Working Paper, v5:
A biodiversity number with explicit speciation

Alternative title: A Neutral Unified Model of Community Evolution

Carlos J. Melián¹*, David Alonso², Stefano Allesina¹, Rampal S. Etienne² and Richard S. Condit³

¹National Center for Ecological Analysis and Synthesis, University of California, 735 State St., Suite 300, Santa Barbara, CA 93101, USA.
³Center for Tropical Forest Science, Smithsonian Tropical Research Institute, Unit 0948, Panama.

*To whom correspondence should be addressed. E-mail: melian@nceas.ucsb.edu, phone: +1-805-892-2527, fax: +1-805-892-2510.
Abstract

Studies focused on speciation have tried to explain the emergence of new species but stopped short of studying what it means for the biodiversity patterns such as abundance or diversity. Community ecologist, on the other hand, have studied how such patterns are maintained, but they did not study the processes that formed the basic components of these patterns. We here take advantage of recent developments in speciation theory to unify neutral theories of molecular evolution and biodiversity. This provides a framework that simultaneously consider community evolution in asexual and sexual populations with explicit mechanisms of speciation. We first calculate the expected speciation rate for two models with asexual and sexual reproduction. We then use maximum likelihood to compare the biodiversity number from those two scenarios with the neutral theory of biodiversity assuming implicit speciation. We found a new biodiversity number for the asexual and sexual reproduction models that takes into account evolving genomes of individuals and a molecular filter that regulates the production of viable offspring. Second, the estimated biodiversity number from the asexual and the sexual reproduction models are consistently larger and lower than the expected from the implicit speciation model, respectively. Asexual evolution generates higher species richness, coexistence and diversity despite its lower speciation rate values when comparing with the sexual model. Thus, diversification rates, diversity and coexistence depends critically on reproduction mode. We discuss the mechanisms that enhance and constrain speciation rate and dramatically alter diversity and coexistence in the context of molecular, sexual–asexual modes of reproduction and ecological factors.

Short running title: A new biodiversity number

Keywords: Neutral Molecular Evolution, Neutral Biodiversity Theory, Genetic graphs, Speciation. Sexual and asexual populations.
Neutral models of community assembly have reinvigorated the study of ecological and evolutionary patterns in an unified framework (Hubbell, 2001; Chave et al., 2002; Chave, 2004; McGill, 2003; Alonso et al., 2006). An essential feature of these models is that they use a minimal set of mechanisms from first principles to test simultaneously multiple levels of biodiversity (i.e., from individual variability in fitness to relative species abundance and community phylogeny). Extensions of models at molecular and ecological levels within the same framework promise to contribute to our understanding of contemporary diversity (Vellend, 2005; Whithman et al., 2006; Johnson & Stinchcombe, 2007; Lankau & Strauss, 2007; Allen & Savage, 2007).

The neutral theory of biodiversity describes speciation in an implicit way by simply introducing a speciation rate at the individual level ($\nu$). In the simplest model, speciation is modeled as point mutation: a single mutation is assumed to lead to a new species. The fundamental biodiversity number in this model is a composite of the speciation rate and the metacommunity size $J_M$: $\theta = 2J_M\nu$ or $\theta = \frac{\nu}{1-\nu}(J_M - 1)$, depending on whether one assumes the Wright-Fisher or Moran model for the dynamics of the metacommunity (Etienne & Alonso 2007). $\theta$ drives the species diversity in the community. Although the speciation rate ($\nu$) is crucial to the model (without it diversity cannot be maintained), the speciation parameter is simply assumed and has no basis in biological processes. Studies of alternative modes of speciation in the neutral framework have similarly assumed a single speciation parameter (Hubbell, 2001; Etienne et al., 2007b; Mouillot & Gaston, 2007; Haegeman & Etienne, 2009). This is a major weakness of these models which otherwise are based on straightforward and easily measured processes (birth, death, dispersal).
It is well known that reproduction mode (i.e., sexual or asexual) and molecular evolution can have a strong influence on genetic diversity, but its effects on speciation rate, species diversity and coexistence of species in evolving communities with explicit speciation have been rarely studied simultaneously (Barraclough et al., 2003; Nee, 2005; Fraser et al., 2009). In this paper we study how explicit description of asexual and sexual reproduction in a DNA evolution model affects the neutral predictions on diversity. Particularly we aim to formulate net speciation rates that can be compared to the phenomenological speciation rates in the original neutral model.

