Diversification and biodiversity dynamics of hot and cold spots

Carlos J. Melián[1,*], Ole Seehausen[1],
Víctor M. Eguíluz[2], Miguel A. Fortuna[3], and Kristy Deiner[4],
[1]Fish Ecology and Evolution Department,
Center for Ecology, Evolution and Biogeochemistry,
Swiss Federal Institute of Aquatic Science and Technology, Switzerland.
[2]Instituto de FÃŋsica Interdisciplinar y Sistemas Complejos,
Spanish Research Council, Universitat de les Illes Balears,
Palma de Mallorca, Spain.
[3]Integrative Ecology Group
Estación Biológica de Doñana (EBD-CSIC), Seville, Spain
[4]Department of Aquatic Ecology,
Swiss Federal Institute of Aquatic Science and Technology, Switzerland.

Keywords: Metacommunity dynamics, landscape genetics, speciation theory, asymmetric gene flow, assortative mating, ecological drift.
Type of Article: Article
Number of figures: 2 (in color); Number of tables: 1
∗ Corresponding author: Phone: +41 58 765 2208; Fax: +41 58765 2168;
e-mail:carlos.melian@eawag.ch.

Abstract

The determinants that shape the distribution of diversity of life on Earth have been long discussed and many mechanisms underlying its formation have been proposed. Yet connecting the biogeography of hot and cold spots of diversification and current biodiversity patterns to the microevolutionary processes remains largely unexplored. Here, we combine a landscape genetics model based on demographic stochasticity with a speciation model that can be interpreted as a model of the evolution of premating incompatibility or assortative mating to map diversification rates in a spatial context. We show that landscape structure and the intensity and directionality of gene flow strongly influence the formation of hot and cold spots and its connection to patterns in species richness. Specifically, hot and cold spots form in landscapes in which gene flow is sufficiently strongly structured that the metacommunity nearly breaks up into several disconnected metacommunities. In such a landscape structure, speciation hot spots originate in the center or in the periphery of the landscape depending on whether the direction of net gene flow is from the periphery to the center or viceversa, respectively. However, for any given level of gene flow intensity, diversification rates are approximately twice higher in the center than in the periphery of the landscape. These results suggest that sinks may form diversification hot spots with higher probability than sources, in particular, those sinks surrounded by highly diversified sources in different locations of the landscape. Joining mechanistically microevolutionary and macroevolutionary processes on landscapes present many fascinating challenges and opportunities to connect the biogeography of diversification with biodiversity dynamics.

Introduction

Species richness varies strongly in geographical space (Gaston, 2000). The rates at which new species are produced and existing species go extinct too vary in space (Rabosky, 2009). Yet, it is still unclear how variation in speciation and extinction rates relates to variation in species richness although it has fascinated scientists since the birth of modern evolutionary biology (Wallace, 1855, 1876; Simpson, 1944; Dobzhansky, 1950). Species richness at any spatial scale reflects a balance between the tendency of lineages to form new species and species to go extinct (i.e., the diversification rate), and the variety of species that an environment can support (Rabosky, 2009). At a first glimpse, one might expect higher species richness where the net diversification rate is higher (i.e., a "hot" spot in terms of the number of speciation events, or a cradle of species diversity) but this is not necessarily so, because after their origin, species may change their distribution through range expansion and regional extinction, and speciation and extinction rates may vary independently of each other in space. Thus, diversity may accumulate in places with relatively low diversification rates, for instance in places with zero speciation and low extinction but high immigration (i.e., a diversification "cold" spot). Consequently, there is still considerable debate about how diversification relates in space to variation in species richness (Ricklefs, 2012; Mannion *et al.*, 2014).

By reconstructing phylogenetic trees of lineages of birds and mammals along a latitudinal gradient, a higher recent speciation rate but also higher recent extinction rates have been found at higher latitudes compared to the tropics (Weir & Schluter, 2007; Jetz *et al.*, 2012). These results are based on the average time since divergence of sister species, and intraspecific phylogroups and haplotypes, which tend to be shorter with much smaller variance in temperate regions than in tropical regions (but see (Tobias *et al.*, 2008)). Thus, these results suggest that the larger richness in the tropics in birds and mammals is due to lower extinction rates. In

a major review of the latitudinal species richness gradient, Mittelbach et al. (2007) come to the slightly different conclusion that current evidence points to the tropics as both a museum and a cradle, where taxa preferentially originate in the tropics and tend to also persist there longer (Mittelbach *et al.*, 2007). Consistent with this view, some studies have shown that the vast majority of bivalve lineages in the Sea had originated in the tropics, in spite of the strong sampling bias towards temperate zones (Jablonski *et al.*, 2006). Taken together these observations suggest that high latitudes have likely experienced more extinction and most in situ speciation events are recent, but rapid recent speciation is not uncommon in the tropics either (see also (Wagner *et al.*, 2014).)

Other studies found larger differences in diversification rates between Eastern and Western hemispheres than across latitudinal gradients (Jetz *et al.*, 2012). For instance, while cichlid fish, a group that diversifies predominantly in lakes, diversified faster in Africa than in the Neotropics (Genner *et al.*, 2007) mostly because of recent high speciation rates in African lakes, characins, another unusually species rich group of tropical freshwater fish is much more diversified in the extensive and geologically old river networks of the Neotropics and accumulated its diversity through low extinction rates without high speciation (Albert & Reis, 2011). Such observations suggest that the geographic and environmental structure of the landscape and its persistence through time play an important role in driving net diversification rate variation but that effects may vary between lineages with different ecological characteristics. Additional evidence for the importance of the geographical structure of the landscape comes from investigations of the dynamics of the latitudinal diversity gradient through geological time using fossil records. Such studies have shown that the equatorial peak and poleward decline in species richness has not been a persistent pattern but it is restricted to the past 30 million years and a few earlier intervals (Mannion *et al.*, 2014) associated perhaps with the constellation of landmasses and

climate. Given the complexity of the empirically inferred relationships between diversification rates and diversity, but also the problems associated with inferring these relationships from incompletely sampled biotas and heterogeneous taxonomic knowledge, we suggest that theoretical modeling may be helpful to infer conditions under which relationships between variation in diversification rate and species richness in space are expected or not.

