Hardy, O. J., L. Maggia, E. Bandou, P. Breyne, H. Caron, M-H. Chevallier, A. Doligez, C. Dutech, A. Kremer, C. Latouche-Hallé, V. Troispoux, V. Veron, B. Degen (2006). Fine-scale genetic structure and gene dispersal inferences in ten neotropical tree species. *Molecular Ecology* 15: 559-571.
doi: 10.1111/j.1365-294X.2005.02785.x
Publisher version: http://onlinelibrary.wiley.com/doi/10.1111/j.1365-294X.2005.02785.x/full

Fine-scale genetic structure and gene dispersal inferences in ten neotropical tree species

Olivier J. Hardy[*†], Laurent Maggia[‡], Eric Bandou[§], Peter Breyne[¶], Henri Caron[‖], Marie-Hélène Chevallier[**], Agnès Doligez[††], Cyril Dutech[‖], Antoine Kremer[‖], Céline Latouche-Hallé[‖], Valérie Troispoux[‡], Vincent Veron[§], Bernd Degen[§‡‡].

[*] Université Libre de Bruxelles, Service d'Eco-Ethologie Evolutive, CP 160/12, 50 Av. F. Roosevelt, B-1050 Brussels, Belgium

[‡] CIRAD Forestry Department, UMR Ecologie des Forêts de Guyane, Campus agronomique, BP701, 97387 Kourou cedex, France

[§] INRA, UMR Ecologie des Forêts de Guyane, Campus agronomique, BP709, 97387 Kourou cedex, France

[¶] Institute for Forestry and Game management (IBW), Gaverstraat 4, 9500 Geraardsbergen, Belgium

[‖] INRA, UMR Biodiversité, Gènes et Ecosystèmes, 69 Route d'Arcachon, 33612 Cestas Cedex, France

[**] CIRAD Forestry Department, Centre d'Ecologie Fonctionnelle & Evolutive, 1919 route de Mende, 34293 Montpellier cedex 5, France

[††] INRA, UMR Diversité et Génome des Plantes Cultivées, Equipe Génétique Vigne, Bât 6, 2 place Viala, 34060 Montpellier Cedex 01, France

[‡‡] Bundesforschungsanstalt für Forst- und Holzwirtschaft (BFH), Institut für Forstgenetik und Forstpflanzenzüchtung, Sieker Landstrasse 2, D-22927 Grosshansdorf, Germany

Keywords: tropical trees, spatial genetic structure, gene dispersal, seed dispersal, French Guiana

Running title: GENE DISPERSAL IN TROPICAL TREES

ABSTRACT

The extent of gene dispersal is a fundamental factor of the population and evolutionary dynamics of tropical tree species, but directly monitoring seed and pollen movement is a difficult task. However, indirect estimates of historical gene dispersal can be obtained from the fine-scale spatial genetic structure of populations at drift-dispersal equilibrium. Using an approach which is based on the slope of the regression of pairwise kinship coefficients on spatial distance and estimates of the effective population density, we compare indirect gene dispersal estimates of sympatric populations of 10 tropical tree species. We re-analyzed 26 data sets consisting of mapped allozyme, SSR, RAPD or AFLP genotypes from two rainforest sites in French Guiana. Gene dispersal estimates were obtained for at least one marker in each species, although the estimation procedure failed under insufficient marker polymorphism, limited sample size, or inappropriate sampling area. Estimates generally suffered low precision and were affected by assumptions regarding the effective population density. Averaging estimates over data sets, the extent of gene dispersal ranged from 150 m to 1200 m according to species. Smaller gene dispersal estimates were obtained in species with heavy diaspores, which are presumably not well-dispersed, and in populations with high local adult density. We suggest that limited seed dispersal could indirectly limit effective pollen dispersal by creating higher local tree densities, thereby increasing the positive correlation between pollen and seed dispersal distances. We discuss the potential and limitations of our indirect estimation procedure and suggest guidelines for future studies.

INTRODUCTION

Species dispersal is a determining factor affecting the dynamics of populations, communities, and ecosystems. Accordingly, tropical tree seed and pollen dispersal has received much attention in studies of population genetic structure (Sork *et al.* 1999), population demography (Nathan & Muller-Landau 2000), and community assembly (Condit *et al.* 2002).

Neutral genetic markers are increasingly being used to infer patterns of gene dispersal. To this end different methods have been developed which differ by 1) the contemporary or historical nature of gene flow estimates, and 2) their dependence with respect to an underlying model whose assumptions are assumed to hold. These approaches are generally termed 'direct' when using exact genotypes or 'indirect' when using gene pool data (Smouse & Sork 2004; V. Sork, pers. comm.). The most direct approaches use genetic markers to monitor contemporary movements of individuals or propagules by reconstructing parentage relationships, as in paternity/parentage analyses (Jones & Ardren 2003) which are model-free beyond assuming Mendelian inheritance of the markers. The most indirect approaches estimate historical dispersal parameters from the observed genetic structure by fitting it to theoretical population models generally assuming drift-migration equilibrium. Some of these methods exploit coalescence theory (e.g. Beerli & Felsenstein 2001; Hey & Nielsen 2004). Indirect methods to estimate contemporary gene flow, such as 'TwoGener' (Smouse *et al.* 2001; Austerlitz & Smouse 2002), or direct methods based on model fitting, such as the 'neighbourhood model' (Burczyk *et al.* 2002), have also been developed. In the following we focus on indirect approaches of historical gene flow and their application to tropical tree species in continuous populations.

Indirect approaches to assess historical gene flow employ the principle that genetic structure displayed by neutral markers is essentially caused by local genetic drift, the effect of which is counterbalanced by gene dispersal. For an island model, Wright (1965) proposed the famous $F_{ST} = 1/(1 + 4Nm)$ relationship, illustrating that genetic differentiation among populations (F_{ST}) depends on a balance between local genetic drift (N, the effective size of a population) and dispersal (m, the per generation population migration rate). Isolation-by-distance models, which treat space explicitly, make similar predictions: genetic differentiation increases with distance at a rate proportional to $1/D\sigma^2$, where D is the effective population density and σ^2 the second moment of dispersal distance, expressing the spatial extent of gene dispersal (Rousset 1997). Isolation-by-distance models can be used to interpret spatial genetic structure (SGS, i.e. the spatial distribution of alleles) on a fine scale within populations (Rousset 2000; Vekemans & Hardy 2004). Intuitively, one can imagine that the observed level of fine-scale SGS within a population depends on the degree of overlap between 'gene shadows' (i.e. the spatial distribution of seed and pollen mediated gene dispersal events around each parent) of adjacent individuals. The lower the overlap, the greater the probability nearby progenies are sibs. In this way, SGS builds in successive generations until an equilibrium SGS is reached. The product $D\sigma^2$ expresses the degree of individual gene shadow overlap. It implies that a same level of SGS can be obtained under higher density and lower dispersal distance, or vice versa.

A major advantage of indirect methods is that they are relatively easy to apply in natural populations. Unlike parentage or paternity analyses, SGS assessment does not require exhaustive population sampling and is less demanding in terms of marker polymorphism. However, indirect methods do not provide as detailed information as parentage / paternity analyses. For example, they cannot provide detailed descriptions regarding the shape of the

dispersal distribution or make the distinctions between the extent of seed and pollen dispersal, unless both uniparentally and biparentally inherited markers are available (but see Heuertz et al. 2003). Moreover, SGS depends upon the balance between gene dispersal and local genetic drift, so that the estimation of gene dispersal per se requires an independent assessment of local genetic drift through the determination of the effective density. Finally, their reliability relies on whether the assumptions of the underlying models hold in real populations (i.e. demographic equilibrium). Despite many potential limitations of indirect methods, numerical simulations (Hardy 2003; Leblois et al. 2003, 2004) and comparisons between direct and indirect gene dispersal estimates (e.g. Fenster *et al.* 2003) indicate that fine-scale SGS analyses can provide reliable estimates of gene dispersal distances.

