Impact of selective logging on inbreeding and gene dispersal in an Amazonian tree population of *Carapa guianensis* Aubl.

D. CLOUTIER,* M. KANASHIRO,† A. Y. CIAMPI‡ and D. J. SCHOEN*

*Biology Department, McGill University, 1205 Ave. Docteur Penfield, Montréal, Québec, Canada, H3A 1B1, †Embrapa Amazônia Oriental, Trav. Dr Eneas Pinheiro, S/N Marco, CP 48, CEP 66095-100, Belém, Pará, Brazil, ‡Embrapa Cenargen, PqEB, Final Av. W5 Norte, CP 2372, CEP 70770-900, Brasília, DF, Brazil

Abstract

Selective logging may impact patterns of genetic diversity within populations of harvested forest tree species by increasing distances separating conspecific trees, and modifying physical and biotic features of the forest habitat. We measured levels of gene diversity, inbreeding, pollen dispersal and spatial genetic structure (SGS) of an Amazonian insect-pollinated *Carapa guianensis* population before and after commercial selective logging. Similar levels of gene diversity and allelic richness were found before and after logging in both the adult and the seed generations. Pre- and post-harvest outcrossing rates were high, and not significantly different from one another. We found no significant levels of biparental inbreeding either before or after logging. Low levels of pollen pool differentiation were found, and the pre- vs. post-harvest difference was not significant. Pollen dispersal distance estimates averaged between 75 m and 265 m before logging, and between 76 m and 268 m after logging, depending on the value of tree density and the dispersal model used. There were weak and similar levels of differentiation of allele frequencies in the adults and in the pollen pool, before and after logging occurred, as well as weak and similar pre- and post-harvest levels of SGS among adult trees. The large neighbourhood sizes estimated suggest high historical levels of gene flow. Overall our results indicate that there is no clear short-term genetic impact of selective logging on this population of *C. guianensis*.

Keywords: genetic diversity, mating system, microsatellite loci, pollen dispersal, selective logging, spatial genetic structure

Received 9 July 2006; revision accepted 7 October 2006

Introduction

The exploitation of a proportion of forest trees chosen for the economic value of their wood, i.e. selective logging, has a direct impact on the demography of harvested tree populations, as it: (i) removes large individuals that likely contribute more to reproduction than smaller individuals, and (ii) reduces the overall population density of reproductive individuals. Additionally, tree harvesting may result in ‘collateral damage’ to the forest habitat (e.g. a shift in the abundance, diversity and behaviour of animal pollinators) that in turn can have an impact on the reproductive biology of remaining tree populations. From a population genetics perspective, such impacts may induce changes in patterns of genetic diversity, inbreeding, gene flow, and the effective size of populations. Two expected genetic consequences of small population size include increases in the level of genetic drift and inbreeding (Ellstrand & Elam 1993). The potential losses of genetic variation associated with logging may decrease the ability of populations to adapt to directional abiotic and biotic environmental changes as well as reduce their flexibility to respond to short-term challenges such as pathogens and herbivores (Lowe et al. 2005). Thus, reductions in population size and increase in inbreeding resulting from selective logging may pose significant threats to the long-term viability of harvested tree populations.

The directly measurable population genetic effects of selective logging could fall into several categories. First,
selective harvesting could lead to an immediate loss of genetic diversity as a consequence of removal of adult trees; or in the seed progeny generation, as a consequence of reduced availability of pollen donors. Such changes, if they occur, should be detectable with marker genes. A study of *Carapa guianensis* in Central America found that marker allelic richness in saplings was lower than in the adult population in logged forests, but not in unlogged forests (Dayanandan et al. 1999; but see Hall et al. 1994). Second, coupled with the reduction in density following tree harvest, there may be a reduction in the number of individuals available to donate pollen. This may restrict pollen movement and increase inbreeding, two factors that could lead to loss of genetic diversity in the long term. In the review by Lowe et al. (2005), six out of eight experimental studies that assessed progeny arrays for levels of inbreeding highlighted significant anthropogenic impacts, including a study of *Carapa procera* in French Guiana where lower outcrossing rates were found in logged plots compared to undisturbed plots (Doligez & Joly 1997). Alternatively, one may hypothesize that pollinators are able to cope with the increased intertree distance following selective logging, and that gene dispersal could be maintained (or even increased). This alternate hypothesis is supported by studies where high levels of pollen dispersal were discovered in an undisturbed population of *C. guianensis* (Cloutier et al. in press), while increased pollen flow has been reported following habitat fragmentation in *Swietenia humilis* (White et al. 2002). Third, population thinning due to logging could modify the spatial genetic structure of reproductive adults, potentially leading to an increased proportion of matings among unrelated individuals in the population. Support for this hypothesis comes from studies where self-thinning in tree populations resulted in a reduction of spatial genetic structure from younger to older cohorts (Hamrick et al. 1993).

The goals of the present study were to assess the level of genetic diversity, inbreeding, gene flow and spatial genetic structure in a natural population of the ‘small-insect’-pollinated tree species *C. guianensis* managed for timber harvest. For the purposes of forest management in the Brazilian Amazon, *C. guianensis* is classified as a high-density and fast-growing climax species. Seed progenies were collected over a 22-month period, before and after selective logging occurred, allowing us to test the following broad genetic hypotheses: (i) selective logging may change levels of genetic diversity; (ii) selective logging may change gene flow through pollen and change levels of inbreeding; and (iii) selective logging may change levels of spatial genetic structure among adult trees. The results are discussed with respect to demographical and ecological aspects of *C. guianensis*, the historical context of the study population, and the short-term consequences of logging for the genetic resource base of *C. guianensis*.

