Evidences for multiple maternal lineages of *Caryocar brasiliense* populations in the Brazilian Cerrado based on the analysis of chloroplast DNA sequences and microsatellite haplotype variation

ROSANE G. COLLEVATTI,*† DARIO GRATTAPAGLIA*† and JOHN. D. HAY‡

*Pós-Graduação em Ciências Genômicas, Universidade Católica de Brasília, SGAN 916, Mod. B, Brasília, DF. Brasil 70790–160,
†Laboratório de Genética de Plantas, EMBRAPA Recursos Genéticos e Biotecnologia, C.P.02372, Brasília, D.F. Brasil, 70770–900,
‡Departamento de Ecologia, Universidade de Brasília, Brasília, D.F. Brasil, 70910–900

Abstract

In this work we report on the phylogeography of the endangered tree species *Caryocar brasiliense* based on variability in two classes of maternally inherited chloroplast DNA sequences with different rates of molecular evolution. Eleven sequence haplotypes of a noncoding region between the genes *trnT* and *trnF* and 21 distinct 10-locus microsatellite haplotypes could be identified in a total of 160 individuals, collected in 10 widespread populations of *C. brasiliense*. An AMOVA indicated that most of the variation can be attributed to differences among populations, both for DNA sequence (87.51%) and microsatellites (84.38%). Phylogeography based on a median-joining network analysis of the noncoding region showed a sharp difference from the analysis of microsatellite haplotypes. Nevertheless, both analyses indicated that multiple lineages may have contributed to the origin of *C. brasiliense* populations in Brazilian Cerrado. Incongruences in the microsatellite haplotypes network suggest that homoplasy, which emerged from recurrent and independent mutations, greatly influenced the evolution of the *C. brasiliense* chloroplast genome. We hypothesize that our results may show the outcome of the restriction of ancient relic populations to moist refugias during extended droughts coinciding with glaciation in the northern hemisphere. The subsequent spread to favourable areas throughout Central Brazil may have caused contact between different lineages during the interglacial periods. The extinction of megafauna dispersers in the last glaciation may have caused a restriction in seed movement and currently, gene flow has been occurring mainly by pollen movement.

Keywords: *Caryocar brasiliense*, Caryocaraceae, cpDNA, microsatellites, phylogeography, tropical tree

Received 23 May 2002; revision received 23 August 2002; accepted 30 August 2002

Introduction

Species that undergo long–term biogeographic barriers to gene flow may be composed of geographical populations that belong to different gene genealogies and are correlated to geographical boundaries (Avise *et al.* 1987; Avise 1994). Phylogeography is based on the spatial distribution of gene genealogies, and provides a means of detecting the correlation between geographical distribution of haplotypes and their genealogical relationships (Avise *et al.* 1987). This area of study relies on the variability at more conserved cytoplasmic genomes with uniparental inheritance, low levels of mutation rate and no recombination, such as the chloroplast for plants (cpDNA), and the mitochondrial genome (mtDNA) for animals (Birky 1988; Avise 1994). Additionally, because of the haploid nature and mode of inheritance, the effective population size for the chloroplast genome is expected to be one-half that for the nuclear genome, leading to a stronger effect of
R. G. COLLEVATTI, D. GRATTAPAGLIA and J. D. HAY

Materials and methods

Populations, sampling and DNA extraction

Ten populations throughout the whole geographical distribution of Caryocar brasiliense, were surveyed (Fig. 1): AGE — Águas Emendadas Ecological Station, Brasília; CGR — Campus of the Federal University of Mato Grosso do Sul, Campo Grande; CNV — State Park of Caldas Novas, Goiás; FAL — Água Limpa Experimental Station, Brasília; GSV — Grandes Sertões Veredas National Park, Minas Gerais; ITI — Itirapina Ecological Reserve, São Paulo; MTR — Rondonópolis, Mato Grosso; PNB — Brasília National Park, Federal District; TOC — Porto Nacional, Tocantins; and URU — Serra da Mesa Hydroelectric affected area, Uruaçú. In each of these populations at least 30 individuals were sampled; expanded leaves were collected and stored at −80 °C. Genomic DNA extraction followed the standard CTAB procedure (Doyle & Doyle 1987).

In addition, two individuals of a vicariant species, Caryocar villosum, were analysed and included as an outgroup in the phylogenetic analysis.

