Genetic characterisation of *Annona crassiflora* as basis for conservation and breeding

Maria Fernanda Maia Ferreira[1], Adelson Lemes da Silva Júnior[1] (iD),

Sérgio Bruno Fernandes[1] (iD), Lucimara Cruz de Souza[2] (iD),

Lucas Amaral de Melo[1] (iD) and Dulcinéia de Carvalho[1]

[1]Departamento de Ciências Florestais, Escola de Ciências Agrárias de Lavras, Universidade Federal de Lavras, Lavras, Minas Gerais, Brazil and [2]Departamento de Biologia, Setor de Genética, Universidade Federal de Lavras, Lavras, Minas Gerais, Brazil

Abstract

Annona crassiflora is a fruit-bearing tree species native to the Cerrado that has ecological and economic potential, mainly due to the production of attractive and useful fruits for a number of species, including humans. To provide a basis for its conservation and breeding, the objective was to assess the diversity and genetic structure of natural populations of the species using inter-simple sequence repeat (ISSR) molecular markers. Eight populations were analysed in the state of Minas Gerais, Brazil, with 24 individuals randomly sampled in each population, reaching a total of 192 trees. Ten ISSR primers were used, resulting in high genetic diversity for the combined data ($H^* = 0.35$ and $I^* = 0.52$). However, a wide range of representative values was observed in the individual assessment of populations, with JAN, GM and MC standing out for their low genetic diversity, resulting in H^* of 0.19, 0.23 and 0.24 and I^* of 0.28, 0.34 and 0.35, respectively. An analysis of molecular variance showed greater variation within populations, indicating gene flow (N_m), but genetic differentiation between populations was moderate. Bayesian analysis, although resulting in four genetic groups, revealed the presence of a majority group for the GM and JAN populations. We propose measures to maintain these populations, such as raising awareness of local extractivism and planting genetically divergent seedlings. Furthermore, we recommend including of all populations in conservation and breeding programmes, aiming to cover the maximum genetic variation for the species.

Introduction

The Cerrado biome covers 22% of the Brazilian territory, is characterised by its great biodiversity of species and recognised as the richest savannah in the world. However, of all the hotspots in the world, it has the lowest percentage of fully protected areas (8.21%), mainly due to agribusiness pressure and the clearing of production areas (Ministério do Meio Ambiente, 2022).

This predatory exploitation has created a scenario of vulnerability for forest species, in addition to the loss of products such as fruits, resins and oils, among others, which could be used in food production, pharmacological products and cosmetics through sustainable management (Strassburg *et al.*, 2017). One species native to the Cerrado that has potential for sustainable management is *Annona crassiflora* Mart., popularly known as marolo or araticum (Cohen *et al.*, 2010).

A. crassiflora is a fruit-bearing tree species of the Annonaceae family with a wide distribution in Brazilian regions. It has hermaphrodite flowers, an allogamous reproductive system favoured by insects (mainly beetles) and zoochoric seed dispersal (Cavalcante *et al.*, 2009). The species provides ecological services and has economic potential, mainly through its fruits. They serve as a food source for animal species, mainly birds. They also have good sensory acceptance for human consumption, and high nutritional quality, containing bioactive compounds (Arruda *et al.*, 2016). The fruits can be eaten fresh or processed, and are a source of food for communities in Cerrado areas, as well as being sold in local markets (Pereira *et al.*, 2022). In addition, its wood contains alkaloids such as atherospermidine and liriodenine, which have antimicrobial activities against human hepatoma cells (Inoue *et al.*, 2010).

Despite the potential of the species, the exploitation of native Cerrado areas and the fragmentation of this biome (MMA, 2022) have resulted in the spatial isolation of forest remnants, which may lead to a decrease in the number of *A. crassiflora* individuals. Spatial isolation hinders or prevents migration between populations, reducing gene flow and increasing differentiation between them (Santos and Oliveira, 2020). In addition, reduction in population size favours genetic drift, which randomly alters allele frequencies and promotes allelic fixation and loss. These factors reduce genetic diversity and alter the genetic structure of populations,

making them more vulnerable to adverse events such as pests, diseases and climate change (Ellegren and Galtier, 2016).

In this context, molecular markers can detect genetic variation, assisting decision-making related to the conservation and management of natural populations (Filippos, 2016). Different types of molecular markers are available, but inter-simple sequence repeats (ISSRs), characterised as dominant, universal and multilocus, are recognised for covering a high level of polymorphism distributed throughout the genome (Turchetto-Zolet *et al.*, 2017). Furthermore, ISSRs have been successfully used to determine the diversity and genetic structure of the genus *Annona* (Gwinner *et al.*, 2016; Samaradiwakara *et al.*, 2020; Sá *et al.*, 2022).

Therefore, the aim of this study was to assess the diversity and genetic structure of natural populations of the *A. crassiflora* species using ISSR molecular markers to provide information for use in conservation and genetic breeding programmes for the species.

Materials and methods

Sampling and collection sites

The sampling sites were selected in the state of Minas Gerais, covering the areas where *A. crassiflora* grows. To this end, eight populations were identified, varying between ecological reserves and pastures, both with an area of less than 50 ha, where 24 individuals of the species were randomly sampled per population, for a total of 192 adult trees (Table 1). Variations in climatic conditions were observed depending on the location of the collection area, with mean annual temperatures ranging from 18.2 (Carmo da Cachoeira) to 22.7°C (Januária), and annual precipitation from 972 (Grão Mongol) to 1582 mm (Carmo da Cachoeira) (Alvares *et al.*, 2013).

