ConservaGen software: A useful tool for genetic conservation of germplasm

Andrei Caíque Pires Nunes[1*]

Abstract: *ConservaGen software can assist germplasm conservation projects in terms of population genetics. It can be used for both in situ and ex situ germplasm conservation and can generate parameters to assist in decision-making in these projects. ConservaGen is freely available and can be downloaded from https://gpfsb.webnode.com/software/.*

Keywords: *Genetic diversity, inbreeding rate, population genetics, quantitative genetics, restoration projects*

INTRODUCTION

The vast majority of germplasm conservation, reforestation and/or restoration projects have been conducted without considering the population genetics of the target species. Most of these actions are based only on the ease of obtaining seeds, and the lack of knowledge regarding the genetic parameters causes inappropriate assembly of plantations. Seeds are usually collected from one or a few seed-trees and planted in the area to be restored or conserved, without quantifying the genetic impacts of such actions. In the long term, these actions can lead to the contraction of the genetic base of these populations, thereby causing genetic drift, and consequently the extinction of the species involved in these projects (Resende and Vencovsky 1990, Resende 2002, Sonstebo et al. 2018).

To accurately implement the germplasm conservation, reforestation and/or restoration projects, the aspects related to population genetics and genetic diversity must be taken into consideration. Therefore, establishing a suitable effective population size (N_e), low inbreeding rates, reduce the expected decrease in heterozygosity and limiting the number of individuals of the same family in the projects, would ensure the maintenance of rare alleles with adaptive significance in the conserved population (Vencovsky and Crossa 1999, Vencovsky et al. 2007, Arantes et al. 2010, Sonstebo et al. 2018, Castro et al. 2019, Guimarães et al. 2019). These alleles are important for the longevity of the population in terms of genetic diversity and the maintenance of evolutionary potential over several generations (Resende 2002, Angeloni et al. 2011, Snowdon et al. 2015, Mistro et al. 2019).

The retention of rare alleles in plant populations can be assessed and must be considered in the genetic conservation and reforestation programs. In the case of perennial plants, inappropriate decisions related to the establishment of a genetic pool of the population may delay the reproductive success of the plants, thereby compromising the genetic diversity of populations and natural

*Corresponding author:
E-mail: andrei.nunes@gfe.ufsb.edu.br
ORCID: 0000-0002-9536-5226

[1] Universidade Federal do Sul da Bahia, Centro de Formação em Ciências Agroflorestais, 45.613-204, Itabuna, BA, Brazil

fitness of individuals in the subsequent generations (Hallander and Waldmann 2009). Since in populations with low diversity, genetic erosion tends to compromise the future adaptability of the plants, therefore, assessing the conservation efficiency using a genetic aspect guarantees successful implementation of these strategies (Batista et al. 2012, Souza et al. 2017). Additionally, it helps to preserve a population with optimal N_e, which is capable of generating viable seeds (Ottewell et al. 2016, Souza et al. 2017).

Despite the relevance of considering the parameters of population genetics, the use of these concepts has been highly restricted to the scientific, public, and theoretical research. Considering these aspects, the ConservaGen software has been developed to provide practical knowledge on population genetics and assist in the decision-making process in the germplasm conservation projects.

CONSERVAGEN SOFTWARE AND ITS APPLICATIONS

General information

ConservaGen software has been developed in C# language, and it is free and operates on the Windows operating system interface. It has been developed to assist in the decision-making process in both *in situ* and *ex situ* germplasm genetic conservation projects. ConservaGen software would help the decision makers of a given conservation project to answer questions such as:

- How many seed-trees must be sampled to guarantee a genetic basis of the future populations?

- How many individuals per seed-tree should be planted to guarantee an ideal genetic basis of a population?

- What is the expected decrease in heterozygosity of the conserved populations?

- What is the minimum frequency of alleles retained in the population?

- How efficient is the genetic conservation of populations?

