Genetic structure in fragmented populations of *Solanum lycocarpum* A. St.-Hil. with distinct anthropogenic histories in a Cerrado region of Brazil

T.M. Moura¹, K. Martins², P.S. Sujii³, A.M. Sebbenn⁴ and L.J. Chaves⁵

¹Programa de Pós-Graduação em Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP, Brasil
²Departamento de Biologia, Universidade Federal de São Carlos, Campus Sorocaba, Sorocaba, SP, Brasil
³Programa de Pós-Graduação em Genética e Biologia Molecular, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP, Brasil
⁴Instituto Florestal de São Paulo, São Paulo, SP, Brasil
⁵Setor de Melhoramento Vegetal, Universidade Federal de Goiás, Goiânia, GO, Brasil

Corresponding author: T.M. de Moura
E-mail: tmariamoura@gmail.com

Received November 10, 2011
Accepted April 29, 2012
Published July 10, 2012
DOI http://dx.doi.org/10.4238/2012.July.10.16

**ABSTRACT.** *Solanum lycocarpum* is a woody tree widely distributed in the Cerrado that reaches high population densities in disturbed environments. We examined the genetic diversity and population differentiation of six *S. lycocarpum* populations with different degrees of human disturbance in order to determine if they are negatively affected by anthropogenic activity. Three populations located in southern and three located in southeastern regions of Goiás State, Central Brazil, were genotyped with five microsatellite markers. The population located in a protected area had higher number of alleles (26) than the remaining populations (19 to 21...
alleles). It indicates that extensive and continuous areas of preserved native vegetation contribute positively to the conservation of genetic diversity, even with *S. lycocarpum* that easily adapts to disturbed environments. The three southeastern populations, although fragmented, had preserved native vegetation and were not significantly different from each other ($\theta_p = 0.002$). All other population pairs compared were significantly divergent ($\theta_p$ varied from 0.03 to 0.11 between pairs, $P < 0.05$). We found three distinct sets of allele frequencies. The three southeastern populations shared similar gene pools, as well as the two disturbed southern populations, which are secondary vegetation. The southern population located in protected area had the most dissimilar gene pool. In conclusion, populations showing a higher degree of human disturbance tends to show a larger population differentiation than expected from the isolation by distance model, which in the current scenario of the Cerrado destruction points out to a threat to the long-term conservation of the species.

**Keywords:** Wolf fruit; Genetic conservation; Population differentiation; Genetic diversity; Microsatellite marker; Gene flow

**INTRODUCTION**

The Cerrado is the second largest Brazilian biome, and it is biologically the richest savannah in the world. Around 57% of the original Cerrado vegetation has been completely destroyed, and half of the remaining areas have been extensively degraded. The annual rate of deforestation is estimated to be 1.5%, which corresponds to three million hectares annually, a rate ten times higher than that estimated for the Atlantic Forest (Machado et al., 2004). Because of endemism, where 44% of vascular plant species are endemic, and due to the current rate of devastation, the Cerrado is considered one of the 25 biodiversity hotspots for global conservation (Myers et al., 2000). Up to 3% of the Cerrado range is in protected areas (Aguiar et al., 2004), and the distribution of these areas is not homogeneous throughout the biome. As a consequence, an important part of the diversity is not embedded in the national network of protected areas (Silva and Bates, 2002). Thus, one of aims for this biome’s conservation is to identify priority areas for conservation and to plan the establishment of new protected areas.

The conservation of genetic diversity is a necessary precondition for the maintenance of all levels of biodiversity and is an essential component of the sustainability of populations (Boyle, 2000; Namkoong et al., 2002). The reduction in natural habitats and the subsequent spatial isolation of populations cause changes in basic evolutionary processes of the populations, such as genetic drift, selection and migration, which result in changes in genetic diversity and structure. Therefore, to maintain levels of genetic diversity and to avoid extinction, it is necessary to conserve these evolutionary processes (Namkoong et al., 2002).

In the last decade, several studies have evaluated the genetic structure in Cerrado plant species (Zucchi et al., 2003; Soares et al., 2008; Moreno et al., 2009; Collevatti et al., 2001, 2010; Telles et al., 2003, 2010; Tarazi et al., 2010; Martins et al., 2006, 2011; Moura et al., 2009, 2011; Moraes and Sebbenn, 2011). Most of them addressed the genetic diversity and population differentiation, aiming to determine the conservation status of natural populations, which would then make it possible to infer historical events related to anthropogenic
disturbance, gene flow between populations and the effect of ecosystem fragmentation. In small populations, genetic drift is a significant determinant of the genetic structure; unlike in large populations, where the main factor influencing the genetic structure of plant species is reproductive biology, especially because reproductive biology determines the mating systems and the extent of the gene flow among populations. 