1 The Model

We consider a DNA evolution model with explicit speciation, where the per capita speciation rate is a function of mutation rate ($\mu$), genetic drift, and mating. We assume haploid individuals. The genome of each haploid individual is represented by a sequence of $L$ sites. Each site represents a nucleotide which contains either a purine (state +1) or a pyrimidine (-1) base, resulting in two possible states (+1 or −1). The genome of each individual $i$ can be then written in vector notation as $(S^i_1, S^i_2, ..., S^i_L)$, where $S^i_u$ is the state of the $u^{th}$ site for the individual $i$. The genetic similarity of two individuals, $i$ and $j$, is defined as

$$q^{ij} = \sum_{u=1}^{L} S^i_u S^j_u$$  \hspace{1cm} (1)

We represent the genetic similarity values between all pairs of individuals in a community in a genetic similarity matrix, $Q$ which has elements $q^{ij}$. We assume a molecular constraint as the basis for reproductive isolation (Mayr, 1970; Coyne, 1992; Higgs & Derrida, 1992; Wu, 2001; Gavrilets, 2004; Mallet, 2006): reproductive isolation occurs when the genetic similarity
drops below a minimum similarity $q^{\text{min}}$ (Higgs & Derrida, 1992), i.e., when $q^{ij} < q^{\text{min}}$. In other words, the $q^{\text{min}}$ value works as a molecular filter that defines species according to the biological species concept because viable offspring only arise if the genetic similarity between individuals $i$ and $j$ is larger than $q^{\text{min}}$ (i.e., $q^{ij} > q^{\text{min}}$). Genetic incompatibility driven by the parameter $q^{\text{min}}$ thus implicitly captures the ecological (i.e., prezygotic) and genetic (i.e., postzygotic) mechanisms of speciation (Rice, 1987; Coyne, 1992; Gavrilets, 2004; Coyne & Orr, 2004; Schluter, 2009).

As in the neutral theory of biodiversity we assume ecological drift and symmetric competitive interactions as the only other processes in the metacommunity (Hubbell, 2001). By combining the Higgs-Derrida model on explicit speciation with the neutral model of biodiversity, a theory of community properties (speciation rate, diversity, abundances) can be built entirely from traits of sexual and asexual individuals. It will also allow us to link the fundamental diversity number from the neutral theory of molecular evolution (Kimura, 1968; King & Jukes, 1969; Higgs & Derrida, 1992) with the biodiversity number from the neutral theory of biodiversity (Hubbell, 1979, 2001).

We look at two versions of the model: the one parent model and the two parents model, reflecting asexual and sexual reproduction, respectively (see methods). Here we detail the one parent model and refer to the appendix for the two parents model. In the one parent model (i.e., asexual reproduction), one individual dies at each time step with equal probability for all individuals. Then, one individual reproduces, also with equal probability for all individuals. We assume the infinite sites limit $L \rightarrow \infty$. The expected fraction of nucleotides in the offspring $k$ shared with each individual $i$ in the population is then given by (see section
A.1 appendix):

\[
\begin{align*}
E[q^{ki}] &= e^{-2\mu} (q^{G(k)i}), \\
E[q^{kk}] &= 1.
\end{align*}
\tag{2}
\]

We assume that all individuals have identical sequences initially. Thus the genetic similarity matrix contains elements that are all equal to 1. The evolution of the matrix $Q$ can be formulated explicitly because in the infinite sites limit each pair of nucleotides contributing to the similarity of each pair of individuals in the population ($q^{ki}$) comes with equal probability from the parent of $k$ and each individual $i$ in the population (Higgs & Derrida, 1992) (see appendix).

Higgs and Derrida (1992) numerically studied the $Q$ matrix and found species formation for minimum similarity values larger than the expected mean of the $Q$ matrix (i.e., $q^{\text{min}} > Q^*$). If the minimum genetic similarity value ($q^{\text{min}}$), the mutation rate per genome ($\mu$) and population size are sufficiently large, then the evolutionary dynamics are characterized by a continuous splitting of the population into different reproductively isolated genetic clusters. For example, using the expected genetic similarity of the matrix $Q$ at equilibrium ($Q^* = \frac{1}{\theta + 1}$, where $\theta = 4J\mu$ and $J$ is the effective number of individuals in the population, see appendix), then if $q^{\text{min}} = 0.3$ and $\mu = 10^{-7}$, the population size $J$ must be larger than $12 \cdot 10^6$ for speciation to occur.