Several approaches have been developed to merge the geographic variation of diversification rates to landscape structure (Thompson, 2005; Lawson, 2013). Some have emphasized the connection between the microscopic analysis of ecological systems to the emergence of macroecological patterns (see for example, an island biogeography model developed by Gilpin and Diamond in the 70s (Gilpin & Diamond, 1976); a theory to connect abundance and distribution patterns with diversification rates across a variety of taxa (Brown, 1984); a general model to connect, in addition to the structure of the landscape, landscape dynamics as a factor affecting species persistence (Keymer *et al.*, 2000); and how landscape complexity alter species diversification and sorting in mammals (Vrba, 1992)). Yet, linking microevolutionary processes that generate reproductively isolated species to the macroecological patterns in species distributions and richness remain at an incipient stage mostly because of computational complexity.

In the present study, we investigate the connections between gene flow and landscape structure with the spatial variation in diversification rates and species richness. We also study the emergence of hot and cold spots in speciation rates and how they relate to the distribution of species richness in the landscape. We define hot spots of diversification and species richness as geographic regions with high diversification rates or high species richness, respectively, while cold spots are geographic regions with low diversification rates or species richness. We ask, does the spatial structure of interconnected populations together with the intensity and directionality of gene flow influence the formation of hot and cold spots in diversification and species

richness? If so, can we make predictions about the biogeography of hot and cold spots? In the present study we take a landscape genetics approach. We implement a neutral population genetic model for speciation to explore connections between spatial genetic population structure and gene flow directionality with the emerging dynamics of hot and cold spots in diversification and how these relate to spatial variation in species richness. Such an approach wants to make contributions to understanding of how macroecological patterns of species richness may emerge simply as a consequence of microevolutionary processes that act on individuals in populations causing them to diverge, speciate and go extinct in a persistent landscape, i.e., without the need to invoke changing landscapes through time.

Our results suggest that landscape structure and the intensity and directionality of gene flow strongly influence the formation of hot and cold spots of biological diversification and its connection to patterns in species richness. Specifically, hot and cold spots of diversification form in landscapes in which gene flow is sufficiently strongly structured that the metacommunity nearly breaks up into several disconnected metacommunities. Hot spots originate in the center or in the periphery of such landscapes depending on whether the direction of net gene flow is from the periphery to the center or viceversa, respectively. However, for any given level of gene flow intensity, diversification rates are approximately twice higher in the center when the direction of gene flow is to the center than in the periphery of the landscape when the direction of gene flow is away from the center. These results suggest that under the assumptions of our models, gene flow sinks may form diversification hot spots with higher probability than sources, in particular, those sinks surrounded by many sources. On the flip side of the coin, landscapes with symmetric gene flow of any intensity, landscapes with many isolated populations or densely connected landscapes with asymmetric gene flow do not produce hot spots in our model. We discuss these results in the context of understanding the connection between

geographical variation in diversification and species richness patterns.

The model

Population genetics models of speciation and phenotypic models of speciation take into account, in most of the existing variations, three main processes: Migration dynamics, viability selection, and mating (Felsenstein, 1981; Kondrashov & Kondrashov, 1999; Dieckmann & Doebeli, 1999; Kirkpatrick & Ravigné, 2002; Van Doorn *et al.*, 2009). The order of these three processes and the complexity considered within each of these three components may have an effect on the dynamics of speciation (Felsenstein, 1981). In this study, we combine mutation, migration and drift with similarity-based mating in landscape genetics models to explore conditions under which hot and cold spots of diversification form in space. New species arise if a population becomes genetically too distant from its nearest relatives, species spread and go extinct through demographic stochasticity. In the following sections we introduce the landscape genetics models using random geometric landscapes and demographic stochasticity in three different gene flow scenarios, the speciation and extinction concepts used in this study and the method used to track diversification rate.

Random geometric landscapes

We consider landscapes consisting of randomly located sites (i.e., nodes with a spatial location given by x_i and y_i) connected by dispersal events (i.e., links). This spatial network is embedded in an area of 1000x1000 km^2 (Figure 1) with a total number of sites, $\mathcal{S}$, equal to 1000. Two sites i and j are connected by dispersal events if the Euclidean distance between them, d_{ij}, is equal or smaller than a threshold distance, d_{max}. The largest fraction of connected sites, the giant component, depends on the distance threshold considered. We explored a range of values

for the distance threshold (20 km $< d_{max} <$ 100 km) that recovers the full range of sizes for the giant component. That is, we explore landscape structure going from landscapes in which all sites are completely disconnected to landscapes in which all sites are connected in a single component (Figure 1) (Penrose, 2003).

Population dynamics and gene flow

At the beginning of the simulations in each replicate we have an initial population that spreads instantaneously across the whole landscape. We assume that all sites are fully occupied and have the same carrying capacity, i.e., population size at a given site i, N_e^i, is equal to the site environmental carrying capacity. The overall number of individuals at a given site is fixed, and the total number of individuals in the landscape is $N_e = N_e^1 + N_e^2 + N_e^3 + N_e^4$,..., $+ N_e^S$, with $\mathcal{S}$ the total number of sites.

The population evolves on the spatial network under a zero-sum birth and death process in overlapping generations. This means that at each time step an individual dies from a randomly chosen site i. This individual is replaced with a offspring coming from another site (i.e., migrant) or from the same site than the death individual. Parents are chosen with probability m from outside site i and with probability $(1 - m)$ from the site i. To explore the effect of the intensity (i.e., probability m) and directionality (i.e., center to periphery or viceversa) of gene flow on hot and cold spot formation we need to take into account the spatial location of the parents coming to site i. We considered three models differing in the directionality of gene flow: symmetric, centripetal and centrifugal gene flow (Brown, 1957; Lemmon & Lemmon, 2008). In the symmetric model parents are chosen from the subset of j sites connected to the site i that satisfy the condition $d_{ij} \leq d_{max}$. In the centripetal gene flow scenario, parents are chosen from the j sites connected to the site i that satisfy the condition

$d_{ij} \leq d_{max}$ as in the symmetric scenario, but now we only choose those j sites that are more in the periphery than site i. Peripheral sites are those sites whose total sum of geographic pairwise distance is equal or larger than for site i. In the centrifugal gene flow scenario, parents are chosen from the j sites connected to the site i that satisfy the condition $d_{ij} \leq d_{max}$ as in the previous scenarios, but now we only choose those j sites that are more in the center than site i. Center sites are those sites whose total sum of geographic pairwise distance is equal or lower than for site i.

Model 1: Symmetric gene flow

In the simplest scenario for those i and j connected populations (i.e., geographic distance lower or equal than the radius, $d_{ij} \leq d_{max}$), we consider gene flow of species k is only a function of the geographic distance between site i and j. Gene flow between site j and site i is

$$m_{ij} = \frac{m}{d_{ij}}, \tag{1}$$

where d_{ij} is the geographical distance between site i and j, and m determines the intensity of gene flow.

