Running head: Beta-diversity in neutral theory

Title: Beta-diversity in spatially implicit neutral models: a new way to assess species migration

François Munoz$^{1,2}$, Pierre Couteron$^{1,2}$ and B.R. Ramesh$^1$

Affiliation:

$^1$ French Institute of Pondicherry (FIP) – 11, Saint Louis Street – Pondicherry 605 001 – India.

$^2$ IRD, UMR AMAP (botAnique et bioinforMatique de l'Architecture des Plantes) – Boulevard de la Lironde, TA A-51/PS2 – F-34398 Montpellier Cedex 5,– France.

François Munoz (corresponding author), francois.munoz@cirad.fr,
Pierre Couteron, pierre.couteron@mpl.ird.fr,
B.R. Ramesh, ramesh.br@ifpindia.org.

Key words:

Beta-diversity; diversity partitioning; \( G_{ST} \); immigration rate; neutral theory; sampling; spatially implicit neutral model; similarity estimators.

Article paper

Date of draft: 1/18/2008
Abstract (Word count = 158)

The spatially implicit neutral model (SINM) of S.P. Hubbell predicts species’ abundance distributions at two levels: local communities where extinction balances immigration, characterized by the immigration number $I$, and the metacommunity, a source pool of migrants where speciation balances extinction. Previously, a plot’s $I$ was estimated from its species abundance distribution. Here we relate neutral theory to the additive partitioning of species diversity and calculate the immigration rate into different plots from the variation in species composition among them. We revisit the $G_{ST}$ statistic of population genetics to introduce a new version, $G_{ST}(k)$, conditional on each community sample, $k$. We derive an analytical expectation of $G_{ST}(k)$ as a function of the local immigration number, $I(k)$, under a generalized version of the SINM, which allows the pool of migrants to deviate from the large-scale speciation-extinction balance. Simulations and field data suggest that $G_{ST}(k)$ provides reasonable estimates of immigration numbers, which were compared with the results from alternative likelihood-based estimations.
**Introduction**

Whittaker (1972) analyzed biodiversity in terms of species diversity within local sampling areas (alpha-diversity), variation in species composition between locations (beta-diversity) and global diversity (gamma-diversity). This distinction has been central to community ecology ever since. The last decade has seen a renewed interest in the topic, thanks to additive schemes of diversity partitioning (Lande 1996), which allow the expression of beta-diversity as a function of environment variables and/or of distance between sampling locations (Couteron and Pélissier 2004; Pélissier and Couteron 2007). Most of these efforts, however, have focused on methods of analyzing field data rather than modeling community composition and diversity dynamics. In fact, few dynamical theories have been developed to model species composition and relative abundances in species-rich communities: this may be one reason why models of neutral communities have aroused such great interest among ecologists (Bell 2000; Hubbell 2001; Chave and Leigh 2002; Chave 2004; Hu et al. 2006).

These neutral models can predict diversity-related parameters, assuming that all trees have equal prospects of dying or reproducing whatever their species or habitat (Hubbell 1979; 2001). Whether this assumption holds or not (Hubbell 2006; Zillio and Condit 2007), neutral models can help ecologists grasp how the interaction between migration and local “ecological drift” (Hubbell 2001) may shape taxonomic composition. Conversely, one may be interested in the reverse problem, that is, using the neutral theory to estimate immigration rates in local communities from the variation in species composition among a set of community samples (beta-diversity), as Wright (1931; 1943) and Slatkin (1985) inferred migration rates within a species from variation in genetic composition in a subdivided population.
Basically, published neutral models are either implicit or explicit regarding the role of local dispersal limitation in broad-scale migration patterns. The spatially-explicit modeling approach is appealing to explore the influence of fine-scale processes of dispersal limitation on community structure (Condit et al. 2002). Expectations for a similarity statistic closely related to beta-diversity have been derived by Chave et al. (2002) and Chave and Leigh (2002) from the spatially explicit framework of Malécot (1948), which assumes space homogeneity and spatial stationarity of the dispersal process. On the other hand, the spatially implicit neutral model (SINM) proposed by Hubbell (2001) also appears useful to investigate, at broad scales, the relationship between ecological communities and their biogeographical background. Analytical expectations have been established for characterizing the species abundance distribution (hereafter denoted SAD) in a local community using the SINM modeling framework (Etienne 2005), which assumes that all the local communities receive migrants from a same source pool. This assumption is debatable (Leigh 2007), but when field datasets from heterogeneous landscapes are to be dealt with, it is an open question to determine which of the common migrant pool and spatial stationarity assumptions is less realistic.

In fact, the use of neutral models as a null hypothesis in community ecology is in its infancy, and it is worth investigating several modeling options to cope with diversified datasets and ecological situations. One interesting feature of Hubbell’s SINM is that it is hierarchical. Therefore it allows us to decouple processes in a large-scale metacommunity from the migration-drift balance in local-scale communities. This is not only a simplifying assumption that allows analytical predictions, it is also a way to investigate different patterns or processes at relevant scales, while methods centered on a particular scale may fail to grasp some of the hierarchical features (Allen and Starr 1982; Levin 1992; Ricklefs 2004). Furthermore, the neutral assumption may be more or less plausible at different scales (Alonso et al. 2006: 455). For instance, it may
locally hold within a particular forest type and fail at larger scales if strong environmental heterogeneities, determining distinct forest types, are encountered (an interpretation proposed by Condit et al. 2002 for their results). But at a very large biogeographic scale, the neutral theory may be anew relevant to account for macroscopic diversity features, where environmental control is quite loose (Hubbell 2001; Ricklefs 2004).

Beta-diversity measures have not yet been investigated in the spatially-implicit context and our overall aim in this paper is to propose a way to express beta-diversity within the SINM framework by revisiting the $G_{ST}$ statistic, introduced by Nei (1973; 1977) in population genetics, to quantify allelic differentiation in subdivided populations. Many metrics are available to quantify beta-diversity (Chao et al. 2005). Here we use the beta fraction of Simpson’s (1949) diversity (Lande 1996) and express the fundamental $G_{ST}$ statistic as a ratio of beta- to gamma- species Simpson’s diversity, to use ecological wording. This is related to Wright’s (1943) intersample statistic $F_{ST}$ as a measure of the similarity of allelic compositions across separate samples (Excoffier 2001).

Under island neutral models (Wright 1931; Wright 1943: 114), there are straightforward expectations for such statistics, which are used for inferring migrant fluxes between populations in population genetics (Weir and Cockerham 1984; see review by Slatkin 1985).

