BOOK REVIEWS

FITNESS LANDSCAPES AND THE ORIGIN OF SPECIES[1]

N. H. BARTON

Institute of Evolutionary Biology, University of Edinburgh, EH9 3JT, United Kingdom
E-mail: n.barton@ed.ac.uk

Recently, there has been an upsurge of interest in mechanisms of speciation—reflected in a spate of conferences, prominent papers, and in reviews such as Howard and Berlocher (1997) and in *Trends in Ecology and Evolution* 2001, Volume 16, number 7. Of course, the origin of species has been a central issue throughout the history of evolutionary biology, but it was relatively neglected during the last decades of the 20th century as attention focused on the causes of molecular variation, on measuring natural selection in the wild, and on reconstructing phylogenies. Recent work on speciation is scattered over many disciplines. Empirical work has come from phylogenetics and biogeography, laboratory experiments, field studies of diverging populations, and the classical genetics of differences between (mostly) *Drosophila* species. Theoretical work has been similarly scattered, drawing on evolutionary ecology, population genetics, and, most recently, physics. Thus, the publication last year of two books that draw together this disparate field is timely and welcome. Coyne and Orr (2004) review the experimental and comparative evidence, while Gavrilets (2004) summarizes the various theories. Treatment at book length is especially necessary in this area, because many different lines of argument need to be synthesized in a coherent way.

Gavrilets' book is divided into three sections: fitness landscapes, the Bateson-Dobzhansky-Muller (BDM) model, and the joint action of disruptive selection and nonrandom mating. The first section establishes the general properties of fitness landscapes and argues for the inevitability of *nearly neutral networks*—sets of genotypes that are mutationally accessible and that have similar and high fitness. The second section reviews a model of speciation introduced by Bateson in 1909, but usually associated with Dobzhansky and Muller. Here, reproductive isolation evolves as a side effect of divergence across nearly neutral networks; crucially, isolation need not be expressed during divergence. The final section reviews models of sympatric speciation, in which disruptive selection drives the evolution of assortative mating. In such models, speciation is an adaptation that prevents interbreeding between distinct, well-adapted genotypes.

How useful are ''fitness landscapes'' for understanding the origin of species? The idea was introduced by Sewall Wright in 1931, primarily as a metaphor for understanding how populations can adapt when gene interactions cause multiple fitness peaks—hence, Wright's original term, *adaptive landscape*. This way of understanding evolution has been criti-

cized for several reasons (Provine 1986). First, Wright introduced two quite different versions of his adaptive landscape, without making the distinction clear. One can either plot individual fitness against individual genotype, or alternatively, one can define the adaptive landscape as a graph of population mean fitness against the state of the population, as measured by allele frequency or trait means. Second, simple visualizations of the adaptive landscape are misleading, because evolution can occur in very many dimensions—roughly equal to the number of alternative alleles available, which is many times larger than the number of bases in the genome. Finally, the adaptive landscape corresponds to quantitative models only under very restrictive conditions. In the second population-based version, natural selection only takes populations to local peaks of mean fitness if linkage disequilibria can be ignored, so that the population can be properly represented by the set of allele frequencies. In terms of quantitative traits, we must assume a fixed genetic variance, so that evolution consists of changes in trait means.

Despite these criticisms, the principle behind the adaptive landscape is crucial to understanding speciation: the problem is to see how a population, visualized as a cloud of high-fitness genotypes, can split into two distinct clusters, separated by intermediates of low fitness. The key idea is of a mapping of fitness onto genotypes, and of a geometry defined by similarity between genotypes. Gavrilets uses the metaphor of a fitness landscape very effectively: though the arguments are fundamentally mathematical, the mental picture of a multidimensional landscape can be used to present the ideas in a nontechnical way that will be accessible to a wide readership. (One nice innovation is to use the landscape to follow prezygotic as well as postzygotic isolation, simply by plotting the fitness of *pairs* of genotypes).