Chloroplast sequencing analysis

Two fragments corresponding to a noncoding intergenic region of cpDNA, between the genes trnT and trnF, were sequenced in 16 individuals in each population. Using two
pairs of primers, ‘c-d’ and ‘e-f’, as described by Taberlet et al. (1991), fragments were amplified by polymerase chain reaction (PCR) in a 20-µL volume containing 1.0 ng of each primer, 1 U Taq DNA polymerase (Gibco), 200 µM of each dNTP, 1x reaction buffer (10 mM Tris–HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl₂), bovine serum albumin 3.25 µg and 10.0 ng of template DNA. Amplifications were performed using a GeneAmp PCR System 9700 (Applied Biosystems) with the following conditions: 96 °C for 2 min (one cycle), 94 °C for 1 min, 56 °C for 1 min, 72 °C for 1 min (30 cycles); and 72 °C for 7 min (one cycle). PCR products were sequenced on an ABI Prism 377 automated DNA sequencer (Applied Biosystems) using a DYEEnamic™ ET terminator cycle sequencing kit (Amersham Pharmacia Biotech), according to the manufacturer’s instructions. Both, ‘c-d’ and ‘e-f’ fragments were sequenced in forward and reverse directions.

Chloroplast microsatellite analysis

Ten chloroplast microsatellite loci developed by Weising & Gardner (1999) for dicotyledoneous angiosperms were used to genotype 160 individuals (16 in each population). For each locus the annealing temperature was optimized to produce clear and robust DNA products and an initial screening for polymorphism was carried out. Microsatellite marker amplifications were performed in a 13-µL volume containing 0.3 µM of each primer, 1 U Taq DNA polymerase (Gibco), 200 µM of each dNTP, 1x reaction buffer (10 mM Tris–HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl₂), bovine serum albumin 3.25 µg and 10.0 ng of template DNA. Amplifications were performed using a GeneAmp PCR System 9700 (Applied Biosystems) with the following conditions: 96 °C for 2 min (one cycle), 94 °C for 1 min, 56 °C for 1 min, 72 °C for 1 min (30 cycles); and 72 °C for 30 min (one cycle). Polymorphism detection in the screening step was carried out in 4% denaturing polyacrylamide gels stained with silver nitrate (Bassam et al. 1991). Preliminary allele sizing was carried out by comparison to a 10-base pair (bp) DNA ladder standard (Gibco) on a computer screen.

For those loci that amplified clear and polymorphic products across a sample of C. brasiliense trees, the forward primer was labelled with a fluorescent dye. Four loci
(ccmp1, ccmp2, ccmp5 and ccmp6) were labelled with 6-FAM, three with TET (ccmp4, ccmp7 and ccmp8) and three with HEX (ccmp3, ccmp9 and ccmp10). In the fluorescent genotyping procedure, PCR amplifications were performed in a 10-μL volume containing 0.4 μM of each primer, 1 U Taq DNA polymerase (Gibco), 200 μM of each dNTP, 1x reaction buffer (10 mm Tris–HCl, pH 8.3, 50 mm KCl, 1.5 mm MgCl2), bovine serum albumin 3.25 μg and 5.0 ng of template DNA. Amplifications were performed as in the screening step. PCR amplifications were performed separately for each locus. Reactions were then diluted 1:5 in three multiplexes with three loci each, labelled with different dyes, except for locus ccmp2 which was diluted separately. One microlitre of the 1:5 diluted reaction was added to 0.25 μL GeneScan 500 internal lane standard (ROX, Applied Biosystems), 0.45 μL of loading buffer (25 mm ethylenediaminetetraacetic acid and 50 mg/mL Blue-Dextran) and 2.3 μL deionized formamide. The reactions were then heated to 95 °C for 3 min, chilled on ice and electrophoresed in 5% denaturing polyacrylamide gel in an ABI Prism 377 automated DNA sequencer (Applied Biosystems). Fluorescent PCR products were automatically sized using GENESCAN AT GENOTyper softwares (Applied Biosystems).

Because of potentially inaccurate fragment size determination caused by Taq polymerase slippage and gel-to-gel variation (Haberl & Tautz 1999), we used two individuals from the CNV population as a reference sample in all PCR amplifications and electrophoresis runs. Additionally, for all microsatellite loci, PCR reactions and electrophoresis analyses were repeated for four individuals of each population, to verify allele sizing variation.

