

LETTERS

Global patterns of speciation and diversity

M. A. M. de Aguiar^{1,2}, M. Baranger^{1,3}, E. M. Baptistini², L. Kaufman^{1,4} & Y. Bar-Yam¹

In recent years, strikingly consistent patterns of biodiversity have been identified over space, time, organism type and geographical region^{1,2}. A neutral theory (assuming no environmental selection or organismal interactions) has been shown to predict many patterns of ecological biodiversity^{2,3}. This theory is based on a mechanism by which new species arise similarly to point mutations in a population without sexual reproduction. Here we report the simulation of populations with sexual reproduction, mutation and dispersal. We found simulated time dependence of speciation rates, species–area relationships and species abundance distributions consistent with the behaviours found in nature^{1–13}. From our results, we predict steady speciation rates, more species in one-dimensional environments than two-dimensional environments, three scaling regimes of species–area relationships and lognormal distributions of species abundance with an excess of rare species and a tail that may be approximated by Fisher's logarithmic series. These are consistent with dependences reported for, among others, global birds⁴ and flowering plants⁵, marine invertebrate fossils⁶, ray-finned fishes⁷, British birds^{8,9} and moths¹⁰, North American songbirds¹¹, mammal fossils from Kansas¹² and Panamanian shrubs¹³. Quantitative comparisons of specific cases are remarkably successful. Our biodiversity results provide additional evidence that species diversity arises without specific physical barriers^{6,11,14}. This is similar to heavy traffic flows, where traffic jams can form even without accidents or barriers¹⁵.

Speciation studies have identified conditions under which speciation events can occur. Allopatry is considered the dominant form of speciation. Studies of partial barriers (parapatric speciation) and debates about sympatric speciation focus on whether such types of speciation are possible and, if so, whether they are plausible in nature^{16,17}. Simulations of homogeneous spatial environments have suggested the necessity of trait-based variation of sexual selection, competition or diffusion rate, or use outbreeding depression^{18–20}. Studies exist indicating these may not be necessary even in sympatry, at least where there is no genetic linkage^{21,22}. Significantly for the results we report here, no simulations of speciation that include sexual mating have yet been shown to yield the observed patterns of species diversity.

We simulated the evolution of a population whose members, at the beginning, are uniformly distributed in space and have identical genomes. The population evolves under the combined influences of sexual reproduction, mutations and dispersal. During reproduction, potential mates are identified from among those in a spatial region around an individual (specified by a spatial mating distance, S) whose genomes are sufficiently similar to that of the individual (specified by a genetic mating distance, G). This is a minimal form of sexual selection, essential (necessary but not sufficient) for speciation, called assortative mating²³ (postzygotic genetic incompatibilities may have a role but are not essential²⁴). A mate is chosen from this set at random. Reproduction with crossover and mutation occurs. An offspring is then dispersed within a region around the originating and expiring

parent. Genetic variation grows over time, due to mutation and recombination. We identify a species as a group of organisms reproductively separated from all others by the genetic restriction on mating and connected among themselves by the same condition, without requiring all members of the group to be able to mate with each other. An example of such a case is a ring species¹⁴, in which progressive differences along a chain of individuals result in individuals at one end not being able to mate with those at the other end. We consider this a single species, owing to the ongoing gene flow. No condition on spatial proximity is imposed on the members of a species.

We simulated many variants of the model; the one we report here uses haploid and hermaphroditic individuals placed randomly in the space (for convenience annotated as a 128×128 or 256×256 lattice). Multiple individuals can exist at the same site but typically do not, owing to low densities. Unlike parapatric simulations that are otherwise similar^{24–26}, our density of organisms is less than one per site (there are no a-priori-defined groups whose genetic divergence is a possible mechanism of speciation), so our speciation may be termed 'topopatric'. Genomes consist of binary strings of length 125. Variation in the number of offspring of an individual was included by randomly choosing a neighbour to reproduce, instead of the expiring parent, in $Q = 30\%$ of cases. Other assumptions, including variations of parameters or more basic changes such as separating two sexes, give qualitatively similar results (Methods).

Depending on the model parameters, we find sexual isolation of subpopulations in genome space (speciation) and spatial inhomogeneity in spite of the absence of geographical barriers, resource gradients or natural selection. We show an example in Fig. 1. The first branching into two species occurs after about three hundred generations; additional speciation events take place later. Each of the species is coloured differently. Species extinctions also occur, owing to random variation in population size and resulting in a variable number of species over time. Although species are not homogeneously distributed and tend to be concentrated in different regions, there is considerable overlap among their domains. Newly branched species are genetically close to their parent species, whereas species formed in the far past reach the maximum average genetic distance in the genome space. The species that form are completely separated by genetic difference according to the mating condition. The spatial and genetic clustering occurs even though the underlying mating behaviour couples organisms in both physical and genetic space.

