ORIGINAL PAPER

The role of sex separation in neutral speciation

Elizabeth M. Baptestini · Marcus A. M. de Aguiar · Yaneer Bar-Yam

Abstract Neutral speciation mechanisms based on isolation by distance and assortative mating, termed topopatric, has recently been shown to describe the observed patterns of abundance distributions and species–area relationships. Previous works have considered this type of process only in the context of hermaphroditic populations. In this work, we extend a hermaphroditic model of topopatric speciation to populations where individuals are explicitly separated into males and females. We show that for a particular carrying capacity, speciation occurs under similar conditions, but the number of species generated is lower than in the hermaphroditic case. As a consequence, the species–area curve has lower exponents, especially at intermediate scales. Evolution results in fewer species having more abundant populations.

Keywords Neutral speciation · Assortative mating · Biodiversity patterns · Sex separation

Introduction

Natural selection is a key process in the adaptation of species to changes in the environment and to changes in other species. Whether it plays an important role in shaping the observed patterns of biodiversity, however, has been questioned. Theories based on drift, migration, and statistical fluctuations in

E. M. Baptestini · M. A. M. de Aguiar (✉)
Instituto de Física Gleb Wataghin, Universidade Estadual de Campinas,
Campinas, São Paulo 13083-970, Brazil
e-mail: aguiar@ifi.unicamp.br

M. A. M. de Aguiar · Y. Bar-Yam
New England Complex Systems Institute,
Cambridge, MA 02142, USA

population size without selection, called *neutral theories*, have been very successful in reproducing the observed abundance distributions, which exhibit remarkable universal features (Hubbell 2001; Kopp 2010; Etienne and Haegeman 2011; Rosindell et al. 2011).

Speciation is the ultimate driver of biodiversity. The neutral theory of biogeography developed initially by Hubbell (2001) and others (Kopp 2010; Ter Steege 2010; O'Dwyer and Green 2010) included speciation as random point mutations, without specifying an underlying mechanism. The assumption of random speciation events may be reasonable for describing island biogeography, where "speciation" represents the arrival of new species from the continent. However, in other situations, the point mutation model is a more radically simplifying assumption. Requirements for multiple individuals in a viable sexually reproducing population and the role of subpopulation divergence should be considered. Hence, the process of speciation merits a discussion of its own in the neutral framework (Banavar and Maritan 2009; Rosindell et al. 2010; Etienne and Haegeman 2011). Recently, explicit speciation mechanisms have been introduced in the context of neutral models (de Aguiar et al. 2009; Rosindell and Phillimore 2011; Desjardins-Proulx and Gravel 2012).

Of the many types of speciation processes, allopatry is considered to be the dominant form (Mayr 1988). Allopatric speciation happens when geographic barriers block the genetic flow between groups of individuals. These isolated groups evolve independently, either by selection or drift, eventually acquiring incompatibilities leading to reproductive isolation. In this context, the evidence for the role of natural and sexual selection in promoting reproductive isolation has been observed both in laboratory experiments and in nature. Neutral divergence, due to drift alone, has also been observed in plants and mammals, although in fewer cases (Coyne and Orr 2004).

Sympatric speciation, on the other hand, is triggered by ecological interactions taking place in a single spatial domain and even in the same niche (Rosenzweig 1997). The key driver is the coupling between ecological and mating

traits, which may lead to disruptive selection and ultimately to speciation (Dieckmann and Doebeli 1999; Leimar et al. 2008; Baptestini et al. 2009; Pinho and Hey 2010).

A neutral theory of speciation relying on isolation by distance (Wright 1943), without geographic barriers or ecological interactions, has been recently demonstrated (de Aguiar et al. 2009; Gavrilets and Cruzan 1998; Gavrilets et al. 2000; Hoelzer et al. 2008). The mechanism, termed topopatry, was shown to describe the universal features observed in abundance distributions and species–area relationships. In this context, assortative mating driven by spatial and genetic distances (also similar to sexual selection) with neutral and independent genes is sufficient to promote speciation (de Aguiar et al. 2009; Melian et al. 2010). The demonstration that speciation can happen *even* in homogeneous environments also suggests that speciation can be accelerated by the presence of partial barriers, selection, and gene interactions. Ring species are interesting examples of such, where geography plays a crucial role in physically shaping the ring but does not block dispersal or gene flow along the ring (Irwin et al. 2001, 2005; Ashlock et al. 2010). In topopatric speciation, isolation by distance plays the role of the physical barrier in allopatry, significantly reducing gene flow between geographically distant individuals. Therefore, the genetic and ecological mechanisms leading to reproductive isolation in allopatric processes may act without the need for an initial period of geographical separation. Allopatry and sympatry can be viewed as the extremes of a continuum of speciation modes (Fitzpatrick et al. 2009). Topopatry, based on isolation by distance, is a new example of an intermediate case.

The topopatric model described in de Aguiar et al. (2009) relies on several limiting assumptions that should be relaxed for comparison with real ecologies. In particular, the model assumed individuals to be haploid and hermaphroditic. In this paper, we consider haploid males and females explicitly and study the effects of sex separation in the process of speciation.

Sex separation in evolution has been shown to either favor or hinder speciation, depending on the mechanisms driving it (Gorelick and Heng 2011; Melian et al. 2012). An important case is that of diploid organisms where Dobzhansky–Muller incompatibilities related to sex chromosomes exist (Orr 1997; Orr and Presgraves 2000; Turelli and Orr 2000; Kondrashov and Kondrashov 2001; Coyne and Orr 2004; Haerty and Singh 2006). According to this perspective, sexual differentiation facilitates speciation through unviable hybrids according to Haldane's rule. On the other hand, the separation of individuals into males and females may give rise to sexual dimorphism, where significant phenotypic differences between the two sexes develop. Recent studies have shown that adaptive speciation and ecological sexual dimorphism may compete as outcomes of assortative mating, reducing the likelihood of speciation (Bolnick and Doebeli 2003; Parker and Partridge 1998).

In this work, we consider the primary effects of sex separation on topopatric speciation. Ecological traits conferring advantages to individuals, direct competition, or epistatic effects between sexual and asexual chromosomes are not considered. Our treatment, although simplified, has the advantage of isolating the effects of sex separation in the process of reproduction. We find that for a particular carrying capacity, speciation occurs under similar conditions in the sex-separated case and in the hermaphroditic case. However, the number of species with sex separation decreases by a factor that ranges between two and four for different parameter-specified conditions. Evolution in this case results in fewer but more populous and stable species. As a consequence, the exponents of the power law describing the species–area curve decrease, especially at intermediate scales.