At each time step the diversity (of our “genetic” species) can be determined by finding the components of the undirected graph in the genetic similarity matrix $Q$. Each pair of individuals are defined to be in the same component if there exists a direct or indirect path between them. Thus, all the individuals connected at least by one link belong to the same species. The meaning of a link in the sexual model is straightforward: each pair of individuals
with a link can have viable offspring. A link in the asexual model can be interpreted as the possibility of horizontal transfer of plasmids or of recombination. Thus the genetic species concept also applies to asexual organisms (Roberts & Cohan, 1995; Cohan, 1998; Coyne & Orr, 1998).

2 Results

2.1 Speciation modes

In our model two speciation modes emerge: fission and mutation–induced speciation mode (see methods). Fission happens after the death of an individual that was the only link between two genetic clusters. Mutation–induced speciation happens when offspring cannot mate with any individual due to an isolating mutation. Following equation 1, the minimum mutation rate ($\mu_{\text{min}}$) for the mutation-induced speciation mode in the asexual reproduction model is (see methods):

$$\mu_{\text{min}} = -\frac{\log(q_{\text{min}})}{2}$$

For example, if $q_{\text{min}} = 0.95$, the minimum mutation rate to have mutation–induced speciation is $\mu_{\text{min}} \approx 2 \cdot 5 \cdot 10^{-2}$. This value increases with lower is even $q_{\text{min}}$, e.g. for $q_{\text{min}} = 0.90$, $\mu_{\text{min}} = 5.2 \cdot 10^{-2}$ and for $q_{\text{min}} = 0.85$ $\mu_{\text{min}} = 8 \cdot 10^{-2}$. These values represent biologically unrealistic high mutation rates; indeed they are larger than all the values we explored in the simulations (i.e., from $10^{-4}$ to $4.9 \cdot 10^{-3}$). Similar values are obtained in the sexual reproduction model (see section A-3 in the appendix). Note that lower mutation rates can still produce mutation-induced speciation but $\mu_{\text{min}}$ is an expected lower bound especially for
\( q_{\text{min}} \) much closer to 1. Thus fission must account for most of the speciation events.

### 2.2 Expected speciation rate

In order to estimate the speciation rate, let us assume in the asexual model that we have just one individual reproducing itself in a sequence (i.e., individual \( A_1 \) is the offspring of \( A \), thus \( A \rightarrow A_1 \rightarrow A_2, \ldots, A_n \), see Fig. 1). What is the number of steps \( n \) at which \( q^{\text{min}}_\text{AA}\) \( > q^{\text{min}}_\text{A} \) \( \) From equation 1 the number of steps to drop a link between \( A \) and \( A_n \) is (i.e., dashed line in Fig. 1):

\[
  n = -\frac{\log(q^{\text{min}})}{2\mu},
\]

(4)

and the rate of dropping links that is proportional to the per capita speciation rate is:

\[
  \frac{1}{n} = -\frac{2\mu}{\log(q^{\text{min}})},
\]

(5)

which is the equation to approximate the speciation rate \( \nu_{\text{asex}} \) in the figure 2a. The effective speciation rate can be accurately predicted for the two models (see Methods). Figure 2a and 2b represent the speciation rate per generation. Results for the speciation rate per capita are qualitatively the same (data not shown). The functions that best describe the simulated speciation rate values are a power law \( (r^2 = 0.998, p < 0.05) \) and a linear function \( (r^2 = 0.99, p < 0.01) \) for the asexual and the sexual reproduction model, respectively. The per capita speciation rate for the asexual model can be approximated by:

\[
  \nu_{\text{asex}} = \alpha \left[ \frac{-2\mu}{\log(q^{\text{min}})} \right]^\beta,
\]

(6)

where \( \alpha \) is equal to 0.83 and the exponent \( \beta \) is equal to 1.7. The per capita speciation rate
for the sexual model can be approximated by

\[ \nu_{\text{sex}} = \alpha + \beta \left[ -2\mu + \log(\frac{1+E[X]}{2}) \right], \]

(7)

where \( \alpha \) is equal to \(-0.1215\), the slope \( \beta \) is equal to 0.466 and \( X \) is the expected similarity between the two parents. Because the similarity is in the range \([q_{\text{min}}, 1]\), assuming a uniform distribution on the similarity leads to an expected similarity of \( X = \frac{1+q_{\text{min}}}{2} \) (see appendix).