ConservaGen software comprises the procedures for assessing allogamous, autogamous and mixed mating system species. Additionally, it is possible to consider the collection of seeds in the seed-tree located in one or several independent locations, as well as their genetic conservation both *in situ* or *ex situ*. The software also has two additional procedures for establishing genetic improvement experiments. In these modules, the decision maker can generate random numbers, determine the number of repetitions necessary to obtain adequate selective accuracy, and measure the degree of genetic diversity through selecting an effective population size to be used for the experiments. This determination is made to orientate the selection optimization of recombination orchards to test the progenies of full and half sibling families.

Estimation of genetic representativeness of populations

The genetic representativeness of a population depends on the number of seed-trees sampled (Nf) and number of individuals sampled per seed-tree (kf). This representativeness can be measured using the effective population size (N_e) and frequency of the retained alleles (FRA) (Resende and Vencovsky 1990, Resende 2002).

ConservaGen software has been developed to meet the demands of the plant germplasm conservation projects and allows the estimation of the genetic representativeness of a population. To set this estimation, it is necessary to consider the different reproductive systems of the target species. For example, in case of monoecious allogamous species, with an equal number of individuals collected per seed-tree, N_e can be estimated as follows (Resende 2002):

$$N_e = \frac{4N_f k_f}{k_f + 3}$$

When different number of individuals collected per seed-tree is involved, the effective population size of monoecious allogamous species can be estimated as follows (Resende 2002):

$$N_e = \frac{4N_f \overline{k}_f}{\overline{k}_f + 3 + (\sigma_{kf}^2 / \overline{k}_f)}$$

where: $\overline{k}_f$: average number of individuals selected per seed-tree, σ_{kf}^2: the variance of the number of individuals selected per seed-tree.

Considering dioecious allogamous species, with an equal number of individuals collected per seed-tree, female and male gametic control and equal proportions of male and female offspring, N_e can be estimated as follows (Resende 2002, Vencovsky et al. 2012):

$$N_e \approx \frac{4N_f k_f}{k_f + 1}$$

In case of mixed mating system, N_e can be estimated as follows (Resende 2002):

$$N_e = \frac{2(2-S)N_f k_f}{(1+S)^2 k_f + (3-2S-S^2)}$$

where S is the self-fertilization rate.

For autogamous species N_e can be estimated as follows (Resende 2002):

$$N_e = 0.5N_f$$

The values of N_e can be used to measure the degree of genetic diversity in a population (Resende 2002). These expressions above estimated the N_e for progeny arrays, assuming: 1) that seed-trees are not genetic related; 2) individuals within family are half-sibs. Based on this, the expected decrease in heterozygosity (F) or potential inbreeding rate when only autozygous genotypes are involved follows:

$$F = \frac{1}{2N_e}$$

Sampling in independent populations can elevate the genetic diversity of the target germplasm. This procedure does not apply when the reference population is structured in subpopulations. Gathering R independent samples in quantities of propagules or individuals, each with arbitrary effective sizes N_{e1}, N_{e2},.., N_{eR}, the effective size (Resende 2002) of the composite sample (N_{ec}) can be estimated as follows:

$$N_{ec} = \frac{R^2}{\sum_j^R \frac{1}{N_{ej}}} = R \times \bar{N}_e$$

where N_{ej}: the effective size of each sample, where each independent sample represents a provenance for seed-tree sampling, $\bar{N}_e$: the harmonic mean of N_{ej}, and j: 1, 2, 3, ... R.

Frequency of retained alleles (FRA) and germplasm conservation efficiency

With the estimation of N_e value, it is possible to infer the frequency of alleles in the original population that was captured in the sample through calculating the minimum FRA for each N_e. The minimum FRA represents the lower limit of the confidence interval (CI) for the allele frequency in a given sample (Resende 2002). To obtain the lower and upper limits of CI, the following expression can be applied:

$$C.I. = p_0 \pm z \left\{ \frac{[p_0(1-p_0)]}{[2Ne]} \right\}^{1/2}$$

where, z: the tabulated value of the standard normal distribution associated with a certain degree of confidence, equivalent to 1.96 to 95% confidence, and p_0: the parametric frequency of the alleles in the original population.