Gavrilets' central argument in this first section is that under a wide range of assumptions, populations can diverge across a large nearly neutral network. Thus, populations can evolve from one genotype to a very different genotype while maintaining high fitness along the way, and crosses between populations generate recombinant genotypes that have never been produced before, and so tend to be unfit. For example, the simplest model has L loci, each with two alleles, and assigns genotypes as *fit* or *unfit* at random. If the probability that a genotype is fit is not too small (roughly between $1/L$ and $1/2$), then a very large network of genotypes exists, which is within a few mutational steps of all possible genotypes; yet, most recombinant genotypes will be unfit, so that reproductive isolation readily evolves. Introducing a continuous range of fitness and correlations between similar genotypes does not change the conclusions qualitatively, and more

---

250   BOOK REVIEWS

biological models, such as stabilizing selection on an additive trait, have similar properties. The argument is compelling, and intuitively plausible: there are many ways to make a functioning organism, but it is unlikely that random mixtures between them should also function well. However, we do not have many theoretical models that predict fitness as a function of genotype, and there is even less evidence from actual organisms. The best-understood studies are of RNA molecules, where widely divergent sequences can fold in the same three-dimensional structure, and hence maintain the same function. Gavrilets gives a fascinating review of this area.

Gavrilets' argument appears hard to test: how can one determine the fitness of a very large number of genotypes, including those that were part of the diverging chain of ancestors, and those that were never actually realized? However, this can be done (at least, for a sample of genotypes) using experimental evolution of bacteria or RNA molecules, and fitness measurements on crosses between natural populations give us similar information, albeit with an obscure genetic basis. Another way of looking at the problem is to ask about the selection coefficients experienced during evolution: are they typically small or large (10^{-4} or 10^{-1}, say). Evolution must necessarily have been along ridges of high fitness, but the typical magnitude of selection determines whether random drift plays any significant role in divergence. Examples where sequence diversity is reduced in a wide region around putatively selected genes suggests quite strong selection (e.g., Schlotterer 2002) but we have very little evidence of this kind as yet.

In the simplest form, the BDM model encapsulates the idea that species can form when they evolve along ridges of high fitness. One population replaces allele *a* by *A* at one locus, which then replaces *b* by *B* at another locus. Though *ab*, *Ab*, and *aB* must all be fit, *AB* may not be, in which case reproductive isolation has evolved. Although this simple genetic model is nearly a century old, it has hardly been discussed in the literature until recently. This may be because there has been little formal mathematical analysis of such models, with only a few papers prior to Turelli and Orr (1995). So, though the qualitative argument may have been widely accepted, it received little explicit attention. In contrast, models in which random drift knocks populations across fitness valleys invite quantitative analysis. As Gavrilets makes clear in chapter 3, that analysis has shown that drift is unlikely to generate much reproductive isolation. However, the mere existence of formal models has drawn attention to the idea.

Gavrilets has developed models in which random drift causes divergence across nearly neutral networks, without much opposition from selection. These yield simple approximations for the expected time to speciation, the time for speciation to be completed, and so on. Gavrilets' models are valuable because they show that drift can drive speciation when divergence is not opposed by appreciable selection and, moreover, that divergence can occur across a spatially continuous species' range (i.e., in parapatry) as well as in strict allopatry. However, the actual expressions for rates of speciation should not be taken too seriously: they rely on fitness differences between genotypes being very small, which is a feature of the highly symmetric models that Gavrilets uses. I prefer to keep to the basic idea that because the selection

generated by reproductive isolation is negligible in BDM-type models, divergence can be caused by a variety of processes, at a rate which is not affected by whether isolation will be expressed in the future, in crosses between divergent populations. What causes genetic divergence, and what subset of genetic differences are responsible for reproductive isolation, are primarily empirical questions, about which pure theory has little to say.

Mathematical models can say much more about the joint action of disruptive selection and nonrandom mating, which is covered in the final section of the book. Gavrilets gives an excellent critical review of this area. There have been many prominent papers over the past few years that present simulations in which sympatric speciation occurs readily. These have received considerable attention, and have played a large part in increasing the popularity of sympatric speciation among evolutionary biologists. However, simulations can rarely explore the full range of parameters and are not easy to interpret: often, the mechanisms responsible for divergence are not identified correctly. As far as possible, Gavrilets uses simple analytical models, which give a much better understanding. For example, he develops a two-locus model first analyzed by Udovic (1980) to show how the combined strength of assortative mating and disruptive selection determine whether the two become coupled together despite recombination, and thus strengthen reproductive isolation. Perhaps the most important conclusion from this section is that cost-free assortative mating can readily lead to reproductive isolation even without disruptive selection—though some ecological divergence is necessary for distinct types to coexist. The real issue, then, is a biological one: how often can assortment evolve without causing sexual selection against rare types, which would prevent divergence?