DNA extraction

Leaf tissue samples from each selected individuals were used to extract genomic DNA using the method of Moog and Bond (2003). The concentration and purity of the extracted DNA was then assessed by spectrophotometry in a NanoDrop 2000C (Thermo Scientific), using the A_{260}/A_{280} ratio ($1.80 \leq A_{260}/A_{280} \geq 2.00$) as a quality parameter (Aguilar *et al.*, 2016).

ISSR genotyping

Polymerase chain reactions (PCRs) were performed using DNA aliquots from individuals at a final concentration of 10 ng/μl and 10 ISSR primers (A16; JOHN; MANNY; and UBCs 807; 834; 835; 840; 841; 855 and 857) developed by the University of British Columbia, Vancouver, Canada. The total reaction volume was 12 μl, containing 2 μl of genomic DNA, 1.2 μl of 10× PCR buffer (500 mM Tris-HCl pH 8.0, 200 mM KCl, 2.5 mg/ml BSA, 200 mM tartrazine and 1% Ficol), 1.2 μl of dNTP + MgCl$_2$ (2.5 mM dNTP, 25 mM MgCl$_2$), 0.15 μl of Taq DNA polymerase (5 U/μl) and 2 μl of each primer (2 mM).

Amplifications were performed in a thermocycler (GeneAmp PCR System 9700) according to protocols previously used for other forest species (Silva Júnior *et al.*, 2020; Vieira *et al.*, 2022), including an initial denaturation step (5 min at 94°C), followed by 37 cycles of denaturation (15 s at 94°C), annealing (30 s at 47°C) and extension (1 min at 72°C). At the end of the last cycle, a final extension was performed (7 min at 72°C).

The amplified fragments were separated by electrophoresis on a 1.5% agarose gel with 1× TBE buffer (10.8 g/l Tris base; 5.5 g/l boric acid; 0.83 g/l EDTA) at 100 V for 4 h. The gels were stained with ethidium bromide (5 mg/ml) and photographed under ultraviolet light in a photodocumenter (UVP DigiDoc-It System) linked to UVP Doc-Itls image analysis software. The size of the fragments was estimated by comparison with the molecular weight marker 100 pb ladder (Ludwig Biotec).

Data analysis

As ISSR markers are dominant markers, a binary matrix was generated where the presence of fragments is scored as 1, i.e. the locus is considered homozygous dominant or heterozygous, and the absence as 0, indicating homozygous recessive (no amplification). The binary matrix was then used to perform a bootstrap analysis, which assesses the optimal number of fragments required to perform the study. For each locus, the number of fragments obtained was counted and the total number of fragments (N_F) was also determined. In addition, the informativeness of the ISSR markers was assessed through the polymorphic information content (PIC), which was carried out for each primer and for the entire set of primers. The Genes program (Cruz, 2016) was used for the optimal number of fragments and PIC analyses.

Genetic dissimilarity (DG) between pairs of individuals was determined using the Jaccard coefficient. For populations and combined data, pairs of individuals with minimum (IDG$_{min}$) and maximum (IDG$_{max}$) genetic dissimilarities were identified and the corresponding values were found (DG$_{min-max}$). The DG matrix between individuals was used to construct a dendrogram, obtained by the unweighted grouping of means (UPGMA) method, with a cut-off point proposed by Mojena (1977), establishing the coefficient $k = 1.25$. The consistency between the dissimilarity matrix and the clusters presented in the dendrogram was evaluated by the cophenetic correlation coefficient (CCC). The Genes program (Cruz, 2016) was used for these analyses. R software (R Core Team, 2016) was used to generate the circular dendrogram, using the packages vegan (Oksanen *et al.*, 2018), cluster (Maechler *et al.*, 2019), dendextend (Galili *et al.*, 2020), extrafactor (Kassambara and Mundt, 2017), ggpubr (Kassambara, 2020), cowplot (Wilke, 2019) and gridExtra (Auguie and Antonov, 2017).

Levels of genetic diversity were measured for populations and combined data using Popgene 3.2 software (Yeh and Boyle, 1997). The parameters calculated were the number of observed alleles (A_O), number of effective alleles (A_E), Nei diversity index (H^*) (Nei, 1978) and Shannon index (I^*) (Shannon and Weaver, 1949).

Genetic structuring of the populations and combined data was assessed by the analysis of molecular variance (AMOVA) with two hierarchical levels (Excoffier *et al.*, 1992). In addition, considering populations pairwise, gene flow values (N_m) were determined using Arlequin 3.5 software (Excoffier and Lischer, 2010) and geographic distance in kilometres by GPS TrackMaker software (Trackmaker, 2022), which were evaluated for correlation by the Mantel test with 1000 permutations (Mantel, 1967) using the Genes program (Cruz, 2016). Finally, a Bayesian approach was used to determine the number of genetic groups (K) using Structure 2.3 software (Falush *et al.*, 2007). The number of groups (K) was set to vary between 1 and 11, with 20 samples taken for each K value and 10,000 Monte Carlo interactions via Markov chains. Data were exported to Structure Harvester software (Earl and Vonholdt, 2012), and the best value of K was indicated by the ad hoc ΔK method (Evanno *et al.*, 2005).

