A Neutral Theory for Interpreting Correlations between Species and Genetic Diversity in Communities

Fabien Laroche,[1,2,*] Philippe Jarne,[1] Thomas Lamy,[1,3] Patrice David,[1] and Francois Massol[1,4]

1. Centre d'Ecologie Fonctionnelle et Evolutive, Unité Mixte de Recherche (UMR) 5175, CNRS, Université de Montpellier, Université Paul-Valéry Montpellier, Ecole Pratique des Hautes Etudes, 1919 Route de Mende, 34293 Montpellier, Cedex 5, France; 2. AgroParisTech, 19 Avenue du Maine, 75732 Paris, Cedex 15, France; 3. Département de Sciences Biologiques, Université de Montréal, Montréal, Québec, Canada; 4. Laboratoire de Génétique et Evolution des Populations Végétales, UMR 8198, CNRS, Université Lille 1, Bâtiment SN2, 59655 Villeneuve d'Ascq Cedex, France

Submitted March 4, 2014; Accepted July 19, 2014; Electronically published December 16, 2014

Online enhancements: appendixes, data file.

abstract: Spatial patterns of biological diversity have been extensively studied in ecology and population genetics, because they reflect the forces acting on biodiversity. A growing number of studies have found that genetic (within-species) and species diversity can be correlated in space (the so-called species-gene diversity correlation [SGDC]), which suggests that they are controlled by nonindependent processes. Positive SGDCs are generally assumed to arise from parallel responses of genetic and species diversity to variation in site size and connectivity. However, this argument implicitly assumes a neutral model that has yet to be developed. Here, we build such a model to predict SGDC in a metacommunity. We describe how SGDC emerges from competition within sites and variation in connectivity and carrying capacity among sites. We then introduce the formerly ignored mutation process, which affects genetic but not species diversity. When mutation rate is low, our model confirms that variation in the number of migrants among sites creates positive SGDCs. However, when considering high mutation rates, interactions between mutation, migration, and competition can produce negative SGDCs. Neutral processes thus do not always contribute positively to SGDCs. Our approach provides empirical guidelines for interpreting these novel patterns in natura with respect to evolutionary and ecological forces shaping metacommunities.

Keywords: neutral theory, SGDC, coalescence, community genetics, diversity pattern, mainland-island model.

Introduction

It has been recognized for several decades that the diversity patterns of genes within species and those of species within communities are not independent. Understanding their interactions is the main goal of "community genetics" (Antonovics 1976), an interdisciplinary field that has recently seen a burst of interest (Wares 2002; Agrawal 2003; Neuhauser et al. 2003; Bernhardsson et al. 2013). The rise of environmental genomics and long-term surveys of populations and communities has enhanced the opportunity to confront these two organizational levels (Gugerli et al. 2013). In particular, it is becoming common practice to compute "species-genes diversity correlation" (SGDC; Vellend 2003), which consists in quantifying the link between the genetic diversity in several populations of a species (the focal species) and the species diversity of the local communities within which these populations are embedded. This has been done in more than 40 studies since the seminal work of Vellend and coworkers (see Vellend et al. 2014 for a review).

SGDCs provide information on both fundamental and applied issues with regard to biodiversity. From a fundamental perspective, investigating the generality of SGDC patterns can help to uncover determinants of ecological processes shaping diversity at different levels. For example, several empirical studies have shown that site area (Vellend 2003), which often constitutes a proxy for carrying capacity in community ecology, and site connectivity (Lamy et al. 2013) contribute markedly to positive SGDCs. This suggests that drift and migration might have a strong impact on both species and genetic diversity. From a more applied perspective, detecting positive SGDCs might be useful, in conservation studies, to infer diversity from one level to the other (e.g., predicting species diversity based on genetic data; Papadopoulou et al. 2011).

The growing number of empirical studies on SGDCs constitute a strong incentive to build a quantitative theoretical basis that would help interpreting observed patterns. Vellend and Geber (2005) made a conceptual advance on this issue by envisioning three types of relationships between diversity levels that may generate interpretable signals in community genetics: causal effects of genetic diversity on species diversity, causal effects of

* Corresponding author; e-mail: fabien.laroche@cefe.cnrs.fr.

Am. Nat. 2015. Vol. 185, pp. 59–69.
DOI: 10.1086/678990

species diversity on genetic diversity, and simultaneously parallel effects of external factors on both levels. Neutral theories of molecular evolution (Kimura 1984) and of biodiversity (Hubbell 2001) provide some conceptual elements regarding these potential parallel effects. Indeed, both theories consider limited dispersal and drift to be the main drivers of diversity patterns, and they both predict that carrying capacities and immigration rates should be positively related to diversity (Wright 1931; Hubbell 2001). A positive SGDC should then arise from any external factor generating variation in carrying capacity and connectivity across sites, as has been supported by simulation work (Vellend 2005).

However, even under a neutral framework, the interpretation of positive SGDCs may not be as straightforward as suggested above, because of interactions between the focal species (i.e., the one studied for genetic variation) and other species of the community within sites. In particular, the local abundance of a species may be positively linked to its genetic diversity but also negatively linked to the abundance of other species, and thus to species diversity, as a consequence of competition for limited space. This might produce a negative SGDC (Vellend 2005; Wehenkel et al. 2006; Odat et al. 2010) under specific circumstances that remain to be characterized quantitatively. To our knowledge, no analytical model predicts the sign and magnitude of SGDCs when accounting for the two effects mentioned above, namely, (i) local competition dynamics and (ii) variation in carrying capacity and connectivity among sites. A first objective here is to propose such a model.

A complete quantitative theory of SGDCs also has to include the forces generating diversity, namely, mutation and speciation. These processes have indeed been neglected when discussing SGDC on the grounds that they are often too slow compared with ecological processes (Vellend and Geber 2005). This is true when these rates are negligible with respect to migration. For speciation, this assumption may be challenged when considering archipelagoes (Losos and Schluter 2000) but remains correct when studying metacommunities at limited spatial scale (e.g., a pond network in a single island; Lamy et al. 2013). Here we focus on situations where speciation can be neglected. However, even in this context, assuming that mutation has a negligible impact on genetic diversity is still questionable, especially when using highly mutable markers such as microsatellites (Jarne and Lagoda 1996; Ellegren 2002). Such markers are commonly used in studies reporting SGDCs (Cleary et al. 2006; He et al. 2008; Struebig et al. 2011; Blum et al. 2012; Lamy et al. 2013). Our second objective is thus to provide insights on how mutation may affect SGDCs, even at rather limited spatial and temporal scales.

