Holey Fitness Landscapes and the Maintenance of Evolutionary Diversity

Greg Paperin, Suzanne Sadedin, David Green, Alan Dorin

Faculty of Information Technology, Monash University
Monash University Clayton Campus, Building 63, Wellington Road
Clayton, 3800 Victoria, Australia
gpaperin@infotech.monash.edu.au

Abstract

Analytical models show that high-dimensional fitness landscapes form “holey” rather than “rugged” topographies, but the implications of this finding for biological and artificial life systems remain largely unexplored. One of the reasons for this gap can be attributed to serious difficulties in the implementation of individual-based holey fitness landscape (HFL) models. Here, we introduce a method for simulating HFLs in spatially explicit individual-based models that overcomes these difficulties. We examine how the HFL changes predictions for the maintenance of genetic diversity in the face of migration. Previous models suggest that ecologically-based reproductive isolation will rapidly collapse under migration. Our results indicate that an underlying HFL can often maintain diversity in this situation. Hybrid species emerge frequently when HFL genetics are simulated, but are usually doomed to extinction because of small population sizes. However, hybridisation can also lead to novel adaptations and potentially the exploitation of new ecological niches. More generally, the results imply that HFL genetics should not be neglected in studies of adaptation and diversity.

Introduction

The processes underlying the emergence and persistence of diversity form a key topic in evolutionary theory. Analytical models have provided considerable insight into these issues, but integrating the findings from different theoretical approaches remains a formidable challenge. In particular, the relationship between genetic diversity and reproductive isolation – widely considered the defining feature of biological species [3, 4] – remains controversial [5-9]. Here, we explore the dynamics of reproductive isolation (RI) in a genetically realistic fitness landscape within an individual-based, spatially explicit model.

Reproductive isolation (RI) is often seen as a requirement for biological diversification because it permits the coexistence of different lineages with co-adapted genomes. However, the origin and persistence of RI requires special circumstances. A mutant individual that is reproductively isolated from the surrounding population will rarely be successful. For this reason, speciation is usually thought to occur between spatially separated populations that acquire incompatible alleles through drift or selection [10-12]**.** However, even in this scenario, the maintenance of RI presents a theoretical challenge: even moderate migration between the two populations leads to selection against incompatible alleles, and the extinction or merging of incipient species is likely. Likewise, when RI is based on ecological divergence or mating barriers, it is often transient, collapsing when selection pressures change.

Recent theoretical advances suggest that assumptions about the relative fitness of different combinations of traits have profound implications for our understanding of these problems [11]**.** In particular, Gavrilets and Gravner [13] showed that when fitness landscapes have high dimensionality (as is likely for real organisms), the topology of the landscape changes from “rugged” to “holey”. Several implications of this insight for speciation theory are explored by [11]**.** However, integrating the HFL into simulation models that incorporate spatially and ecologically plausible assumptions remains a challenge. In the sections that follow, we examine the notion of the fitness landscape and its implications for genetic diversity. We then present a method for integrating HFL genetics into a spatially explicit, individual-based model. Using this model, we explore conditions for maintenance of RI, genetic variation, and for the emergence of hybrid species.

Fitness Landscapes and Speciation

The term “fitness landscape” (FL) was coined by Wright [14] to represent the fitness of all conceivable individuals relative to their traits. He envisaged a rugged landscape, where peaks represented combinations of traits with high fitness separated by valleys of low-fitness trait combinations. On this landscape, selection drives populations uphill. Since Wright’s work, several critiques of the FL concept have been made. Fitness landscapes are usually treated as static networks, but in reality, fitness is the ability to survive and reproduce in a dynamic environment that is constantly changing through co-evolutionary dynamics and external disturbances [2]. Some models account for this by using a FL that changes with time to reflect changes in the environment (e.g. [15]). However, the effects of genes underlying species differences and RI are, in general, not strongly affected by the environment and FLs are, therefore, widely accepted as a useful abstraction in theoretical biology [11].

In terms of FLs, the problem of speciation is that part of a population located at a fitness peak must cross the fitness valley surrounding the peak in order for the diverged genes not to be selected out. Stochastic factors such as genetic drift may act against natural selection and help overcoming fitness valleys, particularly for small populations, however, such factors can only account for selected types of speciation. It has been shown that speciation due to stochastic crossing of fitness valleys is, in general, extremely unlikely [10, 11].

Peaks in low-dimensional spaces become saddle points in higher-dimensional spaces. This led to the suggestion that highly multi-dimensional biological FLs may actually possess a single global maximum that can be reached by hill climbing from (almost) any point [16]. Although this model is useful in some cases, it does not apply in general: the local-maxima-to-saddle-point transformations are outnumbered by the appearance of new peaks in higher dimensions [11].

On a biochemical level, most genetic changes are fitness-neutral. This led to the suggestion that the fitness landscapes may be largely flat [17] and that the main force behind speciation is stochastic genetic divergence, i.e. genetic drift. However, an overwhelming proportion of biochemically conceivable genotypes are, in fact, inviable because they contain deleterious genes or groups of incompatible genes. Neutral fitness landscapes fail to account for this fact.