Table 1. Sampled populations of *A. crassiflora* in the state of Minas Gerais, including average annual temperatures, annual precipitation and conservation status of the area and intensity of extractivism of the species

Population/ location	Code	Geog. coord.	Temp. (°C)	Prec. (mm/year)	Description
Curvelo	CVA	18°88′45″ 44°44′64″	21.0	1.374	Fragment of Cerrado limited by roads, pastures and *Eucalyptus* plantations.
Curvelo	CVB	18°82′34″ 44°52′73″	21.0	1.374	Pasture area with the presence of the species *A. crassiflora*, including natural regenerants.
Curvelo	CVC	18°77′03″ 44°49′16″	21.0	1.374	Pasture area with the presence of the species *A. crassiflora*, including natural regenerants.
Morro da Garça	MG	18°50′18″ 44°57′93″	21.2	1.366	Pasture area with the presence of the species *A. crassiflora*, with traces of clearings.
Carmo da Cachoeira	CC	21°44′92″ 45°08′99″	18.2	1.582	Fragment of Cerrado *sensu stricto*, occurring in a rural property, modified in a silvopastoral system.
Montes Claros	MC	16°71′67″ 43°86′66″	21.2	1096	Fragment of Cerrado *sensu stricto*, belonging to a biological reserve.
Grão Mongol	GM	16°56′66″ 42°89′98″	20.6	972	Agroextractive reserve called Projeto de assentamento Americana, impacted by selective cutting and fires until the 1980s. It is currently in the process of regeneration and *A. crassiflora* fruits are collected by residents of the region.
Januária	JAN	15°48′33″ 44°36′ 65″	22.7	1.056	Área de Proteção Ambiental do Rio Pandeiros, with areas of pastures in the surroundings and eventual penetration of cattle along the edges. *A. crassiflora* fruits are also extracted for the production and commercialisation of food products by cooperatives in the region.

Results

Efficiency of ISSR markers

The analysis of the optimal number of fragments resulted in a power of 0.049 and a correlation of 0.948, where 41 polymorphic fragments would be sufficient to assess the diversity and genetic structure between and within populations of the *A. crassiflora* species. Furthermore, the efficiency of the ISSR markers was also evaluated by the PIC, with values ranging from 0.34 for the John primer to 0.47 for the A16 primer, with an average of 0.39. The number of fragments per primer was also determined, ranging from 3 to 8, with a total of 61 fragments (Table 2).

Table 2. Descriptive analysis of ISSR primers selected for evaluation in *A. crassiflora*

Primers	Sequence (5′–3′)[a]	N_F	PIC
A16	CACACACACACAR	3	0.47
JOHN	AGAGAGAGAGAGAGYC	6	0.38
MANNY	CACCACCACCACRC	7	0.34
UBC 807	AGAGAGAGAGAGAGAGT	8	0.39
UBC 834	AGAGAGAGAGAGAGAGYT	7	0.37
UBC 835	AGAGAGAGAGAGAGAGYC	8	0.44
UBC 840	GAGAGAGAGAGAGAGAYT	6	0.38
UBC 841	GAGAGAGAGAGAGAGAYC	5	0.46
UBC 855	ACACACACACACACYT	3	0.38
UBC 857	ACACACACACACACYG	8	0.42
Total	–	61	0.39

N_F, total number of fragments; PIC, polymorphic information content.
[a]H = A, T or C; R = A or G; V = A, C or G and Y = C or T.

Dissimilarity and genetic diversity

The DG calculated between individuals within populations showed equality in the genetic makeup of some pairs of individuals ($IDG_{min} = 0.00$) occurring in the Morro da Garça (MG), Carmo da Cachoeira (CC), Grão Mongol (GM) and Januária (JAN) populations, and the maximum value ($IDG_{max} = 0.55$) was observed between the pairs of individuals CVB3 × CVB23 and MG10 × MG23, located in the Curvelo B (CVB) and MG populations, respectively. For the combined data, the maximum genetic dissimilarity ($IDG_{max} = 0.63$) was observed between the pair of individuals CVB4 × CC21, located in the CVB and CC populations, respectively (Table 3).

The cluster analysis of the DG between individuals, obtained by the UPGMA method, resulted in 11 groups, with a cut-off point of 0.317. The largest group, called G1, consisted of 70 individuals, 18 from the GM population, 17 from the JAN, 10 from the Montes Claros (MC), 8 from the MG, 5 from the Curvelo C (CVC), 5 from the CC, 4 from the CVB and 3 from the Curvelo A (CVA). Other large groups were G2, G3, G4 and G5 with 40, 16, 30 and 19 individuals. The two smaller groups consisted of only one individual each, both identified in the MG population. The CCC was 87% (Fig. 1).

When the populations were evaluated individually, the highest values for the number of observed (A_O) and effective (A_E) alleles were found for the CVB population, while the lowest values were found for the JAN population The genetic diversity index of Nei (H^*) and Shannon (I^*) determined for the individual populations resulted in higher values of genetic diversity for the CVB population, which also indicated high levels contained in the CVA, MG and CC populations. For the combined data, there was an increase in the number of observed alleles ($A_O = 2.00$); however, the number of effective alleles ($A_E = 1.61$) remained lower than the resulting alleles in the CVB and MG populations. The genetic diversity parameters, on the other hand, were higher when calculated for the combined data, resulting in high genetic diversity, as indicated

Table 3. Parameters of dissimilarity and genetic diversity estimated for populations and combined data of the species *A. crassiflora* using ISSR markers