In the context of community ecology, we shall explore the potential of this statistic for estimating migration among sample plots and propose a comprehensive framework by defining a new $G_{ST}$-like statistic, $G_{ST}(k)$ for each sample $k$ instead of dealing with an average over all samples. We shall provide new results on the apportioning of diversity/dissimilarity indices (Rao 1982; Lande 1996; Couteron and Pélassier 2004; Pélassier and Couteron 2007) in order to establish explicit expectations for $G_{ST}(k)$ regarding the spatially implicit neutral models (SINM).
In the SINM formalism, the fundamental immigration number $I$, also named the fundamental dispersal number (Etienne and Alonso 2005; Alonso et al. 2006), features the number of migrants from the metacommunity competing with local offspring to replace a newly dead individual in a local community. Most recent methods estimate the immigration number from species abundance distributions in local communities (Etienne 2005; Etienne 2007; Munoz et al. 2007). In contrast, here we shall provide a method integrating beta-diversity information through the $GST(k)$ statistic to estimate the immigration number $I(k)$ for any sample $k$ of a local community. Our approach models migrants as random draws from a single source pool, but it accepts that the source pool deviates from a speciation-drift equilibrium, for either neutral or non-neutral reasons, by referring to an extended SINM framework, which includes Hubbell’s two-level model (noted hereafter 2L-SINM) as a particular case.

Applying the approach to simulated community samples and to fifty evergreen forest plots in South India, we shall eventually assess the potential of the $GST(k)$ statistic for estimating immigration numbers $I(k)$ in local communities. Furthermore, we shall compare the $GST(k)$-based approach with an alternative likelihood-based approach (Munoz et al. 2007) that calculates the immigration numbers from local species abundance distributions in the context of the 2L-SINM.

**Methodological background**

**Similarity statistics**

Very often, ecological communities cannot be exhaustively characterized, and we consider here community samples (denoted by $k$), as a type of data frequently obtained from fieldwork, for instance by enumerating tree species abundances in forest plots distant enough to be in different communities. In order to measure taxonomic similarity (or dissimilarity) within and between
community samples, one may define statistics in analogy with those used in population genetics, where intra- and interpopulation similarities measure the co-occurrence of alleles. Let us consider the probability, \( F_{\text{intra}}(k) \), that two arbitrary individuals in sample \( k \) are conspecific, i.e., belong to the same species, and \( F_{\text{global}} \), the probability that two arbitrary individuals from the lumped set of samples are conspecific. To quantify intersample similarity, let \( F_{\text{inter}}(k,l) \) be the probability that an individual drawn from sample \( k \) and one from sample \( l \) are conspecific.

We also introduce two other similarity statistics, relative to a given sample \( k \), which are the probability that an arbitrary individual in \( k \) is conspecific with an individual from the pooled dataset, including \( k \), called \( F_{\text{global}}(k) \), or is conspecific with an individual from any sample distinct from \( k \), called \( F_{\text{inter}}(k) \). Applying the classical principle of conditional probability yields:

\[
F_{\text{global}}(k) = P(k/k)F_{\text{intra}}(k) + (1 - P(k/k))F_{\text{inter}}(k),
\]

where \( P(k/k) \) notes the conditional probability to draw a second individual from sample \( k \).

Similarities can be estimated using either the most popular approximate estimators (sampling with replacement), or exact estimators (sampling without replacement), as introduced in Appendix A, where we further establish that Eq.1 is exact for both types of estimators. An important aspect of the generality of Eq.1 is that it can accommodate unequal sample sizes, in contrast with simplified formulas frequently used in population genetics (see Slatkin 1985: 398).

**The \( G_{ST} \) statistic**

The \( G_{ST} \) statistic, as introduced by Nei (1973), measures genetic differentiation among populations from the ratio of alleles’ beta- to gamma-diversity (sensu Simpson 1949) computed from an entire set of populations. As far as species enumerated in a field sample, e.g., a plot, are
analogous to a very large set of alleles recorded at a given locus, the $G_{ST}$ statistic, though largely ignored in ecology (but see Hardy and Sonké 2004), appears appealing as a synthetic measure of the variation among samples in species composition. Although Nei’s initial definition was introduced for equal sample sizes and sampling with replacement, we show in Appendix B that the concept can be kept without any restrictive assumption, neither on sampling nor on sample size, leading to the following relationship between $G_{ST}$, diversity and similarity statistics:

$$G_{ST} = \frac{\text{Div} \_ \beta}{\text{Div} \_ \gamma} = \frac{\bar{F}_{\text{intra}} - F_{\text{global}}}{1 - F_{\text{global}}},$$  \hspace{1cm} \text{Eq.2a}

where $\bar{F}_{\text{intra}}$ is the average intrasample similarity, defined as $\bar{F}_{\text{intra}} = \sum_k P(k)F_{\text{intra}}(k)$ (Rao 1982), and relates to Simpson’s (1949) alpha-diversity as $\text{Div} \_ \alpha = 1 - \bar{F}_{\text{intra}}$. Simpson’s gamma-diversity likewise reads $\text{Div} \_ \gamma = 1 - F_{\text{global}}$ and Nei’s (1973) fundamental result linking beta-diversity to similarities, i.e., $\text{Div} \_ \beta = \bar{F}_{\text{intra}} - F_{\text{global}}$, holds with no restrictive assumption about sampling and sample sizes (Appendix B).

In population genetics, expectations for $G_{ST}$ under island models for subdivided populations (Wright 1943; Takahata and Nei 1984; Slatkin 1985) are simple functions of the migration onto an island. This result is used (Wright 1943) to estimate the average flux of migrants between populations from the measure of the $G_{ST}$ statistic.

**A novel sample-based $G_{ST}(k)$ statistic**

Because it may be useful to estimate immigration at a fine level, that is, for each community sample, we propose a comprehensive framework, which is based on the statistic:
The relationship between the numerator and beta-diversity fraction conditional on \( k \) is established in Appendix B.

**The three-level spatially implicit neutral model (3L-SINM)**

The spatially implicit neutral models (SINM), as developed by Bell (2000) and Hubbell (2001), have presented speciation, migration and drift as the possible primary drivers of community dynamics, and assumed a top-down coupling of processes that conforms to the hierarchical theory of Allen and Star (1982). While very rare speciation events are modeled at an upper, very large spatial and temporal scale (the metacommunity), migration and ecological drift are assumed to shape the composition of local communities. Until now, most works referring to the SINM postulated that migrants establishing in local communities are randomly drawn from the entire metacommunity, which represents the biogeographical context. This constitutes the hereafter-denoted two-level SINM (2L-SINM; Fig. 1A). We consider here a more general three-level SINM (3L-SINM; Fig. 1B) where the migrants originate in a common source pool, which may correspond to a particular subregion of a very large, hypothetical metacommunity but which can substantially deviate from the SAD of the metacommunity (sensu Hubbell 2001), i.e., from Fisher’s log-series. The gain of generality is twofold. Firstly we do not need to assume random migration across the entire metacommunity, as we just need to assume a common source-pool at a far smaller, regional scale (the scale issue). Secondly, the pool of migrants may reflect limited migration from the metacommunity (see Etienne 2007: 616) for neutral as well as non-neutral processes taking place at the scale of the region to which the pool refers (the process issue).
Hubbell’s 2L-SINM is thus a very specific case of the 3L-SINM where the regional pool of migrants is just a random sample of the metacommunity.

