

## Neutral Models with Generalised Speciation

Bart Haegeman<sup>a,\*</sup>, Rampal S. Etienne<sup>b</sup>

<sup>a</sup>*MERE Research Team, INRIA Sophia Antipolis—Méditerranée, UMR Systems Analysis and Biometrics, 2 place Pierre Viala, 34060 Montpellier, France*

<sup>b</sup>*Community and Conservation Ecology Group, Centre for Ecological and Evolutionary Studies, University of Groningen, P.O. Box 14, 9750 AA Haren, The Netherlands*

Received: 17 April 2008 / Accepted: 10 February 2009 / Published online: 12 March 2009  
© Society for Mathematical Biology 2009

**Abstract** Hubbell’s neutral theory claims that ecological patterns such as species abundance distributions can be explained by a stochastic model based on simple assumptions. One of these assumptions, the point mutation assumption, states that every *individual* has the same probability to speciate. Etienne et al. have argued that other assumptions on the speciation process could be more realistic, for example, that every *species* has the same probability to speciate (Etienne, et al. in *Oikos* 116:241–258, 2007). They introduced a number of neutral community models with a different speciation process, and conjectured formulas for their stationary species abundance distribution. Here we study a generalised neutral community model, encompassing these modified models, and derive its stationary distribution, thus proving the conjectured formulas.

**Keywords** Biodiversity · Ewens sampling formula · Metacommunity · Neutral model · Speciation · Species abundance distribution

### 1. Introduction

Since their introduction a few years ago (Hubbell, 2001; Bell, 2001), neutral community models have gained much interest in community ecology. These models aim to explain commonly observed macro-ecological patterns, such as species abundance distributions and species area relationships, by simple stochastic rules and the assumption of functional equivalence between conspecific as well as heterospecific individuals (see, e.g., Etienne, 2005; Rosindell and Cornell, 2007). Hubbell’s model describes the endless repetition of randomly assigning an individual to die, and immediately replacing it by another individual. One distinguishes two spatial scales. On the large scale, the so-called metacommunity, the replacing individual originates either by birth or by speciation (i.e., birth of a mutant). On a smaller scale, the so-called local community, the replacing individual originates by birth or by immigration from the metacommunity. This results in community composition

\*Corresponding author.

E-mail addresses: [bart.haegeman@inria.fr](mailto:bart.haegeman@inria.fr) (Bart Haegeman), [r.s.etienne@rug.nl](mailto:r.s.etienne@rug.nl) (Rampal S. Etienne).

dynamics governed by species extinction and immigration in the local community, and by extinction and speciation in the metacommunity.

Here we focus on the stationary species abundance distribution predicted by the neutral metacommunity model, and in particular how it is affected by the speciation process. In Hubbell's model there is a fixed probability  $\nu$  that a death event is followed by a speciation event; the probability that a death event is followed by a birth event is then  $1 - \nu$ . This means that all *individuals* have the same probability to speciate. However, this assumption has been scrutinised by Etienne et al. (2007), because speciation theorists usually assume that rather all *species* have the same probability to speciate (Stanley, 1975; Etienne and Apol, 2009), which is supported by, albeit scarce, data (Makarieva and Gorshkov, 2004).

Etienne et al. (2007) showed that alternative assumptions on the speciation process can be incorporated in neutral metacommunity models, without drastically modifying the mathematical framework. They proposed a stochastic model, where repeatedly a pair of individuals are sampled from the metacommunity. As in Hubbell's model, the first individual dies, and the second individual reproduces leaving either a conspecific individual or a new species (due to mutation). However, the abundance  $\ell$  of the species the second individual belongs to, is used to determine the probability  $\nu(\ell)$  of a speciation event. Taking  $\nu(\ell)$  independent of  $\ell$  ( $\nu(\ell) = \nu_1$ ), as in Hubbell's model, can be considered as speciation with a rate proportional to the number of individuals in the community, i.e., every individual has the same probability to speciate. Taking  $\nu(\ell)$  inversely proportional to  $\ell$  ( $\nu(\ell) = \frac{\nu_0}{\ell}$ ) corresponds to a speciation rate proportional to the number of species in the metacommunity, i.e., every species has the same probability to speciate.

In this paper we investigate the neutral metacommunity model with a general function  $\nu(\ell)$ . We first describe the model as a continuous-time Markov process, so that the corresponding mathematical theory can be applied, guaranteeing existence and uniqueness of the stationary species abundance distribution. Next, we try to compute this stationary distribution, by imposing the so-called detailed balance conditions. They can be satisfied for a restricted class of functions  $\nu(\ell)$ , including the two special cases mentioned above:  $\nu$  constant and  $\nu(\ell) \propto \frac{1}{\ell}$ . In the former case, the stationary distribution is given by Ewens' sampling formula (Ewens, 1972). In the latter case, we retrieve the formula given in Etienne et al. (2007). Etienne et al. (2007) only showed that this latter formula is correct for small community size, and that it produces the correct mean values for any community size. We provide a formal proof of this formula for any community size.

