CHAPTER 4 / THE GENETICS OF SPECIES FORMATION

It is an irony of evolutionary genetics that, although it is a fusion of Mendelism and Darwinism, it has made no direct contribution to what Darwin obviously saw as the fundamental problem: the origin of species. I do not mean to say that the theory and observations of population genetics have not influenced and even permeated theories of species formation. One has only to read Dobzhansky's *Genetics and the Origin of Species* (1951) or Mayr's *Animal Species and Evolution* (1963) to see how population genetics informs modern ideas about speciation. Mayr's whole thesis, that speciation is the tearing apart of a group of genetically interconnected populations, normally held together by strong cohesive forces of migration and natural selection, comes directly out of the balance school of population genetics, which he has assimilated into the theory of geographic speciation. But it is a long way from describing speciation in general genetic terms to constructing a quantitative theory of speciation in terms of genotypic frequencies. While it is a question of elementary population genetics to state how many generations will be required for the frequency of an allele to change from q_1 to q_2, we do not know how to incorporate such a statement into a speciation theory, in large part because *we know virtually nothing about the genetic changes that occur in species formation.*

How much of the genome is involved in the early steps of divergence between two populations, causing them to be reproduc-

tively isolated from each other? We do not know. Mayr writes of a "genetic revolution" in speciation, but we cannot put quantitative limits on this revolution (which may after all turn out to be only a minor reform) until we begin to characterize the *genetic* differences between populations at various stages of *phenotypic* divergence. The whole concept of a genetic revolution arises out of the undoubted truth that every gene affects every character, that "no gene frequency can be changed, nor any gene added to the gene pool, without an effect on the genotype as a whole, and thus indirectly on the selective value of any other genes" (Mayr, 1963, p. 269).

But it does not follow that every gene substitution really matters. It may be true "that thou canst not stir a flower without troubling of a star," but the computer program for guiding a space capsule does not, in fact, have to take my gardening into account. General principles are not the same as quantitative relationships.

The problem of making quantitative statements about the multiplication of species has been that we have been unable to connect the phenotypic differentiation between populations, races, semispecies, and species with particular genetic changes. It is our old problem in a new context. The "stuff of evolution," the subtle changes in cell physiology, developmental pattern, behavior, and morphology that lead to reproductive isolation and ecological differentiation are the observables, but the only entities for which we have been able to construct a dynamic theory are genotypes. Until we are able to specify the genotypic differences between populations at various levels of phenotypic differentiation, we will not have the beginnings of a quantitative genetic theory of speciation. Even when we are able to measure genotypic differentiation, however, it will be only a beginning, since in the end we must know in what way particular genetic differences are related to the particular reproductive and ecological properties that separate two species.

The general theory of geographic speciation postulates a multistage process after an initial geographical isolation. Populations must become geographically isolated from each other because even a small amount of migration will prevent genetic differentiation between populations unless some extraordinarily strong selection virtually fixes different alleles in different populations. Although it is theoretically possible for disruptive selection to lead to reproductive isolation without geographical isolation, Mayr makes a strong and convincing case that such a phenomenon must be rare in nature.

Thoday and Boam (1959) and Thoday and Gibson (1962) succeeded in producing stable sexual isolation by disruptive selection without isolation, but some unusual genetic condition must have underlain their result, since several attempts to produce the phenomenon again have failed (for example, Scharloo, 1971). Attempts to produce sexually isolated populations within a species by selecting against hybrids between them have succeeded, but the isolation that appeared during the course of the selection disappears if the populations are again allowed to interbreed (Wallace, 1954; Knight, Robertson, and Waddington, 1956). If there is any element of the theory of speciation that is likely to be generally true, it is that geographical isolation and the severe restriction of genetic exchange between populations is the first, necessary step in speciation.

The first stage of speciation is the appearance, after some period of isolation, of genetic differences that are sufficient to restrict severely the amount of gene exchange that can take place between the populations if they should again come into contact. In addition, there is some divergence in ecological niche and, indeed, this divergence might be the direct cause of the reproductive isolation. Changes in feeding preference, pattern of diurnal activity, breeding season, and so on could lead to micro-temporal or micro-spatial overlapping. The question of the relation between ecological divergence and reproductive isolation is open. How often does reproductive isolation arise "accidentally," even in the absence of any other differentiation? How often is it simply an indication of a general genetic divergence for basic developmental patterns? How often is it the direct consequence of ecological divergence? We do not know, because the genetic analysis and ecological study necessary to answer these questions have not been, and for the most part cannot be, carried out on numerous populations in the early phases of the speciation process.

The second stage in speciation occurs if and when the isolated populations come into contact again. If sufficient reproductive isolation of the proper sort developed while the populations were allopatric, there may be a reinforcement of the reproductive barriers by natural selection. If, during isolation, physiological differences have arisen that cause hybrid offspring to be less viable or fertile, then individuals that mate heterospecifically will leave fewer genes to future generations, and there will be selection for characters that reduce the amount of heterospecific contact. There is an interesting

quantitative problem here. How strong must the reproductive barrier be between the two newly sympatric populations in order that the differences they have accumulated in isolation from each other not be swamped out by the crossing between them? The theory of geographic speciation rests on the assumption that even a little migration between populations is enough to prevent them from speciating in the first place. Hence the requirement for geographical isolation. Does this mean that even a little gene flow between populations newly come together will destroy their differentiation, so that successful speciation demands the virtually *complete* rejection of genes flowing from one population to the other?

During the second stage there will also be selection for ecological divergence between the forming species, at least until they have the minimum niche overlap necessary for stable coexistence with each other and with the other species in the community. Although the principle that newly formed species must diverge ecologically from each other has long been asserted and was given an explicit form in Brown and Wilson's concept of character displacement (1956), the foundations of a quantitative and predictive theory of the limiting similarity between coexisting species have only recently been laid by MacArthur and Levins (1967). It is important to realize that selection for niche divergence and selection for secondary sexual isolation are not necessarily independent phenomena, but that, on the contrary, divergence in life habits may in itself usually lead to sexual isolation by reducing temporally and spatially the opportunity for the union of gametes from the two newly formed species.

There is yet a third stage in the evolution of species, which is not generally thought of as part of the speciation process. The newly formed species continue to evolve, but not in response to each other. Each becomes simply a part of separate communities of species undergoing phyletic evolution, splitting, and extinction, with no special relation to each other as the offspring of a common ancestor. Then the degree of genetic similarity between contemporaneous representatives of the two new phyletic lines reflects only the total passage of time since the original split, and the average speed of evolution of a more or less long series of evolutionary episodes.