In Fig. 2, we show the effect on the pattern of speciation of varying the spatial, S , and genetic, G , mating distances. Speciation becomes less likely as the critical mating distances increase. Neither spatial nor genetic restrictions on mating alone leads to speciation. The former is required for the distinction of spatial location, and the latter is required for genetic isolation. The largest number of species arises for small values of both parameters. However, for intermediate values of spatial distance (for example $S = 6$ in Fig. 2), intermediate values of genetic distance yield the most species, as small values of the

¹New England Complex Systems Institute, Cambridge, Massachusetts 02138, USA. ²Universidade Estadual de Campinas, Unicamp, Campinas, São Paulo 13083-970, Brazil.

³University of Arizona, Tucson, Arizona 85719, USA. ⁴Boston University, Boston, Massachusetts 02215, USA.

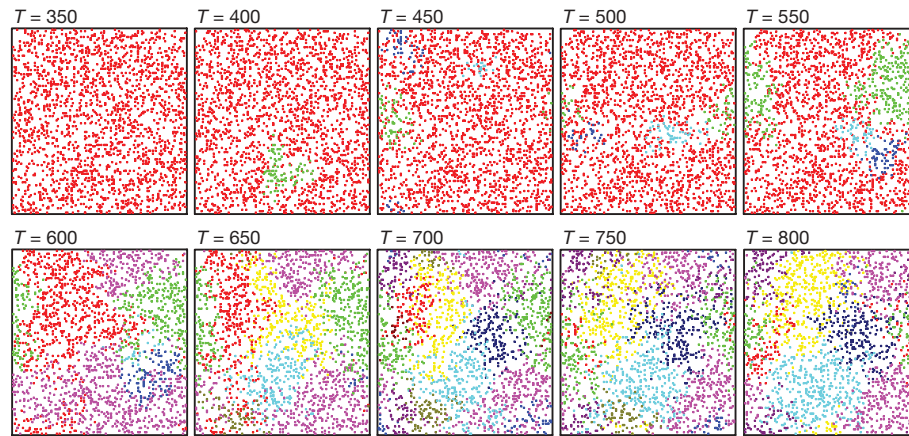


Figure 1 | Time evolution for 2,000 individuals on a 128×128 lattice. Maximum mating distance between two organisms is $S = 6$ lattice cells, and $G = 20$ genetic differences out of 125 total genes. Reproductively isolated species are shown in different colours.

genetic distance reduce the probability of genetic fluctuations arising from mutation and recombination that lead to speciation.

The average spatial density of a species over a population (Supplementary Fig. 2a, b) is consistently well described by a Gaussian

$$\rho(r) = \rho_0 e^{-r^2/R^2}$$

where r is the distance from the geographic centre, ρ_0 is approximately the average density, N/L^2 (where N is the number of organisms and L^2 is the area), and, in the regime of large numbers of species, R is given by

$$R = \alpha \sqrt{\frac{qG}{\mu}} + \beta S^2$$

with $\alpha = 0.18$ and $\beta = 0.26$. By integration, the number of species is $N_S = N/\pi\rho_0 R^2$.

The model describes the spatial dynamics of speciation, from which we can obtain the time dependence, the spatial dependence (species–area relationship) and the distribution of species sizes (species

abundance distribution). In Fig. 3 (see also Supplementary Fig. 2c, d), we summarize the following four key predictions for species diversity consistent with observation.

First, we predict constant rates of new species over long times, as captured by the fossil and genetic data. Figure 3a shows how the number of species varies over time. After a transient period, during which mutation and recombination increase the variation in the initially identical population, the number of species increases rapidly and reaches a steady state subject to statistical fluctuations. The existence of a steady state is sufficient to imply a constant rate of new species arising over long times. This result is consistent with the record of marine invertebrate fossils from the Phanerozoic eon⁶. Over shorter time frames, genetic differences among North American songbirds¹¹ and the fossils of mammals from the Meade basin of Kansas¹² show remarkably steady speciation rates (consistent with our results) and do not correlate with glaciations, disproving this once-dominant view, which was held because of the expectation that speciation is promoted by physical barriers. A constant speciation rate in the absence of landscape forcing is generally supported by the literature on whole-tree cladistic analysis²⁶. This does not discount the potential importance of habitat expansion, such as occurred for North American organisms following glacial retreat.

