SOFTWARE RELEASE

'Gametes Simulator': a multilocus genotype simulator to analyze genetic structure in outbreeding diploid species

Crop Breeding and Applied Biotechnology
20(1): e26482019, 2020
Brazilian Society of Plant Breeding.
Printed in Brazil
http://dx.doi.org/10.1590/1984-70332020v20n1s9

Bettina Porta[1*], Peter Fernández[2], Guillermo A. Galván[3] and Federico Condón Priano[4]

Abstract: *Bulk sampling and subsequent DNA fingerprinting are applied to obtain estimated allelic frequency data and unveil population structure for outbreeding species. We developed 'Gametes Simulator', a routine to simulate gametes from allelic frequencies of unlinked loci, which enables the analysis of population structure through Bayesian analysis. Based on the allelic frequencies in the populations, the software simulates the alleles per population in the right proportions and assigns one allele per marker to each gamete following Mendel's laws that govern gametogenesis.*

Keywords: *Bulk sampling, unlinked loci, allelic frequencies.*

INTRODUCTION

Knowledge of the genetic structure of plant populations and germplasm collections is a crucial factor in the development of conservation strategies (Mutegi et al. 2015) and in the exploitation of the breeding potential of domesticated species (Crystian et al. 2018, Ferreira et al. 2018, Guimarães et al. 2019) and their wild relatives (Dempewolf et al. 2017). In animal populations knowledge of the genetic structure of the populations is also crucial to design efficient conservation strategies (Contina et al. 2019). Genetic structure in outbreeding species is determined by multiple factors, such as assortative mating, bottlenecks, genetic drift, kinship, and natural selection (Thornhill 1993), as well as artificial selection in the case of domesticated species (McCouch 2004). The nonexistence of Hardy-Weinberg equilibrium in a population could be the result of the action of one or more factors and occurs frequently in populations of wild and domesticated species (Waples 2014).

Genetic diversity analysis in outbreeding species requires the sampling of a high number of individuals per population (a minimum of twenty-five to thirty individuals) to accurately estimate allelic frequencies within each population (Hale et al. 2012). A high number of individuals per population increase the associated costs, and these costs increase even more when multiple populations must be analyzed. To solve this resource limitation, bulk DNA fingerprinting can be applied, reducing the costs and time involved as a result of bulking the DNA from fifteen or more individuals into a population DNA sample (Reif et al. 2005, Dubreuil et al. 2006, Porta et al. 2018). From these bulk samples, allelic frequencies are precisely estimated (Reif et al.

***Corresponding author:**
E-mail: bporta@fagro.edu.uy
ORCID: 0000-0001-6835-679X

Received: 11 March 2019
Accepted: 31 October 2019
Published: 13 February 2020

[1] Universidad de la República, Facultad de Agronomía, Departamento de Biología Vegetal, 12900, Montevideo, Uruguay
[2] Instituto Nacional de Investigación Agropecuaria, Programa Nacional de Investigación en Pasturas y Forrajes, 70000, Colonia, Uruguay
[3] Universidad de la República, Facultad de Agronomía, Departamento de Producción Vegetal, Centro Regional Sur, 90100, Canelones, Uruguay
[4] Instituto Nacional de Investigación Agropecuaria, Unidad de Semillas y Recursos Fitogenéticos, Estación Experimental La Estanzuela, 70000, Colonia, Uruguay

2005, Dubreuil et al. 2006, Bedoya et al. 2017) but the ability to recognize the genotype of each of the individuals belonging to the bulk sample is lost.

Bayesian analysis using the Structure algorithm (Pritchard et al. 2000) can be used to detect underlying populational genetic structure, which is difficult to detect using phenotypic traits but significant in terms of population genetic differentiation and evolution. This approach estimates the best-fitting number of subpopulations in a sample and predicts the probability that a certain genotype belongs to a subpopulation. Nevertheless, it requires knowledge of the individual profile of diploid or haploid genotypes (Falush et al. 2003).

In the case of bulk sampling, where allele frequencies are estimated precisely (Dubreuil et al. 2006), the simulation of individual gametes based on observed allelic frequencies arises as a suitable solution (Porta et al. 2018). The simulated haploid gametes can then be used as input into the Structure program (Falush et al. 2003) to detect the genetic structure of the populations.