*Solanum lycocarpum* A. St.-Hil is an interesting plant species for the study of population genetic structure geared towards revegetation: 1) the species is widely distributed in the Cerrado; 2) the species occupies rapidly degraded environments, where it reaches high population densities; 3) the amount of knowledge regarding the reproductive biology of the species is considerable (Oliveira-Filho and Oliveira, 1988; Lombardi and Motta Jr., 1993; Courtenay, 1994; Moura et al., 2010); and 4) pollinators and seed dispersers contribute to long-distance gene flow (Martins et al., 2011). Because of these attributes, we expect no significant population differentiation. If the opposite were observed, we could infer it is a negative effect of human disturbance, as a consequence of genetic drift and/or restriction in gene flow. The establishment of secondary vegetation and the fragmentation are common types of human disturbance found in Cerrado biome, which could cause strong genetic drift and limited gene flow. We aimed to assess the genetic diversity and divergence among *S. lycocarpum* populations with different anthropogenic disturbance histories, to determine if there have been any negative effects and to contribute to the establishment of conservation plans for the Cerrado biome.

**MATERIAL AND METHODS**

**Study species**

*S. lycocarpum* is a woody species which produces functional male and hermaphrodite flowers in large quantities throughout the year. Large bees pollinate them (Oliveira-Filho and Oliveira, 1988). Various mammals disperse seeds over long distances, thus allowing the colonization of new environments. Since the fruits only finish their ripening process after falling to the ground (Lombardi and Motta Jr., 1993) many seeds germinate under the seed tree’s canopy or are scattered over short distances by leaf-cutter ants (Courtenay, 1994). The species occurs at a low density on primary vegetation but it is able to colonize anthropogenic environments, where it reaches higher densities. Although the higher population density would imply an apparent increase in effective population size, the simple increase in the number of individuals is not synonymous with genetic sustainability, as detected by Moura et al. (2011).

**Study site and sampling**

Six *S. lycocarpum* populations were studied in Goiás State, in Central Brazil. Three located in the southeast (called A, B and C) and the other three in the south (D, E and F) (Table 1). The F population has preserved vegetation and is situated in a protected area: the Parque Estadual da Serra de Caldas Novas. The remaining populations are located in fragments belonging to private ranches. Southeast populations were located in the Cerrado region encompassing areas of preserved vegetation and with low anthropogenic influence. The D and E populations consist of pastureland; however, D still retains native vegetation, and E has a predominance of signal grass (*Urochloa* spp) with few native species.
Sixty adult individuals were sampled for each population. Genomic DNA extraction was according to Doyle and Doyle (1987), with modifications. Five polymorphic microsatellite (SSR) primers developed for *Capsicum* spp and already amplified in *S. lycocarpum* (Martins et al., 2006) were used. The conditions of polymerase chain reaction (PCR) and the characterization of loci are described by Martins et al. (2006). After amplification, the DNA fragments were separated using a 4% polyacrylamide gel and visualized after staining with silver nitrate. The details of the procedures can be found in Martins et al. (2006) and Moura et al. (2009).

### Data analysis

We estimated the total number of alleles, the number of rare alleles (frequency of occurrence less than 0.05), and the fixation index (f) for each population. The 95% confidence interval of the fixation index was calculated using 1000 bootstraps over loci. Estimates of other parameters of genetic diversity for these populations were previously published (Martins et al., 2006; Moura et al., 2009). In this study, the average gene diversity (*H*\textsubscript{e}) and allelic richness (*R*\textsubscript{s}) (Petit et al., 1998) were both compared in the southeastern (populations A, B and C) and southern (populations D, E and F) regions by the *G*-test (Goudet et al., 1996) with 1000 randomizations and Bonferroni’s correction. All analyses were carried out using the FSTAT software (Goudet, 1995).

The genetic structure was characterized by the analysis of variance of gene frequencies (Weir, 1996), using the FSTAT software. The estimated parameters were: mean fixation index within populations (f), fixation index of the total set of populations (F) and genetic differentiation among populations (\(\theta_p\)). The 95% confidence intervals were estimated by 1000 bootstraps over loci. Subsequently, we estimated the genetic differentiation between pairwise populations (\(\theta_p\)).

The statistical significance of genetic differentiation between populations was assessed using the *G*-test (Goudet et al., 1996) with 1000 randomizations and Bonferroni’s correction. The indirect gene flow (\(N_{m}\)) between pairs of populations was estimated according to Wright (1951). For these calculations, *F*\textsubscript{ST} was replaced by \(\theta_p\), as suggested by Cockerham and Weir (1993) for an unbiased estimation of gene flow.