Paralleling previous studies comparing among population differentiation of various species using F_{ST} (e.g. Hamrick & Godt 1996), Vekemans and Hardy (2004) proposed a new statistic to quantify fine-scale SGS within populations using pairwise kinship coefficients between individuals. They reanalyzed literature data to compare SGS among 47 plant species and demonstrated that SGS is correlated with the mating system, life-form and population density. Their approach rests on the following prediction of isolation-by-distance models for a two-dimensional space: at drift-dispersal equilibrium and for distances in the range $\sigma_g > d_{ij} >$ ca $20\sigma_g$, where σ_g^2 is half the mean squared parent-offspring distance, the expected kinship coefficient between individuals decreases linearly with $\ln(d_{ij})$ at a rate $b \cong -(1-F_N)/(4\pi D_e \sigma_g^2)$, where D_e is the effective population density (a function of the rate of coalescence per generation and unit area), and F_N the kinship coefficient between neighbouring (competing) individuals (Rousset 2000; Vekemans & Hardy 2004). With mapped genotypes, b and F_N can be estimated using pairwise kinship coefficients. Therefore, the statistic $Sp = b/(F_N-1)$ was proposed as a way to synthesize SGS intensity, because it is expected to be equal to $1/(4\pi D_e \sigma_g^2) = 1/N_b$, the reverse of Wright's neighbourhood size, N_b (Wright 1946), a synthetic measure of the balance between drift and gene flow at a local scale (Fenster *et al.* 2003). This expectation holds when SGS is measured at the appropriate spatial scale, is solely due to isolation-by-distance, and has reached equilibrium. Even when the latter conditions are not met, Vekemans and Hardy (2004) argue that *Sp* remains a useful statistic to make comparisons among species or populations because, as long as the kinship curve is approximately linear with $\ln(d_{ij})$ over the spatial scale investigated, it is less sensitive to the sampling scheme than other synthetic measures of SGS intensity (e.g. distance of zero relatedness).

In their SGS comparisons among species, Vekemans and Hardy (2004) did not attempt to estimate gene dispersal distances (σ_g) from their *Sp* values because reliable effective population density estimates (D_e) were not available for most reviewed SGS studies. However, they proposed a refined approach to estimate gene dispersal distances when D_e can be assessed. In this paper, we apply their refined SGS analysis to estimate past gene dispersal in 10 neotropical tree species from two sites in French Guiana, using all genetic data available for these species and knowledge of adult population densities. Different types of nuclear markers (codominant or dominant) were considered: allozymes, SSR (i.e. microsatellites), RAPD, and AFLP, totalling 26 data sets each representing a different species by site by marker combinations. We check the consistency of our estimates among markers within species and site, and discuss the potential sources of estimation bias and the precautions to be taken when applying the approach. We also compare gene dispersal estimates among species and relate the differences to the respective dispersing agents of seed and pollen as well as to population density.

MATERIALS AND METHODS

The French institutions CIRAD (*French Agricultural Research Centre for International Development)* and INRA (*French National Institute for Agricultural Research*) have undertaken a joint effort to study the genetic structure of 15 neotropical tree species exhibiting contrasting life history traits using a variety of molecular markers. For this study, we compiled the available nuclear genotypic data for 10 species from two sites.

Study sites. Situated 35 km apart and about 15 km from the coast, the Paracou and Counami sites belong to the virtually uninterrupted forest cover of the Guiana shields and are representative of lowland moist tropical forest ecosystems.

Paracou (5°18'N, 52°55'W) contains 16 permanent study square plots (15 of 6.25 ha plus one of 25 ha; Fig. 1A) maintained since 1984 by CIRAD Forestry Department to study forest regeneration in natural conditions and under logging (Gourlet-Fleury *et al.* 2004; Schmitt & Bariteau, 1990). All trees above 10 cm diameter at breast height (DBH) have been identified, mapped and measured in each plot. Although several plots were logged, the SGS examined here refers to trees present prior to logging.

Counami (5°22'N, 53°14'W) is a pristine forest of ca 12000 ha where an inventory of trees (DBH > 37.5 cm) was carried out by CIRAD Forestry Department using 411 small sample plots (40 m by 75 m) situated at the nodes of a rectangular grid (adjacent plots were 500 m apart in one direction and 400 m in the perpendicular direction; Fig.1B) (Couteron *et al.* 2003). Samples for genetic analyses came from these plots (or less than 100 m from a plot), collecting, for each studied species, one individual per plot where the species was present.

Study species. Variation at nuclear markers was assessed for nine species at Paracou and four species at Counami, totalling 10 different species (Table 1). Hereafter, these species will be referred to by genus only. The species exhibit a range of seed dispersal mechanisms (gravity, mammals, birds, bats, or wind, Table 1), so that the extent of seed shadows (i.e. the spatial distribution of dispersed seeds around the maternal tree) is expected to vary substantially. In contrast, pollen dispersal is not well characterized, although the types of pollinating agent can be assessed from floral traits (Table 1). By tropical rainforest standards, most of the studied species may be considered as fairly abundant within the study sites (original adult densities > 1 ha^{-1}; Table 2). At Paracou, we could also estimate the local density of adults, defined here as the mean density within a circle of 50 m radius around each adult tree of the focal species. The ratio of local to overall density ($\Omega_{0\text{-}50}$) expresses the degree of spatial aggregation of each species (Condit *et al.*, 2000), values above unity indicating spatial aggregation. Note that as several plots at Paracou were logged, we computed local and global densities from pre-logging inventories.

Genetic markers. Depending on the species, one to three of the following nuclear genetic markers were used: allozymes, SSR, RAPD, AFLP (Table 2). The first two markers are codominant, whereas the last two are dominant. All samples came from trees with DBH > 10 cm, and consisted of cambium tissue or fresh leaves stored at –80°C until enzyme or DNA extraction. Laboratory protocols are detailed elsewhere (Caron 2000; Degen *et al.* 2001a,b; Doligez & Joly 1997; Dutech *et al.* 2002). Sample sizes varied widely among species and markers (n = 32 to 202; Table 2), and the number of polymorphic loci varied greatly among marker types (Table 2). Within species, different samples of individuals were used for the different markers, although they overlapped to varying degrees. In total, there were 26 species by marker by site combinations (Table 2).

Characterizing spatial genetic structure. SGS was assessed for each data set following the procedure described in Vekemans and Hardy (2004), based on pairwise kinship coefficients between individuals, F_{ij}, using the software SPAGeDi (Hardy & Vekemans 2002; http://www.ulb.ac.be/sciences/ecoevol/spagedi.html). F_{ij} measures the extent of genetic similarity between individuals *i* and *j* relative to the mean genetic similarity between random individuals in the sample. The magnitude of the relatedness is thus relative to the genetic background of the area sampled. F_{ij} values were estimated as in Hardy (2003) using eq. 15 for dominant markers, and eq. 16 for codominant markers. The estimation of F_{ij} with dominant markers requires the inbreeding coefficient, *F* (Hardy 2003). Therefore, for each species, we used *F* estimated on the basis of a codominant marker when one was available, otherwise we set $F = 0$ as the studied species are essentially outbreeders (Caron *et al.* 2004).

To visualize SGS, F_{ij} values were averaged over a set of distance classes (*d*), giving $F(d)$, and plotted against the logarithm of the distance. F_{ij} values were also regressed on $\ln(d_{ij})$, where d_{ij} is the spatial distance between *i* and *j*, to provide the regression slope, *b*. To test for SGS, the spatial positions of the individuals were permuted 10,000 times in order to get the frequency distribution of *b* under the null hypothesis that F_{ij} and d_{ij} were uncorrelated (cf. Mantel test). The SGS intensity was quantified by $Sp = b/(F_1\text{-}1)$, where F_1 is the average kinship coefficient between individuals separated by the first distance class (<50m).

Estimating gene dispersal. Estimating historical gene dispersal from SGS must be distinguished from SGS characterization because, here, we implicitly assume that the observed SGS represent equilibrium isolation-by-distance patterns. We must also make a distinction between the estimation of Wright's neighbourhood size, $N_b \equiv 4\pi D_e \sigma_g^2$, where the relative role of drift and dispersal are not separated, from the estimation of gene dispersal distances, σ_g, which requires an independent estimate of drift (D_e).

Wright's neighbourhood size can be estimated as $\hat{N}_b = -(1-F_1)/b_r$ where b_r is the restricted regression slope of F_{ij} on $\ln(d_{ij})$ in the range $\sigma_g > d_{ij} > 20\sigma_g$. F_1 is here taken as an approximation for F_N. As $F_1 << 1$ in general (e.g. $F_1 \approx 0.1$ in allogamous species; Vekemans & Hardy 2004), the $(1\text{-}F_1)$ term should not cause substantial bias. An obvious difficulty with this formula is that b_r depends on σ_g, which is a parameter to estimate. Therefore, we used an iterative procedure suggested by Vekemans and Hardy (2004): 1) N_b was estimated from the global regression slope of F_{ij} on $\ln(d_{ij})$: $\hat{N}_b = (F_1 - 1)/b$; 2) σ_g was estimated from this N_b estimate knowing D_e : $\hat{\sigma}_g = \sqrt{\hat{N}_b/(4\pi D_e)}$; 3) N_b was re-estimated from the regression slope in the range $\hat{\sigma}_g > d_{ij} > 20\hat{\sigma}_g$: $\hat{N}_b = (F_1 - 1)/b_r$; steps 2) and 3) were repeated until the successive N_b estimates converged. In some cases the procedure failed to converge because b_r became null or positive at one step, or $\hat{\sigma}_g$ became larger than d_{ij} for all *i-j* pairs, in which case no estimate was obtained.