The importance for us of distinguishing the three stages of species divergence is that the genetic questions involved at each stage are very different and will be totally confounded if we ask only What is

the genetics of speciation? or What is the genetic difference between two species? At the first stage we want to know how much and what kind of genetic differentiation is required for primary mechanisms of reproductive isolation to arise. Is it true in general, as Mayr believes, that a genetic revolution occurs as a result of a reduction in population size, followed by selection of quite new genotypes, and that reproductive isolation is the incidental concomitant of this revolution? Or is a very small fraction of the genome involved, so that at the end of the first stage much the greatest part of the genome is unchanged? Even if there are substantial changes in the greater part of the genome, is the reproductive isolation a result of differentiation of a few loci only?

At the second stage we want to know how much more genetic divergence must occur to produce ecologically differentiated, stable members of the species community. If only minor differentiation occurs during geographical isolation, the major portion of genetic diversification during speciation may occur during the second, sympatric stage when natural selection is operating to push the new entities apart. Alternatively, stage two may be only a fine-tuning process in which the finishing touches are put on an already accomplished speciation.

Finally, the questions about the third stage are really questions of genetic systematics. That is, if we consider species judged by the criteria of evolutionary systematics to be more closely and less closely related, and therefore more recently and less recently diverged from a common ancestral population, how much genetic similarity is there between them? What is the rate of independent genetic divergence in absolute and "taxonomic" time?

For historical and pedagogical reasons it is better to take up these questions in reverse order—to begin with the genetic differences between species, about which we know something, and to end with the first steps in speciation, which are almost totally obscure.

GENETIC DIFFERENCES BETWEEN SPECIES

Once again we are faced with the methodolocial contradiction. Species are, by definition, reproductively isolated from each other, but genetic investigations are carried out by making crosses. The contradiction is not absolute, because there do exist some pairs of

species which, although completely isolated reproductively in nature, can be made not only to hybridize, but even to yield F_2 generations and backcrosses in the laboratory or garden. It is nevertheless true that as long as doing genetics meant making successful crosses, investigations of the genetics of species differences had to be restricted to very closely related, partly compatible species. The problem is less severe in plants than in animals because isolating mechanisms in plants are more often meiotic than developmental, so that if enough crosses are made, a few seeds will be set and these can be grown for further analysis.

Most of the analyses of genetic differences between species (see Mayr, 1963, p. 543; Dobzhansky, 1970, pp. 261–63) have concentrated on the morphological characters that are used to differentiate them. In a few cases a single gene substitution with some modifiers may account for a clear-cut species difference. More often there appear to be many, but an unknown number of, gene differences involved. In any event, it is impossible to estimate what proportion of the genome differentiates the species.

A second major line of investigation has been on the nature of the sterility in hybrids between species (see Dobzhansky, 1951, chapter 8). Nearly all of these analyses have been at the chromosomal level, showing that the haploid sets from the two parental species fail to pair properly at meiosis, or that there is an incompatibility between the chromosomes of one species and the cytoplasm of another, or that sex determination is disturbed by mixtures of sex chromosomes or autosomes from different species.

Perhaps the case that was best worked out, given the limitations of the methods available, was Dobzhansky's analysis of male sterility in the hybrids between *Drosophila pseudoobscura* and *D. persimilis* (1936). Crosses between these two sibling species are easy to make and produce large numbers of fully fertile F_1 females and completely sterile F_1 males. When the cross is made of *D. persimilis* ♀ × *D. pseudoobscura* ♂, the sterile F_1 males have testes about one-fifth normal size, while the reciprocal cross gives sons with normal-sized testes, although these males too are completely sterile. Backcrosses of F_1 females to either parental species are possible, and the backcross sons have testes of varying size and are variously fertile, testis size and fertility being correlated. By means of marker chromosomes, Dobzhansky was able to show that there are at least

two genes on each of the large chromosomes influencing testis size and that an interaction between the sex chromosome of one species and the autosomes of the other was a predominant effect.

Whether they are concerned with morphological differences or sterility, such studies are essentially attempts to work out the genetics of complex quantitative characters. Even under the best conditions, within a species that has an excellent array of marker genes and chromosomal aberrations available as tools, the genetic dissection of a quantitative character is a difficult and somewhat slippery affair, and the results, although framed in terms of "genes," are really in terms of chromosome segments.

For species crosses in which generations beyond the F_1 are difficult to produce, in which there is a paucity of mutant markers, and in which many of the critical genotypes may barely survive, if at all, one can usually do very little beyond assigning influences to different chromosomes. Moreover, no matter how fine the analysis, the dissection of a particular morphological difference between species, or even of the sterility barrier between them, is not really an answer to the question of how much genetic difference there is between two species, of what the relative genetic divergence is between taxa of different evolutionary relationship.

The ambiguity of continuously varying phenotypes can be overcome by studying simple Mendelizing differences between species; by analogy with the same problem in the study of genetic variation within a species, antigenic differences come immediately to mind. If nonspecific antigenic tests are made—for example, by preparing antiserum against tissue extracts of one species and measuring the intensity of reaction of this antibody against other species—nothing has been gained since we are dealing again with a genetically unanalyzable quantitative character. It is necessary to make antibodies against single antigens specified by single genes and then to survey species for the possession of homologous antigens. This technique was utilized by Irwin (1953) to study several genera of doves (Columbidae). He first found nine antigenic differences between the blood cells of *Streptopelia chinensis* and *S. resoria*, establishing by species crosses that these were the result of nine separate gene differences. He then tested the blood cells of 23 other species against the nine specific antisera, with results shown in table 35. Antigen d-1, for example, seems widely distributed over the

TABLE 35
Number of species in different genera of the Columbidae that possess red blood cell antigens homologous to nine antigens of *Streptopelia chinensis*.

	Genus Streptopelia		*Genus Columba*		*Other genera*	
Antigen	*present*	*absent*	*present*	*absent*	*present*	*absent*
d-1	6	2	5	6	5	1
d-2	3	5	4	7	0	6
d-3	4	4	2	9	0	6
d-4	5	3	3	8	3	3
d-5	2	6	5	6	0	6
d-6	2	6	4	7	0	6
d-7	2	6	9	2	0	6
d-11	6	2	8	3	0	6
d-12	4	4	4	7	0	6
Total	34	38	44	55	8	46

Note: From Irwin (1953).

family whereas d-3 is nearly restricted to *Streptopelia*. Congeners of *S. chinensis* have slightly more similarity to it than do species of *Columba*, and the two genera together are more similar than are other genera. One species, *Streptopelia senegalensis*, was homologous to *S. chinensis* for all nine antigens.

