Evolutionary Genetics

From Molecules to Morphology

Edited by

R. S. SINGH
McMaster University

C. B. KRIMBAS
University of Athens

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK www.cup.cam.ac.uk
40 West 20th Street, New York, NY 10011-4211, USA www.cup.org
10 Stamford Road, Oakleigh, Melbourne 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain

First published 1999

Printed in the United States of America

Typeface New Baskerville 9.75/12 pt. *System* LaTeX 2_ε [tb]

A catalog record for this book is available from
the British Library.

Library of Congress Cataloging-in-Publication Data

Evolutionary genetics : from molecules to morphology /
edited by R. S. Singh and C. Krimbas.
p. cm.
Includes bibliographical references and index.
ISBN 0-521-57123-5
1. Evolutionary genetics. 2. Lewontin, Richard C., 1929– .
I. Singh, R. S. II. Krimbas, Costas B.
QH390.E96 1999
576.8–dc21 99-14415
CIP

ISBN 0 521 57123 5 hardback

Contents

CHAPTER ONE

The Problems of Population Genetics

richard lewontin

The science of population genetics is the auto mechanics of evolutionary biology.

Organic evolution, in the Darwinian scheme, is a consequence of the conversion of variation among members of an ensemble into differences between ensembles in time and space. Classically those ensembles are collections of individual organisms, populations in the usual sense, but the Darwinian scheme can be applied as well to ensembles of organelles within cells or to collections of populations that make up a species. The essential features of the Darwinian scheme that determine both the subject and form of population genetics are that

1. there are processes that produce variation among individuals within a population,
2. there are processes that result in changes in the relative frequencies of the variants within a generation,
3. there is a hereditary process across generations that may result in further change in the relative frequencies of the different variants, but population frequencies are correlated across generational lines so that the frequency distribution of variants in any time interval is some nontrivial function of their distribution in the previous interval of time.

It is the relation between the processes of the generation and the modulation of variation within generations in features 1 and 2 and the processes occurring between generations in feature 3 that both create the science of population genetics and pose its methodological dilemmas and its shape as an inquiry. These relations are most easily understood as the transformations in the state of the population, shown in Fig. 1.1 (from Lewontin, 1974). The population change can be represented both in a genotypic and a phenotypic space, and the complete laws of transformation make use of both these spaces.

R. S. Singh and C. B. Krimbas, eds., *Evolutionary Genetics: From Molecules to Morphology*, vol. 1.
 ISBN 0-521-57123-5.

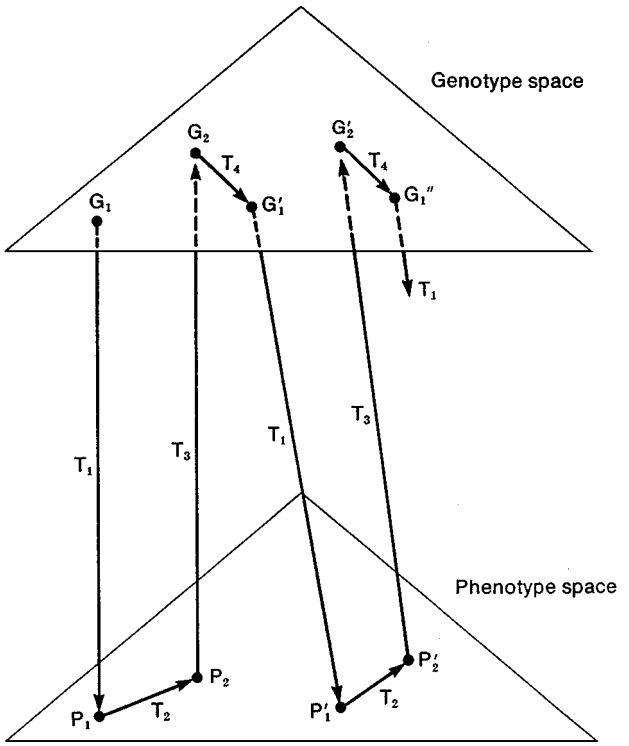

Figure 1.1. Schematic representation of the paths of transformation of population genotype from one generation to the next. G and P are the spaces of genotypic and phenotypic description. G_1, G_1', G_2, and G_2' are genotypic descriptions at various points in time within successive generations. P_1, P_1', P_2, and P_2' are phenotypic descriptions. T_1, T_2, T_3, and T_4 are laws of transformation. Details are given in the text. (From *Evolutionary Genetics* by R. C. Lewontin. Copyright © 1974 by Columbia University Press. Reprinted with permission of the publisher.)

