Classification: Biological Sciences - Evolution

Title: Evolution and Stability of Ring Species

Authors: Ayana B. Martins[1], Marcus A. M. de Aguiar[2,3] and Yaneer Bar-Yam[3*]

Affiliations:
[1] Instituto de Biociências, Universidade de São Paulo, USP, 05508-090, São Paulo, SP, Brazil

[2] Instituto de Física Gleb Wataghin, Universidade Estadual de Campinas, Unicamp, 13083-970, Campinas, SP, Brazil

[3] New England Complex Systems Institute, 238 Main St., Suite 319, Cambridge, MA 02142, USA

*Corresponding author:

Yaneer Bar-Yam

New England Complex Systems Institute, 238 Main St., Suite 319

Cambridge, MA 02142, USA

Telephone: 1-617-547-4100

e-mail: yaneer@necsi.edu

Abstract: Neutral models, in which genetic change arises through random variation without fitness differences, have proven remarkably successful in describing observed patterns of biodiversity, despite the manifest role of selection in evolution. Here we investigate the effect of barriers on biodiversity by simulating the expansion of a population around a barrier to form a ring species, where the two ends of the population are reproductively isolated despite ongoing gene flow around the ring. We compare the spatial and genetic properties of a neutral agent-based population model to the greenish warblers' complex, a well-documented example of an actual ring species in nature. Our results match the distribution of subspecies, the principal components of genetic diversity and the linear spatial-genetic correlation of the observed data, even though selection is expected to be important for traits of this species. We find that ring species are often unstable to speciation or mixing but can persist for extended times depending on species and landscape features. For the greenish warblers our analysis implies that the expanded area near the point of secondary contact is important for extending the duration of the ring, and thus for the opportunity to observe this ring species. Nevertheless it also suggests the ring will break up into multiple species in 10,000 to 50,000 years. These results imply that simulations can be used to accurately describe empirical data for complex spatial-genetic traits of an individual species.

\body

Introduction

There is a growing paradox in our understanding of the core issues of speciation and biodiversity.

First, the manifest empirical importance of selection in the traits of species appears to be counter to the ability to describe biodiversity using neutral models of speciation (1–4). Second, geographical heterogeneity, including geographical isolation, is no longer considered necessary to account for the emergence of new species. Traditionally the emergence of new species has been explained by a period of geographic isolation of subpopulations. In the absence of gene flow, genetic differences accumulate and reproductive isolation eventually arises (5). However, if population ranges are large enough, genetic divergence can lead to speciation despite ongoing gene flow (6), as demonstrated by empirical and theoretical results (7–14). Moreover, it has been shown that existing patterns of diversity are consistent with such speciation with ongoing gene flow and without trait selection or spatial heterogeneity, as long as mating is constrained by spatial and genetic distances (15). Here we show that neutral models of speciation can also account for the genetic diversity of a single species in a context where geographical barriers play a central role in that diversity. We compare the diversity of a neutral model to empirical findings about a ring species formed when population expansion occurred around a geographical barrier. The genetic patterns are consistent even though selection is expected to be important for traits of this species (11, 16, 17).

Currently, the best documented example of a ring species is that of the greenish warblers which inhabit a ring around the Tibetan Plateau (11). In the northernmost area of this ring are two Siberian taxa that are reproductively isolated and occupy overlapping geographical regions (18). Gene flow is still possible between them over multiple generations through a chain of reproductively linked taxa around the ring. One noteworthy characteristic is that the Siberian taxa, which occur in the area of secondary contact (*i.e.* ring closure) of the expanding population, have much larger distribution ranges than do other subspecies. The correlation between range size and latitude is a strong pattern found in many terrestrial groups of the Old World (19).

We simulated the formation of a ring species (Fig. 1), explicitly including the ring topography and allowing a small initial population to grow as it expands and differentiates around a geographical barrier (Fig. S1). We used an individual based model based on neutral replacement with local mating, migration and mutation (see Methods). Reproductive isolation is modeled by a multilocus generalization of the Batenson-Dobzhansky-Muller model where the population evolves through nearly constant fitness ridges (20, 21). The initial population is set within a starting area, and the individuals whose local mating area is underpopulated generate two rather than one offspring before expiring. This enables the population to grow and spread up to the carrying capacity of the entire available space.

Results and discussion

The genetic variation plot (Fig. 1, C and D) reveals a gradual genetic change over geographical space (see also Figs. S2 and S3). Simulated and actual warbler populations have a similar decomposition of variance (15.9 versus 19.4% for the first principal component and 10.8 versus 5.6% for the second principal component, respectively). The genetic distance between individuals increases linearly with their geographic distance as measured around the ring in both cases (Fig. 1, E and F). Different behaviors are expected for a single mixed species or multiple species (Fig. S4). For the sampling shown in Figure 1 the average genetic distance within populations is approximately 3 and the genetic distance between the populations at the area of secondary contact is close to 7. This is in agreement with AFLP data (11) where AFLP distances between populations at secondary contact is about twice the average distance within populations. The number of genetic differences in the sampled model corresponds to the same proportion of AFLP differences in the sampled experimental observations. This indicates that populations separated by half the ring's length should already be reproductively isolated. The terminal populations came into contact after ~2000 generations (Fig. 2, middle line). The number of generations is consistent with the timing of habitat recovery due to tree cover expansion—an important factor in the spread of the greenish warbler. The forest range expanded starting about 10,000 years ago and achieved its current area approximately 2,000 years later (17). Given the greenish warblers' annual life cycle and assuming that their range expansion accompanied this forest recovery (19), the rates of population expansion in our simulations coincide, consistent with empirical evidence (17). A sharp boundary between reproductively isolated subspecies in the area of secondary contact is maintained by the difficulty in finding compatible mates when an individual is surrounded by the other type. The spatial and genetic restrictions in mating lead to selection against rare types in favor of common types in the region of overlap. The U-shaped pattern in Figs. 1C and 1D, and Fig. S3B, is a consequence of the gene flow around the ring, that couples the otherwise independent expanding fronts. If the fronts begin to mix into a single species the bending increases and the U-shape turns into an O that closes into a single spot.

There seems to be an empirical mismatch between the expected spread of the population around the ring 10,000 years ago and the genetic divergence of types measured using mitochondrial DNA clocks, which is 2 million years (11). This may be resolved through the observation that asexual replication of mitochondria diverges genetically even without speciation, while genetic bounds on reproduction limit the range of the sexual component of the genome. Prior to ring formation when the population evolved while located solely in the southernmost area, the diversity of mitochondrial DNA would diverge, but would be spatially correlated in local patches reflecting branches of the ancestral tree (23, 24). The spreading of the population from one patch around one arm and another patch around the second arm of the ring would result in large divergences between the mitochondrial DNA of the types that meet after ring closure. The divergence would be characteristic of the time to common ancestor of that DNA, but not of the time of ring formation. This scenario is consistent with simulations (Fig. S5).

Although natural and/or sexual selection appear to be important for greenish warblers traits (11, 16, 17) and other potential ring species (13, 18, 25–29), our results show that their role in the emergence of reproductive isolation and maintenance of ring species (as suggested in (11)) is not yet well understood. Consistent with the importance of neutral divergence in the formation of ring species, for the salamander *Ensatina eschscholtzii*, a second well documented ring species (reviewed in (30)), divergence in neutral nuclear markers is a better predictor of reproductive isolation than ecological divergence (31).

