Evolution and speciation in a hyperspace: the roles of neutrality, selection, mutation and random drift

Sergey Gavrilets
Department of Ecology and Evolutionary Biology
and Department of Mathematics, University of Tennessee,
Knoxville, TN 37996-1610, USA

Contents

Abstract

The world as we perceive it is three dimensional. Physicists currently believe one needs on the order of a dozen dimensions to explain physical world. However, biological evolution occurs in a space with millions dimensions. Sewall Wright's powerful metaphor of rugged adaptive landscapes with its emphasis on adaptive peaks and valleys is based on analogies coming from our three-dimensional experience. Because the properties of multidimensional adaptive landscapes are very different from those of low dimension, for many biological questions Wright's metaphor is not useful or is even misleading. A new unifying framework that provides a plausible multidimensional alternative to the conventional view of rugged adaptive landscapes is emerging for deepening our understanding of evolution and speciation. The focus of this framework are percolating (nearly) neutral networks of well-fit genotypes which appear to be a common feature of genotype spaces of high dimensionality. A variety of important evolutionary questions have been approached using the new framework.

The problem of speciation

Between 1.4 and 1.8 million species have been described[124]. Current estimates are on the order of 10 million species with some estimates going as high as 100 million species[36,60,84]. There are 950,000 insect species of which 350,000 are beetles, 230,000 species of flowering plants, 69,000 fungal species, 25,000 bony fishes, 13,000 species of nematodes, 9,000 species of birds, about 4,200 mammal species, 1,814 species of rodents, 986 species of bats, hundreds of endemic species of Hawaiian Drosophila, 300 cichlid species in the Lake Victoria and 250 species of gammarids in Lake Baikal[35,129,151]. It has been argued that the living species represent less than 1% of the number of extinct species[110]. This gives the "average rate of speciation" on the order of 3 new species per year[123]. [It is interesting that this number is very close to the rate of 1 new species per year estimated by Lyell (1832; cited in Ref. 16).] Of course any "average" rates of speciation are somewhat misleading for speciation takes place simultaneously in many different geographic locations and its rates vary between different groups of organisms. Table 1 shows rates of genus origination (number of originations/standing diversity per 1 million years [My]) in marine animals in the fossil record for three major faunas[123]. As a rule of thumb, species origination rates are one order of magnitude higher than the corresponding genus origination rate. Thus, for example, an "average" species of Paleozoic crinoids produced a new species in 1 My with a probability of 60%.

The maximum speciation rates known are much higher. These are the rates of speciation in lakes and on islands. The following are some examples compiled by McCune[90]. The Hawaiian Islands have existed for about 5.6 My. During this time a large number of endemic species have originated there: 250 species of crickets, 860 species of drosophilids, 47 species of honeycreepers, 100 species of spiders and 40 species of plant bugs. Fourteen endemic species of finches are known on Galapagos Islands, which have existed for 5-9 My. Other examples are for fishes in lakes. Six species of semionotids originated in Lake P4, Newark Basin, in 5,000-8,000 years, five species of cyprinodontids in Lake Chichancanab in 8,000 years, twenty two species of cyprinodontids in Lake Titicaca in 20,000-150,000 years. Speciation of cichlids in Great African Lakes has been extremely rapid: 5 species in Lake Nabugabo in 4,000 years, 400 species in Lake Malawi in .7-2.0 My, dozens of species in Lake Tanganyika in 1.2-12 My, 11 species in Lake Borombi Mbo in 1.0-1.1 My, and arguably the most spectacular speciation event known - speciation of 300 cichlid species in Lake Victoria in 12,000 years.

Table 1: Rates of genus origination (number of originations/standing diversity per 1 My) in marine animals in the fossil record (after Ref. 123).

Fauna	Taxonomic class	Rate of origination
Cambrian	Trilobites	0.13
	“Monoplacophorans”	0.12
	Hyoliths	0.07
	“Inarticulates”	0.06
Paleozoic	Cephalopods	0.11
	Articulates	0.06
	Crinoids	0.06
	Corals	0.05
	Ostracodes	0.04
	Stenolaemates	0.03
Modern	Echinoids	0.03
	Crustaceans	0.03
	Foraminfera	0.03
	Gastropods	0.03
	Bivalves	0.03

Speciation is a universal biological phenomenon that can be very rapid. Speciation has traditionally been considered to be one of the most important and intriguing processes of evolution. [Recall that Darwin’s book[32] title begins with ”The origin of species ...”] In spite of this consensus and significant advances in both experimental and theoretical studies of evolution, understanding speciation still remains a major challenge[27,30,62,89,126,127]. The main reason for this situation is the ineffectiveness of direct experimental approaches because of the time scale involved. Experimental work necessarily concentrates on distinct parts of the process of speciation intensifying and simplifying the factors under study[116,128]. In situations where direct experimental studies are difficult or impossible, mathematical modeling has proved to be indispensable for providing a unifying framework. Although numerous attempts to model parts of the process of speciation have been made, a quantitative theory of the dynamics of speciation is still missing. Currently, verbal theories of speciation are far more advanced than mathematical foundations.

Rugged adaptive landscapes

Speciation is an extremely complex process influenced by a large number of genetical, ecological, environmental, developmental and other factors. When one is trying to understand a very complex phenomenon, it is very helpful to have a simple model of this phenomenon. A minimal model for discussing evolution and speciation considers an organism as a sequence of genes that has some probability to survive to the age of reproduction. An individual’s genes and the probability of survival are referred to as its genotype and fitness, respectively. The set of all possible genotypes is referred to as genotype space. [Wright referred to the “field of possible gene combinations”.]

Genotype space can be mathematically represented by the vertices of a (generalized) hypercube or an undirected graph[51,111,112]. It is useful to visualize each individual as a point in this genotype space. Accordingly, a population will be a cloud of points, and different populations (or species) will be represented by different clouds. Selection, mutation, recombination, random drift and other factors change the size, location and structure of these clouds.

The relationship between genotype and fitness is one of the most important factors in determining the evolutionary dynamics of populations. This relationship can be visualized using the metaphor of the adaptive landscape[147]. In what follows an adaptive landscape represents fitness as a function defined on the genotype space. To construct an adaptive landscape one assigns "fitness" to each genotype (or each pair of genotypes) in genotype space. Different forms of selection and reproductive isolation can be treated within this conceptual framework. For example, fitness can be a genotype's viability (in the case of viability selection); it can be fertility or the probability of successful mating between a pair of genotypes (in the case of fertility selection or premating isolation, respectively). Following Wright, adaptive landscapes are usually imagined as "rugged" surfaces having many local "adaptive peaks" of different height separated by "adaptive valleys" of different depth. Adaptive peaks are interpreted as different (potential) species, adaptive valleys between them are interpreted as unfit hybrids[8]; adaptive evolution is considered as local "hill climbing"[72], and speciation is imagined as a "peak shift"[147].

However, there are problems with this description and several of its implicit assumptions can be questioned. For instance, do different species really have different fitnesses (cf. Ref. 70) ? Are small differences in fitness important in speciation? [Note that Wright himself believed that "the principal evolutionary mechanism in the origin of species must ... be an essentially nonadaptive one", p.364 in Ref.147.] Are local peaks attainable given mutation and recombination, which destroy "good" combinations of genes, and finite population size, which results in the confinement of the population to an infinitesimal portion of the genotype space? Does formation of a new species always imply a (temporary) reduction in fitness? It does not seem that there are compelling reasons for positive answers to these questions. Finally there is a fundamental problem, realized already by Wright himself: how can a population evolve from one local peak to another across an adaptive valley when selection opposes any changes away from the current adaptive peak?

A possibility of escaping a local peak that has received most attention is provided by stochastic fluctuations in the genetic composition of the population (random genetic drift). Random genetic drift is always present if the population size is finite. The following two examples illustrate the difficulties of stochastic transitions in bistable systems arising in modeling evolution on a rugged landscape.

Fixation of an underdominant mutation. Let us consider a finite population of diploid organisms where a single diallelic locus controls fitness (viability). Let N be the population size and $w_{\mathbf{AA}} = 1, w_{\mathbf{Aa}} = 1 - s, w_{\mathbf{aa}} = 1$ be the fitnesses of genotypes **AA, Aa** and **aa**, respectively (with $s > 0$). An adaptive landscape corresponding to this model has two "peaks" represented by the homozygous genotypes **AA** and **aa** which are separated by a "valley" represented by the heterozygous genotype **Aa**. Assume that initially all N organisms are homozygotes **AA** and consider the fate of a single allele **a** introduced in the population by mutation. If the new allele is neutral, that is if $s = 0$, the probability of its eventual fixation is $1/(2N)$ (Ref. 73). Lande[75] has shown that if $s > 0$, the probability that allele **a** will be fixed in the population (and, thus, the population will shift to a new peak) is approximately

$$U = \frac{e^{-Ns}\sqrt{4Ns/\pi}}{erf(\sqrt{Ns})}$$

times smaller than the probability of fixation of a neutral allele. Here $erf(x)$ is the error function. Some numerical examples: if $Ns = 5$ then $U \sim 0.017$, if $Ns = 10$ then $U \sim 10^{-4}$ and if $Ns = 20$ then $U \sim 10^{-8}$. This shows that if the population is at least moderately large ($N > 200$) and the adaptive valley is at least moderately deep ($s > 0.1$), the probability of a stochastic transition across the valley is extremely small.