Pop. (code)	No. of Ind.	IDG$_{min}$	IDG$_{max}$	DG$_{min-max}$	A_O	A_E	H^*	I^*
CVA	CVA1–CVA24	CVA6–CVA9	CVA11–CVA17	0.03–0.43	1.72	1.55	0.31	0.44
CVB	CVB1–CVB24	CVB1–CVB8	CVB3–CVB23	0.04–0.55	1.82	1.65	0.35	0.51
CVC	CVC1–CVC24	CVC9–CVC16	CVC4–CVC6	0.01–0.46	1.65	1.52	0.28	0.41
MG	MG1–MG24	MG13–MG14	MG10–MG23	0.00–0.55	1.80	1.64	0.35	0.50
CC	CC1–CC24	CC5–CC7	CC3–CC20	0.00–0.50	1.72	1.56	0.30	0.44
MC	MC1–MC24	MC9–MC16	MC4–MC22	0.01–0.42	1.60	1.42	0.24	0.35
GM	GM1–GM24	GM17–GM23	GM11–GM21	0.00–0.39	1.60	1.40	0.23	0.34
JAN	JAN1–JAN24	JAN8–JAN9	JAN3–JAN4	0.00–0.31	1.50	1.32	0.19	0.28
Comb. data	CVA1–JAN24	–	CVB4–CC21	0.00–0.63	2.00	1.61	0.35	0.52

IDG$_{min}$, pair of individuals with minimal genetic dissimilarity; IDG$_{max}$, pair of individuals with maximum genetic dissimilarity; DG$_{min-max}$, minimum and maximum genetic dissimilarity; A_O, number of alleles observed; A_E, number of effective alleles; H^*: Nei genetic diversity index; I^*: Shannon index.

by the values of H^* and I^*, which were 0.35 and 0.52, respectively (Table 3).

Genetic structure

The AMOVA gave a global estimate of Φ_{ST} equal to 0.1439, i.e. 14.39% of the total genetic variation was found between populations. Therefore, most of the genetic variation (85.61%) is within populations. The highest value determined for the average number of migrants ($N_m = 7.03$) between paired populations was observed between the CVC and MG populations, which are geographically separated by approximately 31.11 km, indicating gene flow between them. The lowest value ($N_m = 2.15$) was between the CVC and JAN populations, separated by 364.03 km (Table 4). The value of N_m for the combined data was 2.03.

The Mantel test revealed a negative (−0.5576) and significant correlation at 1% probability between the values of gene flow (N_m) and geographic distances (km), indicating that the smaller the geographic distance, the greater the gene flow between populations.

The Bayesian approach resulted in four genetic groups ($k = 4$) (Fig. 2(a)). The genetic groupings represented by graphs for each population show that they all have the four genetic groups. However, for the GM and JAN populations there was a predominance of one genetic group indicated in dark blue (Fig. 2(b)).

Discussion

Genetic dissimilarity

Most primers evaluated individually were moderately informative (PIC values between 0.25 and 0.45 according to Tatikonda et al., 2009). Furthermore, when evaluated together (Table 2), they were also classified as moderate and were therefore sufficient for genetic characterisation in *A. crassiflora* populations. However, it is important to note that although ISSR markers are useful for analysing genetic diversity, they have some limitations. These include their dominant nature, which prevents homozygous and heterozygous individuals from being distinguished, and their sensitivity to amplification conditions, which can lead to difficulties in reproducing results (Turchetto-Zolet et al., 2017). These limitations must be taken into account when interpreting the data, especially in studies requiring high genetic precision.

Analysis of DG between individuals of *A. crassiflora* revealed pairs with values equal to 0 (Table 2), suggesting possible self-fertilisation and/or asexual reproduction. The species is characterised as allogamous (Cavalcante et al., 2009), with self-fertilisation occurring at very low rates (<5%) in the absence of self-incompatibility. According to Kiill and Costa (2003), *A. crassiflora* is self-incompatible; however, it is dichogamous and the seeds resulting from geitonogamy have lower viability, which does not make self-fertilisation impossible. These results may indicate the possibility of asexual reproduction, as suggested by Pimenta et al. (2017) and Souza et al. (2020), since DG values equal to 0 represent clones.

Despite genetically similar individuals, there was high genetic variability, confirmed by the maximum genetic dissimilarity between individuals (IDG$_{max}$) within populations (Table 2) and by the formation of 11 groups in the UPGMA cluster analysis (Fig. 1). These groups mix individuals from different populations, indicating a sharing of alleles between them.

Genetic diversity

The higher values for observed alleles (A_O) and effective alleles (A_E) in the CVB and MG populations indicate better allele coverage and distribution compared to the other populations. The increase in these values for the combined data indicates the occurrence of private alleles between the populations. The A_O value in the combined data is similar to that observed in diploid species, with an average of 1.61 for A_E, indicating a good distribution of alleles (Silva Júnior et al., 2020).