To apply this procedure, D_e is assumed to be known. However, D_e is not simply the adult population density as it also depends on the variance of lifetime reproductive success among individuals (effective density). D_e can be approximated as $D.Ne/N$ where Ne/N is the ratio of the effective over the census population sizes, *D* being the census population density (Vekemans & Hardy 2004). Demographic studies suggest that Ne/N ratio in adult plant populations typically ranges between 0.1 and 0.5 (Frankham 1995). Therefore, for each

species by marker by site combination, we applied the iterative procedure for three assumed D_e values (D/2, D/4, D/10), D being the observed density of trees with a DBH at least equal to the minimum DBH of flowering trees (Tables 1, 2).

For a fixed D_e, approximate lower and upper bounds for the 95% confidence interval (CI) of $\hat{N}_b$ were computed as $(F_1\text{-}1)/(b_r+2SE_b)$ and $(F_1\text{-}1)/(b_r\text{-}2SE_b)$, respectively, SE_b being the standard error of the b_r estimates obtained by jackknifing over loci (when $b_r < 2SE_b$, the upper bound was reported as infinite, ∞). The 95% CI of $\hat{\sigma}_g$ was obtained similarly as $\sqrt{\hat{N}_b/(4\pi D_e)}$ using the upper and lower $\hat{N}_b$ bounds. Note that the confidence intervals refer to the populations sampled and do not constitute adequate confidence intervals at the species level.

The *Sp* and $\hat{\sigma}_g$ estimates were compared among markers within species and site, between sites within species, and among species. In the latter case, we tested whether differences could be explained by seed dispersal vectors or population density. An objective ranking of the species with respect to seed dispersal distance is difficult to achieve, but as propagule weight is likely to be negatively correlated with seed dispersal ability (Hammond *et al.* 1996), we tested the rank correlation between $\hat{\sigma}_g$ and propagule weight.

RESULTS

Performance of the estimation procedure. Non-random SGS was detected in 19 of the 26 species by site by marker combinations and in nine of the 10 species for at least one marker (Table 3). Failure to detect SGS was apparently due to the low information content of some data sets (e.g. low marker polymorphism and few loci for allozymes in *Chrysophyllum*, *Eperua* and *Virola*) but may also be due to weak actual SGS (e.g. *Jacaranda*, *Sextonia*). Most species displayed SGS patterns apparently consistent with isolation-by-distance expectations, where F_{ij} decreases approximately linearly with $\ln(d_{ij})$, although some data sets tend to show an asymptotic behaviour at large distance (Fig. 2). Note that as there is no test to demonstrate that SGS conforms to equilibrium isolation-by-distance patterns, hereafter we assume that they do. The procedure used to estimate dispersal parameters converged in 14 of the 26 species by site by marker combinations whatever the assumed *De*/*D* ratio, and in three additional cases for the highest assumed ratio ($De = D/2$). Thus, at least one estimate was obtained for each species (Table 3). Estimations of neighbourhood size ranged between 20 (*Moronobea*) and ca. 340 (*Jacaranda* and *Sextonia*). Failure of the estimation procedure could be attributed to low marker polymorphism (few loci and low polymorphism per locus; four cases with allozymes), too narrow sampling scale (one case with *Dicorynia*), or weak SGS probably due to extended gene dispersal (two cases with *Sextonia*). Approximate confidence intervals on *Nb* and σ_g estimates were always wide, the upper bound often being undefined, and the lack of precision enforced under weak SGS (Table 3).

Consistency among markers and sites. Judging from three comparisons within species and site, RAPD and AFLP markers gave congruent SGS patterns (similar *Sp*) and, accordingly, *Nb* and σ_g estimates had overlapping CI (Table 3). Allozymes displayed less SGS than other markers. Apart from one species (*Sextonia*), SSR provided SGS patterns diverging from those obtained by RAPD or AFLP, displaying less SGS in *Symphonia*, but more SGS in *Vouacapoua*. However, in the case of the *Vouacapoua* population at Counami,

there was sufficient overlap between the SSR and RAPD samples to subsample the same set of individuals (n = 154), in which case the SGS patterns for the two markers became highly congruent ($Sp \pm SE = 0.0114 \pm 0.0067$ with SSR, and 0.0092 ± 0.0027 with RAPD).

Comparisons between the two sites can be made for three species. *Sextonia* gave the lowest *Sp* values at both sites (Table 3). According to RAPD markers, *Dicorynia* showed more SGS at Counami ($Sp \pm SE = 0.019 \pm 0.003$) than at Paracou ($Sp \pm SE = 0.011 \pm 0.002$), whereas the reverse occurred in *Vouacapoua*, both with RAPD and SSR (Table 3). Overall densities do not seem to explain these differences: for *Dicorynia* there is no difference between sites and, for *Voucapoua*, density is lower at Counami than at Paracou (Table 2) while stronger SGS would be expected for a lower density population (Vekemans & Hardy 2004). Note that the sites might differ in terms of local densities (not recorded at Counami).

Differences among species. Despite the variability of estimates within species for different markers or sites, and the wide confidence intervals, interpretation of differences among species can be attempted using average estimates over markers and sites, assuming a fixed *De* (*De*=*D*/4). Stronger SGS and lower σ_g estimates were typically obtained in species where seed dispersal is ensured by gravity and scatter-hoarding rodents, such as *Carapa*, *Moronobea* or *Vouacapoua* ($\sigma_g \leq 200$ m), whereas the reverse occurred in species where birds or monkeys are likely to ensure greater seed dispersal, such as in *Chrysophyllum*, *Sextonia* or *Virola* ($\sigma_g >$ 400 m; Fig. 2; Table 4). *Eperua*, which should undergo very limited seed dispersal, and *Symphonia*, which can be dispersed by bats, seem to deviate from this pattern. There are two wind dispersed species and one of them, *Dicorynia*, has heavy propagules whereas the other, *Jacaranda*, has much lighter ones (Table 1) that are widely dispersed (Jones *et al*., in press). Accordingly, *Dicorynia* displayed more SGS leading to a lower σ_g estimate (σ_g = 1180 m and 203 m for *Jacaranda* and *Dicorynia*, respectively). It must be noted that σ_g (gene dispersal) should be contributed by seed and pollen dispersal, so that a tight correlation with seed dispersal alone is not necessarily expected.

When species were ranked according to the dry mass of their propagules (Table 1), grouping species with very similar masses (*Chrysophyllum*, *Sextonia* and *Virola* in one class; *Carapa*, *Moronobea* and *Symphonia* in another class), the species rank was significantly correlated with the mean *Sp* statistic over markers (excluding allozymes) and sites (Spearman R = 0.68, P = 0.038), as well as with the mean σ_g estimates over the data sets where the iterative procedure succeeded using *De* = *D*/4 (Spearman R = -0.72, P = 0.018). The σ_g estimates were also correlated with the overall adult density (Spearman R = -0.85; P = 0.002), and the local adult density (Spearman R = -0.83; P = 0.003). As the local adult density was correlated with the propagule mass (Spearman R = 0.69; P = 0.027), the impact of propagule mass and local adult densities are partially confounded.

DISCUSSION

Our data analysis has two main advantages compared to most previous studies of SGS in tropical plants: i) it allows a comparison of SGS using the same metric (i.e. by pairwise kinship coefficients) for dominant and codominant markers, and ii) it attempts to estimate historical gene dispersal distances by fitting the observed SGS patterns to an isolation-by-distance model. Our allozyme data were usually inefficient for that purpose (note that we had generally few allozyme loci), but RAPD, AFLP and SSR data provided estimates in most cases, although these estimates generally suffered low precision. It is worth noting that the

theory on which our gene dispersal estimates are based holds whatever the probabilistic form of the gene dispersal curve. However, when rare long distance gene dispersal events occur relative to the scale of observation (highly leptokurtic dispersal), σ_g would correspond to a truncated gene dispersal distribution, not accounting for dispersal events beyond the sampled area (Rousset 2001).