Irwin could further test whether the homologies were indeed identities by saturating an antiserum with the cells of one species and then seeing whether the antiserum still retained any power against the cells of its own species. If not, then presumably the cells of the two species had identical antigens. Except for *S. senegalensis*, four of whose antigens were identical with *S. chinensis*, nearly all other cases of homology failed to show identity. The failure of this approach to the genetic similarity between species is, as Irwin points out, that we do not know how many red cell antigens the species are identical, because the tests depended upon finding an antigenic difference between *S. chinensis* and *S. risoria* in the first place. As I pointed out in chapter 3, this is the problem in general of using antigens to measure genetic differences. There has to be a difference to begin with for a gene to be detected at all.

The methodological problems of measuring the differences between the genomes of different species are the same as for comparing different individuals within a species except that there is the added problem that most species cannot be crossed with each other.

There is all the more reason, then, to have a method which does not depend totally on classical genetic analysis and in which the invariant genes can be detected as well as differentiated ones. We are again led to the study of a random sample of specific enzyme and protein molecules by means of a comparison of their electrophoretic mobilities. By use of proteins, especially enzyme proteins, whose genetics have been established by intraspecific study, species can be sampled even when they cannot be crossed, although crosses should be made whenever possible to establish gene homologies.

The pioneering work in this field was the study of *Drosophila virilis* and its relatives by Hubby and Throckmorton (1965). They compared the electrophoretic mobility of the soluble proteins from ten species of the *virilis* group. Different numbers of bands appeared in different species, from 29 in *D. lacicola* and *D. flavomontana* to 42 in *D. texana*. Because no genetic work had then been done, protein bands could not be assigned to loci, so that only presence or absence of a band at a particular location in the gel could be scored. If, for example, a locus had ten electrophoretically different alleles, each fixed in a different species, a total of ten potential band positions would be scored and each species would show "presence" for one and absence for the other nine. If, on the other hand, each species had a unique locus specifying a protein, not shared by the other species, there would again be ten potential positions in the total study and each species would have a band at one and lack it at the nine other positions. There is thus a confounding between different alleles at one locus and the presence or absence of the products of different genes.

A second problem is that with so many protein bands, a change in electrophoretic mobility of the protein from one locus could simply move the band into a position superimposed on a protein from another locus.

Third, it is not certain that each band within a species is coded by a different gene. Indeed, if any of the strains tested were polymorphic (only one strain per species was tested), a difference between species might be assumed that was only a chance difference between strains.

Finally, if the proteins form complexes with small molecules that affect the net charge of the complex, then a whole group of bands from one species might be shifted relative to another species, although the gene products were the same.

All of these difficulties make the analysis imprecise, but they do not create any major biases (except for the last problem, which, if widespread, would be serious). Each band is, to a first approximation, a gene product, and a different mobility means an elementary genetic difference. Unfortunately, as in all electrophoretic work, a lack of difference is ambiguous since many gene changes are not reflected in charge. *This last fact means that all estimates of genetic differences between species based on electrophoretic studies of proteins are underestimates*, by perhaps a factor of three.

Table 36 is a summary of Hubby and Throckmorton's findings. As little as 2.6 percent of the proteins of a species may be unique to it (*D. virilis*) or as many as 28.2 percent, with an average of 14.3 percent. This does not mean, however, that 14 percent of the proteins of these species have arisen since they diverged from their common ancestor, or even since the common ancestor of two very closely related forms. In any pair of species that have derived from a common ancestor, one of the unique alleles may be ancestral.

The probable phylogenetic tree of the ten species in Hubby and Throckmorton's study, based on chromosomal rearrangements and protein similarities, shows that the ten extant species most probably

TABLE 36

Percentages of larval proteins that are unique to a species, shared with other species in the phylad, and shared with some member of the species group as a whole, for ten species of the *Drosophila virilis* group

Species	*Total bands*	*% Unique*	*% Common to phylad*	*% Common to species group*
Virilis phylad				
D. americana	38	5.3	23.7	71.1
D. texana	42	21.4	16.7	61.9
D. novamexicana	38	7.9	21.1	71.1
D. virilis	38	2.6	21.1	76.3
Montana phylad				
D. littoralis	39	28.2	25.6	46.2
D. ezoana	35	8.6	25.7	65.7
D. montana	37	18.9	37.8	43.2
D. lacicola	29	20.7	20.7	58.6
D. borealis	42	19.0	28.6	40.6
D. flavomontana	29	10.3	37.9	51.7
Average	36.6	14.3	25.9	59.8

Note: Data are from Hubby and Throckmorton (1965).

trace back to four immediate ancestral forms, so that 40 percent of unique proteins could nevertheless be ancestral. Making this maximum allowance, we conclude that at a minimum 8.5 percent of the proteins in the extant species have arisen since their speciation. How much of this has been "phase 3" evolution, and how much stems from the initial speciation events, we cannot tell. A further 25.9 percent of the proteins are unique to one of the two phylads that make up the species group but common to species within the phylad. By the same reasoning, half of these could have been present in the ancestral form, so that as little as 15 percent of the protein species may have arisen since the origin of the two phylads and, adding this to the minimum estimate from unique species proteins, we get a minimum estimate of 23.5 percent for the proportion of proteins that have changed from the ancestral form. Again we must bear in mind that this is doubly a minimum, since it takes into account only the electrophoretically identifiable changes.

Recently J. Hubby, L. Throckmorton, and R. Singh have shown that "alleles" identified by electrophoresis are, in fact, heterogeneous classes in some cases. By determining the heat stability of *xanthine dehydrogenase* and *octanol dehydrogenase* alleles in *Drosophila virilis* and its close relatives, these workers have detected amino acid substitutions within electrophoretic classes, thus multiplying the number of different alleles segregating within a species by a factor of 2.5 and in the species group as a whole by a factor of 3.8. In turn, some species thought to be identical with respect to allelic composition at the *octanol dehydrogenase* or *xanthine dehydrogenase* loci turn out to be carrying different alleles that are detected by the heat-stability criterion.