Model 2: Centripetal gene flow

Under this scenario, gene flow from site j to site i depends not only on the distance between the two sites, d_{ij}, but parents are chosen preferentially from the periphery. Gene flow from site j to site i is

$$m_{ij} = \begin{cases} \frac{m}{d_{ij}} & \text{if } \sum_{k=1}^{\mathcal{S}} d_{ik} \leq \sum_{k=1}^{\mathcal{S}} d_{jk}, \\ 0 & \text{if } \sum_{k=1}^{\mathcal{S}} d_{ik} > \sum_{k=1}^{\mathcal{S}} d_{jk} \end{cases} \tag{2}$$

Formula (2) assumes that dispersal is directional from sites with higher total geographic distance (where total geographic distance is defined as the sum over all distances between the focal site and all other sites) to sites with a lower total geographic distance. If core sites are defined as

those that have lower total geographic distance than peripheral sites, then formula (2) implies that offspring move from the periphery to the center of the landscape. (We remark there is no net flow of individuals from the periphery to the center because the number of individuals in each site remains constant.)

Model 3: Centrifugal gene flow

This is the opposite scenario than the above one: parents are now chosen preferentially from the center. Gene flow from site j to site i is

$$m_{ij} = \begin{cases} \frac{m}{d_{ij}} & \text{if } \sum_{k=1}^{\mathcal{S}} d_{ik} \geq \sum_{k=1}^{\mathcal{S}} d_{jk}, \\ 0 & \text{if } \sum_{k=1}^{\mathcal{S}} d_{ik} < \sum_{k=1}^{\mathcal{S}} d_{jk} \end{cases} \tag{3}$$

In this case, offspring move preferentially from the center to the periphery of the landscape.

Species concept

Each individual in our model contains a haploid genome consisting of an infinite string of nucleotides. All genomes are identical at the outset and each offspring genotype is produced from freely recombined parental haplotypes. This means there is no standing genetic variation at the start of each run. The genetic similarity between two individuals k and l, q_{kl}, is defined as the proportion of identical nucleotides along the genome. The genetic similarity matrix, $\mathcal{Q} = [q_{kl}]$, contains all the pairwise similarity values, q_{kl}. In our models, this genetic similarity matrix evolves as a consequence of mutation, mating and free recombination that produces a haploid offspring that differs from both parents (Higgs & Derrida, 1992; Melián *et al.*, 2010).

We define a species as a group of individuals with the ability of interbreeding that are reproductively isolated from other groups (Nei *et al.*, 1983). This means any two individuals can mate and have fertile offspring if and only if their genetic similarity value is larger or equal to the minimum value, q_{min}. Thus, for replacing a dead individual in site i parents k and l with

larger genetic similarity than the threshold, $q_{kl} \geq q_{min}$, are chosen with probability m from any other j connected site, $d_{ij} \leq d_{max}$, and with probability $(1 - m)$ from the site i. Mating pairs that fulfill the condition $q_{kl} \geq q_{min}$ have identical fitness, hence there is no selection on mating compatibility in our model. Mating pairs that do not fulfill the condition $q_{kl} \geq q_{min}$ do not form, hence there is assortative mating determined by the genetic similarity threshold.

However, the ability of interbreeding does not imply that all conspecific individuals can have fertile offspring between themselves. If we link every pair of individuals whose genetic similarity is larger than the minimum value q_{min} (i.e., population graph), then two individuals connected at least by one pathway through the population graph are considered conspecific, even if the genetic similarity among them is smaller than q_{min}. Speciation occurs in our model when all individuals with intermediate genotypes die. Values of q_{min} ranged from 0.93 to 0.96 in our simulations. We note this definition of species and speciation through compatibility-based assortative mating is not directly applicable to ecological speciation.

Computing genetic similarity matrix to detect speciation events

The genetic similarity matrix, $\mathcal{Q} = [q_{kl}]$, containing all the pairwise genetic similarity values, q_{kl}, has a size for the entire population of $10^6 x 10^6$ (1000 sites with 1000 individuals each). This matrix is symmetric ($q_{kl} = q_{lk}$). We store it as an adjacency list. The adjacency list describes for each individual i the genetic similarity value with all individuals and the sites where these individuals are. At each time step we update this adjacency list as a consequence of changes in the genetic similarity values between the offspring and all the individuals in the population following the method described in (Melián *et al.*, 2010). For each replicate, and after 10000 generations (1 generation is 1000000 time steps) we start to count the components each 10 generations following a depth-first search algorithm (Gabow, 2000). New speciation

events (i.e., components in the adjacency list) are then localized in space and the centroid of the distribution of the new species calculated. This was done for 100000 generations for each replicate and for a total of 100 replicates (see "Simulations" below). Computation time speed up within each replicate with the number of generations because the adjacency list will be decreasing in size due that some pairwise genetic similarity values $q_{kl} < q_{min}$ (zero values in the adjacency list.)

Spatial location of hot and cold spots

In order to compute the diversification rate, we track the spatial location of each speciation and extinction event. This information can be plotted after several speciation and extinction events and so we can map where the hot and cold spots in speciation and extinction events are. Speciation events occur when a population graph split into two or more distinct components (i.e., species). On the contrary, extinction events occur when a population graph disappears because the last individual has died. The spatial location of each speciation event was calculated as the mean geographic distance of all the sites N containing at least one individual that belong to the new species k. This center of gravity of sites containing the new species is given by

$$x = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{4}$$

$$y = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{5}$$

and the speciation rate for each spatial location is given by

$$\lambda_{x,y} = \frac{\#sp_{x,y}}{G}, \tag{6}$$

where G is the number of generations.

The site that harbored the last living individual of a given species is considered the spatial location of an extinction event. The extinction rate is then calculated as

$$\mu_{x,y} = \frac{\#ex_{x,y}}{G}, \tag{7}$$

and the diversification rate at a given spatial location, $\Omega_{x,y}$, can then be calculated as

$$\Omega_{x,y} = \lambda_{x,y} - \mu_{x,y}. \tag{8}$$

Species richness

Mapping each speciation and extinction event in the landscape allow us to track the number of extant species. Thus, as for the diversification rate, we count the number of extant species across generations for each replicate and plot this number as a function of the distance to the center of the landscape.

Simulations

Our simulation is a stochastic, individual-based, zero-sum birth-death model linking gradual genetic changes in populations with overlapping generations (microevolutionary processes) with speciation events based on compatibility-based assortative mating driven by the genetic similarity threshold (macroevolutionary processes). Specifically, we simulated the ecological and evolutionary dynamics of populations of haploid individuals located at discrete sites randomly distributed on a spatial network. Simulations were carried out with an initial population at each site i, N_e^i, of 1000 individuals for a total of 1000 sites. The population size and the number of sites remained constant throughout the simulations.