Our model of ring species formation also provides an opportunity to clarify and analyze the role of geographical barriers. There is a dichotomy that can be readily understood: fast enough differentiation leads to multiple species due to isolation by distance (15), slow differentiation to a single species (Fig. 2). Our question, however, is whether a third intermediate case of ring species can be robustly formed so that they are observed in nature. When a population expands around a spatial barrier that limits its range, the interplay between the velocity of range expansion and genetic differentiation will determine the fate of the terminal populations when they come into secondary contact. Increasing carrying capacity decreases differentiation (32) and increases the velocity. Other features that affect this interplay include population features (*e.g.* growth rate, mating distance, dispersal rate and mobility) and landscape features (*e.g.* corridor width, distance around the barrier). None of these provides stability for a ring species. We tested the robustness of ring species formation to variation in the geographical structure of the ring. We hypothesized that the larger habitable area present near the region of secondary contact enhanced the likelihood of ring species formation. Larger populations of the divergent types inhibit genetic fluctuations that could lead to mating opportunities that would revert the speciation process (33). Simulations confirmed that when the excess area was displaced to a different location in the ring near the point of origin, equivalent to shifting the point of origin from south to north, ring species were less likely to form, even though the time until secondary contact between terminal populations remained the same (Fig. 3). The enlarged regions in the contact area reduce genetic variability over time, and selection against rare types limits the overlap area, reducing the probability of merger.

Fig. 4 shows the occurrence of ring species formation as a function of the dispersal rate and carrying capacity, which are related to the rate of population expansion. The comparison of results in the two different geographical structures highlights the importance of the specific structure of the landscape (not just geographic distance around the ring) in regulating genetic variability in space and thus the outcome of the speciation process.

Even though we find that ring species can form during expansion, given the possibility of speciation without barriers, an important question is whether ring species are ultimately stable either to breaking up into multiple species or to consolidation to a single interbreeding population. The extent of their persistence is important for determining their prevalence in nature. In order to evaluate stability and persistence, we extended simulations that resulted in ring species after 10,000 generations. We performed these simulations for both of the geographical structures previously identified. In the original scenario in which ring species are more likely to form, we found they persist for many generations, though eventually most resolved to single mixed or multiple species. The time until 90% of ring species collapsed, either by speciating or mixing, was 130,000 generations. For the simulation shown in Fig. 1, the ring species is expected to break up into multiple species in 10,000 to 50,000 years (Figs. S6 and S7). Thus, most ring species are not stable, but they can persist for a long time and thus be observed. These results counter the idea of ring species instability reported in previous models (9, 34, 35).

Comparisons between closely related taxa that share a history of range expansion around the same barriers could provide empirical insight into the relation between species' and landscape parameters in ring species formation. Old World leaf-warblers, genus *Phylloscopus*, are forest-dependent species and the range expansion of their ancestral populations is thought to be associated with forest expansion following warming of the climate after Pleistocene glaciations (19). However, the only case of ring species confirmed so far for this group is the greenish warbler *(P. trochiloides*). The species in *P. inornatus* complex, for example, have been considered separate species based on genetic and song divergence (36). Salamanders in western

North America may provide another interesting case. Despite some taxonomic controversy, the *Ensatina eschscholtzii* complex is still one of the few well documented ring species (reviewed in (30)) and it occurs in sympatry with other Salamander genera (27).

The fine-tuning of population, individual and landscape parameters necessary for ring species formation in our simulations, as well as their ultimate instability, suggest that ring species can be expected to be rare in nature. For a given barrier only species within a narrow range of mobility are candidates to display rings. More mobile species, e.g. birds, require larger barriers than do reptiles or small mammals. The key genetic evolutionary parameter is the underlying rate of genetic isolation, the mutation rate divided by the genetic isolation distance. Ring species will form if the period of expansion L/v, where L is the ring size and v the speed of expansion times the rate of genetic isolation w is approximately one, i.e. if $Lw/v \sim 1$. Even under this condition rings will be more or less likely depending on the details of the geography. Given the wide range of observed results, simulations may be used to characterize the dynamics of range expansion genetics for direct comparison in multiple cases, including single, multiple and ring species.

Finally, this work also demonstrates that barriers can dramatically increase speciation, even when they do not cause isolation. Many of the simulations that resulted in multiple species in the ring geometry would result in a single species were the barrier not present. Speciation is due to the increase in effective distance around the ring, as well as the reduction in carrying capacity of the narrow circumferential habitat. This speciation by partial obstruction complements the idea of speciation by partial isolation, i.e. parapatric speciation.

Methods

The simulations use an agent based population model that consists of haploid and hermaphroditic individuals characterized by a binary string of B loci. Reproduction is restricted by a mating area determined by a spatial distance, S and a genetic distance, G, above which other individuals are not considered potential mates. Reproductive isolation caused by this mechanism can be considered a multilocus generalization of the Batenson-Dobzhansky-Muller model where individuals accumulate genetic incompatibilities without the population passing through an adaptive valley (21). Reproductive success depends on the number of genetic differences between the individuals attempting mating, without affecting the fitness of the individuals themselves (20). The genetic distance between two individuals is the number of differing loci along their strings. The geographical space is represented by a 128×142 lattice with barriers defined by sites that cannot be occupied by individuals (Fig. S1). The remaining sites can be occupied by multiple individuals. The carrying capacity, K, determines the maximum population size allowed in the entire population. The expected demographic density, ρ, is calculated as the ratio between K and the number of available sites. The initial population consists of genetically identical individuals randomly placed in a fixed area bellow the larger barrier, shown in Figure S1. The number of individuals in the initial population is given by ρ times the number of sites in this area.

The population evolves in discrete generations. Each individual seeks potential mates for reproduction once and is replaced by the resulting offspring. However, there is a probability Q that the individual will not reproduce at all. In this case, a neighbor is selected at random to reproduce in its place. In each reproductive event, one individual is selected at random from the list of potential mates (meeting both spatial and genetic criteria) and the genetic information of both parents is combined with one genetic crossover for which the breakpoint is randomly selected along the string. There is a mutation rate per locus, μ. Offspring dispersal occurs with probability D; the offspring is placed at one of the 20 neighboring sites, chosen randomly. In addition, if the demographic density in the mating area is lower than 60% of the expected demographic density, ρ, a second offspring is produced. This allows for the spreading of the population towards empty areas, as shown in Figures 3 and S2. After the population reaches the carrying capacity, each individual can have only one offspring. If there are fewer than P potential mates for a specific individual, the mating area for that individual is increased by setting S to $S + 1$. If the number of available mates is still smaller than P, the process is repeated. S may be increased by a maximum of 10 units, in which case, if there are still less than P potential mates, a neighbor is randomly selected to reproduce instead. A highly divergent individual is likely to be replaced by a more common type. This selects against rare types when they coexist with a common type in a geographical region.

For the results reported, $B = 125$, $S = 9$, $G = 20$, $\mu = 0.0003$, $Q = 0.3$, $P = 20$ and the geographical barriers occupy 11796 out of 18176 sites, leaving 6380 available (see Fig. S1)

A species is identified as a group of organisms that are connected by gene flow and separated from all others by the genetic restriction on mating determined by G. This definition does not require all members of a species to be able to mate with each other; two members can be incompatible, as long as they can exchange genes indirectly through other members of the species. As an example, three individuals *A*, *B* and *C* whose genetic distances satisfy $d(A,B) < G$, $d(B,C) < G$ but $d(A,C) > G$ belong to the same species. 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. Considering this strategy to delimitate species, ring species are identified as a population in which we would identify two species when considering only the individuals located in the area of secondary contact (Fig. S1), but a single species when considering the population as a whole.