## 2. Markov process for species abundances

In neutral metacommunity models, species are constantly originating and disappearing through speciation and extinction. If we take two snapshots, separated by a sufficient amount of time, the snapshots will have no species in common. A stationary state is therefore only biologically meaningful when we neglect species identity. Rather than specifying the abundance of every species, it suffices to specify the number of species  $S_k$  with abundance  $k$ . We assume here that the metacommunity size, which we denote by  $J_M$ , is constant over time. Then we can collect these numbers in a vector  $\vec{S}$  of size  $J_M$ ,

$$\vec{S} = (S_1, S_2, \dots, S_{J_M}) \quad \text{with} \quad \sum_{k=1}^{J_M} k S_k = J_M. \quad (1)$$

The metacommunity models we consider, for which the state at time  $t$  is given by a vector  $\vec{S}$ , are Markov processes. This means that the probability to go from the state  $\vec{S}_1$  at the present time  $t_1$  to a state  $\vec{S}_2$  at a future time  $t_2$ , does not depend on states in the past (i.e. before  $t_1$ ). For mathematical convenience we choose a continuous-time description, but the discrete-time description is very similar. We denote the transition rate from  $\vec{S}_1$  to  $\vec{S}_2$  by  $R(\vec{S}_1, \vec{S}_2)$ . Given some initial probability distribution  $P(0, \vec{S})$ , the dynamics determine the probability distribution  $P(t, \vec{S})$  of the system being in state  $\vec{S}$  at time  $t$ . They are described by the so-called master equation (Van Kampen, 1992),

$$\frac{d}{dt}P(t, \vec{S}_1) = \sum_{\vec{S}_2 \neq \vec{S}_1} \left( P(t, \vec{S}_2)R(\vec{S}_2, \vec{S}_1) - P(t, \vec{S}_1)R(\vec{S}_1, \vec{S}_2) \right). \quad (2)$$

The master equation is nothing but a continuity equation for probabilities. It states that the probability of state  $\vec{S}_1$  can only increase by having a transition from any other state  $\vec{S}_2$  to  $\vec{S}_1$  (first term in the right-hand side), and can only decrease by having a transition from  $\vec{S}_1$  to any other state  $\vec{S}_2$  (second term in the right-hand side).

Stationary species abundance distributions  $P(\vec{S})$  should satisfy the equations obtained by putting the right-hand side of (2) equal to zero,

$$\sum_{\vec{S}_2 \neq \vec{S}_1} \left( P(\vec{S}_2)R(\vec{S}_2, \vec{S}_1) - P(\vec{S}_1)R(\vec{S}_1, \vec{S}_2) \right) = 0, \quad \text{for all } \vec{S}_1. \quad (3)$$

As the state space of all possible abundance vectors  $\vec{S}$  is finite (due to fixed community size  $J_M$ ), many results are known about the stationary distribution. For the transition rates  $R(\vec{S}_1, \vec{S}_2)$  that we specify in the next section, it is guaranteed that any initial distribution  $P(0, \vec{S})$  will converge to a unique stationary distribution  $P(\vec{S})$  (Van Kampen, 1992).

Nevertheless, the stationary equations (3) are often too complicated to solve explicitly. A set of equations that is much easier to handle, is given by

$$P(\vec{S}_1)R(\vec{S}_1, \vec{S}_2) = P(\vec{S}_2)R(\vec{S}_2, \vec{S}_1), \quad \text{for all } \vec{S}_1, \vec{S}_2. \quad (4)$$

By imposing these equations, we require that not only the sums in (3) vanish, but each individual term in (3). The conditions (4) can be readily interpreted. They state that the frequency of transitions from  $\vec{S}_1$  to  $\vec{S}_2$  equals the frequency of transitions from  $\vec{S}_2$  to  $\vec{S}_1$ . As every transition should be balanced by its reverse transition, these conditions are known as detailed balance (Van Kampen, 1992, see also Haegeman and Etienne, 2008). Obviously, the set of detailed balance conditions (4) is much stronger than the set of conditions (3). Therefore, if (4) has a solution, it must be the unique stationary distribution  $P(\vec{S})$  we are looking for. However, it can also happen that (4) has no solution at all. We will encounter examples of both cases.

### 3. Metacommunity model with generalised speciation

We define a neutral metacommunity model that generalises Hubbell's model, and the different models studied in Etienne et al. (2007). As in Hubbell's model, all transitions consist of first decreasing the abundance of one species, and next increasing the abundance

of another species. Suppose that before the transition, the first species has abundance  $k$ , and the second species abundance  $\ell$ . Hence, after the transition, we have one species less with abundance  $k$ , one species more with abundance  $k - 1$ , one species less with abundance  $\ell$ , and one species more with abundance  $\ell + 1$ . The transition therefore goes from state  $\vec{S}$  to state  $\vec{S} - \vec{e}_k + \vec{e}_{k-1} - \vec{e}_\ell + \vec{e}_{\ell+1}$  (the unit vector  $\vec{e}_k$  has component  $k$  equal to 1, and all other components equal to zero). Note that this is only valid for  $k > 1$  and  $\ell > 0$ . In the case  $k = 1$ , the first species disappears. In the case  $\ell = 0$ , the second species originates by speciation. The general transition is therefore from  $\vec{S}$  to

$$\vec{S}_{k,\ell} = \vec{S} - \vec{e}_k + \vec{e}_{k-1}\delta_{k>1} - \vec{e}_\ell\delta_{\ell>0} + \vec{e}_{\ell+1}, \tag{5}$$

where we introduced to shorthand notation  $\vec{S}_{k,\ell}$  (the delta function  $\delta_C$  equals 1 if condition  $C$  is satisfied, and 0 otherwise).