Holey Fitness Landscapes

A genetic model that accounts for the above limitations is the holey fitness landscape (HFL) introduced by Gavrilets [10, 11, 13]. Generally, a HFL is "an adaptive landscape where relatively infrequent high-fitness genotypes form a contiguous set that expands throughout the genotype space" [10].

To build some intuition for this model, we first recall a few results from the percolation theory which plays an important role in the analytical treatment of HFLs. Consider a 2-dimensional lattice of cells which can assume one of two states: "black" or "white" (figure 1). Let every cell be black with some probability p independently of all other cells, or white with probability $1 - p$. If p is small, the lattice will contain a few black cells, which may be grouped in a number of small, isolated clusters. As p increases, these clusters grow and merge. Once p crosses a certain threshold p_c, most of the black cells merge together into a single giant cluster that percolates the whole lattice (see figure 1). For a 2-dimensional square lattice this percolation threshold is known to be $p_c \approx 0.5927$ [18]. However, for lattices of higher dimensions the percolation threshold lies around the reciprocal of the lattice dimension [19], meaning that for a high dimension lattice a small proportion of black cells is sufficient for the emergence of a giant percolating cluster of connected black cells.

For the HFL model, we assume that a genotype is viable with probability p independent of all other genotypes, and inviable with probability $1 - p$. For the purpose of this discussion, the exact fitness of a genotype is irrelevant and we generalise to set the fitness of all viable and inviable genotypes to 1 and 0 respectively. Assume that all possible genotypes are ordered in an abstract genotype space in which the distance between the genotypes describes the probability or ease of transformation from one genotype to another. Distance 1 means that two genotypes can be transferred into each other through a single one-point mutation. Consider the space of all possible haploid genotypes with L loci and A alleles at each locus (note that for the purposes of this model, a diploid genotype with L loci can be represented as a haploid genotype with $2L$ loci [20]; for simplicity, we will therefore only consider haploid genotypes). The dimensionality of this genotype space is $D = L \times (A - 1)$, and the corresponding percolation threshold is $p_c = 1/D$. Note that even for short (on biological scales) genotypes a relatively small value of p will result in an extensive network of high-fitness ridges extending through the genotype space (e.g. for $L = 10^5$ and $A = 5$, $p_c \approx 20 \times 10^{-7}$). The traditional picture of rugged highly-dimensional FLs is therefore misleading, as these landscapes are characterised by the existence of percolating nearly neutral networks. It can be shown [11, chap. 4] that if the fitness of the genotypes is not restricted to 1 or 0, a large number of such networks emerges, each containing genotypes from a narrow fitness band. Among these networks, those with high fitness are particularly important as adaptive walks along such networks can proceed very far without any substantial loss to fitness.

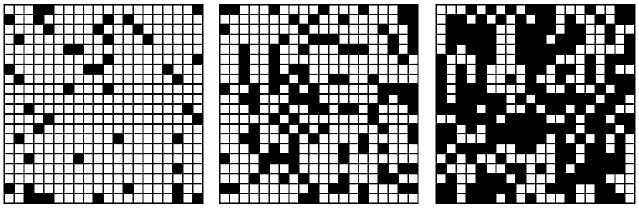

Figure 1. Percolation on a square lattice. The cells are black with probability $p = 0.1$ (left), $p = 0.3$ (middle) and $p = 0.6$ (right).

Holey Fitness Landscape in Simulations

There are a number of analytic models of adaptive radiation based on HFLs (e.g. see [11, part 1]), however they do not incorporate ecological selection and are not explicitly spatial. Other models treat disruptive (diversifying) selection while ignoring the viability and genomic compatibility issues introduced by the HFL (e.g. [21]) or make strong simplifying assumptions about such incompatibilities (e.g. [12]). It is known that diversification occurs easily in large spatial environments with disruptive ecological selection, and there will often be restricted gene flow between the resultant ecotypes, but how enduring RI occurs remains unclear. Gene flow barriers induced by mating barriers – even with strong ecological selection – appear to be transient. Models of adaptive radiation and ecological speciation in general deal with this simply by setting a threshold level of gene flow that they regard as acceptable, but this is unsatisfactory in that such species can merge back together as soon as selective pressures change. HFLs are thought to underlie the evolution of lasting, effective barriers to gene flow that appear during adaptive radiation, however this has not been further explored *in silico*. The reason for this gap is that difficulties arise when realising a HFL in a computer model.

Recall that according to the HFL model, the majority of viable genotypes $G \in \mathbb{V}$ belong to a single largest connected cluster $\mathbb{V}' \subseteq \mathbb{V}$, where $\mathbb{V} \subset \mathbb{G}$ is the set of all viable genotypes and $\mathbb{G}$ is the set of all genotypes. The size of $\mathbb{V}'$ is of the order of $2^L \times p$, and $\mathbb{V}'$ percolates $\mathbb{G}$. The details of the proof can be found in [13]. The proof uses the idea of a surviving branching process to estimate the size of $\mathbb{V}'$. Assume that $p = p_c = 1/(L - 1)$. The probability that the branching process dies at any specific branching point is given by $(1 - p)^{L-1} = (1 - 1/(L - 1))^{L-1} \approx (1 - 1/L)^L$. This means that the above statement holds with a probability 1 when $L \to \infty$. For finite but large L this probability is close to 1, however, for smaller L, the probability of the emergence of the giant connected cluster is smaller.