In all figures where individuals are colored, the color is according to the genetic distance to a reference individual, marked as a star. Dark blue corresponds to potential mates for the reference individual (those whole genetic distance is less than the critical value, *G*). The remaining colors represent genetic distances in the following intervals: light blue $[G, 1.15G]$; green $[1.15G, 1.15^2 G]$; yellow $[1.15^2 G, 1.15^3 G]$; orange $[1.15^3 G, 1.15^4 G]$ and red, greater than or equal to $1.15^4 G$. Ring species formation is sensitive to variations in *S*, *G*, μ, *K* and *D*. The effects of varying *K* and *D* are described in the text and figures; varying other parameters combinations yield qualitatively similar results.

Acknowledgments: We thank Sergey Gavrilets, Trevor Price and Les Kaufman for helpful comments on the manuscript. ABM and MAMA acknowledge financial support from FAPESP and MAMA also thanks CNPq.

References

1. Hubbell SP (2001) *The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32)* (Princeton University Press).

2. Volkov I, Banavar JR, Hubbell SP, Maritan A (2007) Patterns of relative species abundance in rainforests and coral reefs. *Nature* 450:45–49.

3. Muneepeerakul R et al. (2008) Neutral metacommunity models predict fish diversity patterns in Mississippi–Missouri basin. *Nature* 453:220–222.

4. Kopp M (2010) Speciation and the neutral theory of biodiversity: Modes of speciation affect patterns of biodiversity in neutral communities. *Bioessays* 32:564–570.

5. Coyne JA, Orr HA (2004) *Speciation* (Sinauer Associates, Inc.). 1st Ed.

6. Turelli M, Barton NH, Coyne JA (2001) Theory and speciation. *TREE* 16:330–343.

7. Rice WR, Hostert EE (1993) Laboratory Experiments on speciation: What have we learned in 40 Years? *Evolution* 47:1637–1653.

8. Smith TB, Wayne RK, Girman DJ, Bruford MW (1997) A role for ecotones in generating rainforest biodiversity. *Science* 276:1855–1857

9. Gavrilets S, Li H, Vose MD (1998) Rapid parapatric speciation on holey adaptive landscapes. *Proc Biol Sci* 265:1483–1489.

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

11. Irwin DE, Bensch S, Irwin JH, Price TD (2005) Speciation by distance in a ring species. Science 307:414 –416.

12. Marthinsen G, Wennerberg L, Lifjeld JT (2007) Phylogeography and subspecies taxonomy of dunlins (*Calidris alpina*) in western Palearctic analysed by DNA microsatellites and amplified fragment length polymorphism markers RID B-1978-2008. Biol J Linnean Soc 92:713–726.

13. Bensch S, Grahn M, Müller N, Gay L, Åkesson S (2009) Genetic, morphological, and feather isotope variation of migratory willow warblers show gradual divergence in a ring. Mol Ecol 18:3087–3096.

14. Pereira RJ, Wake DB (2009) Genetic leakage after adaptive and nonadaptive divergence in the *Ensatina eschscholtzii* ring species. *Evolution* 63:2288–2301.

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

16. Irwin DE (2000) Song variation in an avian ring species. *Evolution* 54:998–1010.

17. Irwin DE, Irwin JH (2005) in *Birds of two worlds: the ecology and evolution of migration* (JHU Press), pp 27–40.

18. Irwin DE, Irwin JH, Price TD (2001) Ring species as bridges between microevolution and speciation. *Genetica* 112-113:223–243.

19. Price TD, Helbig AJ, Richman AD (1997) Evolution of breeding distributions in the old world leaf warblers (genus *Phylloscopus*). *Evolution* 51:552–561.

20. Gavrilets S (2000) Waiting time to parapatric speciation. *Proc Biol Sci* 267:2483–2492.

21. Gavrilets S (2004) *Fitness landscapes and the origin of species (MPB-41)* (Princeton University Press).

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

23. Rauch EM, Bar-Yam Y (2004) Theory predicts the uneven distribution of genetic diversity within species. *Nature 431:*449–452.

24. Rauch EM, Bar-Yam Y (2005) Estimating the total genetic diversity of a spatial field population from a sample and implications of its dependence on habitat area. *PNAS* 102:9826–9829.

25. Price TD (2010) The roles of time and ecology in the continental radiation of the Old World leaf warblers (*Phylloscopus* and *Seicercus*). *Phil Trans R Soc B* 365:1749–1762.

26. Alexandrino J et al. (2005) Strong selection against hybrids at a hybrid zone in the *Ensatina* ring species complex and its evolutionary implications. *Evolution* 59:1334–1347.

27. Wake DB (2009) What salamanders have taught us about evolution. *Annu Rev Ecol Evol Syst* 40:333–352.

28. Cacho NI, Baum DA (2012) The Caribbean slipper spurge *Euphorbia tithymaloides*: the first example of a ring species in plants. *Proc Biol Sci 279*:3377–3383.

29. Devitt TJ, Baird SJE, Moritz C (2011) Asymmetric Reproductive isolation between terminal forms of the salamander ring species *Ensatina eschscholtzii* revealed by fine-scale genetic analysis of a hybrid zone. *BMC Evolutionary Biology* 11:245.

30. Kuchta SR, Parks DS, Mueller RL, Wake DB (2009) Closing the ring: historical biogeography of the salamander ring species *Ensatina eschscholtzii. J Biogeogr* 36:982–995.

31. Pereira RJ, Monahan WB, Wake DB (2011) Predictors for reproductive isolation in a ring species complex following genetic and ecological divergence. *BMC Evolutionary Biology* 11:194.

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

33. Nosil P, Harmon LJ, Seehausen O (2009) Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution* 24:145–156.

34. Noest AJ (1997) Instability of the sexual continuum. *Proc Biol Sci* 264:1389–1393.

35. 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. *J Theor Biol* 264:1202–1213.

36. Irwin DE, Alström P, Olsson U, Benowitz, Fredericks ZM (2001) Cryptic species in the genus *Phylloscopus* (Old World leaf warblers). *Ibis* 143:233–247.