We applied the procedure described by Slatkin (1993) to assess whether gene flow results in a pattern of isolation by distance. The approach is based on the linear relationship between gene flow (\(N_{m}\)) and logarithms of geographical distance between populations. As discussed by Strand et al. (1996), in the context of isolation by distance and in a stepping-stone model of gene flow, a strong negative correlation indicates that populations are in equilibrium and that the current gene flow results in a pattern of isolation by distance, where genetic differentiation between populations increases with geographic distance. The lack of correlation between gene flow and geographical distance would suggest that populations are not in equilibrium.

<table>
<thead>
<tr>
<th>Region</th>
<th>Population</th>
<th>Municipality</th>
<th>Location</th>
<th>Vegetation condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Southeast</td>
<td>A</td>
<td>Bela Vista de Goiás</td>
<td>17°00’S; 48°46’W</td>
<td>Preserved Cerrado</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>São Miguel do Passa Quatro</td>
<td>17°00’S; 48°35’W</td>
<td>Preserved Cerrado</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>Orizonta</td>
<td>17°03’S; 48°22’W</td>
<td>Preserved Cerrado</td>
</tr>
<tr>
<td>South</td>
<td>D</td>
<td>Morrinhos</td>
<td>17°54’S; 49°11’W</td>
<td>Pastureland, with native vegetation</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>Morrinhos</td>
<td>17°55’S; 49°00’W</td>
<td>Pastureland, with fewer native vegetation</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>Caldas Novas</td>
<td>17°46’S; 48°40’W</td>
<td>Preserved Cerrado</td>
</tr>
</tbody>
</table>

Table 1. Description of the *Solanum lycocarpum* A. St.-Hil. populations studied in Goiás State, Brazil.
suggests that the prevailing gene flow mainly results in historical association among populations than the current inter-population one. Estimates of $\theta_p$ for each pair of populations were used to calculate the gene flow between them. The significance of the correlation between the logarithm of geographic distance and gene flow was evaluated by Mantel $Z$ statistics (Mantel, 1967) using 1000 permutations, with the NTSYS-pc software (Rohlf, 1989). The genetic divergence estimates $G_{ST}$ (Hedrick, 2005) for the whole population were obtained through the GDA_NT software (Degen, 2006).

The individual groups’ separation was determined by a Bayesian approach with the Structure 2.3.1 software (Pritchard et al., 2000). The model used allowed for genome mixing, with the number of groups ($K$) ranging from 1 to 7, with 500,000 resampling by the MCMC (Markov Chain Monte Carlo) method, 100,000 burn-ins and 10 independent repetitions. The most likely number of groups was found using the statistics described by Evanno et al. (2005), based on the rate of change in data probability between successive $K$ values. The peak in the $\Delta K$ value can identify the $K$ value that best represents the populations’ structuring.

**RESULTS**

The F population, located in a protected area, showed the highest number of alleles. The remaining populations had similar numbers of alleles, between 19 and 21 (Table 2). All populations had rare alleles, and these accounted for 33 and 43% of the total number of alleles for the southeastern and southern populations, respectively. We found that allelic richness ($R_s$) and gene diversity ($H_e$) were not significantly different between regions (Table 2). Inbreeding was observed only in the A population, judging by the confidence interval (Table 2).

<table>
<thead>
<tr>
<th>Table 2. Genetic diversity estimates in six <em>Solanum lycocarpum</em> populations in Goiás State based on five SSR loci.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Southeast</strong></td>
</tr>
<tr>
<td><strong>A</strong></td>
</tr>
<tr>
<td>No. of individuals</td>
</tr>
<tr>
<td>Total No. of alleles</td>
</tr>
<tr>
<td>No. of rare alleles</td>
</tr>
<tr>
<td>$f$ (frequency &lt;0.05)</td>
</tr>
<tr>
<td>$H_e$</td>
</tr>
<tr>
<td>$R_s$</td>
</tr>
</tbody>
</table>

$f$ = fixation index with 95% confidence interval (95%CI) obtained by 1000 bootstraps over loci; $H_e$ = gene diversity; $R_s$ = allelic richness. NS = not significant, $P > 0.05$ by the $G$-test (Goudet et al., 1996).

By determining the genetic divergence between pairs of populations (Table 3), we observed that the southeastern C population was more similar to the southern F population than to its geographically closer neighbors (D and E). This shows that F was the most genetically diverse among the populations from the same region, although the genetic divergence between populations D and E ($\theta_p = 0.033$) was also significantly different from zero ($P < 0.05$). With the exception of the comparisons between populations A, B and C, all other population pairs compared were significantly divergent (Table 3).
Three distinct sets of allele frequencies were observed (K = 3) (Figure 1). Populations A, B, and C shared similar gene pools, with the same being observed for the D and E populations. The F population was the one with the most dissimilar gene pool.