In natural populations, L is very large, but in an individual-based simulation, the genotype of each individual must be modelled explicitly, held in computer memory and processed by various operations. In practice, this limits the number of loci L to relatively low values. If L is small, $\mathbb{V}'$ can be expected to be small, i.e. $\mathbb{V}$ can be expected to consist of a large number of small clusters that are not connected to each other. Thus, an adaptive walk starting at some $G \in \mathbb{V}$ cannot proceed far in this case and evolution cannot occur.

Note, however, that for small L, the probability that $\mathbb{V}'$ contains most of $\mathbb{V}$ is not large, but positive. For any given small L and $p \geq p_c$, consider all possibilities for selecting $\mathbb{V}$ from $\mathbb{G}$. For most of such possibilities, $\mathbb{V}'$ is small, but there are some choices for which $\mathbb{V}'$ is large.

Any selection of $\mathbb{V}$ from $\mathbb{G}$, for which the giant connected cluster $\mathbb{V}'$ emerges, is an approximation of a HFL for large L. For any such selection, all crucial properties of $\mathbb{V}'$, $\mathbb{V}$ and $\mathbb{G}$ hold and no assumptions are violated. The results about HFL obtained in [11, 13] hold in these cases. If a way to select $\mathbb{V}$ from $\mathbb{G}$ such that the giant connected cluster $\mathbb{V}'$ emerges can be found, the resulting set of genotypes can be used as a basis for individual-based simulations exploring HFL genetics.

One of the challenges in creating an appropriate set $\mathbb{V}' \subset \mathbb{G}$ that is connected and uniformly distributed in $\mathbb{G}$ is related to the fact that the size of $\mathbb{V}'$ grows exponentially with L. We have developed a number of algorithms that allow creating $\mathbb{V}'$ for relatively large values of L (up to 30) within a few minutes on a common desktop computer. In [20] we give an overview of our approach and provide a numerical analysis of the evolutionary properties of the resulting FL.

In short, we create a set $\mathbb{V}'$ of haploid genotypes with a number of diallelic loci represented as bit-strings. This set adheres to the properties described above. The bit-strings are stored in a manner that allows an efficient implementation of a function *viable(G)* that takes an arbitrary bit-string and returns true iff $G \in \mathbb{V}'$.

In the remainder of the paper we introduce an individual-based simulation model designed to investigate to what extent can HFL-genetics sustain RI between spatially separated sub-population in face of migration.

Simulation model

Our objective is to investigate the extent to which HFL can sustain existing RI between spatially isolated populations under different levels of migration. For this we created an individual-based simulation model in which the individuals are located on a homogeneous landscape consisting of cells. Individuals, whose fitness (viability) is defined by the HFL, mate with other individuals within the same cell and then migrate to a neighbouring cell with a certain probability. As common in biological models (e.g. [21]), we use a number of neutral loci to measure the level of gene flow between the populations in different cells for different migration rates.

Methods

In this model the individuals are represented by their genotype, which consists of two sections: a *coding* section and a *neutral* section. The coding section consists of a number of diallelic loci that are assumed to code for vital traits. The coding section of a genotype is used as a parameter to the *viable* function of the HFL in order to determine whether an individual is viable. We experimented with 20 to 28 coding loci (not shown here) and found that the particular number does not affect the results significantly. In the experiments reported here we use L=26, which represents a trade-off between richer genotypes and computational resources required to complete a large number of simulation runs. The neutral genotype section consists of 5 loci with 128 different alleles possible at each locus. The neutral loci do not affect the fitness (viability) of an individual and are used to measure the genetic divergence between individuals (see figure 2).

0 | 1 | 0 | 0 | 1 | 0 | 0 | . . . | 1 | 1 | 1 | 0 | 0 | 038 | 000 | 112 | 127 | 009
Coding section consisting of 26 diallelic loci
Neutral section consisting of five 128-allelic loci

Figure 2. An example of a model genotype.

The lifecycle of a model individual is reproduction – selection – migration. Generations are non-overlapping.

Reproduction. Individuals mate only with other individuals within the same cell of the spatial landscape. Each individual in a cell is selected once as a mother. For each mother, a partner is uniformly randomly selected from the same cell (selfing is permitted). The number of offspring for each pair is drawn from a Poisson distribution with a parameter λ=4 (values in range λ=2..10 did not affect the results significantly). The genotype of each offspring is determined through free recombination of the parents' genotypes (i.e. the allele at each locus is inherited from each parent with equal probability independent of other loci). Each locus of the offspring is mutated with a probability 10^{-4} (values in the range 10^{-3} to 10^{-5} are commonly used in biological models of this kind, e.g. [2, 21]). If a coding locus is mutated, its binary value is flipped. The neutral loci are subject to a circular stepwise mutation model [22]. If the coding section of an offspring's genotype is determined to be viable by the HFL model, the offspring is added to the new generation, otherwise it is discarded immediately. After all offspring for all pairs of parents have been determined, the old generation is discarded and replaced with the new population.