To specify the model completely, we have to choose the transition rates  $R(\vec{S}, \vec{S}_{k,\ell})$ , for which we use the shorthand notation  $R_{k,\ell}(\vec{S})$  from hereon, for notational convenience. As mentioned in the introduction, we let the speciation probability  $v(\ell)$  depend on the abundance  $\ell$  of the species the newly speciated individual originates from. For transitions consisting of a death in the first species (abundance  $k$ ) followed by a reproduction in the second species (abundance  $\ell$ ), the rate is

$$R_{k,\ell}(\vec{S}) = \frac{kS_k}{J_M} (1 - v(\ell)) \frac{\ell S_\ell}{J_M}, \tag{6a}$$

because there are  $S_k$  species with abundance  $k$ , and  $S_\ell$  species with abundance  $\ell$ . In the case  $k = \ell$ , this rate should be modified to exclude transitions where the same species dies and reproduces (which has no effect in terms of the vector  $\vec{S}$ ):

$$R_{k,k}(\vec{S}) = \frac{kS_k}{J_M} (1 - v(k)) \frac{k(S_k - 1)}{J_M}. \tag{6b}$$

Note that we slightly deviate from Hubbell’s model, because in Hubbell’s model the individual that dies cannot leave off-spring, where here this individual can still contribute to the next generation which is the so-called Moran model (Moran, 1962; Etienne et al., 2007; Etienne and Alonso, 2007). Finally, for transitions consisting of a death in the first species followed by a speciation, the rate is

$$R_{k,0}(\vec{S}) = \frac{kS_k}{J_M} \sum_{m=1}^{J_M} v(m) \frac{mS_m}{J_M}, \tag{6c}$$

where the sum is over all species, because each species may have produced the new species. The total speciation rate  $R_{\text{spec}}(\vec{S})$  in the metacommunity can be obtained by summing the latter transition rate over all abundances  $k$ ,

$$R_{\text{spec}}(\vec{S}) = \sum_{k=1}^{J_M} R_{k,0}(\vec{S}) = \sum_{m=1}^{J_M} v(m) \frac{mS_m}{J_M}. \tag{7}$$

We consider a few special cases. To recover Hubbell's metacommunity model, we take  $\nu(\ell) = \nu_1$  independent of  $\ell$ . In that case, we have

$$R_{\text{spec}}(\vec{S}) = \nu_1 \sum_{m=1}^{J_M} \frac{m S_m}{J_M} = \nu_1, \quad (8)$$

which is a constant speciation rate. Because the number of individuals  $J_M$  is also constant, every individual has the same probability to speciate. To be precise, this is the continuous-time version of the Hubbell–Moran model, called  $M_1$  by Etienne et al. (2007).

The alternative metacommunity models considered in Etienne et al. (2007) are also contained in our framework. Taking  $\nu(\ell) = \frac{\nu_0}{\ell}$  inversely proportional to  $\ell$ , we get

$$R_{\text{spec}}(\vec{S}) = \nu_0 \sum_{m=1}^{J_M} \frac{S_m}{J_M} = \nu_0 \frac{S_M(\vec{S})}{J_M}, \quad (9)$$

with  $S_M(\vec{S})$  the number of species in the metacommunity state  $\vec{S}$ . The speciation rate is then proportional to the number of species, i.e., every species has the same probability to speciate. This model was called  $M_0$  by Etienne et al. (2007).

Etienne et al. (2007) also investigated two models intermediate between  $M_0$  and  $M_1$ . One model,  $M_c$ , has a speciation rate that linearly combines the speciation rates of models  $M_0$  and  $M_1$ ,

$$\nu(\ell) = \frac{\nu_0}{\ell} + \nu_1. \quad (10)$$

Another model,  $M_s$ , assumes a saturation-type dependence for the speciation rate,

$$\nu(\ell) = \frac{\nu_1 x_s}{x_s + \frac{\ell}{J_M}}. \quad (11)$$

The constant  $x_s$  is a characteristic relative species abundance. A species with relative abundance  $\ell/J_M \ll x_s$  has approximately speciation rate  $\nu_1$ , which is independent of its abundance; this corresponds to model  $M_1$ . At the other extreme, a species with relative abundance  $\ell/J_M \gg x_s$  has approximately speciation rate  $J_M \nu_1 x_s / \ell$ , inversely proportional to its abundance  $\ell$ ; this corresponds to model  $M_0$ . Model  $M_s$  is arguably biologically more realistic than model  $M_c$  (Etienne et al., 2007).

#### 4. Solving detailed balance conditions

In this section we study the stationary distribution equations (3) with the transition rates (6a–6c). Although we know that there is a unique distribution satisfying this set of conditions (Van Kampen, 1992), we are only able to solve these equations when the detailed balance conditions (4) holds. In this case, the detailed balance solution is necessarily the unique stationary distribution. Because these equations are easy to solve, we obtain explicitly the stationary distribution. If, on the other hand, conditions (4) cannot be satisfied simultaneously, our method does not provide a solution. Experience tells us that in such

cases there is often no closed-form analytical solution. Fortunately, experience also tell us that solutions to the detailed balance equations can often be found for simple models.