In the following we first discuss methodological aspects to assess the reliability of SGS-based gene dispersal inferences. Then we discuss the implications of our results for understanding of gene dispersal in tropical trees.

Consistency of SGS patterns among markers and sites. Different markers sometimes provided contrasting *Sp*, N_b and σ_g estimates (i.e. with non-overlapping confidence intervals, Table 3) for the same species in the same site. Marker-dependent bias can occur for a number of reasons. However, there was no systematic trend between the extent of SGS displayed by different types of markers, suggesting that the most likely source of discrepancies between markers was sampling. This hypothesis is supported by the *Vouacapoua* population at Counami where SSR and RAPD showed very congruent patterns once the analyses were performed on the same set of individuals.

Among the three species analyzed at both sites, *Sextonia* consistently showed weak SGS, whereas *Dicorynia*, and more particularly *Vouacapoua*, displayed different levels of SGS at the two sites. The differences could be due to a site effect, but also to a sampling effect: the sampling schemes adopted at Paracou covered small distances well but missed large ones (>3 km), whereas the one adopted at Counami missed small distances (< 400 m). Divergent SGS between sites can arise from ecological factors (e.g. differences in the efficiency of seed or pollen dispersing agents, different population densities) but also from historical factors when SGS is not at drift-dispersal equilibrium. Historical factors and mode of recolonization were invoked to explain diverging *Vouacapoua* SGS between two sites sampled at the same scale (Dutech *et al.* 2002).

Reliability of dispersal parameters estimated from SGS. A critical assumption in SGS-based N_b and σ_g estimation is that SGS is representative of an equilibrium isolation-by-distance pattern, and as such some authors have warned that gene dispersal inferences based on the $F_{ST} = 1/(1 + 4Nm)$ relationship are unreliable (Whitlock & McCauley 1999). This is certainly true when SGS is assessed at a scale much larger than that of gene flow, but at a local scale (relative to σ_g), SGS can reach an equilibrium in a few generations (Hardy & Vekemans 1999), so that the equilibrium hypothesis is less critical. Simulations by Leblois *et al*. (2004) confirm that indirect neighbourhood size estimates obtained within continuous populations are fairly robust with respect to temporal and spatial variation of demographic parameters, and they are generally little affected by historical events having occurred more than 10-20 generations ago. Future comparisons between direct and indirect estimates of gene dispersal in undisturbed tree populations should provide additional hints on whether the equilibrium hypothesis is reasonable.

The neighbourhood size, here defined as $N_b = 4\pi.De.\sigma_g^2$, does not disentangle the impact of drift and dispersal on SGS but rather expresses a balance between them. To extract σ_g, *De* must be assessed. Under uniform density, simulations show that $De \approx D.Ne/N$ where *D* is the census population density (O. J. Hardy, unpublished). Note that the relevant *Ne*/*N* ratio here refers to the level of drift occurring within population over a timescale sufficient for the SGS to build up. It is not an asymptotic (long term) *Ne*/*N* ratio at the species level, which can be

extremely small. Under stable population size (density), *Ne*/*N* can be predicted from the variance in lifetime reproductive success among individuals (*V*): $Ne/N \approx 4/(2+V)$ (Kimura & Crow 1963). With overlapping generations, under the conservative assumption that the lifetime reproductive success is proportional to the adult lifetime, $Ne/N = 1/2$ at the limit for low mortality rates between breeding seasons (Nunney 1991). In species with separate sexes, such as *Virola*, *Ne*/*N* is expected to be further reduced if the sex ratio is unbalanced (Nunney 1993). Hence, $De=D/2$ should be a maximum estimate for long-lived hermaphrodite tree species. From these theoretical considerations and experimental results (Frankham 1995), our guess is that *De* should often be closer to *D*/4, notably because trees must generally reach a larger size than the minimum DBH of flowering individuals (the criterion used to assess *D*) to flower regularly and contribute significantly to the next generation. The parameter *V* is difficult to assess in natural populations because tree lifespan is often much longer than that of field biologists. Efforts at estimating *V,* and thus *De,* in tropical trees should constitute a priority in future studies, not only to improve SGS-based σ_g estimation, but also because it would bring many insights into patterns of genetic variation and the dynamics of populations. Because the estimation of σ_g depends on the square root of *De*, the error made on the assumed *De* is somewhat tempered. Assuming $De = D/4$ for all species, our gene dispersal distances estimates (σ_g) were in the order of 150-1200 m (Table 4). These estimates would be multiplied by ca. 1.6 if *De* was actually closer to *D*/10 (Table 3).

The goal of the iterative procedure used to estimate σ_g is to reduce estimation bias by measuring the slope of the kinship curve over a distance range for which theory provides robust expectations. An interesting corollary is that the procedure may fail to converge, as observed in 12 of the 26 data sets analysed when $De = D/10$. Non-convergence is expected, for example in populations sampled at a scale too small with respect to pollen dispersal but where limited seed dispersal generates SGS. In such a situation, the kinship curve is asymptotic, its steepness decreases with distance (Heuertz *et al.* 2003). If σ_g is estimated considering the whole kinship curve, a biased estimate will be obtained. Applying the iterative procedure, σ_g estimates would increase at each iteration until estimates exceed the sampling scale (no convergence). Hence, by failing to converge, the iterative procedure constitutes a way to eliminate data sets not sampled at the appropriate spatial scale, for example the SSR data of *Dicorynia* in Paracou (Table 3).

Besides accuracy (estimation bias), the precision of an estimation procedure is also important. The approximate confidence intervals obtained by jackknifing over loci show that N_b and σ_g estimates suffer low precision for most of our data sets, especially when N_b was high (lack of upper bound). Low precision is partially due to the fact that the SGS expressed by a single allele represents just one realization of a highly stochastic process. When a population is sampled at the appropriate sampling scale (σ_g - $20\sigma_g$), simulations by Hardy (2003) indicate that the variance of $1/N_b$ estimates ($\approx Sp$) is roughly proportional to the total number of alleles minus the number of loci (the standard error and the width of the confidence interval should thus depend on the square root of this number). Because N_b and σ_g estimators are non-linear functions of *Sp*, their standard errors also depend on the level of SGS. Therefore, to obtain sufficiently precise N_b estimates, with an upper bound of a 95% confidence interval not larger than twice the lower bound, the standard error of the slope (b_r) of the kinship curve should be less than 1/6 of the slope itself. Lack of precision is somewhat tempered for σ_g estimates which depend on the square root of N_b estimates (Table 3), but this advantage is essentially annihilated by the additional imprecision on the *De* estimate. The spatial distribution of sampled individuals also matters, and higher precision might be

expected when pairwise distances between samples are homogeneously distributed within the σ_g - $20\sigma_g$ distance range, but optimal sampling strategies deserve to be investigated further.

Comparisons with direct estimates of seed or pollen dispersal. In a two-dimensional space, $\sigma_g^2 = \sigma_s^2 + ½\sigma_p^2$ (Crawford 1984), where σ_s and σ_p refer to effective seed and pollen dispersal, respectively. Using this formula, indirect gene dispersal estimates could be compared to direct estimates of seed and pollen dispersal (e.g. Fenster et al. 2003). Although a true validation of our indirect estimates cannot be achieved because direct estimates of both seed and pollen are lacking for the studied species, our indirect estimates can be compared with direct seed dispersal estimates in six of our species, and with direct estimates of pollen dispersal in other tropical tree species. Reanalyzing data sets monitoring the movement of marked seeds or the distribution of juvenile plants around isolated trees assumed to be the maternal ones (data obtained in French Guiana from the theses of P.-M. Forget, P. A. Jansen and S. Traissac, or in Panama for *Jacaranda*), direct σ_s estimates were much lower than our σ_g estimates (for $De = D/4$), suggesting that pollen dispersal is the major vector of gene flow (Table 4). However, a reanalysis of data sets from published paternity analyses in tropical trees indicate that the apparent contribution by pollen to σ_g ranges from 20 m to 250 m (but exceed 500 m under very low densities in fragmented habitats; Table 5), which is still lower than the range of our SGS-based σ_g estimates (Table 4). The discrepancy might be due to an underestimation of pollen dispersal distances because paternity analyses often fail to describe accurately long dispersal events (the same problem occurs for seed dispersal), or simply because different set of species were considered.