Upon calculating each alleles frequency and respective N_e, the minimum FRA can be obtained (Table 1). The procedure to calculate FRA using ConservaGen software is named *FAR*, and it allows the user to estimate the minimum FRA of a given population with a specific N_e.

For the conservation of natural genetic resources, ConservaGen software can provide both *in situ* and *ex situ* conservation strategies, which can be implemented for sampling the germplasm and maintaining the genetic diversity of the populations.

Considering N_e of 175 or 200, it is possible to capture the alleles with a minimum frequency of 2% (Table 1).

Table 1. Minimum frequency values of the retained alleles with different effective sizes

N_e	Minimum FRA
30	7.0%
40	6.0%
50	5.0%
100	3.0%
150	3.0%
175	2.0%
200	2.0%
250	2.0%
270	2.0%
1000	1.0%

According to Resende (2002), N_e of 200 is recommended for *ex situ* germplasm conservation, while N_e ranging from 500 to 5000 can be applied for *in situ* conservation. From these reference values, the efficiency of the genetic conservation of a germplasm was set in the ConservaGen software.

Genetic breeding procedures

In addition to the procedures directly related to the germplasm conservation, ConservaGen software has useful modules for decision making in genetic improvement. It follows the applications that can be easily managed and assist the breeding programs: a) definition of replication number; b) random number generation; and c) determination of the genetic diversity in breeding populations to subsidize the selection and optimization of hybridization orchards.

To conduct breeding experiments, it is essential to have detailed knowledge regarding the best trial design and number of replications to accurately estimate the genetic parameters and select the potential genotypes (Binkley et al. 2017). Experimental designs must allow genetic selection to occur in an optimized and accurate manner. Determining the ideal number of replications according to high selective accuracy is crucial for the success of the breeding programs (Stanger et al. 2011). Considering these aspects, ConservaGen software can be used to simulate scenarios with different replications and accuracies, which may assist in the decision making of the breeder.

According to the accuracy expression reported by Resende (2002), the experimental quality is directly related to the heritability of the measured traits and number of replicates used. Considering the traits with low (0.10), low-medium (0.20), medium (0.30), and high (0.40) levels of heritability, the lower the genetic control of the trait, the greater the number of repetitions have to be performed to achieve the accuracy above 0.90 (Figure 1).

In addition to determine the accuracy, ConservaGen software assists the breeders in studying the genetic diversity of breeding populations and hybridization orchards. The selection and optimization of hybridization orchards is important for establishing a suitable N_e and limiting the number of individuals of the same family during the crosses. This procedure ensures the maintenance of rare alleles in the base population during subsequent breeding cycles (Sonstebo et al. 2018). It also preserves the elite genotypes with the potential to ensure high gains without an exhausted genetic basis (Castro et al. 2019, Nogueira et al. 2019).
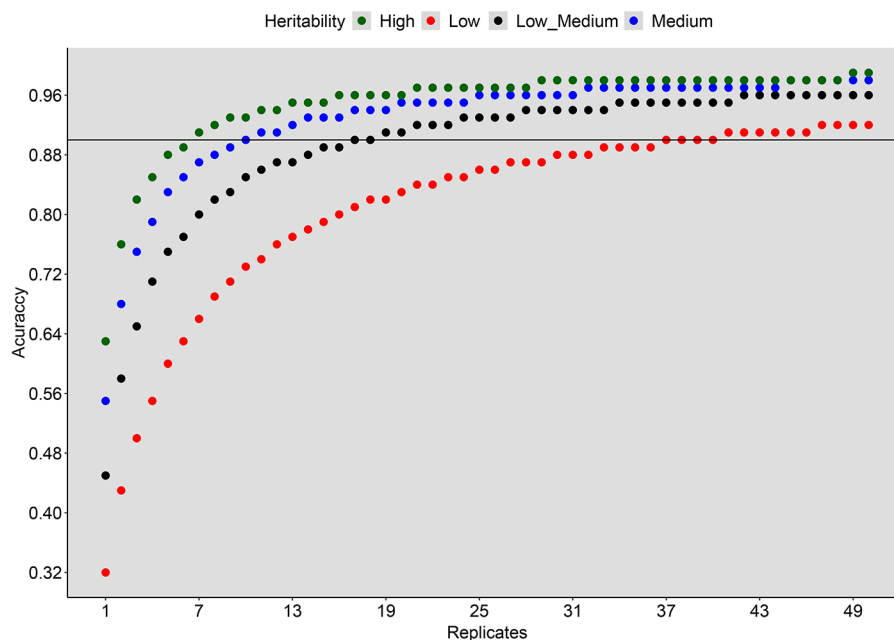