The isolation by distance test showed a moderate negative correlation yet statistically significant between gene flow and logarithms of geographic distances between pairs of populations. Only 55% of gene flow variation was explained by the variation in geographical distance. This was mainly due to the high divergence between the F population and other populations in the southern region (Table 3). A strong correlation would indicate that populations are at equilibrium between genetic drift and gene flow and that the actual gene flow results in a pattern of isolation by distance, where genetic differentiation between populations increases with geographic distance. The correlation was moderate and significant ($r = -0.745; P = 0.011$) showing a trend toward isolation by the distance model and that there was no balance between genetic drift and gene flow.

**DISCUSSION**

The highest number of alleles was observed in the population located in Parque Estadual de Serra de Caldas Novas (F), a protected area. This is an indication that extensive and continuous areas of preserved native vegetation contributes positively to the conservation of genetic diversity, even with *S. lycocarpum*, which easily adapts to degraded environments as noted by Moura et al. (2011). In this study, we found that populations whose native vegetation is preserved (populations A, B and C), but fragmented, had fewer alleles than the population located in a protected area, containing continuous vegetation.

The fixation index of the total set of populations ($F = 0.07$) and average fixation index within populations ($f = 0.002$) did not differ from zero; which indicates that a major portion of
the total inbreeding is due to population subdivision, and thus, it is a result of genetic drift, and restricted gene flow between populations. Population differentiation was low ($\theta_p = 0.068$) but significant. The population differentiation among southeastern populations (A, B and C) ($\theta_p = 0.002$) was not statistically different from zero. This indicates that they consisted of a single panmictic population prior to the Cerrado’s fragmentation. Since nowadays native vegetation is discontinuous in the region, they can be considered subpopulations. In contrast, the divergence among the southern populations (D, E and F) was significant and greater than population differentiation in the total set of populations ($\theta_p = 0.076$), showing that these populations are heterogeneous among themselves. Contrary to our expectations the population differentiation estimated according to Hedrick (2005) ($G_{ST} = 0.041$) was lower than the one estimated by the Weir and Cockerham (1984) statistics. This was probably due to the low polymorphism of our microsatellite loci, in contrast to what is generally observed for this type of marker. The greatest genetic divergence between the southern and southeastern (A, B and C) populations might have been due to their particular anthropogenic history. It is believed that, historically, they had suffered less anthropogenic influence than the D and E populations. This assumption was corroborated by the vegetation’s appearance; D and E populations are degraded areas, possibly, being secondary type vegetation.

We noted that each pool of alleles was related to an environmental condition. The F population is situated in a protected area and had a particular set of genes, and the D and E populations apparently have similar anthropogenic histories and share a similar gene pool, the same being observed for the A, B and C populations. Our results show that even species that easily occupy degraded environments, such as *S. lycocarpum*, have experienced negative effects due to vegetation fragmentation. The significant genetic divergence among populations, as occurred in the southern part of the state where the environment is more anthropogenic, shows that the populations are losing connectivity, probably due to reduced number of populations and restriction in foraging by their pollinators and seed dispersers.

Although studies with the Cerrado’s woody species have recently been published, there are still gaps in our knowledge of the genetic diversity of the Cerrado’s species, which must be filled. Few studies, for example, have assessed mating systems in detail and contemporary pollen flow (see Collevatti et al., 2010; Moraes and Sebbenn, 2011). Although virtually all authors argue about the risks of the rapid rate of deforestation, the effect of habitat fragmentation has not been systematically evaluated. There is a predominant use of indirect approaches for estimating historical gene flow, one reason being that biome fragmentation is recent. Only two studies (Collevatti et al., 2003; Martins et al., 2011) used a comparative approach with nuclear and chloroplast markers to assess the relative contribution of pollen and seeds to total gene flow. Only studies by Moura et al. (2009, 2011) compared populations in fragmented and continuous vegetation. It is also worth emphasizing the remarkable territorial extent of the biome and the need for studies covering larger geographic ranges. The majority of studies were conducted in Goiás State and the Distrito Federal, while the States of Mato Grosso, Mato Grosso do Sul, Tocantins, and São Paulo have received little attention.

**ACKNOWLEDGMENTS**

We are thankful to the staff of LARGEA (Laboratório de Reprodução e Genética de Espécies Arbóreas, ESALQ/USP), and to “Espaço da escrita” translator service at UNICAMP.
Research supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP/Brazil). K. Martins received a PhD fellowship from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq/Brazil).

REFERENCES