### Materials and methods

#### Study species

*Carapa guianensis* Aubl. (Meliaceae) is a widespread Neotropical tree found in the Caribbean islands, Central America and northern South America. The species inhabits upland *terra firme* forests as well as floodplain *varzea* forests in the Amazon basin. In Brazil, it is an important timber tree species, and the oil extracted from the seed is extensively traded and used for its medicinal properties. *C. guianensis* is a monoeccious species, and the small white flowers (3–5 mm diameter) are visited by small insects (e.g. stingless bees, beetles and moths) (D. Cloutier, personal observation). Individual trees above a size of ∼20 cm diameter at breast height (d.b.h.) may be seen with flowers (Dendrogene project, unpublished data) and trees below ∼30 cm d.b.h. typically do not produce fruits in closed undisturbed forests (D. Cloutier, personal observation), while trees above the size of 30 cm d.b.h. may produce no fruits or produce fruits once or several times per year. Both flowers and fruits may be observed simultaneously on the same tree (D. Cloutier, personal observation). At the population level, flowers and fruits are produced year round, with several intrayear peaks of flowering intensity and fruit release (Dendrogene project, unpublished data). During the study period, most fruits were released around the onset (January–February) and the end (May–June) of the wet season, allowing us to sample fruits at each of those times. The fruit of *C. guianensis* contains 4–20 seeds, weighing 20–40 g each, which may be dispersed by water in flood-prone forests. However, in the upland *terra firme* populations, the seeds are dispersed by gravity, and secondarily by medium-sized scatter-hoarding rodents over distances less than 25 m (Guariguata et al. 2002; Cloutier et al. 2005). As a result, seeds and seedlings are typically found directly beneath the maternal parent tree at upland sites.

#### Study site

The study site lies at the Tapajós National Forest (Flona Tapajós), Pará state, Brazil, at 2°51’S, 54°57’W, and 175 m above sea level. The area is a flat plateau covered by a dense *terra firme* forest. The climate is tropical with a mean annual rainfall of 2000 mm (peaks from February to May) and an annual dry season of 2–3 months (August to October). Before the creation of the Flona Tapajós in 1974, the only known recent disturbances were sporadic low-impact activities such as hunting and selective logging of a few tree species, but not of *C. guianensis* (Jennings 1997). Under the auspices of the Dendrogene project (Kanashiro et al. 2002), an intensive study plot (500 ha) is being monitored at Tapajós for the effects of selective logging (reduced impact logging — RIL) on the ecology and genetics of this
and other species. The study species were chosen to have different characteristics such as growth response to light, reproductive ecology, and population demography. This was done in order to place species into groups for management purposes, and to predict (through genetic modelling) impacts on future generations with respect to genetic diversity. Among the species included in the study, *C. guianensis* was classified as a fast-growing climax species, locally abundant relatively to other species, and with reproduction through monoecy and nonsynchronous flowering. Prior to this study, a commercial pre-logging inventory was conducted in the study plot, and every tree larger than 20 cm d.b.h. was identified and mapped.

A mean density of 2.5 *C. guianensis* trees/ha ≥ 30 cm d.b.h. was found in the 400 ha of the central plot area (Fig. 1), indicating that *C. guianensis* has the second highest average tree density at this site. At the end of 2003, the study plot was selectively logged and at least 40 tree species were harvested, *C. guianensis* being one of them. The size below which *C. guianensis* trees were not harvested (cutting size limit) was about 53 cm d.b.h. A total of 257 *C. guianensis* trees were harvested and sent to the sawmill (from the 474

---

**Fig. 1** (A) Spatial distribution of all *Carapa guianensis* adult trees above 30 cm d.b.h. in the central 400 hectares (2000 m × 2000 m) of the study site. Adult trees not sampled for genetic analyses (not genotyped) and those sampled (genotyped) are indicated, respectively, by light grey and dark grey open circles. Filled black circles correspond to the 21 trap trees from which fruits were collected. The size of the circles is proportional to the d.b.h. of the tree. (B) Same as (A) except that the spatial locations of the harvested adult trees at the end of 2003 are indicated by filled light grey circles. (C) Spatial distributions of the juvenile trees (black squares) and the seedlings (black triangles) sampled for genetic analyses.
An important feature of this study is that seed progeny between 53 and 111 cm d.b.h. present prior to logging), i.e. more than 50% of C. guianensis trees of harvestable size were removed, leaving approximately 2.0 trees/ha ≥ 30 cm d.b.h. after logging. The distribution of the harvested and non-harvested C. guianensis trees at the study site is shown in Fig. 1. Trees left standing often had a hollow or poorly formed trunk. It is worth noting that a substantial number of additional trees were likely damaged or killed directly during the logging operations, or indirectly (trees that fell following increased exposure to wind after the logging), but no estimate of this additional damage is available, and therefore the harvested trees in Fig. 1 should be considered as a conservative estimate of the loss of C. guianensis conspecifics.

**Field sampling**

An important feature of this study is that seed progeny were collected from the same C. guianensis trees monitored during 22 months from 2003 to 2005, within which time interval controlled commercial selective logging was conducted at the end of 2003. The samples taken in the field during that time interval can be divided into two types: (i) seeds sampled from 21 reproductive trees monitored before and after logging, and (ii) leaf and cambium sampled before logging from three cohorts of the existing population (e.g. seedling, juvenile and adult).

Seeds that were not previously handled by animals were sampled directly beneath selected seed parents both before and after logging. The seed parents can be described as biological ‘pollen traps’, and are referred to hereafter as ‘trap trees’. We began by regularly visiting a number of potential trap trees, separated by a range of different intertree distances, ranging from trap trees that are close enough to receive pollen from the same source trees, to those far apart and likely to receive pollen from different source trees. Trees that produced at least 10 fruits in both the pre-logging and the post-logging period were selected as trap trees (Fig. 1). The distance among the 21 trap trees varied between 16 m and 1883 m, averaging at 832 m. We sampled seeds from the same 21 unlogged trap trees (before and after logging) aiming to minimize the effects of local spatial variation so as to better examine temporal variation in the population. A separate genetic analysis conducted on C. guianensis multiseeded fruits collected in 2004 at the study site revealed that seeds within individual fruits are significantly more likely to share the same father than are seeds taken from different fruits of the same mother tree (D. Cloutier, unpublished data), suggesting pollen codispersal from the same male parent(s) to single flowers. Thus, to sample as many independent pollination events as possible (a requirement for obtaining good estimates of pollen dispersal; Hardy et al. 2004), we analysed a single seed per fruit sampled. An average of 37 fruits were sampled over the 22-month period from each of the 21 trap trees, varying between 12 and 37 (total of 391) for the pre-harvest period, and between 10 and 33 (total of 390) for the post-harvest period. The seed were sent to the laboratory in closed plastic bags where the small living embryos were extracted for DNA analyses.

