).

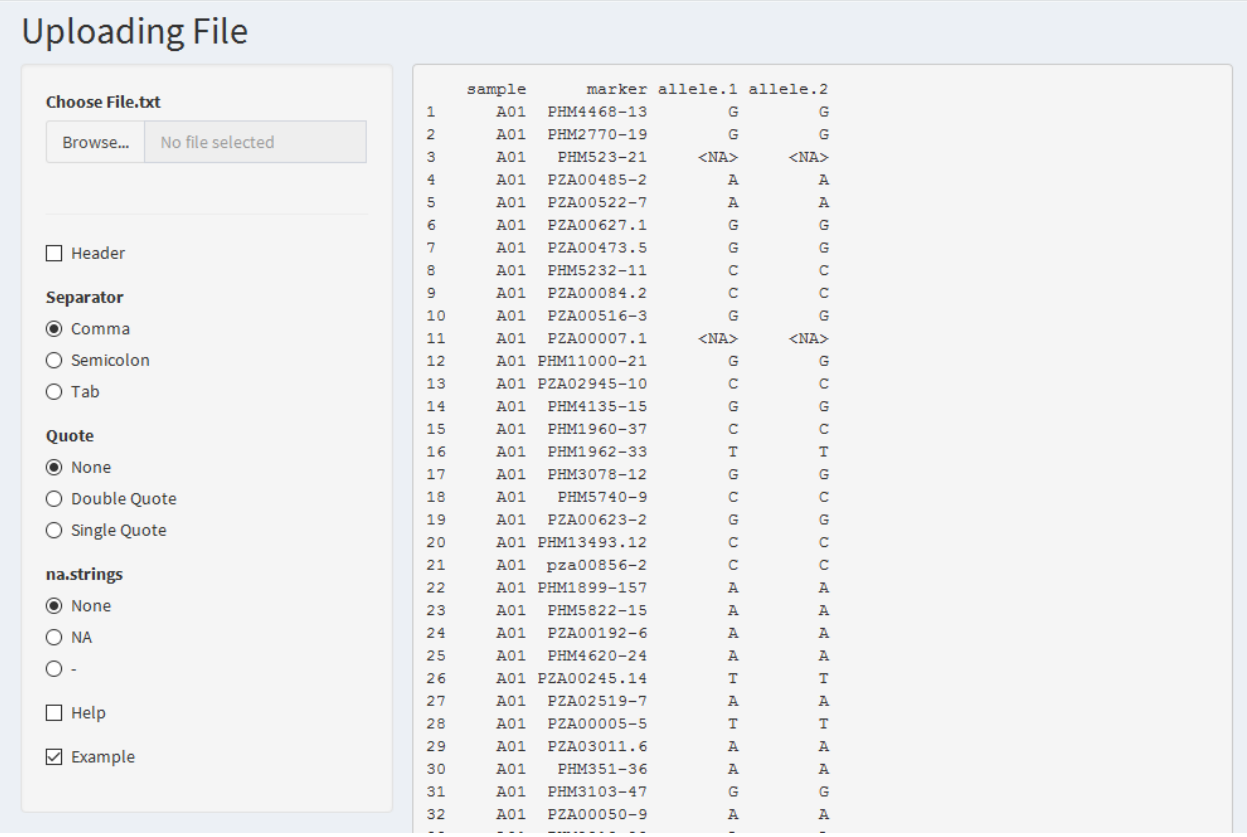


Figure 35. Sample set – Quality control - Genotyping Data tab

In addition, the user must provide the information of marker position, known as "HapMap" matrix. The columns must contain the marker identification, the chromosome where it is, and the position within the chromosome (Figure 36). Finally, to run the analyses, it is required to attribute the desired values for the following parameters (Figure 37):

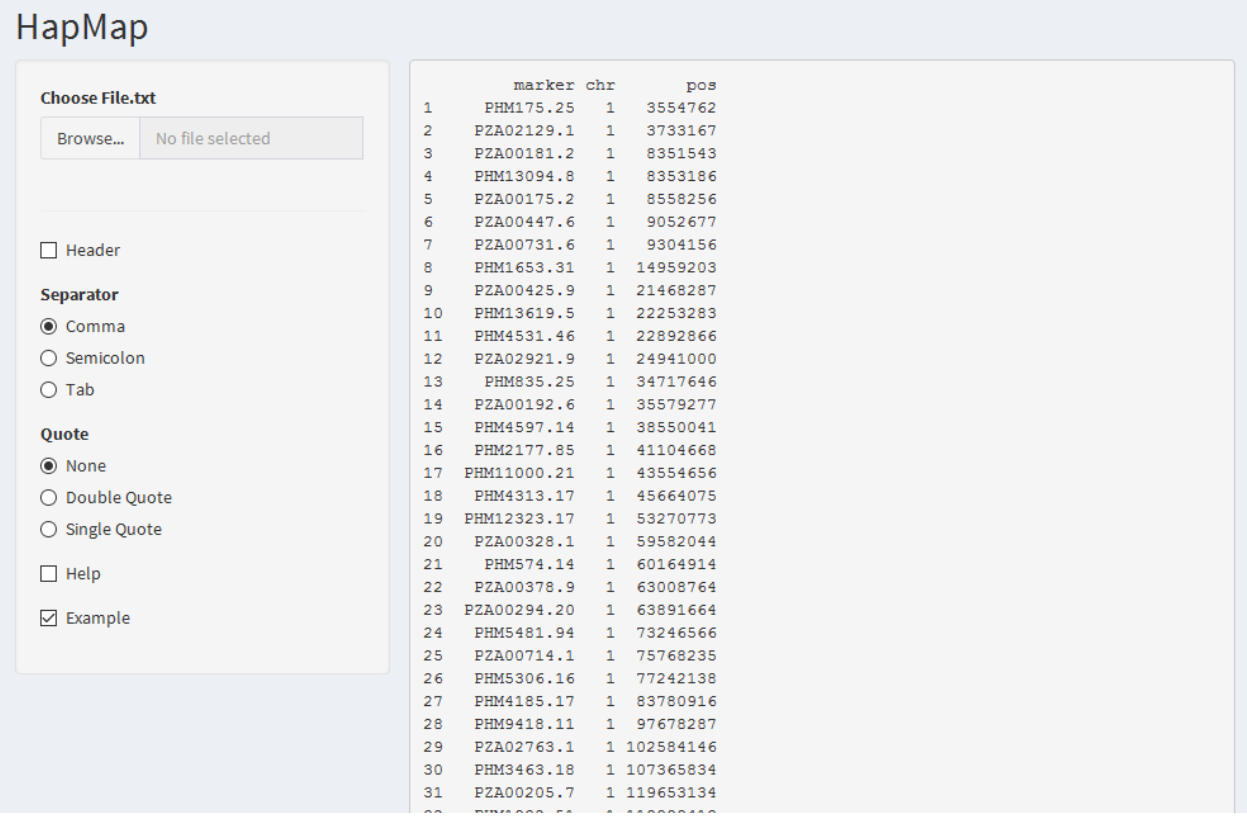


Figure 36. HapMap - Quality control – Genotyping Data tab

- MAF (Minor Allele Frequency): markers with minor allele frequency lower than the threshold defined by the user will be removed;
- Call Rate: markers with a Call Rate lower than the threshold defined by the user will be eliminated;
- Sweep sample: individuals with a Sweep sample lower than the threshold defined by the user will be eliminated.

This tab also allows data imputation, which is carried out from the combined probability of the alleles of a determined $SNP_{(i)}$ (frequencies of p_i and q_i) and the level of homozygosity of the individual that has a marker to be imputed. The output generated by the function is a "clean" matrix, in which the output format can also be defined by the user, with the options of counting a reference allele for each locus (0,1,2), this matrix centered in zero (-1,0,1), or even the matrix in the appropriate format for entry in the Structure software. Thus, the resulting matrix is in the adequate format to be used in other software or packages or, moreover, to proceed with other analysis in the Be-Breeder application itself.