Shift between stable phenotypic states. In the model considered above there is a single major locus controlling fitness. However, the majority of traits affecting fitness are controlled by many loci with small effects. Such traits can be modeled using the standard framework of quantitative genetics. Let us consider a finite population of diploid organisms where a single additive quantitative trait z controls fitness. Assume that the distribution of the trait in the population is normal with a constant variance G. Let the fitness function w(z) has two "peaks" at $z = a$ and $z = b$ with a "valley" between them at $x = \nu$. [For example, one can chose $w(z) = exp(-(z-a)^2/V) + exp(-(z-b)^2/V)$with a sufficiently small V]. Assuming that initially the population is at one peak, the expected time until the peak shift is approximately

$$T = \frac{2\pi}{G}\,(-c_a c_\nu)^{-1/2}\left[\frac{\overline{w}(a)}{\overline{w}(\nu)}\right]^{2N}$$

where $\overline{w}(x)$ and c_x are the average fitness of the population and the curvature of the fitness function at $z = x$ and N is the population size[10,77]. Some numerical examples[77]: if the initial adaptive peak is 1.05 times higher than the valley (that is $\overline{w}(a)/\overline{w}(\nu) = 1.05$), then using realistic values of other parameters if $N = 100$ then $T \sim 10^6$ whereas if $N = 200$ then $T \sim 10^{10} - 10^{11}$.

These two examples show that although stochastic transitions across very shallow valleys may sometimes occur, it is highly improbable that they can be a major mechanism of genetic divergence of populations on a large scale. This is especially so if the population size is larger than a few hundred individuals and if the valley is sufficiently deep (that is if the stochastic transitions are to result in significant reproductive isolation). Natural populations are usually much larger than few hundred individuals and reproductive isolation even between closely related species is strong.

Shifting-balance theory. To solve the problem of stochastic transitions between different adaptive peaks Wright[146,148] proposed a shifting-balance theory. He considered populations to be subdivided into a large number of small subpopulations connected by migration. Because local subpopulations are small and there are many of them, there is a non-negligible probability of a stochastic peak shift in at least some of them. Wright reasoned that having been established in a subpopulation, a new adaptive combination of genes can take over the whole system as a result of differential migration. Wright's argument was mainly verbal. Recent formal analyses of different versions of the shifting-balance theory[11,17,42,75,76] have lead to the conclusion that although the mechanisms underlying this theory can, in principle, work, the conditions are rather strict. The main problem is the third phase of the shifting-balance process - the spread of the new combination of genes from a local subpopulation to the whole system[31,42,59].

Founder effect speciation. Another possibility to escape a local adaptive peak is provided by founder effect speciation[23,87,88,125]. In this scenario, a few individuals found a new population that is geographically isolated from the ancestral species and that expands a new area. Here, a stochastic transition to a new peak happens during a short time interval when the size of the expanding population is still small. An advantageous feature of the founder effect speciation scenario relative to the shifting balance process is that the new combination of genes does not have to compete with old combinations of genes which outnumbers it. The proponents of this scenario proposed only verbal schemes without trying to formalize them. Later, formal analyses of founder effect speciation on rugged adaptive landscapes using analytical models and numerical simulation have however shown that stochastic transitions between peaks after a founder event cannot result in a sufficiently high degree of reproductive isolation with a sufficiently high probability to be a reasonable explanation for speciation[8,10,26].

It appears that the apparently crucial question about the mechanism of stochastic transitions between different adaptive peaks cannot be answered. Something should to be wrong. At this moment it is good to look back and ask ourselves why we ask this question at the first place. The question about peak shifts appears to be a very natural one to ask within the realm of Wright's metaphor of rugged adaptive landscapes. However, one has to realize that the metaphors and simple models we use do not only help to answer questions we have but also define the questions that we believe should be asked. Perhaps something is wrong with this metaphor? Even at a close inspection, the metaphor of rugged adaptive landscapes seems very appropriate. Indeed, if we look around what we see are landscapes with elevations (peaks) and depressions (valleys) and there is no way one can get from one peak to another without first descending to some kind of valley between the peaks. But are analogies coming from our three-dimensional experience any good for biological evolution?

Table 2: Gene number (after Ref.19)

Prokaryotes	1,000-8,000
Eukaryotes (except vertebrates)	7,000-15,000
Vertebrates	50,000-100,000

Nearly neutral networks and holey adaptive landscapes

The dimensionality of sequence space can be defined as the number of new sequences one can get from a sequence by changing single elements of the sequence. Even the simplest organisms known have on the order of thousand genes (see Table 2) and on the order of million DNA base pairs. Each of the genes can be at at least several different states (known as alleles). Thus, the dimensionality of genotype space is at least on the order of thousands. It is on the order of millions if one considers DNA base pairs instead of genes. This results in an astronomically large number of possible genotypes (or DNA sequences) which is much higher than the number of organisms present at any given time or even cumulatively since the origin of life. On the other hand, the number of different fitness values is limited. For example, if the smallest fitness difference one can measure (or that is important biologically) is, say, 0.001, then only 1000 different fitness classes are possible. Even if one wants to create an adaptive landscape for a set of binary sequences in a computer memory using double precision, then the assignment of different numerical fitness values to different sequences is only possible for sequences with the length $L < 64$.

There is an important consequence of this observation. Because of the redundancy in the genotype-fitness map, different genotypes are bound to have very similar (identical from any practical point of view) fitnesses. Unless there is a strongly ”non-random” assignment of fitnesses (say all well-fit genotypes are put together in a single “corner” of the genotype space), a possibility exists that well-fit genotypes might form connected clusters (or networks) that might extend to some degree throughout the genotype space. If this were so, populations might evolve along these clusters by single substitutions and diverge genetically without going through any adaptive valleys.

Another consequence of the extremely high dimensionality of the genotype space is the increased importance of chance and contingency in evolutionary dynamics. Because a) mutation is random (which gene will be altered to which allele is unpredictable), b) each specific mutation has a very small probability, and c) the number of genes subject to mutation is very large, the genotypes present will be significantly affected by the random order in which mutations occur. Thus, mutational order represents a major source of stochasticity in evolution in the genotype hyperspace[82,91,92]. One should expect that even with identical initial conditions and environmental factors different populations will diverge genetically.

The origin of the idea

These ideas are well-established, having been discussed in the literature many times. In particular, Dobzhansky[34] pointed out that if there are multiple genes producing isolation, then reproductive isolation between two species evolving from a common ancestor can arise as a by-product of fixing ”complementary” genes, none of which has to be deleterious individually. To illustrate this he proposed a simple verbal model of a two-locus two-allele system in which well-fit genotypes formed a chain connecting two reproductively isolated genotypes. Dobzhansky noted that “this scheme may appear fanciful, but it is worth considering further since it is supported by some-well established facts and contradicted by none” (p.282). Similar schemes were discussed by Bateson (1909, cited in Ref.106), Muller (1942), Maynard Smith (1983), Nei (1976), Barton and Charlesworth (1984). Kondrashov and Mina[74] expressed this idea in terms of a “complex system of ridges in a genotype space” and illustrated it graphically (their Figure 2). The discussions of all these authors were restricted to the statement that if a specific kind of genetic architecture exists, then the problem of crossing adaptive valleys is solved. Maynard Smith[85] made one step further by concluding that this kind of architecture must be present: “It follows that if evolution by natural selection is to occur, functional proteins must form a continuous network which can be traversed by unit mutational steps without passing through nonfunctional intermediates” (p. 564).

Recently Maynard Smith’s conjecture was put on firmer theoretical grounds. On the one hand, extensive “continuous networks” were discovered in numerical studies of RNA fitness landscapes[38,57,58,68,122] and also for protein fitness landscapes[3]. On the other hand, in analytical studies of different general classes of adaptive landscapes the existence of connected networks of well-fit genotypes has been shown to be inevitable under fairly general conditions[51,111,112,113].