### 2.3 Biodiversity number

We estimated the biodiversity number for communities simulated with the implicit speciation model, the asexual model and the sexual model, using \( \theta = J_M \nu \) where \( \nu \) refers to the speciation rate in the three models respectively. The estimated biodiversity number (\( \theta_{\text{MLE}} \)) is accurately predicted by simulations from the neutral theory of biodiversity (Figs. 3a and 4, and eq. 19 in Methods), as expected.

The estimated biodiversity number from the asexual and the sexual reproduction models are consistently higher and lower than the value expected from the implicit speciation model for low \( \theta \) values, respectively (Fig. 3a). The sexual reproduction model quickly approaches the implicit speciation model for large \( \theta \) values. Note the large coincidence range for large \( \theta \) values between the implicit speciation and the sexual reproduction model when plotting all the replicates for each parameter combination explored (Fig. 4 bottom).

For very small (\( \theta \approx 0 \)) and very large (\( \theta \to \infty \)) \( \theta \) values, the implicit and explicit speciation models should converge (i.e., all individuals are in the same species or each individual
belongs to a different species, respectively), which is indeed what we observe for the sexual model (Figs. 3a and 4 bottom). We don’t see this convergence for the asexual model for very large \( \theta \) values (Figs. 3a and 4 top). We can explain this as follows: The asexual reproduction model has a smaller rate of dropping links than the sexual scenario in the genetic similarity matrix (i.e., \( e^{-2\mu(q^{G(i)}_k)} > e^{-2\mu(q^{G_1(i)}_k + q^{G_2(i)}_k)/2} \)), which implies a lower speciation rate. Perhaps counterintuitively, this lower speciation rate is accompanied by larger incipient species sizes leading to lower extinction rates and generating high values of the estimated \( \theta_{MLE} \).

With the estimated speciation rate we can define a biodiversity number. By multiplying (7) on both sides by the metacommunity size \( J_M \) and by using the regular definition of \( \theta \) for neutral molecular evolution: \( \theta_{NME} = 4J_M\mu \) we find

\[
\theta_{\text{asex}} = J_M\nu_{\text{asex}} = -\alpha \left[ \frac{\theta_{\text{NME}}}{2 \log(q_{\text{min}})} \right]^\beta \left[ \frac{1}{J_M} \right]^{\beta-1}
\]

for the asexual model and

\[
\theta_{\text{sex}} = J_M\alpha + \beta \left[ -\frac{\theta_{\text{NME}}}{2} + J_M \log(\frac{1+X}{2}) \right]
\]

for the sexual model. These approximations accurately predict the simulated values (Fig. 5). Those biodiversity numbers take into account evolving genomes and a molecular filter that regulates the production of viable offspring.

### 2.4 Species richness

We find higher species richness in the asexual model (Fig.3b). As explained above the lower speciation rate is still accompanied by higher richness due to larger incipient species sizes.
leading to lower extinction rates (Fig. 3b). For large values of $\theta$ the high speciation rate of the sexual reproduction model reduces the incipient species size, approaching the value assumed by the neutral theory of biodiversity (i.e., $n_0 = 1$), thus increasing extinction rate and decreasing richness and coexistence.

3 Discussion

By taking advantage of recent developments in speciation theory we found (1) a new biodiversity number for sexual and asexual communities that takes into account evolving genomes of individuals and links explicit mechanisms of speciation with the genetic and species diversity, and (2) that diversification in sexuals is higher than in asexuals, in agreement with previous studies. However, this does not imply higher species richness in sexually reproducing taxa. The sexual model generates incipient species sizes that are smaller than in the asexual model, and thus leads to higher extinction rates than in the asexual model. The asexual model predicts lower speciation rates but combined with larger incipient species size entailing lower extinction rates, it yields higher species richness.