The right assumptions and data selection procedure for the simulation of gametes or individuals are crucial to obtain nonbiased results. Whenever diploid individuals are simulated, on the basis of allelic frequencies, is made the assumption that the population is in Hardy-Weinberg equilibrium. This fact might lead to nonrealistic simulations since deviations from Hardy-Weinberg equilibrium are highly common in natural populations as a result of a variety of biological and non-biological reasons (Waples 2014): population size, assortative mating, selection, migration, etc. By the contrary, the simulation of haploid gametes according to the allelic frequency data is independent of the state of the population with respect to the Hardy-Weinberg equilibrium. Diploid species generally produce haploid gametes (Mable and Otto 1998). Gametes simulation based on the allele frequencies of unlinked loci does not imply any assumption regarding the phase of the alleles in the genotypes, the proportions of homozygous and heterozygous combinations or the existence of linkage disequilibrium, therefore preventing bias in the result.

We developed 'Gametes Simulator', a computer routine to simulate gametes from the allelic frequencies of unlinked loci, which enables the analysis of population structure through Bayesian analysis in Structure software. The objective is to make this routine available to other scientists working on the genetic structure of outbreeding diploid species.

METHODS

Scope and theoretical assumptions

'Gametes Simulator' software simulates haploid gametes produced by diploid organisms, which implies that the gametes are produced due to the gametogenesis process through meiosis I and II. The 1st and 2nd laws of Mendel are the direct descriptions of chromosome movement during meiosis. Furthermore, chromosome movement during meiosis is responsible for allele assortment in the gametes. Consequently, 'Gametes Simulator' software adopts the following theoretical assumptions:

• Mendel´s first law or law of segregation. Each gamete will have only one of the two parental gene copies. The gene copies are randomly allocated to the gametes due to: 1) the random alignment in metaphase I of the homologous chromosomes and its migration towards opposite poles in anaphase I and 2) the orientation and separation of the sister chromatids in anaphase II during the gametogenesis process.

• Mendel´s second law or law of independent assortment. This implies that the alleles of two or n different unlinked genes are assorted independently into gametes. There is no influence of an allele received by a gene on an allele received by another gene and vice versa. The law of independent assortment has its physical basis during meiosis I in the gametogenesis process, when the homologous chromosome pairs line up in random orientations in the middle of the cell during metaphase I. Consequently, the gametes will have different combinations of maternal and paternal homologues, which also determine the alleles carried in each case. This law is applicable in the case of unlinked loci, with the loci in linkage equilibrium, because they are either in different chromosomes or their recombination frequency is 50%, or the map distance is equal to or higher than 50 cM. This element is crucial to the right usage of the software: the selected loci should be known to be unlinked.

In summary, two mechanisms are responsible for the assortment of alleles during gametogenesis: 1) the random alignment of the bivalents in metaphase I, which results in a random combination of maternal and paternal chromosomes allocated to each gamete and therefore the alleles they carry, and 2) the crossing over or recombination during prophase I of meiosis among homologous chromosomes, which leads to the physical assortment of alleles in the genes along a chromosome. The homology between the gametogenesis process and the algorithm followed by 'Gametes Simulator' software is presented in Table 2 as well as the outputs generated per algorithmic step and the biological meaning.

Practical procedures

This Java program uses the matrix of estimated allelic frequencies in populations genotyped through bulk DNA sampling - or other methodologies - as input (Table 1). The matrix contains the allelic frequencies, with the studied populations as columns and each allele as a row. The alleles are identified using the first two columns: 1) the marker name or loci used to genotype the populations and 2) the name of the specific allele. This matrix is an ASCII-type file from Excel. The matrix could be generated by the program FreqsR, specifically designed to determine allelic frequencies in bulk-sampling populations of maize genotyped by microsatellites (Franco et al. 2005), or by any other program able to generate output containing all the information specified in Table 1.