A few words about the terminology. In what follows a **neutral network** is a contiguous set of sequences possessing the same fitness. This definition is in accord with that used in Ref.5,68,97 and is synonymous with “continuous network” used in Ref.85, “networks of neutral paths” used in Ref.122, "neutral nets" used in Ref.66 and with "connected components" used in Ref.51. This definition appears to be preferential to the broader definition of a neutral network that did not assume the connectivity (e.g. Ref.112,114,121) and, thus, is somewhat confusing for the word "network" does imply the connectivity. A **nearly neutral network** is a contiguous set of sequences possessing approximately the same fitness. A **holey adaptive landscape** is an adaptive landscape where relatively infrequent well-fit (or as Wright put it, "harmonious") genotypes form a contiguous set that expands ("percolates") throughout the genotype space[51]. [An appropriate three-dimensional image of such an adaptive landscape is a flat surface with many holes representing genotypes that do not belong to the percolating set.]

Simple models

In this section I illustrate the origin of connected networks of well-fit genotypes in some simple models.

Russian roulette model

Let us assume that an individual's genotype can be completely specified by a binary sequence of length L. [Using the population genetics terminology we consider haploid individuals different with respect to L diallelic loci.] In this case the genotype space is equivalent to a binary hypercube. Let us consider a family of adaptive landscapes arising if genotype fitnesses are generated randomly and independently and are only equal to 1 (viable genotype) or zero (inviable genotype) with probabilities p and $1-p$, respectively. [Here, one might think of the set of all possible genotypes playing one round of Russian roulette with p being the probability to get a blank.] The probability p can be interpreted as the probability to get a viable genotype after combining genes randomly. Thus, from biological considerations p is supposed to be rather small (definitely much smaller than that one in the non-genetic version of the Russian roulette). On the other hand, the number of loci, L, is very large. A counter-intuitive feature of this model is that viable genotypes form neutral networks in the genotype space such that members of a neutral network can be connected by a chain of viable single-gene substitutions. Properties of these networks can be identified using methods from percolation theory and random graphs theory[4,51,111,112]. In general, there are two qualitatively different regimes: sub-critical, which takes place when $p < p_c$, and super-critical, which takes place when $p > p_c$, where the critical values p_c, which is known as the percolation threshold, is approximately $1/L$. At the boundary of these two regimes, all properties of adaptive landscapes undergo dramatic changes, a physical analogy of which is a phase transition. In the sub-critical regime there are many small networks, whereas in the super-critical regime there is a single "giant component" that includes a significant part of all viable genotypes and "percolates" through the whole genotype space. The adaptive landscape corresponding to the super-critical regime is "holey." Biologically that means that there is a possibility for substantial evolution by fixing single mutations without crossing any adaptive valleys. In the sub-critical regime, typical members of a network can be connected by a single sequence of viable genotypes. Thus, there is a single possible "evolutionary path." In contrast, in the super-critical regime, typical members of the percolating neutral network can be connected by many different evolutionary paths. It is very easy to see why the percolation threshold p_c in this model should be approximately equal to $1/L$. Indeed, if the number of loci L is very large and $p > 1/L$, then each viable genotype will have at least several viable neighbors (that is genotypes different in a single gene) and the network of viable genotypes will "percolate." The percolation threshold decreases if there are more than two alleles. With k alleles at each locus, $P_c \approx 1/(L(k-1))$. If one allows for mutational events affecting more than one gene simultaneously, the percolation threshold decreases dramatically. For example, if a genotype cluster is defined as a set of genotypes that can be connected by a chain of viable single- or two-gene substitutions, the percolation threshold becomes equal to $p_c = 2/L^2$. In general, in high-dimensional genotype spaces (that is when $L * (k-1)$ is large), even very small values of p will result in the existence of a percolating neutral network.

Different properties of a diploid version of this model are discussed in Ref.51. Note that instead of assigning fitnesses 0 and 1, one can assign fitnesses 1 and $\sigma > 1$. In this case, the sequences with superior fitness σ will form a percolating neutral network if $p > 1/L$. Reidys et al.[114] studied the error threshold for a molecular quasispecies evolving on a holey adaptive landscape arising in this model.

Uniformly rugged landscape

The assumption that fitness can only take two values might be viewed as a serious limitation. Here, I consider the same genotype space as in the previous section, but now I assume that genotype fitness, w, is a realization of a random variable having uniform distribution between 0 and 1 (Ref.51). The adaptive landscape arising in this model will be called a "uniformly rugged landscape." Let us introduce two threshold values, w_1 and w_2 such that $w_2 - w_1 = p > 0$, and let us say that a genotype belongs to a (w_1, w_2)-fitness band if its fitness w satisfies to $w_1 < w \leq w_2$. According to the results from the previous section if $p > 1/L$, there is a percolating nearly neutral network of genotypes in a (w_1, w_2)-fitness band. The members of this network can be connected by a chain of single-gene substitutions resulting in genotypes that also belong to the network. If one chooses $w_2 = 1$ and $w_1 = 1 - p$, it follows that uniformly rugged landscapes have very high "ridges" (with genotype fitnesses between $1 - p$ and 1) that continuously extend throughout the genotype space. In a similar way, if one chooses $w_2 = p$ and $w_1 = 0$, it follows that uniformly rugged landscapes have very deep "gorges" (with genotype fitnesses between 0 and p) that also continuously extend throughout the genotype space. If p is small, the fitnesses of the genotypes in the (w_1, w_2)-fitness band will be very similar. Thus, with large L extensive evolutionary changes can occur in a nearly-neutral fashion via single substitutions along the corresponding nearly-neutral network of genotype belonging to a percolating cluster. The maximum number of the non-overlapping (w_1, w_2)-fitness bands is $1/p$, which with p just above the percolation threshold is about L. Thus, the maximum number of percolating near-neutral networks of genotypes is L. In this model, there is a percolating network of well-fit genotypes and, thus, the corresponding adaptive landscape is holey.

Multiplicative fitnesses

In a commonly used multiplicative fitness model alternative alleles are interpreted as "advantageous" and "disadvantageous" and the fitness of an individual with k disadvantageous alleles is chosen to be $(1-s)^k$ with $s > 0$. Here, the fitness landscape has a single peak and $L+1$ different fitness values. Any two genotypes from at the same fitness level can be connected by a chain of single substitutions leading not farther than the previous or the next fitness level. Thus, the number of distinct nearly-neutral networks in this model is approximately $L/2$. These networks can be imagined as spherical shells in genotype space at a constant mean Hamming distance from the optimum genotype. In contrast to the previous model, in the multiplicative fitness model different networks have different sizes with the diameter of the network (the maximum Hamming distance between its members) decreasing from L for genotypes with an equal number of advantageous and disadvantageous alleles to 0 for the most fit genotypes. Woodcock and Higgs[145] have studied this model in detail and shown that a finite population subject to mutation reaches a state of stochastic equilibrium staying close to the fitness level corresponding to U/s disadvantageous alleles. Here U is the rate of mutation per sequence. The whole of the population is clustered together in a particular region of genotype space wandering randomly through the corresponding nearly-neutral network. Gavrilets and Gravner[51] have considered a diploid version of this model in that each genotype is assigned fitness 1 with probability proportional to $p = (1 - s)^k$ where k is the number of heterozygous loci.

Stabilizing selection on an additive trait

A common model in evolutionary quantitative genetics is that of stabilizing selection on a trait z determined by the sum of effects of L diallelic loci, $z = \sum \alpha_i l_i$ where α_i is the contribution of the i-th gene to the trait and $l_i = 0$ or 1 for $i = 1, \ldots, L$. The term "stabilizing selection" means that individual fitness $w(z)$ decreases with the deviation of the trait value from some optimum value θ. Assuming for simplicity that $\alpha_i = 1$ for all i and that optimum θ is at the mid value of the trait range ($\theta = L/2$) results in a single-peak fitness landscape with $L/2$ different fitness values. As in the multiplicative fitness model, any two genotypes at the same fitness level can be connected by a chain of single substitutions leading not farther than the previous or the next fitness level. Thus, the number of distinct nearly-neutral networks in this model is approximately $L/4$. In contrast to the multiplicative model where the most fit nearly-neutral network has the smallest diameter, in the present model this network percolates (and has the largest diameter L) and, thus, the corresponding adaptive landscape is holey. This means that extensive nearly-neutral divergence is possible under stabilizing selection. Barton[9] and Mani and Clarke[82] studied the divergence in this model in detail.

NK model

Stabilizing (or any other nonlinear) selection on an additive trait z results in epistatic interactions between effects of different loci on fitness. The order of these interactions depends on the degree of non-linearity of the fitness function $w(x)$ but each locus epistatically interacts with all other $L - 1$ loci. For example, under quadratic stabilizing selection (that is with $w(z) = 1 - sz^2$), there are pairwise additive-by-additive epistatic interactions in fitness between all $L(L - 1)/2$ pairs of loci[50].

A structurally different class of epistatic model is a family of the so-called NK-models[71] where each locus interacts only with a specified number K of other loci in such a way that interactions of all possible orders (from the second through the K-th order) are present. The existence of neutral and nearly-neutral networks percolating through the genotype space in the NK-models was demonstrated and considered in detail[5,97]. Barnett[5] and Ohta[103,104] have numerically studied the patterns of population evolution in these models.