If pollen dispersed much better than seeds in our species ($\sigma_p >> \sigma_s$), as has been suggested for tropical trees (Loveless & Hamrick 1984), gene dispersal would depend essentially on pollen movements and no correlation between seed dispersal efficiency and σ_g would be expected, except if seed and pollen dispersal distances are positively correlated. As species with limited seed dispersal also tended to occur at higher local densities, and pollen probably disperses less under higher density (Stacy *et al.* 1996; Vekemans & Hardy 2004), limited seed and pollen dispersal might be positively correlated in our group of species. This would explain the observed correlations of our σ_g estimates with i) factors closely linked to seed dispersal ability (propagule mass or dispersal vector), and ii) local density (negative correlation). We suggest that this phenomenon, whereby limited seed dispersal favours limited pollen dispersal by causing individuals to be more aggregated, might be common among tropical forest trees because tree aggregation is generally associated with limited seed dispersal (Condit *et al.* 2000). The correlation between the extent of seed and pollen dispersal should be stronger when most mating events occur between near neighbours.

Guidelines to apply the estimation method proposed. The sampling scale is very important to consider at the beginning of such studies. The σ_g estimates obtained here for mixed tropical forests indicate that an appropriate sampling scheme for trees should include pairs of individuals separated by distances ranging from ca. 10-50 m (neighbours) to ca. 10-20 km (20 times our large σ_g estimates, 500-1000 m). Having a wide distance range also allows to evaluate the linearity of the kinship curve. Vekemans and Hardy (2004) suggest that sampling along several long intersecting transects can be a good strategy to gather many pairs of neighbours while covering a large area. The precision of estimates improves substantially under low neighbourhood size. Therefore, the approach should perform best under low population densities, and may be inappropriate for species with very extensive pollen or seed dispersal, or in high density populations. The low precision obtained for most of the data sets analysed here suggests that sample sizes and/or marker polymorphism were generally too

low. Half of our data sets contained less than 70 individuals, which may be sufficient only for small neighbourhood size (≤50) when markers are very informative (many loci like AFLP and/or many alleles per locus like microsatellites). In general, larger samples (200 individuals or more) are likely necessary. The informative content of the markers is crucial and can be evaluated from the total number of alleles minus the number of loci (the difference equals the number of polymorphic loci in the case of dominant markers). Our data sets for which this difference was inferior to 30 (allozymes) always failed to converge. Hence, higher values (100 or more) are advisable. For σ_g inference, assumptions regarding *De*/*D* is critical. Our guess is that *De* should be in the order of *D*/4, but only new detailed parentage analyses assessing how the reproductive success varies among individuals and through time will provide reliable estimates of the range of *De*/*D* ratios for tropical trees.

CONCLUSION

SGS-based gene dispersal estimates are based upon some critical assumptions (e.g. drift-dispersal equilibrium, isotropic dispersal) and may suffer limited precision. However, they have the great advantage of being relatively easy to carry out, requiring only identification of an appropriate spatial scale for sampling and sufficiently polymorphic neutral markers for genotyping. We believe that SGS-based inferences are likely to offer a useful first approach to assessing the extent of historical gene dispersal in tropical tree species and may be used to correlate estimates with life-history traits. Although the estimates provided here for each species should be taken with caution given their low precision, we suggest that the ranges of values observed for the neighbourhood size (N_b = 20 to 350) and gene dispersal distances (σ_g = 150 to 1200 m) are typical for trees in mixed tropical forests. Gene dispersal estimates were smaller (1) in species with presumably not well dispersed diaspores, and (2) in populations with high local adult density. To explain these results, we suggest that limited seed dispersal could indirectly limit effective pollen dispersal by favoring the aggregation of trees (higher local density). This mechanism would enhance the correlation between gene and seed dispersal, despite pollen is likely the main gene dispersal vector in most species. In future research, comparisons between gene dispersal estimates based on SGS and on paternity / parentage analyses or the TwoGener approach will be important, firstly, to assess more precisely the accuracy and reliability of SGS-based estimates, and secondly, to assess the impact of human perturbations in previously stable habitats by comparing historical (SGS-based) and contemporary (paternity / parentage analyses, TwoGener) gene dispersal patterns (e.g. Dutech *et al*. 2005).

REFERENCES

Austerlitz F, Smouse PE (2002) Two-generation analysis of pollen flow across a landscape. IV. Estimating the dispersal parameter. *Genetics* **161**, 355-363.

Beerli P, Felsenstein J (2001) Maximum-likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences USA* **98**, 4563-4568.

Boshier DH, Chase MR, Bawa K (1995) Population genetics of *Cordia alliodora* (Boraginaceae), a neotropical tree. 3. Gene flow, neighborhood and population substructure. *American Journal of Botany* **82**, 484-490.

Burczyk J, Adams WT, Moran GF, Griffin AR (2002) Complex patterns of mating revealed in a Eucalyptus regnans seed orchard using allozyme markers and the neighborhood model. *Molecular Ecology* **11**, 2379-2391.

Caron H (2000) *Organisation et dynamique de la diversité génétique de cinq espèces arborées de la forêt guyanaise.*, Université de Montpellier II.
Caron H, Dutech C, Bandou E (2004) Reproductive phenology and mating system of six tree species in Paracou stand. In: *Ecology and Management of a Neotropical Rainforest. Lessons drawn from Paracou, a long-term experimental research site in French Guiana* (ed. Laroussinie O). Elsevier.
Chase MR, Moller C, Kesseli R, Bawa KS (1996) Distant gene flow in tropical trees. *Nature* **383**, 398-399.
Condit R, Ashton PS, Baker P, *et al.* (2000) Spatial patterns in the distribution of tropical tree species. *Science* **288**, 1414-1418.
Condit R, Pitman N, Leigh EG, *et al.* (2002) Beta-diversity in tropical forest trees. *Science* **295**, 666-669.
Couteron P, Pélissier R, Mapaga D, Molino J-F, Teillier L (2003) Drawing ecological insights from a management-oriented forest inventory in French Guiana. *Forest Ecology and Management* **172**, 89-108.
Crawford TJ (1984) What is a population? In: *Evolutionary Ecology, 23rd Symposium of the British Ecological Society* (ed. Shorrocks B), Leeds.
Dawson IK, Waugh R, Simons AJ, Powell W (1997) Simple sequence repeats provide a direct estimate of pollen-mediated gene dispersal in the tropical tree *Gliricidia sepium*. *Molecular Ecology* **6**, 179-183.
Degen B, Caron H, Bandou E, *et al.* (2001a) Small scale spatial genetic structure of six tropical tree species in French Guiana. In: *Genetic response of forest systems to changing environmental conditions* (ed. Schubert R), pp. 75-92. Kluwer Academic Publishers, Dordrecht.
Degen B, Caron H, Bandou E, *et al.* (2001b) Fine-scale spatial genetic structure of eight tropical species as analysed by RAPDs. *Heredity* **87**, 497-507.
Dick CW, Etchelecu G, Austerlitz F (2003) Pollen dispersal of tropical trees (*Dinizia excelsa*: Fabaceae) by native insects and African honeybeees in pristine and fragmentaed Amazonian rainforest. *Molecular Ecology* **12**, 753-764.
Doligez A, Joly HI (1997) Genetic diversity and spatial structure within a natural stand of a tropical forest tree species, *Carapa procera* (Meliaceae), in French Guiana. *Heredity* **79**, 72-82.
Dutech C, Seiter J, Petronelli P, Joly HI, Jarne P (2002) Evidence of low gene flow in a neotropical clustered tree species in two rainforest stands of French Guiana. *Molecular Ecology* **11**, 725-738.
Dutech C, Sork VL, Irwin AJ, Smouse PE, Davis FW (2005) Gene flow and fine-scale genetic structure in a wind-pollinated tree species, *Quercus lobata* (Fagaceaee). *American Journal of Botany* **92**, 252-261.
Fenster CB, Vekemans X, Hardy OJ (2003) Quantifying gene flow from spatial genetic structure data in a metapopulation of *Chamaecrista fasciculata* (leguminosae). *Evolution* **57**, 995-1007.
Forget P-M (1988) *Dissémination et régénération naturelle de huit espèces d'arbres en forêt guyanaise.*, Université de Paris VI.
Frankham R (1995) Effective population size/adult population size ratios in wildlife: a review. *Genet. Res., Camb.* **66**, 95-107.
Gourlet-Fleury S, Guehl J-M, Laroussinie O (2004) *Ecology and Management of a Neotropical Rainforest. Lessons drawn from Paracou, a long-term experimental research site in French Guiana* Elsevier.
Hammond DS, Gourlet-Fleury S, Hout Pvd, Steege Ht, Brown VK (1996) A compilation of known Guianan timber trees and the significance of their dispersal mode, seed size