The sampling design is assumed to provide a fair representation of the specific composition of the area under study and of the corresponding source pool postulated by 3L-SINM. The samples are assumed to be far enough from one another to belong to distinct communities. According to the coalescent-type reasoning of Etienne and Olff (2004), any individual in a community sample is seen as descended from a migrant ancestor, randomly drawn from the regional pool. The probability that an individual in \( k \) is conspecific with an individual in another sample, i.e. \( F_{\text{inter}}(k) \), is thus equal to the probability that their ancestors are conspecific. This also implies that \( F_{\text{inter}}(k) \) is equal to the similarity between the source pool and \( k \), \( F_{\text{inter}}(\text{pool}, k) \), and to the similarity within the source pool:

\[
F_{\text{inter}}(k) = F_{\text{inter}}(\text{pool}, k) = F_{\text{intra}}(\text{pool}). \tag{Eq.3}
\]

**Expectations of the \( G_{ST}(k) \) statistic for the 3L-SINM**

The immigration number, \( I(k) \), is the number of migrants that are available for competition with a resident offspring in the community including local sample \( k \). The immigration rate for a sample \( k \) made up of \( n(k) \) individuals is \( m(k) = \frac{I(k)}{I(k) + n(k) - 1} \), in accordance with the modification of Moran’s (1958) model introduced by Hubbell (2001) and discussed by Etienne et al (2007), which considers that a dead individual cannot be replaced by its own offspring (contrary to the initial Moran model, see Appendix C for detail). Let us consider \( N_c \) community samples, each of them
belonging to a distinct local community, where the immigration number \( I(k) \) \((k = 1 \ldots N_c)\) differs for different samples \( k \).

Remembering that \( F_{\text{inter}}(\text{pool}, k) \) denotes the probability that an individual in sample \( k \) is conspecific with one in the source pool of migrants, we expect (Appendix C):

\[
F_{\text{intra}}(k) = 1 - \frac{I(k)}{I(k) + 1} \left(1 - F_{\text{inter}}(\text{pool}, k)\right). \tag{Eq.4}
\]

It is interesting to note that this result is true for both the initial and modified versions of Moran’s model (Appendix C). Further, equations 1, 3 and 4 permit derivation of an expectation for \( G_{ST}(k) \) under the 3L-SINM (Appendix D), which presents the great advantage of being independent of \( F_{\text{inter}}(\text{pool}, k) \):

\[
G_{ST}(k) = \frac{1}{1 + I(k)/(1 - P(k/k))}. \tag{Eq.5}
\]

If the total number of individuals in the lumped set of samples is far larger than the number of individuals in \( k \), then \( P(k/k) \approx 0 \) and \( G_{ST}(k) = \frac{1}{1 + I(k)} \). \( G_{ST}(k) \) only depends on the migration number \( I(k) \) (Eq.5) and is independent of any parameter characterizing the source pool of migrants. This result, established at the level of individual samples, resembles the well-known relationship relating the overall \( G_{ST} \) statistic to the average flux of migrants in island models from population genetics (Takahata and Nei 1984), although the 2L-SINM is rather analogous to a mainland–island model (Hu et al. 2006: 549). Moreover, our reference model (3L-SINM) does not restrictively assume that migrants come directly from the entire metacommunity.
Application: inference from simulated and real-world community samples

Simulating neutral communities: a hierarchical framework

The original ‘sequential’ algorithm of Etienne (2005; Appendix S2) can simulate the steady-state SAD in a neutral community sample with given values of $\theta$ (biodiversity number, Hubbell 2001) and $m$. A second version (Etienne 2007) can generate several samples according to the 2L-SINM framework, with the same $\theta$ and potentially varying $m$ values. To obtain community samples conforming to a particular version of the 3L-SINM (Fig. 1B) featuring migration limitation at two scales, we wrote a modified two-step algorithm (Munoz et al. 2007, Appendix) in Matlab language (Mathworks 2004). The first step is a direct implementation of Etienne’s (2005) algorithm to generate the pool of migrants as a very large sample characterized by a biodiversity number of $\theta$ and by a regional immigration rate ($0 < m_{pool} \leq 1$), while the second part of the algorithm is run to generate $N_c$ local samples $k$ made of $n(k)$ individuals with known local immigration rates ($m(k) < 1$), and for which all ancestors belong to the predefined pool (see Munoz et al. 2007 for more details). This way of simulating allows us to illustrate that the estimation process is valid for any regional pool following the zero-sum multinomial abundance distribution parameterized by $\theta$ and $I_{pool}$ (corresponding to $m_{pool}$), which is a more general reference than the log-series distribution of a metacommunity sample parameterized by $\theta$ (2L-SINM). For simplicity sake, this simulation algorithm does not cover the full generality of the 3L-SINM framework, which may also integrate non-neutral processes at regional scale.

To test the reliability and the precision of $G_{ST}(k)$-based inferences, we simulated a large number of rather small community samples ($n(k) = 400$ individuals), with values for $I(k)$ increasing from 10 to 900 by increments of 10. We chose 50, 100 and 150 as alternative values of the biodiversity number $\theta$, to explore the range of biodiversity figures found for tree communities in semi-
evergreen and evergreen tropical forests (Hubbell 2001; Munoz et al. 2007). For each value of \( I(k) \), we simulated five independent migrant pools, with fixed size \( 50n(k) = 20,000 \) individuals, for each of which we finally simulated \( N_c = 50 \) community samples, indexed by \( k \). We thus generated \( 5 \times 50 \) samples for each value of \( I(k) \), hence a total of \( 90 \times 250 = 22,500 \) simulations for each value of \( \theta \). We reiterated the process using \( m_{pool} = 1 \) (the migrant pool is a metacommunity sample, 2L-SINM, \( I_{pool} = \text{infinity} \)) and \( m_{pool} = 0.01 \), i.e. \( I_{pool} \approx 200 \) (the migrant pool undergoes migration limitation from the metacommunity, a particular version of the 3L-SINM). The estimator of the \( GST(k) \) statistic is calculated from the estimators of the similarities (notation with hats), which can be either exact or approximated (Appendix A) as:

\[
\hat{G}_{ST}(I,k) = \frac{\hat{F}_{\text{intra}}(k) - \hat{F}_{\text{global}}(k)}{1 - \hat{F}_{\text{global}}(k)} .
\]

The computation of \( \hat{F}_{\text{global}}(k) \) involves all the simulated community samples generated from the same simulation of the source pool (one replicate), without any use of the known values of \( \theta \) or \( m_{pool} \). Accordingly, the estimated immigration number for each sample \( k \), \( \hat{I}_{GST}(I,k) \), is:

\[
\hat{I}_{GST}(I,k) = \left( \frac{N_e - n(k)}{N_e - 1} \right) \left( \frac{1}{\hat{G}_{ST}(I,k)} - 1 \right) = \left( \frac{N_e - n(k)}{N_e - 1} \right) \left( \frac{1 - \hat{F}_{\text{intra}}(k)}{\hat{F}_{\text{intra}}(k) - \hat{F}_{\text{global}}(k)} \right) . \quad \text{Eq.6}
\]

\( N_e \) is the number of individuals in the lumped set of community samples.