Results for Figure 2 and 3 were obtained after 100 replicates and 100000 generations of a single model run, where a generation is an update of the total number of individuals in the

landscape. Hot and cold spots plotted represent the mean values of 100000 slices ranging from the center to the periphery of the network. We explored a broad range of landscapes using the maximum distance to connect two sites and to determine the mating pool, 20 km $< d_{max} <$ 100 km. Population genetics parameter combinations explored: mutation rate, $\mu \in (3\text{x}10^{-4}, 10^{-5})$, the intensity of gene flow, $m \in (0.3, 0.001)$, and the cut-off values to count speciation events and species richness in the transient and equilibrium dynamics represented as the minimum genetic similarity value to define a species, $q_{min} \in (0.96, 0.93)$. Total number of species reached a mean value after approximately 75000-90000 generations (variation due to the different initial parameter values). At that stage, speciation rate equals extinction rate and all hot spots disappear.

Results

In order to connect the dynamics of speciation hot and cold spots formation to biodiversity patterns, we first generate a range of landscapes, from completely connected (Figure 1A represents a landscape with sites i and j connected if $d_{ij} \leq d_{max} = 75$km) to completely disconnected sites (Figure 1C, $d_{max} = 25$km). Between these two landscape structures we have a clustered landscape with groups of interconnected sites (Figure 1B, $d_{max} = 50$km).

Diversification hot spots are possible only when the spatial network is near the percolation threshold. This means hot spot formation occurred in landscapes that were nearly broken up into several disconnected metacommunities and only with centripetal or centrifugal gene flow. Figure 1B can illustrate the idea of a metacommunity as several sites sparsely connected across the whole spatial network. Thus, each aggregation of sparsely connected sites can be defined as a metacommunity. Centripetal gene flow model produces higher diversification rate in the center than in the periphery of the landscape (Figure 2A, dotted red line with gene flow, m

= [0.1, 0.3]). Centrifugal gene flow model produces higher diversification rate in the periphery than in the center of the landscape (Figure 2A, continuous red line with gene flow, m = [0.1, 0.3]). Diversification rate decreases with low gene flow in both scenarios (Figure 2A, dotted and solid black line with gene flow, $m < 0.1$.)

How do speciation and extinction rate and how do diversification rate and species richness correspond in the landscape? We found speciation and extinction rate spatially corresponding in the landscape (Figure 2B, centripetal, and 2C, centrifugal). Extinction rate peaks in the hot spots of diversification but it does not increase as much as the speciation rate (Figure 2B and 2C). Approximately twice higher extinction rates were observed where the hot spots formed in the centripetal gene flow model (compare blue line in Figure 2B for the centripetal with the Figure 2C for the centrifugal gene flow model). Diversification rate and species richness do also spatially correspond in the landscape (Figure 3). The centripetal gene flow model produced hot spots and high species richness in the center (Figure 2A-B and 3). The centrifugal gene flow model produced hot spots and high species richness in the periphery (Figure 2A-C and 3). Centripetal gene flow predicts approximately twice higher speciation rate, diversification rate and species richness compared to the centrifugal gene flow model (Figure 2A, diversification rate approx. 0.004 vs. 0.002 and Figure 3, species richness 150 vs. 65). Centripetal gene flow model also produced higher γ-species richness than the centrifugal gene flow model, 206$\pm$49 vs. 94$\pm$35.

In summary, our predictions suggest there is an "asymmetry" in the intensity of the hot spots formation. Centripetal gene flow model consistently predicted higher speciation and extinction rate, diversification rate and species richness than the centrifugal gene flow model (Figure 2 and 3). This "asymmetry" in the intensity of the hot spot formation has consequences for biodiversity: the number of species accumulating in the center is consistently higher than

in the periphery. This result indicates that those sectors of the network receiving migrants (i.e., sinks) may form diversification hot spots, and thus acquire higher species richness, than those areas in the network delivering migrants (i.e., sources). This seems particularly true in those sinks close to the center of the species distributions that are surrounded by many sources connected to highly differentiated areas in the landscape.

Discussion

Our study adds to previous attempts to connect microevolutionary dynamics and macroevolutionary patterns in a mechanistic framework (Gavrilets, 2004; Gavrilets & Vose, 2007; de Aguiar *et al.*, 2009; Kopp, 2010; Melián *et al.*, 2010; Rosindell & Phillimore, 2011; Davies *et al.*, 2011). Our goal was to understand patterns in the geography of the hot and cold spots in diversification rate and their connection to geographical variation in species richness. Our analysis showed that the origin of speciation hot and cold spots required directional gene flow in landscapes connected to an extend that they nearly broke up into several disconnected metacommunities (Figure 1B). Centripetal gene flow dynamics with net gene flow from the periphery to the center generated higher diversification rates in the hot spots compared to the centrifugal gene flow model in which net gene flow is from the center to the periphery (Figure 2). Our results showed that the formation of speciation hot spots was inhibited in densely connected landscapes, as well as in landscapes with most communities isolated and in landscapes with symmetric gene flow patterns regardless of the degree of isolation. Our results also suggest that speciation and extinction rate variation (Figure 2B and 2C) together with diversification rate variation (Figure 2A) and standing amounts of species richness (Figure 3) have correspondence in the landscape.

In the following we discuss our model results in the context of island biogeography theory and macroecological species richness gradients. Previous studies suggest that the geographic and

environmental structure of the landscape and its persistence through time play an important role in driving net diversification rate variation (Ricklefs, 2012; Mannion *et al.*, 2014). Island biogeography theory with speciation has been applied to two very different island scenarios: oceanic islands and lakes. Whereas oceanic islands and island archipelagoes tend to be strongly isolated from the nearest mainland with gene flow restricted to rare colonization events, large lakes are often connected to one or several rivers and receive continued gene flow from the rivers. Our models suggest that the geometry of the island network combined with the directionality of gene flow strongly determine whether and where hot spots in diversification and in the corresponding species richness arise. For example, our models predict lower speciation rates when islands are in the periphery of the network of habitat patches than when islands are central to such networks. Our results indicate that diversification hotspots do not form when islands are disconnected from the source sites (i.e., isolated with low gene flow), whereas our models predict hot spots of diversification in islands that are connected to the sources by moderate levels of gene flow. By analogy this may predict larger rates of diversification and larger species richness in continental lakes that are connected to several streams providing colonists and continued gene flow, than in oceanic islands that are in the periphery of the network of habitat islands. More generally these results suggest that the landscape structure combined with the directionality of gene flow may play a critical role in determining whether and where a radiation zone or a hot spot of diversification arises (Gillespie *et al.*, 2008; Rosindell & Phillimore, 2011).