and taxonomic affinity to tropical rain forest management. *Forest Ecology and Management* **83**, 99-116.
Hamrick JL, Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London B*, **351**, 1291-1298.
Hamrick JL, Murawski DA (1990) The breeding structure of tropical tree populations. *Plant Species Biology* **5**, 157-165.
Hardy OJ (2003) Estimation of pairwise relatedness between individuals and characterization of isolation-by-distance processes using dominant genetic markers. *Molecular Ecology* **12**, 1577 –1588.
Hardy OJ, Vekemans X (1999) Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity* **83**, 145-154.
Hardy OJ, Vekemans X (2002) SPAGeDi : a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* **2**, 618-620.
Heuertz M, Vekemans X, Hausman J-F, Palada M, Hardy OJ (2003) Estimating seed versus pollen dispersal from spatial genetic structure in the common ash. *Molecular Ecology* **12**, 2483-2495.
Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **167**, 747-760.
Jansen PA, Bartholomeus M, Bongers F, *et al.* (2002) The role of seed size in dispersal by a scatter-hoarding rodent. In: *Seed dispersal and frugivory: ecology, evolution and conservation* (eds. Levey DJ, Silva WR, Galetti M), pp. 209-226. CABI Publishing, Oxon.
Jones AG, Ardren WR (2003) Methods of parentage analysis in natural populations. *Molecular Ecology* **12**, 1-15.
Jones FA., Chen J, Weng G-J, Hubbell SP (in press) A genetic evaluation of seed dispersal in the neotropical tree, *Jacaranda copaia* (Bignoniaceae). *The American Naturalist*.
Kimura M, Crow JF (1963) The measurement of effective population number. *Evolution* **17**, 279-288.
Konuma A, Tsumura Y, Lee CT, Lee SL, Okuda T (2000) Estimation of gene flow in the tropical-rainforest tree *Neobalanocarpus heimii* (Dipterocarpaceae), inferred from paternity analysis. *Molecular Ecology* **9**, 1843-1852.
Leblois R, Estoup A, Rousset F (2003) Influence of mutational and sampling factors on the estimation of demographic parameters in a "continuous "population under isolation by distance. *Molecular Biology and Evolution* **20**, 491-502.
Leblois R, Rousset F, Estoup A (2004) Influence of spatial and temporal heterogeneities on the estimation of demographic parameters in a continuous population using individual microsatellite data. *Genetics* **166:** 1081–1092.
Loveless MD, Hamrick JL (1984) Ecological determinants of genetic structure in plant populations. *Annual Review of Ecology and Systematics* **15**, 65-95.
Nason JD, Herre EA, Hamrick JL (1998) The breeding structure of a tropical keystone plant resource. *Nature* **391**, 685-687.
Nathan R, Muller-Landau HC (2000) Spatial patterns of seed dispersal, their determinants and consequences for recruitment. *Trends in Ecology and Evolution* **15**, 278-285.
Nunney L (1991) The influence of age structure and fecundity on effective population size. *Proceedings of the Royal Society of London Series B-Biological Sciences* **246**, 71-76.

Nunney L (1993) The influence of mating system and overlapping generations on effective population size. *Evolution* **47**, 1329-1341.
Rousset F (2000) Genetic differentiation between individuals. *J. Evol. Biol.* **13**, 58-62.
Rousset F (2001) Genetic approaches to the estimation of dispersal rates. In: *Dispersal* (ed. Nichols JD), pp. 18-28. Oxford University Press, Oxford.
Schmitt L, Bariteau M (1990) Gestion de l'écosystème forestier guyanais. Etude de la croissance et de la régénération naturelle. Dispositif de Paracou. *Bois et Forêts des Tropiques* **220**, 3-23.
Smouse PE, Dyer RJ, Westfall RD, Sork VL (2001) Two-generation analysis of pollen flow across a landscape. I. Male gamete heterogeneity among females. *Evolution* **55**, 260-271.
Smouse PE, Sork VL (2004) Measuring pollen flow in forest trees: an exposition of alternative approaches. *Forest Ecology and Management* **197**, 21–38.
Sork VL, Nason J, Campbell DR, Fernandez JF (1999) Landscape approaches to historical and contemporary gene flow in plants. *Trends in Ecology and Evolution* **14**, 219-224.
Stacy EA, Hamrick JL, Nason JD, *et al.* (1996) Pollen dispersal in low-density populations of three neotropical tree species. *American Naturalist* **148**, 275-298.
Traissac S (2003) *Dynamique spatiale de Vouacapoua americana (Aublet), arbre de forêt tropicale humide à répartition aggrégée*, Université Claude Bernard - Lyon I.
Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology* **13**, 921-935.
White GM (2002) Increased pollen flow counteracts fragmentation in a tropical dry forest: an example from *Swietenia humilis* Zuccarini. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 2038-2042.
Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: Fst!=1/(4Nm+1). *Heredity* **82**, 117-125.
Wright S (1943) Isolation by distance. *Genetics* **28**, 114-138.
Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* **19**, 395-420.

ACKNOWLEDGMENTS

We thank Pierre Couteron and Frank A. Jones for sharing data as well as Christopher Baraloto, David Boshier, Stephen Cavers, Frank A. Jones, Victoria Sork and three anonymous reviewers for their helpful comments and suggestions. This study was financially supported by FEDER (Fonds européen de développement régional) through the XIth and XIIth Contrat de Plan Etat-Région (CPER), by the European Union (program INCO: TS3-CT94-0316, and STD3 project), by the BRG (Bureau des Ressources Génétiques), and by the French Ministry of Environment (SOFT-project 1995). O. J. H. is a Research Associate from the Belgian National Fund for Scientific Research (FNRS). The theses of A. D. and C. D. were financed by CIRAD and MRT.

Figure legends.

Fig. 1. Maps of the study sites showing the sampling areas (gray areas) at A) Paracou, and B) Counami. At Counami, the sampling areas are not to scale as their actual size is 40m by 75m. Note that the scales of the two maps differ markedly.

Fig. 2. Comparison of SGS among species in the sites of A) Paracou and B) Counami, selecting one marker per species (the most informative one). The type of symbol provides indications on the seed dispersal vector: gravity and/or scatter-hoarding rodents (filled symbols); bats, birds or arboreal mammals (open symbols); wind (crosses or stars). Species are abbreviated by the four first letters of their genus and species names.

Table 1. Studied tree species, some of their life history traits related to reproductive strategies, and their minimal size as adult (based on field observation of flowering trees).

Species (Family)	Sex	Pollinators	Seed dispersers	Dry diaspore mass (g)*	Growth strategy	Min DBH adults (cm)
Carapa procera Aubl. (Meliaceae)	M	I	SR	6.58	H	17
Chrysophyllum sanguinolentum L. (Sapotaceae)	H	Ba, I	M	1.35	T	25
Dicorynia guianensis Hamshoff (Caesalpiniaceae)	H	I	SR, W	1.07	HT	22
Eperua grandiflora Aubl. (Caesalpiniaceae)	H	I	G	27.61	HT	28
Jacaranda copaia **(**Aubl.) D. Don (Bignoniaceae)	H	I	W	0.047	P	20
Moronobea coccinea Aubl. (Clusiaceae)	H	Bi	SR	6.86	HT	20
Sextonia rubra (Mez) van der Werff (Lauraceae)	H	I	Bi	1.17	T	29
Symphonia globulifera L. (Clusiaceae)	H	Bi	AM, Ba, SR	6.64	H	17
Virola michelii Aubl. (Myristicaceae)	D	I	AM, Bi, M	1.25	H	24
Vouacapoua americana Aubl. (Caesalpiniaceae)	H	I	SR	12.39	T	20

Sex: M, monoecious; H, hermaphrodite flowers; D, dioecious.

Pollinators: I, insects; Ba, bats; Bi, birds.

Seed dispersers: AM, nocturnal arboreal mammal; Ba, bats; Bi, birds; G, gravity only; M, monkeys; SR, scatter-hoarding rodents; W, wind.

Growth strategy: P, pioneer; H, heliophilic; T, shade tolerant; HT, hemi-tolerant.

* C. Baraloto (pers. comm.)

Table 2. Adult density, sample size, number of polymorphic loci and total number of alleles (for allozymes and SSR) per marker for each species and site. See Table 1 for complete species names.