For the H^* and I^* indices, which classify the level of genetic diversity, the values found for the combined data were higher than those observed in previous studies with the same species (Gwinner et al., 2016), genus (Samaradiwakara et al., 2020; Sá et al., 2022), family or species with similar characteristics (Vieira et al., 2022). In addition, the values of the Shannon index (I^*) vary between 0 and 1, with values close to 1 indicating high genetic diversity (Lewontin, 1972). Therefore, there is high genetic diversity in CVB and MG populations, moderate in CVA, CVC and CC, and low in MC, GM and JAN. The high diversity in CVB and MG may be related to the smaller geographical distance and gene flow between them, according to the Mantel test. In contrast, the JAN population showed the lowest levels of diversity ($H^* = 0.19$ and
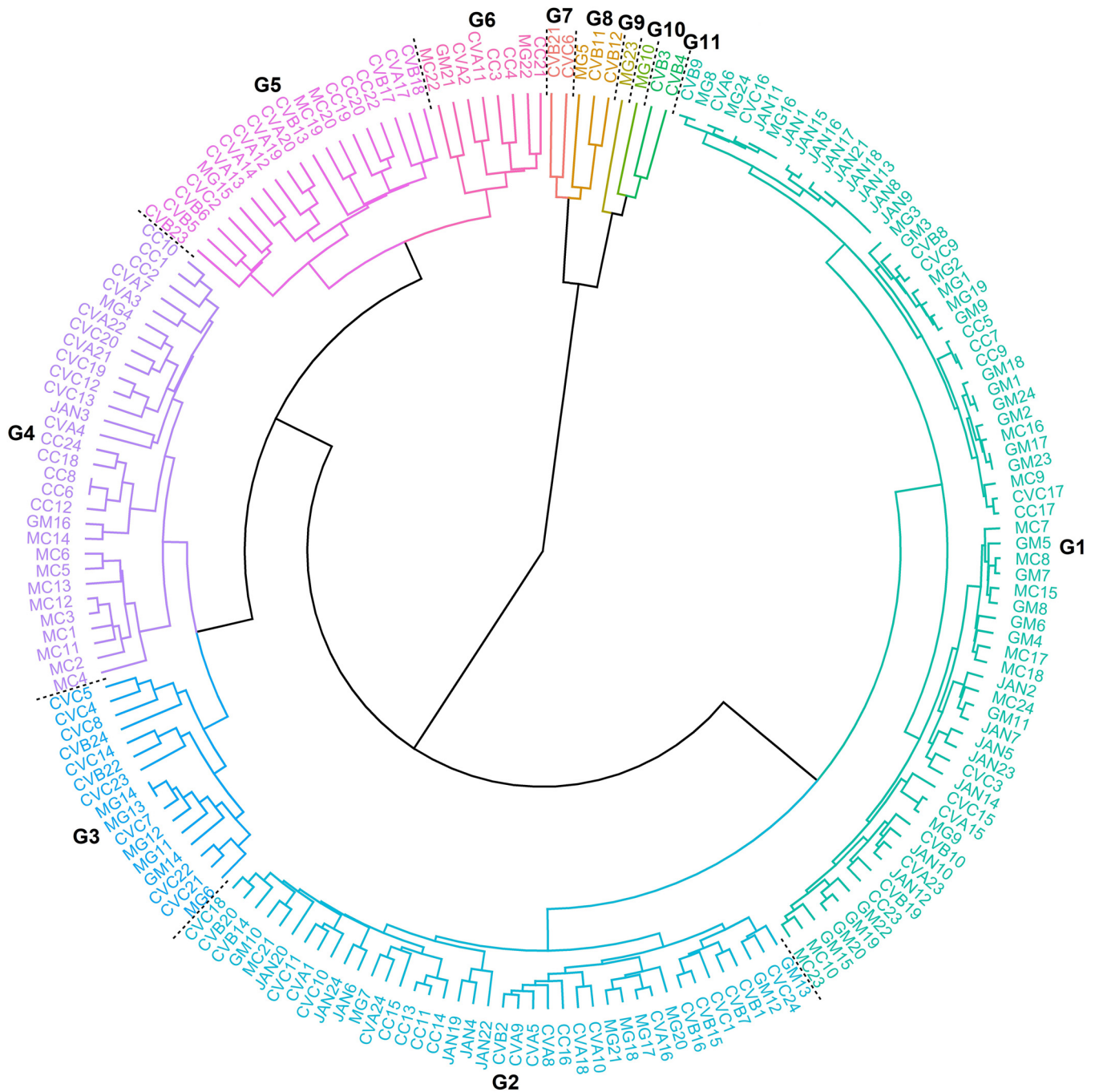
Figure 1. Cluster analysis among 192 individuals of the species *A. crassiflora* evaluated by ISSR markers, obtained by the UPGMA method.

$I^* = 0.28$), possibly due to geographical isolation, anthropogenic influences (Table 1) and genetic drift.

For the combined data, the H^* and I^* indices indicate high genetic diversity for the species, with values higher than those found in previous studies on *A. crassiflora* ($H^* = 0.17$ and $I^* = 0.28$; Gwinner *et al.*, 2016) and the genus *Annona* ($H^* = 0.22$ and $I^* = 0.32$; Sá *et al.*, 2022). Differences in genetic diversity indices between studies can be attributed to the number of individuals sampled and the conservation status of the populations.

The high and moderate genetic diversity in some populations indicates their ability to be maintained over generations in the face of disturbance (Ellegren and Galtier, 2016). All the populations evaluated can be used in breeding programmes to establish a base population, due to the high genetic diversity indices of the combined data. However, in order to conserve the populations with the lowest genetic diversity, it is essential to raise awareness among the local communities, especially where fruit is harvested, and to plant seedlings with genetically divergent material, including from the populations sampled in this study.

Genetic structure

The AMOVA confirms a greater genetic variation within the *A. crassiflora* populations (Table 4), suggesting a sharing of alleles

Table 4. AMOVA between and within populations and gene flow (N_m) estimated for populations pair by pair as a function of approximate geographic distance (km)

Source of variation	Degrees of freedom	Sum of squares	Variance components	Variation (%)
Between populations	7	136.02	0.64	14.39
Within populations	184	710.08	3.85	85.61
Total	191	846.11	4.49	
$\Phi_{ST} = 0.1439$				
Number of migrants (N_m)/geographic distance (km)				