To verify if the amplified products did contain a microsatellite, PCR products of each locus were sequenced using both, forward and reverse primers, on an ABI Prism 377 automated DNA sequencer (Applied Biosystems) using dye-terminator fluorescent chemistry.

Data analysis

An analysis of molecular variance (AMOVA, Excoffier et al. 1992) was performed on both the microsatellite and sequencing data using ARLEQUIN version 2000 (Schneider et al. 2000), to test the hypothesis of genetic differentiation among populations. The population genetic structure parameter θ (Weir & Cockerham 1984) and pairwise differentiation among populations (pairwise θ) were estimated from AMOVA. The significance of θ was tested by a nonparametric permutation test (Excoffier et al. 1992) implemented in the ARLEQUIN software. As most mutations in microsatellites follow the stepwise mutation model, Slatkin’s RsqT was estimated (Slatkin 1995) for chloroplast microsatellite haplotypes to compare with previous data of nuclear microsatellites (Collevatti et al. 2001a).

A significance test of differentiation was performed by randomizing genotypes among samples to obtain the log-likelihood G statistics (Goudet et al. 1996). An AMOVA was also performed to test the hypothesis of differentiation of C. brasiilense populations from C. villosum.

Intraspecific phylogenies for both microsatellite and sequencing data were inferred using median-joining network analysis (Bandelt et al. 1999) performed with the software NETWORK (Forster et al. 2000). In this analysis a minimum spanning network was constructed based on the union of all minimum spanning trees, connecting all given sequences without creating any cycles or inferring an additional ancestral node (Bandelt et al. 1999). Using parsimony criteria, the software finds the median vectors, i.e. the consensus sequences of mutually close sequences, biologically interpreted as possible unsampled sequences or extinct ancestral sequences (Bandelt et al. 1999). Sequences were aligned using CLUSTALX (Thompson et al. 1997), and characters (each base pair) were equally weighted before analysis. For microsatellite data, characters (loci) were equally weighted. Caryocar villosum haplotypes were included in the analysis as an ancestral haplotype for both sequence and microsatellite data.

Results

Chloroplast sequencing and microsatellites characterization

Amplification of the noncoding regions using ‘c-d’ and ‘e-f’ primer pairs generated fragments of 446 bp and 410 bp, respectively. Polymorphisms in the ‘c-d’ fragments were mainly because of point mutations. Twenty-six substitutions were found, in contrast to only six insertions/deletions of 1 bp. The ‘c-d’ fragment is characterized by a high abundance of adenine (40%) and thymine (30%), and the presence of many mononucleotide repeats along the sequence. The longest repeat was a region of three (TA)6 repeats in tandem, spaced by a small sequence of (TA)7TGA(TA)2. The other longest repeats were a (TA)6 one (A)10 and one (A)7. Because the haplotypes in ‘e-f’ fragments differed only by the substitution of one thymine by an adenine or a guanine in different positions of a (T)10 mononucleotide repeated sequence, the fragment was not used in the phylogeographic analysis. Only three different point mutations were added when the sequence of Caryocar villosum was included to the analysis, two insertions and one substitution.

Eleven different haplotypes (Table 1) could be distinguished for the ‘c-d’ fragment. Although most populations displayed only one haplotype (Table 1), in populations ITI and MTR more than one haplotype could be detected. Populations GSV, MTR and TOC shared the same haplotype (Hapl 7), the most frequent one (Table 1). The two
individuals of *C. villosum* showed a common and distinct haplotype from *C. brasiliense*.

All chloroplast microsatellite loci developed by Weising & Gardner (1999) were transferable to *C. brasiliense* and detected intraspecific polymorphism providing a powerful tool for phylogeography analysis. Fluorescent detection of polymorphism at the microsatellite loci allowed a high-resolution analysis where allele differences of one nucleotide repeat unit could be identified and reproduced. Considering the 10 populations analysed in this work, and a total of 160 individuals, 21 distinct haplotypes could be identified for *C. brasiliense* (Table 2). No different haplotype was found when an additional sample of 24 individuals from the FAL population was analysed. All populations but AGE and FAL had more than one haplotype. These two populations shared a common haplotype AGEFAL (Table 2).

Sequencing of the amplified product from all 10 loci showed that the fragments did contain the expected microsatellite. The two individuals of *C. villosum* presented a unique and distinct haplotype from all *C. brasiliense* haplotypes.