Leaf and cambium tissue were sampled in the population to assess levels of genetic diversity and spatial genetic structure in three different cohorts (seedling, juvenile and adult), defined as follows. The adults (n = 199) were reproductive individuals, defined here as trees of size above 30 cm d.b.h. The juveniles (n = 82) were trees between 10 and 30 cm d.b.h., mostly nonreproductive. The seedlings (n = 84) were plants less than 2 years old and less than 1 m in height, and were entirely nonreproductive. The distribution of sampled adults, juveniles and seedlings is shown in Fig. 1. When leaves were accessible, a piece of leaf was taken in the field and dried in silica. Otherwise, a piece of cambium (~150 mg) was extracted and stored in a buffer of 70% ethanol and 0.3% β-mercaptoethanol before being sent to the laboratory. The leaf and cambium samples analysed were randomly selected in the laboratory from the set of field samples available, but since some areas were more intensively sampled than others in the field, the distribution of analysed samples is not uniform (Fig. 1) (e.g. seedlings were sampled along three transects).

**Laboratory analyses**

Samples were brought into the plant genetic laboratory of EMBRAPA-CENARGEN (Brazilian Agricultural Research Corporation — Centre for Genetic Resources and Biotechnology) in Brazilia, Federal District, Brazil. The six microsatellite (SSR) loci used in this study are described by Dayanandan et al. (1999) for loci cg05 and cg07, and by Vinson et al. (2005) for loci cg01, cg06, cg16 and cg17. New reverse primers were designed for cg05 (5’-GAGGAT-CTTGTACGTTGGC-3’) and cg07 (5’-CTGTTCGTTGAAGAACTTGG-3’) in order to obtain shorter polymerase chain reaction (PCR) products. DNA extraction, PCR conditions and electrophoresis were carried out as described in Cloutier et al. (in press). To rule out incorrect assignment of progeny to the mother trees, we compared trap tree and seed genotypes to detect and remove mismatched individuals from the data set (thus 0.9% of seed genotypes were removed).

**Genetic diversity analyses**

Standard genetic diversity parameters were estimated for the seeds collected before and after logging; for the seedling, juvenile and adult cohort; and for the harvested and nonharvested adult trees larger than the size harvest limit. The number of alleles per locus (N_a), the allelic richness...
Multilocus correlation of paternity (estimate of apparent selfing due to biparental inbreeding).

First, we contrasted the seed progenies sampled before and after logging and for linkage disequilibrium among loci were conducted with the program FSTAT version 2.9.3 (Goudet 1995). Exact tests for fixation index estimates \( F_{st} \) (Weir & Cockerham 1984) and for linkage disequilibrium among loci were calculated following the mixed-mating model (Ritland 2002) using the seed together with their known maternal genotypes. If mating among relatives (biparental inbreeding) occurs, \( t_m \) should be higher than \( t_p \), the difference giving a minimum estimate of apparent selfing due to biparental inbreeding. Multilocus correlation of paternity \( r_p \) was calculated to estimate the probability that two outcrossed progenies drawn at random from the same seed family are full-sibs. Mating system analyses were conducted using the program MLTR version 3.1 (Ritland 2002). The default settings were used, except that the parental fixation index was constrained to 0.029 (i.e. the value observed in the adult cohort). We assumed no apomixes and no seed mortality due to inbreeding prior to the assay. Standard errors for \( t_p \), \( t_m \) and \( r_p \) were estimated from 1000 bootstraps at the level of the seed families.

Two different mating system analyses were performed. First, we contrasted the seed progenies sampled before logging \( n = 391 \) vs. those sampled after logging \( n = 390 \), to assess the impact of logging on mating system. Second, we divided the seed progenies into four temporal groups in order to assess whether there was seasonal variation in mating system. The four temporal groups were composed of: (i) seed collected in April–June 2003 \( n = 243 \) and January–February 2004 \( n = 148 \), thus resulting from pre-logging pollination; and (ii) seed collected in May–July 2004 \( n = 158 \) and January–February 2005 \( n = 232 \), thus resulting from post-logging pollination. Assuming that about 4 months separate pollination from fruit release, it is possible that seeds collected from a few trap trees in January–February 2004 were actually resulting from pollination that occurred during the logging because a portion of the study site began to be logged at that period.

Pollen dispersal distance analyses and effective number of pollen donors

Genotyping of every potential pollen donor, as in a paternity analysis, was not practical in this study, since trees were continuously distributed over a very large area. Instead, we analysed pollen dispersal with the twoGener algorithm developed by Smouse et al. (2001), using an implementation of the algorithm programmed by F. Austerlitz (Université Paris-Sud, France). The statistic \( \Phi_{FT} \), which measures the differentiation in allelic frequencies among the pollen clouds sampled by different seed parents (i.e. trap trees, whose spatial coordinates are known), was computed to estimate the pollen dispersal parameters. A formal relationship between \( \Phi_{FT} \) and pollen dispersal parameters has been derived for different pollen flow models (Austerlitz & Smouse 2001). To estimate pollen dispersal parameters either a global \( \Phi_{FT} \), based on average pollen pool differentiation and average physical distance over all sampled mothers, or alternatively, a pairwise \( \Phi_{FT} \) based on pollen pool differentiation and physical distance between each pair of sampled mother can be used — both are employed below, and referred to, respectively, as ‘global’ and ‘pairwise’ estimators.