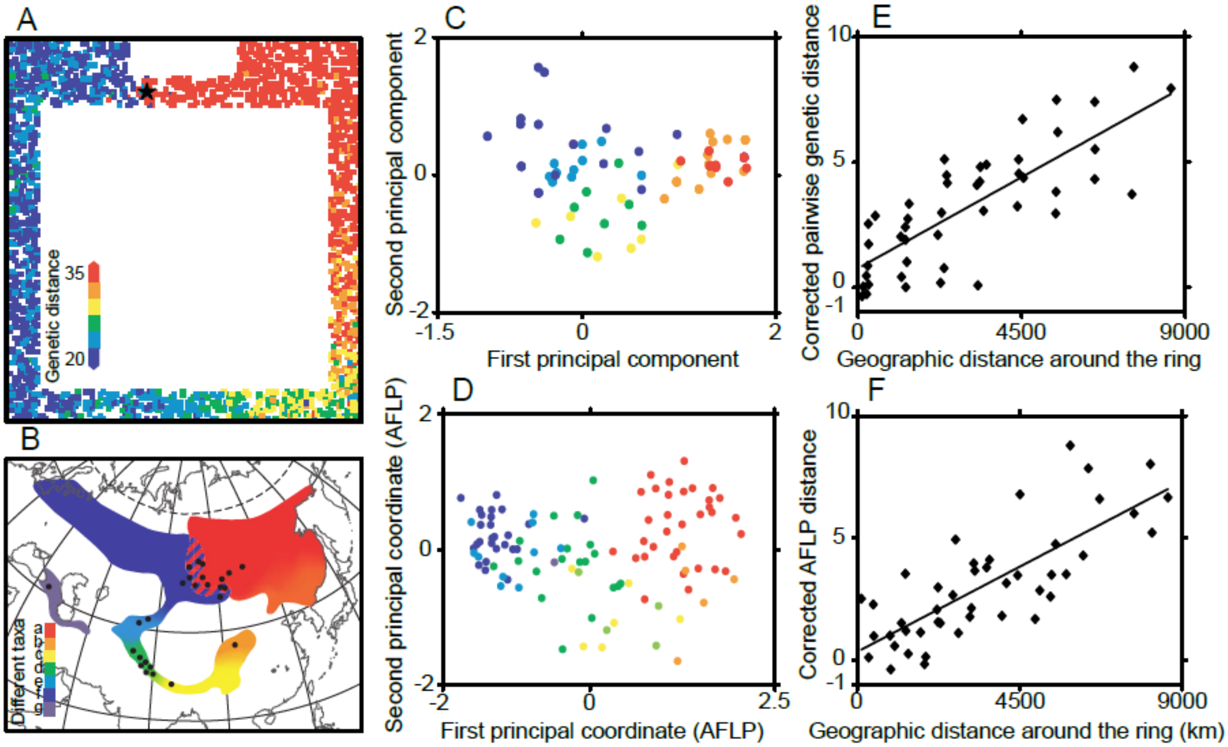

Fig. 1. Results from one simulation (A, C and E) are shown on top of data from the greenish warbler (*Phylloscopus trochiloides*) ring species [B, D and F, adapted from (11)]. The spatial snapshot (A) shows the result of our simulation of ring species formation after 10,000 generations. Colors represent the genetic distance to a reference individual marked as a star (see Methods). The map of Asia (B) shows the range of greenish warblers in the breeding season with sampling sites indicated by black dots. Colors represent different subspecies: a, *plumbeitarsus; b, obscuratus;* c, *trochiloides*; d, *ludlowi*; e and f, *viridanus*; g, *nitidus*. The hatched area in central Siberia indicates the overlap zone between two subspecies and the gap in the distribution of the subspecies shown in orange is probably due to recent habitat destruction (22). C and D show genetic data as summarized by principal component analysis and principal coordinate analysis, respectively. We sampled 25 loci from 85 individuals from the simulated population. For the warblers the genetic data corresponds to 62 AFLP markers (see Fig. S3 for details). Colors represent the same as the panel to the left. E and F show the genetic distance as a function of the geographic distance measured around the ring for the same sample shown in C and D. Corrected average pairwise distances between populations were calculated as the mean number of pairwise differences between two populations minus the average distance between individuals within those populations (see Fig. S4 for details). The lines are fit by least-squares regression to all points. Ring length and genetic distances were re-scaled on the left panel to be comparable with warblers' data.

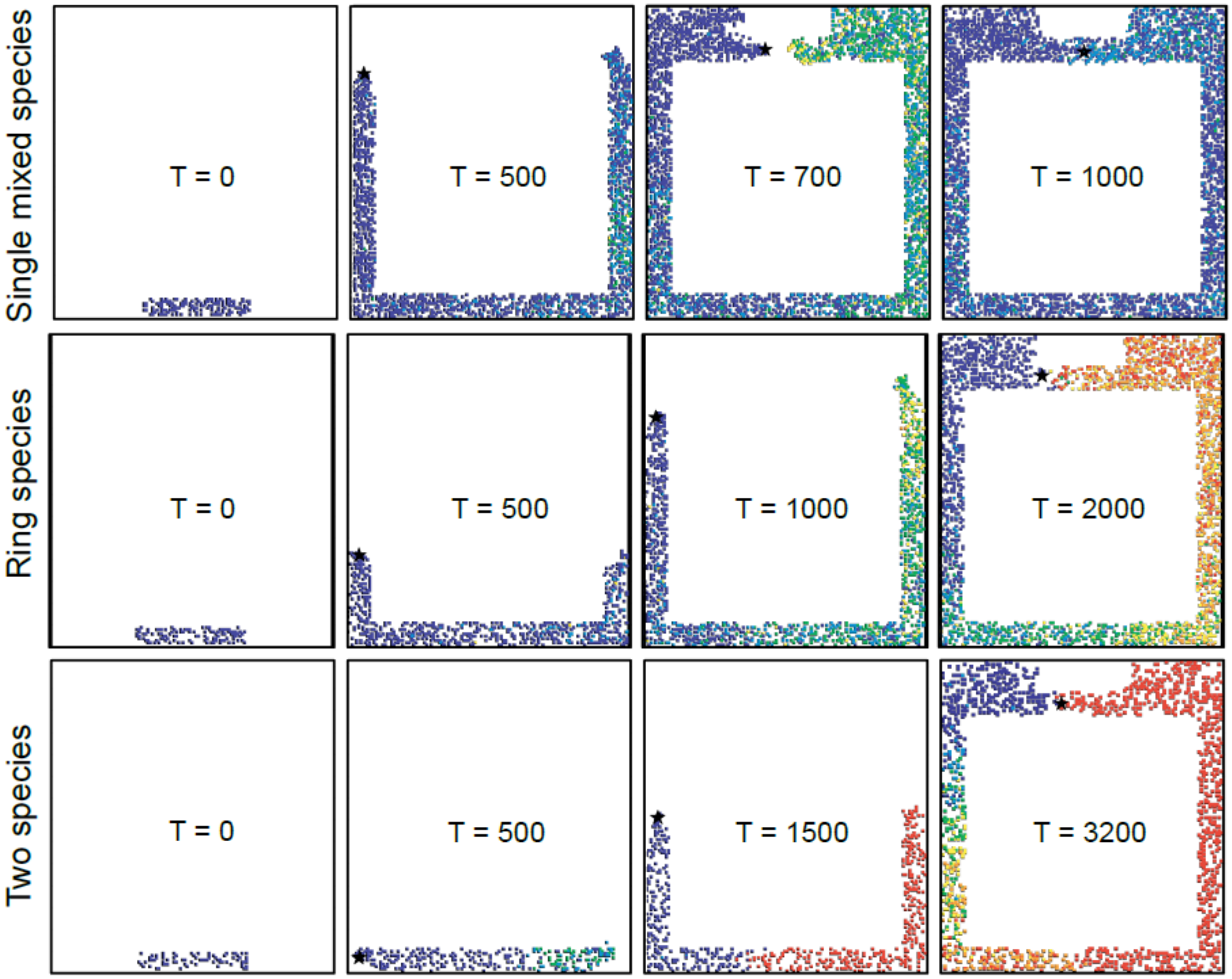

Fig. 2. Three simulations of population expansion around a geographical barrier resulting in single species, ring species and multiple species. Snapshots show the time evolution of the simulated populations as they expand (T is the number of generations) for different values of carrying capacity (*K*) and dispersal rate (*D*). The last snapshot shows the population a few hundred generations after secondary contact. Colors represent the genetic distance to a reference individual marked as a star. In the last panel (T=3200) the genetic variation can be seen on the left in the color variation but not on the right, as all individuals on the right are incompatible with the starred individual, and therefore are marked as red. Single species (top): $K = 3600$ and $D = 0.3$. Ring species (middle): The same simulation shown in Figure 1, $K = 2400$ and $D = 0.15$. Multiple species (bottom): $K = 1500$ and $D = 0.1$, two completely reproductively isolated groups are present after 1500 generations.