How do hot and cold spot formation relate to sources and sinks and species richness? Our results suggest that hot spots occurred with higher probability in those fractions of the network receiving migrants from a highly diversified set of sources and that these sinks also accumulated higher species richness than the areas in the network that delivered migrants (i.e., the sources). This suggests distinct geographic patterns in the relationship between species richness and the

diversification rate may arise (Figure 2 and 3) in the absence of spatial variation in carrying capacities across the landscape. Extending these models by explicitly modeling sources and sinks with differential growth rates (or carrying capacities) and productivity across the landscape may show even more distinct geographic patterns of diversification rate and species richness. Recent investigations of the fossil record have shown that the tropical peak and poleward decline in species richness has not been a persistent pattern throughout the Phanerozoic, but is restricted to intervals of the Palaeozoic and the past 30 million years (Mannion *et al.*, 2014). Our models can also be extended to incorporate climatic regimes and landscape dynamics to provide a dynamic system in which to explore spatiotemporal diversity fluctuations. These extended models can tell us how much complexity is required to make predictions that match periods of peaks or flattened species richness gradients as observed in the fossil record.

We contrasted biodiversity dynamics in models with completely symmetric (Fig. 2A blue line) versus completely asymmetric (Fig. 2A red lines) migration and gene flow. Asymmetric migration patterns are common in nature. Dendritic river networks, inclusive of lakes embedded into them, are a classical example where migration and gene flow in the downstream direction tends to be much stronger than in the opposite direction. Dispersal in the Sea is another classical example where ocean currents determine the dominant direction (Bonfil *et al.*, 2005), and rare long distance dispersal of birds across oceans may often be predictable by the predominant direction of winds (Thorup *et al.*, 2007). The detection of speciation events in space using very similar models have shown that location of speciation in a symmetric migration model strongly depends on the level of genetic variation (Gavrilets *et al.*, 2000). In the future, our modeling framework should be extended to situations that combine symmetric and asymmetric gene flow to explore the dependency of the emergence of diversification hotspots on the extent of deviation from symmetry and genetic variation. In our study, we contrasted dynamics with

361 migration rates of approx. 0.1 (Fig. 2A black lines) with those when migration rates were
362 larger than 0.1 but smaller than 0.3 (Fig. 2A red lines). Another interesting extension would
363 be to simulate a larger range of gene flow values and investigate the shape of the relationship
364 between migration rate, landscape structure and the intensity of diversification hotspots.

365 How do the availability of space or resources and the strength and direction of selection
366 change the results presented in this study about the hot and cold spot formation? Diversification
367 may depend, at least in part, on the availability of space and resources, landscape structure
368 and gene flow among populations, the architecture of the genotypes and the strength and
369 form of selection (Gavrilets & Losos, 2009). Our approach did not explicitly test for sources
370 and sinks because we assumed equal growth rates across the landscape, nor did we assume
371 any asymmetry in competition or trophic interactions as possible mechanisms for structuring
372 diversity in our landscapes, hence a neutral theory of biodiversity was applied. Our approach
373 instead takes into account assortative mating, landscape structure and the directionality of gene
374 flow as the main drivers of speciation events across the landscape (Welch, 2004). Our models
375 cannot capture the microevolutionary dynamics associated with ecological speciation fueled by
376 adaptations to niches (Rundle & Nosil, 2005), or speciation driven by sexual selection (Maan
377 & Seehausen, 2011). Specifically, we have considered infinite genomes with all genes having a
378 weak effect on divergence. It would be interesting to contrast our results from results assuming
379 finite genomes with a small number of genes having a strong effect, where speciation would
380 also involve isolation by distance. Extending the models introduced in this study, by explicitly
381 considering habitat, sexual selection or finite genomes may alter the geography of hot and cold
382 spots presented in new and unexpected ways. Nonetheless the models explored here may be
383 useful as a benchmark to compare their predictions in the geographic variation of hot and cold
384 spots of diversification to biologically more realistic scenarios.

Acknowledgments

We thank Alejandro Rozenfeld and Rebecca Best for valuable suggestions which greatly improved the manuscript. CJM was supported by the Swiss National Science Foundation project 31003A-144162. MAF was supported by a postdoctoral fellowship (JAE-Doc) from the Program "Junta para la Ampliacion de Estudios" co-funded by the Fondo Social Europeo.

References

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

Albert, J.S. & Reis, E.R. (2011). *Historical Biogeography of neotropical freshwater fishes*. University of California Press. Berkeley, USA.

Bonfil, R., Meÿer, M., C., S.M., Johnson, R., O'Brien, S., Oosthuizen, H., Swanson, S., Kotze, D. & Paterson, M. (2005). Transoceanic migration, spatial dynamics, and population linkages of white sharks. *Science*, 310, 100–103.

Brown, J.H. (1984). On the relationship between abundance and distribution of species. *The American Naturalist*, 124, 255–279.

Brown, W.L. (1957). Centrifugal speciation. *Quaterly Review of Biology*, 32, 247–277.

Davies, J.T., Allen, D.P., Borda de Água, L., Regetz, J. & Melián, C.J. (2011). Neutral biodiversity theory can explain the imbalance of phylogenetic trees but not the tempo of their diversification. *Evolution*, 65, 1841–1850.

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

Dobzhansky, T. (1950). Evolution in the tropics. *American Scientists*, 38, 209–221.

Felsenstein, J. (1981). Skepticism towards santa rosalia, or why are there so few kinds of animals? *Evolution*, 35, 124–138.

Gabow, H.N. (2000). Path-based depth-first search for strong and biconnected components. *Information Processing Letter*, 74, 107–114.

411 Gaston, K.J. (2000). Global patterns in biodiversity. *Nature*, 405, 220–227.

412 Gavrilets, S. (2004). *Fitness Landscapes and the Origin of Species*. (Princeton University Press,
413 Princeton).

414 Gavrilets, S., Li, H. & Vose, M.D. (2000). Patterns of parapatric speciation. *Evolution*, 54,
415 1126–1134.

416 Gavrilets, S. & Losos, J.B. (2009). Adaptive radiation: Contrasting theory with data. *Science*,
417 323, 732–737.

418 Gavrilets, S. & Vose, A. (2007). Case studies and mathematical models of ecological speciation.
419 2. palms on an oceanic island. *Molecular Ecology*, 16, 2910–2921.