Second, we predict a high diversity of freshwater versus marine ray-finned fishes when measured by area. Figure 3b shows that the number of species formed increases with the available area if the density of individuals (and corresponding values of other parameters) is kept constant. In one-dimensional environments, the number of species formed is much greater per unit area. Thus, the lower dimensionality of rivers relative to oceans has an enhanced effect on speciation even when specific barriers do not exist. We studied one-dimensional and two-dimensional cases, and river networks may be considered fractal with dimension between one and two, but the same considerations apply. This is consistent with the high diversity of freshwater ray-finned fishes when measured by area and similarly the preponderance of near-shore shallow-water marine species⁷.

Third, we predict a species–area relationship that is largely independent of the type of species or geographical location, is nearly linear on large and small scales and has intermediate scaling behaviour with a low slope that varies among ecological communities^{1,2}. The species–area relationship within a system is shown in Fig. 3c on a log–log plot. The simulated results correspond very well to those found, among others, in birds⁴ and flowering plants⁵ sampled on local to global scales, as well as exhaustive studies of trees and shrubs in Panama¹³ (detailed fit, as shown). On intermediate scales, the slope can be varied with model parameters (inset, Fig. 3c).

Finally, we predict the distribution of species abundance to be well fit by a lognormal distribution with excess rare species. Simulated species abundance distributions (detailed correspondence shown in

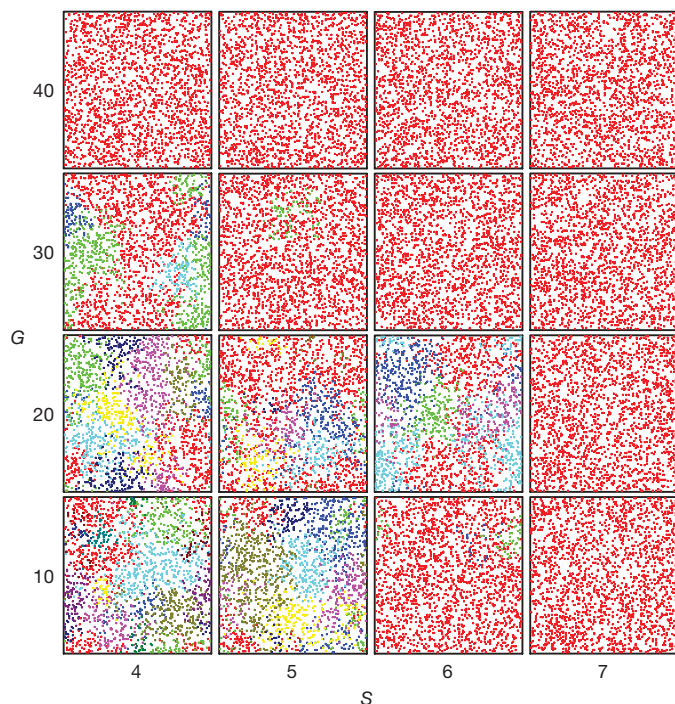


Figure 2 | Spatial snapshots after 1,000 generations for 2,000 individuals on a 128×128 lattice. S and G are as shown. Colours are for different species.

