Ecology Letters, (2009) **12**: 239–248 doi: 10.1111/j.1461-0248.2008.01280.x

LETTER

Inferring the parameters of the neutral theory of biodiversity using phylogenetic information and implications for tropical forests

Franck Jabot* and Jérôme Chave
Laboratoire Evolution et Diversité Biologique, CNRS, Université Paul Sabatier, Bâtiment 4R3, 31062 Toulouse, cedex 4, France
**Correspondence: E-mail: franck.jabot@m4x.org*

Abstract

We develop a statistical method to infer the parameters of Hubbell's neutral model of biodiversity using data on local species abundances and their phylogenetic relatedness. This method uses the approximate Bayesian computation (ABC) approach, where the data are summarized into a small number of informative summary statistics. We used three statistics: the number of species in the sample, Shannon H index of evenness and Shao and Sokal's B_1 index of phylogenetic tree imbalance. Our approach was found to outperform previous methods, illustrating the potential of ABC methods in ecology. Applying it to four large tropical forest tree data sets, the best-fit immigration rates m were found to be two orders of magnitude smaller and regional diversities θ larger than previously reported for the same data. This implies that neutral-compatible regional pools of tropical trees should extend over continental scales, and that m measures, in this context, mostly the frequency of long-distance dispersal events.

Keywords

Approximate Bayesian computation, community phylogenetics, dispersal limitation, neutral theory, parameter inference, phylogenetic imbalance, regional species pool.

Ecology Letters (2009) 12: 239–248

INTRODUCTION

As phylogenetic trees become more widely available through molecular methods, ecologists have an invaluable new dimension of information with which to investigate the mechanisms of community assembly (Losos 1992; Webb *et al.* 2002; Pennington *et al.* 2006). It has long been suggested that ecological processes may have a distinctive fingerprint on the patterns of evolutionary relatedness of coexisting species (Hutchinson 1959; Losos 1992; Wiens & Donoghue 2004). However, most models of species coexistence fail to take into account the phylogenetic structure of a species assemblage, as they are usually restricted to the scale at which individuals interact physically, thus neglecting larger spatial and temporal scales (Ricklefs 2004). One exception is Hubbell's (2001) neutral theory of biodiversity and biogeography, which includes both local and regional processes in a single conceptual framework. Hubbell's neutral model assumes that a local community is connected to a larger pool of species through dispersal, much like in classic island biogeography (MacArthur & Wilson 1967). A single parameter θ, the number of species arising per generation through speciation in the regional species pool, summarizes the diversity in this pool. A second parameter m, the fraction of recruits coming from the source pool into the local community, describes the magnitude of dispersal limitation in the local community.

In many of the recent applications of the neutral theory, the species abundance distribution of a local assemblage has been used to test the neutral theory. Of particular relevance here is the work of Latimer *et al.* (2005), who assessed the evolutionary consequences of the neutral assumption by analysing plant species abundance distributions in the South African fynbos. They inferred the neutral parameters and found neutral migration rates to be two orders of magnitude smaller than those found in tropical forest tree communities. Latimer *et al.* (2005) also found that parameter θ, as inferred from their data set, was much higher than in previous reports. They interpreted this result as a signature of a peculiarly high speciation rate of the fynbos flora. However, Etienne *et al.* (2006) critically reassessed this finding. They commented that the inferred neutral parameters θ and m often have two nearly equally likely values when they are inferred from species abundance data. For the tropical

forest trees of Barro Colorado Island (BCI), Etienne *et al.* (2006) found that the maximum likelihood estimation (MLE) of the neutral parameters (θ, m) yields two different but almost equally likely maxima. This finding is a serious theoretical challenge for the neutral theory, because it suggests that there is no way to estimate accurately these parameters. For the BCI tree data set, it has been assumed that the parameters are close to $\theta = 47.7$ and $m = 0.093$ (Leigh 2007). The second combination of neutral parameters ($\theta = 241.9$, $m = 0.003$, Etienne *et al.* 2006) would imply that the BCI forest is far more dispersal limited, and that the regional diversity is much higher than previously imagined.

Aside from Latimer *et al.* (2005), attempts to examine the neutral theory at evolutionary timescales are scarce (but see, e.g. Lande *et al.* 2003; Lavin *et al.* 2004). Many studies of macroevolutionary patterns are based on simple models of lineage diversification, such as the Yule model, which produces tree topologies assuming that all species have the same chance to give rise to an altogether new species (Yule 1924). Simulated trees are often compared with empirical ones using the balance of the tree topology, i.e. the extent to which nodes define subgroups of equal size (balanced trees are also called 'symmetrical', while imbalanced ones are sometimes called 'pectinate', see Mooers & Heard 1997). It turns out that the Yule model produces trees that are generally more balanced than those reconstructed from biological data (Mooers & Heard 1997). Recently, Mooers *et al.* (2007) used Hubbell's neutral model to produce simulated trees, and they found that it produced trees less balanced than empirical ones. Hubbell (2001) had already noticed that larger θ values were associated with a more even distribution of speciation events among lineages; hence, neutral trees simulated with larger θ should be more balanced. Thus, there are reasons to believe that the trees produced by Hubbell's model encompass a range of balance values, if the parameter θ is free to vary more than in Mooers *et al.*'s (2007) study. Turning this argument around, we speculate that the balance of real phylogenetic trees may provide a direct way to infer the neutral parameter θ, independently from species abundance distributions.

Here, we develop a new sampling theory of Hubbell's neutral model, which takes into account not only the species abundance distribution in a sample, but also the phylogenetic relatedness of these co-occurring species. We first assess whether phylogenetic trees predicted by Hubbell's neutral model are a reasonable fit for real phylogenies, in particular with respect to phylogenetic imbalance. Next, we use simulated data to show that phylogenies enable us to infer the neutral parameters more precisely. We finally apply our new inference method to tropical tree data sets in the Neotropics and in South-East Asia. New parameter estimates are reinterpreted in the light of dispersal biology and biogeographical patterns.