We will proceed as follows. We first write down the detailed balance equations for the model given by the transition rates (6a–6c). We then consider two subsets of all possible transitions that leads to a conjecture for the full stationary distribution. We then prove that this conjecture is indeed correct.

Consider the transition  $R_{k,\ell}(\vec{S})$  that goes from  $\vec{S}$  to  $\vec{S}_{k,\ell}$ . Its reverse is the transition  $R_{\ell+1,k-1}(\vec{S}_{k,\ell})$  that goes from  $\vec{S}_{k,\ell}$  back to  $\vec{S}$ . Detailed balance (4) can then be written as

$$P(\vec{S})R_{k,\ell}(\vec{S}) = P(\vec{S}_{k,\ell})R_{\ell+1,k-1}(\vec{S}_{k,\ell}), \tag{12}$$

for  $k = 1, \dots, J_M$ ,  $\ell = 0, \dots, J_M - 1$ , and all  $\vec{S}$ . Imposing detailed balance (12) yields the ratio of two components of the abundance distribution we are looking for. Repeating this for different combinations  $k, \ell$  and  $\vec{S}$ , we can construct the entire distribution, up to a constant, which can be found by requiring that the components of a probability distribution should sum up to one.

First, we consider transitions that conserve the number of species  $S_M(\vec{S})$ , i.e., transitions without speciation or species extinction. The rates are given by (6a). From (12), we find

$$\begin{aligned} \frac{P(\vec{S}_{k,\ell})}{P(\vec{S})} &= \frac{R_{k,\ell}(\vec{S})}{R_{\ell+1,k-1}(\vec{S}_{k,\ell})} \\ &= \frac{\frac{kS_k\ell S_\ell}{J_M^2}(1 - \nu(\ell))}{\frac{(\ell+1)(S_{\ell+1}+1)(k-1)(S_{k-1}+1)}{J_M^2}(1 - \nu(k-1))} \\ &= \frac{k\ell}{(k-1)(\ell+1)} \frac{S_k S_\ell}{(S_{k-1}+1)(S_{\ell+1}+1)} \frac{1 - \nu(\ell)}{1 - \nu(k-1)}. \end{aligned} \tag{13}$$

This condition should hold for all  $k, \ell$  (with  $\ell \neq k$ ) and  $\vec{S}$  (with fixed  $S_M(\vec{S})$ ). We can iterate this formula to link the different components of the detailed balance distribution (if it exists). We show in [Appendix A](#) that this leads to

$$P(\vec{S}) = C_1(S_M(\vec{S})) \prod_{k=1}^{J_M} \frac{1}{k^{S_k} S_k!} \left( \prod_{m=1}^{k-1} (1 - \nu(m)) \right)^{S_k}, \tag{14}$$

for all vectors  $\vec{S}$  (recall that all considered vectors  $\vec{S}$  have the same number of individuals  $J_M$ ). Note that the proportionality factor  $C_1$  depends on the state  $\vec{S}$  only through the number of species  $S_M(\vec{S})$ .

Next, we investigate transitions in which the number of species changes. For example, consider a speciation transition  $R_{k,0}(\vec{S})$  with rate given by (6c), together with its reverse transition  $R_{1,k-1}(\vec{S}_{k,0})$ , in which a species goes extinct. From (12),

$$\frac{P(\vec{S}_{k,0})}{P(\vec{S})} = \frac{R_{k,0}(\vec{S})}{R_{1,k-1}(\vec{S}_{k,0})}$$

$$\begin{aligned}
 &= \frac{\frac{kS_k}{J_M^2} \sum_{m=1}^{J_M} mS_m \nu(m)}{\frac{(S_1+1)(k-1)(S_{k-1}+1)}{J_M^2} (1 - \nu(k-1))} \\
 &= \frac{k}{k-1} \frac{S_k}{(S_{k-1}+1)(S_1+1)} \frac{\sum_{m=1}^{J_M} mS_m \nu(m)}{1 - \nu(k-1)}. \tag{15}
 \end{aligned}$$

At the same time, from (14), we must have

$$\begin{aligned}
 \frac{P(\vec{S}_{k,0})}{P(\vec{S})} &= \frac{P(\vec{S} - \vec{e}_k + \vec{e}_{k-1} + \vec{e}_1)}{P(\vec{S})} \\
 &= \frac{C_1(S_M(\vec{S}) + 1)}{C_1(S_M(\vec{S}))} \frac{\frac{1}{(k-1)(S_{k-1}+1)} \frac{1}{S_1+1} \prod_{m=1}^{k-2} (1 - \nu(m))}{\frac{1}{kS_k} \prod_{n=1}^{k-1} (1 - \nu(n))} \\
 &= \frac{C_1(S_M(\vec{S}) + 1)}{C_1(S_M(\vec{S}))} \frac{k}{k-1} \frac{S_k}{(S_{k-1}+1)(S_1+1)} \frac{1}{1 - \nu(k-1)}. \tag{16}
 \end{aligned}$$