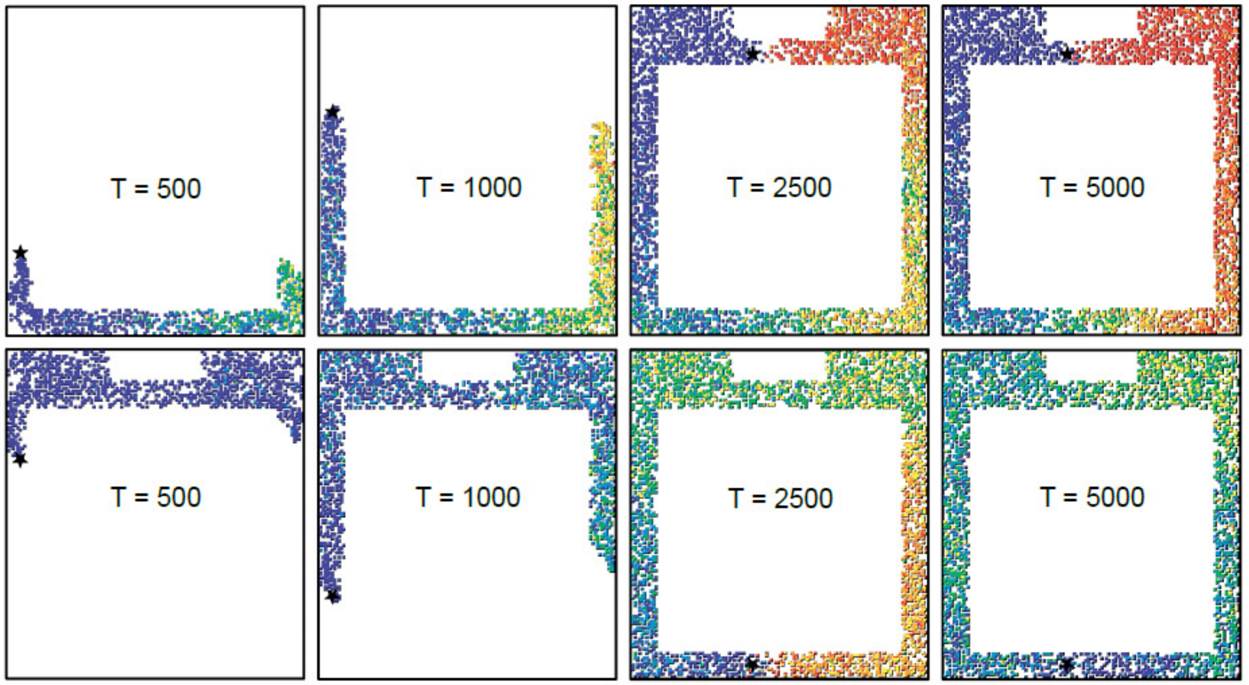

Fig. 3. Spatial snapshots showing the expansion of the simulated populations around a ring. Top and bottom lines differ only in the place of origin of the expanding population, south and north respectively. Since there are larger habitable areas in the north, the top simulation has this larger area at the point of secondary contact, and the bottom has it near the point of origin. Simulation parameters are $K = 3300$ (carrying capacity) and $D = 0.1$ (dispersal rate). Colors represent the genetic distance to a reference individual marked as a star.

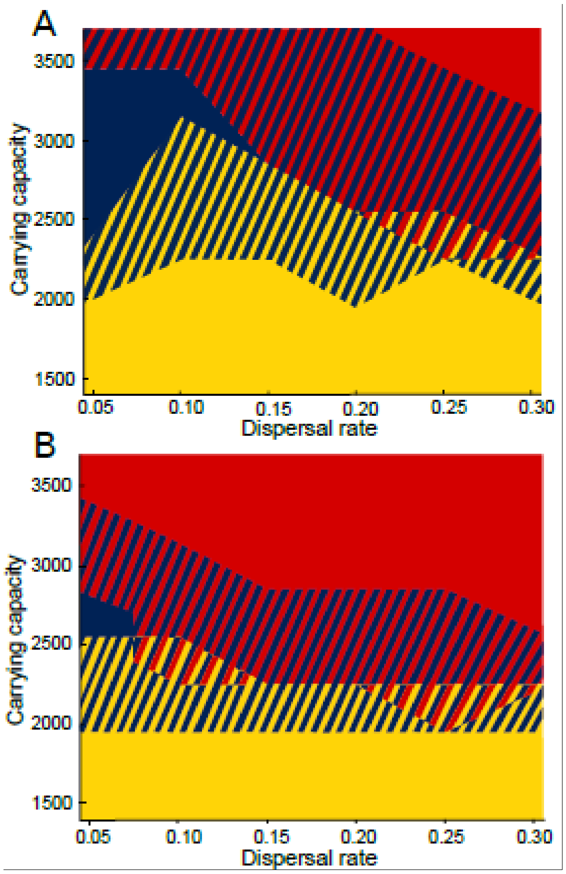

Fig. 4. Phase diagram showing ring species for different values of the dispersal rate and carrying capacity after 10,000 generations (details as in Fig. S7). The blue area represents ring species; the red area, single mixed species, and the yellow area, multiple species. These areas have been extrapolated from the results of 48 simulated points with five replicates for each. A and B differ only in the configuration of geographical barriers: increased distribution ranges near the region of secondary contact (the same as Fig. 3, top) or the region of population origin (the same as Fig. 3, bottom), respectively. In the striped areas where colors overlap different simulation replicates resulted in different outcomes.

Supplementary Materials:

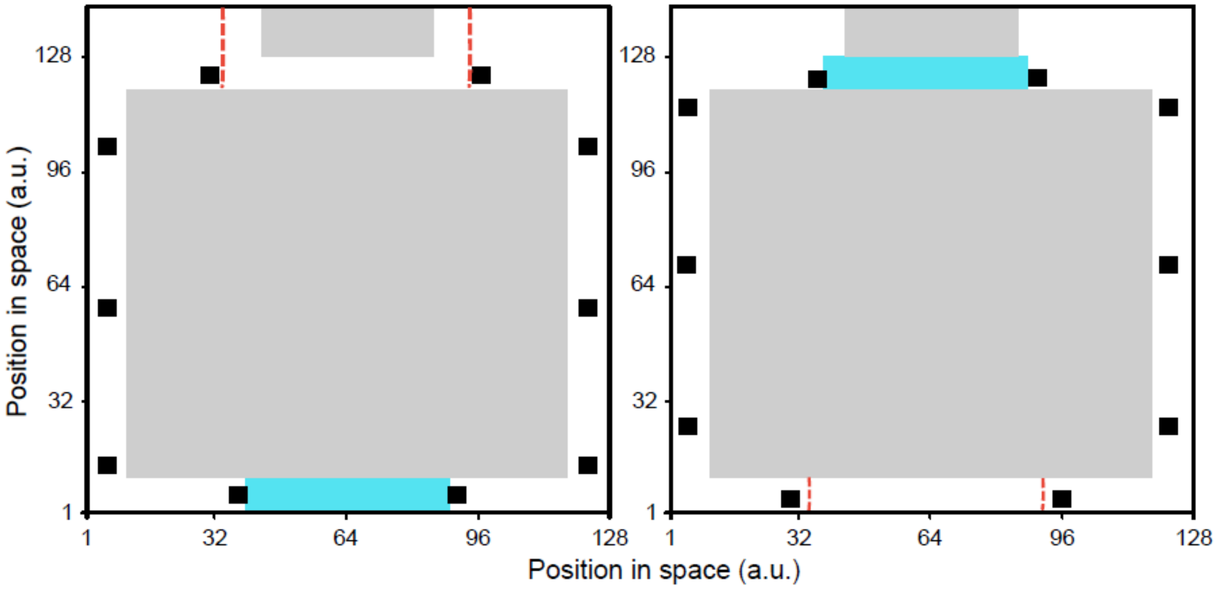

Fig. S1. Geography of the simulation. Gray sites cannot be occupied by individuals. The individuals of the initial population were randomly placed in the blue area, either at the “south” or at the “north” of the ring. The population expands toward open spaces. The configuration shown on the left has more space available in the region of secondary contact. The one on the right has more space available in the region of population origin. Simulation areas represent both physical area and population density with lower density in the north (1). The area of secondary contact, delimited by red dashed lines, has width of 60 lattice sites and is centered horizontally in the simulated space opposite to the to area where the population started. Whenever the simulated populations were sampled, individuals were taken from 10 sampling sites outside the region of secondary contact and evenly distributed around the ring (black squares).

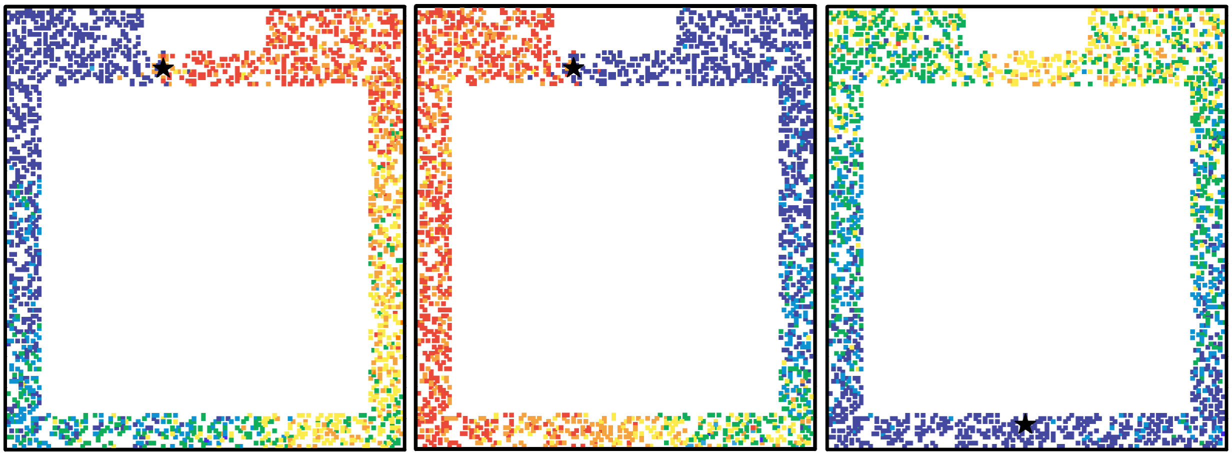

Fig. S2. Spatial snapshots showing genetic distances to different reference individuals (star) for the simulated ring species shown in Figure 1. Dark blue corresponds to potential mates for the reference individual and the most genetically differentiated individuals are shown in red. Terminal populations are far apart genetically (first and second snapshot), and are similarly differentiated relative to individuals in the area of population origin (last snapshot). Details of how ring species were detected are described in the Methods section.

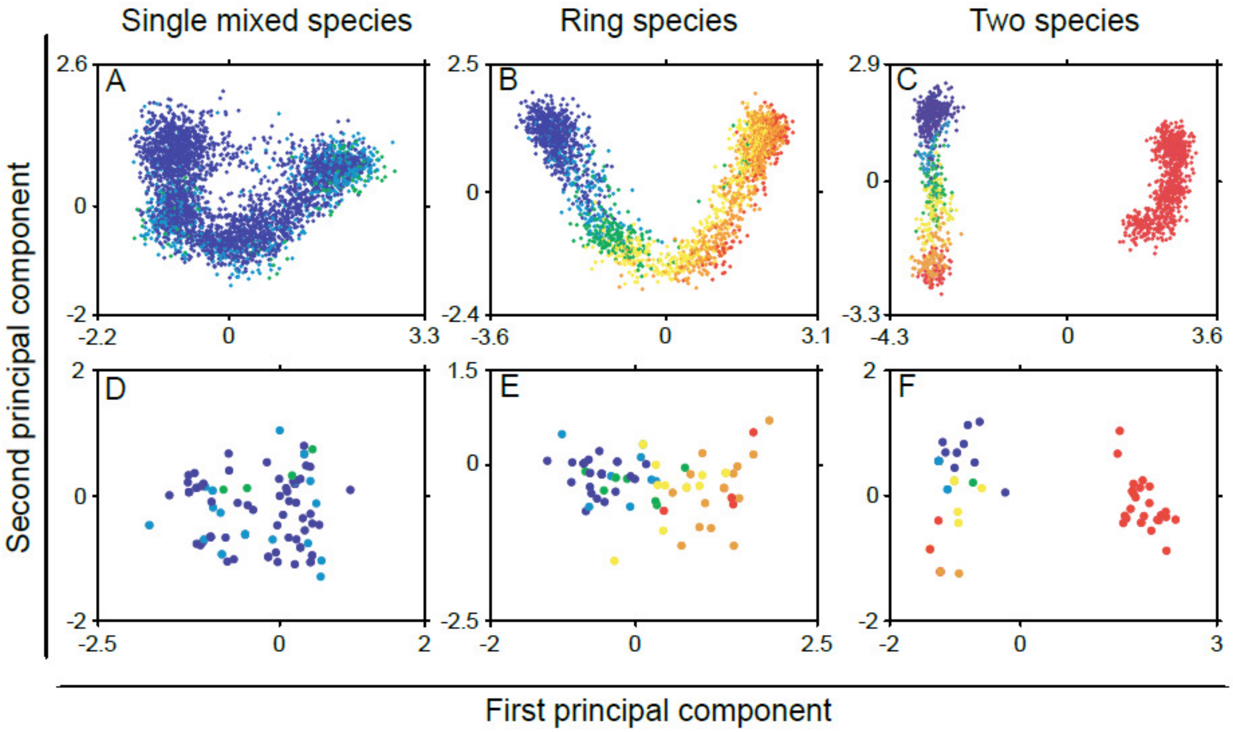

Fig. S3. Genetic data as summarized by principal components from the three simulations shown in Figure 2. The population growth rate was varied indirectly by changing the local carrying capacity. Population carrying capacity and individuals' dispersal rate affect the velocity of range expansion and determine if a ring species will be formed, if all else is equal. The top row shows results using all individuals; the bottom row shows the results from the same method using sampling of 25 loci from individuals located in the 10 sampling sites shown in Figure S1. Single species (A), ring species (B) and multiple species (C) may be clearly identified by visual inspection of principal components plots. Identification of simulated ring species using visual inspection matched those of the computer algorithm described in the Methods section in all cases tested. However, these patterns may be somewhat confounded by sampling effects in real world application (D, E and F). Quantitative evaluation of principal component variance also provide indicators: (A) single mixed species at 1000 generations (ring closed after 700 generations), PC1 (first principal component) explains 6.4% of the variance and PC2 (second principal component), 3.2%; (B) ring species at 2000 generations (ring closed after 1400 generations), PC1 explains 13.3% of the variance and PC2, 5.6% (C) two species at 3200 generations (ring closed after 2700 generations), PC1 explains 28.2% of the variance and PC2, 6.9%. After sampling: (D) single mixed species, PC1 explains 12.2% of the variance and PC2, 9.8%; (E) ring species, PC1 explains 17.4% of the variance and PC2, 8.7%; (F) two species, PC1 explains 46.1% of the variance and PC2, 10.4%.