METHODS

Approximate Bayesian computation

In a classical likelihood framework, inferring parameters of the neutral model using phylogenies would entail the derivation of an exact likelihood function for these data. Currently, there is no simple formula for such a likelihood function. Instead, it is possible to approximate this function through simulation, using a method called approximate Bayesian computation (ABC). This method is commonly used in population genetics (Tavaré *et al.* 1997; Beaumont *et al.* 2002; Marjoram & Tavaré 2006), but to our knowledge, has not been employed in ecology.

The ABC method is useful when the exact likelihood function of a model can be approximated by a large number of independent simulations, across a wide range of model parameter values drawn from a set of prior distributions. Empirical data are summarized into a set of informative statistics called summary statistics. These summary statistics are then computed for each of the simulated data sets and compared with the values observed in the empirical data set. Only the simulations whose summary statistics are close enough to the observed values are retained. The corresponding parameter values, in our case θ and m, then form an approximate joint posterior distribution for the parameters. A computer-intensive approach like the ABC method relies on the ability to quickly simulate samples under the model considered. For Hubbell's neutral model, this is made possible by using the coalescent approach (Wakeley 2007). Details on our simulation algorithm are provided in Appendix S1.

Choice of summary statistics

Various diversity indices may be used to summarize the species abundance distribution (Magurran 2004). Likewise, phylogeny-based diversity indices, and statistics describing the tree topology may be used to summarize the phylogeny of species occurring locally in the community (Mooers & Heard 1997). We included a total of 24 candidate summary statistics in our preliminary tests (Table S1). To assess which statistics were the most informative, we simulated 1 000 000 local communities each of size $J = 20\,000$, where J is the number of individuals in the local community. We sampled the neutral parameters with a uniform prior distribution on $\ln(\theta)$ ($0 \leq \ln(\theta) \leq 7$) and on $\ln(I)$ ($0 \leq \ln(I) \leq 10$), where I is the scaled immigration rate defined as $I = m(J - 1)/(1 - m)$ (Etienne 2005). These priors contain the range of neutral parameters that have been previously estimated for tropical forests (Chave *et al.* 2006).

The number of species S in the local community sample was retained as the first summary statistic, because in the limiting case were $m = 1$, S is a sufficient statistic for θ

(Ewens 1972). We then tested which of the other statistics were the most informative in predicting the neutral parameters. Constraining the simulation outputs to a fixed value of S (in practice, for $S = 200$ species), we regressed the 23 remaining summary statistics against $\ln(\theta)$. All statistics of tree imbalance monotonically correlated with $\ln(\theta)$; hence, we compared their relative predictive power based on the regression coefficient of a linear model. The summary statistics for the species abundance distribution presented a mode for intermediate values of $\ln(\theta)$ (Fig. 1). We then compared their predictive performance based on the adjusted-r^2 of a quadratic model. Among these statistics, two were found to be the most correlated with $\ln(\theta)$ (see Results): the Shannon's index of diversity H defined as $H = -\sum_i p_i \ln(p_i)$, where p_i is the relative abundance of species i in the sample and Shao & Sokal's (1990) phylogenetic tree balance statistic B_1 defined as $B_1 = \sum_i 1/M_i$, where M_i is the maximal number of nodes between the interior node i and the terminal species of the subtree rooted at node i, the summation being on all interior nodes except the root (Shao & Sokal 1990). More balanced phylogenetic trees have higher B_1 values. The definition of the other trial statistics is reported in Table S1. Regression of the statistics against $\ln(I)$ would be very similar to those against $\ln(\theta)$, because the two parameters θ and I are strongly negatively correlated.

We also assessed whether the range of phylogenetic imbalance that the neutral theory is able to predict encompassed realistic values in phylogenetic imbalance. We retrieved the first 2000 published phylogenies in TREEBASE (http://www.treebase.org/), measured the imbalance statistics B_1 and compared this statistics to that obtained from phylogenies simulated with Hubbell's meta-community model (see Appendix S3 for more details).

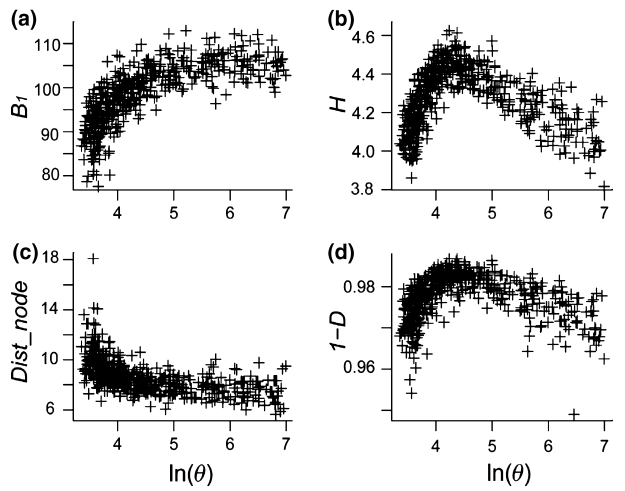

Figure 1 Four summary statistics in simulated data sets for different values of the regional diversity index, $\ln(\theta)$. The goodness of fit between the summary statistics and $\ln(\theta)$ was measured by the adjusted r^2 and the retained statistics correspond to the best fit results. (a) Shao and Sokal's B_1 imbalance index. (b) Shannon's diversity index H. (c) Number of nodes in the phylogeny between two randomly chosen individuals Dist_Node. (d) Simpson's index $1 - D$.