Biodiversity number values from the neutral theory with implicit speciation in metacommunities range between [5-600] (Hubbell, 2001; Volkov et al., 2003; Wootton, 2005; Latimer et al., 2005; Etienne et al., 2006; Dornelas et al., 2006). Our results show that the estimated $\theta$ for the sexual reproduction model approaches the implicit speciation for $\theta$ values larger than 100. This suggest we could estimate the explicit speciation parameters that best fit to the estimated biodiversity numbers (i.e., the mutation rate, $\mu$ and the minimum genetic similarity value to have fertile offspring, $(q^{\min})$) increasing our information on the molecular and
ecological processes that generate diversification and coexistence in ecological communities.

Most estimated $\theta$ values are, however, in the range below 100 for which we found maximum differences between the biodiversity number with implicit and explicit speciation. For example, for $\theta$ implicit $\approx 50$, $\theta_{asex}$ and $\theta_{sex}$ are approximately twice and half this value, respectively. How can we reinterpret the estimated biodiversity numbers with explicit speciation? Most empirical data compared with the neutral expectations have used organisms with sexual and asexual reproduction during their life cycle (i.e., coral reefs, trees, rocky intertidal organisms and fynbos communities). Those numbers with explicit speciation can give us new information about the dominant reproduction mode in those evolving communities. Also, likelihoods from those numbers can test simultaneously several data sets (i.e., species abundance data and molecular diversity) and thus improve the limitations associated with testing neutral expectations using just one data set (Etienne et al., 2006). [Is this speculative?]

Asexual taxa are genetically the most diverse group of life (Fraser et al., 2009), yet there are fewer than $10^4$ described species of Bacteria and Archaea (Oren, 2004). In contrast there are roughly $1.5 \cdot 10^6$ named species of eukaryotes, which are predominantly sexual (i.e., they have sexual reproduction at same stage in their life cycle). Evidently, these numbers may be biased, but one still gets the impression that the eukaryotic domain may be substantially more subdivided than either Bacteria or Archaea (Lynch, 2007), in contrast to our results.

There are, however, several limitations to define species in the same way in the two groups. For example, these groups differ dramatically in horizontal gene transfer between distantly related taxa and variable rates of homologous recombination (Fraser et al., 2007),
and also in the genome size and the rate of base-substitution mutations per nucleotide site (Lynch, 2007). Practical difficulties with sampling bacteria, lack of theory and observations of vast amounts of as yet unclassified microbial diversity have all fuelled the controversy of how one defines bacteria species (Fraser et al., 2009).

Our choice of parameters and model structure inevitably represents a compromise between plausibility and computational limitations, achieved by reducing the effective population size and by using an approximation for the genetic similarity in the infinite genome limit. We chose a particularly simple model exactly because our objective was not (only) to study molecular evolution, but to tackle the implications at the ecological level. This allows us to derive several community-level quantities based on first principles.

Evolutionary graphs have many fascinating extensions and applications. For example, the expected speciation rate estimated here can be tested with the empirical speciation rates for a variety of taxa (Otto & Whitton, 2000; Seehausen, 2006) to answer questions such as: Do the neutral asexual and sexual evolution models predict adaptive radiations from bacteria to fish? What elements should we add to better predict adaptive radiations and coexistence? We have represented models that capture just the two reproductive extremes. There is empirical evidence suggesting that reproductive mode is a continuum mediated by recombination (Fraser et al., 2005). A further question is therefore: How does recombination alter speciation rates, species richness, genetic–species diversity and coexistence in populations with reproduction modes represented as a continuum?

We can foresee extensions of our model that will make use of more refined representations of the molecular and speciation components, especially by taking into account spatial
heterogeneity (Malecot, 1970; Manzo & Peliti, 1994; Gavrilets, 2004; Lynch, 2007), recombina
tion rate (Fraser et al., 2007; Lynch, 2007; Ma et al., 2008) and explicit mechanisms that describe the cause of genetic incompatibilities (Lynch, 2000, 2007; Bikard et al., 2009).