The paper is organized as follows: in "The model" section, we provide a detailed description of the model and of our working definition of species. In "Theoretical results" section, we discuss some theoretical predictions based on results obtained for hermaphroditic populations, and in "Results of simulations" section, we present the results of numerical simulations. Finally, in "Discussion" section, we present our discussions and conclusions.

The model

We use an agent-based model to simulate the neutral evolution of spatially distributed populations. The current approach differs from that used previously (de Aguiar et al. 2009) in that we distinguish male and female individuals and restrict mating accordingly. In this section, we present a detailed description of the model.

The physical and genetic spaces

We consider an initial population of N haploid and genetically identical individuals randomly distributed over a homogeneous environment, represented by a rectangular geographical domain subdivided into $L \times L$ regions. We use periodic boundary conditions so that there are no boundaries or corners. Multiple individuals can exist at the same site but typically do not. The number of individuals is held fixed throughout the simulation, corresponding to an underlying fixed ecological capacity.

Each individual in the population is located at a position (x, y) in the physical space, with $1 \leq x, y \leq L$ and has a haploid genome, of length $B+1$ with independent biallelic genes, which are labeled 0 or 1, as illustrated in Fig. 1. The kth gene of the ith individual is denoted by σ_k^i and the genome by the bit string:
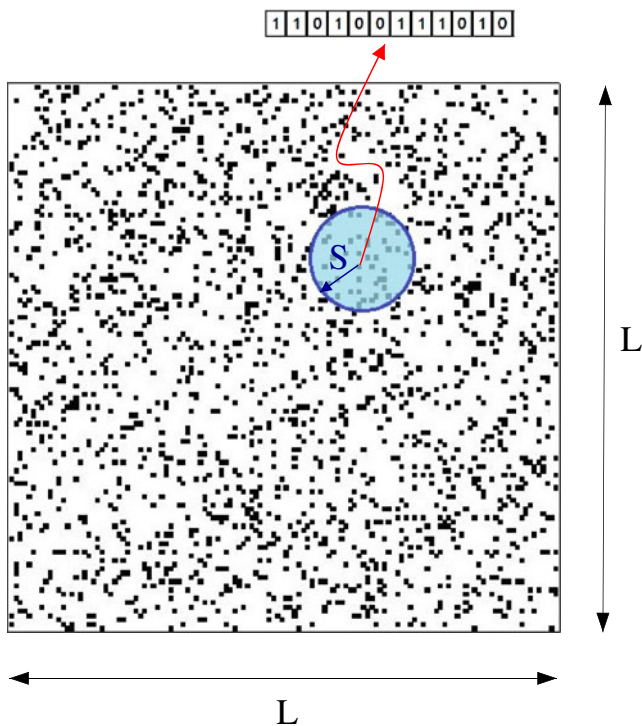
Fig. 1 Snapshot of agent population on a square lattice. Individuals are represented by a *square* at its lattice location. The schematic shows the genome of a female for $B=11$ and its mating neighborhood of radius S

$$g^i = \left(\sigma_1^i, \sigma_2^i, \ldots, \sigma_B^i; \sigma_{B+1}^i \right). \qquad (1)$$

The last position in the genome determines the gender of the individual: $\sigma_{B+1}^i = 1$ for males and $\sigma_{B+1}^i = 0$ for females. At the beginning of the simulation, all individuals have identical genomes with $\sigma_k^i = 0$ for $k=1,\ldots,B$. The value of σ_{B+1}^i is assigned 0 or 1 with equal probability.

The key ingredient of the model is the introduction of assortative mating based on two critical mating distances (de Aguiar et al. 2009): one in physical space and one in genetic space. In physical space, an individual can mate only with others of the opposite sex living in the neighborhood of its location determined by the spatial mating distance S. This type of spatial mating restriction was considered by Wright (1940, 1943) and Kimura and Weiss (1964) and may lead to significant genetic differences between geographically distant individuals of the same species. Striking evidence of this mechanism of "isolation by distance" is provided by ring species (Irwin et al. 2001, 2005). The effect of the spatial constraint on the outcome of speciation depends on the genome size B. In the limit of infinitely, large genomes speciation becomes possible even if the spatial constraint is removed (Higgs and Derrida 1992). Equation 9 below corroborates this result and gives explicit dependence of the threshold for speciation on the model parameters.

We also assume that individuals do not mate with others who have dramatically different genotypes. Among the many reasons for this are structural differences in the sex organs, failure of the sperm to reach or fuse with the egg, and failure of the individual to elicit mating behavior (Coyne and Orr 2004). This restriction can be considered a form of assortative mating, as it has the same effect as behavioral mating preference. Assortative mating is a key ingredient of several of models of sympatric speciation (Dieckmann and Doebeli 2000, Doebeli and Dieckmann 2003; Bolnick and Kirkpatrick 2012), and empirical evidence of its role in speciation has been discussed (Parker and Partridge 1998; Coyne and Orr 2004). Reproductive isolation caused by this mechanism can be considered a multilocus generalization of the Batenson–Dobzhansky–Muller model in which individuals accumulate genetic incompatibilities. As shown previously (de Aguiar et al. 2009), genetic restriction on mating alone does not lead to speciation, but it keeps different existing species genetically isolated from one another. To impose genetic proximity on mating organisms (Gavrilets 2004; Higgs and Derrida 1991), we restrict the number of distinct genes to be no more than the genetic mating distance G. The genetic distance between individuals i and j is measured by the Hamming distance:

$$d(i,j) = \sum_{k=1}^{B} \left| \sigma_k^i - \sigma_k^j \right| \qquad (2)$$

and mating is possible if $d(i,j) \leq G$.

Time evolution

The evolution of each generation is divided into N *time steps*, in which a single individual reproduces. After one generation, or N such time steps, the entire population has been replaced. We start with the ith individual which attempts to reproduce and is successful with a probability $(1-Q)$. We identify all individuals of the opposite sex in its spatial mating neighborhood, specified by the distance S, whose genetic distance is less or equal to G. From this list, a mate individual is selected with at random, say, individual j. There is no mating preference within this list of compatible individuals, including no restriction on mating between relatives.