Test of the ABC method using simulated data

It is impossible to ensure that the chosen set of summary statistics is optimal to infer the parameter in the ABC method (Marjoram *et al.* 2003). However, it is possible to check that the selected summary statistics do summarize most of the information present in the data and thus lead to an efficient estimation method. We simulated 300 neutral data sets with various parameter values and then estimated the neutral parameters by ABC. More precisely, we computed a mean standard error and a bias on the estimated parameters, defined as

$$\mathrm{MSE} = \left(\left\langle \frac{(\ln(\theta_{\mathrm{estimated}}) - \ln(\theta_{\mathrm{simul}}))^2}{\ln^2(\theta_{\mathrm{simul}})} \right\rangle\right)^{1/2} + \left(\left\langle \frac{(\ln(I_{\mathrm{estimated}}) - \ln(I_{\mathrm{simul}}))^2}{\ln^2(I_{\mathrm{simul}})} \right\rangle\right)^{1/2}$$

$$\mathrm{Bias} = \left\langle \frac{\left|\ln(\theta_{\mathrm{estimated}}) - \ln(\theta_{\mathrm{simul}})\right|}{\ln(\theta_{\mathrm{simul}})} \right\rangle + \left\langle \frac{\left|\ln(I_{\mathrm{estimated}}) - \ln(I_{\mathrm{simul}})\right|}{\ln(I_{\mathrm{simul}})} \right\rangle$$

where the brackets represent a mean over the 300 simulated communities. Both statistics describe the estimation efficiency of the ABC method. For comparison, we also computed these statistics with the estimates obtained from an exact maximum-likelihood approach based on the species abundance distribution only (Etienne 2005).

Application to four tropical forest tree data sets

The ABC method was used to infer the neutral parameters from four large tropical forest tree data sets (data from Condit *et al.* 2006). These data sets correspond to a full census of trees greater than 10 cm in trunk diameter at breast height (dbh) in plots of 25–52 ha in size, in central Panama (BCI), Colombia (La Planada), peninsular Malaysia (Pasoh) and Malaysia, Sarawak (Lambir). More details on these study sites may be found in Losos & Leigh (2004).

For each site, a maximally resolved phylogenetic tree subtending the local community was generated using the software PHYLOMATIC (Webb & Donoghue 2005). For the BCI site, we also included additional published data to

produce an improved phylogenetic tree (see Appendix S2). For the BCI phylogeny, about 69% of nodes were resolved in this improved phylogenetic tree, compared with 53% with the default options of PHYLOMATIC. The remaining polytomies were resolved randomly.

For each data set, we ran the ABC method using 200 000 simulated species assemblages, with a uniform prior distribution for $\ln(\theta)$ $(0 \leq \ln(\theta) \leq 10)$ and for $\ln(I)$ $(0 \leq \ln(\mathbf{I}) \leq 10)$. To control for the possible bias due to the incomplete resolution of the phylogenies, we repeated the random resolution procedure 100 times (see Discussion). For each of the 100 random resolutions of the polytomies in the observed phylogenetic tree, we selected the 200 outputs for which the simulated values of S, H and B_1 were the closest to the real ones – by taking the euclidean distance in the space of summary statistics (S, H, B_1). We thus obtained $200 \times 100 = 20\,000$ points in the plane $(\ln(\theta),\ln(I))$, from which we computed an approximate posterior distribution.

Analyses using Etienne's (2005) MLE were performed using the TETAME freeware (http://www.edb.ups-tlse.fr/equipe1/chave/tetame.htm). Post-processing of the ABC simulations (determination of the approximate posterior distribution and mode) were carried out with the R software version 2.7.0 (http://www.r-project.org/), using the routine 'bkde2D' of the library 'KernSmooth'. All R scripts are available upon request.

RESULTS

We found that the summary statistic of species abundances with the largest correlation with $\ln(\theta)$ was Shannon's index H, a measure of the evenness of the species abundance distribution. The phylogenetic summary statistic with the largest correlation with $\ln(\theta)$ was found to be Shao & Sokal's (1990) B_1 imbalance statistic, which measures the level of symmetry in the phylogenetic tree subtending the local community (Fig. 1). All statistics based on branch lengths poorly correlated with $\ln(\theta)$ and were thus found to be uninformative in our inferential framework. Our simulations showed that large θ values lead to more balanced phylogenetic trees (larger B_1 values) than small θ values (Fig. 1). We also found that observed levels of phylogenetic imbalance of the 2000 published trees examined were always within the range of Hubbell's model predictions (Fig. 2).

When only species abundance data were used in the ABC method – by using solely the summary statistics S and H, we found that inference by ABC was nearly as efficient as Etienne's (2005) exact MLE method. The mean standard error on the parameters was equal to 30% with the ABC method, when compared with 25% for the MLE method. However, our inference method was more biased that the MLE method (Bias = 22% with the ABC method versus 9% with the MLE method).

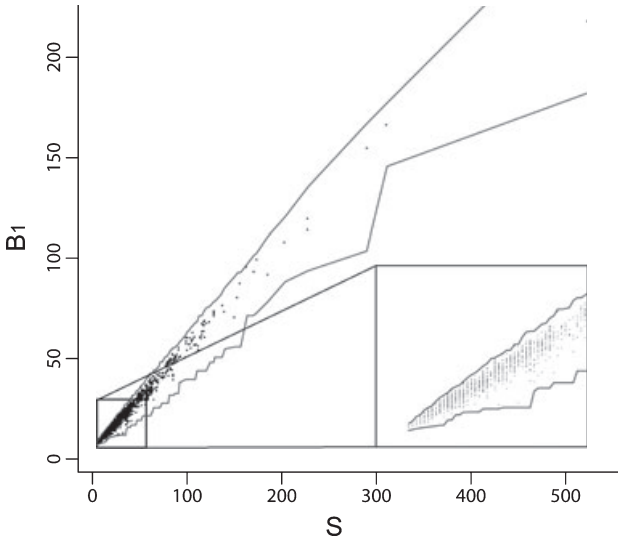

Figure 2 Compatibility of 2000 published phylogenies with Hubbell's neutral theory in terms of phylogenetic tree shape. Phylogenetic tree shape is measured by the statistic of imbalance B_1. Each dot represents a published phylogeny. The lines correspond to minimum and maximum B_1 values obtained in simulated neutral phylogenies.