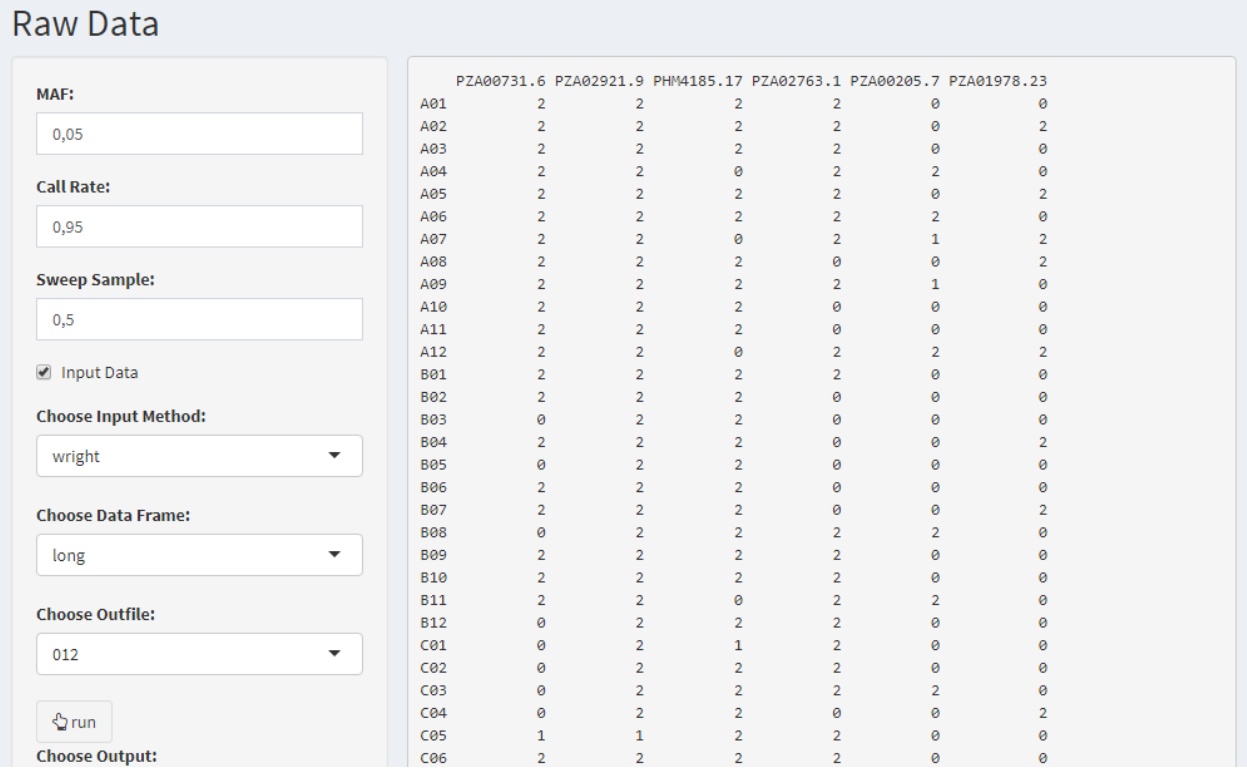


Figure 37. Raw data - Quality control – Genotyping Data tab

3.1.2 Kinship Matrix

This tab uses the *G.matrix* function of the *snpReady* R package (Granato et al. 2018) in order to construct kinship matrices according to four different methods:

- VanRaden (WW): kinship matrix of VanRaden (VanRaden, 2008);
- UAR: unified additive kinship matrix (Yang et al., 2010);
- UARadj: adjusted unified additive kinship matrix (Yang et al., 2010);
- GK: non-linear Gaussian kernel, a reproducing kernel matrix used especially in semi-parametric methods.

In the Z file subtopic, the input file must contain individuals in rows and markers in columns coded as 0-1-2, as showed in the Example option in the tab (Figure 38).

For the outcome, in addition to the possibility of generating the four types of matrices, the user can choose the output format. The ones available are the traditional kinship matrix (wide frame) or the long format (long frame), in which the inverse of the kinship matrix is generated and then organized in columns – it is useful to perform GS analyses in other softwares. Furthermore, the kinship coefficients can be calculated for an additive or a dominant model, except for the GK method (Figure 39). So, the user may observe and download different

number of combinations, according to the options in each parameter (method, frame, and nature of results).

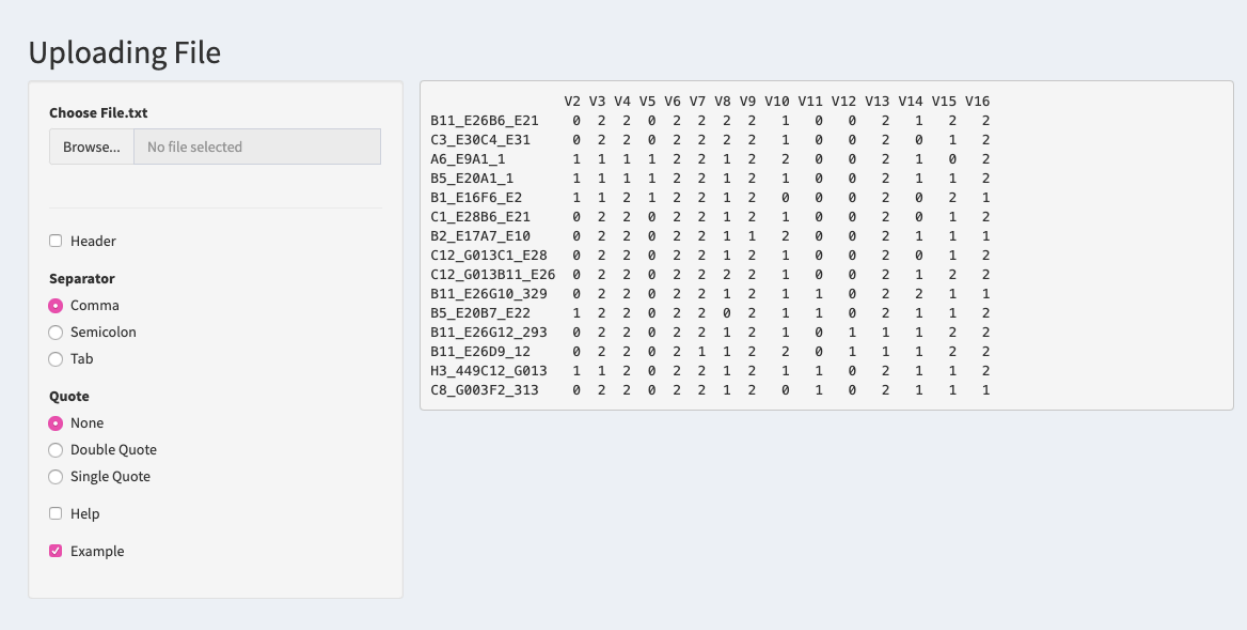


Figure 38. Z matrix – Kinship matrix – Genotyping Data tab from Be-Breeder 2.0

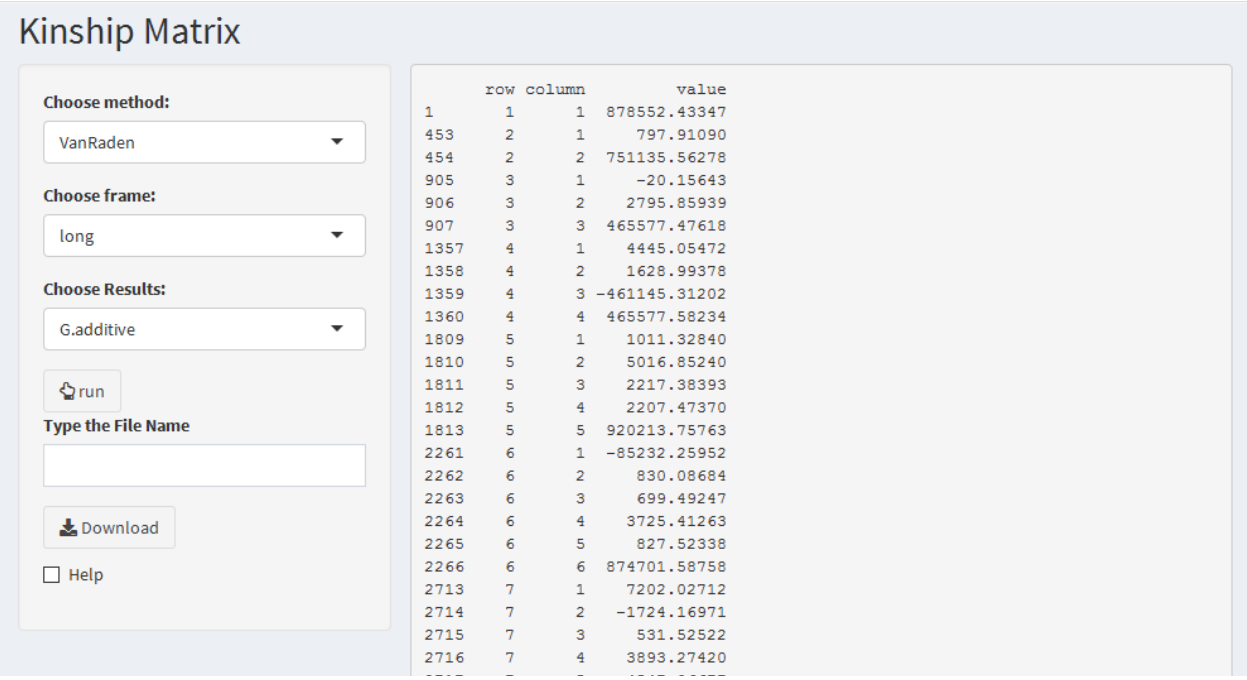


Figure 39. Kinship matrix – Kinship matrix – Genotyping Data tab from Be-Breeder 2.0

3.2 Genomic Selection (GS)

3.2.1 GS analysis

Genome selection (GS) is a powerful tool in plant breeding employed to carry out prediction analyses. The models typically use a training population (TP), which is both genotyped and phenotyped, to calculate genomic estimated breeding values (GEBVs) for individuals in a validation population (VP), which is only genotyped. For that reason, the statistical model must first be trained and validated and for then using the marker effects as a predictive tool. In this tab, the method used is the RR-BLUP, implemented in the rrBLUP package (Endelman, 2011), which assumes the marker effects as random through the Restricted Maximum-Likelihood approach.