Comparing (15) and (16), we see that the proportionality factor  $C_1$  has to satisfy

$$\frac{C_1(S_M(\vec{S}) + 1)}{C_1(S_M(\vec{S}))} = \sum_{m=1}^{J_M} mS_m \nu(m), \tag{17}$$

for all vectors  $\vec{S}$ . However, the left-hand side of this equation depends on the vector  $\vec{S}$  only through the number of species  $S_M(\vec{S})$ . Hence, the detailed balance conditions (12) can be satisfied only if the probabilities  $\nu(m)$  are such that

$$\sum_{m=1}^{J_M} mS_m \nu(m) \text{ depends on } \vec{S} \text{ only through } S_M(\vec{S}). \tag{18}$$

For probabilities  $\nu(m)$  for which this condition holds, we can define the function  $V$  such that

$$V(S_M(\vec{S})) = \sum_{m=1}^{J_M} mS_m \nu(m). \tag{19}$$

Equation (17) can then be rewritten as

$$C_1(S_M(\vec{S})) = C_2 \prod_{m=1}^{S_M(\vec{S})-1} V(m), \tag{20}$$

where the proportionality constant  $C_2$  does not depend on  $\vec{S}$  (but it can still depend on  $J_M$ ). Substituting (20) into (14) leads to

$$P(\vec{S}) = C_2 \left[ \prod_{m=1}^{S_M(\vec{S})-1} V(m) \right] \left[ \prod_{k=1}^{J_M} \frac{1}{k^{S_k} S_k!} \left( \prod_{m=1}^{k-1} (1 - \nu(m)) \right)^{S_k} \right]. \tag{21}$$

Note that we have constructed the distribution (21) by considering only a subset of the possible transitions (e.g., we have not used transition rate (6b)). However, in order that (21) is the stationary distribution, condition (12) should hold for all possible transitions. In Appendix B we prove that this is indeed the case. Therefore, we have arrived at a characterisation of the metacommunity model with generalised speciation in terms of detailed balance:

- If condition (18) holds, the stationary distribution is the detailed balance solution, explicitly given by (21).
- If condition (18) does not hold, the detailed balance conditions (12) are conflicting. We cannot find a closed-form analytical solution for the stationary distribution (although a solution does exist).

We conclude by considering the special cases mentioned previously:

Model  $M_1$  (constant speciation rate per individual) corresponds to  $\nu(\ell) = \nu_1$ . In that case, condition (18) holds,  $V(m) = \nu_1 J_M$ , and solution (21) becomes

$$P(\vec{S}) = C_3 \prod_{k=1}^{J_M} \frac{\theta^{S_k}}{k^{S_k} S_k!} \quad \text{with } \theta = \frac{\nu_1}{1 - \nu_1} J_M. \quad (22)$$

After determining the proportionality constant  $C_3$  by normalising the probability distribution, we recover Eq. (2) in (Etienne et al., 2007), closely related to Ewens' sampling formula (Ewens, 1972).

Model  $M_0$  (constant speciation rate per species) corresponds to  $\nu(\ell) = \frac{\nu_0}{\ell}$ . Condition (18) holds,  $V(m) = \nu_0 m$ , and solution (21) becomes

$$P(\vec{S}) = C_4 (S_M(\vec{S}) - 1)! \prod_{k=1}^{J_M} \frac{\nu_0^{S_k}}{k^{S_k} S_k!} \left( \prod_{m=1}^{k-1} \left( 1 - \frac{\nu_0}{m} \right) \right)^{S_k}. \quad (23)$$

After determining the proportionality constant  $C_4$ , we recover Eq. (4) in (Etienne et al., 2007).

Model  $M_c$  (linear combination of speciation rates  $M_0$  and  $M_1$ ) corresponds to the formula for  $\nu(\ell)$  in (10). Condition (18) holds, and we obtain Eq. (5) in (Etienne et al., 2007).

Model  $M_s$  (saturation characteristic between speciation rates  $M_0$  and  $M_1$ ) corresponds to the formula for  $\nu(\ell)$  in (11). In that case, condition (18) does not hold, and our method does not provide the stationary distribution, even though we know that a unique stationary distribution exists.

## 5. Concluding remarks

We have analysed a set of neutral metacommunity models with point mutation speciation, which is generalised in the sense that any dependence of the speciation rate on species' abundances is allowed. This set of models includes Hubbell's model (Hubbell, 2001) and the models recently introduced by Etienne et al. (2007). Using a master equation approach and imposing the detailed balance condition, we were able to rigorously obtain



the stationary species abundance distribution for a subclass of these models. This subclass includes the simplest speciation assumptions: a total speciation rate proportional to the number of individuals (as in Hubbell's model), and a total speciation rate proportional to the number of species (as proposed by Etienne et al., 2007). For more complicated speciation processes, such as a speciation rate saturating with species abundance (see Etienne et al., 2007), detailed balance is not satisfied and therefore does not yield the stationary distribution.