420 Genner, M.J., Seehausen, O., Lunt, D.H., Joyce, D.A., Shaw, P.W., Carvalho, G.R. & Turner,
421 G.F. (2007). Age of cichlids: New dates for ancient lake fish radiations. *Mol. Biol. Evol.*, 24,
422 1269–1282.

423 Gillespie, R., Claridge, E.M. & Goodacre, S. (2008). Biogeography of French Polynesia: Diver-
424 sification within and between a series of hotspot archipelagoes. *Phil. Trans. Roy. Soc. Lond.*,
425 363, 3335–3346.

426 Gilpin, M.E. & Diamond, J.M. (1976). Calculation of immigration and extinction curves from
427 the species-area-distance relation. *Proceedings of the National Academy of the Sciences of*
428 *the USA*, 73, 4130–4134.

429 Higgs, P.G. & Derrida, B. (1992). Genetic distance and species formation in evolving popula-
430 tions. *Journal of Molecular Evolution*, 35, 454–465.

Jablonski, D., Roy, K. & Valentine, J.W. (2006). Out of the tropics: Evolutionary dynamics of the latitudinal diversity gradient. *Science*, 314, 102–106.

Jetz, W., Thomas, G.H., Joy, J.B., Hartmann, K. & Mooers, A.O. (2012). The global diversity of birds in space and time. *Nature*, 491, 444–448.

Keymer, J.E., Marquet, P.A., Velasco-Hernández, J.X. & Levin, S.A. (2000). Extinction thresholds and metapopulation persistence in dynamic landscapes. *The American Naturalist*, 156, 478–494.

Kirkpatrick, M. & Ravigné, V. (2002). Speciation by natural and sexual selection: models and experiments. *The American Naturalist*, 159, S22–S35.

Kondrashov, A.S. & Kondrashov, F.A. (1999). Interactions among quantitative traits in the course of sympatric speciation. *Nature*, 400, 351–354.

Kopp, M. (2010). Speciation and the neutral theory of biodiversity. *Bioessays*, 32, 564–570.

Lawson, L.P. (2013). Diversification in a biodiversity hot spot: landscape correlates of phylogeographic patterns in the african spotted reed frog. *Molecular Ecology*, 22, 1947–1960.

Lemmon, A.R. & Lemmon, E.M. (2008). A likelihood framework for estimating phylogeographic history on a continuous landscape. *Syst. Biol*, 57, 544–561.

Maan, M.E. & Seehausen, O. (2011). Ecology, sexual selection and speciation. *Ecology Letters*, 14, 591–602.

Mannion, P.D., Upchurch, P., Benson, R.B.J. & Goswami, A. (2014). The latitudinal biodiversity gradient through deep time. *Trends in Ecology and Evolution (http://dx.doi.org/10.1016/j.tree.2013.09.012).*

Melián, C.J., Alonso, D., Vázquez, D.P., Regetz, J. & Allesina, S. (2010). Frequency-dependent selection predicts patterns of radiations and biodiversity. *PLoS Comput Biol*, 6, e1000892.

Mittelbach, G.G., Schemske, D.W., Cornell, H.V., Allen, A.P., Brown, J.M., Bush, M.B., Harrison, S.P., Hurlbert, A.H., Knowlton, N., Lessios, H.A., McCain, C.M., McCune, A.R., McDade, L.A., McPeek, M.A., Near, T.J., Price, T.D., Ricklefs, R.E., Roy, K., Sax, D.V., Schluter, D., Sobel, J.M. & Turelli, M. (2007). Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters*, 10, 315–331.

Nei, M., Maruyama, T. & Wu, C.I. (1983). Models of evolution of reproductive isolation. *Genetics*, 103, 557–579.

Penrose, M.D. (2003). *Random Geometric Graphs*. Oxford Univ. Press, Oxford, UK.

Rabosky, D.L. (2009). Ecological limits and diversification: alternative paradigms to explain the variation in species richness among clades and regions. *Ecology Letters*, 12, 735–743.

Ricklefs, R.E. (2012). Disconnects in diversity. comment on "the global diversity of birds in space and time". *Nature*, 491, 1–2.

Rosindell, J. & Phillimore, A.B. (2011). A unified model of island biogeography sheds light on the zone of radiation. *Ecology Letters*, 14, 552–560.

Rundle, H. & Nosil, P. (2005). Ecological speciation. *Ecology Letters*, 8, 336–352.

Simpson, G.G. (1944). *Tempo and mode in evolution*. Columbia Univ. Press, New York.

Thompson, J.N. (2005). *The Geographic Mosaic of Coevolution*. (University of Chicago Press, Chicago).

Thorup, K., Bisson, I.A., Bowlin, M.S., Holland, R.A., Wingfield, J.C., Ramenofsky, M. & Wikelski, M. (2007). Evidence for a navigational map stretching across the continental u.s. in a migratory songbird. *Proceedings of the National Academy of Sciences of the U. S. A.*, 104, 18115–18119.

Tobias, J.A., Bates, J.M., Hackett, S. & Seddon, N. (2008). Comment on âĂIJthe latitudinal gradient in recent speciation and extinction rates of birds and mammalsâĂİ. *Science*, 319, 901c.

Van Doorn, G.S., Edelaar, P. & Weissing, F.J. (2009). On the origin of species by natural and sexual selection. *Science*, 326, 1704–1707.

Vrba, E.S. (1992). Mammals as a key to evolutionary theory. *Journal of Mammalogy*, 73, 1–28.

Wagner, C.E., Harmon, L.J. & Seehausen, O. (2014). Cichlid species-area relationships are shaped by adaptive radiations that scale with area. *Ecology Letters*.

Wallace, A.R. (1855). On the law which has regulated the introduction of new species. *Annals and Magazine of Natural History*, 16, 184.

Wallace, A.R. (1876). *The Geographical Distribution of Animals: with a study of the relations of living and extinct faunas as elucidating the past changes of the Earth's surface*. Macmillan Co. London, UK.

Weir, J.T. & Schluter, D. (2007). The latitudinal gradient in recent speciation and extinction rates of birds and mammals. *Science*, 315, 1574–1576.

Welch, J.J. (2004). Accumulating Dobzhansky-Muller incompatibilities: Reconciling theory and data. *Evolution*, 58, 1145–1156.