Selection. All individuals within a single cell of the spatial landscape compete to survive to the age of reproduction. Note that this approach is different from the approach commonly used in genetic algorithms, where all individuals survive and then compete to be selected for reproduction. Here all surviving individuals reproduce and their progeny compete to reach a mature age, which normally requires acquiring environmental resources. Each landscape cell is assumed to have a certain maximum carrying capacity C_{mc}, i.e. to provide enough resources for the survival of C_{mc} mature individuals. If a cell is inhabited by no more than C_{mc} individuals, all survive. Otherwise C_{mc} individuals are selected with equal probability and the rest are discarded (as in this HFL model a particular individual is either fit or inviable).

Dispersal. Individuals that reach maturity have a certain probability of migrating to one of the neighbouring spatial landscape cells. To avoid edge artefacts the landscape is represented as a torus. The effect of different migration rates is discussed in the results section.

In order to investigate how spatial distance affects the results we consider different grid layouts. We start with the simplest case (a 1×2 grid) and then gradually increase the grid size (2×2 cells and 3×3 cells). The results (discussed below) imply how the dynamics of RI will behave on larger landscapes. Each cell is initialised with a random viable individual with alleles at neutral loci all set to 0. Initially we disable any migration between the cells and iterate the model for 100 thousand generations in order to allow the allele distribution to reach equilibrium. We then turn on migration at a specific rate (see results section) and iterate the model for 300 thousand further generations. Measurements are taken every 1000 generations.

A quantity of prime interest in this model is the number of reproductively isolated groups (RI groups) present in the model at any one time as well as various attributes of such groups. We are interested in groups of genotypes that could mate successfully, not in groups of individuals who actually do so. Finding such groups is difficult as the groups may be partially overlapping (a genotype can successfully mate with two genotypes that cannot mate with each other) and the genetic distance between groups is initially unknown and may vary. In order to cluster the genotypes of a population into RI groups we employ the Markov Clustering algorithm (MCL) [23], as it does not require a distance threshold parameter and because it has been successfully applied to a similar task – clustering protein sequences into families [24] (we essentially cluster gene sequences into families). For that we first calculate a reproductive success probability matrix for all genotypes in the population. The probability of reproductive success of two genotypes is estimated by simulating a large number of crossovers between the genotypes and considering the proportion of crossovers that result in viable offspring. The matrix is then used as input to the clustering algorithm. To further verify the applicability of MCL to our model we apply this algorithm to a previous model of adaptive radiation that uses the same genetic setup [2]. There we investigated adaptive radiation under disruptive selection caused by ecological niches and RI groups could be determined simply by asserting to which niche genotypes were best adapted. Tests show that the RI groups determined by the clustering correspond to the groups determined by assigning the genotypes to niches (see figure 3).

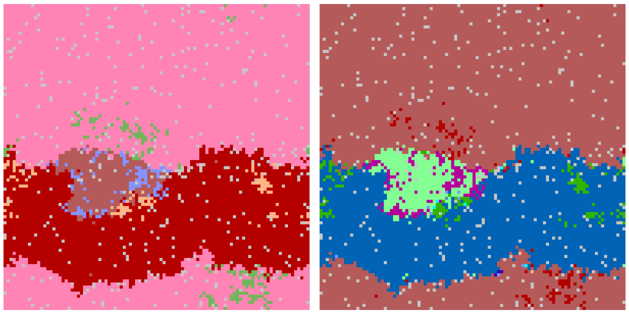

Figure 3. Using Markov Clustering (MCL) for determining RI groups. Depicted is a snapshot of a spatial landscape (100×100 grid) from [2]. Each cell is coloured according to the cluster to which the majority of the genotypes of the individuals inhibiting the cell belong. Left: the genotypes were assigned to RI groups using the MCL algorithm. Right: the genotypes were assigned to RI groups according to the ecological niche to which they are best adapted. Although represented by different colours, both groupings are largely the same.

On the basis of the RI groups we measure the average genetic divergence in neutral loci between the groups using the fixation index F_{st}. A number of slightly different approaches to calculating F_{st} have been proposed. Here we follow the approach taken in [25]: for every pair of genotypes within a group C, we measure the stepwise genetic distance – the minimum number of stepwise mutations necessary to obtain one genotype from the other – and calculate the average genetic distance $d_W(C)$ within the group C. We then measure the pair-wise distances between all genotypes that belong to C and all genotypes that do not belong to C in order to obtain the average genetic distance $d_B(C)$ between C and all other groups. Then, $F_{st}(C) = 1 - d_W(C) / d_B(C)$ and the overall fixation index F_{st} is the average of $F_{st}(C_i)$ for all groups C_i. Note that groups of different sizes are treated equally in this approach.