Conclusions from models

The existence of chains of well-fit genotypes that connect reproductively isolated genotypes was postulated by Dobzhansky and other earlier workers. In contrast, the models just described show it to be inevitable under broad conditions. The existence of percolating nearly-neutral networks of well-fit genotypes which allow for "nearly-neutral" divergence appears to be a general property of adaptive landscapes with a very large number of dimensions. Do existing experimental data substantiate this theoretical claim?

Experimental evidence

Although none of the examples listed below can be viewed as an irrefutable evidence by itself, viewed as a whole these examples provide "substantial and credible evidence" that the genetic architecture leading to the extended nearly-neutral networks of well-fit genotypes is widespread.

i) The most straightforward approach is to analyze the relationships between genotype and fitness[143]. Results of many studies of epistatic interactions in plants, *Drosophila*, mammals and moths[105,138,150] imply the existence of chains of well-fit genotypes connecting genotypes that are reproductively isolated to some degree.

ii) "Ring species" can probably be considered as one of the best manifestations of holey adaptive landscapes. A ring species is a chain of "races" (or subspecies) with gradual transitions and no reproductive isolation between adjacent geographic races but abrupt changes and reproductive isolation where the terminal races come into contact. Nine cases of ring species were described in Ref.87 and more than a dozen additional cases were documented in Ref.88. In ring species, chains of genotypes connecting reproductively isolated forms are recreated in a natural way.

iii) Strong artificial selection in a specific direction usually results in the desired response, but as a consequence of the genetic changes brought about by selection different components of fitness (such as viability or fertility) significantly decrease[65]. Moreover, after relaxing selection, natural selection usually tends to return the population to its original state. These observations stimulated Wright's view of species as occupying isolated "peaks" in an adaptive landscape[109]. However, the size of the experimental populations under selection is usually very small - on the order of few dozen individuals. Small populations will be characterized by low levels of genetic variation and may not "find" ridges in the adaptive landscape even if they were present. Weber[142] performed selection experiments using very large populations of *Drosophila melanogaster* with thousands individuals selecting for the ability to fly in a wind tunnel. Weber was able to change the selected trait (which is obviously non-neutral) by many standard deviation but did not observe any significant reduction in fitness components nor any tendencies to return to the original state after selection was relaxed. A straightforward interpretation of Weber's results is that the large population was able to find a ridge of well-fit genotypes in the adaptive landscape.

iv) Extensive natural hybridization in animals and plants[2,21,117] represents something that is very difficult to reconcile with isolated peaks but is well expected if there are ridges in the genotype space.

v) The existence of fit intermediates between radically different morphologies has been observed in the fossil record[24].

vi) The analyses of RNA sequences and secondary structures have provided abundant evidence for the existence of neutral networks[38–40,57,58,66–68,111,112,120–122].

v) Empirical evidence for extensive functional neutrality in protein space is presented by Martinez with co-workers[83].

vi) Some additional evidence coming from the properties of hybrid zones and patterns of molecular evolution will be considered below.

A metaphor of holey adaptive landscapes

Wright's metaphor of rugged adaptive landscapes puts special emphasis on adaptive peaks and valleys. This metaphor is very useful for thinking about adaptation and optimization. However, its utility for understanding perpetual genetic divergence and speciation is questionable. Overstating to make the point, peaks are largely irrelevant because populations are never able to climb there (cf. Ref.103), and valleys are largely irrelevant because selection quickly moves populations away from there. A finite population subject to mutation is likely to stay mostly within a fitness band determined by the balance of mutation, selection and random drift. Under very general conditions, genotypes with fitnesses within this band form a connected network. Better understanding of the processes of genetic divergence and speciation can be achieved by focusing on these nearly-neutral networks of well-fit genotypes, which are expected to extend throughout genotype space under fairly general conditions. A simplified view of adaptive landscapes that puts special emphasis on these networks is provided by the metaphor of holey adaptive landscapes[43,51]. This metaphor disregards fitness differences between different genotypes belonging to the network of well-fit genotypes and treats all other genotypes as "holes". The justification for the latter is a belief that selection and recombination will be effective in moving the population away from these areas of genotype space on a time scale that is much faster than the time scale for speciation. Accordingly, microevolution and local adaptation can be viewed as the climbing of the population from a "hole" towards the holey adaptive landscape, whereas macroevolution can be viewed as a movement of the population along the holey landscape with speciation taking place when the diverging populations come to be on opposite sides of a "hole" in the adaptive landscape. In this scenario, there is no need to cross any "adaptive valleys"; reproductive isolation between populations evolves as an inevitable side effect of accumulating different mutations. As Charlesworth[25] put it, "the loss in fitness to species hybrids is no more surprising than the fact that a carburetor from a car manufactured in the USA does not function in an engine made in Japan" (p. 103).

Applications

Simple model and metaphors train our intuition about complex phenomena, provide a framework for studying such phenomena and help identify key components in complex systems. Next I briefly review some important biological problems and processes that have been (or can be) studied using approaches focusing on (nearly) neutral networks and holey adaptive landscapes.

Genetic divergence and molecular evolution

One of the consequences of the existence of the percolating nearly-neutral networks of well-fit genotypes is the expectation that biological populations will evolve (and diverge) staying mainly within these networks. The metaphor of "holey" adaptive landscapes neglects the fitness differences between genotypes in the network but these differences are supposed to exist and should be apparent on a finer scale. If one applies a finer resolution, the movement along the network will be accompanied by slight changes in fitness. Evolution will proceed by fixation of weakly selected alleles which can be advantageous, deleterious, over- and underdominant, or apparently neutral depending on the specific area of genotype space the population passes through. Smaller populations will pass faster through the areas of genotype space corresponding to fixation of slightly deleterious mutations whereas larger populations will pass faster through the areas corresponding to fixation of (compensatory) slightly advantageous mutations. These patterns of molecular evolution and genetic divergence, as expected from the general properties of multidimensional adaptive landscapes, are similar to the patterns revealed by the methods of experimental molecular biology[80], which form the empirical basis for the nearly neutral theory of molecular evolution (Ref.99,102; see Ref.55 for an alternative interpretation of the data).

Speciation

A classical view of speciation is that reproductive isolation arises as a by-product of genetic divergence[34,147]. Models incorporating holey adaptive landscapes provide a way to evaluate whether the mechanisms implied by this view may result in (rapid) speciation and to train out intuition about the speciation process.

Nei[95], Wills[144] and Bengtsson and Christiansen[15] initiated formal analyses of the Dobzhansky model. Nei and co-authors[96] studied one- and two-locus multi-allele models with step-wise mutations and considered both postmating and premating reproductive isolation. Genotypes were reproductively isolated if they were different by more than 1 or 2 mutational steps. In their model, speciation was very slow. They conjectured, however, that increasing the number of loci may significantly increase the rate of speciation. Wagner and co-authors[138] considered a two-locus, two-allele model of stabilizing selection acting on an epistatic character. For a specific set of parameters, the interaction of epistasis in the trait and stabilizing selection on the trait resulted in a fitness "ridge". The existence of this ridge simplified stochastic transitions between alternative equilibria. Gavrilets and Hastings[53] formulated a series of two- and three-locus Dobzhansky-type viability selection models as well as models for selection on polygenic characters. They studied these models in the context of founder effect speciation and noticed that the existence of ridges in the adaptive landscape made stochastic divergence much more plausible. In these models, the resulting reproductive isolation can be very high and can evolve with a high probability on the time scale of dozens or hundreds of generations. For appropriate parameter values, Gavrilets and Hastings results have demonstrated that founder effect speciation is plausible. Similar conclusions were reached by Gavrilets and Boake[48] who studied the effects of premating reproductive isolation on the plausibility of founder effect speciation. The adaptive landscape considered by Gavrilets and Boake was defined for *pairs* of genotypes. They have demonstrated that after a founder event a new adaptive combination of genes may rise to high frequencies in the presence of an old combination of genes that is sympatrically (cf. Ref.149).

The models just discussed were formulated for a small number of loci (or quantitative traits). Higgs and Derrida[63,64] proposed a model with an infinitely large number of unlinked and highly mutable loci. In their model the probability of mating between two haploid individuals is a decreasing function of the proportion of loci at which they are different. Here, any two sufficiently different genotypes can be visualized as sitting on opposite sides of a hole in a holey adaptive landscape which is defined for pairs of individuals. Higgs and Derrida as well as Manzo and Peliti[81] studied this model numerically assuming that mating is preferential. In their models populations undergo a continuous process of splitting into reproductively isolated groups with subsequent extinction and/or hybridization and loss of reproductive isolation. Models where the probability of mating between a pair of haploid individuals are mathematically equivalent to models where fitness (viability) of a diploid individual depends on its heterozygosity (that is the number of heterozygous loci).