The genome of the offspring is obtained by a single recombination of g^i and g^j: a random position k in the parent's genomes is chosen to cross over, and two new genomes, g^a and g^b, are produced:

$$
\begin{aligned}
g^a &= \left(\sigma_1^i, \sigma_2^i, \ldots, \sigma_k^i, \sigma_{k+1}^j, \ldots \sigma_B^j; \sigma_{B+1}^j \right) \\
g^b &= \left(\sigma_1^j, \sigma_2^j, \ldots, \sigma_k^j, \sigma_{k+1}^i, \ldots \sigma_B^i; \sigma_{B+1}^i \right)
\end{aligned} \qquad (3)
$$

One of these is taken with equal probability as the offspring's genome, which is further subjected to mutations, at a rate μ per gene.

The offspring is placed at position (x^i, y^i) with probability $(1-D)$, or within a small region of radius r_D around (x^i, y^i) with probability D. D is the dispersal rate and r_D the dispersal range. After reproduction, the originating parent expires, and the label i is assigned to the offspring. In N time steps, or one generation, all individuals are replaced in sequence, from $i=1$ to $i=N$. Note that this does not imply any spatial ordering in reproduction, since the individuals are randomly placed at the beginning. However, the generations are partially overlapping, since a newly born offspring can be chosen as mate partner of another individual during the same "mating season," leading to some degree of inbreeding. This type of reproduction system is common in species with low dispersal, such as certain marine invertebrates (Thornhill 1993). In these cases, and in our model, reproduction only occurs between spatially closed pairs, and the chance of two closely related individuals mating is larger than in panmictic populations.

Reproduction of the ith individual is, however, only successful with probability $1-Q$. With probability Q, the individual dies without leaving a descendant. The parameter Q may be interpreted as attempts to mate with incompatible individuals or failure to find a mate in the mating season. In this case, the availability of local resources enables another individual, chosen at random within the spatial neighborhood of radius S, to reproduce instead of the original individual. The offspring generated is placed in the position of the original individual or in its neighborhood according to D and r_D. On average, two offspring are born for each parent. If $Q=0$, each individual has at least one offspring and also has a probability of being selected by neighbors as a mate for one of their offspring. If $Q \neq 0$, some individuals have no offspring, and the distribution of numbers of offspring includes those with additional offspring to offset them.

During the selection process restricted by spatial and genetic proximity, it is possible for the number of mates available to the reproducing individual to be very small, possibly zero, preventing it from finding a mate. To avoid this situation, we introduce the parameter P, representing the minimum number of *potential mates*. Given S and G, if the number of mates available to the individual is smaller than P, we relax the spatial constraint by increasing $S \rightarrow S+1$ for the present mating season only, i.e., the individual increases the search area in order to have more choices. If the number of available mates is still smaller than P, the process is repeated until S increases up to 10 units, with no cost to the individual. If the number of mates is still smaller than P, the organism does not reproduce, and a neighbor is picked at random to reproduce in its place. In the paper by de Aguiar et al. (2009), the genetic constraint G was relaxed in

addition to S, i.e., $S \rightarrow S+1$ and $G \rightarrow G+1$. Here, however, we let only S change and keep the genetic restriction fixed at all times. The parameter P avoids the frequent appearance and extinction of species with small number of individuals. Cases of increasing S occur for individuals near the boundary with a different species, where the local number of conspecifics is smaller. The algorithm describing the time evolution of the population is presented as flowcharts in the Electronic Supplementary Material.

Species

Many definitions of species have been proposed that work well for specific groups of organisms but fail or are impractical for others. The most commonly used of these definitions is perhaps Ernst Mayr's Biological Species Concept (BSC) (Mayr 1955), based on the interbreeding ability of the individuals in a group. Another concept is that of genetic cohesion devised by Mallet (1995), termed Genotypic Cluster Species Concept (GCSC), according to which, a species is a genetically distinguishable group of organisms that has no (or few) intermediates when in contact with other such groups. A similar definition is the Cohesion Species Concept developed by Templeton (1989).

For our purposes, a species is defined as a group of individuals related by potential gene flow, which need not be possible in a single generation. Therefore, two individuals in the population can be conspecific while also being incompatible, as long as they can exchange genes indirectly through other conspecifics. As an example, consider three individuals A, B, and C such that $d(A,B)<G$ and $d(B,C)<G$ but $d(A,C)>G$. A mutation occurring in A can be transmitted to the offspring of A and B that can, in turn, pass the mutation on when mating with C or its offspring. This situation is common in ring species (Irwin et al. 2001, 2005), and we find it also occurs in our simulations. In the case of a ring species, the appearance of an advantageous mutation on a few individuals might spread over entire ring, due to its genetic cohesion. This, however, might take multiple generations. According to the BSC, A and C are of different species. However, A, B, and C all belong to the same species according to the GCSC, since in genetic space, the individuals form a cluster that is cohesive and is separated by more than G from all organisms not in the cluster. Thus, our definition is similar to, if not exactly the same as, GCSC.

In order to classify the individuals in the population into species, the following algorithm is applied: we start with individual number 1 (which is arbitrary) and assign it to the first species, Species-1. We collect all others such that $d(1,i) \leq G$ and assign them to Species-1. For each of the individuals i just added to Species-1, we check if $d(j,i) \leq G$ for all unassigned individuals. The individuals satisfying this

condition are also assigned to Species-1. For these new individuals j, we check again if $d(k,j) \leq G$ for all unassigned k. The individuals satisfying this condition are also added to Species-1 and so on. When no more individuals are added, Species-1 is completed. It is a cohesive group and genetically isolated from the unassigned individuals. If there are no unassigned individuals, there is only one species. Otherwise, we take one unassigned individual and assign it to the second species, Species-2, repeating the process. It is straightforward to prove that the species obtained in this way are independent of which individuals are chosen. Note that the only criterion used to define species is the genetic mating distance G. No information about the spatial location of the individuals is taken into account. In terms of network theory, we say that individuals i and j are connected if $d(j, i) \leq G$. A species is a path-connected set, where any two individuals in the set are connected by a path.