We have rigorously proved the stationary abundance distributions presented in Etienne et al. (2007). Based on these distributions Etienne et al. (2007) derived so-called sampling formulas both for the metacommunity and the local community. Sampling formulas give the species abundance distribution for a sample of individuals randomly taken from a community. They are therefore of great importance in community ecology as they allow us to distinguish between different models of community structure based on species abundance data of samples rather than the complete community. Species abundance data are ubiquitous and believed to contain relevant information on the forces shaping ecological communities (McGill et al., 2007). Sampling formulas thus provide access to these forces.

With sampling formulas Etienne et al. (2007) showed that model  $M_1$  almost always outperformed models  $M_0$  and  $M_c$  in fitting species abundance distributions of various tree communities. Only tree communities having very abundant species were better explained by model  $M_0$  as highly abundant species are extremely unlikely under model  $M_1$  because the per species speciation rate increases linearly with abundance in the latter model. Etienne et al. (2007) introduced the intermediate model  $M_s$  that resembles  $M_0$  at high abundances and  $M_1$  at low abundances but were not able to derive the stationary abundance distribution for this more realistic model. Here we have shown that finding a closed-form expression for this distribution is unlikely due to the fact that detailed balance does not hold for  $M_s$ .

The master equation (2) provides the appropriate mathematical framework to study speciation processes in neutral community theory, as it describes Markov processes in terms of abundance vectors for unlabeled species. For example, speciation mechanisms where new species can originate with several individuals at once, such the random fission model (Hubbell, 2001), also have a master equation of the form (2). The straightforward approach to solve the master equation by imposing detailed balance, as used here, does not work for such more complex models. Other analytical techniques, or different model assumptions, will be needed to obtain (possibly only approximately) the stationary species abundance distributions for this type of models.

## Acknowledgements

We thank the anonymous reviewer for constructive comments. RSE acknowledges the financial support of the Netherlands Organisation for Scientific Research (NWO).

## Appendix A

In this appendix we argue that (13) implies (14). Transitions (6a) take the system from state  $\vec{S}$  to state  $\vec{S} - \vec{e}_k + \vec{e}_{k-1} - \vec{e}_\ell + \vec{e}_{\ell+1}$ . A species with initial abundance  $k$  has lost an

individual, whereas a species with initial abundance  $\ell$  has gained an individual. By putting together several transitions of this kind, we can let species lose or gain several individuals. In particular, we consider a composed transition where  $n$  species with initial abundance  $k$  all lose  $m$  individuals, and  $n$  species with initial abundance  $\ell$  all gain  $m$  individuals. The state vector then changes from  $\vec{S}$  to  $\vec{S}' = \vec{S} - n\vec{e}_k + n\vec{e}_{k-m} - n\vec{e}_\ell + n\vec{e}_{\ell+m}$ . This composed transition thus consists of  $mn$  elementary transitions (6a). The ratio  $P(\vec{S}')/P(\vec{S})$  can be computed by taking a sequence of  $mn$  transitions (6a) going from  $\vec{S}$  to  $\vec{S}'$ , and multiplying the corresponding  $mn$  ratios (13) together. This gives

$$\begin{aligned} \frac{P(\vec{S}')}{P(\vec{S})} &= \left[ \frac{k\ell}{(k-m)(\ell+m)} \right]^n \\ &\times \frac{S_k(S_k-1)\dots(S_k-n+1)S_\ell(S_\ell-1)\dots(S_\ell-n+1)}{(S_{k-m}+1)\dots(S_{k-m}+n)(S_{\ell+m}+1)\dots(S_{\ell+m}+n)} \\ &\times \left[ \frac{(1-\nu(\ell))(1-\nu(\ell+1))\dots(1-\nu(\ell+m-1))}{(1-\nu(k-1))(1-\nu(k-2))\dots(1-\nu(k-m))} \right]^n. \end{aligned} \tag{A.1}$$

Note that this result does not depend on the chosen sequence of transitions, which is a necessary condition for detailed balance to hold. Introducing the notation

$$W(r) = \prod_{s=1}^{r-1} (1-\nu(s)), \tag{A.2}$$

we can rewrite (A.1) as follows:

$$\begin{aligned} \frac{P(\vec{S}')}{P(\vec{S})} &= \frac{k^n \ell^n}{(k-m)^n (\ell+m)^n} \\ &\times \frac{S_k! S_\ell! S_{k-m}! S_{\ell+m}!}{(S_k-n)! (S_\ell-n)! (S_{k-m}+n)! (S_{\ell+m}+n)!} \\ &\times \frac{W(\ell+m)^n W(k-m)^n}{W(\ell)^n W(k)^n} \\ &= \frac{k^{S_k-S'_k} \ell^{S_\ell-S'_\ell}}{(k-m)^{S'_{k-m}-S_{k-m}} (\ell+m)^{S'_{\ell+m}-S_{\ell+m}}} \\ &\times \frac{S_k! S_\ell! S_{k-m}! S_{\ell+m}!}{(S'_k)! (S'_\ell)! (S'_{k-m})! (S'_{\ell+m})!} \\ &\times \frac{W(\ell+m)^{S'_{\ell+m}-S_{\ell+m}} W(k-m)^{S'_{k-m}-S_{k-m}}}{W(\ell)^{S_\ell-S'_\ell} W(k)^{S_k-S'_k}}. \end{aligned} \tag{A.3}$$