Orr[105,107] studied possibilities for allopatric speciation in a series of such models for the diploid case. Gavrilets with co-workers[54] have performed individual-based simulations to evaluate whether rapid parapatric speciation is possible if the only sources of genetic divergence are mutation and random genetic drift. Distinctive features of their simulations are the consideration of the complete process of speciation (from initiation until completion), and of a large number of loci, which was only one order of magnitude smaller than that of bacteria. To reflect the idea that reproductive isolation arises simultaneously with genetic divergence, it was posited that an encounter of two haploid individuals can result in mating and viable and fecund offspring only if the individuals are different in no more than K loci (cf., Ref.63,64). Otherwise the individuals do not mate (premating reproductive isolation) or their offspring is inviable or sterile (postmating reproductive isolation). In contrast to Higgs and Derrida approach, the encounters of individuals were random and mutation rates were more realistic (much lower). As a consequence, speciation was irreversible. These numerical results demonstrated that rapid speciation on the time scale of hundreds of generations is plausible without the need for extreme founder events or complete geographic isolation. Selection for local adaptation is not necessary for speciation (cf., Ref.115, 118). The plausibility of speciation is enhanced by population subdivision. Simultaneous emergence of more than two new species from a subdivided population is highly probable. Gavrilets[47] developed some analytical approximations for the dynamics observed in the numerical simulations.

Adaptation

Extended (nearly) neutral networks are important in adaptation for they can be "used" by a population to find areas in genotype space with higher fitness values[121]. Numerical and analytical results for RNA fitness landscapes[39,40,66,68], for a single peak fitness landscape arising in the "Royal Road" genetic algorithm[130,131,132,133], and for multipeak fitness landscapes of the NK model[97] show that evolutionary dynamics of adaptation proceed in a step-like fashion where short periods of jumps to a higher fitness level are interrupted by extended periods during which populations diffuse along neutral networks. Similar behavior has been observed in experiments[79] with bacteria populations adapting to a new environment. Thus, understanding evolution along nearly-neutral networks (and on holey adaptive landscapes) may increase our understanding of local adaptation and microevolution.

Hybrid zones

Hybrid zone is a geographic region where genetically distinct populations meet and interbreed to some extent, resulting in some individuals of mixed ancestry. Analysis of hybrid zones provides insights into the nature of species, the strength and mode of natural selection, the genetic architecture of species differences, and the dynamics of the speciation process[11,12,61]. Many hybrid zones exhibit a gradual change ("cline") in a character or in allele frequency along a geographic transect. Theoretical studies of hybrid zones concentrate on the form of clines and the ability of genes to penetrate hybrid zones[6,14,45,49].

Many hybrid zones are thought to be formed following a secondary contact of different populations, and to be maintained by a balance between selection against hybrids and recombinant phenotypes and dispersal[12]. Gavrilets[44] used the Dobzhansky model to contrast the properties of hybrid zones formed when adaptive peaks are isolated with those formed when adaptive peaks are connected by a chain of well-fit intermediates. A major difference between the two types of hybrid zones should be in the distribution and fitnesses of genotypes in the center of the hybrid zone. If adaptive peaks are isolated, in the center of the hybrid zones besides the high-fitness parental forms one should observe mainly low-fitness hybrids. Moreover, one expects concordant clines in neutral allele frequencies[7,98]. With strong Dobzhansky-type epistatic selection, and low rates of migration reproductive isolation between allopatric populations on opposite sides of the hybrid zone will increase with distance between these populations. F_1 hybrids between individuals from allopatric populations on opposite sides of the hybrid zone will have low fitness. These F_1 hybrids will have genotypes that differ from hybrid genotypes common in the center of the hybrid zone, which will have high fitness. In general, clines in the frequencies of neutral marker alleles linked to selected loci will be disjoint and unsymmetric. Concordant clines are expected for neutral alleles unlinked to selected loci.

Hybrid zones with apparently discordant clines and apparently well-fitted recombinant genotypes present are known for house mouse, grasshopper, common shrew, burney moth and field vole[18,61,69,98,134]. A grasshopper hybrid zone studied by Virdee and Hewitt[134] is especially interesting in this regard. Here crosses between the two pure taxa (*Chorthippus parallelus parallelus* and *Chorthippus parallelus erythropus*) result in sterile male offspring whereas no such dysfunction has been detected in hybrid males collected through the center of the hybrid zone. Crosses have revealed noncoincident clines for dysfunction near the center of the hybrid zone.

RNA and proteins

For RNA sequences neutral networks are defined as contiguous sets of sequences that fold into the same secondary structure. Different biological implications of the existence of neutral networks in both RNA sequence space and protein space are explored in much detail elsewhere[38-40,57,58,66-68,83,111,112,120-

Gene and genome duplication

Conrad[29] puts forward an idea of an "extra-dimensional bypass" on adaptive landscapes. According to Conrad an increase in the dimensionality of an adaptive landscape is expected to transform isolated peaks into saddle points that can be easily escaped resulting in continuing evolution. A straightforward mechanism for increasing the dimensionality of the adaptive landscape is an increase in the size of genome by gene or genome duplication. Similar ideas are discussed by Gordon[56]. The results on the existence of percolating nearly-neutral networks of well-fit genotypes reviewed above provide a formal justification of the idea of an "extra-dimensional bypass" (see Ref.51). In general, the percolation threshold decreases with increasing the dimensionality of genotype space which to a large degree is controlled by the genome size. Increasing the latter by gene or genome duplication[100,101,141] should result in increasing the connectivity of the networks of well-fit genotypes which in turn will increase the possibilities for evolutionary change. Interestingly, it has already been argued that Cambrian explosion was a result of an increase in the gene number[94]. Increasing the genome size increases the redundancy in proteins and DNA[28,137] which may facilitate evolution and result in increased canalization of development.

Canalization of development

There is a general belief that biological systems ought to evolve to a state of greater stability[119,135,140]. One usually distinguishes between genetic canalization (insensitivity to mutations) and environmental canalization (insensitivity to environmental variation). Evolution of environmental canalization has been considered elsewhere[52,140]. The metaphor of holey adaptive landscapes is useful for thinking about the evolution of genetic canalization. From general considerations, one should not expect complete symmetry of "real" adaptive landscapes which are supposed to have areas varying with respect to the width and concentration of ridges of well-fit genotypes. Numerical simulations show that populations tend to spend more time in areas of high concentration of well-fit genotypes[37,67,108]. One of the biological manifestations of this effect will be apparent reduction in the probability of harmful mutations, that is, evolution of genetic canalization[136]. Another manifestation will be a change in the ability of random genetic variation to produce phenotypic changes, that is, evolution of evolvability[1,139]. There is a controversy regarding effects of recombination and sex on the ability of populations to find areas of high concentration of well-fit genotypes. Peliti and Bastolla[108] and Finjord[37] results suggest that only sexual populations tend to find these areas and hang there. On the other hand, Huynen and Hogeweg[67] observed the same effect in modeling asexual populations.

Morphological macroevolution

So far the discussion has been limited to genotype space where individuals were represented as sequences of genes or DNA and RNA base pairs. However, the results on the existence of nearly-neutral networks and holey adaptive landscapes should be valid for any sequence space of high dimensionality. In particular, instead of sequences of genes one can consider sequences of discrete morphological characters and study morphological evolution.

Empirical studies of long-term morphological evolution are typically based on a large number of discrete characters. A common null model in interpreting patterns of morphological changes observed in the fossil record is random diffusion in morphospace[20]. An implicit assumption of this model is that all possible directions for evolution are equally probable. In terms of adaptive landscapes, this corresponds to a flat landscape of neutral evolution[33,73]. In general, because of genetic, developmental, or ecological constraints some of the possible character combinations can be prohibited. In this case, the morphospace will be mathematically equivalent to a hypercube with "holes" (with "holes" representing prohibited character combinations) and the corresponding adaptive landscape will be "holey" rather than "flat". If the proportion of holes is not extremely high, "harmonious" character combinations will form a nearly-neutral network extending throughout the whole morphospace. A characteristic signature of a random walk on a holey hypercube appears to be a stretched exponential dependence of the overlap between the current and initial positions of the walker on time (e.g. Ref.22,78). Gavrilets[46] has developed a model describing the dynamics of clade diversification on a morphological hypercube and applied this model to Foote's data[41] on the diversification of blastozoans. The fitting of the stretched exponential curve to blastozoan data has led to inconclusive results: although the fit is good, it is not better than the fit of a simple exponential curve expected for flat landscapes. More detailed data sets and theoretical results on random walks on hypercubes with holes are needed for more precise conclusions.

Conclusion

Although most attempts to use the approaches discussed here are relatively recent, the list of applications is already impressive. Still we have only started and this list will definitely grow in the nearest future. Among applications that appear to be especially important are adaptations of the approach for the case of continuous (morphological) hyperspace, incorporation of changes in the adaptive landscapes brought about by biotic and abiotic factors, development of a dynamical theory of random walks on neutral networks, analyses of cluster formation in hyper-spaces with emphasis on the origin of hierarchies, and bridging results on the dynamics on neutral networks with the methods for reconstructing phylogenies.