We estimated the effective number of pollen donors contributing to the progeny of the average mother tree \( N_{e,m} \), which is derived directly from the global estimate of \( \Phi_{FT} \) as \( N_{e,m} = 1/2 \Phi_{FT} \) (Smouse et al. 2001), as well as the parameter \( \delta \) (i.e. the average distance of realized pollen dispersal, which equals \( \sigma(\pi/2)^{0.5} \) for the normal distribution) (Austerlitz & Smouse 2001; Austerlitz & Smouse 2002). The global \( \Phi_{FT} \) (Austerlitz & Smouse 2001), and the pairwise \( \Phi_{FT} \) (Austerlitz & Smouse 2002) were used to estimate \( \delta \) by assuming an underlying density of reproducing trees \( d \) and a pollen dispersal model (i.e. normal, exponential or power exponential). Because \( N_{e,m} / N \) ratios in plant populations may vary from values close to one, to as low as 1/10 (e.g. due to varying fertility and nonsynchronous flowering) (Frankham 1995), we estimated \( \delta \) from global \( \Phi_{FT} \) assuming an upper bound estimate of the density of reproducing trees \( d_{\max} \) (i.e. the global density of conspecific trees above 30 cm d.b.h. in the field), and a lower bound estimate of \( d \) \( d_{\min} = d_{\max}/10 \). The pairwise \( \Phi_{FT} \) was used also to jointly infer \( \delta \) and \( d \). To assess the 95% confidence intervals of the pollen dispersal parameters, we created bootstrap data sets by resampling the 21 trap tree families. Because of the prohibitive amount of computer processing time required for the calculations, we used 200 bootstrap samples (this required 52 Apple Macintosh G4 CPUs over 2 months). While the ‘power exponential’ model has been suggested to approximate the shape of the dispersal curve (Austerlitz et al. 2004), this model repeatedly failed to converge to realistic values using the original and bootstrapped data sets and therefore only results based on the ‘normal’ and ‘exponential’ dispersal model are reported here.

Detection of bottlenecks in pollen pool allele frequencies

To assess the magnitude of logging-mediated paternal bottlenecks, we estimated the extent of differentiation (or
divergence) between the estimates of pollen allele frequencies and of adult tree allele frequencies, using the estimator $F_k$ (Pollak 1983; Waples 1989). This statistic is related to the ‘variance effective population size’ of the pollen donor pool. We estimated, separately, pollen pool allele frequencies before and after logging using mltr version 3.1 (Ritland 2002), while the actual allele frequencies of the 199 adults sampled were used for the adult allele frequencies. A 95% confidence interval on $F_k$ was derived from the chi-square distribution following Waples (1989).

Spatial genetic structure analyses
To take advantage of different approaches for assessing the consequences of historical gene dispersal, spatial genetic structure (SGS) was estimated using two measures of genetic similarity. First, we used a relative kinship coefficient ($F_{ij}$) described in Loiselle et al. (1995). $F_{ij} = (Q_{ij} - Q_{mm})/1 - Q_{mm}$, where $Q_{ij}$ is the probability of identity in state for random genes from individuals $i$ and $j$, and $Q_{mm}$ is the probability of identity in state for genes from random individuals in the population. Second, we used the relationship coefficient $R_{ij}$, computed as the correlation between individual allele frequencies (i.e. Moran’s I autocorrelation statistic), as described in Hardy & Vekemans (1999). We used the program spagedi version 1.2 (Hardy & Vekemans 2002) for all SGS analyses.

To visualize SGS, $F_{ij}$ values were plotted between individuals according to the physical distance separating them, but since using all pairwise $F_{ij}$ values on a graph would obscure the overall trend of the SGS, we presented average pairwise $F_{ij}$ for pairs of individuals falling into a set of specific distance classes. The upper limit of each distance class was 3.6, 7.6, 14.4, 24, 38, 67, 109, 191, 362, 612, 1014, 1536 m for the seedlings; 30, 50, 70, 91, 118, 156, 211, 303, 450, 722, 1213, 2745 m for the juveniles; and 30, 46, 61, 80, 102, 135, 175, 222, 273, 332, 397, 464, 542, 635, 740, 864, 1020, 1220, 1579, 2945 m for the adults. The distance classes were chosen to facilitate the visualization of fine-scale SGS on a logarithmic graph, while ensuring a minimum of 30 pairs in each distance class. To test for significant deviations from random SGS, observed values for each distance class were compared to the 95% confidence interval derived from 1000 permutations of individuals among locations.

To estimate historical gene dispersal from SGS in terms that are equivalent to Wright’s neighbourhood size ($N_b$), we used two approaches. For the first approach, we followed Rousset (2000), who showed that under isolation by distance in a two- dimensional space, $F_{ij}$ is expected to be approximately linearly related to the logarithm of distance between individuals, and therefore the slope of the regression ($b_k$) of $F_{ij}$ values over ln(distance) estimates $1/4\pi D\sigma^2$, where $\sigma^2$ is the second moment of the parent–progeny dispersal distance and $D$ is the effective density of individuals. The product $4\pi D\sigma^2$ is related to Wright’s neighbourhood size, allowing us to estimate $N_b = -(1 - F_{ij}^{\text{mltr}})/b_k$ (applying the correction proposed by Vekemans & Hardy 2004), where $F_{ij}^{\text{mltr}}$ is the average kinship coefficient between neighbouring individuals. For the second approach, we followed Epperson (2005), who described a nearly linear relationship between the natural logarithm of $N_b$ and the finest scale spatial autocorrelation of individual allele frequencies, as measured by Moran’s I relationship coefficient ($R_{ij}$) between neighbours. This allows one to estimate $N_b = \exp\left[0.544 - R_{ij}^{\text{mltr}}/0.102\right]$, where $R_{ij}^{\text{mltr}}$ is the average Moran’s I coefficient between neighbour individuals. Both approaches require one to define what ‘neighbour’ implies, i.e. what is the maximal distance between two ‘neighbouring’ individuals. Following Epperson (2003), we defined this distance relative to the average distance between contiguous individuals, as if they were uniformly distributed on a lattice model, including pairs of individuals separated by rook and bishop moves, and using 1.5 times the square root of the inverse of adult density, which for this C. guianensis population gave as neighbours, individuals separated by less than 95 m. $F_{ij}^{\text{mltr}}$, $R_{ij}^{\text{mltr}}$, and $b_k$ values were tested against the null hypothesis of random distribution of genotypes by randomizing the spatial positions 1000 times. Standard errors for $F_{ij}^{\text{mltr}}$, $R_{ij}^{\text{mltr}}$ and $b_k$ were obtained by jackknifing genetic loci. The 95% confidence interval for $N_b$ was estimated as $-(1 - F_{ij}^{\text{mltr}})/(b_k \pm 2SE_{b_k})$, and for $N_b$ as $\exp\left[0.544 - (R_{ij}^{\text{mltr}} \pm 2SE R_{ij}^{\text{mltr}})/0.102\right]$.