G_1, G_2, P_1, and P_2 are genotypic and phenotypic descriptions at different points within a generation, while the same quantities with primes are the descriptions of the equivalent states in the next generation. The laws of transformation are

T_1: a set of epigenetic laws that gives the distribution of phenotypes, including fitness, that results from the development of various genotypes in various environments.

T_2: the laws of mating, migration, and natural selection including stochastic elements that transform the phenotypic array of potential reproducing units in a population within a generation.

T_3: the set of reverse epigenetic transformations that allows inferences about the distribution of genotypes G_2 corresponding to the distribution of phenotypes P_2.

T_4: the phenomenology of genetics such as segregation, recombination, mutation, horizontal transfer, etc., including stochastic elements, that allows us to predict the probability distribution of genotypes in the next generation produced from gametogenesis and fertilization, given an array of parental genotypes.

This representation of the structure of population genetic inference points immediately to two major sources of difficulty that have shaped the problems of population from its beginnings and still do so. First there is the problem of development. T_4, the complete apparatus of genetics that carries the population transformation across generations, is operating in the genotypic space to transform genotypic frequencies and cannot be framed in phenotypic terms. But T_2 that specifies the transformation of frequencies within generations is a set of physiological, ecological, and behavioral relations that operates in the phenotypic space. In the absence of a knowledge of the epigenetic transformations, the circuit is cut and nothing can be done. Nor can one depend on developmental genetics to provide the missing laws, because the main direction of developmental biology has never been, and shows no promise of ever being, to understand the environmental contingency of variations in development. Much of the past and the present problems of population genetics can be understood only as an attempt to finesse the unsolved problem of an adequate description of development.

The second problem that appears in Fig. 1.1 is the question of rates. Generally the movement in the phenotypic and the genotypic spaces is extremely small, the number of interacting genetic and phenotypic variables is large, and the number of generations and individuals over which observations can be taken is extremely limited. Thus the possibility of inferring the forces that are operating from the differences between G_1 and G_2 or G_1' or between P_1 and P_2 is very remote. This has led to the other major preoccupation of population genetics, the possibility of estimating dynamic forces without ever actually measuring them. To the extent that population genetics has succeeded in this effort, it has provided a uniquely powerful methodology to evolutionary reconstruction that is not available from studies of physiology, ecology, and behavior, whose resolving power is necessarily low compared with the size of the actual forces operating in nature.

Finally, there is a problem that is not so much a difficulty within population genetics, but rather lies in its relationship with other branches of biology and with the rest of evolutionary theory. In these latter fields there is a strong value on finding universals, or at least in claiming universals even when there is no compelling evidence for them. So, even in a historical science like evolutionary biology, an overwhelming emphasis has been placed on selective, and even adaptive, explanations for all phenomena, to the exclusion of other possibilities, partly as a consequence of the belief that a science is validated to the degree that it can make universal claims. But population genetics has developed in a different direction. The schematic given in Fig. 1.1 is one of

the historical trajectories of a character change as a result of the stochastic realizations of the interaction of a large number of forces that alter gene frequencies. Moreover, the strength of these various forces and the signal-to-noise ratio in their dynamics depend on circumstances that are different for different species and different genes. There are no universal claims, or even great generalizations, that can be made about selection intensities, population sizes, migration rates, or about the temporal history of these forces, which are certainly varying in time, or about the amount of genetic variance likely to be present at any moment, on which these forces act. Thus the population genetic research problem is short on general hypotheses to be tested. There are some very large statistical characterizations that can be achieved by repeated observations on a variety of species and genes, and, at the other extreme, there is the possibility of finding examples of the operation of particular forces like balancing selection or neutral evolution, but neither statistical generalities nor individual cases satisfy the model of what a powerful science is supposed to be.

1.1. Coping with Development

Population genetics has taken two opposite paths to solving the problem of reconciling the phenotypic and the genotypic spaces, paths that essentially finesse the problem by operating in only one of them. The first of these, characteristic of investigation before the emergence of molecular biology but now almost entirely abandoned, is to operate entirely in the phenotypic space. Phenotypic observations were used to make rough inferences about the genotypic space, but without providing a detailed description of the population in terms of genotypic frequencies. The tools and concepts were those of biometrical genetics: selection experiments on phenotypic traits in a variety of environments and a variety of selection schemes, estimates of components of genetic variance and covariance for phenotypic characters, studies of norms of reaction of fitness components, observations on fitness components or morphological traits from segregations, and recombinations involving large parts of the genome.

While lacking in any genetic details, these observations and experiments produced an immense richness of information about heritable phenotypic variation in populations and left unresolved a large number of important problems that are untouched by the more recent observations of molecular population genetics. The concentration on a molecular genetic description of population variation has depauperized the problematic of population genetics by marginalizing studies of the phenotype.