We build a spatially implicit model of a metacommunity using a unifying neutral framework for both genetic and species dynamics to generate theoretical expectations on SGDCs. Our approach takes into account drift and migration at both diversity levels, as well as mutation, while speciation is neglected. We consider a set of local communities receiving migrants from a larger regional community (Hubbell 2001). This model allows distinguishing within- and among-site effects on SGDCs and thus disentangling the effects of competition within local sites from those of drift and migration among sites. When mutation is neglected, the SGDC turns out to be positive. However, high mutation rates, compared with immigration rates, can produce negative SGDCs. Even under neutral assumptions, the sign of SGDCs can be labile, and understanding SGDCs is therefore not straightforward. On the basis of our framework, we provide some empirical guidelines for interpreting SGDCs.

Material and Methods

Modeling the Dynamics of Species and Gene Diversity in a Site

Our work is based on an individual-based model derived from the classical neutral model of ecological communities (Hubbell 2001). We describe it hereafter following the standardized "overview, design concepts, and details" protocol (Grimm et al. 2010).

Purpose. The model aims at simultaneously providing the species composition of a sample taken from a community and the genotypes of the individuals that belong to the focal species in this sample. Model predictions are based on two features of the sampled site: the carrying capacity (K) and the immigration rate from the regional pool (m). The symbols used are summarized in table 1.

Entities, State Variables, and Scale. The model contains three types of entities: a site, its individuals, and a regional pool serving as a source of migrants. Individuals are described using two state variables: the species they belong to and, for individuals that belong to the focal species, their allelic state at a given locus (under the assumption of haploidy). The latter variable is ignored for individuals that do not belong to the focal species. The site is described by parameters K and m (which are permanent characteristics) and the list of individuals that it contains (which varies in time). The regional pool of individuals is characterized by a set of constant state variables, including the relative abundances of B species $\mathbf{f} = (f_1, f_2, \ldots, f_B)$ and a parameter θ which quantifies the mutation-drift ratio in the regional population of the focal species (app. A; apps.

Table 1: Symbols used in model to predict species-gene diversity correlation (SGDC) in a metacommunity

Symbol	Definition
K	Carrying capacity of a local site (variable across sites)
m	Probability of immigration in a local site (variable across sites)
I	Effective number of migrants in a local site (variable across sites)
σ_K, σ_m	Mean values of $\log(K-1)$ and $\log(m/(1-m))$ over local sites
σ_K^2, σ_m^2, σ_I^2	Variance in $\log(K-1)$, $\log(m/(1-m))$, and $\log(I)$ across local sites
ρ_{Km}	Correlation between $\log(K-1)$ and $\log(m/(1-m))$ across local sites
C_{Im}	Covariance between $\log(I)$ and $\log(m/(1-m))$ across local sites
B	Number of species in the regional pool
$\mathbf{f}$	Relative abundances of species in the regional pool
f_e	Relative abundance of the focal species in the regional pool
θ	Drift-mutation parameter of the regional population of the focal species (weak mutation)
μ	Probability of mutation at the genetic locus ($\mu = 0$ under weak mutation)
S	Number of individuals sampled per site (constant across sites)
$\mathbf{s}$	Composition of species sample
k	Number of individuals in the genetic sample (constant across sites)
$\mathbf{u}$	Composition of genetic sample
R_{spe}	Number of species in the species sample ($\mathbf{s}$)
R_{all}	Number of alleles in the genetic sample ($\mathbf{u}$)
C_{sg}	Expected covariance between R_{all} and R_{spe}
C_{within}	Contribution of stochastic competition for space within sites to the expected covariance
C_{among}	Contribution of variation in K and m among sites to the expected covariance
SGDC	Expected correlation between R_{all} and R_{spe}

Note: Boldface type indicates vectors.

A–C available online). When not neglected, mutation is characterized by a per-birth mutation rate μ in the focal species.

Process Overview. The model is characterized by discrete death-birth cycles in the site. At the beginning of each cycle, the site contains exactly K individuals (i.e., is saturated). An individual is then randomly chosen, discarded, and replaced by the offspring of a reproducer which either belongs to the site, with probability $1 - m$, or to the regional pool, with probability m. When the reproducer belongs to the site, the offspring inherits its species identity. Its genotype (focal species) is either the same as the reproducer's genotype (with probability $1 - \mu$) or a mutated allele not already present in the species (with probability μ; see below for additional discussion of the mutation regime). When the reproducer belongs to the regional pool, the species identity of the immigrant offspring is randomly drawn from the distribution of the species relative abundances in the regional pool ($\mathbf{f}$). When the offspring belongs to the focal species (with probability f_e), its genotype is determined as explained below. Note that, in our model, competition among genotypes and species occurs during these cycles, when dead individuals are replaced by offspring of either migrant or local origin (i.e., a lottery competition for space).

Two scenarios are considered with respect to mutation. The first scenario corresponds to a weak mutation regime ($\mu \ll m$; in practice, μ is set to zero) in which mutation is neglected in the local community dynamics. At the regional scale, the allelic frequencies of the focal species are assumed to be at mutation-drift equilibrium and follow a Ewens distribution with parameter θ (Ewens and Tavaré 2006). The second scenario corresponds to a strong mutation regime where mutation at the focal locus cannot be neglected when compared to migration ($\mu \approx m$). Mutation process follows an infinite-allele model: any mutation event generates an allele that never occurred before in the site. As a consequence of high mutation rate, the regional allelic pool is assumed to be infinitely diverse (app. A): immigrants always harbor alleles that did not occur before in the site.

Outputs. Species diversity and allelic diversity are determined through a sampling process designed to mimic a typical SGDC study. S individuals are randomly sampled from the site (the species sample; fig. 1). The species composition of this sample is described by $\mathbf{s} = (s_1, s_2, \ldots, s_B)$, where s_i individuals belong to species i, and $\sum_i s_i = S$. Species diversity is computed as species richness (R_{spe}; i.e., as the number of distinct species occurring in $\mathbf{s}$). In the species sample, the allelic states of the individuals belonging to the focal species e are described by $\mathbf{t} = (t_1, t_2, \ldots, t_n)$, where t_j individuals carry allele j, and

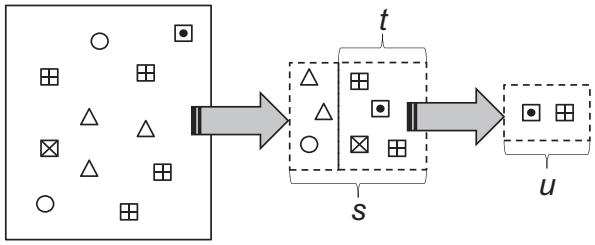

Figure 1: Sampling protocol of a site in the model. The large rectangle on the left depicts a site. Arrows represent random sampling. Dashed rectangles represent samples, and pictograms represent species. The focal species (squares) harbors alleles that are depicted with different graphical patterns (crosses and points). The genetic sample **u** is obtained by subsampling k individuals among individuals of the focal species (**t**) included in the species sample (**s**). Here $S = 7$ and $k = 2$.