Theoretical results

A population whose individuals are genetically identical at time zero develops differences through mutation, which occurs at the rate μ, and recombination. These differences, however, are constrained by sexual reproduction, which tends to contract the genetic spreading caused by mutations. The balance between these two opposing forces results in the natural diversity of the population. When spatial and genetic selection are present, the population may spontaneously break up into multiple species, depending on the values of the many parameters of the model. The coupling between genetic and physical spaces during reproduction gives rise to a process of pattern formation that leads to speciation. Similar results have been observed before in simpler systems (Sayama et al. 2002) and also in speciation models (Gavrilets 2004; Hoelzer et al. 2008).

The number of species formed for a given set of parameters can be estimated by Eq. 12 below. The reasoning leading to this equation is as follows: for a panmictic population in equilibrium with N hermaphroditic individuals, the probability that two individuals picked at random have different alleles for a biallelic gene is

$$P_{ht} = \frac{2\mu N}{1 + 4\mu N} \tag{4}$$

If individuals have B independent genes, the average genetic distance between two individuals is therefore

$$\langle d \rangle = \frac{B}{2}\left(\frac{4\mu N}{1 + 4\mu N}\right) \tag{5}$$

Notice that the numerical model includes recombination with a single crossover, whereas the assumption of

independent genes is equivalent to multiple crossovers. This is a good approximation if the parent's genomes are similar, which is indeed the case.

The effect of the spatial restriction S can be shown to be equivalent to changing the mutation rate to an effective value (de Aguiar and Bar-Yam 2011):

$$\mu_{eff} = \frac{\mu f}{1 + 2\mu f} \tag{6}$$

where $f = L^2/\pi S^2$ is the ratio between the total area and the area available for reproduction.

On the other hand, when mating in panmictic populations is constrained by genetic proximity between individuals, so that pairs whose genetic distance is larger than G are incompatible, the distribution of genetic distances stays very close to $\langle d \rangle = G$, as if the genome had an effective size $B_{eff} = 2G$. However, as S is reduced, the effective mutation rate increases and so does the genetic distance between individuals, as if additional genes were incorporated into the effective genome. When $\langle d \rangle$ becomes larger than about $2G$, the population can no longer hold itself together and splits (de Aguiar and Bar-Yam 2011). The size of this effective genome can be written as:

$$B_{eff} = 2G + (B - 2G)P_r \tag{7}$$

where P_r accounts for the fraction of genes added to the effective genome by the effect of S. Based on numerical simulations, it has been shown that P_r can be approximated by:

$$P_r = \exp\left(-\frac{\pi^2 \rho_0^2 (S - S_{\min})^4}{B^2 \mu^2 \gamma^4}\right), \tag{8}$$

where $\rho_0 = N/L^2$ is the average population density, $S_{\min} = \sqrt{P/\pi\rho_0}$ is the size of a neighborhood containing P individuals, and γ is a parameter obtained by fitting to simulations.

Replacing μ and B in Eq. 5 by μ_{eff} and B_{eff}, Eqs. 6–8, and imposing the condition $\langle d \rangle = 2G$ for the onset of speciation, we obtain an expression for the critical line in the G versus S plane below which speciation occurs. This expression can be simplified if $4\mu_{eff} N \geq 1$, which is verified in most of our simulations. We obtain

$$G_c(S) = \frac{B/2}{1 + P_r^{-1}} . \tag{9}$$

Solving this equation for S, we can also write:

$$S_c = S_{\min} + \gamma L \sqrt{\frac{B\mu}{\pi N}}\left[\log\left(\frac{B - 2G}{2G}\right)\right]^{1/4} \tag{10}$$

Particularly important for the present discussion is the calculation of N_s, the number of species that arise from

speciation. It can be written as $N_s=N/N_i$ where N_i is the average number of individuals in a species. In order to estimate N_i, we note that the instability provoked by S and G rearranges the population into nearly panmictic species with N_i individuals, $<d>=G$ and $B=B_{eff}$. Using Eq. 5, we obtain

$$N_i = \frac{G}{2\mu\left(B_{eff} - 2G\right)} \tag{11}$$

and

$$N_s = \frac{2\mu N}{G}\left(B_{eff} - 2G\right) = \frac{2\mu\rho_0 L^2}{G}(B - 2G)P_r \tag{12}$$

We can further write:

$$N_i = \pi\rho_0 R^2$$

where R is a measure of the species' spatial extent. Assuming that $P_r \approx 1$, the exponential in Eq. 6 can be expanded to first order in its argument resulting in

$$R \simeq \sqrt{\frac{G}{2\pi\mu\rho_0(B - 2G)}}\left[1 + \frac{\pi^2\rho_o^2(S - S_{min})^4}{B^2\mu^2\gamma^4}\right] \tag{13}$$

for the average radius of a species. A similar expression was obtained by de Aguiar et al. (2009) using simulation results.

Equations 9, 12, and 13 fit the numerical data well in the hermaphroditic case. Moreover, Eq. 10 also shows that if $B\mu/N \sim 1$, i.e., if the number of mutations per generation is comparable to the number of individuals in the population, speciation may happen without the spatial constraint because in a single generation, organisms diverge genetically from all other organisms in the population. In this case, we recover the possibility of speciation for infinite length genomes in a panmictic population, which has been previously considered (Higgs and Derrida 1991, 1992).

In the model with sex separation, these estimates also work reasonably well if the equations are properly adapted. In the hermaphroditic case, the population can be seen as a network where each individual is a node and links are established between potential mates (any two individuals who are spatially close and genetically compatible). In the case of sex separation, the nature of the network changes considerably: the nodes represented by females do not link among themselves, but only with nodes representing the males, and vice versa. Networks made up of two disjoint sets of nodes, such that nodes in one set connect only to nodes in the other set are called *bipartite* networks. It is useful to define a female network, where two individuals are connected if they can mate with a common male, and similarly for a male network. For each of these networks, assuming the sex ratio is approximately 1, the average

density is $\rho_0/2$. Also, if the spatial restriction between males and females is S, the maximum separation between individuals of the same sex in a species is $2S$. Moreover, gene flow in the female network is at least a two-step process, since it takes the mating of a first female to a male and, if the offspring is male, it is mating with a second female to combine their genes. Therefore, the time scale of crossover events slows down by a factor of two. Still, the analytic discussion continues to apply to the female or male network so the form of the analytic expressions is valid. We find from simulations described below that Eqs. 9, 12, and 13 can indeed be applied to the sexual model if we change ρ_0 to $\rho_0/2$, S_{min} to $\sqrt{2}S_{min}$, and S to $2S$.