A few among the many discoveries of phenotypic population genetics that have yet to be dealt with at the level of the genotype are

1. Populations sampled from nature can usually be artificially selected for almost any morphological character, with some notable exceptions such as directional

asymmetry, and some physiological characters such as alcohol tolerance, while many physiological characters such as temperature and salt tolerance or speed of development are much less selectable. The relevant evidence on genic polymorphism is either in the opposite direction or almost entirely lacking. Enzyme loci are more genetically variable than genes coding for structural proteins, and artificial selection for adult alcohol tolerance in *Drosophila melanogaster* produces no change in allele frequencies at the *Adh* locus (Cohan and Graf, 1985; Weber, 1986). While homeobox genes and other similar loci active in early development have been extensively studied, nothing is known about the loci responsible for continuous selectable variation in either morphological or physiological phenotype.

2. Little is known about the number of loci segregating and the distribution of gene effects for phenotypic traits with heritable variation. Older gene mapping studies in which sparsely distributed markers were used could localize some gene effects to chromosome arms and occasionally localize a large effect. Modern quantitative trait loci (QTL) mapping studies in which much more densely distributed molecular markers are used have a much greater discriminatory power, but it is a long way from localizing the loci responsible for phenotypic differences between two highly divergent lines to giving a description in allelic frequency terms of the standing heritable variation in a population.
3. Characters with no phenotypic variation may nevertheless have considerable underlying genetic variation that is revealed under conditions of extreme developmental stress. Despite the clear evidence that such canalized characters are common and that the degree of developmental plasticity of a character can also be selected for (Waddington, 1960; Rendel, 1967), including, for example, the amount of fluctuating asymmetry in characters that show no directional asymmetry (Reeve, 1960), nothing is known at the genotypic level. Some glimpse of the possibilities is offered by the recent finding of Gibson and Hogness (1996) that selection for genetic assimilation of the bithorax ether phenocopy is accompanied by a change in the frequency of sequence variants at the *Ubx* locus.
4. Many correlations between morphological characters or between morphology and fitness are a consequence of linkage disequilibrium, but some are not or at least have not been broken by recombination and selection. The entire phenomenon of the resistance to artificial selection by countervailing fitness (usually fertility) effects needs to be investigated at the genotypic level. What is the explanation, for example, of the appearance in the classic selection experiments of Mather and Harrison (1949) of balanced sterile lines at intermediate levels of the selected phenotype so that no selection progress could be made in either direction, despite the presence of genetic variance for the character?
5. There are constraints on the pathway that phenotypic evolution can take under selection. As shown, for example, by Hall (1982) for the *ebg* locus in *Escherichia coli*, the same selected phenotype may have different genotypic bases because of different phenotypically similar mutations, some of which allow for further selectable mutations while others are dead ends for selection because the subsequent one-step mutations do not give appropriate phenotypes.

The mapping of genotypic mutational space onto phenotypic space is a major task for population genetics if an adequate genetic explanation of phenotypic selection and evolution is to be given.

The alternative solution to the problem of development has been to study characters for which there is no morphogenesis, that is, to study the genome directly. This approach is an attempt to collapse the two spaces in Fig. 1.1. It is this ploy that is the contribution of molecular population genetics, beginning, imperfectly, with electrophoretic and immunological studies of proteins and ending with the complete genotypic information contained in DNA sequences. Electrophoretic studies eliminate much of the problem of morphogenesis but there remains a residual ambiguity about genotype because even the most discriminating sequential electrophoresis will not distinguish some amino acid substitutions and there is no information on which or how many amino acid differences separate two electromorphs. There is, of course, no information about silent substitutions. Restriction enzyme studies are completely ambiguous about amino acid substitutions but at least provide information on silent sites. It is only with complete DNA sequences that a complete collapse of the phenotypic into genotypic description is possible.

The euphoria that has accompanied our ability, finally, to give a complete and unambiguous description of the allelic composition of a natural population for some arbitrarily chosen piece of the genome has hindered us from seeing that the problem of epigenesis has not been eliminated. While it is possible now to give a genotypic description of the variation in a population, the scheme of evolutionary transformation in Fig. 1.1 still requires phenotypic information on whole organisms. Formally, what DNA sequencing does is to allow a complete description of G_1 and so we have the illusion that we do not need the morphogenetic rules T_1. But the rules of transformation T_2, the rules of mating, migration, and natural selection, are rules about phenotypes. Gene sequences do not mate, they do not migrate, they do not live or die differentially, except as a consequence of the physiology and the metabolism of their organismic carriers in interaction with the physical and biotic environment. To eliminate totally the problem of morphogenesis from our observations, we must avoid all phenotypic descriptions above the level of genotype, including fitness. A direct attempt to measure the demography and the fitness of genotypes once again introduces the contingencies of development. The epigenetic transformation T_1 is still needed to predict the distribution of phenotypes P_1 that are transformed into the selected phenotypes P_2, and we still require T_3 for carrying the phenotypes back into the genotypic space.