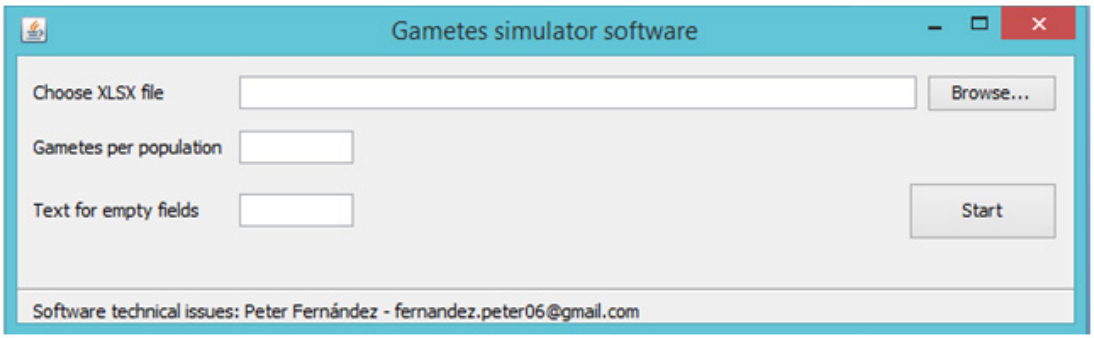
To install 'Gametes Simulator' software, see the instructions in File 1 – 'Instructions to install and use Gametes Simulator Software.docx' at <https://github.com/pfernandezgr/gamete-simulator/>, and read and follow them carefully. Once the program is installed, direct access opens the window shown in (Figure 1A). The input file (the Excel file with the matrix of allelic frequencies) should be browsed and selected. Then, the number of multilocus haploid genotypes (gametes) per population to be simulated must be indicated, filling in the corresponding box 'Gametes per population' (Figure 1A). Finally, this program provides the option to indicate the type or number required for missing data in the output. It is strongly recommended to use the same type or number used for empty cells in the next program in which this output will be used as an input. For example, -9 is used in Structure for empty cells, so it is very useful that the output generated by the 'Gametes Simulator' program has already incorporated the -9 into the empty cells. Once everything is completed, press 'Start'.

After running the program, two Excel files are generated consecutively as output. The first output Excel file allows us to visualize the process by which the gametes within a population are generated per marker according to the allele frequencies within the populations. The information per marker is provided in individual Excel sheets. The first column provides the names of the populations, the second column enumerates the gametes simulated within each population,

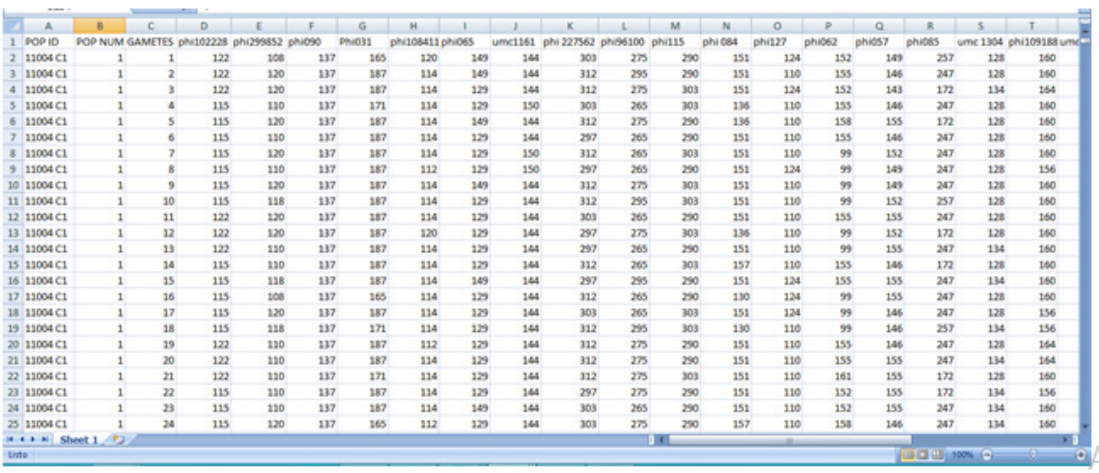
Table 1. Generalized matrix of allelic frequencies found in populations genotyped with molecular markers. Empty cells represent populations not genotyped with the corresponding molecular marker. To see a practical example access to <https://github.com/pfernandezgr/gamete-simulator/> and look at Figure 1 in the file **'**Instructions to install and use Gametes Simulator software.docx**'** or open the file; **'**Example of Input Data Base Gametes Simulator Software.xlxs**'**

