Perspectives

Anecdotal, Historical And Critical Commentaries on Genetics

Edited by James F. Crow and William F. Dove

DOBZHANSKY's *Genetics and the Origin of Species*: Is It Still Relevant?

Richard C. Lewontin

Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138

THE first edition of THEODOSIUS DOBZHANSKY's *Genetics and the Origin of Species* appeared almost precisely 60 years before this issue of GENETICS. It would be hard to find anyone, even envious authors of other *magna opera*, who would disagree with JEFFREY POWELL's 50th anniversary assessment that "it is the most important and influential book on evolution of the twentieth century" (POWELL 1987). It must be remembered, however, that the book was only the concrete form of the Jesup Lectures given the previous year at the invitation of L. C. DUNN and others at Columbia, an invitation that signaled the importance that influential biologists already placed on DOBZHANSKY's previous 10 years of work on genetics and evolution. In an important sense, the publication of the book, to be followed by other Jesup Lectures books by MAYR (1942) and SIMPSON (1944) and eventually STEBBINS (1950), was a manifesto representing a view that was already taking hold.

Nothing would be gained by plagiarizing POWELL's masterful summary of the schism between the communities of genetics and evolutionary biology at the time, and of the highlights of DOBZHANSKY's integration of Mendelism and Darwinism. In this respect, two things that were unique to DOBZHANSKY's book need to be emphasized. First, *Genetics and the Origin of Species* seemed to be essentially a treatise in observational *biology*, speaking in the language and using the biological materials of experimentalists and natural historians. Second, DOBZHANSKY's entire schema began with the origin of variation and culminated with the formation of species, thus seeming to engage DARWIN's outline directly. Some years before, FISHER (1930), WRIGHT (1931), and HALDANE (1932) had already completed syntheses of genetics and evolution at the conceptual level, showing how Mendelism served as a basis for evolutionary change. But their expositions did not rely, as DOBZHANSKY's did, on a large amount of description of observations from nature, and the problem of the origin of species was treated by them only *en passant* (FISHER spent three and a half pages on it), although WRIGHT's picture of alternative adaptive peaks certainly sidled up to the problem.

One historical viewpoint for an appreciation of DOBZHANSKY's book is that of the observer in 1937 looking retrospectively, seeing how DOBZHANSKY's synthesis succeeded in bringing the full apparatus of genetics and of genetic observations in natural populations to bear on the observable facts of speciation and species diversity. It is this synthetic element that was so compelling to his readers of the time. But, like any major scientific synthesis, *Genetics and the Origin of Species* was not simply a compelling reorganization of existing knowledge into a unified structure. The real test of its importance lay in its prospective aspect, in the implicit program for evolutionary genetics from 1937 into the future. As POWELL pointed out in his essay, the book, especially in later editions (1941, 1951), had an important impact in establishing observational population genetics as a scientific field for investigation. In fact, the entire problematic of evolutionary genetics for the last 60 years, including its detailed formulation at present, flows from the organization and content of DOBZHANSKY's Jesup Lectures and the book that embodied them.

DOBZHANSKY's argument, which every graduate student in Zoology at Columbia in his day was expected to reproduce on the written qualifying examination for the Ph.D., was the skeleton of DARWIN's theory of the origin of species. Species are groups of interbreeding organisms that have been cut off, biologically, from sharing heredity with other species with which they share a common ancestry in the remote past. This reproductive isolation is the final step in divergence between geographically separated populations, geographical races, which were originally kept apart only by geography, but which have acquired during their geographical separation sufficient genetic difference to prevent future interbreeding. But for genetic differences to accumulate between populations, there must be genetic variation within populations to begin with. That is, species evolution is a process of the conversion of the variation present between individuals within populations at a given moment into variation between pop-

Author e-mail: dick@mcz.harvard.edu

ulations in time and space. This scheme then places the investigation of intrapopulation genetic variation and polymorphism at the very center of the study of evolutionary dynamics. Even the description of differences between populations is in the form of the statistical description of their polymorphisms rather than by characteristic typological differences. This point of view has differentiated population and evolutionary genetics from all other modes of studying evolution. It is the reason that in a book of 321 pages of text, whose ultimate goal is to explain the origin of species, 178 pages at the beginning are taken up with intrapopulation variation. It is the reason that at present so many population geneticists are skeptical of simple *post hoc* optimality explanations of species characteristics, for they are predisposed to consider the contingency of just the right kind of genetic variation to make the stories work. "I can call monsters from the vasty deep" says that uradaptationist OWEN GLENDOWER. "Why, so can I and so can any man. But will they come when you do call for them?" replies the doubtful population geneticist HOTSPUR. It is the reason that evolutionary geneticists until recently had so neglected a detailed genetic study of the differences that underlie species divergence. After all, species differences are simply the final disposition of the standing genetic variation within species, so it is the nature of that standing variation and of the forces modulating it that is the real stuff of evolutionary genetics. All else is just developmental and molecular biology.