The problem of directly measuring the fitness of genotypes has been considerable, and the results not generally encouraging. The direct estimation of net fitnesses of genotypes in nature and the discrimination of fitness differences from nonrandom mating, segregation biases, migration, and stochastic elements demand the possibility of observing mating pairs and associating their offspring with them, so that fertility as well as viability estimates can be made (see Prout, 1971, and Christiansen and Frydenberg, 1973, for the general

methodology of such estimation). The most favorable material would be annual seed plants or an animal with single clutches that can be associated with parental pairs like birds or ovoviviparous fish. The most complete and sophisticated attempt to measure all the components of the life cycle in nature for genotypes has been, in fact, the work of Christiansen et al. (1973) on the esterase enzyme polymorphism in an ovoviviparous fish, *Zoarces*, but no strong fitness differences were found. Nor should this be very surprising. Given the immense amount of genotypic polymorphism present in natural populations, it is unlikely that large fitness variance can be associated with many individual polymorphic loci, and even when there are cases of strongly selected polymorphisms, it is unlikely that an arbitrary polymorphism, chosen for its ease of observation, will be one of these cases.

As substitutes for net fitness measurements, there have been attempts to measure the physiological properties of defined genotypes and then to argue that such properties are manifest as fitness differences in nature. An example of this approach is the work on the phosphoglucose isomerase enzyme polymorphisms in Colias butterflies (see Watt, 1991, for a review), in which it has been demonstrated that activity differences associated with different alleles appear to cause differences in male flight ability, which in turn must influence male mating success. The question that remains open, however, and a deep difficulty of all such work on fitness in nature, is that this is a *ceteris paribus* argument: "All other things being equal," male mating ability is an important component of fitness, but we do not know whether male mating ability is, in fact, an important component of the standing variance in fitness in nature. The problem is similar to the situation of melanism in *Biston betularia*, in which melanic forms on dark backgrounds are certainly less likely to be taken by predators, but it is by no means clear that predation pressure is sufficiently high to represent an important cause of differential mortality. Moreover, the laboratory demonstration of fitness differences associated with allelic variation makes use of deliberately constructed stress conditions that may be irrelevant in nature. The polymorphic alleles at the *Adh* locus in *D. melanogaster* have very different activities on an ethanol substrate and are selected in the presence of toxic levels of ethanol in the medium (Bijlsma-Meeles and Van Delden, 1976). These ethanol levels, however, are above any known in nature, including those in fermentation sheds in wineries (McKenzie and Parsons, 1974; McKenzie and McKechnie, 1978). The classical work of the Dobzhansky school on fitness in *D. pseudoobscura* was almost entirely in terms of larval competitive viability, yet in nature larvae of this species are never found in crowded conditions, and it seems more likely that the differential ability of females to find an egg-laying site and the ability of larvae to pupate before the small food sources dry up are much more important determinants of fitness variance. The problem of directly measuring fitness differences in nature is one that has plagued population genetics for nearly a century, and our ability to identify genotypes at the DNA level has not made it disappear.

1.2. The Problem of Rates

The difficulty of measuring fitness differences in nature is one facet of the more general issue that the individual forces operating on genotypic composition of a population at a locus are small so that the change from one generation to another, even the real stochastic change, is usually well below the observable limit in samples of feasible size. The question then is how we are to reconstruct the dynamic process in Fig. 1.1 when G and G' are not observably different. The answer has been to invert the process of deduction and to appeal to the cumulative effects of weak forces over long periods. From theoretical considerations we can predict the probability distribution of changes in genotypic composition under various models from one generation to the next. If the forces have remained constant for a very long period, then the population genotypic composition will come to a steady state under these forces, and, if we are lucky, that steady state will bear the unique signature of the forces, making an inference about them possible. We can use, for this purpose, not only the genotypic composition of a given population at a given locus, but the variation among loci, among populations, and among closely related species. Such synchronic or static data replace the direct measurement of the forces, and the direct diachronic observation of temporal change as the preferred method of population genetics.