By contrast, when we estimated the neutral model parameters with the ABC method using all three statistics S, H and B_1 – i.e. including information on the phylogenies, the mean standard error of the inferred parameters with ABC was equal to 17% (versus 25% with the MLE) and the bias was of 2% (versus 9% with the MLE). Hence, phylogenies do add relevant information that improves the quality of parameter inference.

Based only on species abundance data, the likelihood function for the BCI data set has two alternative likelihood maxima, the low θ and high m value being slightly more likely (Fig. 3a, Etienne *et al.* 2006). Similarly, the ABC method yields two alternative maxima when based on the two species abundance statistics (Fig. 3b). By contrast, our new method based on all three summary statistics unambiguously selects the high θ and low m values (Fig. 3c,d). The parameter θ that was selected is one order of magnitude larger than in previous analyses, and the selected m parameter is two orders of magnitude smaller (Table 1). We were able to test whether this result was sensitive to the resolution of the phylogeny with the BCI data set, as a more refined phylogenetic hypothesis is available for this site. We found that our result was independent of the choice of the phylogeny (Fig. 3c,d).

The ABC method was also applied to the La Planada, Pasoh and Lambir data sets (Fig. 4a–c respectively). For all three data sets, the ABC method yielded a single MLE of the neutral parameters. In La Planada, the most likely value of θ

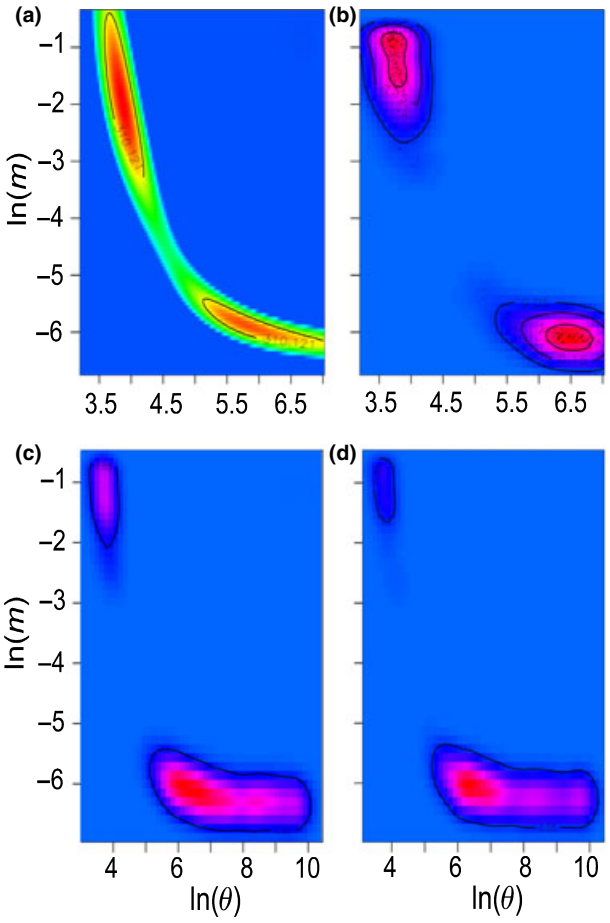

Figure 3 Posterior distributions of the neutral parameters for the BCI tropical tree data set. (a) Likelihood profile for Hubbell's neutral model based on species abundances only and based on Etienne's sampling formula. Likelihood values are colour coded. Solid lines: 95% confidence limits of the parameters. (b) Posterior distribution for the neutral model using the ABC method with species abundances only (i.e. based on the two summary statistics S and H only). Solid lines: density levels (approximate 95% confidence intervals). (c) Posterior distribution for the neutral model using the ABC method with both species abundances and a phylogenetic hypothesis based on the angiosperm phylogeny group. In this phylogeny, 53% of the nodes are resolved. Three summary statistics, S, H and B_1 were used. (d) Same as in (c) but with a better resolved phylogeny where 69% of the nodes are dichotomous.

was 345 versus 30 for the MLE and m was equal to 0.003 versus 0.28 for the MLE (Table 1). In Pasoh and Lambir, the values of θ were fourfold larger than in the MLE and those of m consistently equal to 0.01 (Table 1). In sum, the addition of phylogenetic information to infer the parameters of Hubbell's model led to strikingly larger values for θ and lower values for m when compared with previous inference methods, in all four tropical forest data sets tested here.

DISCUSSION

Neutrality and parameter inference

Etienne (2005) previously showed that the neutral model of biodiversity is endowed with an exact sampling theory, like its counterparts in population genetics (Ewens 2004; Wakeley 2007). This sampling theory relates the full species abundance distribution of one community sample to the neutral model parameters. However, the species abundance distribution contains a limited amount of information and it is not sufficient to jointly estimate both parameters (Etienne *et al.* 2006; Fig. 5). Using simulated data, we first showed that the ABC method based on species abundance only provides results almost as good as Etienne's (2005) exact MLE. The great advantage of the ABC method is that it is easily amenable to generalizations, through the addition of additional summary statistics. Adding phylogenetic information via the imbalance statistic B_1, we inferred the neutral parameters of simulated data sets more precisely than any previous inference method. As B_1 is monotonously related to the neutral parameters, it was successful at discriminating between the two alternative maxima in the likelihood profile (Fig. 5).

More generally, the challenge of estimating the parameters of ecological models based on real data has motivated much recent research (Clark 2005). To our knowledge, our study is the first to make use of ABC to solve an ecological problem. Widely used in population genetics (Marjoram & Tavaré 2006), such computer-intensive inference techniques can allow complex models to be investigated. This should provide an opportunity to expand the range of data used in tests of ecological theories (McGill *et al.* 2007).