To estimate historical gene dispersal distance ($\sigma$) from SGS, we first estimated $\sigma = (b_k/4\pi D)^{0.5}$, using the census density of adults as an estimate of $D$. But the linear relationship of $F_{ij}$ over ln(distance) holds best within a restricted range between 1 and 20 times $\sigma$ (Rousset 2000), therefore we applied an iterative procedure to estimate $\sigma$ using restricted $b_k$ implemented in spagedi (Hardy & Vekemans 2002).

Results
Genetic diversity and heterozygosity
Allele frequencies for both the adults and the seed sampled at the six microsatellite loci are available in the Supplementary material. The total number of alleles for all samples was 4, 12, 13, 9, 17 and 12, respectively, at the loci cg01, cg06, cg05, cg07, cg16 and cg17. Gene diversity ($H_t$) was high at $-0.8$ for cg06, cg05, cg16 and cg17, while lower $H_t$ was found for cg07 ($-0.6$) and cg01 ($-0.4$). No linkage disequilibrium was detected among the loci.

Genetic diversity statistics averaged over loci for the different samples are shown in Table 1. There were minor differences in the mean number of alleles per locus ($N_{all}$) among samples, apparently depending on the sample size.

© 2006 The Authors
Journal compilation © 2006 Blackwell Publishing Ltd
Table 1 Genetic diversity averaged over six microsatellite loci in cohorts of Carapa guianensis

<table>
<thead>
<tr>
<th>Cohort</th>
<th>N</th>
<th>N_all</th>
<th>R (SE)</th>
<th>H_q (SE)</th>
<th>H_e (SE)</th>
<th>F_is (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-harvest seeds</td>
<td>390</td>
<td>10.0</td>
<td>7.7 (2.8)§</td>
<td>0.68 (0.21)</td>
<td>0.69 (0.19)</td>
<td>0.025 (0.022)*</td>
</tr>
<tr>
<td>Post-harvest seeds</td>
<td>391</td>
<td>10.6</td>
<td>7.6 (2.8)</td>
<td>0.67 (0.18)</td>
<td>0.70 (0.17)</td>
<td>0.049 (0.013)*</td>
</tr>
<tr>
<td>Seedlings</td>
<td>84</td>
<td>8.8</td>
<td>7.6 (2.9)</td>
<td>0.65 (0.19)</td>
<td>0.71 (0.19)</td>
<td>0.079 (0.034)*</td>
</tr>
<tr>
<td>Juveniles</td>
<td>82</td>
<td>9.0</td>
<td>7.9 (3.3)</td>
<td>0.68 (0.17)</td>
<td>0.72 (0.15)</td>
<td>0.048 (0.023)*</td>
</tr>
<tr>
<td>Adults</td>
<td>199</td>
<td>10.3</td>
<td>8.1 (2.9)</td>
<td>0.69 (0.19)</td>
<td>0.71 (0.18)</td>
<td>0.029 (0.020)*</td>
</tr>
<tr>
<td>Harvested†</td>
<td>46</td>
<td>8.3</td>
<td>8.0 (3.2)</td>
<td>0.67 (0.24)</td>
<td>0.71 (0.18)</td>
<td>0.056 (0.042)NS</td>
</tr>
<tr>
<td>Unharvested‡</td>
<td>49</td>
<td>9.3</td>
<td>7.7 (2.9)</td>
<td>0.65 (0.18)</td>
<td>0.70 (0.18)</td>
<td>0.059 (0.033)*</td>
</tr>
</tbody>
</table>

N, sample size; N_all, number of alleles; R, allelic richness based on a minimum sample size of 33 diploid individuals; H_q, observed heterozygosities; H_e, gene diversity; F_is, fixation index (test of significance of the fixation index *: P < 0.05 and NS: P > 0.05).
†Harvested adults > 53 cm d.b.h.
‡Adults > 53 cm d.b.h. left unharvested.
§Standard error over loci in parentheses.

Table 2 Mating system of Carapa guianensis as estimated from seeds collected before and after selective logging

<table>
<thead>
<tr>
<th>Mating system parameter</th>
<th>Before logging</th>
<th>After logging</th>
</tr>
</thead>
<tbody>
<tr>
<td>t_m (SE)</td>
<td>0.933 (0.027)</td>
<td>0.927 (0.033)</td>
</tr>
<tr>
<td>t_s (SE)</td>
<td>0.924 (0.028)</td>
<td>0.899 (0.041)</td>
</tr>
<tr>
<td>t_m−t_s (SE)</td>
<td>0.015 (0.015)</td>
<td>0.028 (0.018)</td>
</tr>
<tr>
<td>r_p (SE)</td>
<td>0.050 (0.014)</td>
<td>0.054 (0.013)</td>
</tr>
</tbody>
</table>

Using the measure of allelic richness (R), there is a modest but nonsignificant trend toward higher R values in oldest cohorts (adults and juveniles), and lower R values in the youngest cohorts (seeds and seedlings). Trees above harvest-size limit that were left unlogged had a slightly lower R than logged trees, and consequently, the post-harvest seeds appeared to have a lower R than pre-harvest seeds; however, in all cases the differences were not significant. Gene diversity (H_e) was similar in all samples indicating that common alleles remain at about the same frequencies regardless of logging or age effects. The observed heterozygosities (H_q) were slightly lower than H_e in all samples, and fixation indexes (F_is) were all low, but significantly greater than zero at the 5% level, except for the harvested trees. There were no significant differences in F_is values between the samples when the 95% confidence intervals derived from a bootstrap among loci were compared.

Mating system

Mating system estimates are shown in Table 2. Outcrossing rates (t_m) were high both before (0.94) and after logging (0.93), but were significantly less than 1.00, suggesting the presence of a small amount of selfing in the population. The level of mating between relatives (biparental inbreeding) was not significant, as measured by the lack of difference between single and multilocus outcrossing rates (t_m−t_s). The correlation of outcrossed paternity (r_p) was low, but significantly different than 0, suggesting that most seed within families had a different father (i.e. half-sibs). Overall, no significant differences in mating system parameters were detected between pre- and post-harvest seed progenies.