The second study of Hubby and Throckmorton (1968) was an advance in precision because, in place of the large number of proteins of unknown genetic control, it was restricted to enzymes and larval hemolymph proteins, each of which is the product of a separate gene. Nine triads of species were chosen, two members of each triad being sibling species, the third being a nonsibling member of the same species group. A list of the species is given in the first column of table 37, the first two species in each group being the siblings. The table shows the percentage of genes with identical alleles between siblings (I), the percentage of loci with an allele identical between one of the siblings and the nonsibling (II), and the percentage of loci with identical alleles for all three species. Because

TABLE 37
Percentages of loci with identical alleles between sibling species (I), between a sibling and nonsibling in the same species group (II), and among all three species of a triad (III) for nine triads of species in the genus *Drosophila*

Triads	I	II	III	I + III	II + III
arizonensis *mojavensis* *mulleri*	42.1	6.3	6.3	48.4	12.6
mercatorum *paranaensis* *peninsularis*	55.0	11.8	11.8	66.8	23.6
hydei *neohydei* *eohydei*	43.8	3.2	6.3	50.4	9.5
fulvimaculata *fulvimaculoides* *lemensis*	50.0	13.2	15.8	65.8	29.0
melanica *paramelanica* *nigromelanica*	26.3	10.0	5.3	31.6	15.3
melanogaster *simulans* *takahashii*	52.9	7.9	0.0	52.9	7.9
saltans *prosaltans* *emarginata*	36.8	7.7	10.5	47.3	18.2
willistoni *paulistorum* *nebulosa*	7.1	11.6	15.4	22.5	27.0
victoria *lebanonensis* *pattersoni*	64.3	0.0	21.4	85.7	21.4
Average	42.0	7.9	10.4	52.4	18.3

Note: Data are from Hubby and Throckmorton (1968).

only a single strain was tested for each species, the question of polymorphism within species does not enter into the observations (except in one case), although it does enter into the interpretation of the results, as I shall show.

The percentage of identity between siblings (I + III) varies considerably from group to group. *Drosophila victoria* and *D. lebanonensis*, for example, are nearly identical siblings, with the

same allele at 85.7 percent of their loci, whereas *D. willistoni* and *D. paulistorum* are identical only at 22.5 percent of their loci. There is rather less variation for the comparison between nonsiblings (II + III). On the average, sibling species differ at 47.6 percent of their loci and nonsiblings at 81.7 percent. This study thus shows a much greater degree of genetic divergence than the results on *D. virilis* suggested. There is such immense variation between triads, however, that the *D. virilis* results are well within the range observed.

It is sometimes supposed that sibling species are species *in statu nascendi*, entities in the process of speciation. But as Mayr (1963) explains, that is certainly not the case. Sibling species are those for which the taxonomist is unable, or was unable at one time, to find reliable morphological differences despite clear-cut reproductive isolation. But similarity is in the eye of the beholder (the flies seem to have no difficulty in doing their own taxonomy!), and subsequent more careful study has sometimes led to the discovery of distinctions. However, the close morphological similarity is not without meaning, or the entire practice of morphological systematics would have resulted in a meaningless jumble. Total morphological similarity is, if not an infallible guide, at least a reliable indication of genetic similarity, and the data of Hubby and Throckmorton are an elegant confirmation of this proposition. Sibling species have three times as much genetic similarity as nonsiblings, and only in one case out of nine (*willistoni-paulistorum-nebulosa*) is the genetic similarity greater for the nonsiblings than the siblings

A certain care must be exercised in interpreting the large genetic differences between closely related species in the Hubby and Throckmorton data. Only a single strain, virtually always homozygous as it turned out, was examined for each species. Suppose that two species differ only in allele frequency at a polymorphic locus, one with alleles *A* and *a* in frequency 0.8 and 0.2, respectively, the other with the reverse frequencies. Then a little more than two-thirds of the time, single strains from each species will be fixed for opposite alleles, and the two species will be judged different. But how great, biologically, is the difference between a 0.8:0.2 and a 0.2:0.8 polymorphism? It is certainly not a basis for reproductive or ecological isolation. Is it possible that much of the difference between the species is variation in frequencies of polymorphic loci? Of the species listed in table 37, only four, *D.*

TABLE 38
A comparison of allele frequencies at 13 polymorphic loci in *D. pseudoobscura* and *D. persimilis*

Locus	*Alleles*	*D. persimilis*	*D. pseudoobscura*
Pt-7	*0.75*	1.00	.95
	others	—	.05
Pt-12	*1.18*	1.00	.55 (.20)[a]
	1.20	—	.45 (.80)
Pt-13	*1.23*	—	.06
	1.30	1.00	.94
Malic dehydrogenase	*1.00*	1.00	.97
	1.20	—	.03
Octanol dehydrogenase	*1.00*	1.00	.97
	1.22	—	.03
Acetaldehyde oxidase-2	*1.00*	1.00	.94
	others	—	.06
Pt-8	*0.80*	.02	.02
	0.81	.88	.47
	0.83	.10	.51
Glucose-6-phosphate dehydrogenase	*1.00*	.79	1.00
	1.10	.21	—
Leucine amino peptidase	*1.00*	.71	.89
	1.10	.18	.05
	1.12	.11	—
	others	—	.06
Pt-10	*1.02*	—	.005
	1.04	1.00	.615 (1.00)[a]
	1.06	—	.380

melanogaster, D. simulans, D. willistoni, and *D. paulistorum,* have been tested subsequently for electrophoretic polymorphism; between 42 and 86 percent of their loci are polymorphic. Clearly we need more refined information on the array of alleles present in natural populations of a pair of species before we can have a full appreciation of their specific differences.

The first detailed study of allelic frequency patterns in two closely related species was Prakash's (1969) comparison of natural populations of *D. pseudoobscura* and its sibling species, *D. persimilis.* The range of *D. persimilis* is, as far as is known, completely included within that of *D. pseudoobscura,* being confined to higher and moister localities in California, Oregon, Washington, and British Columbia. There is a broad spatial and temporal overlap in the dis-

Locus	*Alleles*	*D. persimillis*	*D. pseudoobscura*
Xanthine	*0.99*	—	.26
dehydrogenase	*1.00*	.22	.60
	1.02	.68	.01
	1.04	.10	—
	others	—	.13
Amylase	*0.74*	—	.03
	0.84	.02	.29
	0.92	.08	—
	1.00	.54	.68 (1.00)[a]
	1.05	.06	—
	1.09	.03	—
Esterase-5	*0.95*	—	.12
	1.00	—	.43
	1.02	—	.02
	1.03	—	.08
	1.07	.02	.19
	1.09	—	.01
	1.12	.04	.13
	1.16	.06	.02
	1.20	.74	—
	1.29	.02	—
	1.33	.12	—

Note: Data are from Prakash (1969) and Prakash, Lewontin, and Crumpacker (1973).

[a]Frequencies in the Standard gene arrangement of the third chromosome. *D. persimilis* arrangements belong to the Standard phylad of inversions.

tribution of the two species. They can be crossed, and the chromosomes are completely homologous, differing only in a few large inversions. Both species have extensive inversion heterozygosity on chromosome III and share one arrangement, Standard (see figure 14). There is no question about the homology of loci in the species, and it has been shown, for example, that hybrid enzyme is formed between esterase monomers produced by the two species, both in species hybrids and in *in vitro* dimerization (Hubby and Narise, 1967).