$\sum_j t_j = s_e$. A random subset, **u**, of **t** containing k individuals is genotyped and constitutes the "genetic sample" within sites. The genetic diversity is estimated using allelic richness (R_{all}), computed as the number of distinct alleles occurring in **u**. Note that with this sampling procedure, allelic and species richness can be computed only for sites containing more than S individuals ($K > S$) and yielding a sample **s** containing more than k individuals of species e ($s_e > k$).

Modeling and Decomposing Species-Gene Diversity Relationships across Sites

The influence of variation in carrying capacity (K) and immigration rate (m) among sites on SGDC is modeled by considering a set of sites created by independently drawing values of K and m from a bivariate distribution with given variance and covariance (app. B). All the sites are connected to the same regional pool and follow the same mutation dynamics (i.e., weak or strong). R_{all} and R_{spe} are computed using our model in all the sites where $K > S$ and $s_e > k$ (see above). Note that our sampling protocol controls for sample size at both species (through S) and genetic levels (through k), which allows comparing diversity measures among sites. We provide below theoretical expectations about the sign of the expected covariance between R_{all} and R_{spe} computed across sites (C_{sg}). Although C_{sg} is not the SGDC classically estimated in empirical studies (i.e., authors generally use Pearson's correlation coefficient), it provides qualitative information about the sign of the expected relationship between genetic and species diversity. Besides, C_{sg} can be decomposed into two effects. The first one occurs within sites as the result of local competition. The second effect stems from the variation in carrying capacity (K) and migration (m) among sites. Technically speaking, this can be expressed as the decomposition of C_{sg} as the sum of two covariances, C_{within} and C_{among} (app. C), with

$$\begin{cases} C_{sg} = C_{within} + C_{among} \\ C_{within} = \mathbb{E}[\mathrm{Cov}_{K,m}[R_{spe}, R_{all}]] \\ C_{among} = \mathrm{Cov}\left[\overline{R_{spe}}(K, m), \overline{R_{all}}(K, m)\right] \end{cases}, \quad (1)$$

where $\mathrm{Cov}_{K,m}[R_{spe}, R_{all}]$ is the covariance between specific and allelic richness (considered as random variables) within a site with given K and m values, $\mathbb{E}$, and Cov are the expectation and covariance over (K, m) distribution, respectively, and overlined quantities are expectations within sites with given K and m values. C_{among} reflects the effect of (K, m) variation among sites. Importantly, C_{among} is null when K and m do not vary among sites, in which case only local competition (C_{within}) determines C_{sg} and thus the sign of SGDC. From a statistical point of view, this decomposition of C_{sg} can be interpreted as in an analysis of variance framework, C_{among} being the part of covariance explained by K and m and C_{within} being the residual covariance.

Simulating R_{all} and R_{spe} in a Set of Local Sites

Simulations illustrate our theoretical predictions about the sign of SGDC and provide more quantitative information about SGDC (i.e., Pearson's correlation coefficient) variation with respect to K and m distribution among sites. An efficient sampling approach in our model is to simulate the genealogy of the S individuals per sample backward in time (coalescence approach; Rosindell et al. 2008). This simulation strategy is used here to generate **s** and **u** samples within local sites, from which R_{all} and R_{spe} are computed. More details about the simulation algorithm are provided in supporting material.

In all the simulations, the regional community contains 20 species, the relative abundances of which are derived from a truncated geometric distribution with parameter 0.2 (i.e., $f_i = (1 - 0.2) \times 0.2^{i-1}/(1 - 0.2^{20})$). The most abundant species in the regional pool is chosen as the focal one ($f_e \approx 0.8$) to avoid discarding many sites because of unsuccessful sampling; this is a reasonable assumption with regard to empirical studies reporting SGDCs, which generally analyze genetic diversity in common species. Under weak mutation, θ is set to 10. Under strong mutation, μ is set to 10^{-3}, in line with what is known for microsatellite markers (Jarne and Lagoda 1996; Ellegren 2002). Landscapes considered here are sets of 100 sites. K and m per site are determined by sampling $(\log(K - 1), \log[m/(1 - m)])$ in a "discretized" bivariate Gaussian distribution with mean (α_K, α_m), marginal variances

(σ_K^2, σ_m^2), and covariance $\rho_{Km}\sigma_K\sigma_m$ (app. B). The size of the species sample **s** is set to $S = 50$ individuals, and that of the genetic sample **u** is set to $k = 5$ individuals. In each site, R_{all} and R_{spe} values are obtained by simulating the above-mentioned coalescent process (pseudo code available online) with an algorithm implemented in Java (Jdk 7u17, Oracle; code available from the authors upon request). SGDC is computed from the values of R_{all} and R_{spe} across sites, using Pearson's correlation coefficient.

Results

An important outcome of our work is to provide a decomposition of the covariance between diversity levels (C_{sg}) into effects occurring within (C_{within}) and among sites (C_{among}) and derive predictions about each of them. We separately analyzed each type of effects in our model and generated conclusions about the overall sign and value of SGDC. Because the mutation-to-migration ratio strongly affects both C_{within} and C_{among}, we considered the weak mutation regime and the strong mutation regime separately.

Weak Mutation Regime

The behavior of the within- and among-site components of C_{sg} can be analyzed by considering the joint probability of **s** and **u**. Under the weak mutation regime, we established that the compositions of **s** and **u** are probabilistically independent within a site (app. A), so that C_{within} is null. This result is a consequence of controlling for genetic sample size (through the parameter k here) when estimating genetic diversity, thus dampening any effect of local population size on R_{all}. We performed repeated simulations of a single site with given K and m values to estimate the relation between genetic and species diversity and the abundance of the focal species in the site. Because s_e provides a proxy for the abundance of the focal species within sites, species and genetic diversity were actually sorted as a function of s_e (fig. 2). As predicted by our theoretical analysis (see app. A), R_{spe} decreases with s_e, but R_{all} does not show any trend with respect to s_e.

As the within-site component is null, the covariance between species and genetic diversity under the weak mutation regime reduces to the C_{among} component, which depends on the variation in (K, m) among sites. It turns out that the latter influences **s** and **u** compositions only through the variation in $I = (K - 1)m/(1 - m)$, the so-called effective number of migrants (app. A; Etienne and Olff 2004; Etienne and Alonso 2005), which quantifies the relative strength of drift and immigration within sites. The expression of C_{among} in equation (1) can therefore be rewritten as

$$C_{among} = \mathrm{Cov}\left[\overline{R_{spe}}(I), \overline{R_{all}}(I)\right], \quad (2)$$

where overlined quantities are expectations in sites with parameter I. It can be shown (app. A) that both $\overline{R_{spe}}(I)$ and $\overline{R_{all}}(I)$ increase with I, so that C_{among} is expected to be positive.