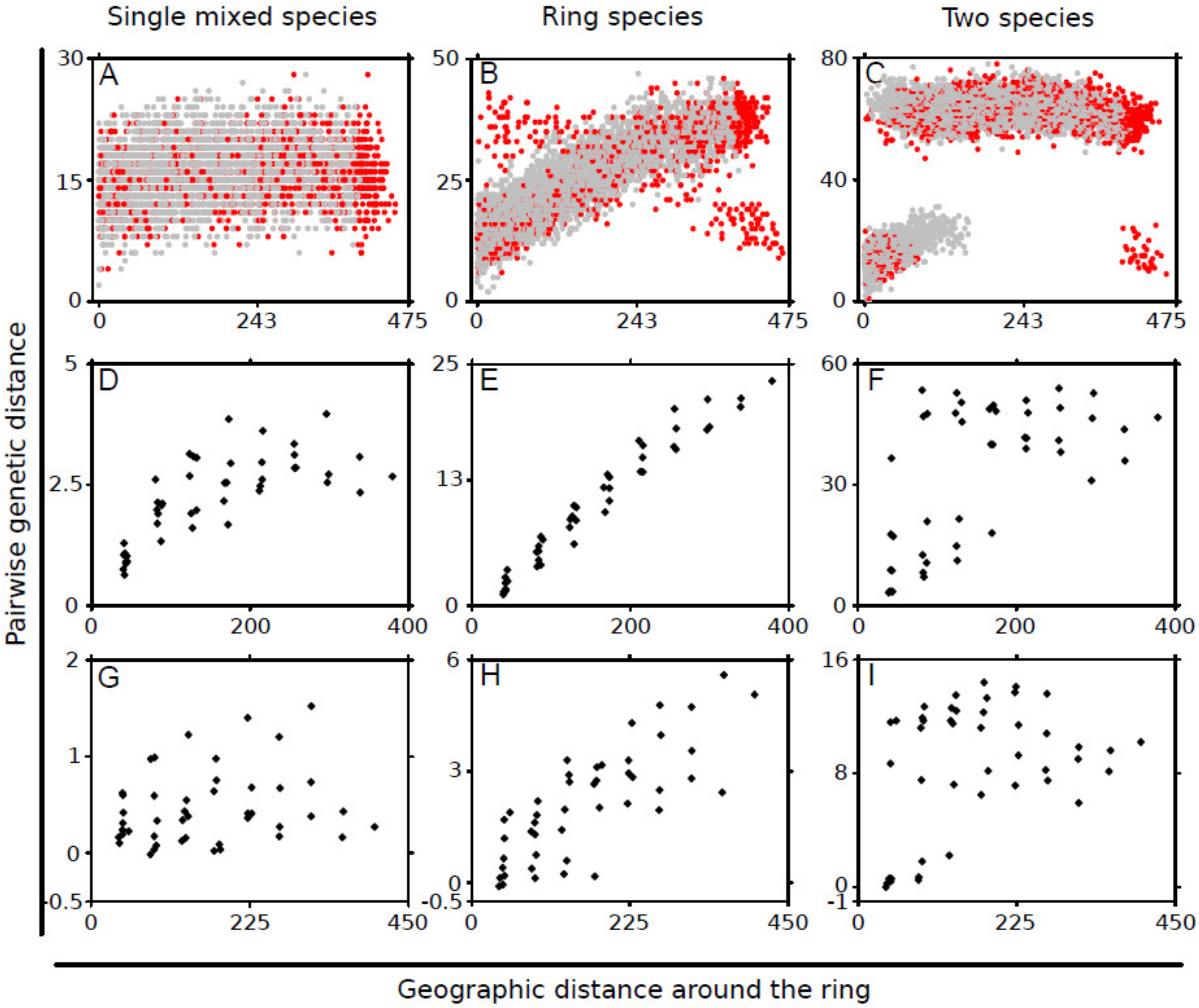

Fig. S4. Genetic distance as a function of the geographic distance measured around the ring from the three simulations shown in Figure 2 after 10,000 generations. Individual pairwise genetic distance and geographic distance measured around the ring for 5000 randomly chosen pairs (A, B and C). After secondary contact, the terminal portions of the population overlap in space. Here, pairs for which at least one of the individuals is located in the area of secondary contact are shown in red (see Fig. S1). In B and C, points appearing in the left upper quadrant and in the right lower quadrant represent pairs for which at least one of the individuals has crossed the midpoint of the area of secondary contact, inverting the expected correlation. Corrected pairwise distances were calculated as the average genetic distance between pairs of individuals in different geographical regions minus the average genetic distance between pairs of individuals within the same regions (D, E and F). The geographical space around the ring was divided in ten evenly distributed regions of the same size starting in the center of the upper corridor. The region corresponding to the area of secondary contact, where the average genetic distance between individuals is too large, was excluded. This method was also performed in a sample of 25 loci from individuals located in the 10 sampling sites shown in Figure S1 (G, H and I). The plot shown in Figure 1 (bottom, left) is the same as H here after axis rescaling.