Neutrality and phylogenetic imbalance

Hubbell's neutral model has often been rejected outright because it was felt that it is much too simplistic to even remotely reflect the complex evolutionary dynamics of species assemblages (e.g. Ricklefs 2006). Although our work does not address this point directly (see Lande *et al.* 2003; Allen & Savage 2007), we found that Hubbell's model is able to predict phylogenetic tree imbalance. This suggests that, although crude, this model may be sufficient to capture basic features of the speciation–extinction balance. Classic diversification models like the Yule and Hey models, which assume an equal rate branching probability among lineages, predict consistently too balanced phylogenies (Mooers & Heard 1997). By contrast, Hubbell's model, which assumes that the branching probability of a lineage is proportional to its abundance, produces phylogenetic balance consistent with observed ones (Fig. 2, Appendix S3). This suggests that the neutral theory's assumption of a speciation rate proportional to species abundance might be a less crude diversification model than the Yule model (Webb & Pitman

Table 1 Neutral parameter estimates inferred from local species abundances and phylogeny in four tropical forest plots: Barro Colorado Island (BCI, Panama), La Planada (Colombia), Pasoh and Lambir (Malaysia)

Site	J	S	θ_1	m_1	θ_2	m_2	W	θ_E	m_E
BCI*	20 788	236	724	0.002	43	0.32	0.78	48	0.14
BCI†	20 788	236	571	0.002	44	0.36	0.88	48	0.14
La Planada	14 100	164	345	0.003	31	0.39	0.8	30	0.28
Pasoh	29 257	674	534	0.01	–	–	1	194	0.07
Lambir	29 890	990	2491	0.008	–	–	1	282	0.13

J and S are the number of individuals and the number of species in the sample respectively. The neutral parameters estimated by the ABC method are (θ_1, m_1). A lower peak is often observed in posterior distributions, whose values are (θ_2, m_2). The relative weight of the mode (θ_1, m_1) compared with the other peak is W, defined as the number of ABC simulations in the confidence interval of the mode divided by the total number of retained ABC simulations. (θ_E, m_E) are the estimates obtained by the exact likelihood formula that uses only species abundances (Etienne 2005).

*Phylogeny based on the compilation of phylogenies made by the software PHYLOMATIC. In this tree, 53% of the nodes are resolved.

†Phylogeny based on additional compilation of phylogenies for species encountered in BCI. In this revised tree, 69% of the nodes are resolved.

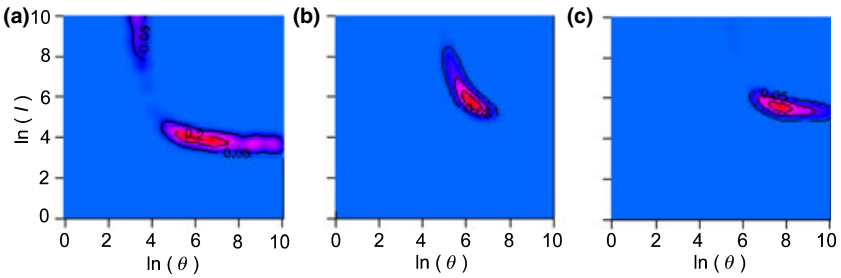

Figure 4 Posterior distributions of the neutral parameters in three tropical forest sites using the ABC method. (a) Posterior distribution of the neutral parameters at La Planada. Solid lines: the density levels (approximate 95% confidence intervals). (b) Posterior distribution of the neutral parameters at Pasoh. (c) Posterior distribution of the neutral parameters at Lambir.

2002). As in Hubbell's model regional pools with larger θ have more even regional species abundances (Hubbell 2001), the corresponding phylogenies are more balanced (Fig. 6). This explains why Mooers *et al.* (2007) found that Hubbell's model predicted too imbalanced phylogenies compared with the observed ones, as they only used a small value of θ ($\theta = 10$) in their simulations.

Our result sheds light on studies of the phylogenetic tree shape in other species groups. For instance, Heard & Cox (2007) recently compared primate phylogenies across continents and they found that the phylogeny of New World primates was more balanced than that of Old World primates. They explained this pattern as a consequence of different biogeographic histories: repeated speciation events in connection with stepwise dispersal, or massive extinctions, may lead to less balanced phylogenies. Conversely, vicariance events should lead to more balanced phylogenies. However, it may also be argued that South America is a richer regional pool of primate species ($S = 80$) than Asia ($S = 57$), Africa ($S = 52$) and Madagascar ($S = 28$). Hence, the more balanced phylogeny observed in South America when compared with other regional phylogenies is consistent with neutral expectations, even in the absence of differential speciation or extinction mechanisms.

Regional assembly of tropical rainforest trees

We found values of θ in four large tropical forest tree plots that were up to one order of magnitude larger than estimates based on previous methods (Table 1). In Hubbell's model, θ is the product of the regional pool size and of the speciation rate. Larger values of θ therefore mean that the regional pool size is larger than previously thought, or that the speciation rate is larger. Latimer *et al.* (2005) found remarkably comparable values of θ in the South African fynbos that they studied. To interpret their result, they reasoned that the regional pool of the fynbos is roughly the Cape Floristic Region, which extends over 50 000 km^2. This extent leads to a regional pool size of about 1.3×10^{11} individuals (Appendix S4). Using the same logic, our new estimates of θ imply that the regional pools of our neotropical tree plots should extend over areas of the size