Results of simulations

The purpose of the present numerical simulations is to study the patterns of abundance resulting from the explicit introduction of males and females and compare them with those obtained with the hermaphroditic model. We refer to the former as the *sex-separated* model.

Since the number of model parameters is quite large, we keep many of them constant throughout the simulations. Variations in these parameters alter the results according to Eqs. 8 and 9, which determine the range of parameters where speciation occurs. For purposes of comparison with the hermaphroditic model (de Aguiar et al. 2009), we fixed the parameter values: mutation rate $\mu=0.001$; length of genome $B=125$; diffusion rate $D=0$; minimum number of potential partners $P=5$; and probability of no reproduction $Q=0.3$. In most cases, we will also use $S=5$, $G=20$, $N=8000$, and $L=256$. In all figures, time is measured in number of generations. Because there is no difference in the fitness assigned to males and females, the sex ratio is nearly constant across generations, fluctuating around $N/2$. Figures S1 and S2 in the Electronic Supplementary Material show examples of simulations with different values of B and diffusions rates, respectively.

Figure 2a shows the number of individuals between r and $r+1$ as measured from the geographic center of a species and averaged over all species (squares) compared with the fit obtained by

$$n(r) = r\exp\left(-\frac{r^2}{R^2}\right) .$$

Figure 2(b) shows the genetic distance between individuals of a species as measured from a reference individual: the one situated closest to the geographic center of the species. This shows clearly the strong correlation between spatial distance and genetic distance. It also shows that the central individual can mate with any
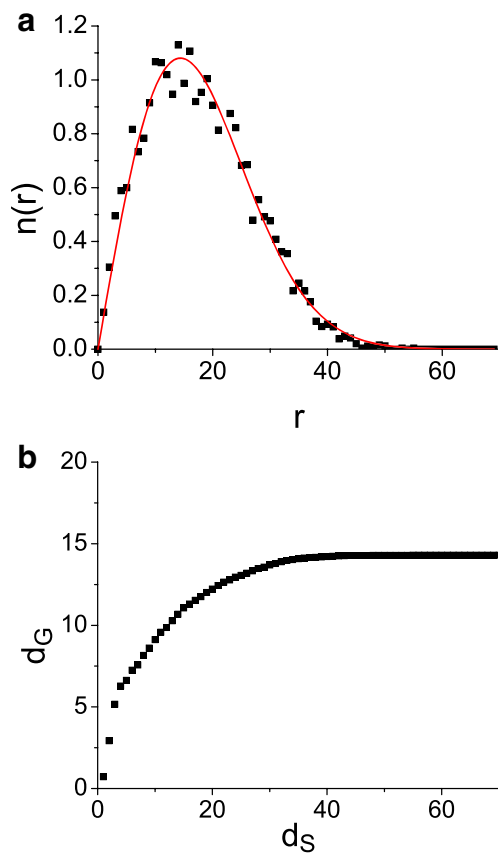
Fig. 2 **a** Spatial distribution of individuals, $n(r)$, for the sex-separated model averaged over all species. The simulation is shown as *black squares* and the Gaussian fit by the *solid line* ($R^2 = 0.989$). **b** Correlation between genetic (d_G) and spatial (d_S) distances between individuals within a species averaged over all species. In both cases, $N=8,000$, $L=256$, $G=20$, and $S=5$

member of the population as far as the genetic constraint is concerned (i.e., the average asymptotic genetic distance is 14.3 which is less than $G=20$), just as is necessary for the species definition according to BSC. In many cases, however, individuals at opposite spatial extremes of a species may have genetic distance larger than G and would not be able to mate even if brought spatially close to each other, a feature also observed in the hermaphroditic model.

One of the strengths of the topopatric speciation model is its ability to replicate observed distributions of abundance. Typical abundance distributions are well fit by the lognormal function with excess rare species (May 1975; Sugihara 1980). In Fig. 3a we compare the results of simulations (black squares) with a lognormal curve (solid line; sampled area equal to the total area of the lattice, 512×512). Figure 3b shows the distribution for sampling area corresponding to only one eighth of the total area available (128×128), displaying a clear excess of rare species as compared to the lognormal distribution. The distribution obtained for even smaller sample

areas converge to a Fisher curve (Fig. 3c; 64×64). Figure 3d shows the abundance–rank plot corresponding to panels (a)–(c), displaying the typical S-shaped curve for large sampling areas. All results were normalized.

Figure 4 shows a comparison of the species area relationship (SAR) (Preston 1960; May 1975) for the hermaphroditic model, (a), and the model with sex separation, (b). Both display a triphasic pattern (Rosenzweig 1995; Tjorve 2003) of the form A^z (Arrhenius 1921), with larger exponents at both the smallest and largest area regimes. We find that the exponents in the sex-separated populations are smaller than in the corresponding hermaphroditic cases in all case studies, particularly at the intermediate scale, where the difference in Fig. 4 is about 20 %. Figure S3 of the Electronic Supplementary Material shows another comparison.
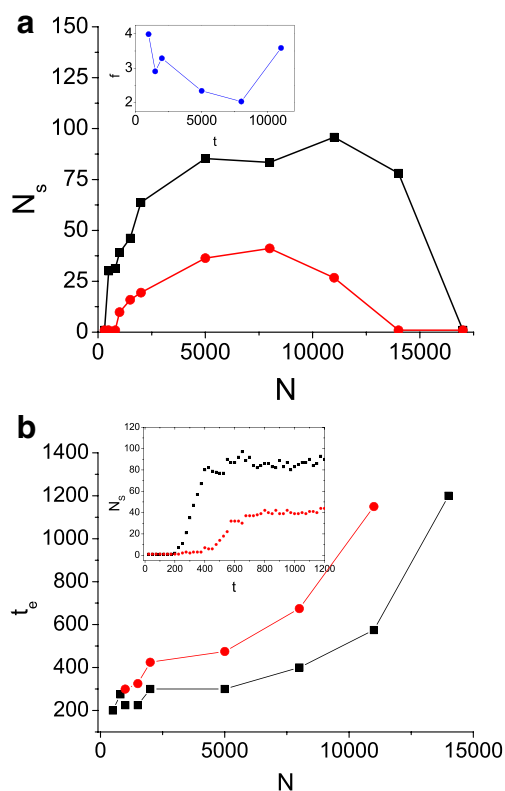
Figure 5a shows the number of species generated after 1,500 generations (when equilibrium has already been reached) as a function of the total number of individuals in the population. The higher curve (black squares) represents the hermaphroditic model, whereas the lower curve (red circles) displays the result of the sex-separated model. In both cases, speciation is not possible at very low densities, since no (or very few) mating partners can be found in the search area delimited by S. More importantly, speciation is also inhibited at high densities, since fluctuations in the population-averaged genotype across geographical location are suppressed. These fluctuations play an important role in the speciation process. The separation of individuals into males and females reduces the number of species formed and also prevents speciation for smaller populations. Figure 5b shows the time elapsed until equilibrium is reached as a function of the number of individuals in the population. The inset shows how the number of species changes with time, showing the equilibration process. For the sex-separated case, equilibration takes about twice as long, and the final number of species is halved, as expected from the theoretical arguments. Figure S4a and b of the Electronic Supplementary Material shows similar plots for $S=6$.