For each of the scenarios discussed below we have performed 10 independent model runs and averaged the results.

Simulation Results

Consider first the 2×2 layout. As a basis for comparison we performed a set of runs with a migration rate of 0%. As expected, the number of RI groups corresponds to the number of cells (4), the divergence at neutral loci grows (F_{st} approaches 1) and the number of distinct coding genotype sections in the population fluctuates around a value slightly higher than the number of RI groups – due to viable mutants and drift (see figure 4). In one of the runs, two of the cells appeared not to be reproductively isolated as the random founder individuals were genetically similar by chance.

In the next scenario we increased the migration rate to 1% after the first 100,000 generations. This lead to a slight increase in the number of distinct coding sections in the population which is due to viable hybrids resulting from breeding with immigrants. Some of these hybrids spontaneously form RI groups, however such groups cannot persist due to low population numbers in comparison to native populations. These viable hybrids facilitate a limited gene flow between the populations: after 300,000 generations F_{st} has decreased to ca. 0.8 (figure 4.D).

The turnover of viable coding genotype sections in the population over time (the number of distinct viable coding sections that have been present in the model population from the start until a given time) can be used to describe the rate at which novel adaptive phenotypes are evolved. In the first 100,000 generations, when migration rate is 0%, the turnover increases at a small rate due to genetic drift. Once migration is enabled, the turnover grows at a higher rate which suggests that more new viable genotypes are discovered through hybridisation than through generic drift. While this result is sensitive to the mutation rate, it is even more pronounced for higher migration rates (figure 4.C).

In the next scenario the migration rate was set to 5% after the first 100,000 generations. Qualitatively, the results are similar to the 1% scenario. Quantitatively, the gene flow between the populations is higher (F_{st} falls to ca. 0.7, not shown). The higher migration rate leads to an increased probability for formation of RI hybrid groups (figure 4.A). Genetic drift within a larger number of RI groups as well as hybridisation between more diverse individuals leads to a larger number of coding genotype sections in the population (figure 4.B) and to a higher rate of discovering new viable adaptations (figure 4.C). Further rises in the migration rate to 10% (not shown), 15% and 20% (figure 4) increase the strength of the above effects.

When the migration rate is set to 25% or more the RI can be no longer sustained. A large number of reproduction events that lead to inviable offspring have a destabilising effect on the population size. Under such conditions, there is a high chance of extinction for any native cell population. Once an immigrant population has become established in a cell, a positive feedback loop is created: For individuals of the native population the chance of having viable offspring is decreased by the presence of the invaders, as they may be selected as mating partners. At the same time, the chance of new invaders to successfully reproduce is increased. As seen in figure 4.A the number of RI groups collapses to 1 under 25% migration. Sporadically small RI groups arise due to drift, but do not persist long enough to achieve a significant divergence in neutral loci (figure 4.D). The main population evolves as a single RI group. As a consequence, the number of distinct coding sections in the population is very small (figures 4.B & 4.C).