### Phylogeography

The chloroplast genome showed a high and significant level of differentiation among populations (*P* < 0.0001) for both the microsatellites haplotype and the DNA sequence of the ‘c-d’ noncoding region. The analysis based on the microsatellite haplotype indicated that 84.38% of the total variation may be attributed to differences among populations. Although we expected a higher percentage of differentiation among populations for the ‘c-d’ noncoding region than for the microsatellite haplotype, this value (87.51%) was as high as that obtained for microsatellite haplotypes. Although a high percentage of variation among groups was found for the ‘c-d’ noncoding region (32.61%, Table 3), when we analysed the divergence among groups (*C. villosum* and *C. brasiliense*), it was not statistically significant. Population differentiation based

### Table 1

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Popln</th>
<th>No. of indiv.</th>
<th>Haplotype</th>
<th>Popln</th>
<th>No. of indiv.</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTR</td>
<td>MTR</td>
<td>8</td>
<td>Hapl7</td>
<td>GSV, MTR, TOC</td>
<td>40</td>
</tr>
<tr>
<td>FAL</td>
<td>FAL</td>
<td>16</td>
<td>CNV</td>
<td>CNV</td>
<td>16</td>
</tr>
<tr>
<td>PNB</td>
<td>PNB</td>
<td>16</td>
<td>ITI1</td>
<td>ITI</td>
<td>2</td>
</tr>
<tr>
<td>AGE</td>
<td>AGE</td>
<td>16</td>
<td>ITI2</td>
<td>ITI</td>
<td>8</td>
</tr>
<tr>
<td>CGR</td>
<td>CGR</td>
<td>16</td>
<td>ITI3</td>
<td>ITI</td>
<td>6</td>
</tr>
<tr>
<td>URU</td>
<td>URU</td>
<td>16</td>
<td>Vill</td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>

**Table 2** Microsatellite cpDNA haplotypes observed for 10 populations of *Caryocar brasiliense* and two individuals of *C. villosum*

| Haplotype | Population | ccmp1 | ccmp2 | ccmp3 | ccmp4 | ccmp5 | ccmp6 | ccmp7 | ccmp8 | ccmp9 | ccmp10 | No. of indiv. |
|-----------|------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|---------------|
| CNV1      | CNV        | 120   | 109   | 103   | 119   | 116   | 87    | 117   | 117   | 117   | 114   | 2      |
| CNV2      | CNV        | 121   | 109   | 103   | 119   | 116   | 87    | 117   | 116   | 117   | 114   | 2      |
| CNV3      | CNV        | 120   | 109   | 103   | 119   | 116   | 87    | 117   | 116   | 117   | 114   | 4      |
| CNVTOC1   | CNV, TOC   | 120   | 109   | 103   | 119   | 117   | 87    | 117   | 116   | 117   | 115   | 9      |
| Hapl6     | CNV, TOC, PNB, URU | 121 | 109 | 103 | 119 | 117 | 87 | 117 | 116 | 117 | 115 | 12   |
| Hapl7     | CNV, TOC, PNB, URU | 121 | 109 | 103 | 119 | 117 | 87 | 117 | 116 | 117 | 115 | 10   |
| PNBTOC    | PNB, TOC   | 121   | 109   | 103   | 120   | 117   | 87    | 117   | 116   | 117   | 115   | 6      |
| TOC5      | TOC        | 121   | 109   | 103   | 121   | 117   | 87    | 117   | 117   | 117   | 115   | 3      |
| TOC7      | TOC        | 120   | 109   | 103   | 119   | 116   | 87    | 117   | 117   | 117   | 115   | 1      |
| GSV1      | GSV        | 121   | 109   | 105   | 120   | 116   | 80    | 114   | 114   | 114   | 114   | 14     |
| 114GVSITI | GSV, ITI   | 122   | 109   | 105   | 121   | 116   | 80    | 114   | 114   | 114   | 114   | 14     |
| ITI3      | ITI        | 122   | 109   | 105   | 120   | 116   | 81    | 114   | 114   | 114   | 114   | 11     |
| CGRMTR1   | CGR, MTR   | 121   | 108   | 106   | 121   | 116   | 87    | 117   | 116   | 117   | 115   | 21     |
| CGRMTR2   | CGR, MTR   | 121   | 108   | 106   | 120   | 117   | 87    | 117   | 116   | 117   | 115   | 7      |
| MTR2      | MTR        | 121   | 108   | 106   | 120   | 117   | 87    | 117   | 117   | 117   | 115   | 1      |
| MTR4      | MTR        | 121   | 108   | 106   | 120   | 117   | 87    | 117   | 117   | 117   | 115   | 1      |
| MTR6      | MTR        | 121   | 108   | 106   | 121   | 117   | 87    | 117   | 117   | 117   | 115   | 2      |
| URU1      | URU        | 121   | 109   | 103   | 119   | 116   | 87    | 117   | 117   | 117   | 115   | 4      |
| URU2      | URU        | 121   | 109   | 103   | 120   | 116   | 87    | 117   | 117   | 117   | 115   | 6      |
| PNBURU    | PNB, URU   | 121   | 109   | 103   | 120   | 117   | 87    | 117   | 117   | 117   | 115   | 5      |
| AGEFAL    | AGE, FAL   | 119   | 109   | 101   | 119   | 119   | 87    | 117   | 119   | 117   | 118   | 32     |
| Vill      | *C. villosum* | 120 | 109 | 103 | 116 | 116 | 87 | 116 | 116 | 116 | 114 | 2      |