This effect can also be seen in Fig. 6a and b, which shows the number of species formed in terms of the parameters S and G responsible for the assortative mating. Speciation occurs for both models at about the same range of S and G for large S and G, although fewer species are formed in the sex-separated case. The thin line shows the critical speciation curve according to the prediction of the hermaphroditic model (Eq. 4). The solid thick curve shows the same theoretical prediction with the changes of ρ_0 to $\rho_0/2$ and S to $2S$. The shape of the level curves is similar, but in the sex-separated case, speciation is more severely hindered at low values of G—an effect already noted in de Aguiar et al. (2009) in the hermaphroditic model but which is not as evident for the present values of parameters.

Fig. 3 Species abundance distribution for 32,000 individuals, for $S=5$ and $G=20$. Simulations are shown as *black squares* and fits as *solid lines*: (**a**) for 512×512 lattice (lognormal fit, $R^2=0.99$), (**b**) for 128×128 sub-lattices (lognormal fit, $R^2=0.86$), and (**c**) for 64×64 sub-lattices (Fisher distribution, $R^2=0.85$). **d** The abundance versus species rank for the same cases. Data was generated running ten simulations for 1,500 generations and then normalized. Lognormal and Fisher formulas and parameters are given in the Electronic Supplementary Material
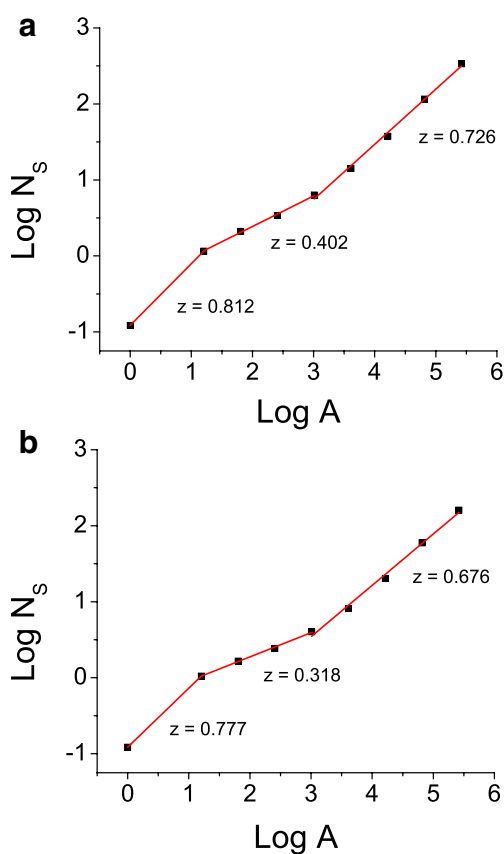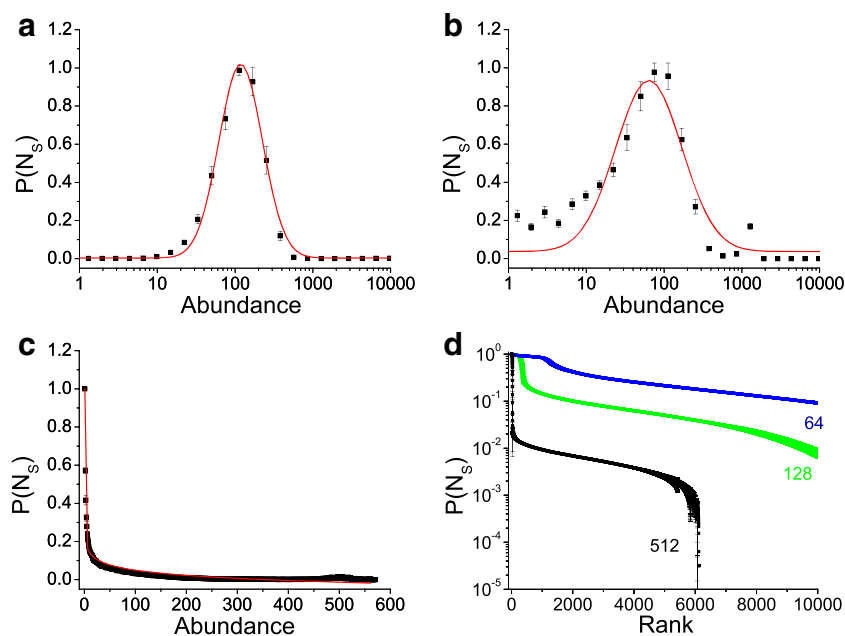
Fig. 4 The classical triphasic species area curve (SAR) for (**a**) hermaphroditic model and (**b**) the sex-separated model. Parameters of the simulation are $S=5$, $G=20$, $N=32,000$, and $t=1,500$ for a 512×512 lattice

Fig. 5 a Number of species formed as function of the number of individuals in the population for the hermaphroditic (*black squares*) and sex-separated (*red circles*) models. *Inset* shows the ratio f between the number of species of the hermaphroditic and the sex-separated models as function of the number of individuals in the population. **b** Time to equilibration as function of the number of individuals in the population. *Inset* shows time evolution of the number of species in the population for $N=8,000$. Parameters: $S=5$, $G=20$, $L=256$, $t=1,500$

Discussion

Speciation can be triggered by several processes, including geographic isolation, competition for resources, and genetic drift, among others. If the genes involved in speciation do not affect the fitness of the individuals, the speciation is termed "neutral." The idea of neutral evolution, where the role of natural selection is secondary, has been challenged by many. Hubbell (2001), however, demonstrated that realistic patterns of abundance distribution can be obtained within a neutral theory of biogeography in which species originate randomly (Banavar and Maritan 2009; Kopp 2010; Ter Steege 2010; Etienne and Haegeman 2011; Rosindell et al. 2011).