	CVA	CVB	CVC	MG	CC	MC	GM	JAN
CVA	–	5.07	4.62	5.21	4.73	3.03	3.14	2.93
CVB	10.91	–	5.49	5.56	4.35	2.21	3.09	2.33
CVC	13.92	6.92	–	7.03	3.52	2.32	3.13	2.15
MG	44.97	36.24	31.11	–	5	2.75	3.37	2.71
CC	291.84	296.65	303.12	330.87	–	3.77	4.13	2.97
MC	247.71	243.51	236.84	211.42	539.43	–	5.29	3.31
GM	304.53	303.64	296.78	278.76	587.63	104.52	–	4.9
JAN	376.51	370.12	364.03	334.85	664.79	146.61	197.54	–

N_m values are shown on the upper diagonal, and values of geographic distances between populations are shown on the lower diagonal.
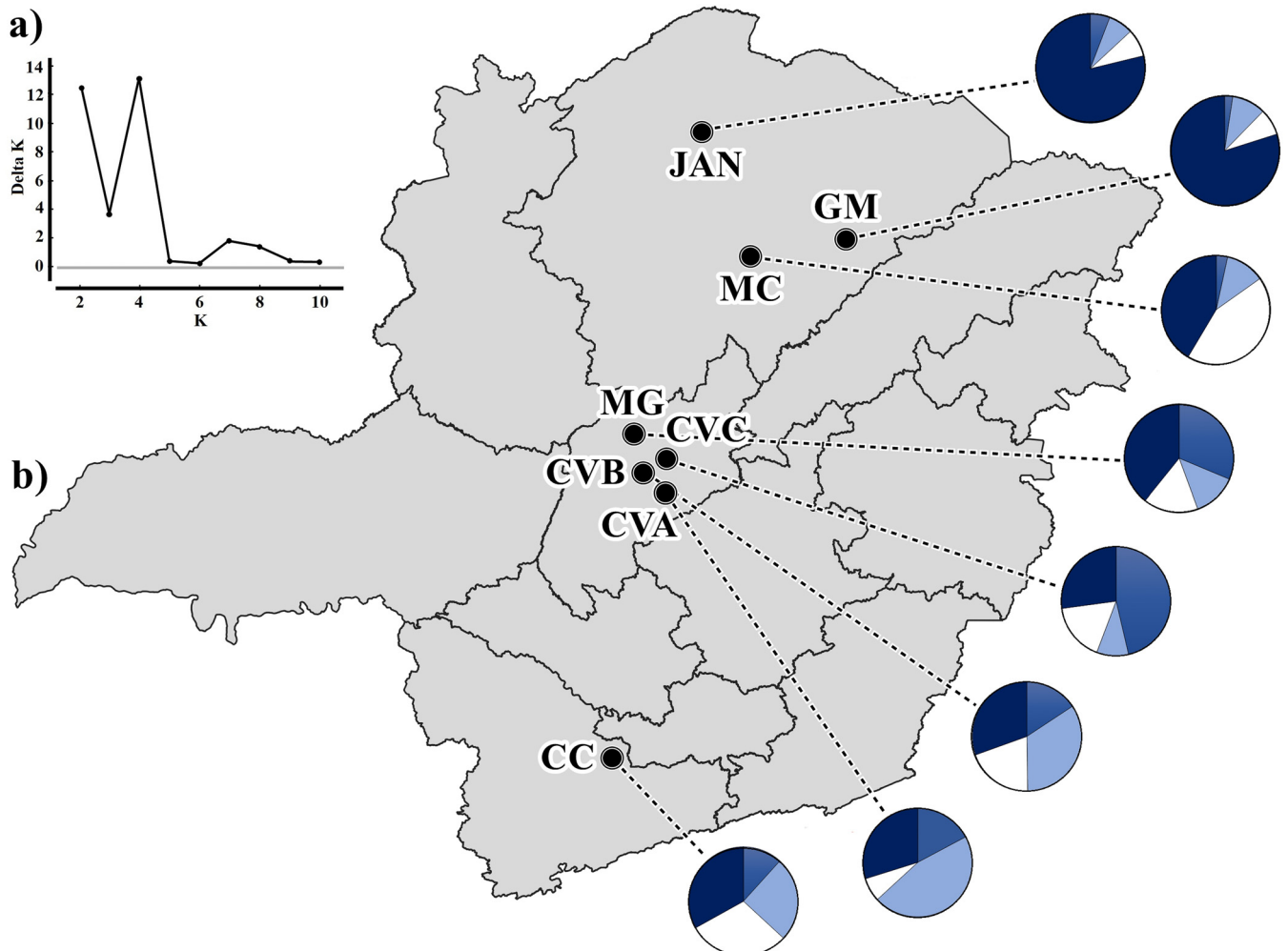

Figure 2. Bayesian approach involving the total sampling (192 individuals) of *A. crassiflora*. (a) Graph of Δ*K* values for each *K* value. Best *K* (4), following the criterion proposed by Evanno *et al*. (2005). (b) Graphs showing the clusters in relation to the populations of *A. crassiflora* identified in the state of Minas Gerais, Brazil.

between them. However, the Φ_{ST} value indicates moderate genetic differentiation (Wright, 1978), possibly influenced by populations with less genetic diversity and the presence of private alleles.

The N_m values indicate significant gene flow between populations (Table 4), with values greater than 1 as suggested by Wright (1951), although this reflects past gene flow. A negative correlation between genetic and geographic distance was also observed, which was confirmed by the Mantel test, showing that geographically closer populations have greater gene flow and are more genetically similar than would be expected by chance (Santo-Silva et al., 2016). This pattern is consistent with the characteristics of *A. crassiflora*, an allogamous species with predominantly beetle pollination and seed dispersal by zoochory (Cavalcante et al., 2009). Consequently, as the spatial distance between populations increases, pollen and seed dispersal by these agents becomes more restricted.