Prakash studied 24 loci in *D. persimilis* from Mather, California. Of these, 11 are monomorphic in *D. pseudoobscura*; they turned out to be *identically* monomorphic in *D. persimilis*. The remaining 13 loci are shown in table 38, compared with allelic frequencies in the

Strawberry Canyon population of *D. pseudoobscura*. For third chromosome loci, gene frequencies in the Standard arrangement of *D. pseudoobscura* are also given, since *D. persimilis* is related to the Standard phylad of inversions.

The results are remarkable. Except for the *esterase-5* locus, where the overlap in gene frequencies is small, there is very little differentiation in gene frequencies. Indeed, in 10 out of the 13 cases the major allele is the same in both species. If we add the 13 cases where both species are identically monomorphic, there is only slight or no gene frequency differentiation at 21/24 or 88 percent of loci, clear quantitative differentiation in allele frequencies at 2 loci, *pt-8* and *xanthine dehydrogenase*, and something approaching qualitative differentiation at 1 locus, *esterase-5*. There are sporadic occurrences of alleles that are unique to one species or the other, but except for the *esterase-5* locus no unique allele reaches a frequency greater than 25 percent. Most important, there is not a single case of fixation or even near-fixation for alternative alleles in the two species.

If there are "species-distinguishing" genes as indeed we suppose there must be, since these species are ecologically differentiated and reproductively isolated, they have not been picked up in a random sample of 24 loci. Thus, even if such species-differentiating genes are large in absolute number, they must be a small fraction of the whole genome, almost surely less than 10 percent of it. An alternative is that there are no such species-distinguishing genes but that the difference between species lies in the accumulation of quantitative differences in allelic frequencies, as in the case of the *esterase-5* locus. This latter hypothesis is not particularly attractive because it assumes that species differences simply represent very low probabilities of total genetic identity between individuals. Yet with the degree of polymorphism within species that has been revealed, the probability of genetic identity within a species is already essentially zero. For example, using only the 20 most polymorphic genes known at present in man, the probability of genetic identity between two Englishmen is already less than 10^{-6}. It seems more reasonable to suppose that *D. persimilis* and *D. pseudoobscura* do indeed differ completely at certain loci, like those found by Dobzhansky in his tudy of sterility in their hybrids, but that only a special part of the genome is involved, while most of the genome remains undifferen-

tiated. The discovery of considerable genetic heterogeneity within electrophoretic classes, through heat denaturation studies (see p. 169), might alter this conclusion significantly.

Ayala and Powell (1972) have asked a somewhat different question about such data. If we look at the species with the eyes of the systematist, we can ask whether a locus such as the *esterase-5* gene would be a good diagnostic character for the species. This can be done by calculating the frequency of the diploid genotypes at a locus in each species and then adding up the overlap in the distributions. If unknown individuals were assigned to that species in which their genotype was the more frequent, the error of assignment would be half the overlap. We can call a locus "diagnostic" if the probability of error is small, 0.01 or less. On this criterion only the *esterase-5* locus is diagnostic in the Prakash sample, the next best locus being *xanthine dehydrogenase,* with a 5 percent overlap in diploid genotypes.

The value of Ayala and Powell's approach is that it draws attention to the diploid genotypic distributions, which are, after all, what matter to the organisms, rather than to the gene frequency distributions, which are more of an abstraction. Ayala and Powell report finding 3 more loci, not examined by Prakash, that are diagnostic for *D. pseudoobscura* and *D. persimilis* although they are not as differentiated as *esterase-5*, and 12 more loci that do not differentiate the species. Thus there are 4 diagnostic loci out of 39.

Another pair of sibling species, *D. melanogaster* and *D. simulans*, were examined by Kojima, Gillespie, and Tobari (1970). The result of studying 17 enzyme loci was much the same as for *D. pseudoobscura*. No loci were fixed for alternative alleles in the two species, although one case, *aldehyde oxidase,* was highly polymorphic in *D. simulans* but monomorphic in *D. melanogaster* for an allele with frequency less than 0.01 in the former species. Several polymorphic loci had different major alleles in the two species, but only 2 (11 percent) were diagnostic at the 0.01 level. On the whole there is more differentiation of allelic frequencies at the highly polymorphic loci than is the case for *D. pseudoobscura* and *D. persimilis*, and unique alleles are a more common occurrence. The general picture, however, is not different in the two cases.

The most extensive and, in many ways, the most interesting, detailed comparison of gene frequencies among species is the work of

TABLE 39
Frequencies of alleles at 12 "diagnostic" loci in four species of *Drosophila*. Several alleles that occur with low frequencies are not included

Gene	Alleles	*Drosophila willistoni*	*Drosophila tropicalis*	*Drosophila equinoxialis*	*Drosophila paulistorum*
Lap-5	*0.98*	.09	.02	—	—
	1.00	.29	.19	—	—
	1.03	.50	.63	.004	.004
	1.05	.09	.15	.21	.08
	1.07	.007	.01	.71	.86
	1.09	—	—	.07	.04
Est-5	*0.95*	.03	—	.03	.03
	1.00	.96	—	.94	.84
	1.05	.01	—	.02	.13
Est-7	*0.96*	.02	.02	—	—
	0.98	.16	.11	—	—
	1.00	.54	.62	—	.002
	1.02	.23	.23	—	.08
	1.05	.05	.03	—	.78
	1.07	.003	.001	—	.09
Aph-1	*0.98*	.02	—	—	—
	1.00	.84	.05	.02	.01
	1.02	.08	.90	.92	.93
	1.04	.06	.04	.06	.03
Acph-1	*0.94*	.05	.95	.01	—
	1.00	.92	.03	.17	—
	1.04	.02	.006	.81	.16
	1.06	—	—	—	.21
	1.08	—	—	.004	.62
Mdh-2	*0.86*	.001	.994	.003	.001
	0.94	.02	.005	.994	.993
	1.00	.97	—	.004	.006

Ayala and Powell (1972) on the four sibling species of the *Drosophila willistoni* group, *D. willistoni, D. paulistorum, D. equinoxialis*, and *D. tropicalis*. These tropical South and Central American species differ very much in their ecological latitude and geographical range. *D. willistoni* is the most inclusive species, ranging from southern Florida and Mexico to northern Argentina. *Drosophila paulistorum* is absent from the northern part of this range but is very common in South America, being dominant in the drier and moderately humid areas, whereas *D. willistoni* is the major species in the superhumid tropics. *Drosophila tropicalis* and *D. equinoxialis* are absent both in southern Brazil and in Florida and