Figure 1. Different accuracy values for distinct numbers of repetitions as a function of heritability magnitudes, considering the traits with low (<0.10), low-medium (<0.20), medium (<0.30) and high (<0.40) levels of heritability.

Using the software

For genetic conservation procedures, the user must select the species' reproductive system, the number of seed collection sites (from one to five) and if the number of sampled seedlings for seed-tree is equal or variable. The seedlings of each seed-tree are used to set the conservation or restoration project. Then, the user must fill in the blanks with the number of seed-tree to be sampled and the number of individuals per seed-tree. After clicking calculate bottom, N_e, F and conservation efficiency values will appear, in order to assist decision making and simulate the ideal situation to be set (Nunes et al. 2021). The conservation efficiency value is based on the allele retention capacity according to FRA, referring to each ideal N_e for an *ex situ* or in situ conservation.

The applications that can be easily managed and assist the breeding programs are: a) definition of replication number; b) random number generation; and c) determination of the genetic diversity in breeding populations to subsidize the selection and optimization of hybridization orchards. To define the number of repetitions for a clonal test, the user must select the pre-experiment window and fill in the blanks with the number of repetitions and the heritability value of the target trait. After clicking calculate bottom, the value of accuracy and medium heritability appear, in order to assist decision making and simulate the ideal situation to be set. Inside the pre-experiment procedure window, the user can easily simulate different sequences of random numbers just filling in the information requested in this procedure. The post-experiment window allows de user to access the determination of the genetic diversity of hybridization orchards. To set that, the user must select the type of recombination orchard and fill in the blanks with the number of families to be maintained in the orchard and the number of individuals per family. After clicking calculate bottom, N_e, F and efficiency values will appear, in order to assist decision making and simulate de ideal situation to be set.

FINAL CONSIDERATION

ConservaGen software was designed to boost attitudes towards genetic conservation and restoration of degraded areas considering the genetic diversity of populations. Currently Brazilian law allows the restoration projects to be composed by seedlings from a few or a single tree, which will certainly result in genetic drift and severe negative effects of the genetic load in the long term. The program is intended to be used by researches as well as the community outside the scientific public. Thus, ConservaGen software is easy to use and interpret, and allows efficient handling of the most common situations related to the population genetic representativeness of plant germplasm for many different plant species, especially with varied mating systems from different seed collection areas. Moreover, ConservaGen software can be applied in genetic breeding decision-making with additional applications for assessing the number of replications, and selection and optimization of hybridization orchards.

REFERENCES

Angeloni F, Ouborg NJ and Leimu R (2011) Meta-analysis on the association of population size and life history with inbreeding depression in plants. **Biological Conservation 144:** 35-43.

Arantes FC, Gonçalves PDS, Scaloppi Junior EJ, Moraes MLTD and Resende MDVD (2010) Ganho genético com base no tamanho efetivo populacional de progênies de seringueira. **Pesquisa Agropecuária Brasileira 45**: 1419-1424.

Batista CM, Freitas MLM, Moraes MA, Zanatto ACS, Santos PC, Zanata M and Sebbenn AM (2012) Estimativas de parâmetros genéticos e a variabilidade em procedências e progênies de *Handroanthus vellosoi*. **Pesquisa Florestal Brasileira 32**: 269-276.