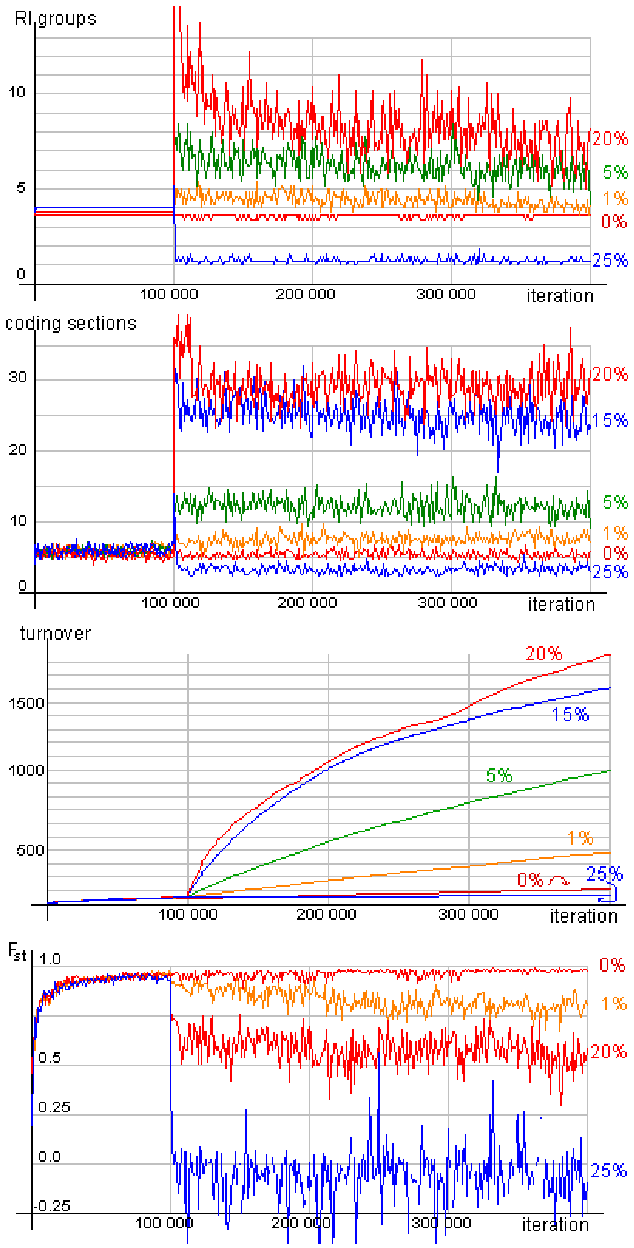

Figure 4. Evolution on a 2×2 grid for the migration rates 0% (red), 1% (orange), 5% (green), 15% (blue), 20% (red) and 25% (blue). Data averaged over 10 runs. Some values omitted for clarity.

A (top): The number of RI groups increases when the migration rate is higher. For very high migration rates the whole model population collapses into a single reproductive group.

B (2nd from top): The number of distinct coding genotype sections in the population increases when the migration rate is high. As the population collapses to a single reproductive group at very high migration rates, the number of coding sequences falls.

C (3rd from top): The rate of evolving new viable coding genotype sections increases when migration rate is higher due to drift in a larger number of IR groups and due to hybridisation between more RI groups. As the population collapses into a single reproductive group at very high migration rates, the number of coding sequences falls.

D (bottom): Genetic divergence between RI groups measured using the fixation index. Higher migration rates lead to increased gene flow and this lower genetic divergence.

In order to investigate how spatial distance affects the above results we have repeated the experiments on a 1×2 grid. In large the model behaviour is similar, however the migration rate has a larger impact on the smaller landscape.

Readily a migration rate of 1% causes F_{st} to decrease to ca. 0.5 after 300,000 generations of migration (figure 5.C). A migration rate of 10% causes the generic divergence of the two RI groups to decrease to insignificant levels within 50,000 generations of migration. However, RI can be sustained at 10% and 15% migration – the number of RI groups stays around 2 which shows that the significant gene flow is not sufficient to break RI and must occur through viable hybrids, who, however, cannot establish a separate RI population. This can also be seen in that the number of distinct coding sections in the population remains small (not shown) suggesting that hybrids occur between the same genotypes. This conclusion is further supported by the turnover rate of the coding sections (figure 5.B): After an initial increase similar to the 2×2 scenarios, the turnover rate slows down to a level close to the rate before migration was turned on, showing that the two populations have reached an equilibrium and that further genetic innovation is due to drift. At 20% migration RI collapses rapidly and the entire model population evolves as a single reproductive group (figure 5).

Next, we repeated the experiments on a 3×3 grid. As expected, larger grid makes it possible to sustain RI at higher migration rates. At 30% migration RI is sustained and the number of RI groups lies above 40. At 35% migration, RI collapses in a way similar to the previous scenarios (not shown).

In order to give our HFL simulations a basis for comparison, we simulated all of the above scenarios without the HFL. In these control runs all individuals are viable and selection is thus random. In this context, RI cannot be defined and the number of RI groups and F_{st} cannot be measured. However, a related measure is the average genetic divergence D_B at neutral loci between all individuals of the entire model population. In the presence of groups without gene flow between them, D_B is expected to grow as the neutral loci in such groups will diverge. We measured D_B for all grid sizes discussed earlier. As expected, for a migration rate of 0%, D_B steadily grows. However, for all grid sizes, a migration rate of 1% is sufficient to cause D_B to drop sharply and to remain low for the rest of the simulation (figure 6). This indicates that without HFL-genetics (or other reproductive barriers occurring in nature) RI cannot be sustained even for small migration rates.

Conclusions

The role of spatial separation in facilitating RI is well known [14]. The difference between the three spatial scenarios demonstrates this effect. In order for an allele to pass from one cell to another non-adjacent cell it must first become established in the intermediate locations. Strong RI induced by the HFL enhances this effect. Thus, hybrid zones and divergent satellite populations may provide a stronger barrier to gene flow than often assumed.