	Density	Sample size (no. of polymorphic loci; total no. of alleles)			
	(N/ha)	Allozyme	SSR	RAPD	AFLP
Paracou					
Carapa	4.13	131 (10; 22)		65 (46)	
Chrysophyllum	1.84	54 (4; 13)		68 (48)	41 (82)
Dicorynia	3.51	135 (4; 12)	181 (6; 45) *	70 (51)	
Eperua	1.79	50 (3; 6)		54 (42)	40 (161)
Moronobea	1.63			68 (40)	
Sextonia	1.73		202 (7; 83)	39 (28)	
Symphonia	6.12		156 (3; 72)	57 (66)	
Virola	1.01	111 (2; 4)		80 (32)	32 (104)
Vouacapoua	6.27		187 (8; 45)	59 (40)	
Counami					
Dicorynia	3.60			140 (44)	
Jacaranda	0.75			54 (67)	
Sextonia	0.74		73 (4; 49)		
Vouacapoua	4.16		164 (7; 42)	197 (57)	

* sampling restricted to the southern plots of Paracou

Table 3. Estimates of SGS and gene dispersal parameters for each species by site by marker combination: inbreeding coefficient (F), average kinship coefficient between individuals separated less than 50m (F_1), SGS intensity (Sp) and its standard error (SE), neighborhood size (N_b) and 95% confidence interval, gene dispersal distance (σ_g) and 95% confidence interval for three effective densities (De) estimated from the density of adults (D). Lack of values indicates that the estimation procedure did not succeed (see text). N_b was little affected by the assumed De so that the value given is the mean estimate under the three assumed De (values in italics when the estimation procedure did not succeed for all assumed De), and the confidence interval considers the lowest inferior and highest superior interval values.

	SGS parameters			Gene dispersal parameters			
Species Site[§] - marker	F	F_1	Sp (SE)	N_b	σ_g (m) ($De=D/2$)	σ_g (m) ($De=D/4$)	σ_g (m) ($De=D/10$)
Carapa procera							
P - allozymes	0.143***	-0.002 [ns]	0.001 [ns](0.003)	-	-	-	-
P - RAPD		0.076***	0.028*** (0.007)	42 (24-436)	122 (95-204)	182 (138-368)	294 (214-917)
Chrysophyllum sanguin.							
P - allozymes	-0.005[ns]	0.041[ns]	0.008[+] (0.004)	-	-	-	-
P - RAPD		0.032**	0.014*** (0.003)	132 (38-∞)	311 (203-∞)	502 (280-∞)	777 (403-∞)
P - AFLP		0.032*	0.015** (0.004)	61 (27-∞)	241 (176-668)	323 (231-1381)	492 (345-∞)
Dicorynia guianensis							
P - allozymes	-0.199***	0.044***	0.005** (0.001)	*179 (78-∞)*	285 (188-∞)	-	-
P - SSR [#]	0.005[ns]	0.100***	0.026*** (0.004)	-	-	-	-
P - RAPD		0.062**	0.011*** (0.002)	-	-	-	-
C - RAPD		0.030*	0.019*** (0.003)	52 (34-116)	158 (128-227)	203 (173-257)	339 (293-416)
Eperua grandiflora							
P - allozymes	-0.297**	0.021[ns]	0.004[ns] (0.003)	-	-	-	-
P - RAPD		0.030**	0.012*** (0.004)	*86 (30-∞)*	262 (180-∞)	412 (231-∞)	-
P - AFLP		0.039***	0.019*** (0.003)	29 (17-67)	166 (140-217)	234 (189-346)	343 (273-527)
Jacaranda copaia							
C - RAPD		0.015[ns]	0.004[ns] (0.004)	327 (90-∞)	836 (438-∞)	1182 (620-∞)	1869 (980-∞)

Moronobea coccinea							
P - RAPD		0.114***	0.053*** (0.015)	20 (11-74)	134 (104-230)	195 (148-380)	332 (232-∞)
Sextonia rubra							
P - SSR	0.076***	0.008^{ns}	0.005*** (0.001)	344 (50-∞)	728 (257-∞)	743 (328-∞)	848 (479-∞)
P - RAPD		0.005^{ns}	0.006^{ns} (0.006)	-	-	-	-
C - SSR	0.113***	0.007^{ns}	-0.001^{ns} (0.005)	-	-	-	-
Symphonia globulifera							
P - SSR	0.170***	0.025***	0.007*** (0.003)	-	-	-	-
P - RAPD		0.057***	0.028*** (0.005)	40 (22-120)	113 (90-177)	141 (116-193)	209 (170-297)
Virola michelii							
P - allozymes	0.214**	0.002^{ns}	0.000^{ns} (0.010)	-	-	-	-
P - RAPD		0.047*	0.015*** (0.005)	120 (21-∞)	565 (218-∞)	505 (271-∞)	768 (411-∞)
P - AFLP		0.033**	0.016*** (0.007)	*41 (15-∞)*	254 (155-∞)	-	-
Vouacapoua americana							
P - SSR	0.036^{ns}	0.130***	0.051*** (0.009)	18 (11-35)	69 (57-94)	90 (74-125)	139 (117-183)
P - RAPD		0.032**	0.014*** (0.004)	70 (26-610)	134 (98-394)	178 (126-∞)	312 (182-∞)
C - SSR	0.186***	0.058***	0.012*** (0.006)	61 (16-∞)	122 (82-∞)	165 (110-∞)	456 (310-∞)
C - RAPD		0.023***	0.007*** (0.002)	128 (56-5500)	228 (163-1461)	269 (208-472)	539 (421-900)

[§] Sites: P = Paracou; C = Counami. [#] Sampling restricted to the southern plots of Paracou.

P values: ^{ns} for $P > 0.1$; $^{+}$ for $0.1 \geq P > 0.05$; * for $0.05 \geq P > 0.01$; ** for $0.01 \geq P > 0.001$; *** for $P \leq 0.001$

Table 4. Comparison across species between adult density, adult aggregation, seed dispersal distances on basis of field observations (σ_s), and SGS-based gene dispersal estimates (average σ_g estimate over markers and sites where the estimation procedure succeeded for *De* = *D*/4). See Table 1 for complete species names.

Species	Global density (ha^{-1})	Aggreg. $\Omega_{0\text{-}50m}$ *	Seed dispersal σ_s † (ref.)	Gene dispersal Mean σ_g
Seed dispersal by gravity, rodents, or wind with heavy diaspores				
Carapa	4.1	1.2	21 m (1)	182 m
Dicorynia	3.5	2.9	21 m (2)	203 m
Eperua	1.8	3.7	9 m (2)	323 m
Moronobea	1.6	2.8	-	195 m
Vouacapoua	5.2	2.8	31 m (3)	176 m
Seed dispersal by bats, arboreal mammals, birds, or wind with light diaspores				
Chrysophyllum	1.8	2.4	-	413 m
Jacaranda	0.7	1.0	94 m (4)	1180 m
Sextonia	1.7	1.4	-	743 m
Symphonia	6.1	1.3	-	141 m
Virola	1.0	2.0	76 m (2)	505 m

* Ratio of the local density (within 50 m of focal trees) over the global density at Paracou (Condit *et al.*, 2000).

† $\sigma_s = \sqrt{\sum r_i^2 / 2N}$, where r_i is the distance crossed by the i^{th} dispersal event (i = 1 to N). Data from: (1) Jansen *et al.* 2002, (2) Forget 1988, (3) Traissac 2003, (4) Jones *et al*., in press.

Table 5. Gene dispersal contributed by pollen in several tropical forest tree species, as assessed by paternity analyses (in many cases these estimates are minimal ones).

Species	Adult density (ha^{-1})	Gene dispersal induced by pollen σ_g †	Reference
Cordia alliodora	20.85	58 m	Boshier *et al.* 1995
Gliricidia sepium	11.50	21 m	Dawson *et al.* 1997
Swietenia humilis	1.43 / 0.09*	248 m / 918 m*	White 2002
Pithecellobium elegans	1.02	96 m	Chase *et al.* 1996
Neobalanocarpus heimii	0.71	125 m	Konuma *et al.* 2000
Spondias mombin	0.17	43 m	Stacy *et al.* 1996
Dinizia excelsa	0.17 / 0.03*	100 m / 620 m*	Dick *et al.* 2003
Platypodium elegans	0.12	236 m	Hamrick & Murawski 1990

* in heavily fragmented landscapes

† $\sigma_g = \sigma_p / \sqrt{2} = \sqrt{\sum r_i^2 / 4N}$, where r_i is the distance crossed by the i^{th} pollination event (i = 1 to N)

Fig 1.