Our models ignore several of these heterogeneities that may arise in populations but may nonetheless provide a preliminary comparative description of the origin and coexistence of diversity by genetic–ecological drift and explicit speciation in evolving asexual and sexual communities. The additions of these heterogeneities could unify several speciation modes with genome evolution and biodiversity in communities with a continuum of reproduction modes.
4 Methods

4.1 The expected speciation rate in the neutral asexual model

Let us assume that we have just one individual reproducing itself in a sequence (i.e., individual $A_1$ is the offspring of $A$, thus $A \rightarrow A_1 \rightarrow A_2 \ldots$, $A_n$, see Fig. 1), what is the number of steps ($n$) at which $q^{min} > q^{AA_n}$? From equation 1 we can represent the first step from $A$ to $A_1$ as:

$$q^{AA_1} = x_1 = e^{-2\mu},$$ (10)

and the second from $A$ to $A_2$ as:

$$q^{AA_2} = x_2 = e^{(-2\mu)^2},$$ (11)

thus

$$q^{AA_n} = x_n = e^{(-2\mu)n},$$ (12)

then

$$q^{min} > e^{(-2\mu)n},$$ (13)

and applying logarithms we obtain:

$$n = -\frac{\log(q^{min})}{2\mu},$$ (14)

which is the number of steps to drop the link between $A$ and $A_n$ (i.e., dashed line in Fig. 3).

The rate of dropping links that is proportional to the per capita speciation rate is:
\[
\frac{1}{n} = -\frac{2\mu}{\log(q_{\min})},
\]  
(15)

which is the equation used to approximate the speciation rate \((\nu_{\text{asex}})\) in the figure 2a. Using this approximation we can ask what is the minimum mutation rate \((\mu_{\min})\) for the mutation-induced speciation mode, thus \(q^{ki} < q^{\min}\)? From equation 1 we have:

\[
q^{\min} > e^{-2\mu} \left(q^{G(k)i}\right),
\]  
(16)

then the equation becomes:

\[
\log(q^{\min}) = -2\mu_{\min},
\]  
(17)

and the minimum mutation rate \((\mu_{\min})\) to have mutation-induced speciation is given by:

\[
\mu_{\min} = -\frac{\log(q^{\min})}{2},
\]  
(18)

which is the equation 3 in the main ms.

### 4.2 Comparing implicit and explicit speciation: Maximum likelihood estimation

We use maximum likelihood to discern the effect of explicit mechanisms of speciation with the implicit speciation assumption in the neutral theory of biodiversity. We estimate the parameter \(\theta\) from the neutral theory of biodiversity that best describes the data from the simulation for each model.
The vector \((C_n|\mu, q^{\text{min}}, J) = (C_1(n), \ldots, C_n(n))\) satisfies \(\sum_{j=1}^{n} JC_j(n) = n\). We then can use the Ewens sampling formula to estimate \((C_n|\theta, J)\) as

\[
P(C_n = a_n|\theta, J) = \frac{n!}{\theta(\theta + 1)\cdots(\theta + n - 1)} \prod_{j=1}^{n} \frac{\theta^{a_j}}{j^{a_j}a_j!},
\]

(19)

where \(a_n = (a_1, a_2, \ldots, a_n)\) is a vector of nonnegative integers satisfying \(a_1 + 2a_2 + \ldots + na_n = n\) and the biodiversity number \(\theta\) is equal to \(J\nu\).

### 4.3 Sampling the models

Our simulations consider a zero-sum birth and death stochastic individual based model with overlapping generations and age independent birth and death rates in the context of neutral mutations and large genome size per individual (effectively infinite gene sequences) (Higgs & Derrida, 1992; Gavrilets, 2004). Individual interactions are introduced using a single and large-homogeneous patch (or metacommunity) in which there is a complete mixing and all individuals have the same chance of potentially interacting with each other (Hubbell, 2001; McKane & Newman, 2004).

Individuals within the community are considered hermaphrodite. Site size inside the patch is defined so that each one contains one individual. All individuals are considered in the reproductive age. These numbers are kept constant by assuming zero-sum dynamics.

An important remark is that our simulations run much faster under the zero-sum rule. Zero-sum models are equivalent to their non zero-sum counterparts at stationarity (Etienne et al., 2007a). Although we have run all our simulations under zero-sum dynamics, we are
quite confident that our main results are robust and do not rely on the specific implementation of the zero-sum rule.