Seasonal variation in mating system is summarized in Fig. 2. All parameters t_m, t_m−t_s, and r_p showed little variation over the 22-month period. However, slight nonsignificant changes were measured in seeds collected around January 2004 for r_p [0.112 (0.139)] and May–June 2004 for t_m [0.859 (0.122)]; both samples corresponding to periods where fruit production in the population was relatively low. Overall, no significant temporal variation was detected over the time period investigated despite potential for variation in population phenology, pollen production, and pollinator composition.

Pollen dispersal

Global pollen pool differentiation (Φ_FT) was estimated at 0.044 before logging, with a 95% confidence interval (95% CI) of 0.024−0.063, and at 0.054 after logging, with a 95% CI of 0.033−0.070. These values translate into an effective number of pollen donors (N_q) of 11.3 (95% CI: 7.9−20.8) before logging, and of 9.3 (95% CI: 7.1−15.2) after logging. These differences before vs. after logging are not significant.

Pollen dispersal distance analyses are summarized in Table 3. The minimum (δ_min) and maximum (δ_max) estimates of average distance of pollen dispersal using global Φ_FT were 75 m and 265 m before logging, when both the normal and the exponential model were considered, while the estimates were similar after logging, with a minimum
of 76 m and a maximum of 268 m. Before logging, the joint estimation of \(d\) and \(\delta\) using pairwise \(\Phi_{FT}\) gave an effective density of pollen donors (\(d\)) of 0.30 and 0.25 trees/ha, and an average pollen dispersal distance (\(\delta\)) of 204 and 252 m, for the normal and exponential model, respectively. After logging, the estimates were slightly different, as \(d\) increased to 0.75 and 0.55 trees/ha, and \(\delta\) decreased to 123 m and 159 m, for the normal and exponential model, respectively. The bootstrap 95% confidence intervals on the parameters were large, however, and no statistically significant differences were detected between pre- and post-harvest periods. Minimum and maximum estimates derived from the pairwise \(\Phi_{FT}\) are not shown since they were very similar to the estimates from the global \(\Phi_{FT}\) reported in Table 3.

**Detection of bottlenecks in pollen pool allele frequencies**

The differentiation in allele frequencies (\(F_{ST}\)) between the adult and the pre-harvest seeds was 0.0084, with a 95% confidence interval of (0.0056–0.0118), while the differentiation between adult and post-harvest seeds was 0.0066 with a 95% confidence interval of (0.0044–0.0093). Therefore, no significant differences between before and after logging were detected.

**Spatial genetic structure**

The SGS in the different cohorts is shown in Fig. 3. The estimates of average \(F_{ij}\) decreased with the logarithm of

---

**Table 3** Pollen dispersal analyses based on estimates of pollen pool differentiation before and after selective logging in *Carapa guianensis*

<table>
<thead>
<tr>
<th>Type of estimate</th>
<th>Dispersal model</th>
<th>Parameter estimated</th>
<th>Before logging* (95% CI)</th>
<th>After logging† (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation of (\delta) using global (\Phi_{FT})</td>
<td>Normal</td>
<td>(\delta_{\text{min}})§</td>
<td>75 (63–102)‡</td>
<td>76 (67–98)‡</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\delta_{\text{max}})¶</td>
<td>237 (206–329)</td>
<td>240 (214–311)</td>
</tr>
<tr>
<td></td>
<td>Exponential</td>
<td>(\delta_{\text{min}})</td>
<td>85 (71–115)</td>
<td>86 (75–110)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\delta_{\text{max}})</td>
<td>265 (230–369)</td>
<td>268 (240–348)</td>
</tr>
<tr>
<td>Joint estimation of (d) and (\delta) using pairwise (\Phi_{FT})</td>
<td>Normal</td>
<td>(d)</td>
<td>0.30 (0.06–3.01)</td>
<td>0.75 (0.19–2.08)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\delta)</td>
<td>204 (2–358)</td>
<td>123 (5–230)</td>
</tr>
<tr>
<td></td>
<td>Exponential</td>
<td>(d)</td>
<td>0.25 (0.04–9.36)</td>
<td>0.55 (0.10–581)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\delta)</td>
<td>252 (18–534)</td>
<td>159 (5–306)</td>
</tr>
</tbody>
</table>

\(\Phi_{FT}\), pollen pool differentiation; \(\delta\), average distance of pollen dispersal (m); \(d\), adult effective density (trees/ha).

*Adult field density > 30 cm d.b.h. before logging, 2.5 trees/ha.
†Adult field density > 30 cm d.b.h. after logging, 2.0 trees/ha.
‡Estimated 95% confidence intervals based on bootstrapping the seed families.
§Assuming adult field density (\(d_{\text{max}}\)).
¶Assuming adult field density \(\times 1/10\) (\(d_{\text{min}}\)) (see methods for details).
distance in all cohorts, as expected under a model of isolation by distance. For the seedling cohort, average $F_{ij}$ values at distances from 2 to 60 m departed significantly from the hypothesis of absence of SGS, while few or no values exhibited significant SGS for the adult and juvenile cohorts. Moreover, the average $F_{ij}$ values for the seedling cohort at short distances were higher than observed in the juvenile and adult cohorts. The overall pattern of SGS for pre-harvest and post-harvest adult population was nearly indistinguishable, indicating that adult tree SGS is not significantly impacted by selective logging.

Estimates of SGS and neighbourhood sizes ($Nb$) are presented in Table 4. The global regression slope ($b_k$) of the adult and the seedling cohorts is significantly different from a random spatial structure, confirming the presence of isolation by distance. Significant positive genetic similarity between neighbours separated by less than 95 m was detected in the adult and seedling cohorts, but not in the juveniles. The seedling cohort gave estimates of $Nb$ that are significantly smaller than the adult and juvenile cohorts (Table 4). The estimates of neighbourhood size are slightly different according to the approach used, but agree in indicating large values for $Nb$.

The historical gene dispersal distance estimate was $\sigma = 238$ m, assuming $D = 2.5$ trees/ha and unrestricted $b_k$. Using the iterative procedure that estimates $\sigma$ within a restricted range, we obtained $\sigma = 273$ m. This latter value should be considered as the lower bound estimate for historical gene dispersal distance because $D$ is likely to be lower than the census density of 2.5 individuals/ha (i.e. the iterative procedure applied with $D < 0.8$ led to an estimate of gene dispersal that is infinite).