Currently, the mathematics of high-dimensional spaces is very abstract and lacks any real applications. Genotype space arising in evolutionary biology is an example of a hyperspace that is both real and very important. The discovery of complex behaviors of simple ecological models a quarter of a century ago has stimulated the impressive development of the theory of low-dimensional dynamical systems. One can hope that recent advances in evolutionary biology reviewed here will have a similar effect on the development of mathematical theories of the structure of and the dynamics in hyperspace.

Acknowledgments

I am grateful to Jim Crutchfield and Peter Schuster for inviting me to participate in this workshop and to Mitch Cruzan for helpful comments on the manuscript. Partially supported by National Institutes of Health grant GM56693 and Santa Fe Institute.

References

1. Altenberg, L. “Genome Growth and the Evolution of the Genotype-Phenotype Map.” In *Evolution and Biocomputation: Computational Models of Evolution* edited by W. Banzhaf and F. H. Eeckman, 205-259. Springer-Verlag Lecture Notes in Computer Sciences vol. 899, 1995.

2. Arnold, M. L. *Natural Hybridization and Evolution.* Oxford: Oxford University Press, Oxford, 1997.

3. Babajide, A., I. L. Hofacker, M. J. Sippl, and P. F. Stadler. “Neutral Networks in Protein Space: A Computational Study Based on Knowledge-Based Potentials of Mean Force.” *Folding and Design* **2** (1997): 261-269.

4. Balobás, B. *Random Graphs.* New York: Academic Press, 1985.

5. Barnett, L. “Tangled Webs: Evolutionary Dynamics on Fitness Landscapes with Neutrality” MSc thesis, University of Sussex, 1997.

6. Barton, N. H. “Gene Flow Past a Cline.” *Heredity* **43** (1979): 333-339.

7. Barton, N. H. “Multilocus clines.” *Evolution* **37** (1983): 454-471.

8. Barton, N. H. “Founder Effect Speciation.” In *Speciation and Its Consequences* edited by Otte, D. and J. A. Endler, 229-256. Sunderland, MA: Sinauer, 1989.

9. Barton, N. H. “The Divergence of a Polygenic System Subject to Stabilizing Selection, Mutation and Drift.” *Genetical Research* **54** (1989): 59-77.

10. Barton, N. H., and B. Charlesworth. “Genetic Revolutions, Founder Effects, and Speciation.” *Annual Review of Ecology and Systematics* **15** (1984): 133-164.

11. Barton, N. H., and K. S. Gale. “Genetic Analysis of Hybrid Zones.” In *Hybrid Zones and the Evolutionary Process* edited by R. G. Harrison, 13-45. Oxford: Oxford University Press, 1993.

12. Barton, N. H., and G. M. Hewitt. “Adaptation, Speciation and Hybrid Zones.” *Nature* **341** (1989): 497-503.

13. Barton, N. H., and S. Rouhani. “Adaptation and the Shifting Balance.” *Genetical Research* **61** (1993): 57-74.

14. Bengtsson, B. O. “The Flow of Genes Through a Genetic Barrier.” In *Evolution Essays in Honor of John Maynard Smith* edited by Greenwood J. J., P. H. Harvey, and M. Slatkin, 31-42. Cambridge: Cambridge University Press, 1985.

15. Bengtsson B. O., and F.B. Christiansen. “A two-locus mutation selection model and some of its evolutionary implications.” *Theoretical Population Biology* **24** (1983): 59-77.

16. Bennett, K. D. *Evolution and Ecology: the Pace of Life.* Cambridge: Cambridge University Press, Cambridge, 1997.

17. Bergman, A., D. B. Goldstein, K. E. Holsinger, and M. W. Feldman. “Population Structure, Fitness Surfaces, and Linkage in the Shifting Balance Process.” *Genetical Research* **66** (1995): 85-92.

18. Bert, T. M., and W. S. Arnold. “An Empirical Test of Predictions of Two Competing Models for the Maintenance and Fate of Hybrid Zones: Both Models Are Supported in a Hard-Clam Hybrid Zone.” *Evolution* **49** (1995): 276-289.

19. Bird, A. P. “Gene Number, Noise Reduction and Biological Complexity.” *Trends in Genetics* **11**(1995): 94-100.

20. Bookstein, F. L. “Random Walks and the Biometrics of Morphological Characters.” *Evolutionary Biology* **9** (1988): 369-398.

21. Bullini, L. “Origin and Evolution of Animal Hybrid Species.” *Trends in Ecology and Evolution* **9** (1994): 422-426.

22. Campbell, I. A., J. M. Flessels, R. Jullien, and R. Botet. “Random Walks on a Hypercube and Spin Glass Relaxation.” *Journal of Physics C* **20** (1987): L47-L51.

23. Carson, H. L. “The Population Flush and its Genetic Consequences.” In *Population Biology and Evolution* edited by R. C. Lewontin, 123-137. Syracuse, NY: Syracuse University Press, 1968.

24. Carroll, R. L. *Vertebrate Paleontology and Evolution.* New York: Freeman, 1988.

25. Charlesworth, B. “Speciation.” In *Palaeobiology: a Synthesis* edited by D. E. G. Briggs and P. R Crowther, 100-106. Oxford: Blackwell Scientific Publications, 1990.

26. Charlesworth, B. “Is Founder-Flush Speciation Defensible?” *American Naturalist* **149** (1997): 600-603.

27. Claridge, M. F., H. A. Dawah, and M. R. Wilson, eds. *Species. The Units of Biodiversity.* London: Chapman and Hall, 1997.

28. Conrad, M. “The Mutation Buffering Concept of Biomolecular Structure.” *Journal of the Biosciences (Supplement)* **8** (1985): 669-679.

29. Conrad, M. “The Geometry of Evolution.” *BioSystems* **24** (1990): 61-81.

30. Coyne, J. A. “Genetics and Speciation.” *Nature* **355** (1992): 511-515.

31. Coyne, J. A., N. H. Barton, and M. Turelli. “A Critique of Sewall Wright’s Shifting Balance Theory of Evolution.” *Evolution* **51** (1997): 643-671.

32. Darwin, C. *The Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life.* New York: Modern Library, 1859.

33. Derrida, B., and L. Peliti. “Evolution in a Flat Fitness Landscape.” *Bulletin of Mathematical Biology* **53** (1991): 355-382.

34. Dobzhansky, T. H. *Genetics and the Origin of Species.* New York: Columbia University Press, 1937.

35. Eldredge, N. *Life in the Balance: Humanity and the Biodiversity Crisis.* Princeton: Princeton University Press, 1998.

36. Ehrlich, P. R., and E. O. Wilson. “Biodiversity Studies: Science and Policy” *Science* **253** (1991): 758-562.

37. Finjord, J. “Sex and Self-Organization on Rugged Landscapes.” *International Journal of Modern Physics C* **7** (1996): 705-715.

38. Fontana, W., and P. Schuster. “A computer model of evolutionary optimization.” *Biophysical Chemistry* **26** (1987) 123-147.

39. Fontana, W., and P. Schuster. “Continuity in Evolution: On the Nature of Transitions.” *Science* **280** (1998): 1451-1455.

40. Fontana, W., and P. Schuster. “Shaping Space: the Possible and the Attainable in RNA Genotype-Phenotype Mapping.” *Journal of Theoretical Biology* **194** (1998): 491-515.

41. Foote, M. “Paleozoic Record of Morphological Diversity in Blastozoan Echinoderms.” *Proceedings of the National Academy of Sciences USA* **89** (1992): 7325-7329.

42. Gavrilets, S. “On Phase Three of the Shifting-Balance Theory.” *Evolution* **50** (1996): 1034-1041.

43. Gavrilets, S. “Evolution and Speciation on Holey Adaptive Landscapes. *Trends in Ecology and Evolution* **12** (1997): 307-312.

44. Gavrilets, S. “Hybrid Zones With Dobzhansky-Type Epistatic Selection.” *Evolution* **51** (1997): 1027-1035.

45. Gavrilets, S. “Single locus clines.” *Evolution* **51** (1997): 979-983.

46. Gavrilets, S. “Dynamics of Clade Diversification on the Morphological Hypercube.” preprint adap-org/9809002 (1998)

47. Gavrilets, S. “A Dynamical Theory of Speciation on Holey Adaptive Landscapes.” *American Naturalist* **000** (1999): 000-000.

48. Gavrilets, S., and C. R. B. Boake. “On the Evolution of Premating Isolation After a Founder Event.” *American Naturalist* **152** (1998): 706-716.

49. Gavrilets, S., and M. B. Cruzan. “Neutral Gene Flow Across Single Locus Clines” *Evolution* **52** (1998): 1277-1284.