Reorganising this equation, we get

$$\begin{aligned} &\frac{k^{S_k} \ell^{S_\ell} (k-m)^{S_{k-m}} (\ell+m)^{S_{\ell+m}} S_k! S_\ell! S_{k-m}! S_{\ell+m}!}{W(k)^{S_k} W(\ell)^{S_\ell} W(k-m)^{S_{k-m}} W(\ell+m)^{S_{\ell+m}}} P(\vec{S}) \\ &= \frac{k^{S'_k} \ell^{S'_\ell} (k-m)^{S'_{k-m}} (\ell+m)^{S'_{\ell+m}} S'_k! S'_\ell! S'_{k-m}! S'_{\ell+m}!}{W(k)^{S'_k} W(\ell)^{S'_\ell} W(k-m)^{S'_{k-m}} W(\ell+m)^{S'_{\ell+m}}} P(\vec{S}'). \end{aligned} \tag{A.4}$$

The left-hand side is a function of  $\vec{S}$  (i.e., it does not depend on  $\vec{S}'$ ), while the right-hand side equals the same function evaluated in  $\vec{S}'$ . Denoting this function by  $F_1$ ,

$$F_1(\vec{S}) = \left( \prod_{r \in \{k, \ell, k-m, \ell+m\}} \frac{r^{S_r} S_r!}{W(r)^{S_r}} \right) P(\vec{S}), \tag{A.5}$$

Eq. (A.4) says that the detailed balance solution (if it exists) satisfies

$$F_1(\vec{S}) = F_1(\vec{S}'), \tag{A.6}$$

for states  $\vec{S}$  and  $\vec{S}'$  that are linked by the composed transitions we have been considering, for variable  $n$  and constant  $k, \ell$  and  $m$ . Note that the set  $\{k, \ell, k - m, \ell + m\}$  in (A.5) contains the indices of the components of  $\vec{S}$  that change when going from  $\vec{S}$  to  $\vec{S}'$ . If we vary  $k, \ell$  and  $m$ , the other components of  $\vec{S}$  can be modified, and we get a more general condition for the detailed balance solution. Defining the function  $F_2$  by

$$F_2(\vec{S}) = \left( \prod_{r=1}^{J_M} \frac{r^{S_r} S_r!}{W(r)^{S_r}} \right) P(\vec{S}), \tag{A.7}$$

this more general condition reads

$$F_2(\vec{S}) = F_2(\vec{S}'), \tag{A.8}$$

for states  $\vec{S}$  and  $\vec{S}'$  that are linked by composed transitions, for variable  $k, \ell, m$  and  $n$ . In fact, any two states that have the same number of species (they all have the same number of individuals  $J_M$ ) are linked in this way. Therefore, detailed balance implies that condition (A.8) holds for all states  $\vec{S}$  and  $\vec{S}'$  for which  $S_M(\vec{S}) = S_M(\vec{S}')$ . In other words, the detailed balance solution (if it exists) satisfies

$$P(\vec{S}) = C_1(S_M(\vec{S})) \prod_{r=1}^{J_M} \frac{W(r)^{S_r}}{r^{S_r} S_r!}, \tag{A.9}$$

where the proportionality constant  $C_1$  depends on the state  $\vec{S}$  only through the number of species  $S_M(\vec{S})$ . This is Eq. (14).

### Appendix B

In this appendix we verify that the distribution (21) satisfies the detailed balance conditions (12) for the transition rates (6a–6c). These rates can be summarised as

$$R_{k,\ell}(\vec{S}) = \begin{cases} \frac{k\ell S_k S_\ell}{J_M^2} (1 - \nu(\ell)) & \text{if } \ell > 0 \text{ and } k \neq \ell, \\ \frac{k^2 S_k (S_k - 1)}{J_M^2} (1 - \nu(k)) & \text{if } \ell > 0 \text{ and } k = \ell, \\ \frac{k S_k}{J_M^2} V(S_M(\vec{S})) & \text{if } \ell = 0, \end{cases} \tag{B.1}$$

for  $k = 1, 2, \dots, J_M$  and  $\ell = 0, 1, \dots, J_M - 1$ .

As detailed balance (12) links a transition to its reverse, it suffices to consider transitions  $R_{k,\ell}$  with  $k > \ell$ , as their reverses include all  $R_{k',\ell'}$  with  $k' \leq \ell'$ ,  $k' = \ell + 1$  and  $\ell' = k - 1$ . Moreover, transitions  $R_{k,\ell}$  with  $k = \ell + 1$  leave the vector  $\vec{S}$  invariant, so that condition (12) is trivially satisfied. For the remaining cases, this condition can be proved as follows:

1. For  $k > \ell + 2$  and  $\ell > 0$ ,

$$\begin{aligned}
 P(\vec{S}_{k,\ell})R_{\ell+1,k-1}(\vec{S}_{k,\ell}) &= P(\vec{S}) \frac{k\ell S_k S_\ell}{(k-1)(\ell+1)(S_{k-1}+1)(S_{\ell+1}+1)} \frac{1-\nu(\ell)}{1-\nu(k-1)} \\
 &\quad \times \frac{(k-1)(\ell+1)(S_{k-1}+1)(S_{\ell+1}+1)}{J_M^2} (1-\nu(k-1)) \\
 &= P(\vec{S}) \frac{k\ell S_k S_\ell}{J_M^2} (1-\nu(\ell)) \\
 &= P(\vec{S})R_{k,\ell}(\vec{S}).
 \end{aligned} \tag{B.2}$$