The Bayesian analysis confirmed the previous analyses and identified four groups in all populations (Fig. 2(a)), confirming gene flow. Most of these have a mixture of genetic groups, while GM and JAN have a majority group (Fig. 2(b)). For these populations, factors such as allelic loss and fixation and the consequent increase in homozygosity may reduce adaptive value through the expression of deleterious alleles, especially in the absence of sustainable management and conservation efforts (Ellegren and Galtier, 2016).

Conclusion

There is high genetic diversity for the combined data from all the sampled populations of *A. crassiflora* in the state of Minas Gerais, Brazil, but levels of diversity in the individual populations vary from high to low, possibly due to geographical isolation, anthropogenic disturbance and evolutionary factors. The greatest genetic variation has occurred within the populations due to gene flow between them, with the presence of four genetic groups. However, there is moderate genetic differentiation, with the dominance of one genetic group, particularly in the GM and JAN populations. It is recommended to raise the awareness of the local communities, especially in the areas where the fruit is harvested. In addition, knowledge of the genetic variability of *A. crassiflora* can be used to collect seeds and produce seedlings from genetically divergent mother trees, thus improving the maintenance of future populations in the face of natural selection, genetic drift and inbreeding. Finally, the establishment of germplasm banks, including permanent collections of pollen, seeds, tissue cultures and seed orchards, would benefit both conservation and breeding programmes.

References

Aguilar LG, López AMS, Aceituno CB, Ávila JAC, Guerreiro JAL and Quesada RA (2016) DNA source selection for downstream applications based on DNA quality indicators analysis. *Biopreservation and Biobanking* **14**, 264–270.

Alvares CA, Stape JL, Sentelhas PC, Gonçalves JLM and Sparovek G (2013) Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift* **22**, 711–728.

Arruda HS, Botrel DA, Fernandes RVB and Almeida MEF (2016) Development and sensory evaluation of products containing the Brazilian savannah fruits araticum (*Annona crassiflora* Mart.) and cagaita (*Eugenia dysenterica* Mart.). *Brazilian Journal of Food Technology* **19**, 1–7.

Auguie B and Antonov A (2017) Miscellaneous functions for 'Grid' graphics. R package version 2.3. Available at https://cran.r-project.org/web/packages/gridExtra/gridExtra (accessed 14 January 2022).

Cavalcante TRM, Naves RV, Franceschinelli EV and Silva RP (2009) Polinização e formação de frutos em araticum. *Bragantia* **68**, 13–21.

Cohen KO, Sano SM, Silva JCS and Melo JT (2010) Avaliação das características físicas e físico-químicas dos frutos de araticum procedentes de Cabeceiras, GO. Embrapa Cerrados, Planaltina, Distrito Federal. 17 p.

Cruz CD (2016) Genes software – extended and integrated with the R, Matlab and Selegen. *Acta Scientiarum Agronomy* **38**, 547–552.

Earl DA and Vonholdt BM (2012) Structure Harvester: a website and program for visualizing structure output and implementing the Evanno method. *Conservation Genetics Resources* **4**, 359–361.

Ellegren H and Galtier N (2016) Determinants of genetic diversity. *Nature Reviews Genetics* **17**, 422–433.

Evanno G, Regnaut S and Goudet J (2005) Detecting the number of clusters of individuals using the software Structure: a simulation study. *Molecular Ecology* **14**, 2611–2620.

Excoffier L and Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564–567.

Excoffier L, Smouse PE and Quattro JM (1992) Analysis of molecular variance inferred from metric distance among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–449.

Falush D, Stephens M and Pritchard J (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes* **7**, 574–578.

Filippos A (2016) Conservation and monitoring of tree genetic resources in temperate forests. *Current Forestry Reports* **2**, 119–129.

Galili T, Benjamini Y, Simpson G, Jefferis G, Gallota M, Renaudie J, Hornik K, Ligges U, Spiess AN, Horvath S, Langfelder P, Loo MVD, Vries A, Gu Z, Cath G, Ma J, Krzysiek G, Hummel M, Clark C, Graybuck L, Jdetribol J, Ho B, Perreault S, Hennig C, Bradley D, Huang H and Schupp P (2020) Extending 'dendrogram' functionality in R. R package version 1.13.4, 2020. Available at https://cran-r-project.org/web/packages/dendextend/dendextend (accessed 14 January 2022).

Gwinner R, Setotaw TA, Rodrigues FA, França DVC, Silveira FA, Pio LAS and Pasqual M (2016) Population structure of *Annona crassiflora*: an endemic plant species of the Brazilian Cerrado. *Genetics and Molecular Research* **15**, 1–13.

Inoue MH, Santana DC, Vilhena KSS, Souza Filho APS, Guilhon GMSP, Possamai ACS, Silva LE and Dallacort R (2010) Avaliação do potencial alelopático de substâncias isoladas em sementes de araticum (*Annona crassiflora*). *Planta Daninha* **28**, 735–741.

Kassambara A (2020) 'ggplot2' based publication ready plots. R package version 0.4.0, 2020. Available at https://cran-r-project.org/web/packages/ggpubr/ggpubr. (accessed 14 January 2022).

Kassambara A and Mundt F (2017) Factoextra: extract and visualize the results of multivariate data analyses. R package version 1.0.5, 2017. Available at https://CRAN.R-project.org/package=factoextra (accessed 14 January 2022).