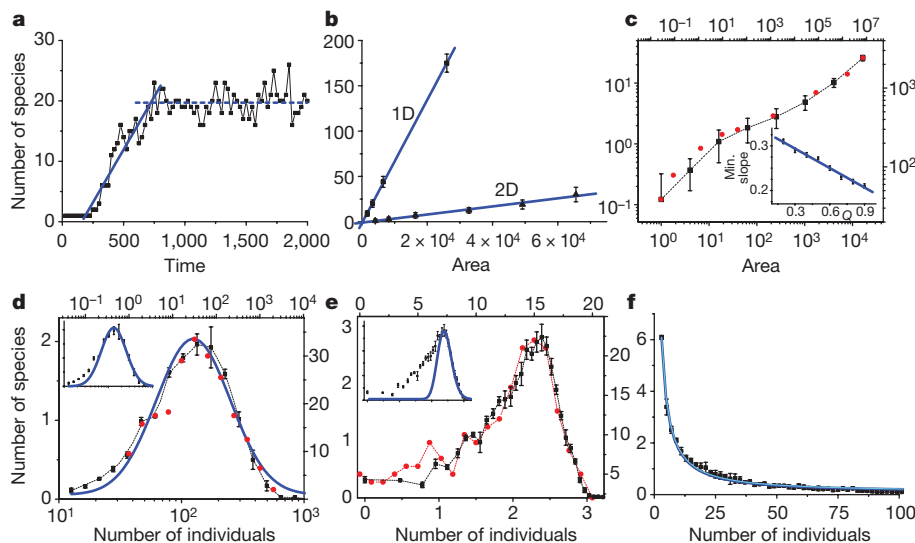


Figure 3 | Species abundance. **a**, Simulated number of species as a function of time ($S = 5$, $G = 20$, also used for **b**, **c**, **d**). Blue fits are a guide to the eye for the increase (solid) and the steady state (dashed). Time measured in generations. **b**, Simulated number of species for increasing system size in one (1D) and two (2D) dimensions ($P < 0.0001$). Area measured in lattice units. Error bars, s.e.m. **c**, Empirical species–area relationship from Panamanian trees¹³ (red; upper and right-hand axes) compared with simulation (black; lower and left-hand axes) of samples with lattice sizes of up to 128×128 , taken from a single simulation of 8,000 individuals on a 256×256 lattice (ranges shown are standard deviations across multiple simulations) fitted by anchoring lowest and highest points ($R^2 = 0.88$). Inset shows smallest slope of simulations as a function of Q (error bars, s.e.m.). **d**, Empirical species abundance distribution of Panamanian trees³ (red; upper and right-hand

Fig. 3d, e) fit observations of Panamanian trees³ and British birds^{8,9}, and are of the form found for moths¹⁰ and various studies of trees², among others^{2,3}. As the number of samples decreases relative to the number of species, the right-hand tail of the distribution, consisting of common species, dominates the sampling, leading to a distribution that can be fit both by the tail of the lognormal and by Fisher's logarithmic series²⁷ (Fig. 3f), as found among many observations^{1,2} (Supplementary Fig. 1). Our species abundance rank distributions also display the observed S-shaped curve².

The key success of neutral theory with a point-mutation model of species formation^{2,3} has been its ability to describe the forms of species–area relationships and species abundance distributions found among many species. Figure 3c–f shows that our model yields similar results. These results are found even though our mechanism includes not only the appearance of localized species, but also the splitting (fission) of a larger species into smaller populations. This does not correspond to allopatric speciation, owing to the absence of geographic barriers. Such splitting leads to key differences from the point-mutation model. For example, the rate at which the steady state arises is much faster because each new species need not start from a single individual. Moreover, the formation of species is not independent of genealogical history. Species fission might be expected to significantly change the behaviour of the species abundance distributions, but we nevertheless find a lognormal distribution with excess rare species and detailed agreement with observed species abundance and species–area relationships.

We note that there are also regimes of the model in which speciation is quite difficult. This accounts for the existence of some observations where geographic barriers and selection do have an important role; for example, observations of greenish warblers¹⁴ have shown the direct impact of geographic barriers and genetic variation across continent-wide environmental gradients.

We have shown that our distributions of species reach a steady state rapidly, with extinctions balanced by speciation consistent with

axes) compared with simulation (black, with error bars; lower and left-hand axes; $\chi^2 = 3.1$, $P = 0.01$) and a lognormal distribution (blue; $\chi^2 = 3.9$, $P = 0.08$). Inset shows a lognormal distribution fitted only to highest and rightmost simulation points. Error bars, s.e.m. **e**, As in **d** (lognormal only in inset, and dashed red line added for clarity), but for British birds ($S = 6$, $G = 10$, $N = 8,000$, $Q = 0.5$) ($\chi^2 = 5.5$, $P < 0.0001$). The distribution is consistent with a larger effective mating distance, although other parameters may affect this conclusion. Even migratory birds satisfy the model assumptions of mating proximate to their locations of birth³⁰. Error bars, s.e.m. **f**, Simulated species abundance for sampling over a 64×64 subset of a 256×256 lattice ($S = 6$, $G = 10$, $N = 8,000$; Supplementary Fig. 1), fitted to two essentially overlapping curves: lognormal (blue; $\chi^2 = 0.009$, $P < 0.0001$) and log-series (cyan; $\chi^2 = 0.005$, $P < 0.0001$). Error bars, s.e.m.

observations, and that species–area relationships and species abundance distributions are consistent with those frequently found in nature. The general forms of the distributions are as observed, and specific simulations are in remarkable correspondence with specific observations. Underlying the agreement is a simple model of a sexually reproducing population in a spatial context with mating restrictions depending on genetic and spatial distance—these are the essential properties for the dynamic biodiversity patterns. Our model dynamically subdivides spatial regions, which may be called topographic speciation. The spontaneous creation of new spatial patterns is an example of the general process of spontaneous pattern formation and symmetry breaking found in many systems²⁸, including traffic jams¹⁵. Pre-existing geographic features and variation may influence patterns, but are not necessary for their formation. The relationship between this process and niche-ordering models²⁹ could be explored.