Inference from static data has a long history in population genetics. In its simplest form, an observed intermediate allelic frequency of a polymorphism could be taken as evidence of balancing selection or of an equilibrium between mutation and selection. In the case of sickle-cell hemoglobin, in which the homozygote S/S is lethal, the inferred mutation rate would have to be of the order of 10^{-1} in West Africa but virtually absent elsewhere to explain the facts of the polymorphism. Clearly the postulate of balancing selection would make more sense even in the absence of the geographical correlation with malaria. The use of B/A and D/L ratios in the 1950s and 1960s were attempts to make inferences about the dominance of deleterious fitness effects when these effects could not be directly measured (see Lewontin, 1974, for a review of these and other similar methods). But the wholesale use of static data as the bread and butter of population genetics has depended, first, on the provision of a rich and precise genotypic data set by molecular survey methods and second, on the development of an articulated stochastic theory of gene frequency evolution that can, in principle, predict different static signatures for different mixtures of forces.

The first attempts to match stochastic theory to molecular data involved the use of the Ewens (1972) and Watterson (1977) distributions to test whether the frequency distributions of electrophoretic alleles at multiallelic loci like *Est5-B* and *Adh* in *Drosophila* were compatible with nonselective processes, with balanced polymorphism, or with purifying selection (Keith, 1983; Keith et al., 1985). This method assumed, as do all subsequent forms of the same approach, that

1. the experimental method distinguishes all genotypes,
2. the forces of selection and effective population size have been constant for a long time so that

3. the current allelic frequency distribution is at steady state, and
4. the current allelic frequency distribution is independent of the initial conditions.

These earlier attempts were not continued partly because the method of electrophoresis does not fulfill requirement 1 and partly because the tests of Watterson and Ewens turned out not to have sufficient power for such data sets to distinguish many selective hypotheses from nonselective hypotheses. In the case of *Est-5b* the data were not compatible with any of the alternatives, suggesting that the steady-state assumption was wrong or that a more complex selective hypothesis needed to be invoked. It soon became apparent that electrophoretic data simply did not have the necessary structure to distinguish hypotheses among forces. In the first place, most polymorphic loci are, at the electrophoretic level, biallelic or triallelic, and there is usually insufficient evidence in the frequencies of two or three alleles to test most hypotheses. There are exceptions. For example, the repeatable altitudinal clines in the two common alleles of the *Adh* (Grossman et al., 1969) and similar geographical clines in different continents of *Gpdh* and *Tpi* electromorphs (Oakshott et al., 1984) make a selective rather than historical explanation extremely likely. Second, even for the occasional locus like *Xdh* or *Esterase* in *Drosophila*, which has 20 or more alleles present in a single sample in which the pattern is of one or two common alleles and a large menu of rare alleles, extreme similarity of distribution between different populations can be explained either as repeatable selection or as a consequence of a small amount of migration between the populations. Third, the bulk of loci that are electrophoretically homozygous provide no information at all about the causes of their homozygosity. It is this problem that lies at the crux of the difficulty, a problem that has been solved by a fortunate quirk of nature.

The problem of genetic identity is to distinguish two possible cases for apparent allelic identity. It may be that two gene copies are identical because they stem from a recent common ancestor, so recent that no mutations have appeared since their genealogical splitting. That is, they are identical by descent. The other possibility is that their most recent common ancestor was indeed very ancient and that many mutations have occurred since that ancestor, but these mutations have been selected against, leaving only the allowable gene form. That is, the present genes are identical by state. Nor does this distinction apply at only a completely monomorphic locus. If there are two alleles in intermediate frequency at a polymorphic locus, one might be the ancient form kept uniform by a constant purifying selection, while the other might be a recent mutation that is selectively favored and has spread rapidly since its origin, having reached a polymorphic equilibrium with the older allele or being in the process of replacing it. On the other hand, both alleles may have been around for a very long time by chance because both are acceptable by natural selection while all other mutations have been rejected. The difficulty is that two alleles do not carry with them any information about how long they have been separated, and this is true whether they are identified by electrophoresis or by total amino acid sequencing. If only alleles also had temporal information that allowed us to distinguish identity by recent descent from enforced identity by

state. But, of course, they do as a consequence of a lucky quirk of nature: the redundancy of the DNA code.

Alleles that are identical in the amino acid state may nevertheless differ in their nucleotides and if, as a first order of approximation, natural selection does not distinguish among alternative codons for the same amino acid, then gene copies can accumulate silent nucleotide substitutions unimpeded by natural selection. The number of such silent differences is then an index to the time that has passed since their common ancestor and can be used to distinguish among competing theories. It is vital to understand that DNA sequencing as the method of choice for population genetics does not stem from the fact that somehow DNA is a more basic or fundamental state description than amino acid sequences, but from the many–one correspondence between DNA sequences and amino acid sequences. Had the relation turned out to be one–one, there would be no advantage to DNA sequencing except as a cheap way to determine protein sequence.