Kiill LHP and Costa JG (2003) Biologia floral e sistema de reprodução de *Annona squamosa* L. (Annonaceae) na região de Petrolina-PE. *Ciência Rural* **33**, 851–856.

Lewontin RC (1972) The apportionment of human diversity. *Evolutionary Biology* **6**, 381–398.

Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K, Studer M, Roudier P, Gonzalez J, Kozlowski K, Schubert R, Murphy K (2019) 'Finding groups in Data': cluster analysis extended Rousseeuw et al. R package version 2.1.0, 2019. Available at https://cran-r-project.org/web/packages/cluster/cluster (accessed 14 January 2022).

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**, 209–220.

Ministério do Meio Ambiente – MMA (2022) O bioma Cerrado. Available at https://antigo.mma.gov.br/biomas/cerrado.html (accessed 14 January 2022).

Mojena R (1977) Hierarchical grouping methods and stopping rules: an evaluation. *The Computer Journal* **20**, 359–363.

Moog RJ and Bond JM (2003) A cheap, reliable and rapid method of extracting high-quality DNA from plants. *Molecular Ecology Notes* **3**, 666–668.

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

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, Mcglinn D, Minchin PR, O'hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E and Wagner E (2018) Vegan: community ecology package. R package version 2.4-5, 2018. Available at https://CRAN.R-project.org/package=vegan (accessed 14 January 2022).

Pereira JCM, Lemes RCGG, Castro JDB, Peixoto JC, Safadi GMVV and Oliveira MF (2022) Valoração do Araticum (*Annona crassiflora* Mart.). *Research, Society and Development* **17**, 1–15.

Pimenta AC, Amano E and Zuffellato-Riba KC (2017) Estaquia e anatomia caulinar de *Annona crassiflora* Mart. *Caderno de Ciências Agrárias* **9**, 1–7.

R Core Team (2016) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Sá GH, Lima FCD, Viana JPG, Lopes ACA, Carvalho LCB, Valente SES and Lima PSC (2022) Genetic diversity and structure of an active germplasm collection of *Annona squamosa* L. *Ecological Genetics and Genomics* **25**, 1–8.

Samaradiwakara SHMRNP, Samarasinghe WLG, Shantha PGS, Jayarathna KGCN, Dehigaspitiya P and Ubeysekera NM (2020) Genetic diversity assessment of selected *Annona muricata* L. germplasm in Sri Lanka. *Asian Journal of Agricultural and Horticultural Research* **7**, 1–13.

Santo-Silva EE, Almeida WR, Tabarelli M and Peres CA (2016) Habitat fragmentation and the future structure of tree assemblages in a fragmented Atlantic Forest landscape. *Plant Ecology* **217**, 1129–1140.

Santos IG and Oliveira MS (2020) Fatores que alteram o equilíbrio de Hardy-Weinberg. In Oliveira MS and Cruz CD (eds), *Genética de Populações com o Aplicativo GPOP*. Curitiba, PR: Brazil Publishing, pp. 97–141.

Shannon CE and Weaver WA (1949) *Mathematical Model of Communication*. Urbana: University of Illinois Press.

Silva Júnior AL, Cabral RLR, Sartori L, Souza LC, Miranda FD, Caldeira MVW, Moreira SO and Godinho TO (2020) Evaluation of diversity and genetic structure as strategies for conservation of natural populations of *Dalbergia nigra* (vell.) Allemão ex Benth. *Cerne* **26**, 435–443.

Souza JLC, Souza ERB, Naves RV, Guimarães RN and Melo APC (2020) Propagação de araticum por estaquia. *Agrossistemas* **12**, 223–236.

Strassburg BBN, Brooks T, Feltran-Barbieri R, Iribarrem A, Crouzeilles R, Loyola R, Latawiec AE, Oliveira Filho FJB, Scaramuzza CAM, Scarano FR, Soares Filho B and Balmford A (2017) Moment of truth for the Cerrado hotspot. *Nature Ecology & Evolution* **1**, 1–3.

Tatikonda L, Wani SP, Kannan S, Beerelli N, Sreedevi TK, Hoisington DA, Devi P and Varshney RK (2009) AFLP-based molecular characterization of an elite germplasm collection of *Jatropha curcas* L. biofuel plant. *Plant Science* **176**, 505–513.

Trackmaker (2022) GPS TrackMaker Version 13.9.608. Geo Studio Technology. Available at https://www.trackmaker.com/main/pt/licenca-gps-software-gratis (accessed 9 September 2022).

Turchetto-Zolet AC, Turchetto C, Zanella AM and Passaia G (2017) Marcadores Moleculares na Era genômica: Metodologias e Aplicações. Sociedade Brasileira de Genética, Ribeirão Preto.

Vieira AAR, Souza AC, Silva Júnior AL, Alves BQ, Miranda FD, Moreira SO and Caldeira MVW (2022) Diversity and genetic structure of *Astronium concinnum* Schott ex Spreng. in conservation units. *Plant Genetic Resources: Characterization and Utilization* **19**, 530–537.

Wilke CO (2019) Streamlined plot theme and plot annotations for 'ggplot2'. R package version 1.0.0, 2019. Available at https://cran.r-project.org/web/packages/cowplot/cowplot (accessed 14 January 2022).

Wright S (1951) The genetical structure of populations. *Annals of Eugenic* **15**, 323–354.

Wright S (1978) *Evolution and the Genetics of Populations: Variability within and among Natural Populations*. Chicago, IL, USA: University of Chicago press.

Yeh FC and Boyle TJB (1997) Population genetic analysis of co-dominant and dominant markers and quantitative traits. *Belgian Journal of Botany* **129**, 156–157.