METHODS SUMMARY

We used agent-based simulations in which agents identified by geographical location and genotype undergo sexual reproduction in pairs limited by geographical and genetic proximity. Each offspring replaces a parent and differs in genotype according to genetic inheritance from both parents, with crossover and mutation, and in geographical location by dispersal. According to tests of multiple model variants, including parameter variation, sequential and synchronous mating, hermaphroditism and two sexes, single or multiple crossovers, direct assortative mating and separate fitness and sexual selection genetic components, our results apply quite generally, with the key properties needed being the limitations on spatial and genetic distances of mating.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 February; accepted 22 May 2009.

1. Rosenzweig, M. L. *Species Diversity in Space and Time* (Cambridge Univ. Press, 1995).
2. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, 2001).

3. Volkov, I., Banavar, J. R., Hubbell, S. P. & Maritan, A. Patterns of relative species abundance in rainforests and coral reefs. *Nature* **450**, 45–49 (2007).
4. Preston, F. W. Time and space and the variation of species. *Ecology* **41**, 611–627 (1960).
5. Schmida, A. & Wilson, M. V. Biological determinants of species diversity. *J. Biogeogr.* **12**, 1–20 (1985).
6. Alroy, J. *et al.* Phanerozoic trends in the global diversity of marine invertebrates. *Science* **321**, 97–100 (2008).
7. Moyle, P. B. & Cech, J. J. *Fishes: An Introduction to Ichthyology* 5th edn (Benjamin Cummings, 2003).
8. Gibbons, D. W., Reid, J. B. & Chapman, R. A. (eds) *The New Atlas of Breeding Birds in Britain and Ireland: 1988–1991* (Poyser, 1993).
9. Gregory, R. D. Species abundance patterns of British birds. *Proc. R. Soc. Lond. B* **257**, 299–301 (1994).
10. Williams, C. B. *Patterns in the Balance of Nature and Related Problems of Quantitative Biology* (Academic, London, 1964).
11. Klicka, J. & Zink, R. M. The importance of recent ice ages in speciation: a failed paradigm. *Science* **277**, 1666–1669 (1997).
12. Rosenzweig, M. L. Tempo and mode of speciation. *Science* **277**, 1622–1623 (1997).
13. D'Arcy, W. E. *Flora of Panama: Checklist and Index* Vols 1–4 (Missouri Botanical Garden, 1987).
14. Irwin, D. E., Bensch, S., Irwin, J. H. & Price, T. D. Speciation by distance in a ring species. *Science* **307**, 414–416 (2005).
15. Nagel, K. & Schreckenberg, M. A cellular automaton model for freeway traffic. *J. Phys. I* **2**, 2221–2229 (1992).
16. Doebelli, M., Dieckmann, U., Metz, J. A. J. & Tautz, D. What we have also learned: adaptive speciation is theoretically plausible. *Evolution* **59**, 691–695 (2005).
17. Gavrillets, S. “Adaptive speciation”—it is not that easy: a reply to Doebeli *et al.* *Evolution* **59**, 696–699 (2005).
18. Payne, R. J. H. & Krakauer, D. C. Sexual selection, space, and speciation. *Evolution* **51**, 1–9 (1997).
19. Doebeli, M. & Dieckmann, U. Speciation along environmental gradients. *Nature* **421**, 259–264 (2003).
20. Hoelzer, G. A., Drewes, R., Meier, J. & Doursat, R. Isolation-by-distance and outbreeding depression are sufficient to drive parapatric speciation in the absence of environmental influences. *PLoS Comput. Biol.* **4**, e1000126 (2008).
21. Higgs, P. G. & Derrida, B. Stochastic models for species formation in evolving populations. *J. Phys. A* **24**, L985–L991 (1991).
22. Kondrashov, A. S. & Shpak, M. On the origin of species by means of assortative mating. *Proc. R. Soc. Lond. B* **265**, 2273–2278 (1998).
23. Lewontin, R., Kirk, D. & Crow, J. Selective mating, assortative mating and inbreeding: definitions and implications. *Eugen. Q.* **15**, 141–143 (1968).
24. Dobzhansky, T. *Genetics and the Origin of Species* (Columbia Univ. Press, 1937).
25. Gavrillets, S., Li, H. & Vose, M. D. Patterns of parapatric speciation. *Evolution* **54**, 1126–1134 (2000).
26. Ricklefs, R. E. Estimating diversification rates from phylogenetic information. *Trends Ecol. Evol.* **22**, 601–610 (2007).
27. Fisher, R. A., Corbet, A. S. & Williams, C. B. The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* **12**, 42–58 (1943).
28. Bar-Yam, Y. *Dynamics of Complex Systems* (Perseus, 1997).
29. Sugihara, G., Bersier, L.-F., Southwood, T. R. E., Pimm, S. L. & May, R. M. Predicted correspondence between species abundances and dendrograms of niche similarities. *Proc. Natl Acad. Sci. USA* **100**, 5246–5251 (2003).
30. Greenwood, P. J. & Harvey, P. H. The natal and breeding dispersal of birds. *Annu. Rev. Ecol. Syst.* **13**, 1–21 (1982).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge internal support by the New England Complex Systems Institute. M.A.M.d.A. and E.M.B. acknowledge financial support from the Fundação de Amparo à Pesquisa do Estado de São Paulo and the Conselho Nacional de Desenvolvimento Científico e Tecnológico, and L.K. from the Marine Management Area Science Program of Conservation International and the Gordon and Betty Moore Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to Y.B.-Y. (yaneer@necsi.edu).