From these results, C_{sg} (equal to C_{among}) always takes positive values. Simulations (fig. 3) illustrate these theoretical expectations: simulated SGDCs are always positive. Moreover, SGDCs increase with the variance in I and, for a given variance in I, they show very little variation. These results are consistent with our theoretical prediction: both R_{spe} and R_{all} depend on site parameters through the value of I only and increase with I. The variance in I is positively related to the variances in both K and m as well as to the covariance between K and m. Note, however, that a large variance in both K and m, associated to a strong negative covariance between these two parameters, generates a low variance in I, leading to weak values of C_{sg} and SGDCs (app. B).

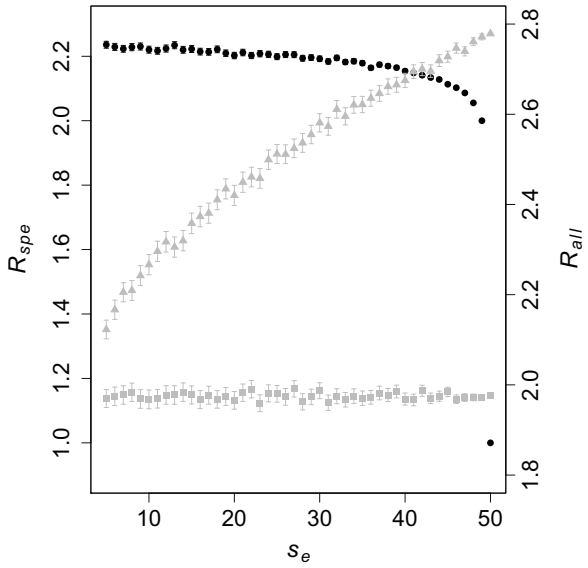

Figure 2: Variation in mean R_{spe} and R_{all} as a function of s_e (sampling size of the focal species) in a site with given carrying capacity (K) and connectivity (m). A total of 10,000 simulations were performed under each mutation regime. Simulation outputs are sorted according to s_e value, and for each s_e value, the mean values of observed R_{spe} (black circles), R_{all} under weak mutation (gray squares), and R_{all} under strong mutation (gray triangles) are reported. For R_{spe}, the output of the 20,000 simulations are considered together to compute mean values for each s_e (because R_{spe} does not depend on the mutation regime). A 95% confidence interval (1.96 times the standard error) is given with the R_{spe} and R_{all} mean values. Other parameters are set to $K = 1{,}000$, $m = 0.001$, $B = 20$, $f_i = (1 - 0.2) \times 0.2^{i-1}/(1 - 0.2^{20})$, $e = 1$, $f_e \approx 0.8$, $\theta = 10$ (for weak mutation), $\mu = 10^{-3}$ (for strong mutation), $S = 50$, and $k = 5$.

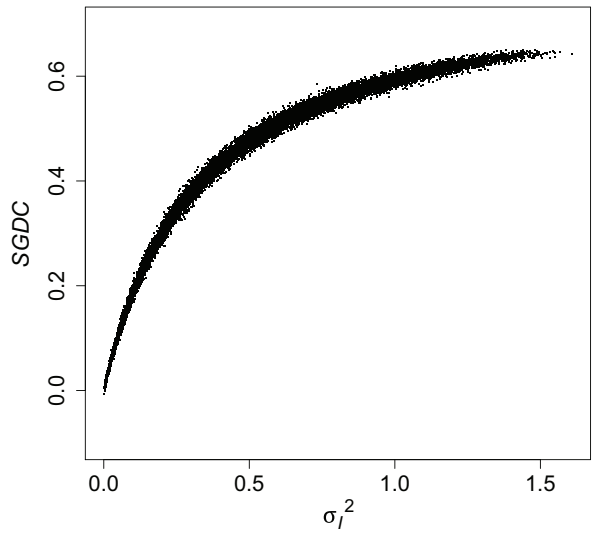

Figure 3: Mean species-gene diversity correlation (SGDC) as a function of σ_I^2 (the variance in $\log_{10} I$ across sites) under the weak mutation regime for a set of simulated landscape. Each point corresponds to a landscape. For each landscape, *(K, m)* follows a bivariate discretized lognormal distribution (see "Material and Methods"). The mean of SGDCs obtained in 500 independent simulations is represented as a function of σ_I^2, recalling that $\sigma_I^2 = \sigma_K^2 + \sigma_m^2 + 2\rho_{Km}\sigma_K\sigma_m$ (see app. B). σ_m^2 and ρ_{Km} are numerically explored over the $[0, 1] \times [-1, 1]$ space by steps of 0.01, leading to 19,900 combinations. Other parameters are set to $\alpha_K = 3, \alpha_m = -3, \sigma_K^2 = 3$, $B = 20$, $f_i = (1 - 0.2) \times 0.2^{i-1}/(1 - 0.2^{20})$, $e = 1$, $f_e \approx 0.8$, $\theta = 10$, $S = 50$, and $k = 5$.

Strong Mutation Regime

Under the strong mutation regime, C_{within} is not necessarily zero anymore. Indeed, R_{all} increases with s_e, whereas R_{spe} tends to decrease (fig. 2), generating negative expectations for C_{within}. This clearly appears when simulating homogeneous landscapes, with the same (K, m) values in all sites (i.e., $C_{\text{among}} = 0$; fig. 4). The SGDC is negative, especially when the carrying capacity K of sites is large and their immigration rate m is small. A delta method to order 0 on C_{within} in equation (1) yields the following approximation:

$$C_{\text{within}} \approx \text{Cov}_{\text{floor}(\bar{K}),\bar{m}}\,[R_{\text{spe}}, R_{\text{all}}], \tag{3}$$

where $\overline{K}, \bar{m}$ are means of K and m across the landscape and floor() is the integer part operator. Other notations are similar to equation (1). We expect C_{within} to depend mostly on mean carrying capacity and immigration in the set of sites and not on variance and covariance of K and m across sites. Thus the negative impact observed in the homogeneous landscapes (K and m constant across sites) detailed in figure 4 should also occur in more complex landscapes that share the same $(\overline{K}, \overline{m})$ values. However, the total covariance will result from the addition of C_{within} and C_{among}, which may have different signs.