More broadly, it is the presence, side by side in the same short gene region, of coding sites together with silent sites, with introns, and with untranslated and nontranscribed flanking DNA, all of which are subject to the same temporal processes of mutation and reproduction, but that differ from one another in the physiological consequences of their variation, that allows us to distinguish selective from purely historical similarities and differences. While the rationale for DNA sequencing is often stated to be that silent sites in codons are selectively neutral and so can serve as a selection-free molecular clock against which amino acid substitutions can be calibrated, that neutrality is not necessary (fortunately, since it is not true). All that we require is that different DNA positions have physiological constraints operating at different levels of intensity and in different contextual situations, so that the static variation observed within and between populations bears the distinguishable signatures of different temporal processes.

The most striking use of the temporal information in DNA sequences and the one with the greatest implications for our understanding of dynamical processes was, in fact, the first application of the method by Kreitman (1983) in his study of the standing variation at the *Adh* locus in *D. melanogaster*. Except for the single Fast / Slow (F / S) electrophoretic polymorphism that was deliberately introduced into the sample by the choice of lines, there was not a single amino acid polymorphism in a geographically very diverse sample, yet silent sites and introns were approximately 6% polymorphic (14% in exon 3). Thus the identity of the amino acid sequences could not be the result of recent common ancestry, but must be the outcome of a selection process in which all amino acid replacements, except the single widespread F / S polymorphism, have been removed by selection.[1] This result was made even more striking by that of Schaeffer and Miller (1992), which showed that in a sample of 99 genomes of *D. pseudoobscura*, there is no amino acid polymorphism at all at *Adh*, with the exception of one line with an isoleucine / valine change, despite the fact that there is ~6% silent nucleotide polymorphism. The implications of these results

for our understanding of natural selection are powerful. Three-quarters of all random single nucleotide changes in coding regions will cause an amino acid substitution, yet every one of these has been screened out by natural selection in a protein that is, for example, 28% leucine/isoleucine/valine. While selection differences need not be large for any substitution, since population sizes for these species are estimated to be, conservatively, in excess of 10^5, selection has been discriminating enough to weed out every substitution. The problem is how every amino acid substitution can make a physiological difference that is ultimately translated into an average difference in viability and fertility. Nor is this discriminatory power of natural selection confined to a small enzyme like alcohol dehydrogenase. When similar studies were done on the very large and electrophoretically very polymorphic enzymes xanthine dehydrogenase (Riley et al., 1992) and esterase-5b (Veuille and King, 1995) a similar but less extreme result was found. Approximately 90% of all amino acid substitutions in xanthine dehydrogenase and 85% in esterase-5b have been rejected by natural selection.

If these results are indeed general, then a major problem has been posed for cellular physiology and metabolism and for evolutionary biology. Given environmental contingency and the accidents of development, how can the amino acid composition of proteins be translated so exquisitely into differential fertility and viability of individuals? If that translation does occur, how can the amino acid compositions of enzyme proteins change by multiple amino acids during speciation, as they indeed have?

Of course, speciation may be accompanied by drastic reduction in population size so that Ns is temporarily small and random fixations of amino acids differences may occur. But that cannot be the whole story. As pointed out by McDonald and Kreitman (1991), neutral fixations during speciation should reproduce between species the same ratio of silent to replacement substitutions as are polymorphic within species. Yet in some cases, for example glucose-6-phosphate dehydrogenase in *D. melanogaster* and *D. simulans* (Eanes et al., 1993) there is a great excess of amino acid differences compared with the intraspecific polymorphism.

The McDonald–Kreitman approach to detecting selectively driven divergence between species is another opening into an old problem made possible by the redundancy of the DNA code. It is widely recognized by systematists, even those trained to avoid typologies, that many characters that distinguish taxa have no detectable variation within species. The genus *Drosophila* is distinguished from other members of the *Drosophilidae* by the possession, by every individual in every species in the genus, of one proclinate and two reclinate orbital bristles. Given the lack of intrataxon variation, where did the variation for taxonomic differentiation come from and where did it go after the event? Many more or less plausible stories can be told, but there is no available method for demonstrating that natural selection was involved. For nucleotide sequences the possibility of implicating natural selection exists because we can make internal comparisons of polymorphism and species divergence for nucleotide positions of different function.

Regional heterogeneity in the density of polymorphism along a DNA sequence can arise for reasons other than the direct constraints on the sequence itself because of the extremely tight linkage among the nucleotides in short genomic sections. Thus a region of unusually high nucleotide variation can be taken as evidence that random fixation of mutations has been resisted by some form of balancing selection. Kreitman and Hudson (1991) used this approach to infer heterosis for the F/S electrophoretic alleles of *Adh* in *D. melanogaster*, although their data show an excess nucleotide polymorphism even within the Slow allelic class, so that something else besides heterosis of the two electrophoretic forms must also be happening.