All the following estimation results are based on exact estimators (Appendix A). Approximate estimators, although widely used in population genetics and ecology, lead to a notable bias on the \( GST(k) \)-based estimations, as illustrated in Appendix E.
The Western Ghats dataset

In addition, we estimated immigration rates into each of a set of 50 one-hectare plots established in the wet tropical forests of the portion of the Western Ghats mountain range within the Karnataka State, south India (Pascal 1988; Ramesh and Swaminath 1999; Munoz et al. 2007). The plots lay on the western slope of the Ghats that receives the southwest monsoon from June through October. In addition, thermic convection rainfall occurs in May and November. Climate variation induced by the relief dramatically influences both physiognomy and composition of vegetation, but we focus here on the forests dominated by evergreen species that mostly grow in the wettest areas (in terms of total annual rainfall) of the “Allepey-Mangalore” climatic regime (Pascal 1988: 14-17). Plots mostly belong to the medium elevation forest (Pascal 1988: 131) and receive an annual rainfall average of 4,100 mm. In Mangalore, there are 5-6 months for which the median monthly rainfall is less than 100 mm. The mean basal area is 30 m²/ha and, following Leigh et al. (1993) and Hubbell (2001), only trees with more than 10 cm dbh were considered (19,555 individuals), so that there were on average 391 individuals and 45.2 species per plot. The average intra-plot species similarity is $\hat{F}_{intra} = 0.108$, with values ranging from 0.046 to 0.44 and a standard deviation of 0.07.

Comparison with an approach based on species abundance distributions

Etienne and Olff (2004) and Etienne (2005) proposed a likelihood-based estimation of the 2L-SINM parameters, $\theta$ and $I$ (or $m$), using species abundance distributions in individual samples. The method was first designed to investigate a single, rather large plot featuring thousands of individuals, and two distinct adaptations have recently been proposed to deal with networks of smaller plots. Etienne (2007) considers in practice a constant $I$ for all samples, while the “two-
stage approach” of Munoz et al. (2007) allows \( I(k) \) to vary across communities, in a way comparable with the \( G_{ST}(k) \)-based approach. Published applications of both likelihood-based approaches assume that the pool of migrants is a random sample from the metacommunity (2L-SINM), while the \( G_{ST}(k) \)-based estimation works whatever the nature of the source pool.

**Results**

**\( G_{ST}(k) \)-based estimation on simulated community samples**

We estimated \( \hat{I}_{GST}(I,k) \) and \( \hat{m}_{GST}(I,k) \) for each combination of \( \theta \) (50, 100 or 150) and \( m_{pool} \) (0.01 or 1), with a very good agreement, on average, over the entire range of expected \( I \) or \( m \) values (Fig. 2, bold crosses). As expected from the analytical reasoning, introducing regional immigration limitation in the pool of ancestors (a particular version of the 3L-SINM, \( m_{pool} = 0.01 \), Fig. 2C and 2D) had no visible influence on the quality of estimation (compare Fig. 2C-D to Fig. 2A-B) and we have further checked that the estimation remains unbiased when \( m_{pool} \) is made to vary between 0.001 and 1 (i.e., \( I_{pool} \) between 20 and infinity).

Furthermore, the variance of estimated values decreases when \( \theta \) increases, and for \( \theta = 100 \) (results not shown) the width of the envelope shrinks considerably in comparison to \( \theta = 50 \). The variance likewise increases when \( m_{pool} \) becomes very small, that is, when there are far fewer species in the pool of migrants. Higher species richness in the samples may convey more information on the diversity components, and indeed we devise our approach to study species-rich communities.
Comparing the $G_{ST}(k)$-based and the likelihood-based estimation

We further compared the $G_{ST}(k)$-based estimation with the two-stage likelihood-based estimation of Munoz et al. (2007) for simulated neutral communities generated with parameter values (2L-SINM, $\theta = 50$, $I$ in $[0, 50]$ and $n(k) = 400$) comparable with the figures previously found for our field dataset in the Western Ghats (Munoz et al. 2007). The overall correlation is very good (Pearson’s $r = 0.84$), with a very slight positive discrepancy (Fig. 3A).

Finally, we compared the $G_{ST}(k)$- and the two-stage likelihood-based estimations of immigration numbers for our field dataset (50 one-hectare tropical wet evergreen forest plots from southwest India). Using the two-stage method, we first got a mean value of $\theta = 62$ from many samples randomly drawn from all plots (one individual per plot, $n = 50$ for each resampling), and at the second step we used Etienne’s (2005) likelihood to get the immigration number for each plot, according to the fixed $\theta = 62$ (Munoz et al. 2007). There was a degree of agreement between the $G_{ST}(k)$- and likelihood-based estimations (Pearson’s $r = 0.66$), but the $G_{ST}(k)$-based estimations were clearly shifted towards lower values (Fig. 3B), and averaged $\hat{I}_{GST} = 14.5$ ($\hat{m}_{GST} = 0.045$), with range $[1.3;35]$, instead of $\hat{I}_{MLE} = 23.4$ ($\hat{m}_{MLE} = 0.066$), with range $[4.7;50]$, for likelihood-based estimations. This negative difference contrasts with the only slightly positive bias noted on the simulated samples.

Discussion

In this paper, we have introduced the idea that diversity (sensu Simpson 1949) or, equivalently, similarity statistics provide a relevant basis for studying species migration patterns using spatially implicit neutral models (SINM). We propose here to use the variation in species composition across communities to estimate the immigration number in a way that differs from existing
approaches (Hubbell 2001; McKane et al. 2004; Etienne 2005; Munoz et al. 2007), which rely on species abundance distributions (usually noted SAD).