Alleles identified by base pairs.
on allele size variation for microsatellites ($R_{ST} = 0.9822, \ P < 0.0001$) was higher than the estimate based on variance on allele frequencies ($\theta = 0.8438, \ P < 0.0001$).

Because of the failure of parsimony methods in the construction of a meaningful tree, resulting in polytomies, the median-joining network approach was used for both the microsatellites and sequence haplotypes. A large number of equally parsimonious trees were detected when a maximum parsimony optimum criterion was used with pAUP (Swofford 1993). The consensus trees were comb-like with no internal structure. The network approach has been used for human mitochondrial DNA sequences and Y microsatellite haplotypes (e.g. Bandelt et al. 1995; Cooper et al. 1996), and recently in plant phylogeography analysis using chloroplast microsatellites (Clark et al. 2000).

Thirty-two mutations explained the median-joining network based on the ‘c-d’ fragment (Fig. 2), and 26 mutations did so disregarding the torso (comprised by the haplotypes CGR, URU, AGE, HAPL7, and the median vectors mv1 and mv2). Haplotypes of populations localized in Central and Western Brazil tended to be grouped — AGE, GSV, CGR, TOC, URU and MTR (Fig. 2). The ITI population seemed to have multiple origins and was highly diverged from the other Caryocar brasiliense populations in Central Brazil.

Fifty-one mutations were invoked to explain the network found in this study. Haplotypes from populations GSV and ITI were grouped together with a common origin at the bottom of the network (Fig. 3). Additionally, these populations shared a haplotype. This pattern could not be explained by gene flow, as these populations are 800 km

### Table 3 Analysis of molecular variance based on the sequencing of ‘c-d’ noncoding region and on 10 chloroplast microsatellite loci, for 10 populations of Caryocar brasiliense and on two individuals of a vicariant species, C. villosum

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>df</th>
<th>Sum of squares</th>
<th>Variance components</th>
<th>% variation</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequencing data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among groups</td>
<td>1</td>
<td>9,927</td>
<td>1.2195</td>
<td>32.61</td>
<td>0.1838</td>
</tr>
<tr>
<td>Among populations</td>
<td>9</td>
<td>320.675</td>
<td>2.2074</td>
<td>59.02</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Within populations</td>
<td>151</td>
<td>47.250</td>
<td>0.3129</td>
<td>8.37</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Total</td>
<td>161</td>
<td>377.852</td>
<td>3.7398</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microsatellite data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among groups</td>
<td>1</td>
<td>7.528</td>
<td>0.3224</td>
<td>9.31</td>
<td>0.359</td>
</tr>
<tr>
<td>Among populations</td>
<td>9</td>
<td>386.538</td>
<td>2.6538</td>
<td>76.61</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Within populations</td>
<td>151</td>
<td>73.688</td>
<td>0.4880</td>
<td>14.09</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Total</td>
<td>161</td>
<td>467.753</td>
<td>3.4642</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

df, degrees of freedom.
apart, and in clearly different biogeographic zones. Independent mutation events converging to a shared haplotype could explain this connection. Attempts were made to explore the problem by down-weighting rapidly mutating characters, such as ccpm4, ccpm5 and ccpm8, or increasing the $e$-value ($e = 1$ and $e = 2$) but the problem persisted. The major structure of the network was not influenced by the shared haplotype between GSV and ITI. Although the removal of these haplotypes from the analysis caused a reduction in the torso size and number of median vectors, the main links among the other haplotypes were not lost.