50. Gavrilets, S., and G. de Jong. “Pleiotropic Models of Polygenic Variation, Stabilizing Selection and Epistasis.” *Genetics* **134** (1993): 609-625.

51. Gavrilets, S., and J. Gravner. “Percolation on the Fitness Hypercube and the Evolution of Reproductive Isolation.” *Journal of Theoretical Biology* **184** (1997): 51-64.

52. Gavrilets, S., and A. Hastings. “A Quantitative Genetic Model for Selection on Developmental Noise.” *Evolution* **48** (1994): 1478-1486.

53. Gavrilets, S., and A. Hastings. “Founder Effect Speciation: a Theoretical Reassessment.” *American Naturalist* **147** (1996): 466-491.

54. Gavrilets, S., H. Li, and M. D. Vose. “Rapid Speciation on Holey Adaptive Landscapes.” *Proceedings of the Royal Society of London B* **265** (1998): 1483-1489.

55. Gillespie, J. H. *The Causes of Molecular Evolution.* Oxford: Oxford University Press, 1991.

56. Gordon, R. “Evolution Escapes Rugged Fitness Landscapes by Gene or Genome Doubling: the Blessing of Higher Dimensionality” *Computer Chemistry* **18** (1994): 325-331.

57. Grüner, W., R. Giegerich, D. Strothmann, C. Reidys, J. Weber, I. L. Hofacker, P. F. Stadler, and P. Schuster. “Analysis of RNA Sequence Structure Maps by Exhaustive Enumeration: Neutral Networks.” *Monatshefte fur Chemie* **127** (1996): 355-374.

58. Grüner, W., R. Giegerich, D. Strothmann, C. Reidys, J. Weber, I. L. Hofacker, P. F. Stadler, and P. Schuster. "Analysis of RNA Sequence Structure Maps by Exhaustive Enumeration: Structure of Neutral Networks and Shape Space Covering." *Monatshefte fur Chemie* **127** (1996): 375-389.

59. Haldane, J. B. S. "Natural Selection" In *Darwin's Biological Work. Some Aspects Reconsidered* edited by P. R. Bell, 101-149. New York, Wiley, 1959.

60. Hammond, P. M. "The Current Magnitude of Biodiversity" In *Global Biodiversity Assessment* edited by V. H. Heywood, 113-138. Cambridge: Cambridge University Press, 1995.

61. Harrison, R. G. "Hybrid Zones: Windows on Evolutionary Process." In *Oxford Surveys in Evolutionary Biology, vol.7* edited by D. Futuyma, and J. Antonovics, 69-128. Oxford: Oxford University Press, 1990.

62. Harrison, R. G. "Molecular Changes at Speciation." *Annual Review of Ecology and Systematics* **22** (19914): 281-308.

63. Higgs, P. G., and B. Derrida. "Stochastic Model for Species Formation in Evolving Populations." *Journal of Physics A: Mathematical and General* **24** (1991): L985-L991.

64. Higgs, P. G., and B. Derrida. "Genetic Distance and Species Formation in Evolving Populations." *Journal of Molecular Evolution* **35** (1992): 454-465.

65. Hill, W.G., and A. Caballero, "Artificial Selection Experiments." *Annual Review of Ecology and Systematics* **23** (1992): 287-310.

66. Huynen, M. A. "Exploring Phenotype Space Through Neutral Evolution." *Journal of Molecular Evolution* **43** (1996): 165-169.

67. Huynen, M. A., and P. Hogeweg. "Pattern Generation in Molecular Evolution - Exploitation of the Variation in RNA Landscapes" *Journal of Molecular Evolution* **39** (1994): 71-79.

68. Huynen, M., P. F. Stadler, and W. Fontana. "Smoothness Within Ruggedness: the Role of Neutrality in Adaptation." *Proceedings of the National Academy of Sciences USA* **93** (1996): 397-401.

69. Jaarola, M., H. Tegelström, and K. Fredga. "A Contact Zone with Noncoincident Clines for Sex-Specific Markers in the Field Vole (*Microtus Agrestis*)." *Evolution* **51** (1997): 241-249.

70. Jongeling, T. B. "Self-Organization and Competition in Evolution: a Conceptual Problem in the Use of Fitness Landscapes." *Journal of Theoretical Biology* **178** (1996): 369-373.

71. Kauffman, S. A. *The Origins of Order.* Oxford: Oxford University Press, 1993.

72. Kauffman, S. A., and Levin, S. "Towards a General Theory of Adaptive Walks on Rugged Landscapes." *Journal of Theoretical Biology* **128** (1987): 11-45.

73. Kimura, M. *The Neutral Theory of Molecular Evolution.* Cambridge: Cambridge University Press, 1983.

74. Kondrashov, A. S., and S. I. Mina. “Sympatric Speciation: When Is It Possible?” *Biological Journal of the Linnean Society* **27** (1986): 201-223.

75. Lande, R. “Effective Deme Sizes During Long-Term Evolution Estimated From Rates of Chromosomal Rearrangements.” *Evolution* **33** (1979): 234-251.

76. Lande, R. “The Fixation of Chromosomal Rearrangements in a Subdivided Population with Local Extinction and Colonization.” *Heredity* **54** (1985): 323-332.

77. Lande, R. “Expected Time for Random Genetic Drift of a Population Between Stable Phenotypic States.” *Proceeding of the National Academy of Sciences USA* **82** (1985): 7641-7645.

78. Lemke, N., and I. A. Campbell. “Random Walks In a Closed Space.” *Physica A* **230** (1996): 554-562.

79. Lenski, R. E., and M. Travisano. “Dynamics of Adaptation and Diversification: A 10 000-Generation Experiment with Bacterial Populations.” *Proceedings of the National Academy of Sciences USA* **91** (1994): 6808-6814.

80. Li, W.-H. *Molecular Evolution.* Sunderland, MA: Sinauer Associates, 1997.

81. Manzo, F., and L. Peliti. “Geographic Speciation In the Derrida-Higgs Model of Species Formation.” *Journal of Physics A: Mathematical and General* **27** (1994): 7079-7086.

82. Mani, G. S., and B. C. C. Clarke. “Mutational Order: a Major Stochastic Process in Evolution.” *Proceedings of the Royal Society of London B* **240** (1990): 29-37.

83. Martinez, M. A., V. Pezo, P. Marlière, and S. Wain-Hobson. “Exploring the Functional Robustness of an Enzyme by *in vitro* Evolution.” *EMBO Journal* **15** (1996): 1203-1210.

84. May, R. M. “How Many Species?” *Philosophical Transactions of the Royal Society of London B* **330** (1990): 293-304.

85. Maynard Smith, J. “Natural Selection and the Concept of a Protein Space.” *Nature* **225** (1970): 563-564.

86. Maynard Smith, J. “The Genetics of Stasis and Punctuation.” *Annual Review of Genetics* **17** (1983): 11-25.

87. Mayr, E. *Systematics and the Origin of Species.* New York: Columbia University Press, 1942.

88. Mayr, E. *Animal species and evolution.* Harvard: Belknap Press, 1963.

89. Mayr, E. *The Growth of Biological Thought.* Cambridge, MA: Harvard University Press, 1982.

90. McCune, A. R. “How Fast is Speciation? Molecular, Geological, and Phylogenetic Evidence from Adaptive Radiations of Fishes.” In *Molecular evolution and adaptive radiation* edited by T. J. Givnish, and K.J. Sytsma, 585-610. Cambridge: Cambridge University Press, 1997.

91. Muller, H. J. “Reversibility in Evolution Considered from the Standpoint of Genetics.” *Biological Reviews* **14** (1939): 261-280.

92. Muller, H. J. "Bearing of the *Drosophila* Work on Systematics." In *The New Systematics* edited by J. Huxley, 185-268. Oxford: Oxford University Press, 1940.

93. Muller, H. J. "Isolating Mechanisms, Evolution and Temperature." *Biological Symposia* **6** (1942): 71-125.

94. Nash, J. M. "When Life Exploded." *Time* **146** (1995): 66-74 (N 23).

95. Nei, M. "Mathematical Models of Speciation and Genetic Distance." In *Population Genetics and Ecology* edited by S. Karlin, and E. Nevo, 723-768. NY: Academic Press, 1976.

96. Nei, M., T. Maruyma, and C-I. Wu. "Models of Evolution of Reproductive Isolation." *Genetics* **103** (1983): 557-579.

97. Newman, M. E. J., and R. Engelhardt "Effects of selective neutrality on the evolution of molecular species" *Proceedings of the Royal Society of London B* **265** (1998): 1333-1338.

98. Nurnberger, B., N. H. Barton, C. McCallum, J. Gilchrist, and M. Appleby. "Natural Selection on Quantitative Traits in the *Bombina* Hybrid Zone." *Evolution* **49** (1995): 1224-1238.