To perform GS analysis, the user must load the phenotypic and the Z matrix files in their respective tabs, or set the Example option. The outputs are displayed in the rrBLUP tab into three types of results: marker effects, breeding values, and accuracy (Figure 40). Finally, it is possible to download the outcomes with the desired file name.

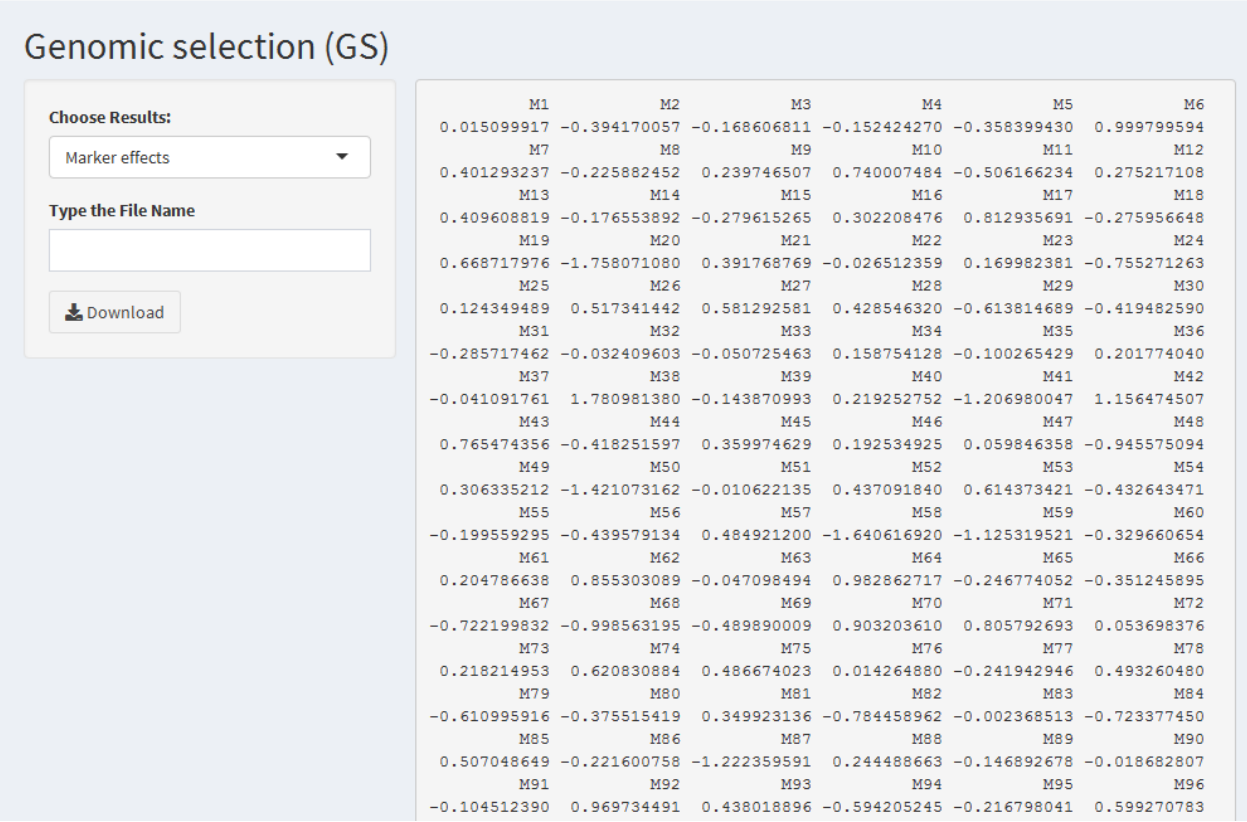


Figure 40. rrBLUP – GS analysis – Genomic Selection tab from Be-Breeder 2.0

3.2.2 Prediction and Selection

In the Prediction and Selection tab, the user must provide the marker effects matrix (Figure 41), the phenotypic dataset (must be entered in the "Phenotypic file" tab in "GS analysis"), and the marker matrix (Z file), with the individuals in the rows and the markers in the columns coded as (-1,0,1) (Figure 42). The former may contain markers of non-phenotyped individuals in order to predict their breeding values using genomic information.

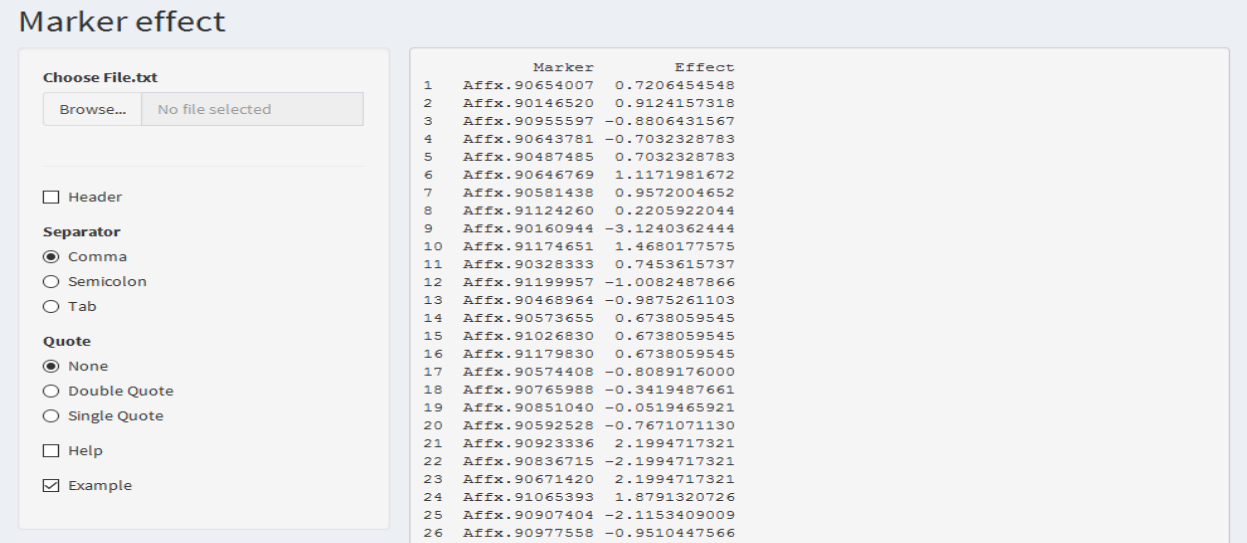


Figure 41. Marker effect - Prediction and Selection - Genomic Selection tab from Be-Breeder 2.0

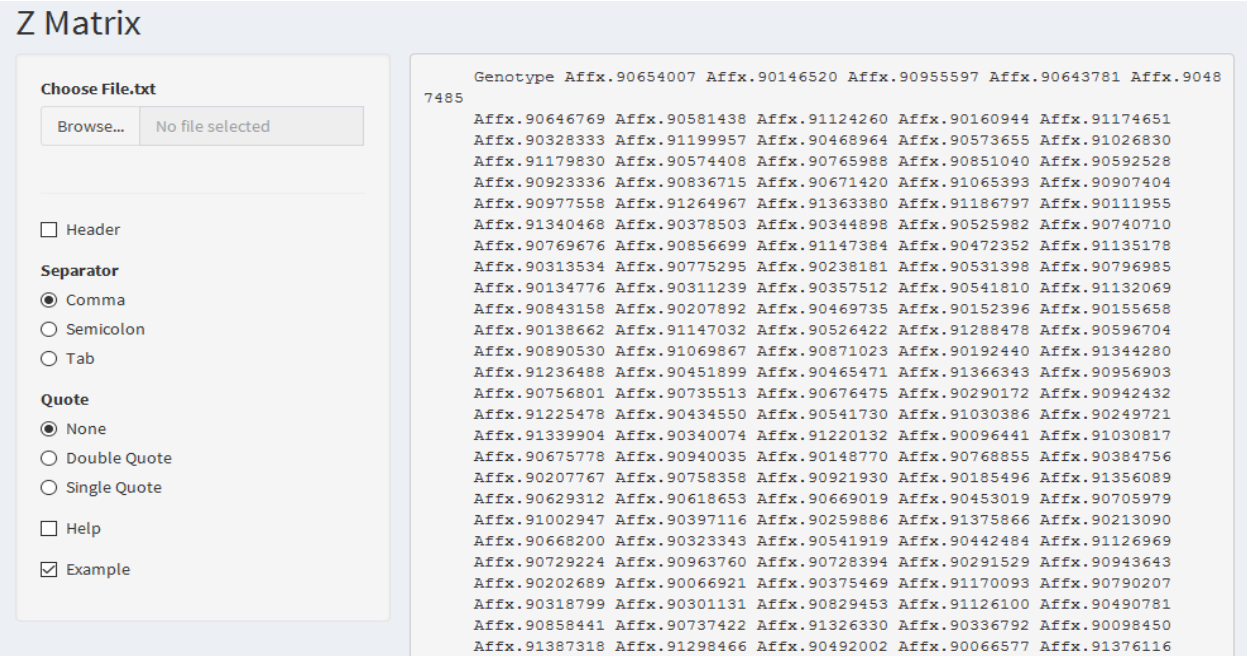


Figure 42. Z matrix – Prediction and Selection – Genomic Selection tab from Be-Breeder 2.0

From these data, the application performs genomic selection according to the selection intensity level provided by the user (Figure 43).