The $G_{ST}$ statistic, borrowed from population genetics, explicitly integrates the information conveyed by the intersample beta-diversity. Its useful properties were first outlined for island models, as $G_{ST}$ relates to migration independently from mutation, and as it more rapidly reaches steady state than the similarities $F_{intras}$, $F_{inter}$ and $F_{global}$ (Crow and Aoki 1984). Interestingly, those properties were nearly verified under the two-dimensional stepping stone model, according to which migration only occurs between proximate islands (Maruyama 1972; Crow and Aoki 1984; see also Cox and Durrett 2002 for other migration patterns). This suggests that $G_{ST}$ allows investigation of species migration for most real-world systems (Nei 1973; Slatkin 1985).

The version at sample scale, $G_{ST}(k)$, is new, both in fields of population genetics and neutral ecology, and displays appealing properties for inference, because it is expected to be, in the case of the SINM, a simple function of the influx of immigrants (Eq.5). This result is all the more interesting in that it is not only valid in the context of the two-level model used by Hubbell (2001) where the source pool is randomly drawn from a hypothetical metacommunity (2L-SINM, Fig. 1A), but also in presence of regional source pools marked by either neutral or non-neutral migration restrictions from broader biogeographic levels (3L-SINM, Fig. 1B). Investigating to what extent $G_{ST}(k)$ conserves its properties under the stepping stone model, for instance using appropriate simulations, as Crow and Aoki (1984) did in population genetics, is a promising perspective that could allow a meeting point between the spatially implicit and explicit neutral approaches to be reached.

In practice, a spatially implicit approach should be useful for investigating data acquired in heterogeneous or fragmented landscapes, for which assuming space continuity and spatial stationarity of the dispersal process may be problematic (see Hu et al. 2006). On the other hand,
assuming a common source pool, even for nearby plots in a specific region, is debatable in terms of biological realism (Leigh 2007). However, we do not claim to mechanistically model how the fine scale process of dispersal limitation from parents determines migration limitation at a community scale, an objective for which more spatially explicit models are probably preferable (Chave and Leigh 2002; Etienne 2005 Appendix; Economo and Keitt 2007). We instead address community networks at broader scales, and aim to quantify, through phenomenological immigration numbers $I(k)$, the extent to which several communities appear more or less isolated from a common biogeographical background. The decoupling of processes across scales is here an important paradigm that calls for interfacing scales of interest and scale-centered models (Levin 1992). The neutral hypothesis may indeed hold at some spatial scales, while failing at others (Alonso et al. 2006), and the variation of immigration numbers, $I(k)$, may help in the future characterize how data depart from neutrality. Spatially implicit and explicit models should be complementary, depending on the ecological context (structure of the environment) and on the scale of the study (size and number of plots, and interplot distances). Migration rates and numbers inferred from both types of models should be further compared and their respective meaning discussed, as attempted by Leigh (2007).

We provided a first illustration of the relevance of using $\text{GST}_k$ to estimate the local immigration number $I(k)$, or equivalently, the immigration rate, $m(k)$, from a large array of neutral community samples of same size, simulated under a particular version of the 3L-SINM (Fig. 1B) with known parameter values. The $\text{GST}_k$-based estimation appeared to be devoid of any serious bias for a very broad range of immigration rates and for the breadth of diversity levels that can be expected for tree communities in tropical evergreen forests ($\theta$ lying between 50 and 150). In the context of the 2L-SINM, the $\text{GST}_k$-based estimations closely agree with those provided by the only other available method that allows investigating varying immigration rates over a network of
community samples (Munoz et al. 2007), even though the latter is far more demanding in terms of computation resources. Furthermore, as expected under the 3L-SINM, the $G_{ST}(k)$-based estimation of immigration in local communities was insensitive to decreasing the $m_{pool}$ parameter featuring migration limitation for the regional source pool ($m_{pool} < 1$), hence it was insensitive to the deviation from the speciation drift equilibrium expected in the metacommunity. This opens up the possibility of studying biogeographic regions and subregions that have been shaped by fragmentation processes, even non-neutral ones, at several scales.

Noticeably, using approximate estimators of the similarity functions in $G_{ST}(k)$ led to a serious bias for high values of the immigration rate (Appendix E), and it appears therefore important to consider the exact estimators (sampling without replacement) introduced in Appendix A, even though working with approximate estimators (sampling with replacement) is a widespread custom. Both estimators are however consistent with the fundamental properties of the additive partitioning of diversity/dissimilarity, even though it was only established previously for sampling with replacement (Nei 1973; Rao 1982; Lande 1996; Couteron and Pélissier 2004).

We also applied the $G_{ST}(k)$-based estimation to a dataset featuring tree species composition in evergreen forests of southwest India, and compared the results with those yielded by the alternative likelihood-based method (Munoz et al. 2007). Interestingly, although both methods provided reasonable, correlated estimates of immigration numbers for the field dataset, a discrepancy was noted (Fig. 3). This discrepancy was probably because the $G_{ST}(k)$-based estimation works for any SAD in the source pool, while the likelihood-based method is built on the 2L-SINM and requires an initial estimation of the biodiversity number $\theta$ from a metacommunity sample supposedly at speciation drift equilibrium. This interpretation is made plausible by the fact that there was no obvious discrepancy between the two estimation methods when considering the simulated communities complying with the 2L-SINM (Fig. 3A). Further in-
depth comparisons between $G_{SK}(k)$-based and likelihood-based approaches (Etienne 2005; Etienne 2007; Munoz et al. 2007), which are beyond the scope of a single paper, will be needed. As for Munoz et al. (2007), the present approach is designed to deal with many scattered community samples and probably cannot be applied to a small number of field plots (as for instance in Etienne 2007).

Because difficulties in estimating the fundamental parameters of the SINM have greatly hampered the development of neutral ecology, in spite of recent breakthroughs, much is still to be done. Apart from the technical aspects of parameter estimation, this paper bridges a gap between the analysis of beta-diversity and the models of species migration, and opens up new perspectives on the neutral theory and its applicability to the characterization of real-world ecological communities.

**Acknowledgments**

We are especially indebted to E.G. Leigh Jr. for his substantial and generous review of the successive versions of the paper, which helped us to broaden the scope. We also thank Champak B. Reddy, Philippe Jarne, François Massol, Raphaël Pélissier, Rampal S. Etienne and one anonymous referee for their valuable help.
Online Appendix A: Estimators of the similarity statistics

Similarities defined in the main text can be estimated in reference to either exact sampling probabilities (without replacement, i.e., exhaustive sampling), or to approximations corresponding to sampling with replacement. To limit bias in the presence of scarce species (as illustrated in Appendix E), it was necessary for us to clarify the properties of the estimators.