Marker name	Allele name	Pop 1	Pop 2	Pop 3	Pop 4	Pop 5	...	Pop n
MK1	MK1a	0.33	0.10	0.40	0.50	0.25	...	Allelic frequency for MK1a in Pop n
MK1	MK1b	0.33	0	0.60	0.30	0.25	...	Allelic frequencyfor MK1b in Pop n
MK1	MK1c	0.33	0.70	0	0.20	0.25	...	Allelic frequency for MK1c in Pop n
...	...	...	...	...	...	...	...	Allelic frequency for ..... in Pop n
MK1	MK1i	0	0.20	0	0	0.25	...	Allelic frequency for MK1i in Pop n
MK2	MK2a	1.0		0.25	0.30		...	Allelic frequency for MK2a in Pop n
MK2	MK2b	0		0.60	0.10		...	Allelic frequency for MK2b in Pop n
MK2	MK2c	0		0.10	0.50		...	Allelic frequency for MK2c in Pop n
MK2	MK2d	0		0.05	0.10		...	Allelic frequency for MK2d in Pop n
MK3	MK3a	0.30	0.10	0.80	0.75	1.0	...	Allelic frequency for MK3a in Pop n
MK3	MK3b	0.70	0.90	0.20	0.25	0	...	Allelic frequency for MK3b in Pop n
MK4	MK4a	0.20	0.30	0.36	0.10	0.50	...	Allelic frequency for MK4a in Pop n
MK4	MK4b	0.60	0	0.14	0.20	0.25	...	Allelic frequency for MK4b in Pop n
MK4	MK4c	0.20	0.70	0.50	0.70	0.25	...	Allelic frequency for MK4c in Pop n
MKn	MKni	...	...	...	...	...	...	Allelic frequency for MKni in Pop n

A



B



POP ID	POP NUM	GAMETES	phi102228	phi299852	phi090	Phi031	phi108411	phi065	umc1161	phi 227562	phi96100	phi115	phi 084	phi127	phi062	phi057	phi085	umc 1304	phi109188
11004 C1	1	1	122	108	137	165	120	149	144	303	275	290	151	124	152	149	257	128	160
11004 C1	1	2	122	120	137	187	114	149	144	312	295	290	151	110	155	146	247	128	160
11004 C1	1	3	122	120	137	187	114	129	144	312	275	303	151	124	152	143	172	134	164
11004 C1	1	4	115	110	137	171	114	129	150	303	265	303	136	110	155	146	247	128	160
11004 C1	1	5	115	120	137	187	114	149	144	312	275	290	136	110	158	155	172	128	160
11004 C1	1	6	115	110	137	187	114	129	144	297	265	290	151	110	155	146	247	128	160
11004 C1	1	7	115	120	137	187	114	129	150	312	265	303	151	110	99	152	247	128	160
11004 C1	1	8	115	110	137	187	112	129	150	297	265	290	151	124	99	149	247	128	156
11004 C1	1	9	115	120	137	187	114	149	144	312	275	303	151	110	99	149	247	128	160
11004 C1	1	10	115	118	137	187	114	129	144	312	295	303	151	110	99	152	257	128	160
11004 C1	1	11	122	120	137	187	114	129	144	303	265	290	151	110	155	155	247	128	160
11004 C1	1	12	122	120	137	187	120	129	144	297	275	303	136	110	99	152	172	128	160
11004 C1	1	13	122	110	137	187	114	129	144	297	265	290	151	110	99	155	247	134	160
11004 C1	1	14	115	110	137	187	114	129	144	312	265	303	157	110	155	146	172	128	160
11004 C1	1	15	115	118	137	187	114	149	144	297	295	290	151	124	155	155	247	134	160
11004 C1	1	16	115	108	137	165	114	129	144	312	265	290	130	124	99	155	247	128	160
11004 C1	1	17	115	120	137	187	114	129	144	303	265	303	151	124	99	146	247	128	156
11004 C1	1	18	115	118	137	171	114	129	144	312	295	303	130	110	99	146	257	134	156
11004 C1	1	19	122	110	137	187	112	129	144	312	275	290	151	110	155	146	247	128	164
11004 C1	1	20	122	110	137	187	114	129	144	312	275	290	151	110	155	155	247	134	164
11004 C1	1	21	122	110	137	171	114	129	144	312	275	303	151	110	161	155	172	128	160
11004 C1	1	22	115	110	137	187	114	129	144	297	275	290	151	110	152	155	172	134	156
11004 C1	1	23	115	110	137	187	114	149	144	303	265	290	151	110	152	155	247	134	160
11004 C1	1	24	115	120	137	165	112	129	144	303	275	290	157	110	158	146	247	134	160