### Discussion

**Impact of tree harvesting on genetic diversity and heterozygosity**

Following logging in this population, no evidence is seen for immediate losses of genetic variation in either the adult trees or their progeny, or for increases in fixation indexes. This may be partially explained by the fact that *Carapa*
guianensis is relatively common in the forest, that its reproductive phenology is asynchronous, that the population has never been previously logged, and that the proportion of trees harvested during the logging operation was not sufficient to cause a detectable decline in genetic diversity or increase in inbreeding. Perhaps more importantly, we have measured genetic variation using a small number of neutral SSR markers. It is possible that selective logging, because it targets the largest and healthiest trees in the population, also reduced the diversity of genes underlying these characteristics, and is therefore dysgenic, leading to erosion of the species’ genetic resources (e.g. quantitative trait genes for rapid growth, or qualitative trait genes for disease- or drought-resistance, etc.). As has been pointed out elsewhere, even when surveys of neutral genetic diversity indicate the presence of substantial polymorphism, genetic diversity for loci important to long-term survival and reproduction may still be in decline (Lowe et al. 2005).

From earlier studies, we might have expected to detect a decrease in $F_{IS}$ with age, indicating the presence of inbreeding depression. Doligez & Joly (1997) found significant excess homozygosity in seeds and excess heterozygosity in adults in Carapa procera, which they explained as due to selection in favour of heterozygotes. There is no such strong trend in our study, however, as the estimated multilocus $F_{IS}$ was 0.025 for the pre-logging seed and 0.049 for the post-logging seed, values that are similar to the seedling (0.079), juvenile (0.048) and adult (0.029) $F_{IS}$. Lack of evidence for significant inbreeding depression in this population could be explained by a low level of inbreeding, potentially due to the presence of a partial prezygotic incompatibility system (M. Maues, personal communication).

**Impact of tree harvesting on inbreeding, pollen dispersal distance and effective number of pollen donors**

A reduction in density of adult trees following logging could, in theory, change levels of inbreeding and gene flow, but we found no statistically significant differences in the genetic parameters assessed before vs. after the tree harvest. Estimates of the mating system show that there are low levels of inbreeding in this population, and that inbreeding levels remained unchanged after the harvest. The lack of a logging effect on the mating system contrasts with the results of Doligez & Joly (1997) on C. procera, but could be explained by geographical or biological differences among the two populations studied. We found no biparental inbreeding in this population (Table 1), a result that is in accord with the weak spatial genetic structure observed among adult trees (Table 4). Estimates of global pollen pool differentiation revealed that pollen dispersal is extensive, that there are a large effective number of pollen donors, and that pollen dispersal parameters did not appear to be strongly influenced by logging. For instance, the global $F_{PT}$ calculated with two-gener using seeds obtained before logging was 0.044 (0.010), while after logging increased (nonsignificantly) to 0.054 (0.010) (Table 3). Multilocus $r_{PT}$ estimates, when $t_{sp}$ is constrained to 1.0 in mltr, were 0.067 (0.094) before logging and increased to 0.104 (0.161) after logging. Both approaches yield values that point toward a slight, though nonsignificant decrease in the effective number of pollen donors ($N_{ep}$) following logging. On the other hand, using pairwise $\Phi_{PT}$ with two-gener to jointly estimate $d$ and $\delta$, we obtained (for the normal model) $d = 0.3$ trees/ha and $\delta = 204$ m before logging, and $d = 0.75$ trees/ha and $\delta = 123$ m after logging, a difference that suggests a (nonsignificant) increase in the number of pollen donors in a more restricted area following logging. The reduced differentiation in allele frequencies between adult and progeny ($F_{ST}$) observed after logging also points to an increase in the number of pollen donors. A recent study on another insect-pollinated tree species suggested that selective logging may not negatively affect pollen movement, and has fewer genetic consequences relative to clear-cut logging (Sork et al. 2005). Overall, it appears reasonable to conclude that our results do not provide evidence for strong genetic consequences of selective logging on C. guianensis.

There are a number of potential sources of error when estimating pollen dispersal parameters. It is possible that we could have slightly underestimated average pollen dispersal distances, due to the limited number of trap trees available, and the resulting restricted number of pollen dispersal models that could be applied to the data (Austerlitz et al. 2004). Another potential source of error is the value of tree density assumed. In this study, we defined reproductive individuals as trees larger than 30 cm d.b.h. However, it is likely that some trees above this size were also not reproductive, and that some trees below this size did, in fact, contribute to reproduction. Moreover, some reproductive trees are likely to be more important contributors than others due to interplant variation in flower production. The reproductive phenology of C. guianensis further complicates the matter as trees can be seen flowering intermittently. In this study, we tried to take into account these factors by estimating minimum and maximum levels of pollen dispersal, based on maximum and minimum estimates of effective tree density, but it is nevertheless possible that the actual pollen dispersal distance is higher than the estimated intervals.

At least two ecological factors may help account for the stability of gene dispersal and inbreeding in the face of selective logging. First, C. guianensis trees are relatively abundant at the study site and they are reproductively mature at a relatively small d.b.h. compared to other tree species. Thus, as long as the size harvest limit (53 cm d.b.h. in this logging experiment) is substantially higher than the size at which trees become reproductive, the impact on
mating patterns and gene flow may be negligible in the short term. For instance, logging resulted in a modest decrease of tree density from 2.5 trees/ha to 2.0 trees/ha when all trees above 30 cm d.b.h. are considered. A second factor could lie in features of the reproductive biology of C. guianensis that potentially buffer the effect of logging. For instance, despite likely intrayear fluctuations in the number of flowering trees and the efficiency of the animal pollinators, C. guianensis can successfully produce seeds with high levels of genetic diversity at different times within a year. Its reproductive system may be more sensitive to seasonal factors, and relatively less sensitive to logging. This is suggested in Fig. 2, where seasonal variation in mating system, albeit low, appears to be higher than the variation detected as a result of the logging.