Figure caption

Figure 1. Random geometric landscapes as geographic template for population dynamics. By using a distance threshold d_{max} to connect 1000 randomly located sites across a landscape of 1000×1000 km^2, we obtained spatial networks with different levels of connectivity. Three of them are shown here and also an speciation event in **B**. **A**), Depicts a densely connected network generated using $d_{max} = 75$ km. **B**), A spatial network near the percolation threshold ($d_{max} = 50$ km). The percolation threshold is a connectivity level (given by a minimum threshold distance d_{max}) beyond which all sites that are connected in a single component, called giant component, start to organizing into many smaller disconnected components. In **B** we zoom out an speciation event. The individual with the largest node size dies forming a new "red" species within the site. An speciation event could also occur from an individual dying outside the focal site. **C** , Depicts a fragmented network generated using $d_{max} = 25$ km.

Figure 2. Diversification hot and cold spots. **A**), Diversification rate (Ω, y-axis) as a function of the distance to the center (x-axis) of a 1000×1000 km^2 landscape. Centripetal gene flow model produces higher diversification rate in the center than in the periphery of the landscape (dotted red line represents the mean values after sampling the intensity of gene flow, m, from a uniform distribution with range $\mathcal{U}[0.1, 0.3]$, see inset). Centrifugal gene flow model produces higher diversification rate in the periphery than in the center of the landscape (continuous red line represents the mean values after sampling the intensity of gene flow, m, from a uniform distribution with range $\mathcal{U}[0.1, 0.3]$, see inset). Diversification rate decreases with low gene flow in both scenarios (dotted and solid black line with gene flow, $m < 0.1$.) The maximum distance used in those plots to connect two sites and to determine the mating pool was 35km $< d_{max} <$ 55km. Symmetric gene flow model with all sites isolated does not produce hot spots (Blue line and landscape in **E** show the results for a maximum distance to connect

two sites and to determine the mating pool between, $d_{max} < 25$km.) **B**), Speciation (λ, red line, y-axis) and extinction (μ, blue line, y-axis) rate for the centripetal gene flow model with gene flow, $m = [0.1, 0.3]$. Diversification hot spot is produced in the center of the landscape (**D**). **C**), Speciation (λ, red line, y-axis) and extinction (μ, blue line, y-axis) rate for the centrifugal gene flow model with gene flow, $m = \mathcal{U}[0.1, 0.3]$. Diversification hot spot is produced in the periphery of the landscape (**F**). Plotted lines represent the mean values of 10^5 sections ranging from the center to the periphery of the network.

Figure 3. Species richness in hot and cold spots. Species richness as a function of its distance to the center of a 1000×1000 km^2 landscape. The centripetal gene flow model produces higher richness (dotted black line, represents the mean values after sampling the intensity of gene flow, m, from a uniform distribution with range $\mathcal{U}[0.1, 0.3]$) than the centrifugal gene flow scenario (continuous line, represents the mean values after sampling the intensity of gene flow, m, from a uniform distribution with range $\mathcal{U}[0.1, 0.3]$). Diversification rate and species richness have correspondence in the landscape (compare Figure 2A with 3). Plotted lines represent the mean values of 10^5 slices ranging from the center to the periphery of the network.

Tables

Table 1	**Glossary of mathematical notation**
Notation	**Definition**
$\mathcal{S}$	Total number of sites
N_e^i	Number of individuals in site i
N_e	Total number of individuals in the landscape
(x_i, y_i)	Spatial location of site i
$\mathrm{d_{ij}}$	Euclidean geographical distance between site i and j
$\mathcal{D}$	Geographic distance matrix containing all pairwise distances, d_{ij}
d_{max}	Maximum geographic distance to connect two sites
$\mathrm{q_{kl}}$	Genetic similarity between individual k and l
$\mathcal{Q}$	Genetic similarity matrix containing all the pairwise similarity values, $\mathrm{q_{kl}}$
m	Intensity of gene flow
m_{ij}^k	Gene flow from site j to site i for species k
$\mathrm{q_{min}}$	Minimum genetic similarity to have fertile offspring
μ	Mutation rate per nucleotide per birth-death cycle
$\lambda_{x,y}$	Speciation rate in spatial location (x, y)
$\mu_{x,y}$	Extinction rate in spatial location (x, y)
$\Omega_{x,y}$	Diversification rate in spatial location (x, y)