Let us consider $N_c$ samples of unequal sizes $n(k)$, summing to $N_e = \sum_{k=1}^{N_c} n(k)$ individuals. The probability of randomly drawing an individual from sample $k$ is thus $P(k) = \frac{n(k)}{N_e}$. Furthermore, let us suppose that there are $S$ species with abundances $(n_1…n_S)$ in the lumped set of samples and $(n_1(k)…n_S(k))$ in a given sample $k$, so that $n(k) = \sum_{s=1}^{S} n_s(k)$ and $N_e = \sum_{s=1}^{S} n_s$.

Exact estimators

The exact estimators are $\hat{F}_{\text{global}} = \sum_{s=1}^{S} \left( \frac{n_s}{N_e} \right) \left( \frac{n_s - 1}{N_e - 1} \right)$ (Simpson 1949) for the global similarity

and $\hat{F}_{\text{global}}(k) = \sum_{s=1}^{S} \left( \frac{n_s(k)}{n(k)} \right) \left( \frac{n_s - 1}{N_e - 1} \right)$ for its version, conditional on sample $k$.

$\hat{F}_{\text{intra}}(k) = \sum_{s=1}^{S} \left( \frac{n_s(k)}{n(k)} \right) \left( \frac{n_s(k) - 1}{n(k) - 1} \right)$ estimates the species similarity in plot $k$,

$\hat{F}_{\text{inter}}(k,l) = \sum_{s=1}^{S} \left( \frac{n_s(k)}{n(k)} \right) \left( \frac{n_s(l)}{n(l)} \right)$ estimates the similarity between plots $k$ and $l$ and

$\hat{F}_{\text{inter}}(k) = \sum_{s=1}^{S} \left( \frac{n_s(k)}{n(k)} \right) \left( \frac{n_s - n_s(k)}{N_e - n(k)} \right)$ the similarity between plot $k$ and all the other plots.
In agreement with sampling without replacement, the conditional probability in Eq.1 (main text) must be: 

\[ P(k/k) = \frac{n(k) - 1}{N_e - 1}. \]

Consider 

\[ \hat{F}_{\text{global}}(k) = \sum_{s=1}^{S} \left( \frac{n_s(k)}{n(k)} \times \frac{n_s - 1}{N_e - 1} \right) = \sum_{s=1}^{S} \left( \frac{n_s(k)}{n(k)} \times \frac{n_s(k) - 1}{N_e - 1} \right) + \sum_{s=1}^{S} \left( \frac{n_s(k)}{n(k)} \times \frac{n_s - n_s(k)}{N_e - 1} \right). \]

Hence 

\[ \hat{F}_{\text{global}}(k) = \frac{n(k) - 1}{N_e - 1} \hat{F}_{\text{intra}}(k) + \frac{N_e - n(k)}{N_e - 1} \hat{F}_{\text{inter}}(k), \]

which is 

\[ \hat{F}_{\text{global}}(k) = P(k/k) \hat{F}_{\text{intra}}(k) + (1 - P(k/k)) \hat{F}_{\text{inter}}(k), \] Eq.1 in the main text.

**Approximate estimators**

Assuming sampling with replacement, approximate estimators for the intrasample and global similarities are the most generally used in both ecology and population genetics, and read:

\[ \hat{F}_{\text{intra}}(k) = \sum_{s=1}^{S} \left( \frac{n_s(k)}{n(k)} \right)^2 \text{ and } \hat{F}_{\text{global}}(k) = \sum_{s=1}^{S} \left( \frac{n_s(k)}{n(k)} \right) \left( \frac{n_s}{N_e} \right). \]

In this case, we must also write:

\[ P(k/k) = \frac{n(k)}{N_e}. \]

\[ \hat{F}_{\text{inter}}(k,l) \text{ and } \hat{F}_{\text{inter}}(k) \text{ are the same as for sampling without replacement.} \]

Then 

\[ \hat{F}_{\text{global}}(k) = \sum_{s=1}^{S} \left( \frac{n_s(k)}{n(k)} \times \frac{n_s}{N_e} \right) = \sum_{s=1}^{S} \left( \frac{n_s(k)}{n(k)} \times \frac{n_s(k)}{N_e} \right) + \sum_{s=1}^{S} \left( \frac{n_s(k)}{n(k)} \times \frac{n_s - n_s(k)}{N_e} \right) \]

and 

\[ \hat{F}_{\text{global}}(k) = \frac{n(k)}{N_e} \hat{F}_{\text{intra}}(k) + \frac{N_e - n(k)}{N_e} \hat{F}_{\text{inter}}(k), \]

which is again 

\[ \hat{F}_{\text{global}}(k) = P(k/k) \hat{F}_{\text{intra}}(k) + (1 - P(k/k)) \hat{F}_{\text{inter}}(k), \] Eq.1 in the main text.
Online Appendix B: Definition of beta-diversity from similarity functions

Average beta-diversity relative to a given sample, \( k \)

We can use Nei’s (1973: 3322) definition to express the intersample diversity \( D_{\text{inter}}(k,l) \), from the similarities:

\[
D_{\text{inter}}(k,l) = \frac{1}{2} (F_{\text{intra}}(k) + F_{\text{intra}}(l) - 2F_{\text{inter}}(k,l)).
\]

Nei (1973) implicitly assumed sampling with replacement, which implies that \( D_{\text{inter}}(k,l) \) reads

\[
\frac{1}{2} \sum_n \left( \frac{n_s(k)}{n(k)} - \frac{n_s(l)}{n(l)} \right)^2
\]

(see also Couteron and Pélissier 2004), and assumed equal sample sizes, but we shall show that the above formula is of a broader generality.

Let us define the average beta-diversity conditional on \( k \) as

\[
D_{\text{inter}}(k) = \sum_{l \neq k} P(l/k) D_{\text{inter}}(k,l).
\]

Whatever the sampling scheme we have:

\[
2D_{\text{inter}}(k) = \sum_{l \neq k} P(l/k) F_{\text{intra}}(k) + \sum_{l \neq k} P(l/k) F_{\text{intra}}(l) - 2 \sum_{l \neq k} P(l/k) F_{\text{inter}}(k,l).
\]

Here we note that the two nested statistics expressing intersample similarity are linked through the relationship:

\[
\sum_{l \neq k} P(l/k) F_{\text{inter}}(k,l) = (1 - P(k/k))F_{\text{inter}}(k).
\]

Thus, the expression for average beta-diversity becomes:

\[
2D_{\text{inter}}(k) = (1 - P(k/k))F_{\text{intra}}(k) + \left[ \sum_{l \neq k} P(l/k) F_{\text{intra}}(l) - P(k/k)F_{\text{intra}}(k) \right] - 2(1 - P(k/k))F_{\text{inter}}(k)
\]

\[
2D_{\text{inter}}(k) = F_{\text{intra}}(k) + \sum_{l} P(l/k) F_{\text{intra}}(l) - 2(P(k/k)F_{\text{intra}}(k) + (1 - P(k/k))F_{\text{inter}}(k))
\]
Considering $F_{\text{global}}(k) = P(k/k)F_{\text{intra}}(k) + (1 - P(k/k))F_{\text{inter}}(k)$ (Eq.1, main text), it yields:

$$2D_{\text{inter}}(k) = F_{\text{intra}}(k) + \sum_{l} P(l/k) F_{\text{intra}}(l) - 2F_{\text{global}}(k).$$

- Assuming sampling with replacement, we have: $\sum_{l} P(l/k) F_{\text{intra}}(l) = \sum_{l} P(l) F_{\text{intra}}(l)$.