Spatial genetic structure and historical gene flow

The level of SGS observed in the adult cohort of C. guianensis is weak, and the different estimates of neighbourhood sizes (Nb) (Table 4) are large, which suggests that gene dispersal was high and that the impact of genetic drift has been historically limited. Moreover, the pattern of SGS for the juvenile and the adult cohorts is similar (Fig. 3), suggesting that gene flow has remained stable in the past. In fact, the SGS in C. guianensis is among the weakest observed to date in plants (Vekemans & Hardy 2004), and is comparable to tropical tree species such as Symphonia globulifera (Dejen et al. 2004), Dicorynia guianensis and Sextonia rubra (Hardy et al. 2006). The existence of high levels of genetic diversity and large Nb values in this undisturbed (and never logged) population at the time of logging may also have helped to buffer negative impacts (Hamrick 2004); it is possible that we would have found different results had the study been conducted in a forest that was previously subjected to one or more cycles of selective logging, as it is often the case for Amazonian forests following road construction in their vicinity. Weak SGS levels were observed overall in the adults, and minor changes in adult SGS caused by artificial thinning are unlikely to have major consequences for mating patterns in this population.

Stronger SGS levels were observed for the seedlings, and the neighbourhood size (Nb) estimates for this cohort are significantly lower than for the juvenile and adult cohorts (Table 4). Similarly, our contemporary estimates of pollen dispersal (Table 3) are apparently lower than the estimated historical level of gene flow. Complete agreement between SGS and twoGEnEr-based estimates of gene dispersal may not be expected, both because the model used to estimate historical dispersal may not be an exact description of the population (Rousset 2000), and because the twoGEnEr estimates may be biased. The fact that the seeds and seedlings examined here were produced over only one or two years, as opposed to many more years for the adults and juveniles, could partially explain the discrepancy. Moreover, density-dependent mortality in the population might explain a large portion of the change in SGS levels among the cohorts. In other studies, natural self-thinning of tree populations was associated with the loss of SGS over time (e.g. Epperson & Alvarez-Buylla 1997). Overall, these results suggest that self-thinning has the potential to erase SGS during the ageing of the different cohorts, but that selective thinning only has a marginal effect on adult SGS in the population studied.

An alternative explanation for the weak SGS observed is that there is overlap of seed shadows, i.e. high levels of seed dispersal. While we have not attempted to directly estimate seed dispersal in this population, several lines of evidence suggest that the contribution of seed dispersal to overall gene flow is negligible. At distances below 20 m, the estimated average kinship coefficient (F0) between seedlings is close to 0.125 (Fig. 3), which is the expected value for maternal half-sib groups. Moreover, the estimates of average F0 between adult-seedling pairs separated by less than 20 m were 0.148 (0.018) (results not shown), which is around the expected genetic similarity for mother–offspring pairs. Of course we cannot rule out that rare long-distance seed dispersal occurs, but this seems unlikely because: (i) the seeds are heavy; (ii) there is no flooding in this area and so seeds cannot be dispersed by water; and (iii) many fruits are seen intact on the forest floor for months, suggesting a lack of animal dispersers. However, the absence of animal dispersers at time of seed collection (2003–2005) may also suggest a recent increase in hunting (facilitated by the construction of the roads to extract the trees a few years before) that may have decimated a significant part of the local animal dispersers’ population at the study site.

Conclusions

When several indicators of genetic diversity are taken into account, this previously undisturbed Carapa guianensis population does not seem to have been strongly impacted by the single episode of selective logging that occurred during the course of this investigation. We have verified that neutral genetic diversity, inbreeding level and gene flow do not measurably change after selective logging of trees in this study, suggesting that pollination of C. guianensis was adequate immediately after the harvest. The relative abundance of C. guianensis at the study site, the small size at which trees become reproductive, and nonsynchronous flowering phenology may potentially explain the lack of a detectable logging effect. However, the high levels of gene flow measured in this study suggest that to conserve present levels of genetic diversity, this species may require large populations that extend beyond the area actually logged at the study site. Over the long run, it remains to be seen
whether selective logging is dysgenic or will cause genetic erosion in the population. It must be emphasized that we monitored this population for only a 2-year period, a short period of time relative to the lifespan of the tree, and possibly insufficient to capture all of the genetic consequences of selective logging for this harvested tree population. Future monitoring of this population should address demographic concerns regarding the establishment and the survival of seedling cohorts in the post-logging forest habitat, as well as possible long-term changes to the mating system that could be a consequence of a more gradual erosion of pollinator density and diversity.

Acknowledgements

We thank IBAMA (Brazilian Institute for Environment and Natural Renewable Resources) for access to study site, and staff at Embrapa in Santarem and Belterra, B Degen and IS Thompson for logistical help. We thank J. Campolina and many others from Belterra and Communidade Franco-1 km 83 for help in sampling fruits, and VCR Azevedo, TN Almeida, VP daSilva, MS Borges and F Saade for help in the lab. This study is part of Dendrogene Project: Genetic Conservation in Managed Forests in Amazônia (www.cpatu.embrapa.br/dendro/index.htm) carried out by Embrapa Amazônia Oriental and its partners, within bilateral cooperation between Brazil and United Kingdom, through Brazilian Cooperation Agency (ABC) and Department for International Development (DFID). This work was carried out with the aid of a grant from the International Development Research Center, Ottawa, Canada (www.idrc.ca). D Schoen acknowledges support from an NSERC Discovery Grant. The Fonds Nature et Technologies, Québec (FQRNT), Natural Sciences and Engineering Council, Canada (NSERC) and the Organization of American States (OAS) provided financial support to D Cloutier.

Supplementary material

The supplementary material is available from http://www.blackwellpublishing.com/products/journals/suppmat/MEC/MEC3193/MEC3193sm.htm

References


© 2006 The Authors
Journal compilation © 2006 Blackwell Publishing Ltd

D. Cloutier is a PhD candidate in biology who is interested in the benefits of genetic variation in trees and medicinal plants for sustainable development. M. Kanashiro is a researcher in forest conservation genetics, coordinator of the Dendrogene Project, and is engaged in the sustainable use and conservation of genetic resources in the Amazon region. A.Y. Ciampi is a researcher in plant genetics, whose research focuses on developing and applying molecular markers in conservation and population genetic studies. D.J. Schoen is a professor of biology, and is interested in the application of theory and molecular data to studies of plant evolution and conservation.