The degree to which the first edition of *Genetics and the Origin of Species* was the enunciation of a problematic for the future, rather than a synthesis of an already adequate body of fact and theory, can be seen in a comparison between the original and later editions. While so much emphasis was placed on the importance of intrapopulation genetic variation in the first edition, the actual evidence was pretty thin. Aside from the few human blood groups then known, DOBZHANSKY and EPLING's (1944) survey of inversion polymorphism in geographical populations of *Drosophila pseudoobscura,* and a few studies of simple Mendelizing morphological polymorphisms and chromosomal lethals by the Dubinin school (see pp. 42–46 in DOBZHANSKY 1937), virtually nothing was known of the frequencies of Mendelian genetic variations in natural populations. DOBZHANSKY's own famous *Genetics of Natural Population* series began to appear only after the Jesup Lectures. By the time he finished the revised third edition in 1951, twenty papers in that series had appeared, comprising a model for how genetical variation in natural populations could be studied. This included observations of temporal variation and stability in polymorphism, estimates of migration and effective population size, evidence for the existence of selective differences in nature, and the creation of laboratory model populations in which selection could be demonstrated and estimated. The third revised edition of 1951 now could refer to 15 years of data from natural and laboratory populations estimating parameters of selection, migration, and breeding structure. Moreover, large quantities of data were now available on the viability and fertility variation among genomes sampled from natural populations of Drosophila, data made possible by an adaptation of MULLER's *ClB* trick for making chromosomes homozygous. Nor was it DOBZHANSKY's school alone that pursued the program, nor Drosophila alone that was the object of study. As a consequence of the medical demands created by the Second World War, great advances had been made in immunological genetics, resulting in an explosion of information on human blood groups and HLA polymorphisms. The most complete model for how to study a Mendelian polymorphism within and between local populations was LAMOTTE's (1951) monograph on the shell color and banding polymorphisms in *Cepaea nemoralis.* A manifesto had become an industry.

The program, while seemingly prosperous, was, however, in deep difficulties. Aside from the occasional, genetically simple morphological or immunological polymorphism, studies of natural genetic variation were dependent on observations of whole chromosomes rather than single physiological and developmentally defined loci. Inversion polymorphism, while serving as a model object of study, could really give no information about the generality of variation on which the genetical theory of evolution depended. Alternatively, the measurement of viability and fertility variation in nature, surely the stuff of evolutionary change, could be assayed only at the whole chromosome level, providing no real information about how much genic variation existed. DOBZHANSKY had created a field and focused investigation on a problematic that seemed impossible to clarify.

The response to this conundrum was the introduction of a method of investigation, protein electrophoresis, that seemed to cut through the difficulty because it (1) provided a phenotype whose variation was easily observable; (2) did not depend on any assumptions on the physiological or developmental consequences of the variation; (3) would detect a large fraction of the variation at a large fraction of loci, locus by locus; and (4) could be applied to any organism irrespective of its amenability to genetic manipulation (HUBBY and LEWONTIN 1966). While it might be flattering to the self-esteem of those who introduced the technique to think of it as "revolutionizing" the field, the truth is quite the opposite. The immense popularity that electrophoretic studies enjoyed for nearly 20 years after their introduction in 1966 was precisely that they seemed to provide the possibility of at last coming to grips with the problematic that had occupied evolutionary genetics since 1937. Unfortunately, the main strength of the method was its fatal flaw. Its essence was that it allowed the assessment of genetic variation unaffected by the physiological and developmental consequences of that variation. But, by liberating the observations from physiology and development, the method also guaranteed that, except in the very extraordinary circumstance that allelic variation at a single locus had a strong marginal effect on fitness, no inferences about the forces operating on the variation could be tested. The Dobzhanskian problematic was even more frustratingly stymied. Now we could describe genetic variation quite generally but seemed barred from explaining it!