Gene	Alleles	*Drosophila willistoni*	*Drosophila tropicalis*	*Drosophila equinoxialis*	*Drosophila paulistorum*
Me-1	*0.90*	—	.03	—	—
	0.94	—	.91	—	.004
	0.98	.02	.06	—	.99
	1.00	.95	—	.005	.005
	1.04	.02	—	.99	—
Tpi-2	*0.94*	.003	.01	—	.02
	1.00	.98	.98	.02	.98
	1.06	.01	.01	.98	—
Pgm-1	*0.96*	.04	—	.01	.02
	1.00	.87	.01	.35	.94
	1.04	.08	.98	.62	.04
Adk-2	*0.96*	.01	.05	—	—
	0.98	.05	—	—	—
	1.00	.88	.92	.04	.98
	1.02	.05	—	—	—
	1.04	.004	.03	.94	.02
Hk-1	*0.96*	.04	.02	.08	—
	1.00	.95	.96	.91	.01
	1.04	.006	.02	.005	.97
	1.08	—	.001	.002	.02
Hk-3	*1.00*	.98	.97	.95	.07
	1.04	.006	.01	.04	.92

Note: Data are from Ayala and Powell (1972). A dash indicates that the allele has not been found in the species.

Lap = leucine aminopeptidase; *Est* = esterase; *Aph* = alkaline phosphatase; *Acph* = acid phosphatase; *Mdh* = malic dehydrogenase; *Me* = malic enzyme; *Tpi* = triose phosphate isomerase; *Pgm* = phosphoglucomutase; *Adk* = adenylate kinase; *Hk* = hexokinase.

Mexico and have their highest proportions in the Greater Antilles and on the Caribbean coast of Colombia and Panama, where they are a major and sometimes dominant proportion of the four species (Burla et al., 1949). On the basis of chromosomal evidence, *D. paulistorum* and *D. equinoxialis* are more closely related to each other than they are to the other two species and, as we will see, this is borne out by the genic evidence.

The data of Ayala and Powell are worth displaying, both for their value in the present context and for our future discussion of the meaning of polymorphism (chapter 5). Table 39 gives the allelic frequencies at 12 out of the 28 loci examined. The loci displayed are

those that are "diagnostic" for at least one species in each case, and so represent the loci at which there is significant differentiation. Even among these diagnostic loci there are some extraordinary similarities, as for example *esterase-5*, which is included, presumably, because no activity at this locus was detected in *D. tropicalis*, but which is nearly identical in frequency distribution in the other three species. Although there is only a single case in which the distributions have no overlap at all (*D. tropicalis* and *D. equinoxialis* at the *me-1* locus), there are many instances among the last 7 loci in the table where species are nearly fixed for alternative alleles, for example *D. tropicalis* and *D. equinoxialis* for *mdh-2*, and *D. equinoxialis* and *D. paulistorum* for *me-1*. Indeed the *me-1* locus almost completely distinguishes all four species from each other.

On the other hand, there are some impressive similarities between pairs of species at highly polymorphic loci. A particularly interesting case is the near-identity of *D. willistoni* and *D. tropicalis* for the 6 alleles at the *esterase-7* locus, in view of their considerable differentiation from *D. paulistorum* at that locus. A similar case is at the *aph-1* locus, for which three species are identically polymorphic and the fourth differs but is still polymorphic. One gains a general impression from the data that *D. willistoni* and *D. tropicalis* form a related pair and *D. equinoxialis* and *D. paulistorum* form a second.

In interpreting the data of table 39, one must not lose sight of the 16 loci not shown, for which all four species were similar or identical in their gene-frequency distributions. When these are taken into account, the proportions of diagnostic loci distinguishing various species pairs are those shown in table 40. The fraction of diagnostic genes varies between 14 and 35 percent.

Generally, the differentiation among the four species of the *D. willistoni* group is higher than for the *melanogaster-simulans* pair or the *pseudoobscura-persimilis* pair. The feature held in common is

TABLE 40

Percentages of loci that are "diagnostic" at the 1 percent level for pairs of species in the *Drosophila willistoni* group

	D. tropicalis	*D. equinoxialis*	*D. paulistorum*
D. willistoni	17.9	21.4	25.0
D. tropicalis		21.4	35.7
D. equinoxialis			14.3

the general absence of alleles that are fixed in one species and lacking in another. Even in the *willistoni* group there are only 5 cases out of 112 in which a dominant allele in one species is not found in at least one other species at a frequency of 0.01 or more. Indeed, in only 45 of 336 cases is the dominant allele in one species not found in *all three* other species at a frequency of 0.01 or greater. There is not a single case of a unique dominant allele.

These observations taken together mean that where species are highly differentiated in their alleles there is at least a low-level polymorphism in one species for the genes that characterize the other. There is then a potential genetic transition between species that does not require the chance occurrence of new variation by mutation. That is, *the overwhelming preponderance of genetic differences between closely related species is latent in the polymorphisms existing within species.* Obviously this generalization becomes less and less true as species diverge farther and farther in the course of their phyletic evolution. We find, for example, no allelic commonality between Drosophila species belonging to different species groups, such as *D. melanogaster* and *D. pseudoobscura.* But our evidence does tell us that this greater differentiation requires only the occasional input of mutational novelties and that the early stages of phyletic divergence make use of an already existing repertoire of genetic variation.

SPECIES *IN STATU NASCENDI*

All of the evidence given so far has concerned species that have long since completed their speciation and are underdoing their third stage of evolutionary differentiation. Evidence for species in their second stage, the reinforcement period when formerly allopatric populations have come together again, is much more difficult to gather. The chief difficulty is one of identification. How are we to know sympatric elements in the second stage of speciation when we see them? Unless there is some morphological or cytological differentiation between the entities, they will not be recognized as distinct. And if they are distinct enough to be recognized, there must be evidence of hybridization between them, hybridization that does not result in a swamping of their differences.

At present I know of only two cases that have been surveyed for

patterns of allelic variation. One is the pair of subspecies of *Mus musculus* in Denmark (Selander, Hunt, and Yang, 1969), and the other is the complex of "semispecies" of *Drosophila paulistorum* in South America (Richmond, 1972).