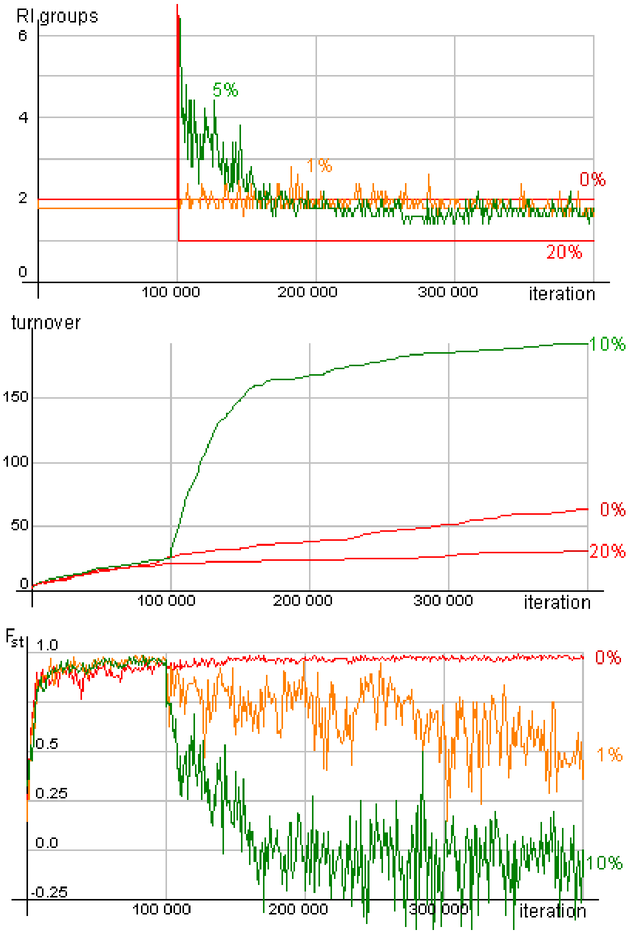

Figure 5. Evolution on a 1×2 grid for the migration rates 0% (red), 1% (orange), 5% (green), 10% (green) and 20% (red).
Data averaged over 10 runs. Some values omitted for clarity.
A (top): Number of RI groups.
B (middle): Turnover of viable coding genotype sections.
C (bottom): Genetic divergence measured using the F_{st}.
(The graphs in this paper were created and processed using the LiveGraph exploratory data analysis and visualisation framework [1].)

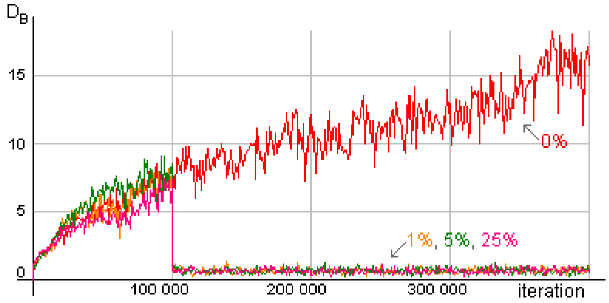

Figure 6. Average genetic divergence D_B at neutral loci between the individuals of the entire population in neutral evolution (without HFL). The average genetic distance grows when migration rate is 0%. The average distance quickly collapses to a small value above 0 (due to drift) for all other migration rates (1%, 5%, 25% are shown). This behaviour is largely the same for all grid sizes considered.

The effect of higher mutation rates on the number of distinct viable coding sections in the population is stronger than on the number of RI groups. This suggests that despite HFL, a small proportion of hybrids is viable and does not exhibit RI from the main population. It is these viable hybrids that facilitate the gene flow between RI populations. However, the small effect of an increasing migration rate on the number of RI groups implies that some hybrid populations exhibit real RI and are not simply fuelled by repeated hybridisation with immigrants.

The common assumption is that hybrid zones are maintained by an interaction between continuous hybridisation and selection against hybrids. RI between the hybrids and the main population is often attributed to ecological preferences to a specific environment within the hybrid zone and not to genetic incompatibility. If such a specific ecological environment is altered, the hybrids become disadvantaged. As a result they become extinct either through selection against them or by adapting to the main environment thus removing RI between the hybrids and the main population. However, hybrid populations that have strong genetic incompatibilities with the main population caused by HFL-genetics are more likely to persist. In our simulations such populations are short-lived because their small initial population size and the absence of prezygotic isolation (RI caused by not mating with members of other groups rather than by offspring inviability) make it unlikely that they successfully reproduce for a large number of consecutive generations. However, in the presence of a free ecological environment niche within the hybrid zone, hybrid groups can multiply in numbers and persist. These populations, once numerous, are less likely to be affected by a disturbance of their specific ecological niche due to the strong genetic RI between the hybrids and the main population. This can allow the hybrid population to further diverge eventually forming prezygotic RI and thus to speciate. Although further data are required, this observation provides potential support for the analogy of novel species to point mutations implicit in some recent ecological [26] and macro-evolutionary [27] theory.

As discussed earlier, for relatively high migration rates, an immigrant (not hybrid) population that became established in a new environment is likely to induce a positive feedback loop leading to the extinction of the native population: A large number of immigrants who act as potential mating partners in the absence of prezygotic RI decreases the chance of native inhabitants to have viable offspring and increases the chance of further invaders to successfully reproduce. This may lead to reinforcement – the evolution of sexual selection and thus prezygotic mating barriers in response to selection against hybrids. Reinforcement is a controversial topic in speciation theory [28, 29]. However, as argued in the previous paragraph and supported by our results, RI generated by the HFL is often resistant to mutations reducing hybrid disadvantage. Thus, reinforcement may be more likely in the context of HFL-genetics than previous models indicate [28, 29].