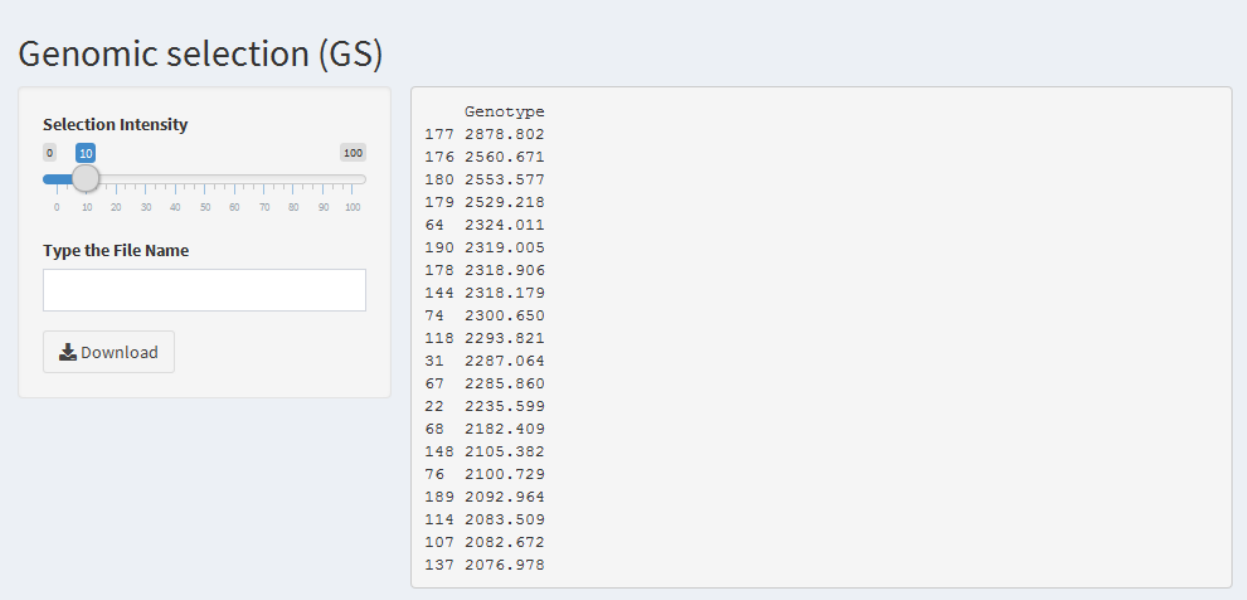


Figure 43. Selection – Prediction and Selection – Genomic Selection tab from Be-Breeder 2.0

3.3 Genomic Association (GWAS)

Determination of genic regions related to a trait (QTL) through molecular markers is an initial step of extreme importance for understanding the regulation and genetic structure of a trait of interest (Korte and Farlow, 2013). To do so, an important tool in breeding is genome wide association studies (GWAS), which uses mixed model equations associated with multiple regression to identify molecular markers correlated with the trait of interest according to a threshold (LOD).

In the *GWAS* section of the Be-Breeder, the user supplies the phenotypic data (Figure 44) and the genotypic information, which can be either a HapMap (marker position in the genome) or the transpose of the Z matrix (Figure 45). From these data and through the "*GWAS*" function of the rrBLUP R package (Endelman, 2011), analysis of association between the markers and the QTL for the trait under study is carried out, considering the model $y = X\beta + Zg + S\tau + \epsilon$, in which y is the phenotype vector, β is the fixed effects vector, g is the genetic effects vector given as random, τ is the additive effects vector of the SNP being considered as fixed, and ε is the residue vector. *X*, *Z*, and *S* are incidence matrices of the model.In addition to the estimates of marker effects (Figure 46), the application displays a Manhattan plot (Figure 47).