In Denmark there is a light-bellied northern race, *Mus musculus musculus,* which meets a southern, dark-bellied race, *M. m. domesticus,* in a narrow zone of overlap and hybridization. Intermediate populations occupy an east-west belt only a few kilometers wide. Selander, Hunt, and Yang surveyed 41 enzyme loci in four populations of *M. m. musculus* and two of *M. m. domesticus,* with the results shown in table 41. Only 16 loci are given, the other 25 being identically monomorphic in both races. The first three *M. m. musculus* samples were from the northern part of the main Jutland peninsula, but the fourth, although morphologically *musculus,* was sampled from three islands, half the sample coming from an island very close to the mainland and far south of the mainland borderline between the subspecies. In several instances the allele frequencies in this population show evidence of introgression from *domesticus;* the most striking examples are *esterase-1,* in which an allele usually missing from *musculus* is present in 5 percent, and *hexose-6-phosphate dehydrogenase,* in which two alleles have been introduced from *domesticus,* in equal and low frequency. Thus the genetic evidence confirms that *domesticus* and *musculus* are genetically differentiated entities with some hybridization and a small amount of introgression in their overlap zone, but not sufficient gene flow to destroy the difference between them. These are exactly the characteristics of populations in stage two of their speciation process.

Table 41 shows a great deal of differentiation. Discounting the introgressed population of *musculus*, there are two cases of fixation for alternative alleles in the two subspecies, *esterase-1* and *hexose-6-phosphate dehydrogenase*, and six additional cases in which opposite alleles are clearly in the majority in the different entities. Thus 8 loci out of 41, or 20 percent of loci, are differentiated between the forming species, and of these, 5, or 12 percent, are diagnostic at the 0.01 level. The very small sample sizes tend to exaggerate the lack of overlap so that the two loci fixed at alternative alleles may not, in fact, have disjunct frequency distributions. Nevertheless, the degree of divergence is not different from that of the fully formed species of the *willistoni* complex in *Drosophila* and, if

TABLE 41

Allele frequencies for 16 loci in samples of *Mus musculus musculus* and *M. m. domesticus*

		M. m. musculus samples				*M. m. domesticus* samples	
Locus	*Allele*	*1*	*2*	*3*	*4*	*1*	*2*
Esterase-1	*a*	1.00	1.00	1.00	.95	—	—
	b	—	—	—	.05	1.00	1.00
Esterase-2	*a*	—	.17	.07	.05	1.00	1.00
	b	1.00	.83	.93	.95	—	—
Esterase-3	*a*	.30	.67	.47	.32	.40	.30
	b	.70	.33	.53	.68	.60	.40
Esterase-5	*a*	1.00	.86	1.00	1.00	.78	.69
	b	—	.14	—	—	.12	.31
Alcohol dehydrogenase	*a*	.83	.70	.90	.60	.97	.97
	b	.17	.30	.10	.40	.03	.03
Lactate dehydrogenase regulator	*a*	.95	.64	.69	—	.38	.29
	b	.05	.36	.31	—	.62	.71
Supernatant malic dehydrogenase	*a*	.37	.37	.37	.47	1.00	1.00
	b	.63	.63	.63	.53	—	—
Mitochondrial malic dehydrogenase	*a*	.95	.93	.80	.71	1.00	1.00
	b	.05	.07	.20	.29	—	—
Hexose-6-phosphate dehydrogenase	*a*	—	—	—	.03	.50	.53
	b	—	—	—	.03	.50	.47
	c	1.00	1.00	1.00	.94	—	—
6-Phosphogluconate dehydrogenase	*a*	.33	.83	.93	.76	1.00	1.00
	b	.67	.17	.07	.24	—	—
Isocitrate dehydrogenase	*a*	.17	.23	—	.08	.93	.98
	b	.83	.77	1.00	.92	.07	.02
Indophenol oxidase	*a*	.02	.13	—	.42	1.00	1.00
	b	.98	.87	1.00	.58	—	—
Phosphoglucomutase-1	*a*	.75	.43	.40	.37	1.00	1.00
	b	.25	.57	.60	.63	—	—
Phosphoglucomutase-2	*a*	.03	—	.23	.13	1.00	1.00
	b	.97	1.00	.77	.87	—	—
Phosphoglucose isomerase	*a*	1.00	1.00	1.00	1.00	.93	.93
	b	—	—	—	—	.07	.07
Hemoglobin	*a*	.10	.10	.63	.50	—	.23
	b	.90	.90	.37	.50	1.00	.77

Note: Data are from Selander, Hunt, and Yang (1969).

anything, is greater than for the sibling species *melanogaster-simulans* and *pseudoobscura-persimilis*, although, of course, a comparison of evolutionary rates in such divergent forms as Diptera and mammals is bound to be shaky.

Drosophila paulistorum is distributed through Central and South America from Costa Rica and the Greater Antilles to southern Brazil. Over that range it is divided into six semispecies that overlap in distribution more or less, depending upon geography. Thus the northernmost, "Centro-American," and the southernmost, "Andean-Brazilian," occupy large exclusive areas but overlap at their margins with the four central semispecies, which are much more broadly overlapping with each other. The "Interior" semispecies, for example, inhabits the overlap zone between the "Amazonian" and the Andean-Brazilian.

The semispecies are not morphologically distinguishable although there has been cytological differentiation and reproductive isolation. Even in laboratory conditions females of one semispecies usually refuse to accept males of another although no choice of mate is offered, and when crosses do occur the F_1 hybrid males are sterile. Gene passage between semispecies happens, however, through the "Transitional" race, which will cross with other races and give fertile sons (see Dobzhansky, 1970, pp. 369–72, for a review and references concerning this fascinating group).

A study of 17 enzyme loci in the six semispecies from a great variety of localities has been made by Richmond (1972). The species as a whole turns out to be as polymorphic as its sibling, *D. willistoni*, with about two-thirds of its loci polymorphic in any population and about 20 percent heterozygosity per individual. Of the 17 loci only 4 show any differentiation between semispecies, and even in those cases the differentiation is small. For 1 locus, three of the semispecies (C, T, and AB) are segregating for two alleles in a ratio of about 0.9:0.1 while the other three (O, I, and A) are in the ratio 0.3:0.7. In a second case the semispecies C, T, and AB are segregating in varying proportions for four alleles while O, I, and A are nearly monomorphic for one of them. At a third locus that has 7 alleles segregating, the most frequent allele in the A race is different from the one that is common in the other races, but there is a broad overlap in the frequency distributions.

Finally, a locus that is nearly monomorphic in the rest of the

semispecies is polymorphic in the AB race, with about 25 percent of a unique allele. Moreover, the O race, represented unfortunately only by laboratory stocks, is monomorphic for an allele that is found segregating only in the C race. With the possible exception of this last case, then, there are no diagnostic loci for any of the semispecies comparisons. It is not surprising that in two of the cases it is the Centro-American, Andean-Brazilian, and Transitional races that are similar to each other but dissimilar to the other three groups. It is precisely the Centro-American and Andean-Brazilian semispecies that are cross-fertile with the Transitional one.