We sampled the steady state for $10^2$ replicates with $2 \times 10^3$ generations each satisfying the condition $q_{min} > Q^*$ for all the combinations. Given $J$ individuals in the initial population, a generation is an update of $J$ time steps. We have explored initial population size, $J = [10^3, 2 \times 10^3]$ individuals, mutation rates, $\mu$, from $10^{-4}$ to $5 \times 10^{-3}$, and the genetic similarity value ($q_{min}$) from 0.85 to 0.95.
References


5 Acknowledgments

We thank Andrew P. Allen, Stanley Harpole, Miguel A. Fortuna, Bill Langford and Jai Ranganathan for useful comments and ideas on the development of the present study. CJM and SA were supported by a Postdoctoral Fellowship at the National Center for Ecological Analysis and Synthesis, a Center funded by NSF (Grant DEB-0553768), the University of California, Santa Barbara, and the State of California. CJM also acknowledges the support by Microsoft Research Ltd., Cambridge, United Kingdom. DA and RSE acknowledge the support of the Netherlands Organization for Scientific Research (NWO).

Support for Richard here!
6 Figure Legends

- Figure 1 represents the approximation to estimate the speciation rate in the asexual reproduction model. We assume individual $A$ reproduces itself in the sequence $A \rightarrow A_1 \rightarrow A_2, \ldots, A_n$, where $A_1$ is the offspring of $A$, $A_2$ is the offspring of $A_1$, etc. We can estimate the speciation rate by approximating the number of steps at which the link ($x_n$) between $A$ and $A_n$ drops.

- Figure 2a represents the approximation of the speciation rate per generation for the asexual reproduction model (results not shown for the speciation rate per capita are qualitatively the same). The x-axis is the approximation given by eq. 5 (i.e., $-2\mu/\log(q_{\text{min}})$). The y-axis is the simulated speciation rate for all the parameter combinations explored (black circles). The best fit is a power law (continuous line). Figure 2b represents the approximation for the sexual reproduction model using $(-2\mu + \log(1 + X^2))/\log(q_{\text{min}})$ (see appendix). The y-axis is the simulated speciation rate for all the parameter combinations explored (black circles). The best fit to this is a linear function. Black circles represent the mean after $10^2$ replicates for all parameter combinations explored.

- Figure 3. Biodiversity number and species richness. a) The x-axis represents the biodiversity number values ($\theta$) from simulations of the neutral theory of biodiversity (solid line), the asexual (thick line) and the sexual reproduction model (dotted lines). The y-axis represents the best estimation represented as the mean after $10^2$ replicates using likelihood from equation 19 (see methods). b) The mean species richness after $10^2$ replicates as a function of the minimum genetic similarity value ($q_{\text{min}}$) and speciation rate per generation ($\nu$) for all the mutation values explored, $\mu = [10^{-4}, 5 \cdot 10^{-4}, 9 \cdot 10^{-4}, 1.3 \cdot 10^{-3}, 1.7 \cdot 10^{-3}, 2.1 \cdot 10^{-3}]$. 
2.5·10^{-3}, 2.9·10^{-3}, 3.3·10^{-3}, 3.7·10^{-3}, 4.1·10^{-3}, 4.5·10^{-3}, 4.9·10^{-3}). Note that diversity and species richness values are higher in the asexual model despite the lower speciation rate values for each combination of $q_{\text{min}}$ and $\mu$.

- Figure 4 represents the data for $J = 10^3$ (left) and $2 \times 10^3$ (right) individuals for all the parameter combinations and the $10^2$ replicates explored. The x-axis represents the biodiversity number values ($\theta$) from simulations of the neutral theory of biodiversity (red line is the mean after $10^2$ replicates and red circles represent the $10^2$ replicates). White circles in top and bottom represent the asexual and sexual model, respectively. The y-axis represents the estimations for the $10^2$ replicates using maximum likelihood from equation 19 ($[\theta_{\text{MLE}}]$).

- Figure 5. The mean of the simulated biodiversity number after $10^2$ replicates for the asexual (panel a) and the sexual (panel b) reproduction model (dots). The solid line is the approximation of the biodiversity number given by equations 8 and 9, respectively. The x-axis, y-axis and the z-axis represent the $q_{\text{min}}$, $\mu$ and $\theta_{\text{asex}}$ ($\theta_{\text{sex}}$) values, respectively.
Fig. 1
Fig. 2

- $-2\mu / \log(q_{\text{min}})$

- $(-2\mu + \log((1 + X)/2)) / \log(q_{\text{min}})$
Fig. 3
Fig. 4
Fig. 5