The notion of holey fitness landscapes, while largely unchallenged, has arguably received insufficient attention from theorists. The current model shows that simulating plausible fitness landscapes can considerably change predictions about the maintenance of diversity and the emergence of new adaptations and species. The approach described here may be useful in further exploring these issues and related problems of adaptive radiation, evolvability and evolutionary search. From the perspective of artificial life research, representing fitness landscapes in a biologically plausible way may facilitate ongoing adaptive exploration and the continuous generation of novelty in evolving populations.

References

[1] (2007). *LiveGraph - a framework for real-time data visualisation, analysis and logging.* Retrieved on 01.03.2008 from: http://www.live-graph.org.

[2] G. Paperin, D. G. Green, S. Sadedin and T. G. Leishman (2007). A Dual Phase Evolution model of adaptive radiation in landscapes. In M. Randall, H. A. Abbass and J. Wiles (eds.), *The Third Australian Conference on Artificial Life (ACAL'07)*, pp. 131-143 Springer.

[3] T. G. Dobzhansky (1937). *Genetics and the origin of species*. Columbia University Press, New York.

[4] E. Mayr (1942). *Systematics and the origin of species*. Columbia University Press, New York.

[5] K. De Queiroz (1998). *The general lineage concept of species, species criteria, and the process of speciation: a conceptual unification and terminological recommendations*. In *Endless forms: Species and speciation*. D. J. Howard and S. H. Berlocher (eds.), pp. 57-75. Oxford University Press, New York.

[6] R. G. Harrisson (1998). *Linking evolutionary pattern and process: the relevance of species concepts for the study of evolution*. In *Endless forms: Species and speciatiopn*. D. J. Howard and S. H. Berlocher (eds.), pp. 19-31. Oxford University Press, New York.

[7] J. Mallet (1995). A species definition for the modern synthesis. *Trends in Ecology and Evolution*. 10 (7): pp. 294-299.

[8] K. L. Shaw (1998). *Species and the diversity of natural groups*. In *Endless forms: Species and speciation*. D. J. Howard and S. H. Berlocher (eds.), pp. 44-56. Oxford University Press, New York.

[9] A. R. Templeton (1998). *Species and speciation: geography, population structure, ecology and gene trees*. In *Endless forms: Species and speciation*. D. J. Howard and S. H. Berlocher (eds.), pp. 32-43. Oxford University Press, New York.

[10] S. Gavrilets (2003). Models of Speciation: What have we learned in 40 years? *Evolution*. 57 (10): pp. 2197-2215.

[11] S. Gavrilets (2004). *Fitness Landscapes and the Origin of Species*. In *Monographs in Population Biology*. Princeton University Press, Princeton / Oxford.

[12] H. A. Orr (1995). The Population Genetics of Speciation: The Evolution of Hybrid Incompatibilities. *Genetics*. 139 (4): pp. 1805-1813.

[13] S. Gavrilets and J. Gravner (1997). Percolation on the Fitness Hypercube and the Evolution of Reproductive Isolation. *Journal of Theoretical Biology*. 184 (1): pp. 51-64.

[14] S. Wright (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. In D. F. Jones (ed.), *6th International Congress of Genetics*, pp. 256-366.

[15] S. A. Kauffman (1993). *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, USA.

[16] W. B. Provine (1986). *Sewall Wright and Evolutionary Biology* University of Chicago Press.

[17] M. Kimura (1983). *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge, New York.

[18] M. E. J. Newman and R. M. Ziff (2000). Efficient Monte Carlo Algorithm and High-Precision Results for Percolation. *Physical Review Letters*. 85 (19): pp. 4104-4107.

[19] G. R. Grimmett (1999). *Percolation*. Springer.

[20] G. Paperin, D. G. Green and A. Dorin (2007). Fitness Landscapes in Individual-Based Simulation Models of Adaptive Radiation. In T. D. Pham and X. Zhou (eds.), *2007 International Symposium on Computational Models for Life Science (CMLS'07)*, pp. 268-278 American Institute of Physics.

[21] S. Gavrilets and A. Vose (2005). Dynamic patterns of adaptive radiation. *Proceedings of the National Academy of Sciences USA*. 102 (50): pp. 18040-18045.

[22] T. Ohta and M. Kimura (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research*. 22 (2): pp. 201-204.

[23] S. Van Dongen (2000): "Graph Clustering by Flow Simulation". University of Utrecht: Utrecht.

[24] A. J. Enright, S. Van Dongen and C. A. Ouzounis (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research*. 30 (7): pp. 1575-1584.

[25] R. R. Hudson, M. Slatkin and W. P. Maddison (1992). Estimation of Levels of Gene Flow From DNA Sequence Data. *Genetics*. 132 (2): pp. 583-589.

[26] S. P. Hubbell (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press.

[27] S. J. Gould (2002). *The Structure of Evolutionary Theory*. Belknap Press, Harvard.

[28] H. G. Spencer, B. H. Mcardle and D. M. Lambert (1986). A Theoretical Investigation of Speciation by Reinforcement. *The American naturalist*. 128 (2): pp. 241-262.

[29] R. Butlin (1987). Speciation by reinforcement. *Trends in Ecology & Evolution*. 2 (1): pp. 8-13.