Figures

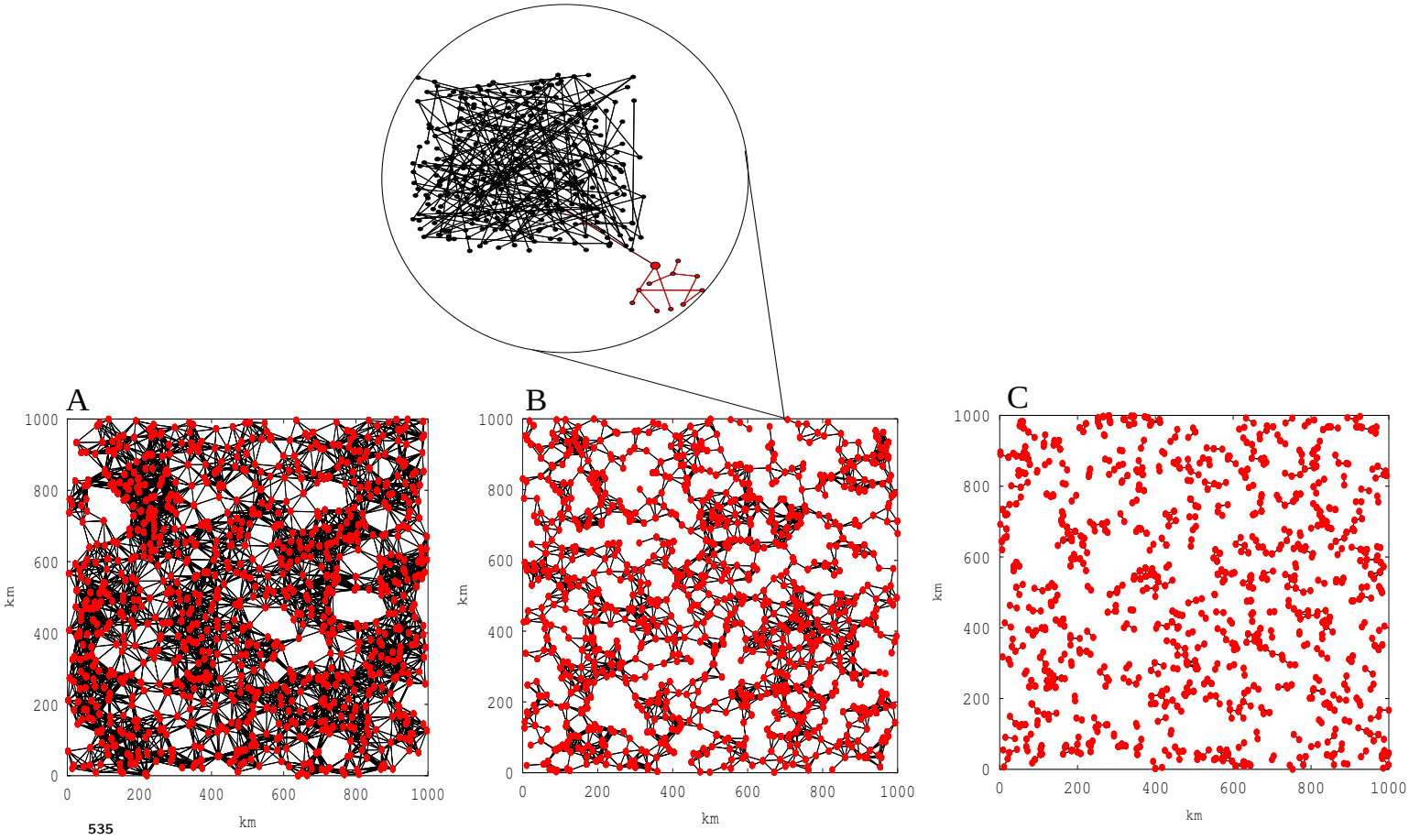


Figure 1

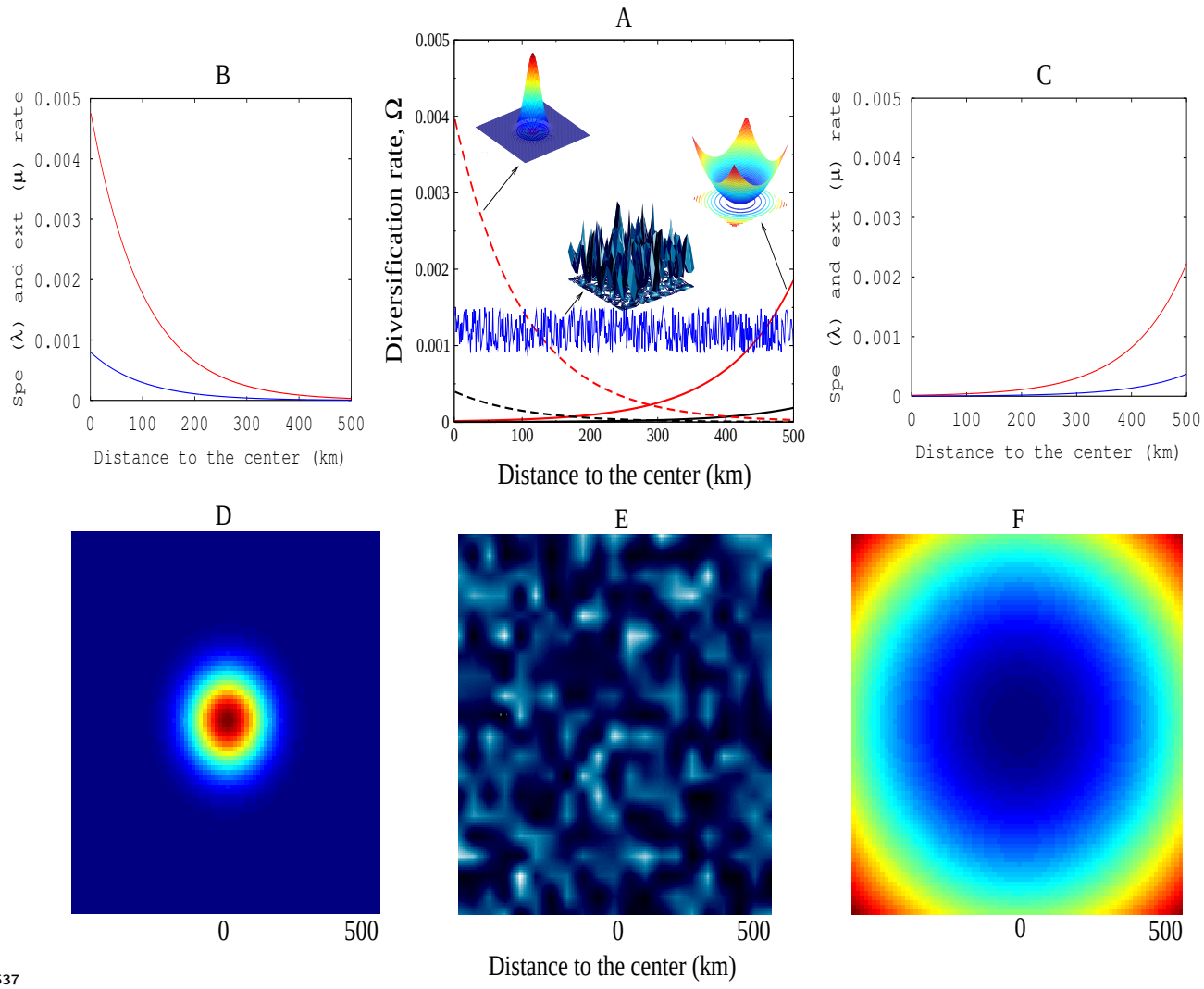
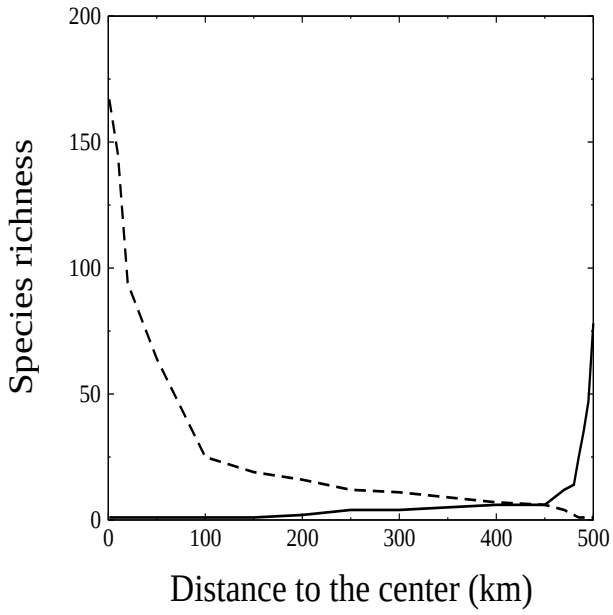

A
B
C
D
E
F
Diversification rate, Ω
Spe (λ) and ext (μ) rate
Distance to the center (km)
0.005
0.004
0.003
0.002
0.001
0
100
200
300
400
500

Figure 2



539

540 Figure 3