Mutation also has an impact on C_{among}. As under the weak mutation regime, C_{among} is affected by variation in I, but also by variation in migration alone (m), independently from I. This is because, when migration is high, mutation events have less impact on within-site diversity, which depends mainly on new alleles brought by immigrants (app. A). The expression of C_{among} is more complex than under the weak mutation regime (eq. [2]), because genetic diversity depends on both I and m as follows:

$$\begin{aligned} C_{\text{among}} &= \text{Cov}\,[\overline{R_{\text{spe}}}(I), \overline{R_{\text{all}}}(I, m)] \\ &\approx \frac{\partial \overline{R_{\text{spe}}}}{\partial I} \times \frac{\partial \overline{R_{\text{all}}}}{\partial I}\text{Var}\,(I) \\ &+ \frac{\partial \overline{R_{\text{spe}}}}{\partial I} \times \frac{\partial \overline{R_{\text{all}}}}{\partial m}\text{Cov}\,[I, m], \end{aligned} \tag{4}$$

where the notation is identical to that in equation (2), ∂ is the symbol for partial derivative, and the approximation

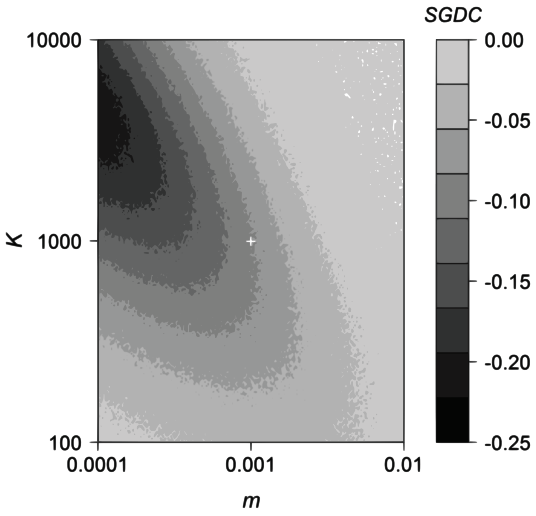

Figure 4: Values of species-gene diversity correlations (SGDCs) with respect to carrying capacity (K) and immigration rate (m) of sites for simulated homogeneous landscapes (i.e., σ_K^2 and σ_m^2 are set to zero) under strong mutation. For each landscape, 500 simulations were performed, and the average SGDC was reported as a dot in the (m, K) space, with a shade of gray indicating the magnitude of the associated value. The white dots on the upper-right corner represent simulations for which a positive SGDC was obtained because of numerical noise. The white cross indicates the points used as a mean in site distributions of figures 5 and 6. K and m were numerically explored on logarithmic scales, leading to 40,401 sets of parameters (landscapes). Other parameters were set to $B = 20$, $f_i = (1 - 0.2) \times 0.2^{i-1}/(1 - 0.2^{20})$, $e = 1$, $f_e \approx 0.8$, $\mu = 10^{-3}$, $S = 50$, and $k = 5$.

is based on the delta method. Equation (4) shows that C_{among} is influenced by I at both organizational levels (first term in the approximation) but also by the spatial covariation between I and m (second term in the approximation). As mentioned in the section on weak mutation, $\overline{R_{spe}}(I)$ is an increasing function of I. $\overline{R_{all}}(I, m)$ is an increasing function of I (m being constant; app. A) and a decreasing function of m (I being constant; app. A). The variation in I (first term) therefore contributes positively to C_{among} (first term), whereas a positive correlation between I and m among sites has a negative impact on C_{among} (second term).

The sign of C_{sg} depends on the relative values and sign of C_{within} and C_{among} (eq. [1]). When the variance in I across sites is very low, equation (4) implies that C_{among} is close to 0, and C_{within} is the dominant term in equation (1). C_{sg} is then expected to be negative. Our simulations confirm this prediction: figure 5*A* illustrates that negative SGDCs occur when the variance in m is low and the correlation between K and m negative; this corresponds to the region where the variance in I is the lowest in figure 5*B*. When the variance in I increases, equation (4) suggests that C_{among} increases, whereas equation (3) suggests that C_{within} remains unchanged as it is mostly determined by the means of K and m over sites and not by variance or covariance of K and m (or equivalently I and m) across sites. Therefore, theory predicts that C_{sg} increases and becomes positive when the effect of C_{among} exceeds that of C_{within}. We retrieve these results through simulations when considering the SGDC rather than C_{sg}: the SGDC increases with the variance in I and eventually becomes positive (fig. 5*B*).

A comparison of figure 3 (weak mutation) with figure 5*B* (strong mutation) indicates that the variance in I is not as good a predictor of the value of SGDC in the latter as in the former case. Equation (4) shows a negative effect of the covariance between I and m on C_{among} under strong mutation, in addition to variance in I. Representing the SGDC as a function of both the variance in I and the covariance between m and I (or equivalently between $\log[m/(1-m)]$ and $\log(I)$, C_{Im}; fig. 6) corroborates that, for constant σ_I^2, the SGDC consistently decreases with C_{Im}, as predicted by equation (4). Although the parameterization in (I, m) makes computation simpler, the initial parameterization in (K, m) is more accessible to intuition. In terms of (K, m) distribution, increasing C_{Im} value for a constant σ_I^2 value can be achieved, for instance, by increasing the variance in m, keeping variance in K constant, and making ρ_{Km} more negative (app. A). This effect of C_{Im} explains the wider spread of the values of SGDC for a given σ_I^2 in figure 5*B* than in figure 3.

Discussion

Community genetics is a rising field of research that has developed along several lines, such as studying relation-

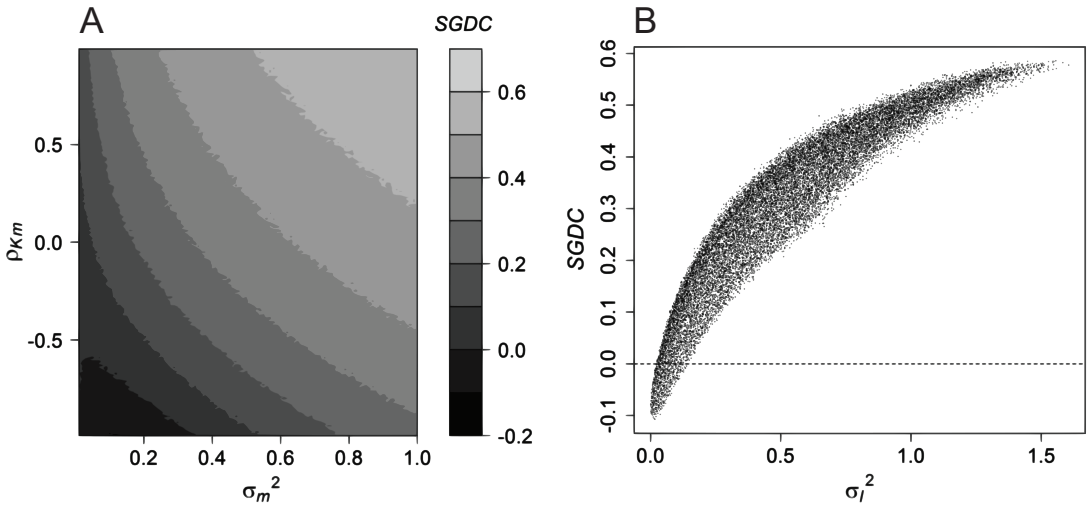

Figure 5: Values of expected species-gene diversity correlation (SGDCs) with respect to the variation (and covariation) in K, m and I across sites for simulated landscapes under the strong mutation regime. A total of 500 simulations were performed for each of the sets of parameters, and the mean SGDC was computed. Panel *A* depicts the mean value of SGDC as a function of σ_m^2 and ρ_{Km}. Panel *B* depicts the mean value of SGDC as a function of σ_I^2. σ_m^2 and ρ_{Km} were numerically explored over the $[0, 1] \times [-1, 1]$ space by steps of 0.01, leading to 19,900 combinations. Other parameters were set to $\alpha_K = 3$, $\alpha_m = -3$, $\sigma_K^2 = 3$, $B = 20$, $f_i = (1 - 0.2) \times 0.2^{i-1}/(1 - 0.2^{20})$, $e = 1$, $f_e \approx 0.8$, $\mu = 10^{-3}$, $S = 50$, and $k = 5$.