- Sampling without replacement yields: $\sum_{l} P(l/k) F_{\text{intra}}(l) = \sum_{l} \frac{n(l)}{N_e} F_{\text{intra}}(l) + \frac{n(k) - 1}{N_e - 1} F_{\text{intra}}(k) = \sum_{l} \frac{n(l)}{N_e - 1} F_{\text{intra}}(l) - \frac{1}{N_e - 1} F_{\text{intra}}(k)$

$(n(k)$ is the number of individuals in sample $k)$.

Let us take the classical definition for the average intrasample similarity (Rao 1982), that is:

$$\bar{F}_{\text{intra}} = \sum_{k} P(k) F_{\text{intra}}(k).$$

We finally have:

- For sampling with replacement: $2D_{\text{inter}}(k) = \bar{F}_{\text{intra}} + F_{\text{intra}}(k) - 2F_{\text{global}}(k)$.

- Without replacement: $2D_{\text{inter}}(k) = \frac{N_e}{N_e - 1} \bar{F}_{\text{intra}} + \frac{N_e - 2}{N_e - 1} F_{\text{intra}}(k) - 2F_{\text{global}}(k)$.

The previous equations show how the conditional similarity and diversity statistics are closely related.

**Overall beta-diversity:** $D_{\text{inter}} = \sum_{k} P(k)D_{\text{inter}}(k) = \sum_{k} P(k)\sum_{l \neq k} P(l/k)D_{\text{inter}}(k,l)$.

Assuming sampling without replacement, we have:
\[ 2 \text{Div}_{-} \beta = \frac{N_e}{N_e - 1} \widetilde{F}_{\text{intra}} \sum_k P(k) + \frac{N_e - 2}{N_e - 1} \sum_k P(k) F_{\text{intra}}(k) - 2 \sum_k P(k) F_{\text{global}}(k) \]

\[ 2 \text{Div}_{-} \beta = \widetilde{F}_{\text{intra}} \left( \frac{2N_e - 2}{N_e - 1} \right) - 2F_{\text{global}} = 2\widetilde{F}_{\text{intra}} - 2F_{\text{global}}. \]

Hence \( \text{Div}_{-} \beta = \widetilde{F}_{\text{intra}} - F_{\text{global}} \). It is easy to check that sampling with replacement also leads to this equation.

This is the well-known result of Nei (1973), which further yields \( \text{Div}_{-} \alpha + \text{Div}_{-} \beta = \text{Div}_{-} \gamma \), i.e., the classical additive apportioning of gamma-diversity into alpha- and beta- fractions (Lande 1996; Couteron and Pélissier 2004; Pélissier and Couteron 2007). But, here, both results are established without any restrictive assumption regarding sampling or sample sizes.
Online Appendix C: Expectation of the time-stationary intrasample similarity $F_{\text{intra}}(k)$

As one may not often be able to assess the species composition at the scale of an entire ecological community, we refer here to a community sample $k$, featuring $n(k)$ individuals. Let $F_{\text{intra}}(k, t)$ be the probability that two selected individuals are conspecific at time $t$. This probability is changing through time due to successive replacements of dead individuals. In agreement with Moran’s model (1958) we assume that a single individual is chosen at random to die at any time step (between time $t$ and time $t + 1$) and a randomly selected pair of individuals may contain or not contain the dead individual with respective probabilities of $2/n(k)$ and $(1 - 2/n(k))$. In the latter case, the probability that the pair is conspecific at time $t + 1$ remains $F_{\text{intra}}(k, t)$, otherwise, using coalescent-type reasoning (Etienne and Olff 2004; Etienne 2005), the replacing individual is either descended from a lineage not represented in the sample at $t$ (new immigrant ancestor), with probability $m(k)$ (see Etienne 2005, Appendix S2), or belongs to a lineage already present in sample $k$ at $t$, with probability $1 - m(k)$.

- If the replacing individual belongs to a new lineage, the probability that the two selected individuals are conspecific at $t + 1$ is $F_{\text{inter}}(\text{pool}, k, t)$, where ‘pool’ refers to the pool of migrants available for sample $k$.

- If the replacing individual does not belong to a new lineage, it can be descended from the second individual of the selected pair (with probability $1/n(k)$), and hence the probability that the two selected individuals are conspecific at time $t + 1$ is 1. Alternatively, the replacing individual can be descended from another lineage present in the sample (with probability $1 - 1/n(k)$), and hence the probability that the selected individuals are conspecific at time $t + 1$ is $F_{\text{intra}}(k, t)$.

In a historical perspective, one may consider two slightly differing versions of Moran’s model.

- In the original model (1958), the dying individual can provide offspring, hence
\[
F_{\text{intra}}(k, t + 1) = F_{\text{intra}}(k, t) \left(1 - \frac{2}{n(k)}\right) + \frac{2}{n(k)} \left(m(k)F_{\text{inter}}(\text{pool}, k, t) + (1 - m(k))\left(\frac{1}{n(k)} + \left(1 - \frac{1}{n(k)}\right)F_{\text{intra}}(k, t)\right)\right)
\]

with \( m(k) = \frac{I(k)}{I(k) + n(k)} \).

- The modified version introduced by Hubbell (1979) and used by followers (Hubbell 1979; Hubbell 2001; Etienne and Olff 2004; Etienne 2005) alternatively assumes that the dying individual does not provide any offspring and that the replacing individual is not descended from the dead one, hence:

\[
F_{\text{intra}}(k, t + 1) = F_{\text{intra}}(k, t) \left(1 - \frac{2}{n(k)}\right) + \frac{2}{n(k)} \left(m(k)F_{\text{inter}}(\text{pool}, k, t) + (1 - m(k))\left(1/(n(k) - 1) + \left(1 - \frac{1}{n(k) - 1}\right)F_{\text{intra}}(k, t)\right)\right)
\]

with \( m(k) = \frac{I(k)}{I(k) + n(k) - 1} \).