The impasse was broken, at least in part, by a lucky fact of nature: the lack of a one-to-one correspondence between the DNA sequence and the amino acid sequence of proteins. The degeneracy of the code, the existence of introns, of transcribed but untranslated and of nontranscribed DNA, all mean that within the same small genic region there are classes of nucleotides with very different relationships to the physiology and development of the organism. Different patterns of genetic variation of these different classes could then provide *internal* evidence about the cumulative effect of selection, which should operate differently on the different classes, as opposed to forces of mutation, population structure, and recombination, which should affect all classes equally. Beginning with the original demonstration by KREITMAN (1983) of the unique power of DNA sequence studies to detect even very weak natural selection unambiguously, the central problematic of evolutionary genetics seemed once again to be accessible. More than a dozen years of population genetics at the nucleotide level have clearly shown that selective constraints exist for all classes of nucleotides including so-called "silent" positions in codons, as well as introns and flanking sequences. (see, as an example, RICHTER *et al.* 1997). Moreover, it has been possible to detect, in patterns of haplotypes, traces of migration among populations (RICHTER *et al.* 1997) and the constraints imposed on variation by differing amounts of recombination (BEGUN and AQUADRO 1992). But, more than this, the study of nucleotide variation has allowed evolutionary genetics to proceed to the next set of questions posed by the schema outlined in *Genetics and the Origin of Species*.

The existence of genetic variation and the modulation of its pattern within a population at any time are only the beginning of the process of species evolution, according to DOBZHANSKY. It is not sufficient that local populations are simply different in gene frequency, for every population must differ from every other one in the real world of finite assemblages. Species are not simply assemblages of organisms that are not interbreeding, but are distinct life forms with distinct relations to the environment, making a living in distinct ways. Nor can this ecological differentiation commonly be a process that follows after reproductive isolation has already occurred, for then we would often observe that partially reproductively isolated populations would show no adaptive differentiation. Unless the populations have come to occupy different peaks in the adaptive landscape, the local populations or geographical races are not likely to be in the preliminary stages in species formation. This, then, poses a second set of problems for population genetics: to demonstrate that natural selection has played a role in population differentiation. It is easy enough to show that strains drawn from different populations have different norms of reaction and different fertilities and viabilities in different experimental circumstances. The first edition of *Genetics and the Origin of Species* uses precisely such evidence to demonstrate genetic differences between local populations. It is a very different matter, however, to link these divergences to specific genetic differences, to show that they matter in nature and that they have been established by some process of adaptive natural selection.

Because of the evident difficulty of such demonstrations, this critical next element in the Dobzhanskian program has been the subject of a great deal of talk but only limited action. The geographical variation of shell patterns in Cepaea provided the opportunity for a long struggle between the English school, which explained all variation as a consequence of local variations in environmental conditions (see, for example, CAIN and SHEPPARD 1950), and the French school, which interpreted the results as a consequence of genetic drift (LAMOTTE 1951). During the heyday of electrophoretic studies, a number of cases of geographical or altitudinal clines in the frequencies of variants were found, and these were correlated with various environmental variables, usually temperature. The most detailed studies linking the frequencies of variants with their enzymatic kinetics, physiology, and behavior, while successful in demonstrating such a relationship, are unable to deal with the basic issue facing all who study natural selection in natural populations, namely, the question of which aspects of the organism's biology account for variance of fitness *in nature*. That is, it may be that, *ceteris paribus*, an increase in egg-laying rate of females would increase the fitness of their genotype, but if females in nature lay so few eggs that differences in physiological potential are irrelevant, then differential physiological fertility is not a significant component of fitness variance. The challenge of studying adaptive variation in nature is that one has to know so much about the biology of the organism. Thus, it would seem that the second phase of the Dobzhanskian project, to show that genetic differentiation has occurred by natural selection, seems to evade us. Once again, studies of nucleotide variation have provided a possibility of progress. By finding short regions of the genome that are markedly depauperate of nucleotide variation for silent sites and introns, as compared with other regions in the same genome, a strong case can be made for a selective gene fixation in the relatively recent past. A striking example is the demonstration by BERRY *et al.* (1991) of a selective sweep on the microchromosome of *Drosophila melanogaster.* Of course, we do not know the biological cause of the sweep nor which sites within the region are responsible for it, as opposed to being carried along by hitchhiking. The demonstration of adaptive differentiation can be carried even further to look for evidence of adaptive divergence between species, detected by an excess of amino acid replacements as compared with silent divergences between them (MCDONALD and KREITMAN 1991). There is, in principle, no limit to how much of the genome could be investigated in this way, choosing particular genes or gene regions and sequencing them within and between species. We could then estimate, for any collection of populations or related species, how much of their differentiation has been driven by selective sweeps within populations and selective divergences between species. In this way, the second phase of DOBZHANSKY's general scheme could be realized. The only question is whether it is the investigators or the granting agencies that would grow tired first.