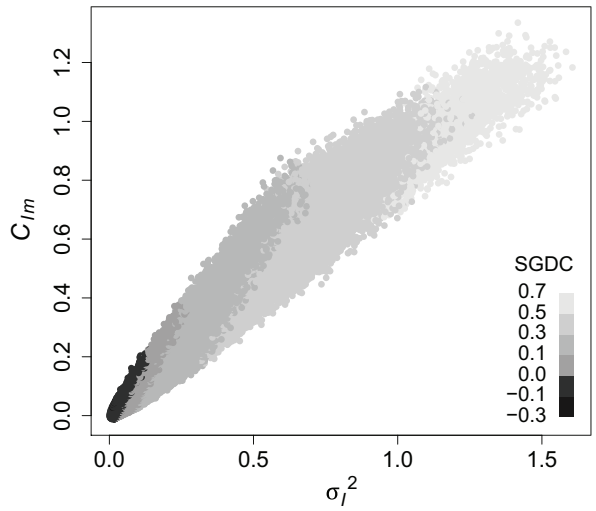

Figure 6: Values of expected species-gene diversity correlations (SGDCs) with respect to the variation in I and the covariation between m and I across sites (i.e., the $[\sigma_I^2, C_{Im}]$ space) for simulated landscapes under the strong mutation regime. A total of 500 simulations were performed for each landscape, and mean SGDC was represented here by a dot, with shades of gray indicating the associated value. The parameters were set to $\alpha_K = 3, \alpha_m = -3, \sigma_K^2 = 3$. σ_m^2 and ρ_{Km} were numerically explored in $[0, 1] \times [-1, 1]$ by step of 0.01, leading to 19,900 combinations. Other parameters were set to $B = 20$, $f_i = (1 - 0.2) \times 0.2^{i-1}/(1–0.2^{20})$, $e = 1$, $f_e \approx 0.8$, $\mu = 10^{-3}$, $S = 50$, and $k = 5$.

ships between herbivore communities and the genetics of plants supporting these communities at small geographic scale (Bernhardsson et al. 2013; McArt and Thaler 2013) or using phylogeography to better understand past processes that have shaped present communities at larger scale (Wares 2002; Webb et al. 2002). The study of correlations between genetic diversity at the species level and species diversity at the community level (SGDCs; Vellend and Geber 2005) is one of these offshoots. Up to now, predictions on SGDCs have essentially been formalized on the basis of verbal models. We propose here a theoretical framework encompassing both species and genetic levels to more fully analyze and interpret SGDCs. We based our work on genealogical approaches and sampling formulae that are now commonly used in both population genetics (Wakeley 2008) and community ecology (Etienne and Olff 2004) to infer processes from patterns. Those techniques proved very useful, for instance, when searching for selective processes (Fu and Li 1993; Etienne 2005, 2007). Although sampling formulae are available for modeling both genetic and species dynamics (Etienne and Alonso 2005), we know of no previous work generating simultaneous predictions at both levels on the basis of a generalized coalescent of genes and species.

An important result of our work is that the covariance between species richness and allelic richness (C_{sg}) can be additively decomposed into (i) the effect of competition between the focal species and other species within local sites (the C_{within} term) and (ii) the parallel effect of variation in carrying capacity and immigration rate of sites on allelic and species richness (the C_{among} term; eq. [1]). Both effects had previously been identified in the literature. Local competition was thought to negatively affect SGDCs (i.e, $C_{within} < 0$ in our framework; Vellend 2005; Wehenkel et al. 2006), while variation in area and isolation among sites was thought to generate positive SGDCs (i.e., $C_{among} > 0$; Vellend 2003). The dominant effect could, in principle, be inferred from the SGDC sign without further need of a quantitative framework. However, our predictions are partially at variance with these intuitions. When the mutation rate is much lower than the immigration rate, C_{among} is positive, as expected, but C_{within} is always zero. When the mutation rate is comparable to or higher than the immigration rate, C_{within} is negative, which corresponds to expectations, but C_{among} can take both signs, which does not.

Because mutation drastically changes the predictions on SGDC patterns, an important aspect in empirical studies should be to determine the mutation-to-migration ratio before interpreting SGDCs. In particular, many estimates of SGDCs are based on microsatellites when evaluating the genetic diversity (Cleary et al. 2006; He et al. 2008; Struebig et al. 2011; Blum et al. 2012; Lamy et al. 2013). These markers may have high mutation rates (Jarne and Lagoda 1996; Ellegren 2002) that are potentially high enough to compare with immigration rates, especially in isolated sites. Insights on the mutation-to-migration ratio can be obtained by computing the relationships between a proxy of the number of migrants in a site (I), a proxy of the local carrying capacity (K), and the allelic richness of the focal species. If allelic richness increases with carrying capacity when controlling for the number of migrants (which can be assessed with a partial correlation analysis for instance), this suggests that mutation contributes strongly to the build-up of variation in these sites. Alternatively, one can also directly use the genetic polymorphism of the focal species to assess the relative strength of mutation and migration processes. For instance, when considering microsatellites, testing for a significant difference between R_{ST} and F_{ST} estimators of genetic structure in the focal species could help in evaluating whether mutation could be neglected (Hardy et al. 2003). We note that speciation may have an impact on SGDC patterns similar to that of mutation. Speciation was not considered in our model, because we focused on a temporal/spatial scale at which it is unlikely to generate species variation to a significant extent. When the immigration and speciation rates are of

similar magnitude (e.g., in isolated metacommunities, such as archipelagoes), a larger number of endemic species should be generated in sites with lower immigration rates (Rosindell and Phillimore 2011). Interactions between the effects of speciation and connectivity on SGDC should then be similar to those detected in our model about mutation.

When mutation is weak compared with immigration, our model predicts that local competition should not affect the SGDC pattern. We emphasize here the importance of the sampling protocol. Earlier studies predicted a negative C_{within}, because genetic diversity of the focal species is expected to increase with its population size (Vellend 2005; Wehenkel et al. 2006). The sampling protocol of our model (fig. 3) did not allow this positive relationship to occur, because a fixed number of individuals of the focal species were genotyped (k individuals in sample $\boldsymbol{u}$; smaller samples were disregarded). Other sampling protocols can generate the same disconnection between the allelic richness and the population size of the focal species as long as they incorporate a control of the genetic sample size. For instance, the same disconnection occurs when genetic diversity is measured by genotyping all individuals belonging to the focal species ($\boldsymbol{t}$) and computing a rarefied richness indicator (Petit et al. 1998), as most empiricists do. Under weak mutation, controlling for the size of species and genetic samples filters out the influence of local competition, which facilitates the interpretation of observed patterns.