Gavrilets' book is admirably comprehensive: his simple classification covers essentially all population genetic models of speciation. The only area that seems to me relatively neglected is the role of gene flow and recombination in impeding divergence. Several striking examples have recently been reported, in which genes for reproductive isolation are closely associated with chromosome rearrangements that prevent recombination (Noor et al. 2001; Rieseberg 2001; Feder et al. 2003). This tells us that gene flow must have occurred during speciation, since chromosomal differences could have no influence in strict allopatry. Although one expects that blocks of genome that cannot be broken up and introgressed though hybrid populations would diverge less readily, models in which divergence is driven by strong selection do not predict a strong effect of gene flow. The observed pattern thus helps us to distinguish alternative mechanisms of divergence, and opens up many possibilities for both empirical and theoretical research.

Gavrilets begins the book with a nice historical survey; he points out that although theoretical population genetics was well developed by the middle of the last century, theoretical models of speciation came very late, beginning in the 1970s. Gavrilets states (p. 4) that the ''situation has now changed, and theoretical studies well outnumber the experimental studies.'' This is only true if one defines experimental studies very narrowly. The numbers of biological papers that mention speciation in title or abstract has increased markedly over the

TABLE 1. Numbers of papers on speciation, and the number of these that were theoretical. Estimated from numbers of papers with *speciation* in title or abstract, from Web of Knowledge (Thomson Scientific). Most such papers are on chemical speciation, not biological speciation. These numbers are estimates based on a sample of 500 papers from each year, classified as nonbiological; biological; and if biological, theoretical.

Year	Papers on speciation	
	Total	Theoretical
2004	955	50
2003	959	59
2002	752	38
2001	638	26
1999	502	67
1994	319	52

past decade (Table 1), and greatly outnumbers theoretical studies. This may just reflect a greater popularity of the term, and we can argue about how much light these papers actually shed on the mechanism of speciation. Nevertheless, it is clear that theoretical studies on speciation are very much a minority—though they can seem overwhelmingly numerous if one tries to read them all.

Theoretical papers on speciation are scattered over a wide range of journals, which makes Gavrilets' synthesis especially valuable. It is remarkable—bizarre, even—that a substantial and increasing fraction (20–30%) are published in physics journals, and are cited almost entirely by physicists. Some are excellent—for example, on understanding the nature of stochastic clustering (Higgs and Derrida 1991). However, the great majority analyze idiosyncratic models that

Book Review Editor: D. Futuyma

make virtually no connection with biological facts. This is a strange situation: in the unlikely event that population geneticists could apply their methods to quantum mechanical problems, would they publish in *Evolution* rather than in a physics journal? Gavrilets' book may help bridge the strange divide that has developed: it provides an excellent, compact overview that is accessible to physicists as well as biologists.

LITERATURE CITED

Coyne, J. A., and H. A. Orr. 2004. *Speciation*. Sinauer Associates, Sunderland, MA.

Feder, J. L., S. H. Berlocher, J. B. Roethele, H. Dambroski, J. J. Smith, W. L. Perry, V. Gavrilovic, K. Filchack, J. Rull, and M. Aluja. 2003. Allopatric genetic origins for sympatric host-plant shifts and race formation in *Rhagoletis*. Proc. Natl. Acad. Sci USA 100:10314–10319.

Higgs, P. G., and B. Derrida. 1991. Stochastic models for species formation in evolving populations. J. Phys. A 24:985–992.

Howard, D. J., and S. H. Berlocher. 1997. *Endless forms: species and speciation.* Oxford Univ. Press, Oxford, U.K.

Noor, M. A. F., K. L. Grams, L. A. Bertucci, and J. Reiland. 2001. Chromosomal inversions and the reproductive isolation of species. Proc. Natl. Acad. Sci. USA 98:12084–12088.

Provine, W. 1986. *Sewall Wright and evolutionary biology.* Univ. of Chicago Press, Chicago.

Rieseberg, L. H. 2001. Chromosomal rearrangements and speciation. Trends Ecol. Evol. 16:351–358.

Schlotterer, C. 2002. Towards a molecular characterization of adaptation in local populations. Curr. Opin. Genet. Dev. 12: 683–687.

Thomson Scientific. ISI Web of knowledge. Available via http:// isiwebofknowledge.com.

Turelli, M., and H. A. Orr. 1995. The dominance theory of Haldane's rule. Genetics 140:389–402.

Udovic, D. 1980. Frequency-dependent selection, disruptive selection, and the evolution of reproductive isolation. Am. Nat. 116: 621–641.