Despite the differences in the network obtained by DNA sequence and microsatellite haplotypes, especially in the topology and the presence of more haplotypes in the microsatellite-based network, they shared some general aspects. In both networks there was a tendency to group the populations of Central Brazil, such as CGR, MTR, TOC, URU and CNV.

**Discussion**

High levels of genetic differentiation were detected for the chloroplast genome ($\theta = 0.8438$ and $R_{ST} = 0.9822$), in
sharp contrast to the nuclear genome ($\theta = 0.07, R_{ST} = 0.29$, Collevatti et al. 2001a). As expected, population differentiation for the maternally inherited genome (cpDNA) was greater than for the biparentally inherited nuclear genome (Ennos 1994). Under the migration-drift equilibrium for the maternally inherited genome (cpDNA) island model to obtain an estimate of gene flow (Birky et al. 1983; Ennos 1994), the mean number of migrants per generation for the nuclear genome was equal to $N_{m} = 3.32$.

For the maternally inherited chloroplast genome

$$F_{ST} = \frac{1}{(4Nm + 1)}, \quad Nm = 0.093, \quad \text{overall chloroplast microsatellite loci},$$

and $Nm = 0.071$ for the ‘c-d’ noncoding region. The ratio of pollen to seed flow for population inbreeding (because $f$ was statistically different from zero for Caryocar brasiliense) was $f = F_{IS} = 0.11$, Collevatti et al. 2001a), could be calculated as proposed by Ennos (1994), i.e. (pollen flow)/(seed flow) = [(1/$F_{ST}$) - 1] $(1 + F_{S})$ $2[(1/F_{ST}) - 1] / [(1/F_{ST}) - 1] = 77.68$, comparing nuclear and chloroplast microsatellites, and (pollen flow)/(seed flow) = 101.34, comparing nuclear microsatellites and ‘c-d’ chloroplast noncoding region. These results show that, considering migration drift–equilibrium, pollen movement is significantly greater than seed movement. Therefore the low population differentiation for the nuclear genome is most likely because of pollen movement rather than seed dispersal. Another previous study supports the hypothesis that the connection among populations of Caryocar brasiliense is mainly because of pollen flow (Collevatti et al. 2001b). In fact, despite a high outcrossing rate ($f = 1.0$), Caryocar brasiliense tends to display small panmictic units or demes as a result of a restriction in gene flow, indicated by a high biparental inbreeding, Caryocar brasiliense is pollinated by small territorial bat species with low flight ranges (Gribel & Hay 1993) and the great majority of seeds are dispersed by gravity, so that seeds tend to remain under the mother tree’s canopy (J. D. Hay, unpublished).

Although the phylogeographic analyses based on the ‘c-d’ noncoding region and on the microsatellite loci showed that multiple maternal lineages may have been at the origin of the current populations of Caryocar brasiliense, the network obtained by the analysis of the cpDNA noncoding region showed a sharp difference from the network based on microsatellites. The distribution of the ITI population haplotypes was unexpected. The phylogeographical relationship of ITI suggests that multiple lineages may have been at the origin of populations of Caryocar brasiliense in the Western São Paulo State, the southwestern limit of the Cerrado geographical distribution (Fig. 1). Our results indicate that these populations may have originated from stocks from Goiás (CNV) and Mato Grosso (MTR).

The divergence in the chloroplast genome for the two vicariant species, Caryocar brasiliense and C. villosum, may be lower than the divergence found among some populations of Caryocar brasiliense. As Caryocar brasiliense is a widely distributed species, this pattern may be the outcome of an ‘isolation by distance’ (Wright 1943) process, or of a high mutation rate in the ‘c-d’ noncoding region and in the microsatellite loci studied. The network based on the ‘c-d’ noncoding region indicated that populations from the North, Northwestern and Northeastern (TOC, MTR, GSV) limits of the Cerrado distribution seemed to have originated from the haplotype of the vicariant species C. villosum. MTR and C. villosum had one of the lowest values of pairwise differentiation ($\theta = 0.7803$). In fact, MTR shared the lowest values of pairwise $\theta$ with all analysed populations, ranging from 0.4667 (GSV and TOC) to 0.8545 (FA. This result was expected, because the MTR population is situated in the northwestern limit of the Cerrado biome (Fig. 1) and southeastern limit of the Amazonia forest, where the geographical distribution of C. villosum (a ‘terra-firme’ forest tree) coincides with Caryocar brasiliense (a ‘Cerrado sensu stricto’ tree). It is possible that ancient or ongoing hybridization has been taking place in this region. In fact, the flowers of these two species are indistinguishable, and they are potentially pollinated by the same bat species. The elucidation of this result will require the analysis of other DNA sequences with lower mutation rates, such as coding regions, the analysis of more individuals from C. villosum, and a detailed study of pollination and mating structure for both species.