As the among-sites effect is always positive under weak mutation, our neutral theory yields the simple prediction that SGDC should always be positive and reflects the variance of the effective number of migrants among sites (fig. 3). Empirical studies that (i) demonstrated that mutation is weak, (ii) controlled for sample size in the sampling protocol, and (iii) observed strong variation in size and connectivity among sites should then expect a positive SGDC. When this prediction is not verified, it may mean one of three things. First, variation in size and connectivity may be negatively correlated among sites in such a way that the overall variation in the number of immigrants among sites is low (app. B). Second, there may be a nonneutral process at work. For instance, Derry et al. (2009) illustrated how species-sorting along an environmental gradient may contribute to cancel the positive parallel effects of variation in size and connectivity on SGDC. Finally, some other assumptions of our model may be violated. The last explanation may apply when considering spatially continuous systems (e.g., alpine forest; Taberlet et al. 2012) for which our implicit description of space may prove insufficient to describe the spatial autocorrelation in the system.

Under strong mutation, the analysis of SGDC patterns is different, because the correlation sign predicted by the neutral theory developed here is more labile than under the weak mutation regime. On the one hand, local competition (C_{within} term) has a negative impact on SGDC. Indeed, the positive relationship between the population size of the focal species and the allelic richness of the genetic sample is maintained, and this occurs even when controlling for genetic sample size (fig. 3). On the other hand, C_{among} can take either sign depending on the codistribution of carrying capacities and immigration rates among sites. In particular, a negative C_{among} value emerges when sites tend to receive the same effective number of immigrants per generation irrespective of their carrying capacity (i.e., low variance in I). Such a situation could occur, for example, in fragmented landscapes with patches of different sizes connected by corridors; the effective number of immigrants would primarily depend on the presence of corridors and may be uncorrelated to patch size (which determines its carrying capacity). On the whole, any sign of the SGDC is compatible with the neutral framework when mutation is strong, so that, contrary to the weak mutation regime, neutrality cannot be rejected on the basis of the sign of the correlation only. Note, however, that using markers that show different levels of polymorphism (using polymorphism as a proxy of mutation rate) may provide further tests. If the correlation is positive when using poorly variable markers and negative when considering highly variable ones, the overall observation is compatible with the neutral framework. By contrast, a consistently negative SGDC, whatever the level of polymorphism of the considered marker, may be interpreted as a rejection of our neutral model. When our framework applies, interpreting the SGDC sign under strong mutation is not straightforward. A significantly positive SGDC indicates a strong positive C_{among} and can be interpreted as an effect of high variance in the number of migrants among sites. However, nonsignificant and negative SGDCs lead to ambiguous interpretation. In particular, negative correlations can indicate an effect of local competition but can also result from a negative C_{among}.

One way to progress in the interpretation of SGDCs is to decompose the covariance between species and genetic diversity (C_{sg}) into the C_{within} and C_{among} effects. Some authors suggested statistical methods to analyze the contribution of size and connectivity of sites to the overall SGDC (Vellend 2003; Lamy et al. 2013). Both studies detected significantly positive SGDCs along with a strong contribution of area and connectivity, respectively, to these correlations, which may indicate a strong positive C_{among}. Our model provides a theoretical basis for going one step further in this analysis by allowing a covariance decomposition based on mechanisms (instead of environmental factors) to be performed in empirical studies. One approach could be to estimate I in sites and to directly per-

form covariance decomposition along those estimates instead of using proxies of size and connectivity, as done in former studies. This should provide a more direct assessment of C_{among}. However, estimating I is not straightforward. One solution could be to use loci different from those used to compute SGDCs and to independently assess the migration-drift ratio in populations of the focal species through Nm (N and m are the population size and immigration rate, respectively, in, for example, island models of population structure; Rousset 2004), which should provide a relevant proxy for If_e. Separate estimates of f_e could be obtained by other approaches, such as by pooling all the local species samples to generate a regional sample (Jabot et al. 2008) so that I could be isolated. Ultimately, decomposing SGDC patterns should contribute to a deeper understanding than the sign of SGDC alone. Beyond helping to interpret ambiguous cases such as negative SGDCs under strong mutation, disentangling C_{within} and C_{among} may also provide new tests of our framework: for instance, under low mutation, observing a large positive SGDC but no significant C_{among} may indicate other nonneutral processes, such as positive interactions between the focal species and the rest of the community within sites (e.g., facilitation in plant communities; Brooker et al. 2008).

With the building of an adequate theory, SGDC patterns may be used to study the processes, such as dispersal and drift, acting in metacommunities. Certainly, a further step is the development of neutral models, including a full sampling theory to provide useful null hypotheses to detect selective processes, based on both species count data and genomic sequencing. The spectacular increase in the availability of genomic data opens interesting perspectives. It seems unlikely, however, that comparison of local diversity across levels provides enough information to unravel the complex processes acting in metacommunities, such as niche structure and environmental filtering among sites. Interestingly enough, empirical studies have started to report other patterns, such as correlations between species and genetic β-diversity (Papadopoulou et al. 2011; Baselga et al. 2013). Theoretical analyses, along the line followed here, are certainly required to evaluate their inferential power and to incorporate them in a spatially explicit neutral theory of SGDCs.

Acknowledgments

Many thanks to U. Berger, J. Bronstein, and three anonymous referees for their useful comments on the manuscript. We thank the Centre d'Ecology Fonctionnelle et Evolutive and M.-C. Quidoz (Système d'Information pour l'Ecologie platform) for access to computational resources. F.L. is supported by a fellowship from AgroParistech. This work was supported by the Affairs project of the Agence Nationale de la Recherche (BioAdapt program; to P.D.). Thanks are due to CNRS for supporting the work of P.J., P.D., and F.M.

Literature Cited

Agrawal, A. A. 2003. Community genetics: new insights into community ecology by integrating population genetics. Ecology 84: 543–544.

Antonovics, J. 1976. The input from population genetics: "the new ecological genetics." Systematic Botany 1:233–245.

Baselga, A., T. Fujisawa, A. Crampton-Platt, J. Bergsten, P. G. Foster, M. T. Monaghan, and A. P. Vogler. 2013. Whole-community DNA barcoding reveals a spatio-temporal continuum of biodiversity at species and genetic levels. Nature Communications 4:1892.

Bernhardsson, C., K. M. Robinson, I. N. Abreu, S. Jansson, B. R. Albrectsen, and P. K. Ingvarsson. 2013. Geographic structure in metabolome and herbivore community co-occurs with genetic structure in plant defence genes. Ecology Letters 16:791–798.