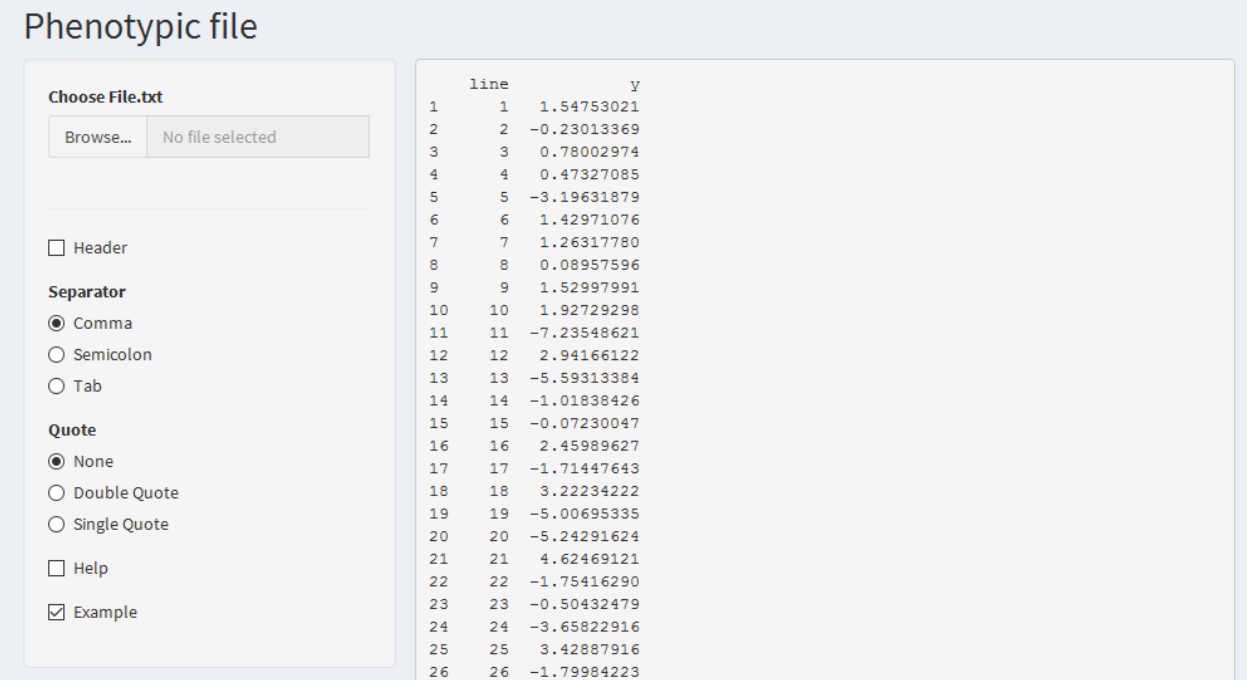


Figure 44. Phenotypic file – GWAS tab from Be-Breeder 2.0

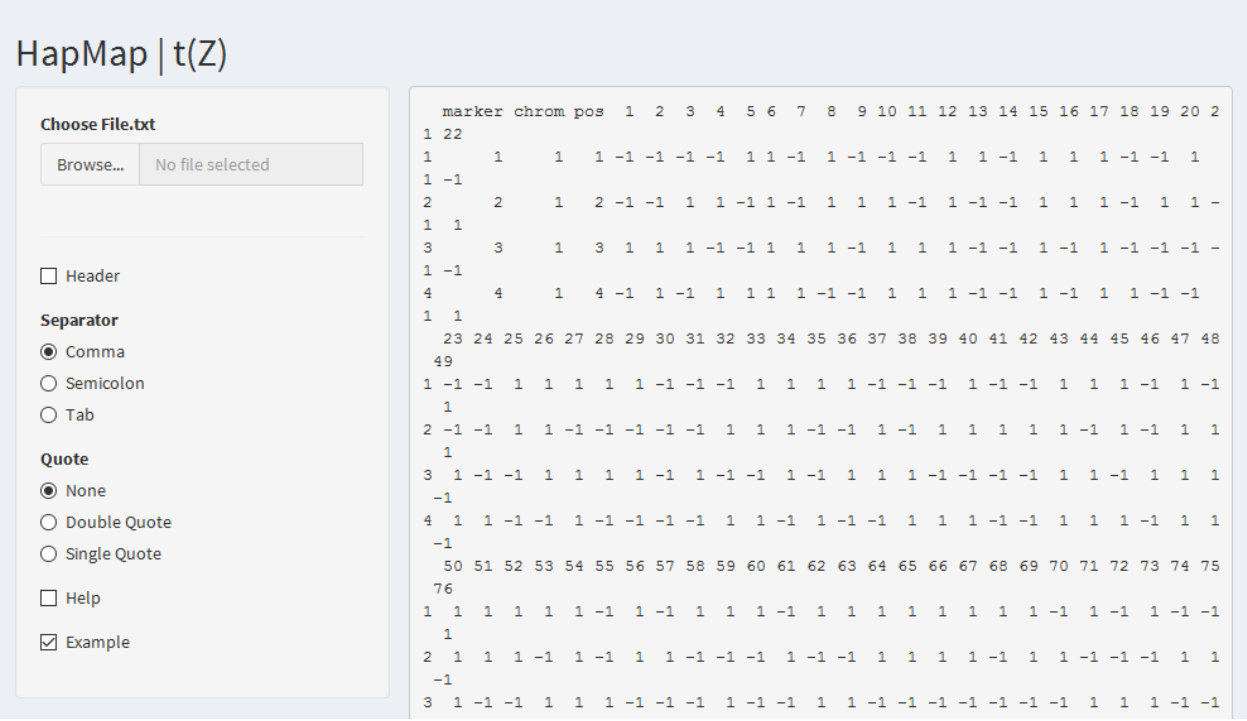


Figure 45. HapMap | t(Z) – GWAS tab from Be-Breeder 2.0

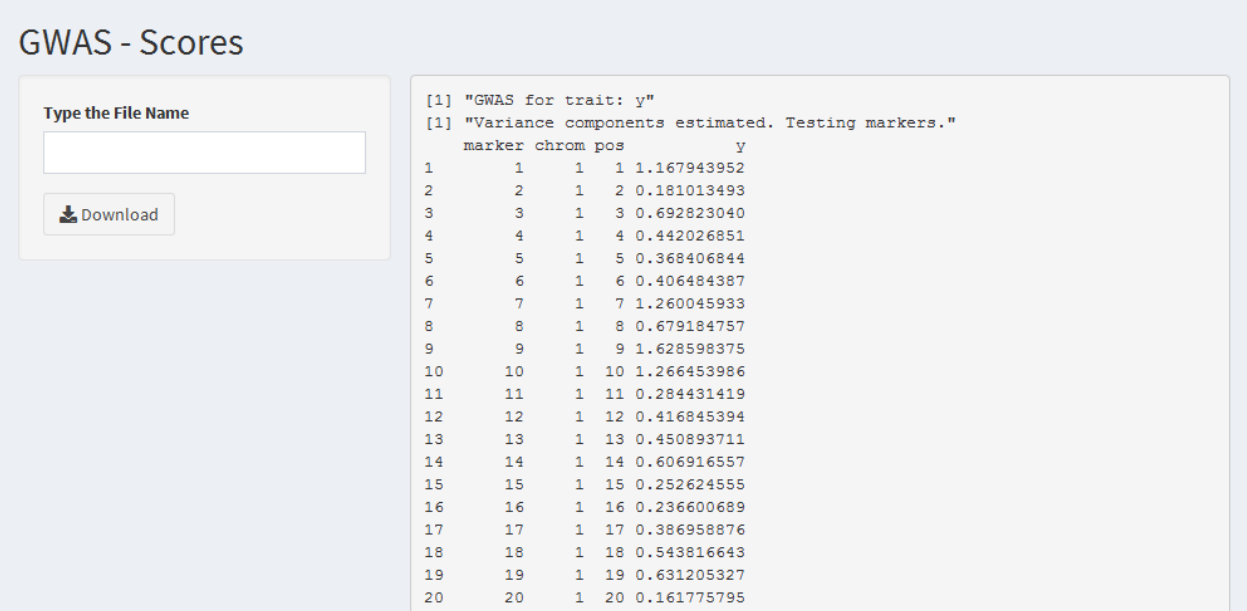


Figure 46. Scores – GWAS tab from Be-Breeder 2.0

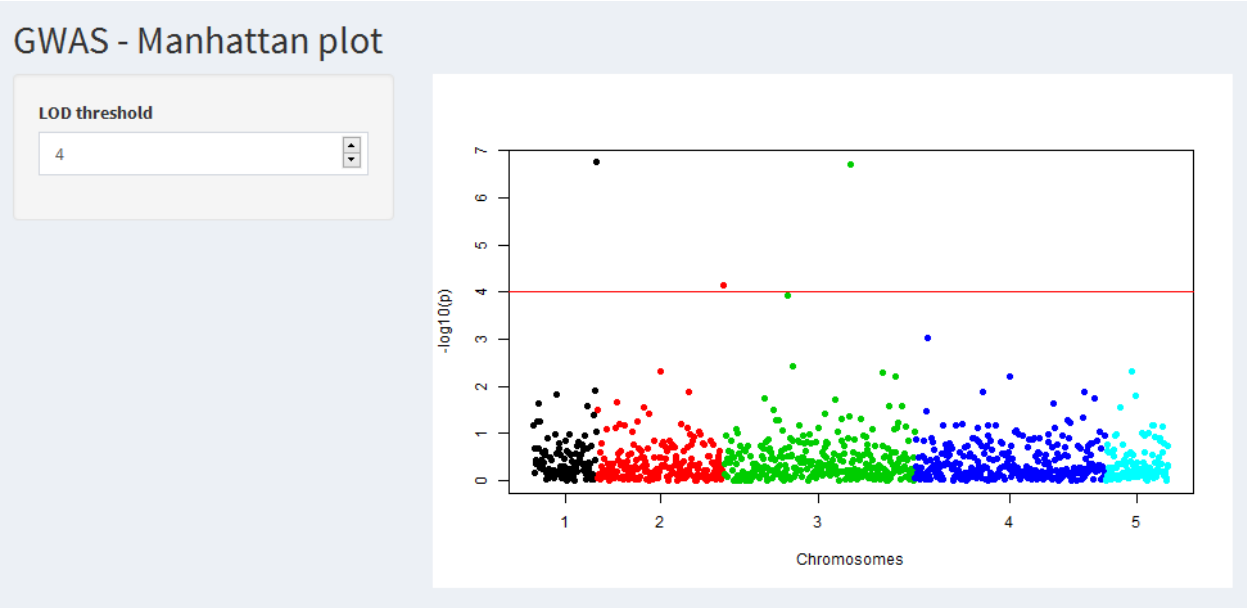


Figure 47. Manhattan plot – GWAS tab from Be-Breeder 2.0

3.4 Diversity Analysis

3.4.1 Genetic Diversity

In this tab, the user can enter files in text format (.txt) generated by the *Genpop* (Raymond and Rousset, 1995) or *Structure* (Pritchard et al., 2000) softwares, as indicated in the Help option and in the Example option file (Figure 48). These analyses were constructed through scripts that associate functions of the *ape* (Paradis et al., 2004), *poppr* (Kamvar et al., 2014),

and *adegenet* (Jombart, 2008) R packages. As output, the user has a general summary with information on the populations and markers (Figure 49). From these estimates, observed heterozygosity (Ho) is obtained by dividing the number of heterozygotes in each locus by the number of individuals and expected heterozygosity (*He*) $He = 1 = \sum_{i=1}^{\neq alleles} p_i^2$, for p in reference to the frequency of the i-th allele (Nei, 1978). The Wright's inbreeding coefficient F is estimated as $F = \frac{He-Ho}{He}$ (Wright, 1965). Several graphic outputs are provided, from Number of alleles per loci to Neighbor joining (Figure 50).

Figure 48. Dataset – Genetic Diversity – Diversity Analysis tab from Be-Breeder 2.0

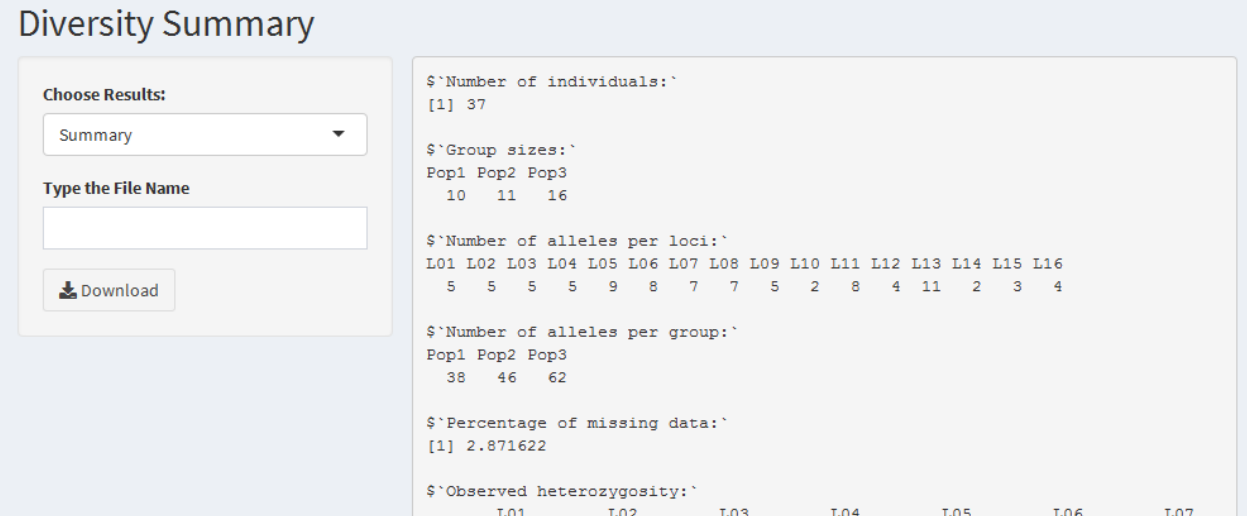


Figure 49. Diversity Summary – Genetic Diversity – Diversity Analysis tab from Be-Breeder 2.0