Both networks showed that homoplasy, potentially emerging from recurrent and independent mutations, may influence the evolution of the chloroplast genome of Caryocar brasiliense. The analysed ‘c-d’ region of the chloroplast genome of Caryocar brasiliense is rich in repeated sequences, especially mononucleotide repeats such as (T), and (A)$_{n}$. The analysis of these regions may fail to contribute to an informative phylogeny because of the high mutation rates, irregularities and asymmetries in mutations, and degradation of microsatellites over time with the substitution and insertion of other nucleotides (Goldstein & Pollock 1997). In fact, many mutations in the noncoding fragment analysed were substitutions or insertions in these repeated areas, leading to some incongruent results in the analysis of the sequence-based network. The same influence of mutation rate, irregularities, asymmetries and microsatellite degradation may be found in the analysis of the microsatellite haplotypes.

In the microsatellite-based network, the common origin of GSV and ITI populations, located in geographically different and distant regions, as well as the common origin of both populations from C. villosum, is most likely caused by a lack of polymorphism to differentiate the populations and by independent mutations. Thus more microsatellite
loci would be necessary to resolve these questions, or other classes of chloroplast molecular markers, including sequences from regions with different rates of evolution. For the same reason, we expected populations AGE, FAL and PNB to be connected by a common ancestral haplotype, because these populations are in the same geographical zone (Fig. 1), and pairwise distances among these populations are less than 20 km. However, contrary to our expectations, we found one of the highest values of pairwise genetic differentiation (θ = 0.983, P < 0.0001) among these populations.

Despite the discrepancies between the results from the noncoding DNA sequence and microsatellite data, our results strongly support the hypothesis that the high population subdivision for the chloroplast genome in *C. brasiliense* may be explained by the restriction in seed movement and possibly by ancient changes in species distribution during glaciation events. The Quaternary glaciations changed the South American landscape, causing a dynamic change in species distribution and the formation of tropical habitat refugia, possibly contributing to the extinction of the megafauna (e.g. Frane 1982; Behling 1998; Ledru et al. 1998; Salgado-Labouriau et al. 1998).

Current patterns of chloroplast population subdivision may be caused by the loss of ancient dispersers, leading to restriction of seed dispersal by gravity, keeping seeds under the mother tree’s canopy and thus isolating maternal genealogies. The megafaunal fruit hypothesis states that many living fruits are botanical anachronisms adapted for dispersal by an extinct megafauna (Janzen & Martin 1982). *Caryocar* appeared in the fossil record of the Tertiary, in the middle Eocene [nearly 50–44 million years ago (Ma)] in the Amazonian Forest of Venezuela (Muller 1981). The mammalian megafauna evolved in the late Tertiary (nearly 20 Ma). It is possible that, as hypothesized for other species, *C. brasiliense*’s large fruit evolved in an environment where megafauna predominated. The last megafauna became extinct in the Quaternary (~ 10 000 years ago) probably because of environmental deterioration caused by the last Wisconsin glaciation, or through overhunting (Martin & Wright 1967; Lewin 1983). The last glaciation (Wisconsin) coincided with a drier climate in the southern and central Cerrado domain (12 000–8000 years ago). Savannah landscapes were dominated by grasslands and frequent fires were recorded. In the fossil pollen record, arboreal and ‘vereda’ (palm swamp) elements are rare and restricted to semi-deciduous forest taxa (Salgado-Labouriau et al. 1998). It is possible that *C. brasiliense*, as some other arboreal savannah taxa, became restricted to sites with moist climatic conditions that served as refugias. After 7000 years ago, the climate became moister and the arboreal pollen record is dominated by savannah vegetation. After 4000 years ago, Cerrado and semi-deciduous forests attained their modern distribution and species composition and the seasonality continued to increase until it reached the current pattern (Salgado-Labouriau et al. 1998). The current pattern of chloroplast haplotype distribution could have been determined by the founder effect of maternal genealogies. Populations restricted to refugias in the last glaciation may have spread and dispersed to favourable areas in the last 7000 years, attaining the present geographical distribution. Because the megafauna was already extinct by 7000 years ago, *C. brasiliense* may have then lost seed dispersers. Seed movement was then restricted to surrounding areas, leading to the phylogeographic break observed in this study, with one lineage in each region and multiple lineages giving rise to the current populations of *C. brasiliense* in the Brazilian Cerrado. Furthermore, the sequence of several glaciations led to an advance of the savannah-like vegetation in Cerrado and a retreat of the tree species and forest vegetation in glaciation, and the opposite in interglacial periods (Salgado-Labouriau et al. 1998). This sequential pattern of climatic changes may have caused several cycles of spreading of *C. brasiliense* lineages and the contact of different lineages during the interglacial periods followed by the retreat of individuals from different previous lineages into a common refuge during the glaciation periods.