Numerical simulations with hermaphroditic populations (de Aguiar et al. 2009) have shown that similar patterns of diversity emerge in explicitly genetic neutral models if reproduction is constrained by spatial and genetic proximity between individuals. Quantitative agreement between observed and simulated diversity was also obtained using this model. Many of the results observed in these numerical simulations were recently derived analytically by mapping the genetic evolution to an influence dynamical process on networks (de Aguiar and Bar-Yam 2011). In this paper, we extended the hermaphroditic neutral model to describe speciation in populations with explicit sex separation.

The distinction between males and females changes the genetic flow in a population considerably. Unlike hermaphroditic species, in which gene flow is allowed between any two members, here the individuals are divided into two separate groups with no direct gene flow within the groups, only between them. The mathematical description of this process, even for a panmictic population, is very different from the hermaphroditic case. Equations similar to the Moran model (Moran 1958; Cannings 1974; Ewens 1979; Gillespie 2004) can be written down, but explicit solutions are not available, except for zero mutation. In this trivial case, it is possible to show that the equilibrium population is composed of identical individuals, and the number of males and females follow a binomial distribution. Here, we find that the analytic solutions obtained for the hermaphroditic model also approximately apply to the sex-separated case if some re-scaling of the parameters is done.

For simulations, the extension of the model from hermaphroditic to sex-separated individuals is constructed by adding an extra "gene" that specifies the sex and by restricting mating to individuals of the opposite sex. The main difference between the two models is that sexually separated individuals have, on average, half the number of potential mates than in a hermaphroditic group with the same density of individuals. This might suggest that gene flow is more restricted when the two sexes are considered explicitly and

that speciation should occur more easily. This, however, is not the case. The reason is that a population consisting of half males and half females is very different from two independent hermaphroditic populations with half the density. Sexual reproduction has a strong effect on the genetic proximity of offspring and is capable of keeping the population united.

We have shown that all basic features of the hermaphroditic model are preserved in the modified version, with few but important changes. Besides a small reduction in the parametric region where speciation occurs (Fig. 6), the most striking feature of the sex-separated model is the decrease in the number of species formed or, conversely, the increase in the average abundance of individuals per species, as shown in Fig. 5. This is an important characterization of the model, which implies smaller extinction and re-speciation rates and, therefore, more stable species.

The results of our simulations cannot be directly applied to treat Dobzhansky–Muller-type genetic incompatibilities
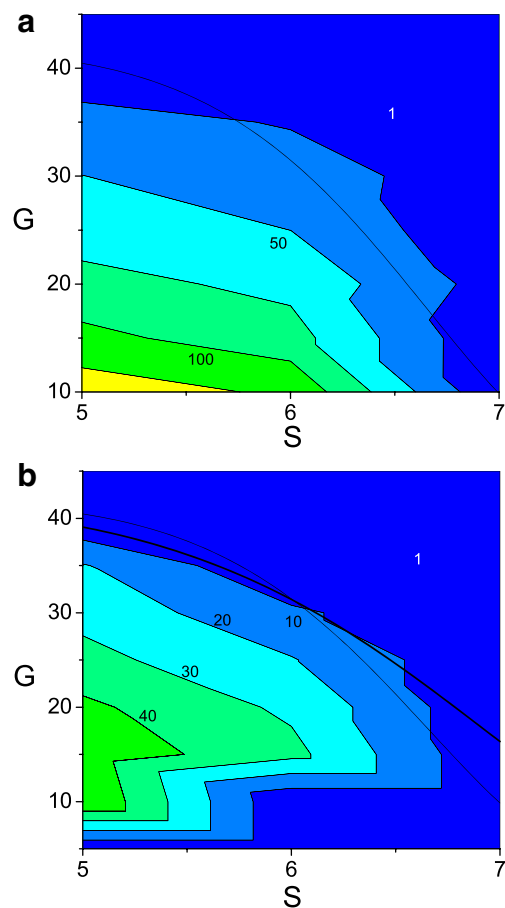
Fig. 6 Contour plot of the number of species formed as function of S and G for (**a**) the hermaphroditic and (**b**) the sex-separated models. Parameters: $\gamma = 5$, $N = 8{,}000$, $L = 256$, and $t = 4{,}000$. The *thin solid line* shows the critical curve according to Eq. 4 with $\gamma = 5$ and the thick line in **b** shows the same curve with $\rho_0 \rightarrow \rho_0/2$, $S \rightarrow 2S$ and $\gamma = 6$

or the development of sexual dimorphism, since the sex chromosomes are not considered explicitly in the model. However, extensions in this direction are possible.

References

Arrhenius O (1921) Specie and area. J Ecol 9:95–99

Ashlock D, Clare EL, von Königslöw TE, Ashlock W (2010) Evolution and instability in ring species complexes: an in silico approach to the study of speciation. JTB 264:1202–1213

Banavar J, Maritan A (2009) Towards a theory of biodiversity. Nature 460:334–335

Baptestini EM, de Aguiar MAM, Araujo MS, Bolnick D (2009) The shape of the competition and carrying capacity kernels affects the likelihood of disruptive selection. JTB 259:5–11

Bolnick D, Doebeli M (2003) Sexual dimorphism and adaptive speciation: two sides of the same ecological coin. Evolution 57:2433

Bolnick DI, Kirkpatrick M (2012) The relationship between intraspecific assortative mating and reproductive isolation between divergent populations. Current Zoology 58(3):484–492

Cannings C (1974) The latent roots of certain Markov chains arising in genetics; a new approach, haploid models. Adv Appl Prob 6:260

Coyne JA, Orr HA (2004) Speciation. Sinauer Associates, Sunderland

de Aguiar MAM, Bar-Yam Y (2011) The Moran model as a dynamical process on networks and its implications for neutral speciation. Phys Rev E 84:031901

de Aguiar MAM, Baranger M, Baptestini EM, Kaufman L, Bar-Yam Y (2009) Global patterns of speciation and diversity. Nature 460:384–387