99. Ohta, T. "Slightly Deleterious Mutant Substitutions in Evolution." *Nature* **246** (1973): 96-98.

100. Ohta, T. "Time to Acquire a New Gene by Duplication." *Proceedings of the National Academy of Sciences USA* **85** (1988): 3509-3512.

101. Ohta, T. "Evolution by Gene Duplication and Compensatory Advantageous Mutations." *Genetics* **120** (1988): 841-847.

102. Ohta, T. "The Nearly Neutral Theory of Molecular Evolution." *Annual Review of Ecology and Systematics* **23** (1992): 263-286.

103. Ohta, T. "The Meaning of Near-Neutrality at Coding and Non-Coding Regions." *Gene* **205** (1997): 261-267.

104. Ohta, T. "Evolution by Nearly-Neutral Mutations," *Genetica* **102/103** (1998): 83-90.

105. Orr, H. A. "The Population Genetics of Speciation: the Evolution of Hybrid Incompatibilities." *Genetics* **139** (1995): 1803-1813.

106. Orr, H. A. "Dobzhansky, Bateson, and the Genetics of Speciation." *Genetics* **144** (1997): 1331-1335.

107. Orr, H.A., and L.H. Orr. "Waiting for Speciation: the Effect of Population Subdivision on the Waiting Time to Speciation." *Evolution* **50** (1996): 1742-1749.

108. Peliti, L., and U. Bastolla. "Collective Adaptation in a Statistical Model of an Evolving Population." *Comptes Rendus de l'Academie des Sciences Series III - Sciences de la Vie - Life Sciences* **317** (1994): 371-374.

109. Provine, W.B. *Sewall Wright and Evolutionary Biology.* Chicago: The University of Chicago Press, 1986.

110. Raup, D. M. “A Kill Curve for Phanerozoic Marine Species.” *Paleobiology* **17** (1991): 37-48.

111. Reidys, C. M. “Random Induced Subgraphs of Generalized n-Cubes.” *Advances in Applied Mathematics* **19** (1997): 360-377.

112. Reidys, C. M., P. F. Stadler, and P. Schuster. “Generic Properties of Combinatory Maps: Neutral Networks of RNA Secondary Structures.” *Bulletin of Mathematical Biology* **59** (1997): 339-397.

113. Reidys, C. M., and P. F. Stadler. “Neutrality in Fitness Landscapes” **00** (1999): 000-000.

114. Reidys, C. M., C. V. Forst, and P. F. Stadler. “Replication and Mutation on Neutral Networks.” *Bulletin of Mathematical Biology* **00** (1999): 000-000.

115. Rice, W. R. “Disruptive Selection on Habitat Preferences and the Evolution of Reproductive Isolation: a Simulation Study.” *Evolution* **38** (1984): 1251-1260.

116. Rice, W. R., and E. E. Hostert. “Laboratory Experiments on Speciation: What Have We Learned in 40 Years?” *Evolution* **47**(1993): 1637-1653.

117. Rieseberg, L. H. “The Role of Hybridization in Evolution: Old Wine in New Skins.” *American Journal of Botany* **82** (1995): 944-953.

118. Schluter, D. “Ecological Causes of Adaptive Radiation.” *American Naturalist* **148** (1996): S40-S64.

119. Schmalhausen, I. I. *Factors of Evolution. The Theory of Stabilizing Selection.* Chicago: The University of Chicago Press, 1949.

120. Schuster, P. “How to Search for RNA Structures: Theoretical Concepts in Evolutionary Biotechnology.” *Journal of Biotechnology* **41** (1995): 239-257.

121. Schuster, P. “Landscapes and molecular evolution” *Physica D* **107** (1997): 351-365.

122. Schuster, P., W. Fontana, P. F. Stadler, and I. L. Hofacker “From Sequences to Shapes and Back: a Case Study in RNA Secondary Structures.” *Proceedings of the Royal Society of London B* **255** (1994), 279-284.

123. Sepkoski, J. J. “Rates of Speciation in the Fossil Record.” *Philosophical Transactions of the Royal Society B* **353** (1998): 315-326.

124. Stork, N. E. “How Many Species Are There?” *Biodiversity and Conservation* **35** (1993): 321-337.

125. Templeton, A.R. “The Theory of Speciation Via the Founder Principle.” *Genetics* **94** (1980): 1011-1038.

126. Templeton, A. R. “The Meaning of Species and Speciation: a Genetic Perspective.” In *Speciation and Its Consequences* edited by D. Otte, and J. A. Endler, 3-27. Sunderland, MA: Sinauer, 1989.

127. Templeton, A. R. "The Role of Molecular Genetics in Speciation Studies." In *Molecular Ecology and Evolution: Approaches and Applications* edited by B. Schierwater, B. Streit, G. P. Wagner, and R. DeSalle, 455-477. Basel: Birkhaüser, 1994.

128. Templeton, A. R. "Experimental Evidence for the Genetic Transilience Model of Speciation." *Evolution* **50** (1996): 909-915.

129. Thompson, J. N. *The Coevolutionary Process* Chicago: The University of Chicago Press, 1994.

130. van Nimwegen, E., J. P. Crutchfield, and M. Mitchell. "Finite Populations Induce Metastability in Evolutionary Search." *Physics Letters A* **229** (1997): 144-150.

131. van Nimwegen, E., J. P. Crutchfield, and M. Mitchell. "Statistical Dynamics of the Royal Road Genetic Algorithm" *Theoretical Computer Science* **000** (1999): 000-000.

132. van Nimwegen, E., and J. P. Crutchfield. "Optimizing Epochal Evolutionary Search: Population-Size Independent Theory." preprint adap-org/9810003 (1998).

133. van Nimwegen, E., and J. P. Crutchfield. "Optimizing Epochal Evolutionary Search: Population-Size Dependent Theory." preprint adap-org/9810004 (1998).

134. Virdee, S. R., and G. M. Hewitt. "Clines for Hybrid Dysfunction in a Grasshopper Hybrid Zone." *Evolution* **48** (1994): 392-407.

135. Waddington, C. H. *The Strategy of Genes.* New York: MacMillan Co., 1957.

136. Wagner A. "Does Evolutionary Plasticity Evolve?" *Evolution* **50** (1996): 1008-1023.

137. Wagner, A. "Redundant Gene Functions and Natural Selection." *Journal of Evolutionary Biology* **000** (1998): 000-000.

138. Wagner A., G. P. Wagner, and P. Similion. "Epistasis Can Facilitate the Evolution of Reproductive Isolation by Peak Shifts - a Two-Locus Two-Allele Model." *Genetics* **138** (1994): 533-545.

139. Wagner, G. P., and L. Altenberg. "Complex Adaptation and the Evolution of Evolvability." *Evolution* **50** (1996): 967-976.

140. Wagner, G. P., G. Booth, and H. Bagheri-Chaichian. "A Population Genetic Theory of Canalization." *Evolution* **51** (1997): 329-347.

141. Walsh, J. B. "How Often Do Duplicated Genes Evolve New Functions?" *Genetics* **139** (1995): 421-428.

142. Weber, K. E. "Large Genetic Change at Small Fitness Cost in Large Populations of *Drosophila melanogaster* Selected for Wind Tunnel Flight: Rethinking Fitness Surfaces." *Genetics* **144** (1996): 205-213.

143. Whitlock, M. C., P. C. Phillips, F. B.-M. Moore, and S. J. Tonsor. "Multiple Fitness Peaks and Epistasis." *Annual Reviews of Ecology and Systematics* **26** (1995): 601-629.

144. Wills, C. J. “A Mechanism for Rapid Allopatric Speciation.” *American Naturalist* **111** (1977): 603-605.

145. Woodcock, G., and P. G. Higgs. “Population Evolution on a Multiplicative Single-Peak Fitness Landscape.” *Journal of Theoretical Biology* **179** (1996): 61-73.

146. Wright, S. “Evolution in Mendelian Populations.” *Genetics* **16** (1931): 97-159.

147. Wright, S. “The Roles of Mutation, Inbreeding, Crossbreeding and Selection in Evolution.” *Proceedings of the Sixth International Congress on Genetics* **1**(1932): 356-366.

148. Wright, S. “The Shifting Balance Theory and Macroevolution.” *Annual Reviews of Genetics* **16** (1982): 1-19.

149. Wu, C-I. “A Stochastic Simulation Study of Speciation by Sexual Selection.” *Evolution* **39** (1985): 66-82.

150. Wu, C-I., and M. F. Palopoli. “Genetics of Postmating Reproductive Isolation in Animals.” *Annual Review of Genetics* **28** (1994): 283-308.

151. Yampolsky, L. Y., R. M. Kamaltinov, D. Ebert, D. A. Filatov, and V. I. Chernykh. “Variation of Allozyme Loci in Endemic Gammarids of Lake Baikal.” *Biological Journal of the Linnean Society* **53** (1994): 309-323.