The lack of genetic differentiation among the semispecies of *D. paulistorum*, entities that are strongly isolated from each other ethologically and by sterility barriers, is in sharp contrast to the considerable differentiation of the group of four sibling species of which *D. paulistorum* is a part. If the semispecies can be regarded as a model for the second step in speciation, which, in the past, led to the formation of the present-day sibling species, then we see that most of the differentiation of the sibling species has occurred since speciation was completed. The differences we observe among the sibling species of the *willistoni* group are best explained as the result of phyletic evolution after successful speciation, whereas the speciation process itself resulted in very little differentiation.

Although we cannot prove that the events in *D. paulistorum* today are a repetition of the history of past speciation in the *D. willistoni* complex, a better comparison of stage two of speciation will never be found. All reconstructions of evolutionary processes that have left no fossil record depend upon the assumption that successive stages in the evolution of a system, in time, can be seen at present in different systems in the various stages of their own evolution. This principle of ergodicity lies at the base not only of biological systematics, but of a great deal of general historical reconstruction as well.

IN FLAGRANTE DELICTO

If so little genetic divergence characterizes races in their second stage of speciation, what will we find during the first stage, when populations have newly acquired reproductive barriers in isolation from each other? At present only one case has been studied, and it

is very illuminating. By a rare combination of luck and intuition, Prakash (1972) has discovered that the Bogotá population of *Drosophila pseudoobscura* is in the first stage of becoming a new species!

When females from Bogotá are crossed with males from any other locality, the F_1 males are completely sterile. The reciprocal cross produces perfectly normal sons. In accord with Haldane's rule (Haldane, 1922), in *Drosophila* male sterility is over and over again one of the primary reproductive isolating mechanisms. Moreover, as we saw on p. 164, the cross between *Drosophila pseudoobscura* and *D. persimilis* produces F_1 males with very small testes when the cross is made in one direction but not in the reciprocal. Thus, the sterility of sons of the cross between Bogotá females and "mainland" males but not from the reciprocal cross, as a first step in speciation, is quite in accord with similar occurrences in the past. Since the Bogotá population is only in the first stage of speciation we do not expect, and Prakash did not find, an ethological isolation between the Bogotá population and the rest of the species. Mating preferences presumably develop in the second stage, after the incipient species come into contact again.

From the evidence in chapter 3, the Bogotá population has all the earmarks of a recent colonization from a small number of propagules. Bogotá is only half as heterozygous as the rest of the species populations; it is usually monomorphic or nearly so for the allele that is most common in the rest of species, except for one locus at which it has a very high frequency of an allele that is usually low; and it is segregating for the two most frequent chromosomal rearrangements found in its nearest neighbors in Guatemala.

Extensive collections of *Drosophila* were made in Colombia in 1955 and 1956 by the expert University of Texas group, and 114 different species were found, but not a single individual of *D. pseudoobscura*. Much of the collecting effort was expended, however, in tropical localities where *D. pseudoobscura* is normally absent. In 1960 *D. pseudoobscura* appeared in traps in the city of Bogotá, and by 1962 this species comprised between 1 and 50 percent of the species trapped in various localities. Although it is impossible to rule out absolutely the presence of *D. pseudoobscura* at a very low level in Colombia for a very long time, the evidence is strong that the species was introduced not much before 1960 and

found an environment to which it was preadapted but from which it had been barred by its previous failure to cross the gap of 1500 miles from Guatemala. The sample of flies on which Prakash's experiments were done was collected in Bogotá in late 1967, so a reasonable guess is that ten years passed between the colonization and the sampling. During those 10 years the first step in speciation was taken, probably as a result of the initial colonization.

The first step in speciation, reproductive isolation, has occurred in the Bogotá population, yet there has been no genetic differentiation at the 24 loci examined by Prakash, Lewontin, and Hubby (1969). If anything, Bogotá is supertypical of *D. pseudoobscura*, since it is often homozygous or nearly homozygous for the most common alleles in the species (see table 26), with the exception of the *pt-8* locus, for which it has a high frequency (87 percent) of an allele that is found at only 1 or 2 percent in other localities. The first step in speciation has been taken, not by a wholesale reconstruction of the genome, but by the chance acquisition of an isolating mechanism that probably has a very restricted genetic basis. Total male sterility in only one of the two reciprocal population crosses would probably not be a sufficient mechanism in itself to cause speciation to be completed if the Bogotá population came in contact with the rest of the species now. But it is a large step toward speciation in a very short time. To what extent such a "quantum" loss of reproductive compatibility without genic divergence is the rule in speciation we cannot know until we accumulate many more cases of the early stages of speciation. The tools are readily available.

THE GENETICS OF SPECIES FORMATION

The evidence is sparse, and the closer we get to the beginning of the speciation process, the sparser the evidence. The most solid and coherent argument comes from the comparison of the sibling species of the *willistoni* group with the semispecies of *D. paulistorum*. The data on the other closely related species, on the races of *Mus*, and on the remarkable case of *D. pseudoobscura* from Bogotá "lend versimilitude to an otherwise bald and unconvincing tale."

The first stage of speciation, the acquisition of primary reproductive isolation in geographical solitude, does not require a major overhaul of the genotype and may result from chance changes in a

few loci. During the second stage, when the isolated populations have again come into contact, there is some genetic differentiation, perhaps as much as 10 percent of the genome having marked differences in the distribution of allelic frequencies. Complete divergence with fixation of alternative alleles is rare, and loss of old or acquisition of new functions rarer still. Given that the allelic classes detected by electrophoresis are themselves heterogeneous and that there are allelic differences between species that are detectable by techniques like heat denaturation, complete divergence is undoubtedly more common than has so far been estimated, probably by a factor of three.

It is only in the third stage of species evolution, the open-ended phyletic change that occurs more or less independently of the evolution of the sister species, that the major divergence occurs. The speed of this divergence is very variable relative to morphological change, at least in the early stages, so that sibling species may differ markedly in as little as 10 percent of their genome or as much as 50 percent, but even then the divergence is predominantly quantitative. It is rare for a species to differ by alleles that are not polymorphic in its very closely related species, so that most of the species divergence in the early stages of phyletic evolution makes use of the already available repertoire of genetic variants and is not limited by the rate of appearance of novelties by new mutation. This last point, that considerable evolutionary change (including speciation and divergence of new full species) occurs without being limited by the rate of appearance of novel genes is the chief consequence, for the process of speciation, of the immense array of genetic variation that exists in populations of sexually reproducing organisms.