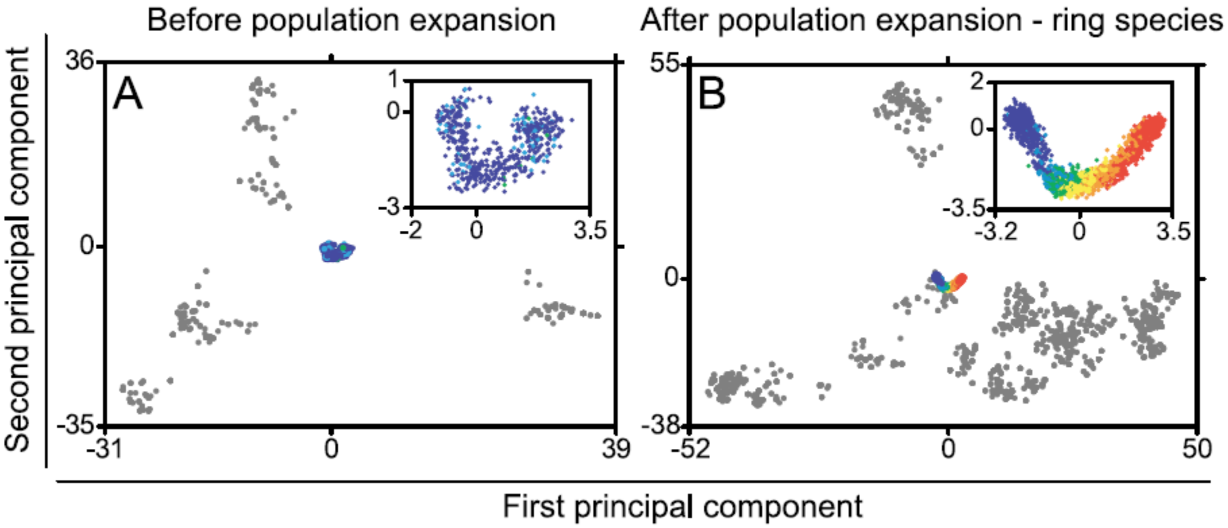

Fig. **S5**. Simulation of mitochondrial genetic divergence clock for a ring species. Genetic data are shown for a simulation with an added asexual genome having 1000 loci. The added loci are neutral, suffer no recombination and are not involved in reproductive isolation. Data, summarized by principal components, show superimposed the asexual loci (gray) and the genetic loci involved in reproductive isolation (colors and other parameters are the same as in Fig. 1). The population evolved for 1 million generations confined to the area of origin below the central barrier and then expanded for another 10,000 generations around the barrier to form the ring. Principal component analysis was performed for each genetic region and the resulting distances were indexed by the largest genetic distance in the corresponding region between a pair of individuals in the population. (A) population after 1 million generations of evolution and before the expansion. (B) 10,000 generations after ring expansion started (ring species). Insets show enlarged images of the sexually reproducing PCA.

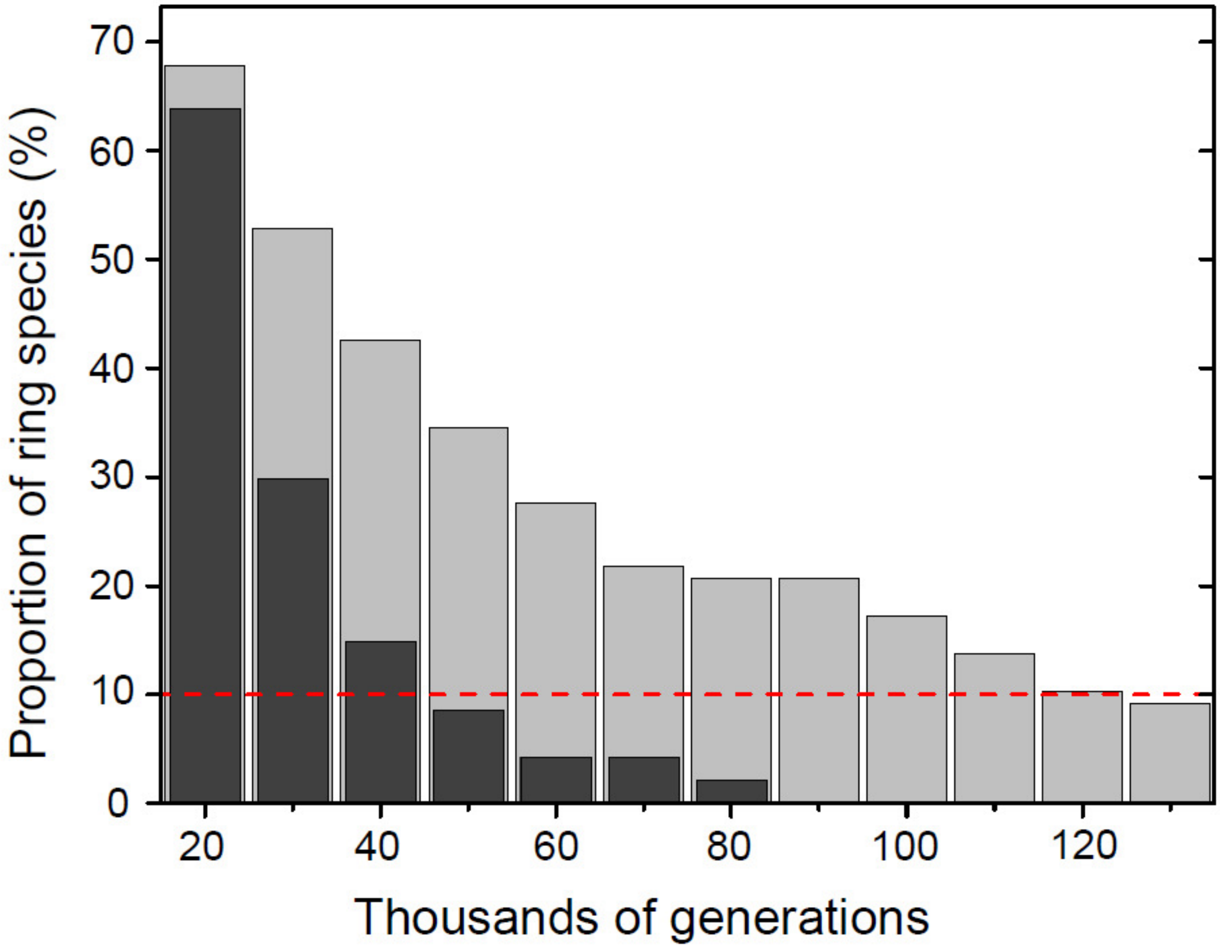

Fig. S6. Ring species detected at 10,000 generations (Fig. 4) were simulated for an additional time period. The histogram shows the proportion of those ring species remaining after additional simulated generations. Light and dark gray reflect increased distribution ranges near the region of secondary contact or the region of population origin, respectively (as shown in the barrier geometries of Fig. S1). Red dashed line indicates 10% as a guide to the eye.

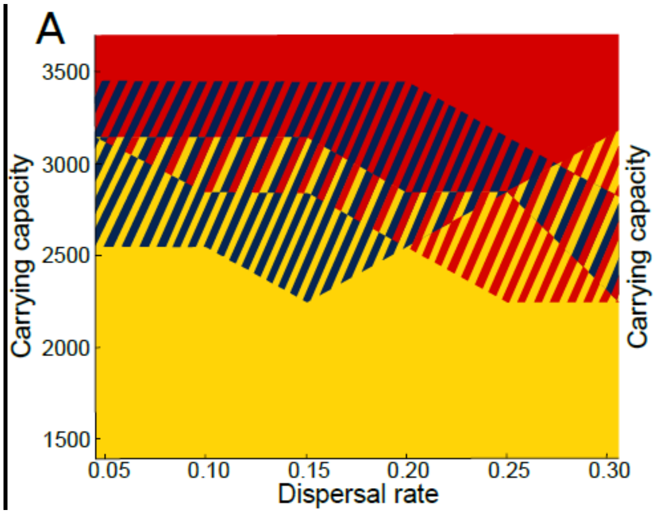

Fig. S7. Phase diagram showing ring species for different values of the dispersal rate and carrying capacity after 50,000 (left) and 100,000 (right) generations. Populations have increased ranges near the area of secondary contact (Fig. S1, left). This is the case in which ring species are more likely to form and more stable. The blue area represents ring species; the red area represents single mixed species, and the yellow area, multiple species. These areas have been extrapolated from the results of 48 evenly distributed simulated points with five replicates for each. Carrying capacity ranged from 1500 to 3600 individuals and dispersal rate ranged from 0.05 to 0.3. In the striped areas, different simulation replicates resulted in the different outcomes represented by the component colors.

References

1. Irwin DE (2000) Song variation in an avian ring species. *Evolution* 54:998–1010.