Blum, M. J., M. J. Bagley, D. M. Walters, S. A. Jackson, F. B. Daniel, D. J. Chaloud, and B. S. Cade. 2012. Genetic diversity and species diversity of stream fishes covary across a land-use gradient. Oecologia 168:83–95.

Brooker, R. W., F. T. Maestre, R. M. Callaway, C. L. Lortie, L. A. Cavieres, G. Kunstler, P. Liancourt, et al. 2008. Facilitation in plant communities: the past, the present, and the future. Journal of Ecology 96:18–34.

Cleary, D. F. R., C. Fauvelot, M. J. Genner, S. B. J. Menken, and A. Ø. Mooers. 2006. Parallel responses of species and genetic diversity to El Niño southern oscillation–induced environmental destruction. Ecology Letters 9:304–310.

Derry, A. M., S. E. Arnott, J. A. Shead, P. D. N. Hebert, and P. T. Boag. 2009. Ecological linkages between community and genetic diversity in zooplankton among boreal shield lakes. Ecology 90: 2275–2286.

Ellegren, H. 2002. Mismatch repair and mutational bias in microsatellite DNA. Trends in Genetics 18:552.

Etienne, R. S. 2005. A new sampling formula for neutral biodiversity. Ecology Letters 8:253–260.

———. 2007. A neutral sampling formula for multiple samples and an "exact" test of neutrality. Ecology Letters 10:608–618.

Etienne, R. S., and D. Alonso. 2005. A dispersal-limited sampling theory for species and alleles. Ecology Letters 8:1147–1156.

Etienne, R. S., and H. Olff. 2004. A novel genealogical approach to neutral biodiversity theory. Ecology Letters 7:170–175.

Ewens, W. J., and S. Tavaré. 2006. Ewens sampling formula. Pages 2131–2135 *in* C. B. Read, N. Balakrishnan, B. Vidakovic, and S. Kotz, eds. Encyclopedia of statistical sciences. Wiley, New York.

Fu, Y. X., and W. H. Li. 1993. Statistical tests of neutrality of mutations. Genetics 133:693–709.

Grimm, V., U. Berger, D. L. DeAngelis, J. G. Polhill, J. Giske, and S. F. Railsback. 2010. The ODD protocol: a review and first update. Ecological Modelling 221:2760–2768.

Gugerli, F., R. Brandl, B. Castagneyrol, A. Franc, H. Jactel, H.-P. Koelewijn, F. Martin, et al. 2013. Community genetics in the time

of next-generation molecular technologies. Molecular Ecology 22: 3198–3207.

Hardy, O. J., N. Charbonnel, H. Fréville, and M. Heuertz. 2003. Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. Genetics 163:1467–1482.

He, T., B. B. Lamont, S. L. Krauss, N. J. Enright, and B. P. Miller. 2008. Covariation between intraspecific genetic diversity and species diversity within a plant functional group. Journal of Ecology 96:956–961.

Hubbell, S. P. 2001. The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton, NJ.

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

Jarne, P., and P. J. L. Lagoda. 1996. Microsatellites, from molecules to populations and back. Trends in Ecology and Evolution 11:424–429.

Kimura, M. 1984. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

Lamy, T., P. Jarne, F. Laroche, J.-P. Pointier, G. Huth, A. Segard, and P. David. 2013. Variation in habitat connectivity generates positive correlations between species and genetic diversity in a metacommunity. Molecular Ecology 22:4445–4456.

Losos, J. B., and D. Schluter. 2000. Analysis of an evolutionary species-area relationship. Nature 408:847–850.

McArt, S. H., and J. S. Thaler. 2013. Plant genotypic diversity reduces the rate of consumer resource utilization. Proceedings of the Royal Society B: Biological Sciences 280:20130639.

Neuhauser, C., D. A. Andow, G. E. Heimpel, G. May, R. G. Shaw, and S. Wagenius. 2003. Community genetics: expanding the synthesis of ecology and genetics. Ecology 84:545–558.

Odat, N., F. H. Hellwig, G. Jetschke, and M. Fischer. 2010. On the relationship between plant species diversity and genetic diversity of *Plantago lanceolata* (Plantaginaceae) within and between grassland communities. Journal of Plant Ecology 3:41–48.

Papadopoulou, A., I. Anastasiou, F. Spagopoulou, M. Stalimerou, S. Terzopoulou, A. Legakis, and A. P. Vogler. 2011. Testing the species–genetic diversity correlation in the Aegean archipelago: toward a haplotype-based macroecology? American Naturalist 178:241–255.

Petit, R. J., A. El Mousadik, and O. Pons. 1998. Identifying populations for conservation on the basis of genetic markers. Conservation Biology 12:844–855.

Rosindell, J., and A. B. Phillimore. 2011. A unified model of island biogeography sheds light on the zone of radiation. Ecology Letters 14:552–560.

Rosindell, J., Y. Wong, and R. S. Etienne. 2008. A coalescence approach to spatial neutral ecology. Ecological Informatics 3:259–271.

Rousset, F. 2004. Inferences from spatial population genetics. Pages 945–979 *in* D. J. Balding, M. Bishop, and C. Cannings, eds. Handbook of statistical genetics. Wiley, Chichester.

Struebig, M. J., T. Kingston, E. J. Petit, S. C. Le Comber, A. Zubaid, A. Mohd-Adnan, and S. J. Rossiter. 2011. Parallel declines in species and genetic diversity in tropical forest fragments. Ecology Letters 14:582–590.

Taberlet, P., N. E. Zimmermann, T. Englisch, A. Tribsch, R. Holderegger, N. Alvarez, H. Niklfeld, et al. 2012. Genetic diversity in widespread species is not congruent with species richness in alpine plant communities. Ecology Letters 15:1439–1448.

Vellend, M. 2003. Island biogeography of genes and species. American Naturalist 162:358–365.

———. 2005. Species diversity and genetic diversity: parallel processes and correlated patterns. American Naturalist 166:199–215.

Vellend, M., and M. A. Geber. 2005. Connections between species diversity and genetic diversity. Ecology Letters 8:767–781.

Vellend, M., G. Lajoie, A. Bourret, C. Múrria, S. W. Kembel, and D. Garant. 2014. Drawing ecological inferences from coincident patterns of population- and community-level biodiversity. Molecular Ecology 23:2890–2891.

Wakeley, J. 2008. Coalescent theory: an introduction. Roberts, Greenwood Village, CO.

Wares, J. P. 2002. Community genetics in the Northwestern Atlantic intertidal. Molecular Ecology 11:1131–1144.

Webb, C. O., D. D. Ackerly, M. A. McPeek, and M. J. Donoghue. 2002. Phylogenies and community ecology. Annual Review of Ecology and Systematics 33:475–505.

Wehenkel, C., F. Bergmann, and H. R. Gregorius. 2006. Is there a trade-off between species diversity and genetic diversity in forest tree communities? Plant Ecology 185:151–161.

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16: 97–159.

Associate Editor: Uta Berger
Editor: Judith L. Bronstein