METHODS

For the simulations reported in the paper, we considered N haploid and hermaphroditic individuals placed randomly on a lattice (128×128 or 256×256) where multiple individuals can exist at the same site. Genomes consist of binary strings of length B , and the genetic distance between two individuals is the number of differences along the genome. Reproduction starts with an individual looking for a mate inside its spatial mating radius, S . To qualify, an individual must have a genetic distance no greater than G from the seeker, reflecting differences in traits or offspring viability. From all such potential mates, we select one at random. If there are fewer than a minimum number of possible mates (an additional model parameter, P), we weaken the spatial and genetic constraints as follows for that mating: $S \rightarrow S + 1$ and $G \rightarrow G + 1$. If the number of available mates is still smaller than P , the process is repeated. (We also investigated letting $S \rightarrow S + 1$ without changing G and, if there were not enough mates in the entire space, choosing a neighbour to reproduce instead, with similar results.)

Each individual is a seeker once in a generation, but there is a probability, Q , that it will not reproduce at all. When that happened, another individual is randomly selected from its spatial neighbourhood to reproduce in its place, with the offspring placed in the location of the original seeker. This maintains an approximately homogeneous demographic density, corresponding to a fixed local carrying capacity. Each reproduction occurs with one genetic crossover (the genome up to a certain point is from one parent and after that is from the other parent), reflecting

genetic linkage. Mutation occurs with a probability μ . Offspring dispersal occurs with probability D ; the offspring is placed at one of the 20 neighbouring sites, chosen randomly. For the results reported, $B = 125$, $\mu = 0.001$, $Q = 30\%$, $P = 8$ and $D = 0.01$. Parameter variation affected details, but not the overall behaviour. For example, robustness can be demonstrated by setting $Q = 0\%$, $D = 0$ and $P = 1$. The effects of varying S and G are described in the text. Results for the number of species and distributions are obtained by time-averaging over 10,000 generations, sampling every 25 generations after reaching a steady state.

The results reported here are for asynchronous mating, although many variations of the model were implemented, with no essential changes in the behaviour. Synchronous mating gave similar results, although speciation times were typically twice as long. We also considered models that indirectly impose the genetic mating distance through selection on genetic differences that affect offspring viability and genetic differences regulating behavioural traits that promote reproductive isolation, models with multiple crossovers, models with adaptive changes in spatial and genetic mating distances when few potential mates were available, and models with populations of two sexes. We also obtained speciation for significantly lower mutation rates. For $\mu = 0.00001$, speciation occurred for $S = 3$, $G = 7$ and $N = 2,500$ in a 128×128 lattice after 7,000 generations. Robustness suggests that improving the realism of the model by including other assumptions (details of mating and recombination, diploidy and dioecism) will not affect the basic results.