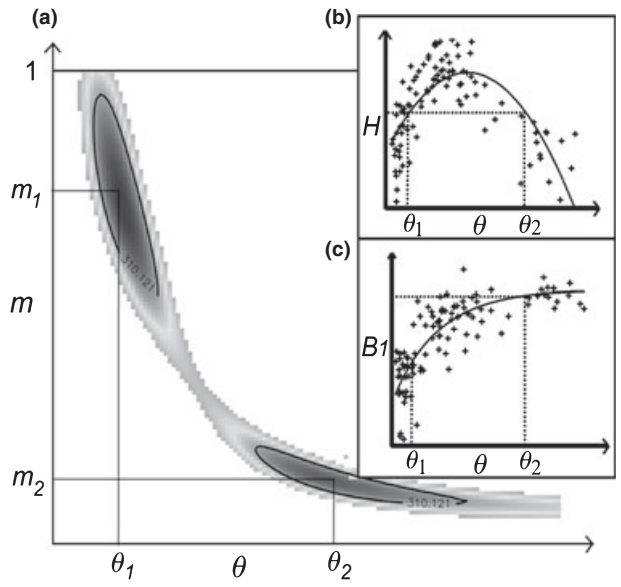

Figure 5 Phylogenies improve the estimation of the neutral theory's parameters. (a) Exact likelihood profile of the neutral parameters using the BCI tree species abundance data (see Fig. 2a). Parameter inference leads to a confidence interval containing two equally likely maxima: one at low θ and high m, the other at high θ and low m. (b) Variation of the evenness (measured by Shannon's H) in simulated neutral communities of the same size as in the BCI data set. The species richness S determines a maximum-likelihood ridge in the parameter space (θ, m). The evenness H contains additional information about the position of the most likely parameters along this ridge. However, as this evenness is unimodally correlated with θ, two parameter combinations yield the same evenness. (c) The phylogenetic tree imbalance statistic B_1 of a local community is positively related to θ. A combination of the statistics H and B_1 yields a single most likely parameter set (θ, m).

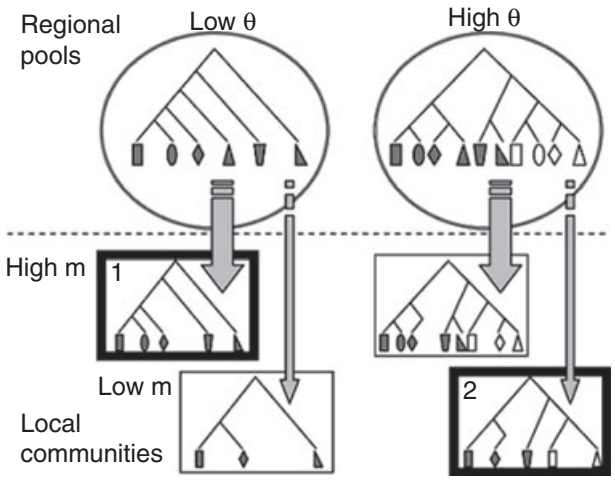

Figure 6 Schematic depiction of the information contained in phylogenies. Each species is denoted by a different symbol. Regional pools are connected to local communities by immigration (grey arrows). When θ is large, regional pools are species rich and have more balanced phylogenies. Based on species abundances only, communities 1 and 2 cannot be distinguished. However, the phylogenetic imbalance of community 1 is greater than that of community 2.

of the entire Neotropics, and the South-East Asian tree pool should extend over an area of the order of the former Sunda Shelf (Appendix S4). An alternative interpretation would be that speciation rates should be extraordinarily high for trees. While limited evidence would support this claim in a few species-rich groups (Richardson *et al.* 2001), this pattern does not seem to hold universally across tropical plant lineages (Pennington & Dick 2004; Pennington *et al.* 2006).

Is a continental extent for the regional pool of tropical trees a biologically sound inference? By definition, the regional pool of an ecological community is the ensemble of species likely to immigrate into the local community. In tropical forests, long-distance dispersal events have been reported based on floristic evidence (Pennington & Dick 2004) and using molecular tools (Dick *et al.* 2008). Although rare, these long-distance dispersal events stir tropical forest pools over wide geographical scales. Further, a regional species pool extending over continental scales is consistent with the fact that numerous Amazonian tree species have a wider distribution than previously thought. For instance, many of the tree species in the family Sapotaceae that were previously reported as having a narrow distribution are now recognized as being pan-Amazonian species (T.C. Pennington, personal communication). Finally, we found comparable values of θ across sites within the same continent, suggesting that these sites indeed share the same pool, and these values were also comparable across continents suggesting that their regional diversity in tree species is comparable as confirmed by independent evidence (Gentry 1988; Fine & Ree 2006). By contrast, previous estimates of θ were one order of magnitude larger in Asia than in South America (Chave *et al.* 2006).

A possible limitation in our analysis is due to the fact that the phylogenies used for parameter inference were not fully resolved. In order to use our inference method, we had to resolve these phylogenies randomly. This may have lead to a bias towards high θ values, because randomly branched trees are more balanced than real ones (Mooers & Heard 1997). However, this potential bias is unlikely to lead to the high estimated value of θ because increasing the resolution of the phylogeny at BCI did not increase the probability of selecting the low θ peak (Table 1, Fig. 3c,d).

Local assembly of tropical rainforest trees

Another finding of our study is that the immigration rate m is much smaller than previously reported for tropical forests (Table 1). This result is a direct consequence of the large

regional diversities θ measured with our new method. If the regional pool has more species, then the local community has to be more dispersal limited from this pool to maintain the same level of local diversity. Does this result make sense biologically? To answer this question, we first emphasize that, in Hubbell's model, an immigration event is not equivalent to an observed immigration event in real continuous landscapes (Alonso *et al.* 2006). In Hubbell's model, immigration events come from anywhere in the regional pool; so, parameter m measures the amount of sampling of this regional pool. In real landscapes, immigration events mostly come from close surroundings of the focal community, and only a small fraction of the regional pool is actually sampled by short-distance dispersal. By contrast, long-distance dispersal events are contributed by the entire regional pool, as assumed in Hubbell's model. Consequently, in real landscapes, one must distinguish short-distance immigration events which constitute the bulk of immigration but do not contribute much to the sampling of the regional pool, and long-distance immigration events which, while rare, are likely to contribute much more to the sampling of the regional pool (Nathan 2006). In this context, our estimate of m predicts that long-distance dispersal events contribute at best 0.2–1% of the within-site recruitment.