Figure 1. **A**. Main window of 'Gametes Simulator' software. Choose the XLSX file that fulfils the format presented in Table 1. Before running the simulation, the user must provide the number of gametes to be simulated within each population and the text that should be used to fill empty spaces as required for the output. **B**. Final output generated by 'Gametes Simulator'software. In the first column is the name of the population repeated as many times as Gs (gametes simulated per population).The second column corresponds to the ordinal number of the population (rows have the same number if they correspond to gametes simulated for the same population). The third column enumerates the gametes simulated within each population. From the fourth column on, the alleles present in each gamete for each molecular marker are detailed in such a way that each row represents the multilocus genotype of a gamete.

the third column contains a binary code assigned to each type of allele within a population, and in the fourth column, each allele within a population is repeated according to its allelic frequency in the population multiplied by the number of gametes simulated within each population. Finally, in the fifth column, the alleles within each population (shown in the fourth column) are ordered randomly within its corresponding population.

The second and final output Excel file uses all the information from the previous Excel file and contains the information required for the input matrix for Structure v.2.3.4 (Figure 1B). To be used in Structure v.2.3.4, this matrix should be transformed according to the instructions given in the document 'Instructions to install and use Gametes Simulator software' in the paragraphs under the title 'Final step'.

History used for the development of 'Gametes Simulator'

This program was created to analyze the genetic structure of populations (landraces) of maize characterized through bulk DNA fingerprinting. A previously designed program called FtoL (Franco et al. 2007) was developed in the R platform

to solve the same problem. However, FtoL assumes that populations are in Hardy-Weinberg equilibrium, and diploid individuals are simulated using the allelic frequencies in each population. Equilibrium might not be a common situation in natural populations or landraces due to many causes: natural selection, selection made by farmers, genetic drift due to small population size, inbreeding due to assortative mating or not random mating (Ramos et al. 2019), or exogamy with other neighbor populations, especially in outbreeding species (Porta et al. 2018). The simulation of diploid individuals in landraces or natural populations assuming Hardy-Weinberg equilibrium could certainly induce bias. To avoid bias, we simulate haploid gametes using the allelic frequencies in the population. The simulation of gametes allows the analysis of the genetic structure of the population but does not reveal evolutionary processes acting in populations or the reproductive success of these gametes in the next generation. Therefore, simulated gametes represent a snapshot of the genetic composition of a population but cannot be used to predict the genetic frequencies of the next generation unless the population is assumed to be in Hardy-Weinberg equilibrium.

Performance characteristics

The 'Gametes Simulator' program uses as input the matrix of allelic frequencies per population in the format shown in Table 1. The program simulates the alleles within each population in the right proportions according to the allelic frequencies within the populations. The software orders the alleles within the populations at random. Each gamete is represented in a row and is built by the union of the alleles in the consecutive columns. Each column represents a different marker used to characterize the populations. Simulated gametes are haploid, with only one allele from each molecular marker. Tens of gametes per population can be easily simulated, but the simulation of hundreds of gametes could be hampered. The maximum number of gametes per population that can be simulated with 'Gametes Simulator'software will depend on the matrix size (number of markers, alleles, and populations) as well as the hardware characteristics, particularly the available RAM (Random Access Memory).