The continuation of DOBZHANSKY's program by sequencing studies reveals its original limitation. Although DOBZHANSKY is usually thought of as the founder of experimental population genetics and his 1937 book as the founding document, there is in fact no experiment described there until the last chapter on hybrid sterility, where experimental crosses and backcrosses between *Drosophila pseudoobscura* and *Drosophila persimilis* using marked chromosomes are discussed. The entire body of evidence marshaled on the control of natural variation within populations and the conversion of that variation into genetic differences in time and space is from static data. It depends upon what inferences can be made from the standing patterns of genetic variation in nature. In this case, the testing of hypotheses that we usually associate with experiments is of a special statistical sort, manifest in the 1937 book, in most of DOBZHANSKY's "experimental" (observational) papers, and in the present state of molecular population genetics. By using population genetic theory, either in explicit mathematical form or more heuristically, a prediction is made of what the distribution of genetic variation should look like under some simple model, say no selection and no migration. The observed standing pattern of variation is then compared with this null prediction, and some inference is made from the agreement or disagreement between the observed and the expected. The observations of inversion clines or regional variations in the viability of strains when tested in a standard laboratory condition are, in this respect, of the same evidentiary nature as the comparison of the standing variation within and between species in the ratio of silent to replacement nucleotide substitutions. In 1951 Lamotte attempted to explain the variation in Cepaea by fitting the distribution of colony gene frequencies to a stationary Wrightian distribution. Today, molecular population geneticists fit the nucleotide polymorphism and diversity between populations at several loci to the predictions of coalescent theory. Of course, one can attempt to show that a genetic difference observed in nature has some consequence for physiology and selection in a laboratory model, just as DOBZHANSKY showed that inversions would be subject to selection in the laboratory under some conditions. But the success or failure of such experiments does not tell us what forces have operated historically or are now operating in nature. Population genetics, then as now, is an observational and statistical science, not an experimental one. As a consequence, while it can offer statistical evidence supporting the past action of one or another of the evolutionary forces having operated, it cannot cash these inferences out in the form of actual biological mechanisms.

An irony of the intellectual history of *Genetics and the Origin of Species* is that DOBZHANSKY came into evolutionary genetics from the study of morphological diversity in nature and so was able to relate the abstractions of genetic theory to the biology of organisms, yet in the end he and the field he founded became captives of the abstractions. Despite 40 years of study of the chromosomal variation in natural populations of Drosophila, DOBZHANSKY published no observations from nature on the possible biological mediation of the natural selection he had detected. There is, in the entire corpus of 43 papers on the *Genetics of Natural Populations,* no paper on the ecology and life history of *D. pseudoobscura.* The closest he came was to measure the rate of movement of genetically marked, laboratory-raised flies along a trap line of attractive banana bait. Nor did he make any pretense that the demonstration of selection of chromosomal inversions in laboratory population cages was meant to reveal the natural biology of this polymorphism. The purpose of those experiments was to show that the allelic contents of the inversions could, indeed, under some circumstances make a large difference to their fitness and that, in these circumstances, heterozygotes were more fit than homozygotes, as he believed them to be in nature. In fact, although there was selection in the cages at 25°, there was none at 18°.

Thus, we see that *Genetics and the Origin of Species,* like DOBZHANSKY's subsequent research career, although seeming to speak in the language of organisms, had the ultimate effect not of uniting genetics with the natural historical and physiological biology, but of building a science that speaks the language of gene frequencies. One of the consequences of that alienation of population genetics from organismal biology was the failure of the projects of the 1960s to build a unified science of population biology out of the elements of ecology and population genetics.

Another consequence of the way in which DOBZHANSKY constructed the problem of the origin of species has been to remove the problem of the actual speciation process from the concern of most population geneticists. All of the issues of natural selection in relation to adaptation are, for Dobzhansky, already dealt with in the problem of adaptive population divergence. The final stage of separation of the species becomes a question of the genetics of reproductive isolation (Orr 1997), a problem in neurobiology and developmental genetics, of why flies don't like each other's looks, or why a particular sperm can't fertilize a particular kind of egg, or why somatic or germline development fails in hybrid embryos. This mechanical view of the problem of ultimate species divergence is already contained in Dobzhansky's adherence to a particular view of species. In *Genetics and the Origin of Species* he reaffirms his previous (1935) definition of a species as "that stage of evolutionary process at which the once actually or potentially interbreeding array of forms becomes segregated in two or more separate arrays which are physiologically incapable of interbreeding" (p. 312). By defining it in this way, Dobzhansky then created the problematic for the study of speciation, the genetic elucidation of an aspect of developmental and neurobiology. Where are the genes? What are their develop mental interactions? What determines when and how they are read? And this is, indeed, the current problematic of speciation studies (Coyne 1992), divorced from the rest of evolutionary genetics until such time as population geneticists finally fold developmental biology into their considerations.