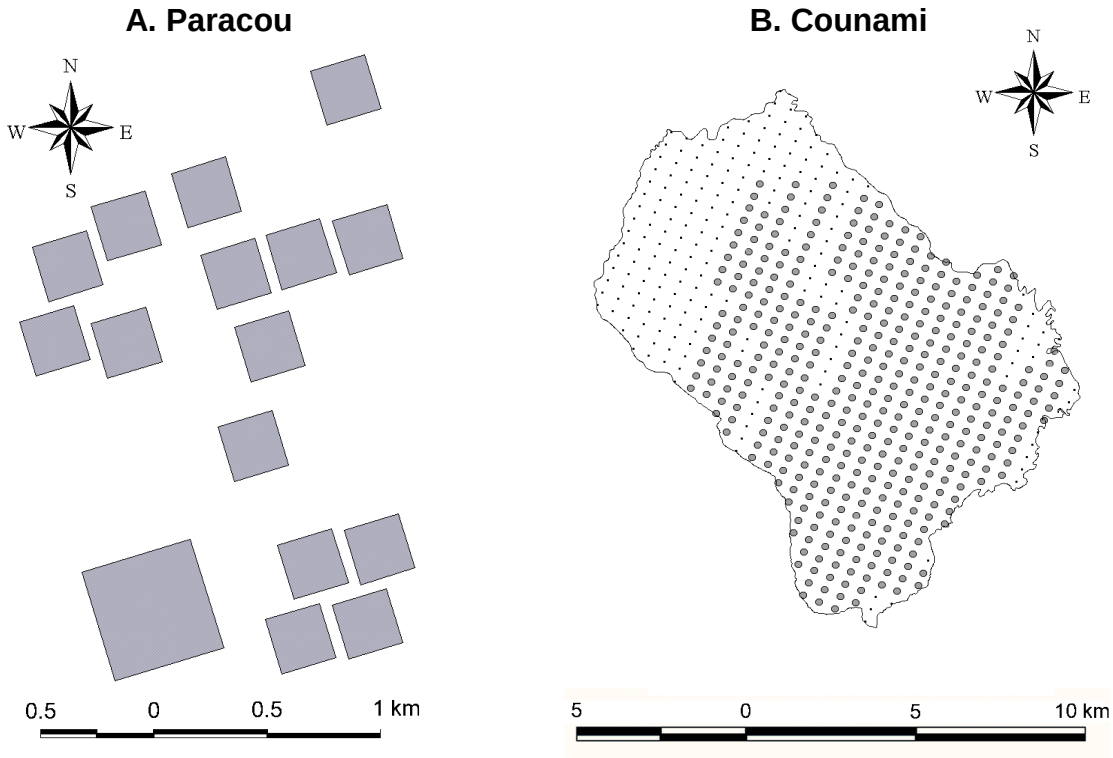
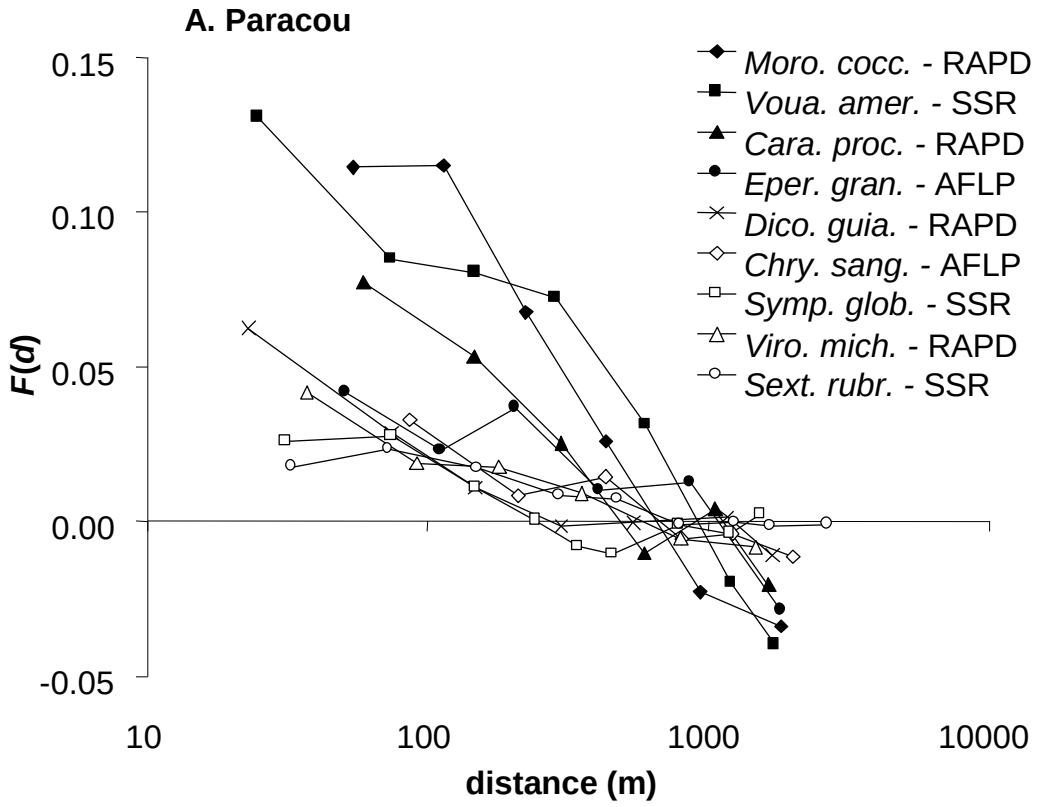
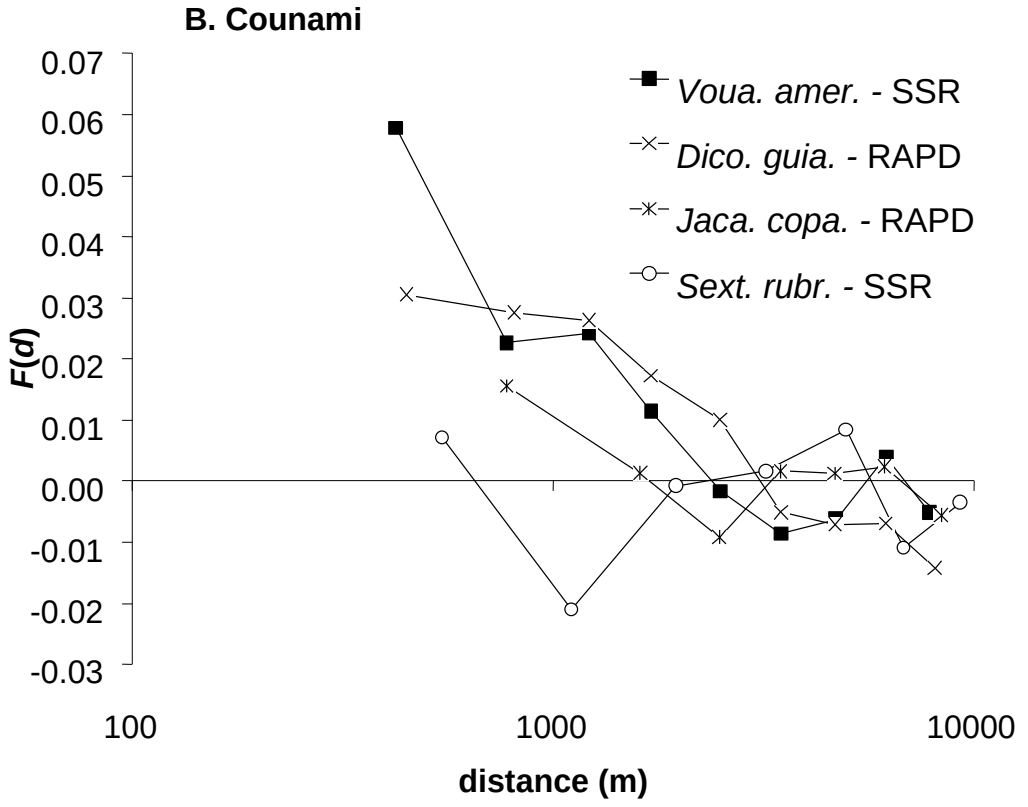


Fig. 2.


A. Paracou
0.15
0.10
0.05
0.00
-0.05
F(d)
10
100
1000
10000
distance (m)
Moro. cocc. - RAPD
Voua. amer. - SSR
Cara. proc. - RAPD
Eper. gran. - AFLP
Dico. guia. - RAPD
Chry. sang. - AFLP
Symp. glob. - SSR
Viro. mich. - RAPD
Sext. rubr. - SSR


B. Counami
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
-0.01
-0.02
-0.03
F(d)
100
1000
10000
distance (m)
Voua. amer. - SSR
Dico. guia. - RAPD
Jaca. copa. - RAPD
Sext. rubr. - SSR