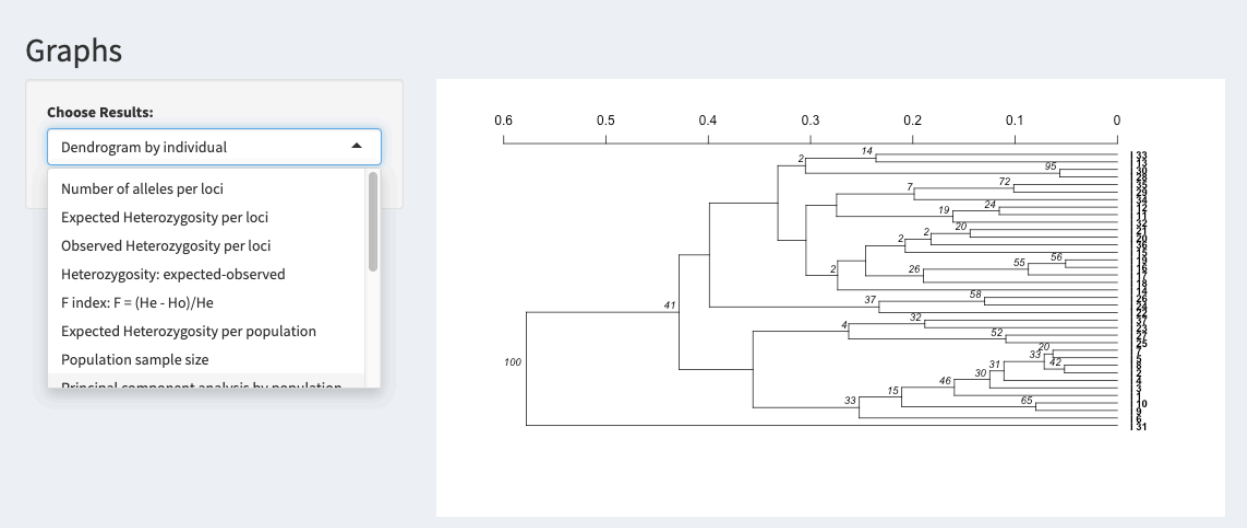
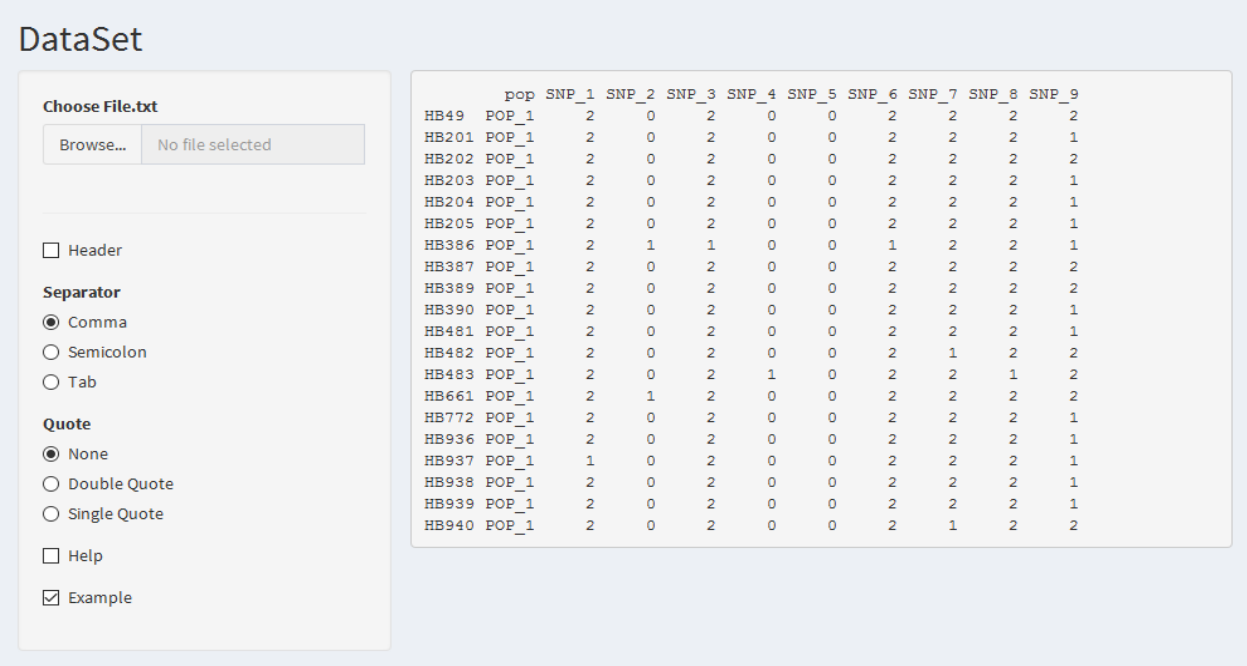


Figure 50. Graphs – Genetic Diversity – Diversity Analysis tab from Be-Breeder 2.0

3.4.2 Discriminant Analysis

The aim in this tab is to graphically analyze population divergence using molecular markers. It can be performed at two levels: individuals (*i*) within a population and (*ii*) among populations. The user must provide a marker matrix coded as (0,1,2), with individuals in the rows and markers in the columns (Figure 51). The outcomes are displayed in four graphs: Barplot of population size, Clusters, Scatterplot, and Compoplot (Figure 52). Packages used in here are *adegenet* (Jombart, 2008) and *ade4* (Dray and Dufour, 2007).



	pop	SNP_1	SNP_2	SNP_3	SNP_4	SNP_5	SNP_6	SNP_7	SNP_8	SNP_9
HB49	POP_1	2	0	2	0	0	2	2	2	2
HB201	POP_1	2	0	2	0	0	2	2	2	1
HB202	POP_1	2	0	2	0	0	2	2	2	2
HB203	POP_1	2	0	2	0	0	2	2	2	1
HB204	POP_1	2	0	2	0	0	2	2	2	1
HB205	POP_1	2	0	2	0	0	2	2	2	1
HB386	POP_1	2	1	1	0	0	1	2	2	1
HB387	POP_1	2	0	2	0	0	2	2	2	2
HB389	POP_1	2	0	2	0	0	2	2	2	2
HB390	POP_1	2	0	2	0	0	2	2	2	1
HB481	POP_1	2	0	2	0	0	2	2	2	1
HB482	POP_1	2	0	2	0	0	2	1	2	2
HB483	POP_1	2	0	2	1	0	2	2	1	2
HB661	POP_1	2	1	2	0	0	2	2	2	2
HB772	POP_1	2	0	2	0	0	2	2	2	1
HB936	POP_1	2	0	2	0	0	2	2	2	1
HB937	POP_1	1	0	2	0	0	2	2	2	1
HB938	POP_1	2	0	2	0	0	2	2	2	1
HB939	POP_1	2	0	2	0	0	2	2	2	1
HB940	POP_1	2	0	2	0	0	2	1	2	2

Figure 51. Dataset – Discriminant Analysis – Diversity Analysis tab from Be-Breeder 2.0

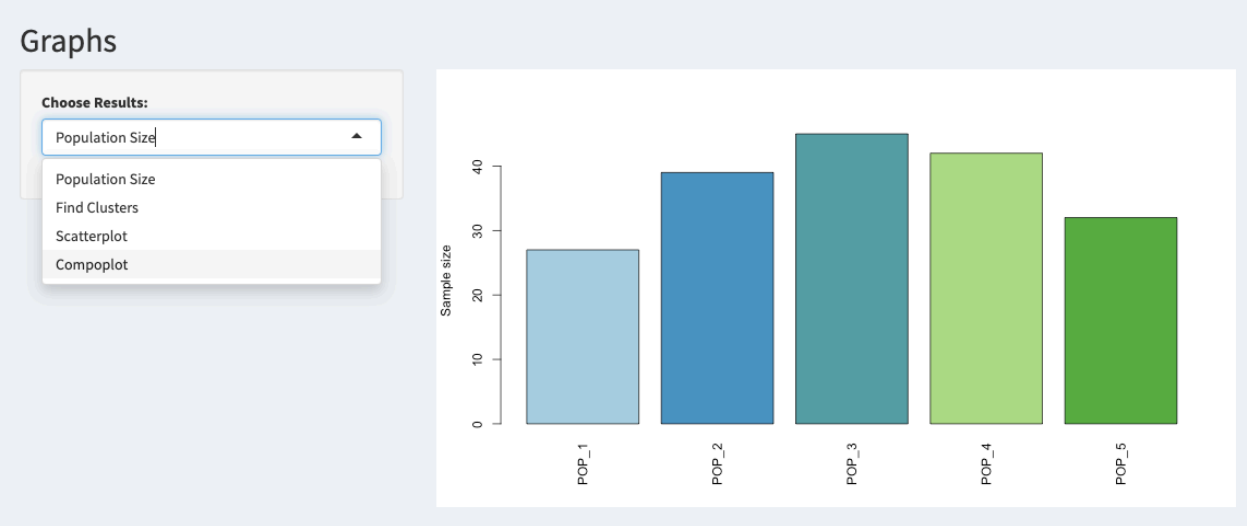


Figure 52. Graphs – Discriminant Analysis – Diversity Analysis tab from Be-Breeder 2.0