The consistency between the neutral parameter m and field data have often made use of average dispersal distances inferred from seed trap data, i.e. short-distance dispersal. Using seed trap counts on BCI, a cross-species mean seed dispersal distance from the parent tree to the propagule's arrival site was estimated to be 39 m (Condit *et al.* 2002). Using this value of mean dispersal distance, Etienne (2005) computed m as the proportion of seeds in the plot coming from outside the plot, and he found that this parameter should be close to 0.1 with a Gaussian seed dispersal kernel. However, as mentioned above, local dispersal should be a poor predictor of the immigration rate m. Hence, the apparent contradiction between average dispersal distances measured by seed trap data and our estimates of m is simply resolved by the fact that the latter measures long-distance dispersal. This quantity is of crucial importance in the context of global change, because long-distance dispersal is what will determine the overall ability of tropical forest species to track environmental changes.

Perspectives

Deviations from the point-wise mutation model assumed here may also contribute to the observed patterns of phylogenetic tree balance. If the new assumption is that a new lineage starts with more than one individual, like in Hubbell's fission model where population are randomly split into two during speciation events (Hubbell 2001), then phylogenetic balance will be higher than with the point-wise mutation model (Mooers *et al.* 2007). Unfortunately, models of speciation with non-point-wise mutation do not possess a simple mapping with a coalescent; so, the present approach cannot be straightforwardly extended to more general speciation models. We hope to return to this question in the future.

Our study paves the road between community modelling and studies of phylogenetic structure. This theme has received a great deal of attention in the recent literature and tests have been devised to compare the phylogenetic structure of local species assemblages to randomly assembled (null) communities from a species pool (Webb 2000; Webb *et al.* 2002). As an illustration, Webb (2000) assumed that the species pool was simply the sum of all species encountered in a surrounding area, because these were considered as potential immigrants into the focal community. Yet, this approach strongly depends on the size of the hypothetical regional pool (Swenson *et al.* 2006) and on the choice of the test statistics (Hardy & Senterre 2007). Further, it makes no use of local species abundances, although species abundances may be informative for testing ecological mechanisms.

A significant improvement of these tests requires building a null theory based at the individual level, rather than at the species level. This theory needs to be endowed with a proper sampling theory, making no explicit reference to a regional pool, for which information on abundances is seldom available. In addition, it needs to take into account the consequences of dispersal limitation on species abundances. Finally, it needs to incorporate demographic stochasticity and to be based on several sampling units. Our work supports the view that the dispersal-limited neutral theory may be used in this research programme (Pennington *et al.* 2006). We could extend it to include multiple plots simultaneously (Jabot *et al.* 2008). Then, by looking at patterns that have not been used to fit the model parameters, such as phylogenetic and taxonomic similarity, one could assess the biological relevance of additional non-neutral processes. Such a model would provide consistent null scenarios to test the hypothesis of community phylogenetics.

ACKNOWLEDGEMENTS

We thank Lounès Chikhi for sharing his expertise on ABC methods, and for comments on a previous version of this manuscript. We also thank Mark Beaumont, Michaël Blum, Nathan Kraft and Christophe Thébaud for useful comments on a previous version of this manuscript. We thank the Editor and three anonymous referees for suggestions that greatly improved this article. We are indebted to the Center for Tropical Forest Science and the numerous field workers who produced the empirical datasets used in this article. FJ was funded by the French Ministry of Agriculture. This work was funded by the ANR-Biodiversité grant BRIDGE,

by a CNRS-AMAZONIE grant and by the Egide Alliance grant no. 12130ZG.

REFERENCES

Allen, A.P. & Savage, V.M. (2007). Setting the absolute tempo of biodiversity dynamics. *Ecol. Lett.*, 10, 637–646.

Alonso, D., Etienne, R.S. & McKane, A.J. (2006). The merits of neutral theory. *Trends Ecol. Evol.*, 21, 451–457.

Beaumont, M.A., Zhang, W.Y. & Balding, D.J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, 162, 2025–2035.

Chave, J., Alonso, D. & Etienne, R.S. (2006). Theoretical biology – comparing models of species abundance. *Nature*, 441, E1.

Clark, J.S. (2005). Why environmental scientists are becoming Bayesians. *Ecol. Lett.*, 8, 2–14.

Condit, R., Pitman, N., Leigh, E.G. Jr, Chave, J., Terborgh, J., Foster, R.B. *et al.* (2002). Beta-diversity in tropical forest trees. *Science*, 295, 666–669.

Condit, R., Ashton, P., Bunyavejchewin, S., Dattaraja, H.S., Davies, S., Esufali, S. *et al.* (2006). The importance of demographic niches to tree diversity. *Science*, 313, 98–101.

Dick, C.W., Hardy, O.J., Jones, F.A. & Petit, R.J. (2008). Spatial scales of pollen and seed-mediated gene flow in tropical rain forest trees. *Trop. Plant Biol.*, 1, 20–33.

Etienne, R.S. (2005). A new sampling formula for neutral biodiversity. *Ecol. Lett.*, 8, 253–260.

Etienne, R.S., Latimer, A.M., Silander, J.A. & Cowling, R.M. (2006). Comment on "Neutral ecological theory reveals isolation and rapid speciation in a biodiversity hot spot". *Science*, 311, 610b.

Ewens, W.J. (1972). The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.*, 3, 87–112.

Ewens, W.J. (2004). *Mathematical Population Genetics. I. Theoretical Introduction*, 2nd edn. Springer, New York, 417 pp.

Fine, P.V.A. & Ree, R.H. (2006). Evidence for a time-integrated species-area effect on the latitudinal gradient in tree diversity. *Am. Nat.*, 168, 796–804.

Gentry, A.H. (1988). Changes in plant community diversity and floristic composition on environmental and geographical gradients. *Ann. Miss. Bot. Gard.*, 75, 1–34.

Hardy, O.J. & Senterre, B. (2007). Characterizing the phylogenetic structure of communities by an additive partitioning of phylogenetic diversity. *J. Ecol.*, 95, 493–506.

Heard, S.B. & Cox, G.H. (2007). The shapes of phylogenetic trees of clades, faunas, and local assemblages: exploring spatial pattern in differential diversification. *Am. Nat.*, 169, E107–E118.