The accuracy of the simulated gametes has been checked and can be checked by any researcher using this software with the reverse procedure: calculating the allelic frequencies from the alleles simulated per marker within populations (final output of the 'Gametes Simulator' software); the result obtained is the same as the data fed into the program in the input matrix. It must be taken into account that in certain cases a non-integer number of gametes is achieved by

Table 2. Summary of 'Gametes Simulator' algorithm, outputs, their homology with the gametogenesis process and the biological meaning

Algorithm	Output	Homology with gametogenesis	Biological meaning
1. Takes the matrix of allelic frequencies per marker and per population as input.			The fraction of all chromosomes in the diploid population that carry each allele
2. Translates the genetic frequencies of each allele in each population into repeated alleles as many times as Af[#] × Gs*	4th column in each marker Excel sheet in the file 'First output of the software'		Real number of allele copies expected in a sample of gametes taken from each population Size of the sample: Gs*
3. Takes the output from 2 and sorts at random the alleles per marker within each population, allocating one allele randomly to each gamete	5th column in each marker Excel sheet in the file 'First output of the software'	Random allocation of each allele per marker (locus) to each future gamete in the gametogenesis process (1st Mendel's law) and independently among unlinked markers (2nd Mendel's law). Also independent among populations because the allelic frequencies per population reported in the input matrix have no influence on one another	Genetic variability among gametes within populations determined by random and by the allelic frequencies present in each population
4. Multilocus haplotype assemblage (gamete genotype): takes the output from the previous step (5th column of each Excel marker sheet) and assembles it considering all the markers. Gametes Simulator does so by gathering the genetic information for each locus per gamete after the randomization process.	Final matrix that has the list of simulated gametes´ haplotypes in each population as rows and all the markers as columns	Chromatid assignment to each future gamete during anaphase II in meiosis II to give place to the gametes. The alleles present in each chromatid were determined previously by random recombination events (in prophase I) and the random alignment and splitting up of homologues (in metaphase I; anaphase I) and sister chromatids (in metaphase II; anaphase II) in meiosis I and II, respectively.	Multilocus haplotype of gametes belonging to a certain population

[#]Af: allelic frequencies reported in the input matrix; *Gs: number of gametes to be simulated per population. Haplotype assemblage: assignment of individual alleles (following Mendel's 1st and 2nd laws using the alleles present in the population) to sets that correspond to each of the different simulated gametes.

multiplying an allelic frequency (Af) by the number of gametes to simulate (Gs), (Af.Gs); if the decimal part of the result is equal to or higher than 0.5 alleles, the software will round the value up to the next integer number, and if it is less than 0.5, it will return the integer number without the decimal part. For example, if Af.Gs ≥ 0.5 but < 1.5, the program will simulate 1 allele, and if Af.Gs is ≥ 1.5 and < than 2.5, it will simulate 2 alleles due to the evident fact that it is not possible to simulate decimal parts of an allele or any decimal part of a gamete (and it is also impossible for this to happen in nature). To downsize the errors due to the use of a small sample size (Gs), we strongly recommend simulate a number of gametes that make possible that the lowest allelic frequency in the input matrix then will be represented at least with one allele in the simulation. To that end, the researcher should take into account that the allele with the lowest frequency in the input matrix will be represented in the population's sample of gametes at least once only if the sample size (Gs: number of gametes to be simulated) is as large as the quotient (q) that represents the inverse of the lowest allelic frequency ($_{lowest}$Af) in the input (q = 1/$_{lowest}$Af). To avoid inaccuracies due to the sample size, a 2q number of gametes can be simulated.

Applications

This program is applicable to analyze genetic diversity in outbreeding species (animal and allogamous plant species) with populations constituted of diploid individuals. It enables the posterior analysis of the genetic structure of populations with a Bayesian approach using the Structure program. The resultant output is the simulation of multilocus gametes genotypes from allelic frequencies of unlinked loci in populations characterized through bulk sampling or other methodologies. This output acts as input for Structure (Pritchard et al. 2000, Falush et al. 2003).

Forms of access

The software was developed using Java® and can be downloaded from

<https://github.com/pfernandezgr/gamete-simulator/>

ACKNOWLEDGEMENTS

The first author is grateful for a research grant from Comisión Sectorial de Investigación Científica (CSIC) and a fellowship from Comisión Académica de Posgrados (CAP), Universidad de la República, Uruguay. B. Porta also thanks Clara Pritsch, Marianella Quezada and Magdalena Vaio for inspirational talks about theoretical concepts regarding the gametogenesis process.