3.5 Population Genetics

This tab uses the *popgen* function from the *snpReady* R package and allows getting information regarding allele frequency per marker (p and q), minor allele frequency (MAF), expected heterozygosity (He) estimated through the formula $He = 2 * p * q$, observed heterozygosity (Ho), genetic diversity (DG), obtained by $DG = -1 - p^2 - q^2$, and polymorphic information content (PIC) estimated by $PIC = -1 - p^2 + q^2 - (2p^2q^2)$ (Hartl and Clark, 2010).

First, the function allows to attribute individuals to subpopulations (Figure 53). If this piece of information is not supplied, the application will consider a single population. Then, the user must provide the marker matrix coded as (0,1,2) (Figure 54). The outcomes are displayed in PopGen Analysis tab (Figure 55), where it is possible to observe each one of the parameters listed above for each subpopulation, as well as the existence of exclusive, absent, and fixed alleles.

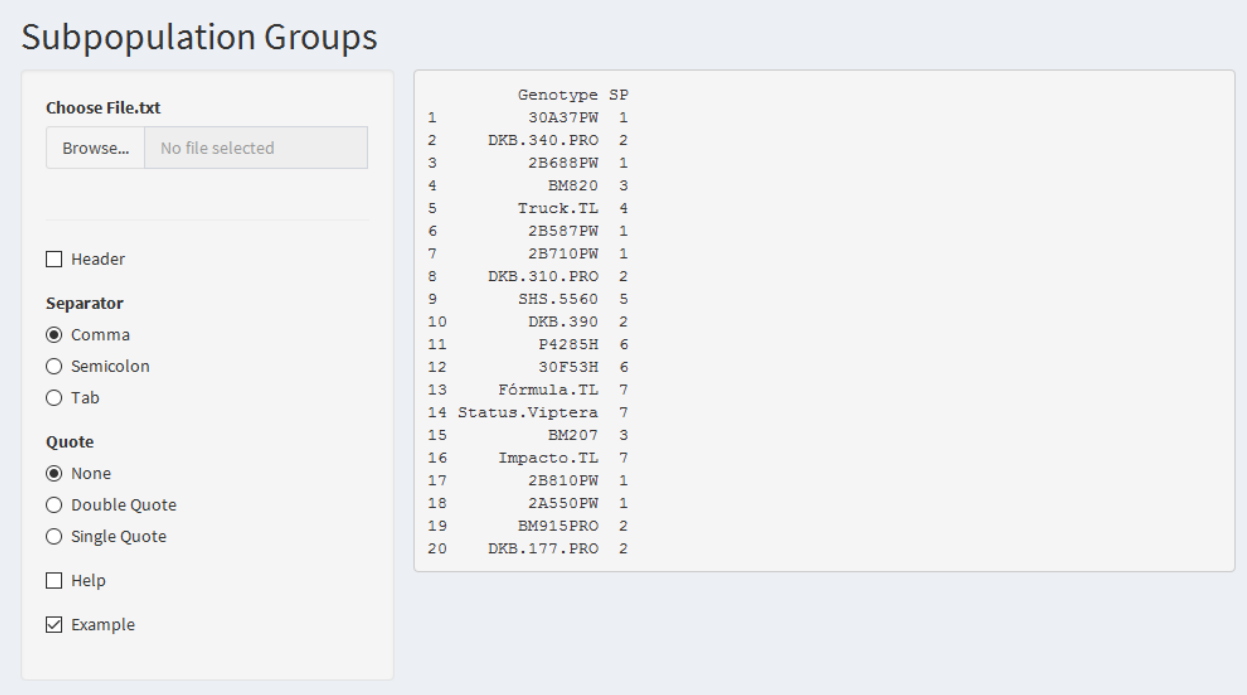


Figure 53. Subpopulation Groups – Population Genetics – Diversity Analysis tab from Be-Breeder 2.0

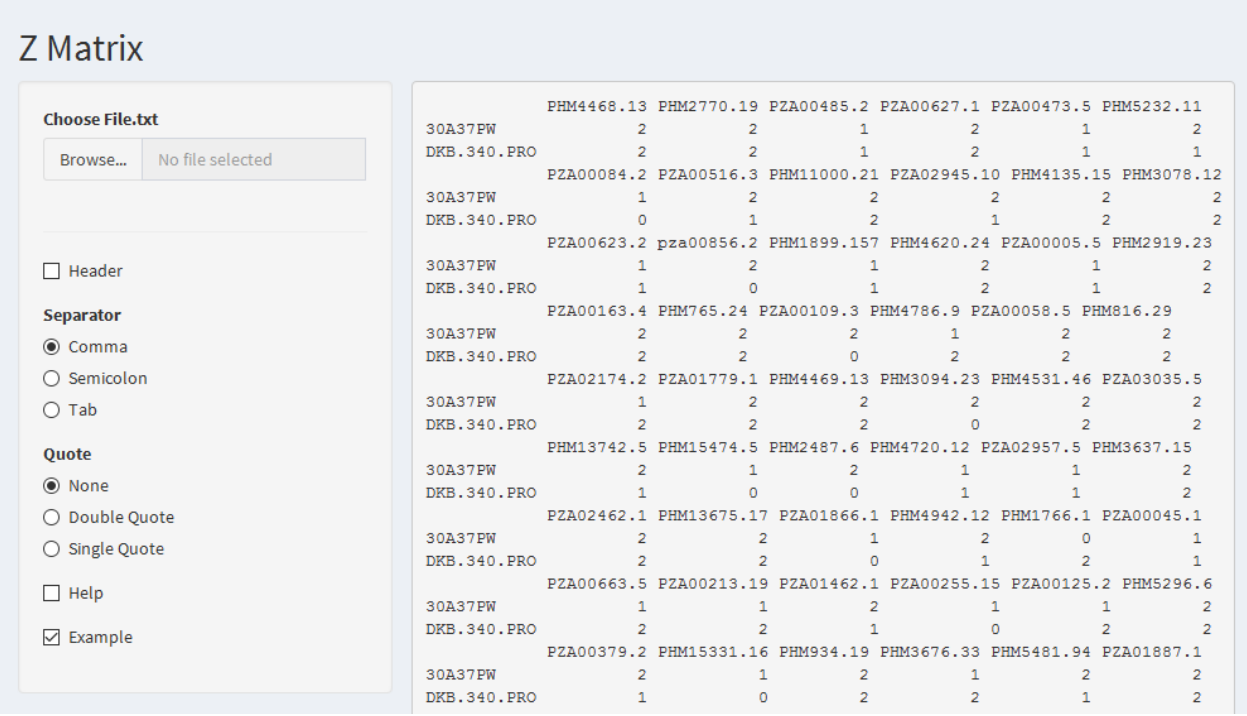
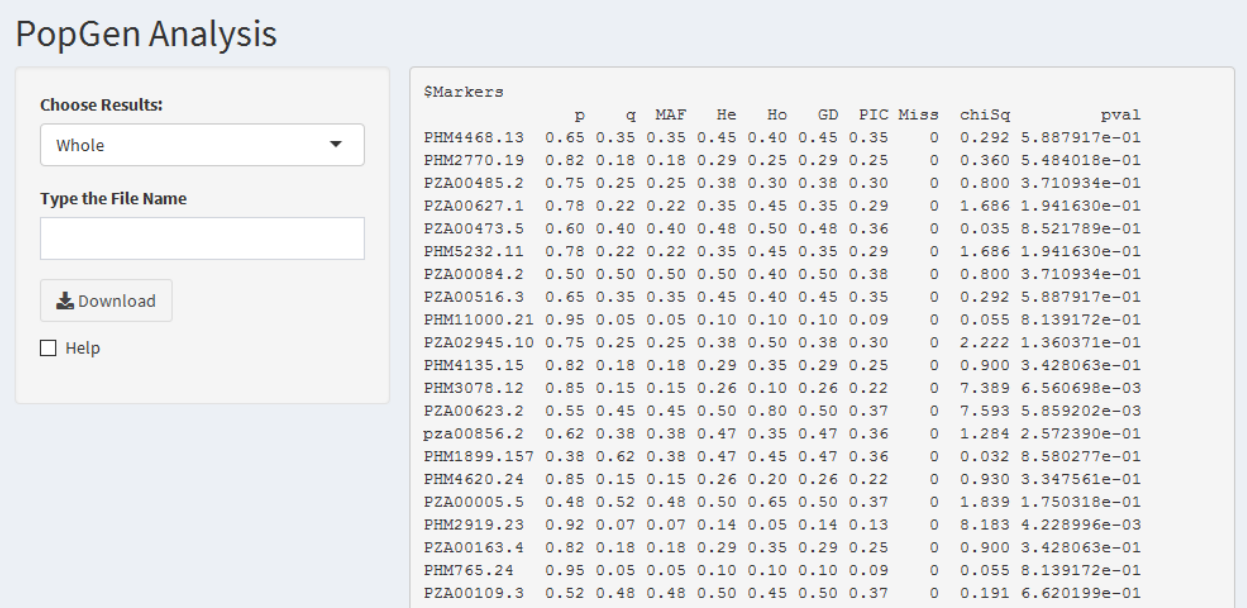