In contrast, extremely low variation in a genomic region might be the equilibrium signature of a high functional constraint on the region, but it might also be the historical leftover of a recent rapid replacement of one or a few sites that carried adjacent nucleotides to fixation by hitchhiking. This explanation by Berry et al. (1991) of nucleotide homogeneity of the fourth chromosome must surely be the correct one. Such a rapid fixation, however, might be the result of a selectively advantageous mutation's having been fixed, a so-called selective sweep, but it might also be the consequence of a segregation distortion. Because the segregation distorters (SDs) that have been studied are necessarily those that are in some sort of balanced equilibrium in populations, like the SD in *Drosophila* and the *t locus* in mice, we do not know how often distorters, unopposed by selection, may sweep through populations. Whether or not one can distinguish selective sweeps from distorter sweeps depends on the rapidity with which the two processes occur and on the time since the event. In the absence of any ancillary evidence about replacement speed and relaxation time, it is hard to see how these processes can be separated.

The problem of distinguishing selection from segregation distortion applies as well to the inference of adaptive evolution between species. Both are illustrations of a deep structural problem in making inferences from static variation. Such data are, at the first level, information on the times of divergence from a common ancestral sequence, in mutational units, between genomes in populations. But the distribution of divergence (coalescence) times is a dual representation of the distribution of rates, in mutational units, of fixation of particular sequences through the population. An unusually low fixation rate of some stretch of genome compared with the average for the genome may mean that there are forces holding sequences in equilibrium against random fixation or that the time scale for the production of variation in the region is more compressed. Such a compression can come from a local excess in mutability or from a migration into the population of genomes that were divergent in this genic region.

In a like manner, a deficiency of variation means either that some genome has spread unusually rapidly through the population, for whatever reason, or that the time scale for the production of variation has been extended because of selective constraints. The decision from among these alternatives and the biological interpretation of the causes of the changes in time rates must be made on other grounds not contained in the static data themselves.

The discovery by molecular biology of different functional classes of nucleotides has not only led to the possibility of obtaining strong evidence about old problems, but has revealed a new set of problems for population genetics, problems that arise from considering the standing variation within those nucleotide classes. These are the problems of explaining regional heterogeneity in DNA sequences even within functional types. Thus, at the *dpp* locus in *Drosophila*, a gene that sits at the center of many of the processes of early embryonic development, there is a major heterogeneity in the pattern of nucleotide polymorphism in the large intron. There is a concentration of polymorphism at the two ends of the intron and little variation in the middle of the intron both within and between species (Richter et al., 1997; Newfeld et al., 1997). Some of this conservation of the middle of the intron coincides with multiple short repressor motifs in the region, but much of it remains to be explained on functional grounds. There is also a region of perfect conservation within and between species in the 3′ untranslated region of Intron 3, in the middle of a region of considerable nucleotide divergence and polymorphism. What are the functional sources of such conservations and how frequent are they?

One of the most interesting and general unresolved problems arising from observations of genomic heterogeneity is the question of codon bias. A reasonable explanation of unequal codon usage is that, especially for highly transcribed genes, unequal availability of tRNA species constrains the codon usage to match it for the most efficient translation. But this explanation raises many problems when different species and different genes are compared. For example, the AUA codon is totally avoided for isoleucine in *Adh* in all species of *Drosophila*, yet for other genes in *D. melanogaster*, it is used on average ~5% of the time. In *D. virilis*, however, it comprises 38% of the isoleucine codons for the *dpp* gene. Moreover, although a clear bias toward G/C ending alternative codons is general for genes in *D. melanogaster*, in *D. virilis* the *dpp* gene is biased toward A/T in eight amino acid groups. Have there been major changes in tRNA availability from one species to another, and, if so, why do different genes in the same species not agree in their codon bias differences from other species? The existence of specific usages of gene and species make simple explanations of the evolution of codon usage doubtful.

The issue of codon usage raises a problem of constraint on DNA sequences that is not ordinarily considered. The secondary structure of message, its stability, and the time course of its translation will be influenced by the nucleotide sequence. We should not suppose that a maximum rate of translation is necessarily optimal for the translation process. The final folding of a protein occurs as a consequence of the formation of folding intermediates during the process of translation, and, presumably, some slowing down and speeding up of chain elongation at appropriate places is necessary for a realization of local free-energy minimizations during folding. These variations in speed will be affected by the particular DNA sequence, partly in its relation to the availability of tRNA, so that the problem is more complicated than that of simply maximizing the use of common tRNA species. Nor should we even suppose that fidelity of