Desjardins-Proulx P, Gravel D (2012) A complex speciation-richness relationship in a simple neutral model. arXiv:1203.3884v1 [q-bio.PE]

Dieckmann U, Doebeli M (1999) On the origin of species by sympatric speciation. Nature 400:354–357

Dieckmann U, Doebeli M (2000) Evolutionary branching and sympatric speciation caused by different types of ecological interactions. Am Nat 156:S77–S101

Doebeli M, Dieckmann U (2003) Speciation along environmental gradients. Nature 421:259–264

Etienne RS, Haegeman B (2011) The neutral theory of biodiversity with random fission speciation. Theor Ecol 4:87

Ewens WJ (1979) Mathematical population genetics I. Theoretical introduction series: biomathematics, Vol. 9. Springer, New York

Fitzpatrick BM, Fordyce JA, Gavrilets S (2009) Pattern, process and geographic modes of speciation. J Evol Biol 22:2342

Gavrilets S (2004) Fitness landscapes and the origin of species. Princeton University Press, Princeton

Gavrilets S, Cruzan MB (1998) Neutral gene flow across single locus cline. Evolution 52(5):1277–1284

Gavrilets S, Arnqvist G, Friberg U (2000) The evolution of female mate choice by sexual conflict. Proc R Soc Lond B 268:531–539. doi:10.1098/rspb.2000.1382

Gillespie JH (2004) Population genetics: a concise guide. The Johns Hopkins University Press, Baltimore

Gorelick R, Heng HHQ (2011) Sex reduces genetic variation: a multidisciplinary review. Evolution 65:1088–1098

Haerty W, Singh RS (2006) Gene regulation divergence is a major contributor to the evolution of Dobzhansky–Muller incompatibilities between species of *Drosophila*. Mol Biol Evol 23(9):1707–1714

Higgs P, Derrida B (1991) Stochastic models for species formation in evolving populations. J Phys A 24:L985–L991

Higgs P, Derrida B (1992) Genetic distance and species formation in evolving populations. J Mol Evol 35:454–465

Hoelzer GA, Drewes R, Meier J, Doursat R (2008) Isolation-by-distance and outbreeding depression are sufficient to drive parapatric speciation in the absence of environmental influences. PLoS Comput Biol 4:e1000126

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

Irwin DE, Bensch S, Price TD (2001) Speciation in a ring. Nature 409:333–337

Irwin DE, Bensch S, Irwin JH, Price TD (2005) Speciation by distance in a ring species. Science 307:414

Kimura M, Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics 49:561–576

Kondrashov FA, Kondrashov AS (2001) Multidimensional epistasis and the disadvantage of sex. Proc Natl Acad Sci USA 98:12089–12092

Kopp M (2010) Speciation and the neutral theory of biodiversity. BioEssays 32:564–570

Leimar O, Doebeli M, Dieckmann U (2008) Evolution of phenotypic clusters through competition and local adaptation along an environmental gradient. Evolution 62(4):807

Mallet J (1995) A species definition for the modern synthesis. Trends Eco Evol 10:294

May R (1975) In: Cody M, Diamond Belknap J (eds) Ecology and evolution of communities. Harvard University Press, Cambridge

Mayr E (1955) In: Arai R, Kato M, Doi Y (eds) Biodiversity and evolution. National Science Museum Foundation, Tokyo

Mayr E (1988) Toward a new philosophy of biology. Harvard University Press, Cambridge

Melian CJ, Alonso D, Vázquez DP, Regetz J, Allesina S (2010) Frequency-dependent selection predicts patterns of radiations and biodiversity. PLoS Comp Biol 6:e1000892

Melian CJ, Alonso D, Allesina S, Condit RS, Etienne RS (2012) Does sex speed up evolutionary rate and increase biodiversity? PLoS Comput Biol 8(3):e1002414

Moran PAP (1958) Random processes in genetics. Proc Cam Phil Soc 54:60

O'Dwyer JP, Green JL (2010) Field theory for biogeography: a spatially explicit model for predicting patterns of biodiversity. Ecol Lett 13:87–95

Orr HA (1997) Haldane's rule. Annu Rev Ecol Syst 28:195–218

Orr HA, Presgraves DC (2000) Speciation by postzygotic isolation: forces, genes and molecules. BioEssays 22:1085–1094

Parker GA, Partridge L (1998) Sexual confict and speciation. Phil Trans R Soc Lond B 353:261–274

Pinho C, Hey J (2010) Divergence with gene flow: models and data. Annu Rev Ecol Evol Syst 41:215

Preston FW (1960) Time and space and the variation of species. Ecology 42:611

Rosenzweig M (1995) Species diversity in space and time. Cambridge University Press, Cambridge

Rosenzweig ML (1997) Tempo and mode of speciation. Science 277:1622–1623

Rosindell J, Phillimore AB (2011) A unified model of island biogeography sheds light on the zone of radiation. Ecol Lett 14:552–560. doi:10.1111/j.1461-0248.2011.01617.x

Rosindell J, Cornell SJ, Hubbell SP, Etienne RS (2010) Protracted speciation revitalizes the neutral theory of biodiversity. Ecol Lett 13:716–727

Rosindell J, Hubbell SP, Etienne RS (2011) The unified neutral theory of biodiversity and biogeography at age ten. Trends Ecol Evol 26:340–348. doi:10.1016/j.tree.2011.03.024

Sayama H, de Aguiar MAM, Bar-Yam Y, Baranger M (2002) Spontaneous pattern formulation and genetic invasion in locally mating and competing populations. Phys Rev E 65:051919

Sugihara G (1980) Minimal community structure: an explanation of species abundance patterns. Am Nat 116:770–787

Templeton A (1989) In: Otte D, Endler J (eds) Speciation and its consequences. Sinauer Associates, Sunderland, pp 3–27

Ter Steege H (2010) How neutral is ecology? Biotropica 42:631

Thornhill NW (1993) The natural history of inbreeding and outbreeding: theoretical and empirical perspectives. The University of Chicago Press, Chicago

Tjorve E (2003) Shapes and functions of specie-area curves: a review of possible models. J Biogeogr 30:827–835

Turelli M, Orr HA (2000) Dominance, epistasis and the genetics of postzygotic isolation. Genetics 154:1663–1679

Wright SJ (1940) Breeding structure of populations in relation to speciation. Am Nat 74:232–248

Wright SJ (1943) Isolation by distance. Genetics 28:114–138