What is now the classical definition of species leads to a problem that Dobzhansky acknowledges. What are we to do with all those asexual organisms where this definition of species is irrelevant? He admits that such organisms are not continuously distributed in phenotypic and genotypic space, but that "there are aggregations of more or less distinct biotypes" and that, just like sexually reproducing forms, these biotypes are "clustered around some of the 'adaptive peaks' in the field of gene combinations" and that "the clusters are arranged in a hierarchical order in a way which is again analogous to that encountered in sexual forms." (p. 320). But, he says, they are not species. So what are they, and why do we lavish so much interest on the problem of reproductive isolation? We are never told, because this is the penultimate paragraph in the book. What Dobzhansky has done is to finesse one of the most interesting questions in evolutionary biology, which is why organisms occupy the phenotypic state space in the hierarchically clustered pattern that we see, sex or no sex. That is, how do organisms acquire new and quite distinct ways of making a living? This is the antecedent question that makes the problem of reproductive isolation relevant for sexual species. Whatever the forces are that cluster organisms in state space, that clustering is destroyed by sexual recombination, so an organism that exploits the advantages of sex has a special problem that asexual ones do not have. In order to allow sexual organisms to maintain the clusters against the disruption of sex, they have to develop isolating mechanisms. Those that fail become extinct from too much compromise.

Dobzhansky's construction of the problem of speciation as solely the problem of reproductive isolation was a piece of scientific synecdoche, substituting the process of reproductive isolation for the speciation process in its entirety. It is a testimony to the influence that *Genetics and the Origin of Species* has wielded over 60 years that we continue to study the speciation process without reference to the world that organisms construct and occupy.

LITERATURE CITED

Begun, D. J., and C. F. Aquadro, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster.* Nature **356:** 519–520.

Berry, A. J., J. W. Ajioka and M. Kreitman, 1991 Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. Genetics **129:** 1111–1117.

Cain, A. J., and P. M. Sheppard, 1950 Selection in the polymorphic land snail *Cepaea nemoralis.* Heredity **4:** 275–294.

Coyne, J. A., 1992 Genetics and speciation. Nature **355:** 511–515.

Dobzhansky, Th., 1935 A critique of the species concept in biology. Philos. Sci. **2:** 344–355.

Dobzhansky, Th., 1937, 1941, 1951 *Genetics and the Origin of Species.* Ed. 1, 2, 3. Columbia University Press, New York.

Dobzhansky, Th., and C. Epling, 1944 *Contributions to the Genetics, Taxonomy, and Ecology of Drosophila pseudoobscura and Its Relatives.* Pub. 554, Carnegie Institute of Washington, Washington, DC.

Fisher, R. A., 1930 *The Genetical Theory of Natural Selection.* Clarendon Press, Oxford.

Haldane, J. B. S., 1932 *The Causes of Evolution.* Harper and Row, New York.

Hubby, J. L., and R. C. Lewontin, 1966 A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura.* Genetics **54:** 577–594.

Kreitman, M., 1983 Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster.* Nature **304:** 412–417.

Lamotte, M., 1951 Recherches sur la structure génétique des populations naturelles de *Cepaea nemoralis* (L). Bull. Biol. Fr. Belg., Suppl. **35:** 1–238.

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

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the *Adh* locus in Drosophila. Nature **351:** 652–654.

Orr, H. A., 1997 Dobzhansky, Bateson and the genetics of speciation. Genetics (in press).

Powell, J. R., 1987 "In the air"—Theodosius Dobzhansky's *Genetics and the Origin of Species.* Genetics **117:** 363–366.

Richter, B., M. Long, R. C. Lewontin and E. Nitasaka, 1997 Nucleotide variation and conservation at the *dpp* locus, a gene controlling early development in Drosophila. Genetics **145:** 311–323.

Simpson, G. G., 1944 *Tempo and Mode in Evolution.* Columbia University Press, New York.

Stebbins, G. L., 1950 *Variation and Evolution in Plants.* Columbia University Press, New York.

Wright, S., 1931 Evolution in Mendelian populations. Genetics **16:** 97–159.