Even though many reports have demonstrated maternal inheritance of the chloroplast in angiosperms (e.g. Corriveau & Coleman 1988; Byrne et al. 1993), some exceptions may exist (Harris & Ingram 1991; Reboud & Zeyl 1994), and eventual paternal chloroplast slippage may influence the intraspecific haplotype diversity, leading to a false ‘maternal multiple lineage’ pattern. A more detailed study of the mode of inheritance of the chloroplast in Caryocaraceae is necessary to exclude possible paternal effects in our results. Furthermore, the identity of some haplotypes, the co-occurrence of multiple lineages and some unexpected results may also be the consequence of human movement between these regions leading to seed dispersal of *C. brasiliense*. The fruit of this species was used by ancient human populations in Central Brazil as a source of food and oil. Presently, it plays a significant role in the local food economy of the inhabitants of Central Brazil who use the yellow mesocarp as an important source of oil for cooking and for home-made recipes for sweets, ice-cream and liqueur.

Finally, this report shows that the comparative analysis of the distribution of genetic variation in different sequences of the chloroplast genome, and also of similar markers in the nuclear genome, allows a more in depth understanding of the detailed genetic structure and dynamics of both ancient and more recent events that shaped the current distribution of genetic variation of tropical tree populations, for which information is typically scarce and very much needed for planning and executing scientifically sound conservation programmes.
Acknowledgements

This study is part of the doctoral thesis of the first author, and has been supported by WWF (World Wildlife Fund), CNPq/PADCT (Brazilian Ministry of Science and Technology) competitive grant 62.00059/97–4 to D.G., and a fellowship to R.G.C., whose assistance we gratefully acknowledge. We wish to thank Rosana P. V. Brondani for continued support in the laboratory, Dr Forster for the authorization for the use of the network software as well as Dr Berthold Heinze for very helpful comments that significantly improved this manuscript. We also thank Aécio A. Santos, Antônio Dimas Costa Jr., Bruno Walter, Eugenia C. Amaral, Fabian Borghetti, Flavia Pinto, Frederico S. Lopes, Galdino R. Santos, João B. Pereira, Maura C. Almeida, Neilton R. Oliveira, Sebastião P. Cordovil, for their kind help in field work. We acknowledge the support of IBAMA (Brazilian Ministry of Environment), IEF-SP (State Forest Institute of São Paulo), FEMAGO (State Environmental Intitute of Goiás) and SEMATEC (Environmental Institute of Brasilia) for authorizing plant collections in the Brasilia and Grandes Sertões Veredas National Parks, at Itirapina Ecological Station, in the State Park of Caldas Novas, and at Águas Emendadas Ecological Station.

References


This study is part of the doctoral thesis of Rosane G. Collevatti who is currently involved in the application of molecular markers to population genetics, phylogeny and conservation of tropical tree species. Dr Dario Grattapaglia is a geneticist involved in the development and use of molecular markers for population genetics and genomic mapping in forest trees. Dr John D. Hay is an ecologist and works on the demography and dynamics of tropical tree species. This work is part of a project of the Catholic University of Brasilia and National Institute for Amazonian Research involving the development of molecular tools for the understanding of population genetic structure, mating systems and the phylogeny of tropical tree species to eventually guide conservation strategies.