translation is an unmitigated good. The translational process, like any molecular mechanism, makes mistakes. There must be a population of variant protein molecules in the cell. Are these variants of some physiological use? After all, if one argues that heterosis comes from the advantage of having two slightly different copies of a molecule in each cell, then why not have more than two different copies and in unequal numbers? But all of these considerations lead us to suppose that the DNA sequence itself, even in coding regions, may be constrained over stretches of various length. It is the common assumption that constraints on amino acid variation are the consequence of requirements on the physiological function of the protein in question. But if we suppose that DNA sequence itself is constrained over even short stretches, then the amino acid sequence will be held constant as a consequence of the constancy of the underlying nucleotide sequence. At the moment we do not know how much of the variation and the conservation of protein sequences is a secondary consequence of requirements on the nucleotide sequence itself. Nor do we know how to find out.

There is yet another possibility opened by data from DNA sequence that returns us to a previous stage in the history of inference. The attempt to make judgements about forces from static data on electrophoretic variation foundered for three reasons: ambiguity about the actual genetic identity of the observed classes, insufficiently rich data sets even for the few multiallelic loci, and a stochastic theory that was derived from the consideration of the moments of stationary distributions rather than the richer structure of gene phylogenies. The first two problems are dealt with directly by the nature of DNA sequence data and by an immense amount of haplotypic diversity that appears even in modest data sets. The third problem has been attacked by developments in stochastic theory that take into account the details of haplotypic phylogenies, although still suffering from the need to make many simplifying assumptions (Hudson, 1983; Hudson et al., 1987; Tajima, 1989). Essentially these developments attempt to compare the observed distribution of DNA haplotypic differences within a sample with those expected from purely neutral evolution. The null hypothesis that produces the expected distribution of differences among haplotypes consists, necessarily, of a long list of assumptions, so what is being tested when the data are fit to the null hypothesis is the conjunction of a large number of neutral assertions. Some of these are background assumptions about population history and linkage that are not of primary interest in the test. Others are relevant to whether natural selection needs to be invoked to explain the observed pattern of differences. For example, the method given by Hudson et al. (1987) involves the differences between two species for two or more genomic regions and the polymorphism within these regions. The background assumptions are that generations are discrete, that mutations at different sites occur independently and with a common mean for sites within a locus, that the species are at a stationary state, that they diverged from an ancestral population whose population size was the average of the stationary sizes of the current species, and that there is complete

linkage within gene regions but free recombination between them. The selectively relevant assumptions are that the segregating mutations that are actually segregating within the species are selectively equivalent, that the degree of purifying selection may differ between regions but is the same between species, and that no change in selective regime occurred in the process of speciation. We do not know, in general, how deviations from the background assumptions affect the power of the test compared with the selective assumptions, nor do we know, in general, what the power of the test is to detect one sort of selective deviation as opposed to another. Thus we do not know at present how to interpret observations of significant or nonsignificant test results. In practice, such tests can be subject to an extensive investigation of their operating characteristics. Simonsen et al. (1995) have shown for the simpler test of Tajima (1989) that the power to detect selection is low, a result that is reminiscent of the earlier situation of similar tests on electrophoretic data. The important question is less whether the tests are powerful in any absolute sense, but what their relative power is to detect deviations from the background assumptions as opposed to deviations from the selective assumptions. Tests of low power are only useless. Tests that powerfully detect the wrong deviations are destructive.

1.3. Problems of Generality

Twenty years of electrophoretic surveys and 15 years of DNA sequencing and restriction analysis make it clear that there is a lot of standing genetic variation in populations both at the amino acid level and the DNA level, but there is a great deal of difference among species and gene regions in how much genetic variation is to be found. It is also possible from theoretical considerations to make some predictions about correlations, as for example that regions of the genome with low recombination should turn out to be regions of lower genetic variation, all other things being equal, as indeed has been observed for a couple of chromosomal regions in *D. melanogaster* (Berry et al., 1991; Begun and Aquadro, 1992; Martin-Campos et al., 1992). How well this correlation holds in other organisms with different demographic histories remains to be seen.

The alternative to statistical generalization is the exemplification of particular phenomena without any strong implication about their frequency in nature. So, for example, a convincing case has been made that the *Adh* electrophoretic variation in *D. melanogaster* is a balanced polymorphism (Kreitman and Hudson, 1991), but, unlike the universalist claims about heterosis that were common in population genetics in the 1950s and 1960s, it has never been suggested that *Adh* in *D. melanogaster* provides any general insight into genetic variation in general. In like manner, the strong case made by Eanes et al. (1993) that *G6pdh* diverged selectively in *D. melanogaster* and *D. simulans* is not taken as a general refutation of claims of the neutral evolution of amino acid sequences. But producing an example, or even several, of selective divergence, neutral evolution, balanced