Binkley D, Campoec OC, Alvares C, Carneirod RL, Cegattad I and Stapee JL (2017) The interactions of climate, spacing and genetics on clonal *Eucalyptus* plantations across Brazil and Uruguay. **Forest Ecology and Management 405**: 271-283.

Castro CAO, Nunes ACP, Roque JV, Teófilo RF, Santos OP, Santos GA and Resende M (2019) Optimization of *Eucalyptus benthamii* progeny test based on Near-Infrared Spectroscopy approach and volumetric production. **Industrial Crops and Products 141**: 111786.

Guimarães RA, Miranda KMC, Mota EES, Chaves LJ, Telles MPC and Soares TN (2019) Assessing genetic diversity and population structure in a *Dipteryx alata* germplasm collection utilizing microsatellite markers. **Crop Breeding and Applied Biotechnology 19**: 329-336.

Hallander J and Waldmann P (2009) Optimum contribution selection in large general tree breeding populations with an application to Scots pine. **Theoretical and applied genetics 118**: 1133-1142.

Mistro JC, Resende MDV, Fazuoli LC and Vencovsky R (2019) Effective population size and genetic gain expected in a population of *Coffea canephora*. **Crop Breeding and Applied Biotechnology 19**: 1-7.

Nogueira TAPC, Nunes, ACP, Santos GA, Takahashi EK, Resende MDV and Corradi IS (2019) Estimativa de parâmetros genéticos em progênies de irmãos completos de eucalipto e otimização de seleção. **Scientia Forestalis 47**: 451-462.

Nunes ACP, Resende MDV, Santos GA and Freitas AF (2021) Conservação genética de espécies florestais nativas: número de progênies e indivíduos a conservar para garantir a perpetuação da espécie no ambiente. **Boletim Técnico SIF 5**: 1-6.

Ottewell KM, Bickerton DC, Byrne M and Lowe AJ (2016) A genetic assessment framework for population level threatened plant conservation prioritization and decision-making. **Diversity and Distributions 22**: 174-188.

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

Resende MDV and Vencovsky R (1990) Condução e utilização de bancos de conservação genética de espécies de *Eucalyptus*. **Silvicultura 42**: 435-439.

Snowdon RJ, Abbadi A, Kox T, Schmutzer T and Leckband G (2015) Heterotic haplotype capture: precision breeding for hybrid performance. **Trends in Plant Science 20**: 410-413.

Sonstebo JH, Tollefsrud MM, Myking T, Steffenrem A, Nilsen AE, Edvardsen OM and El-Kassaby YA (2018) Genetic diversity of Norway spruce (*Picea abies* (L.) Karst.) seed orchard crops: Effects of number of parents, seed year, and pollen contamination. **Forest Ecology and Management 411**: 132-141.

Souza TDS, Santos WD, Deniz LD, Alves ADO, Shimizu JY, Sousa VA and Aguiar AV (2017) Variação genética em caracteres quantitativos em *Pinus caribaea* var. hondurensis. **Scientia Forestalis 45**: 177-185.

Stanger TK, Galloway GM and Retief ECL (2011) Final results from a trial to test the effect of plot size on *Eucalyptus* hybrid clonal ranking in coastal Zululand, South Africa. **Southern Forests: a Journal of Forest Science 73**: 131-135.

Vencovsky R and Crossa, J (1999) Variance effective population size under mixed self and random mating with applications to genetic conservation of species. **Crop Science 39**: 1282-1294.

Vencovsky R, Nass LL, Cordeiro CMT and Ferreira MAJF (2007) Amostragem em recursos genéticos vegetais. In Nass LL **Recursos genéticos vegetais**. Embrapa Recursos Genéticos e Biotecnologia, Brasília, p. 231-280

Vencovsky R, Chaves LJ and Crossa J (2012) Variance effective population size for dioecious species. **Crop Science 52**: 79-90.