Figure 54. Z Matrix – Population Genetics – Diversity Analysis tab from Be-Breeder 2.0



$Markers

	p	q	MAF	He	Ho	GD	PIC	Miss	chiSq	pval
PHM4468.13	0.65	0.35	0.35	0.45	0.40	0.45	0.35	0	0.292	5.887917e-01
PHM2770.19	0.82	0.18	0.18	0.29	0.25	0.29	0.25	0	0.360	5.484018e-01
PZA00485.2	0.75	0.25	0.25	0.38	0.30	0.38	0.30	0	0.800	3.710934e-01
PZA00627.1	0.78	0.22	0.22	0.35	0.45	0.35	0.29	0	1.686	1.941630e-01
PZA00473.5	0.60	0.40	0.40	0.48	0.50	0.48	0.36	0	0.035	8.521789e-01
PHM5232.11	0.78	0.22	0.22	0.35	0.45	0.35	0.29	0	1.686	1.941630e-01
PZA00084.2	0.50	0.50	0.50	0.50	0.40	0.50	0.38	0	0.800	3.710934e-01
PZA00516.3	0.65	0.35	0.35	0.45	0.40	0.45	0.35	0	0.292	5.887917e-01
PHM11000.21	0.95	0.05	0.05	0.10	0.10	0.10	0.09	0	0.055	8.139172e-01
PZA02945.10	0.75	0.25	0.25	0.38	0.50	0.38	0.30	0	2.222	1.360371e-01
PHM4135.15	0.82	0.18	0.18	0.29	0.35	0.29	0.25	0	0.900	3.428063e-01
PHM3078.12	0.85	0.15	0.15	0.26	0.10	0.26	0.22	0	7.389	6.560698e-03
PZA00623.2	0.55	0.45	0.45	0.50	0.80	0.50	0.37	0	7.593	5.859202e-03
pza00856.2	0.62	0.38	0.38	0.47	0.35	0.47	0.36	0	1.284	2.572390e-01
PHM1899.157	0.38	0.62	0.38	0.47	0.45	0.47	0.36	0	0.032	8.580277e-01
PHM4620.24	0.85	0.15	0.15	0.26	0.20	0.26	0.22	0	0.930	3.347561e-01
PZA00005.5	0.48	0.52	0.48	0.50	0.65	0.50	0.37	0	1.839	1.750318e-01
PHM2919.23	0.92	0.07	0.07	0.14	0.05	0.14	0.13	0	8.183	4.228996e-03
PZA00163.4	0.82	0.18	0.18	0.29	0.35	0.29	0.25	0	0.900	3.428063e-01
PHM765.24	0.95	0.05	0.05	0.10	0.10	0.10	0.09	0	0.055	8.139172e-01
PZA00109.3	0.52	0.48	0.48	0.50	0.45	0.50	0.37	0	0.191	6.620199e-01

Figure 55. PopGen Analysis – Population Genetics – Diversity Analysis tab from Be-Breeder 2.0

References

Akdemir D, Sanchez JI, Jannink JL (2015) Optimization of genomic selection training populations with a genetic algorithm. **Genetics Selection Evolution.** 47(1): 38. doi: 10.1186/s12711-015-0116-6.

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.

Bos I, Caligari P (2008) **Selection Methods in Plant Breeding**. Springer Netherlands, Dordrecht.

Chang W, Cheng J, Allaire J, Xie Y and McPherson J (2015) Shiny: web application framework for R. **R package version 0.11 1**.

Comstock RE, Robinson HF, Harvey, PH (1949) A Breeding Procedure Designed To Make Maximum Use of Both General and Specific Combining Ability. **Agronomy Journal.** 41(8): 360. doi: 10.2134/agronj1949.00021962004100080006x.

Dray S, Dufour AB (2007) The ade4 package: implementing the duality diagram for ecologists. **Journal of Statistic Software**. 22(4): 1–20.

Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. **The Plant Genome 4**: 250-255.

Falconer DS, Mackay TF and Frankham R (1996) Introduction to quantitative genetics (4th edn). **Trends in Genetics 12**: 280.

Gardner CO, Eberhart AS (1966) Analysis and interpretation of the variety cross diallel and related populations. **Biometrics** 22(3): 439–52..

Granato ISC, Galli G., Couto, E.G.O, Souza MB, Mendonça LF, Fritsche-Neto R (2018). snpReady: a tool to assist breeders in genomic analysis. **Molecular Breeding** 38:102.

Griffing B (1956) Concept of general and specific combining ability in relation to diallel crossing systems. **Australian Journal of Biological Sciences 9**: 463-493.

Hartl DL and Clark A (2010) **Princípios de genética de populações**. Artmed, Porto Alegre, 660p.

Hartl DL, Clark AG (2007) **Principles of population genetics**. Sinauer Associates.

Heffner EL, Lorenz AJ, Jannink J-L and Sorrells ME (2010) Plant breeding with genomic selection: gain per unit time and cost. **Crop Science 50**: 1681.

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. **Theoretical and Applied Genetics**. 38(6): 226–231. doi: 10.1007/BF01245622.

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.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. **Bioinformatics 24**: 1403-1405.

Jones DF (1918) The effect of inbreeding and crossbreeding upon development. **Proceedings of the National Academy of Sciences 4**: 246-250.

Kamatani Y (2016) Genome wide association study: its theory and methodological review. **Clinical Calcium 26**: 525.

Kamvar ZN, Tabima JF and Grünwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. **PeerJ 2**: 281.

Korte A and Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. **Plant Methods 9**: 1.

Lynch M and Walsh B (1998) **Genetics and analysis of quantitative traits***.* Sinauer Associates, Sunderland, 980p.

Mendiburu F (2017) agricolae: Statistical Procedures for Agricultural Research. R Packag. version 1.2-8. doi: https://cran.r-project.org/package=agricolae.

Mendiburu Fd (2016) Agricolae: statistical procedures for agricultural research. **R package version 1.2-4**: 1-6.

Meuwissen THE, Hayes BJ and Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. **Genetics 157**: 1819-1829.

Muñoz F, Sanchez L (2017) breedR: Statistical Methods for Forest Genetic Resources Analysts. https://github.com/famuvie/breedR.

Nass L (2001) **Recursos genéticos e melhoramento-Plantas***.* Fundação MT, Rondonópolis, 1183p.

Nei M (1972) Genetic distance between populations. **American Naturalist**: 283-292.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. **Genetics 89**: 583-590.

Nietlisbach P, Keller L and Postma E (2016) Genetic variance components and heritability of multiallelic heterozygosity under inbreeding. **Heredity 116**: 1-11.

Paradis E, Claude J and Strimmer K (2004) APE: analysis of phylogenetics and evolution in R language. **Bioinformatics 20**: 289-290.

Pritchard JK, Stephens M and Donnelly P (2000) Inference of population structure using multilocus genotype data. **Genetics 155**: 945-959.

Ramalho M, Santos JB and Pinto CB (2008) **Genética na agropecuária***.* UFLA, Lavras, 463p.

Raymond M and Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. **Journal of Heredity 86**: 248-249.

Resende MDV (2002) **Genética biométrica e estatística no melhoramento de plantas perenes***.* Embrapa Informação Tecnológica, Brasília, 975p.

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 (2019) Gain reports. Available at <http://gain.fas.usda.gov/Pages/Default.aspx>. Access on July 28, 2016.

VanRaden P (2008) Efficient methods to compute genomic predictions. **Journal of Dairy Science 91**: 4414-4423.

Vencovsky R and Barriga P (1992) **Genética biométrica no fitomelhoramento***.* Revista Brasileira de Genética, Ribeirão Preto, 486p.

Wright K (2017) corrgram: Plot a Correlogram. R Packag. version 1.12. doi: https://cran.r-project.org/package=corrgram.

Wright S (1923) The theory of path coefficients a reply to Niles's criticism. **Genetics 8**: 239-255.

Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. **Evolution**: 395-420.

Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG and Montgomery GW (2010) Common SNPs explain a large proportion of the heritability for human height. **Nature Genetics 42**: 565-569