REFERENCES

Bedoya CA, Dreisigacker S, Hearne S, Franco J, Mir C, Prasanna BM, Taba S, Charcosset A and Warburton ML (2017) Genetic diversity and population structure of native maize populations in Latin America and the Caribbean. **PLoS ONE 12**: e0173488.

Contina A, Alcantara JL, Bridge ES, Ross JD, Oakley WF, Kelly JF and Ruegg KC (2019) Genetic structure of the Painted Bunting and its implications for conservation of migratory populations. **Ibis 161**: 372-386.

Crystian D, Santos JMD, Barbosa GVDS and Almeida C (2018) Genetic diversity trends in sugarcane germplasm: Analysis in the germplasm bank of the RB varieties. **Crop Breeding and Applied Biotechnology 18**: 426-431.

Dempewolf H, Baute G, Anderson J, Kilian B, Smith C and Guarino L (2017) Past and future use of wild relatives in crop breeding. **Crop Science 57**: 1070-1082.

Dubreuil P, Warburton ML, Chastanet M, Hoisington D and Charcosset A (2006) More on the introduction of temperate maize into Europe: large-scale bulk SSR genotyping and new historical elements. **Maydica 51**: 281-291

Falush D, Stephens M and Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. **Genetics 164**: 1567-1587

Ferreira F, Scapim CA, Maldonado C, and Mora F (2018) SSR-based genetic analysis of sweet corn inbred lines using artificial neural networks. **Crop Breeding and Applied Biotechnology 18(3)**: 309-313.

Franco J, Warburton M, Dubreuil P and Dreisigacker S (2005) **User's manual for the FREQS-R Program for estimating allele frequencies for fingerprinting and genetic diversity studies using bulked heterogeneous populations**. CIMMYT, Mexico, 6p.

Franco J, Warburton M and Dreisigacker S (2007) **User's manual for the FtoL-R Program for generating a "dummy" data set consisting of allele lengths for hypothetical individuals**. CIMMYT, Mexico, 9p.

Guimarães RA, Miranda KMC, Mota EES, Chaves LJ, Telles MPDC 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.

Hale M L, Burg T M and Steeves TE (2012) Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. **PloS ONE 7**: e45170.

Mable BK and Otto SP (1998) The evolution of life cycles with haploid and diploid phases. **Bio Essays 20**: 453-462.

McCouch S (2004) Diversifying Selection in Plant Breeding. **PLoS Biology 2**: e347.

Mutegi E, Snow AA, Rajkumar M, Pasquet R, Ponniah H, Daunay MC and Davidar P (2015) Genetic diversity and population structure of wild/weedy eggplant (Solanum insanum, Solanaceae) in southern India: Implications for conservation. **American Journal of Botany 102**: 140-148.

Porta B, Condón F, Franco J, Iriarte W, Bonnecarrère V, Guimaraens-Moreira M, Vidal R and Galván GA (2018) Genetic structure, core collection, and regeneration quality in white dent corn landraces. **Crop Science 58**: 1644-1658.

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

Ramos SLF, Lopes MTG, Lopes R, Dequigiovanni G, Macêdo JLVD, Sebbenn AM, da Silva EB and Garcia JN (2019) Mating system analysis of Açaí-do-Amazonas (*Euterpe precatoria* Mart.) using molecular markers. **Crop Breeding and Applied Biotechnology 19(1)**: 126-130.

Reif JC, Hamrit S, Heckenberger M, Schipprack W, Peter Maurer H, Bohn M and Melchinger AE (2005) Genetic structure and diversity of European flint maize populations determined with SSR analysis of individual and bulks. **Theoretical and Applied Genetics 111**: 906-913.

Thornhill NW (1993) **The natural history of inbreeding and outbreeding: theoretical and empirical perspectives.** University of Chicago Press, Chicago, 575p.

Waples RS (2014) Testing for Hardy–Weinberg proportions: have we lost the plot? **Journal of Heredity 106**: 1-19.