Hubbell, S.P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton, NJ.

Hutchinson, G.E. (1959). Homage to Santa Rosalia or why are there so many kinds of animals? *Am. Nat.*, 93, 145–159.

Jabot, F., Etienne, R.S. & Chave, J. (2008). Reconciling neutral community models and environmental filtering: theory and an empirical test. *Oikos*, 117, 1308–1320.

Lande, R., Engen, S. & Saether, B.-E. (2003). *Stochastic Population Dynamics in Ecology and Conservation*, Oxford Series in Ecology and Evolution. Oxford University Press, Oxford UK, 212 pp.

Latimer, A.M., Silander, J.A. Jr & Cowling, R.M. (2005). Neutral ecological theory reveals isolation and rapid speciation in a biodiversity hot spot. *Science*, 309, 1722–1725.

Lavin, M., Schrire, B.P., Lewis, G., Pennington, R.T., Delgado-Salinas, A., Thulin, M. *et al.* (2004). Metacommunity process rather than continental tectonic history better explains geographically structured phylogenies in legumes. *Philos. Trans. R. Soc. Lond. B*, 359, 1509–1522.

Leigh, E.G. Jr (2007). Neutral theory: a historical perspective. *J. Evol. Biol.*, 20, 2075–2091.

Losos, J.B. (1992). The evolution of convergent structure in Caribbean anolis communities. *Syst. Biol.*, 41, 403–420.

Losos, E.C. & Leigh, E.G. Jr (2004). *Tropical Forest Diversity and Dynamism. Findings from a Large-Scale Plot Network*. University of Chicago Press, Chicago, IL.

MacArthur, R.H. & Wilson, E.O. (1967). *The Theory of Island Biogeography*. Princeton University Press, Princeton, NJ, 224 pp.

Magurran, A.E. (2004). *Measuring Biological Diversity*. Blackwell, Oxford, UK, 256 pp.

Marjoram, P. & Tavaré, S. (2006). Modern computational approaches for analyzing molecular genetic variation data. *Nat. Rev. Gen.*, 7, 759–770.

Marjoram, P., Molitor, J., Plagnol, V. & Tavaré, S. (2003). Markov chain Monte Carlo without likelihoods. *Proc. Natl Acad. Sci. USA*, 100, 15324–15328.

McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K. *et al.* (2007). Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecol. Lett.*, 10, 995–1015.

Mooers, A.O. & Heard, S.B. (1997). Inferring evolutionary process from phylogenetic tree shape. *Q. Rev. Biol.*, 72, 31–54.

Mooers, A.O., Harmon, L.J., Blum, M.G.B., Wong, D.H.J. & Heard, S.B. (2007). Some models of phylogenetic tree shape. In: *Reconstructing Evolution: New Mathematical and Computational Advances* (eds Gascuel, O. & Steel, M.). Oxford University Press, Oxford, pp. 149–170.

Nathan, R. (2006). Long-distance dispersal of plants. *Science*, 313, 786–788.

Pennington, R.T. & Dick, C.W. (2004). The role of immigrants in the assembly of the American rainforest tree flora. *Philos. Trans. R. Soc. B*, 359, 1611–1622.

Pennington, R.T., Richardson, J.E. & Lavin, M. (2006). Insights into the historical construction of species-rich biomes from dated plant phylogenies, neutral ecological theory and phylogenetic community structure. *New Phytol.*, 172, 605–616.

Richardson, J.E., Pennington, R.T., Pennington, T.C. & Hollingsworth, P.M. (2001). Rapid diversification of a species-rich genus of neotropical rainforest trees. *Science*, 293, 2242–2245.

Ricklefs, R.E. (2004). A comprehensive framework for global patterns in biodiversity. *Ecol. Lett.*, 7, 1–15.

Ricklefs, R.E. (2006). The unified neutral theory of biodiversity: do the numbers add up? *Ecology*, 87, 1424–1431.

Shao, K.T. & Sokal, R.R. (1990). Tree balance. *Syst. Zool.*, 39, 266–276.

Swenson, N.G., Enquist, B.J., Pither, J., Thompson, J. & Zimmerman, J.K. (2006). The problem and promise of scale dependency in community phylogenetics. *Ecology*, 87, 2418–2424.

Tavaré, S., Balding, D.J., Griffiths, R.C. & Donnelly, P. (1997). Inferring coalescence times from DNA sequence data. *Genetics*, 145, 505–518.

Wakeley, J. (2007). *Coalescent Theory. An Introduction*. Roberts & Company Publishers, Greenwood Village, CO.

Webb, C.O. (2000). Exploring the phylogenetic structure of ecological communities: an example for rain forest trees. *Am. Nat.*, 156, 145–155.
Webb, C.O. & Donoghue, M.J. (2005). PHYLOMATIC: tree assembly for applied phylogenetics. *Mol. Ecol. Notes*, 5, 181–183.
Webb, C.O. & Pitman, N.C.A. (2002). Phylogenetic balance and ecological evenness. *Syst. Biol.*, 51, 898–907.
Webb, C.O., Ackerly, D.D., McPeek, M.A. & Donoghue, M.J. (2002). Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.*, 33, 475–505.
Wiens, J.J. & Donoghue, M.J. (2004). Historical biogeography, ecology and species richness. *Trends Ecol. Evol.*, 19, 639–644.
Yule, G.U. (1924). A mathematical theory of evolution based on the conclusions of Dr. J.C. Willis. *Philos. Trans. R. Soc. Lond. B*, 213, 21–87.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table S1 Summary statistics explored for the ABC method
Appendix S1 ABC algorithm.
Appendix S2 Compilation of phylogenies for the BCI plot.
Appendix S3 Compatibility of neutral theory with 2000 published phylogenies.
Appendix S4 Computation of the regional pool sizes for the four tropical tree plots.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Editor, Thomas Crist
Manuscript received 6 October 2008
First decision made 12 November 2008
Manuscript accepted 25 November 2008