At steady state, we should have \( F_{\text{intra}}(k, t + 1) = F_{\text{intra}}(k, t) = F_{\text{intra}}(k) \). Interestingly, one may check that both models provide the same solution, which is:

\[
F_{\text{intra}}(k) = 1 - \frac{I(k)}{I(k) + 1}(1 - F_{\text{inter}}(\text{pool}, k)).
\]

It is to be noted that \( I(k) \) has constant expectation for any sample drawn in a given community, whatever its size, since it is a characteristic feature of the entire community. In contrast, \( m(k) \) is dependent on the size of the sample drawn in the community.
Online Appendix D: Time-stationary $G_{ST}(k)$ for the three-level spatially implicit neutral model

Supposing that migrants in the study area originate from the same source pool, whatever its SAD may be (3L-SINM), then $F_{\text{inter}}(\text{pool}, k) = F_{\text{inter}}(k)$ (Eq. 2, main text). From Appendix C and Eq. 1 (main text), we have:

\[ 1 - F_{\text{global}}(k) = 1 - P(k / k) \left( 1 - \frac{I(k)}{I(k) + 1} (1 - F_{\text{inter}}(\text{pool}, k)) \right) - (1 - P(k / k))F_{\text{inter}}(\text{pool}, k) \]

\[ 1 - F_{\text{global}}(k) = (1 - F_{\text{inter}}(\text{pool}, k)) \left( P(k / k) \frac{I(k)}{I(k) + 1} + (1 - P(k / k)) \right). \]

On the other hand:

\[ F_{\text{intra}}(k) - F_{\text{global}}(k) = F_{\text{intra}}(k) - P(k / k)F_{\text{intra}}(k) - (1 - P(k / k))F_{\text{inter}}(k) = (1 - P(k / k))(F_{\text{intra}}(k) - F_{\text{inter}}(k)) \]

and then, because $F_{\text{inter}}(\text{pool}, k) = F_{\text{inter}}(k)$:

\[ F_{\text{intra}}(k) - F_{\text{global}}(k) = (1 - F_{\text{inter}}(\text{pool}, k))(1 - P(k / k)) \frac{1}{I(k) + 1}. \]

Hence $G_{ST}(k) = \frac{F_{\text{intra}}(k) - F_{\text{global}}(k)}{1 - F_{\text{global}}(k)} = \frac{(1 - P(k / k)) \frac{1}{I(k) + 1}}{P(k / k) \frac{I(k)}{I(k) + 1} + (1 - P(k / k))}$,

which simplifies into: $G_{ST}(k) = \frac{1}{1 + I(k)/(1 - P(k / k))}$. 
Online Appendix E: Estimation of immigration using approximate estimators of similarities

As shown in the main text, there is no notable bias if the exact estimators of similarities (Appendix A) are used to estimate \( \hat{G}_{STR}(k) \) and eventually the immigration rates. But we shall illustrate hereafter that using approximate estimators determined a bias that systematically and substantially widened with expected \( m \) (underestimation, Fig.A1). Although the variance of estimated \( \hat{m}_{GST}(I,k) \) increased with expected \( m \) values in both cases, it was smaller using the approximate rather than the exact estimators, hence our results suggest a trade-off in bias and variance (Fig. A1 is to be compared with Fig. 2 in the main text).

The exact version of \( \hat{F}_{\text{intra}}(k) \) removes the contribution of singleton species in \( k \) through the multiplicative term \( \left( n_s(k) - 1 \right) \) and this may be mostly responsible for the increased variability. Using either exact or approximate \( \hat{F}_{\text{global}}(k) \) had no detectable influence on the results (not shown), as is expected because the main consequence of the choice is to suppress the influence of singleton species at the level of the lumped set of samples, via the term \( (n_s - 1) \).
Literature cited


Figure captions

Figure 1: Two versions of a spatially implicit neutral model, featuring (A) two levels (2L-SINM, conforming to Hubbell 2001) and (B) three levels (3L-SINM, present paper). Available field community samples in a study area receive migrants competing with local offspring at each individual death event. In (A), the migrants are randomly drawn from a large biogeographical source (Hubbell’s metacommunity), while in (B) they are random draws from an intermediate regional level, depicted as a pool of migrants that can itself undergo migration limitation from biogeographical sources and non-neutral processes. The one-way arrows represent the fluxes of migrants that provide new ancestors into the local community lineages (as in Etienne and Olff 2004), measured by $I(k)$, where $k$ refers to a sample in a given community (it is further assumed that each sample belongs to a distinct community).

Figure 2: $G_{ST}(k)$-based estimation, $\hat{m}_{GST}$, minus the corresponding expected value, $m$ (ordinates), according to $m$ (abscissa), for $m_{pool} = 1$ (no migration limitation for the migrant pool, top, A and B) and $m_{pool} = 0.01$ (strong migration limitation for the migrant pool, bottom, C and D), and for $\theta = 50$ (A and C) and 150 (B and D). Exact estimators of similarities (sampling without replacement) are used. Each bold cross notes the value of $\hat{m}_{GST} - m$ corresponding to the mean of $G_{ST}(k)$ over 250 simulations. Dotted lines delineate the confidence envelopes of the difference $\hat{m}_{GST} - m$, corresponding to the 66%-wide quantile interval [0.17; 0.83] of $G_{ST}(k)$.
Figure 3: Estimating immigration numbers using the two-stage maximum likelihood estimation ($\hat{I}_{MLE}$, Munoz et al. 2007) and the $G_{ST}(k)$-based ($\hat{I}_{GST}$) approaches, for (A) communities simulated with parameters ($\theta = 50$, $I$ in $[0, 50]$ and $n(k) = 400$) and (B) fifty one-hectare forest plots from the evergreen forests in the Western Ghats of India. Parameters used for simulations are comparable with those found for the field dataset. In (A) the scale of gray levels expresses the density of results lying in cells. On both figures, the straight line represents equality.

Online Figure A1: $G_{ST}(k)$-based estimation, $\hat{m}_{GST}$, minus the corresponding expected value, $m$, of migration rates (ordinates), according to $m$ (abscissa), and using approximate estimators for $\hat{F}_{intra}(k)$ and $\hat{F}_{global}(k)$ for $\theta = 50$, $m_{pool} = 1$. Each cross notes the value of the difference $\hat{m}_{GST} - m$ obtained for each expected $m$ value (abscissa) and corresponding to the mean of $G_{ST}(k)$ over 250 simulations. Dotted lines delineate the confidence envelopes of the difference $\hat{m}_{GST} - m$, as computed from the 66%-wide quantile interval [0.17; 0.83] of $G_{ST}(k)$. 
Figure 1

(A) 2L-SINM

(B) 3L-SINM

Sampled study area

Upper levels biogeographical processes (speciation)

Community-scale processes

Pool of migrants
Figure 2

(A) \( \hat{m}_{GST} - m \)

(B) \( \hat{m}_{GST} - m \)
Figure 3

(A)

(B)
Figure A1

\[ \hat{m}_{\text{GST}} - m \]