2. For  $k = \ell + 2$  and  $\ell > 0$ ,

$$\begin{aligned}
 P(\vec{S}_{k,k-2})R_{k-1,k-1}(\vec{S}_{k,k-2}) &= P(\vec{S} + 2\vec{e}_{k-1} - \vec{e}_k - \vec{e}_{k-2}) \\
 &\quad \times R_{k-1,k-1}(\vec{S} + 2\vec{e}_{k-1} - \vec{e}_k - \vec{e}_{k-2}) \\
 &= P(\vec{S}) \frac{k(k-2)S_k S_{k-2}}{(k-1)^2(S_{k-1}+2)(S_{k-1}+1)} \frac{1-\nu(k-2)}{1-\nu(k-1)} \\
 &\quad \times \frac{(k-1)^2(S_{k-1}+2)(S_{k-1}+1)}{J_M^2} (1-\nu(k-1)) \\
 &= P(\vec{S}) \frac{k(k-2)S_k S_{k-2}}{J_M^2} (1-\nu(k-2)) \\
 &= P(\vec{S})R_{k,k-2}(\vec{S}).
 \end{aligned} \tag{B.3}$$

3. For  $k > 2$  and  $\ell = 0$ ,

$$\begin{aligned}
 P(\vec{S}_{1,k-1})R_{k,0}(\vec{S}_{1,k-1}) &= P(\vec{S}) \frac{k S_k}{(k-1)(S_{k-1}+1)(S_1+1)} \frac{V(S_M(\vec{S}))}{1-\nu(k-1)} \\
 &\quad \times \frac{(k-1)(S_{k-1}+1)(S_1+1)}{J_M^2} (1-\nu(k-1)) \\
 &= P(\vec{S}) \frac{k S_k}{J_M^2} V(S_M(\vec{S})) \\
 &= P(\vec{S})R_{k,0}(\vec{S}).
 \end{aligned} \tag{B.4}$$

4. For  $k = 2$  and  $\ell = 0$ ,

$$\begin{aligned}
 P(\vec{S}_{1,1})R_{2,0}(\vec{S}_{1,1}) &= P(\vec{S} - \vec{e}_2 + 2\vec{e}_1)R_{1,1}(\vec{S} - \vec{e}_2 + 2\vec{e}_1) \\
 &= (\vec{S}) \frac{2S_2}{(S_1 + 2)(S_1 + 1)} \frac{V(S_M(\vec{S}))}{1 - \nu(1)} \\
 &\quad \times \frac{(S_1 + 2)(S_1 + 1)}{J_M^2} (1 - \nu(1)) \\
 &= P(\vec{S}) \frac{2S_2}{J_M^2} V(S_M(\vec{S})) \\
 &= P(\vec{S})R_{2,0}(\vec{S}).
 \end{aligned} \tag{B.5}$$

## References

- Bell, G., 2001. Neutral macroecology. *Science* 293, 2413–2418.
- Etienne, R.S., 2005. A new sampling formula for neutral biodiversity. *Ecol. Lett.* 8, 253–260.
- Etienne, R.S., Alonso, D., 2007. Neutral community theory: how stochasticity and dispersal-limitation can explain species coexistence. *J. Stat. Phys.* 128, 485–510.
- Etienne, R.S., Apol, M.E.F., Olff, H., Weissing, F.J., 2007. Modes of speciation and the neutral theory of biodiversity. *Oikos* 116, 241–258.
- Etienne, R.S., Apol, M.E.F., 2009. Estimating speciation and extinction rates from diversity data and the fossil record. *Evolution* 63, 244–255.
- Ewens, W.J., 1972. The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.* 3, 87–112.
- Haegeman, B., Etienne, R.S., 2008. Relaxing the zero-sum assumption in neutral biodiversity theory. *J. Theor. Biol.* 252, 288–294.
- Hubbell, S.P., 2001. *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton.
- Makarieva, A.M., Gorshkov, V.G., 2004. On the dependence of speciation rates on species abundance and characteristic population size. *J. Biosci.* 29, 119–128.
- McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K., Dornelas, M., Enquist, B.J., Green, J., He, F., Hurlbert, A., Magurran, A.E., Marquet, P.A., Maurer, B.A., Ostling, A., Soykan, C.U., Ugland, K., White, K., 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecol. Lett.* 10, 995–1015.
- Moran, P.A.P., 1962. *Statistical Processes of Evolutionary Theory*. Clarendon Press, Oxford.
- Rosindell, J.L., Cornell, S.J., 2007. Species-area relationships from a spatially explicit neutral model in an infinite landscape. *Ecol. Lett.* 10, 586–595.
- Stanley, S.M., 1975. A theory of evolution above the species level. *Proc. Natl. Acad. Sci. USA* 72, 646–650.
- Van Kampen, N.G., 1992. *Stochastic Processes in Physics and Chemistry*. Elsevier, Amsterdam.