Population genetics and evolution theory

1

Population genetics

R. C. LEWONTIN

Population genetics is one of the few biological sciences that has both a theoretical and an observational aspect. In this it resembles ecology, but it is unique in the degree to which the theoretical and the observational are tied to each other, at least in principle and in motivation, if not in useful results. There is no other biological science in which observations are so often directly motivated by formal theory and in which formal theory is so often constructed in an explicit attempt to make sense of the observations. The comparison with ecology is instructive. Theoretical population ecology is almost entirely the elaboration of a single underlying model, the logistic equation of population growth, for which there is virtually no empirical justification. At its most general, population ecological modelling does not take the logistic seriously, but supposes an unspecified multispecies interaction model which is then expanded in a Taylor's series, yielding, to the second term – the logistic model! Virtually all of observational ecology, on the other hand, is phenomenological. Do species interact? How? Can predation, competition, weather be shown to be causally efficacious or not in the determination of numbers of coexistence?

In contrast, population genetics begins with the undoubted facts of Mendelism, of chromosomal recombination, of mutation, of inbreeding, and builds a theoretical structure that is unassailable in its general outline. When called upon, it can even accommodate itself to non-Mendelian mechanisms of inheritance, to gene duplication and amplification, to any of the myriad genetic phenomena that are uncovered in the course of mechanical genetic investigation, as for example segregation distortion or gene conversion. At the same time, observational studies estimate the genetic variation that is the subject of the theoretical structure, or attempt to estimate the parameters of selection, migration, mutation, inbreeding, mating that are prescribed as relevant by the theoretical equations. The collaboration between Sewall Wright and Th. Dobzhansky in the observation and analysis of genic and chromosomal variation in natural popula-

tions has no parallel elsewhere in biology.* Nor was that union of theory and practice unique. The number of population geneticists, like John Maynard Smith, whose own research work has consisted of significant contributions to both theoretical and observation work, is quite remarkable.†

In view of the extraordinary interconnection between theory and observation in population genetics, we would expect to see a coherent field, well on its way to closure, having answered in an unambiguous fashion most or all of its leading questions. Instead, we see a field in disarray, with contending schools of explanation, apparently no closer to agreement on outstanding issues than they were 35 years ago when I first entered the professional study of the subject. At the moment, an appearance of calm pervades the population genetic arena, but a sharp ear will detect the panting of the out-of-breath contestants, as they try to regain some energy for yet another frustrating and indecisive round. What is truly remarkable is that, even when a question is decided, the result may have no perceptible effect on practice. It has now been 20 years since Moll, Lindsey and Robinson showed conclusively that overdominance is not the cause of heterosis in hybrid corn, a result with which no one apparently disagrees (Moll, Lindsey & Robinson, 1964), yet plant breeding continues to use a method designed for overdominance. How are we to explain the lack of real progress toward a consensus on the problems of population genetics, and what are the prospects for a better future?

In a previous work (Lewontin, 1974), I attempted to explain the unsatisfactory state of evolutionary genetics by the difficulty in measuring the actual quantities that appear in theoretical structures. The equations of population genetics deal with genotype frequencies, yet we have been obliged for most of the history of population genetics to study either phenotypes of unknown genotypic basis, or genotypes of no real interest in evolution. One would have thought that the last 20 years of observations on protein variation would have got around that epistemological paradox since protein variation usually has a well-defined genetic basis, and that variation is clearly something that lies at the base of a lot of non-trivial evolution. Yet struggles over what forces are operating on genetic variation in evolution still continue and, indeed, have been exacerbated by the recent interest in so-called punctuated equilibrium. Moreover, population genetics is not coextensive with evolutionary genetics, although people sometimes seem to think so. A large part of human population genetics,

* For details of the motivation, joint planning, and analysis of experiments by Dobzhansky and Wright, see Lewontin *et al.* (1981).

† In addition to Fisher, Wright, Haldane and H. J. Muller, I was able to list 19 in a quarter of an hour. A little more bibliographical effort would no doubt produce an equal number of others. Even a pure theoretician like M. Kimura is constantly motivated by and tied to experimental observations.

especially since the discrediting of eugenics in the 1950s, has concentrated on uncovering the causes of normal and abnormal variation in humans, in an attempt to be relevant to medicine and public health. Heritability studies of alcoholism and the 'genetic epidemiology' of feeble-mindedness use the apparatus of population genetics and are in at least as much contention as the problem of the evolution of enzymes. Agricultural population genetics, so called 'biometrical genetics', seems becalmed in Doldrums from which not even the combined huffing and puffing of statisticians and molecular geneticists has succeeded in moving it. Population genetics as a whole seems as unable as ever to solve the problems it has set for itself.

The difficulty is that in the development of population genetics there has been an inversion of the relationship of theory and observation. Population genetic theory begins as a deductive process. The phenomena of Mendelian segregation and chromosomal recombination, of mutation, migration, mating pattern, of the contingent development of the organism given the genotype and the environment, and of the differential survival and reproduction of genotypes, all subject to stochastic fluctuations, are built explicitly into the formal models of population genetic theory. All these phenomena are forces that are assumed to operate in a precisely defined way on the genotypic and phenotypic composition of the population. The straightforward problem of population genetics, then, is to compute the genetical state of the population at some future time from the present state and the forces. In formal terms, the state of the population at time t in the future is given by

$$x_t = g(x_0, \pi, t), \tag{1}$$

where π is the set of parameters of the process g. The process may also be computed backwards by solving the equation for x_0, given x_t. Adding a stochastic element changes (1) to a probabilistic statement,

$$\Pr\{x_t = x\} = h(x_0, \pi, t), \tag{2}$$

but otherwise the problem remains the same.

Given in this form, population genetics is a computational science. Tell me the genetics of the trait, the mutation rate, selective forces, population size, migration rates, etc., and I will tell you the trajectory (including the equilibrium) of the system. This computation can be carried out, however, only if the values of the parameters, π, are known. The straightforward approach would then be to devise independent methods of estimating the parameters, say of the amount of migration, by actually following individual organisms in their lifetimes, and to substitute the values found into (1) or (2). This deductive programme, however, has not been carried out for a number of reasons. First, it is very difficult to measure the actual rates of mutation and migration, the actual patterns of mating relations, actual $l(x)$ and $m(x)$ schedules of phenotypes, and the norms of reaction of the

various genotypes, even if the genotypes could be identified. Second, the genotypes cannot always be recognized, especially for metric characters. Third, developmental patterns, probabilities of survival and reproduction, and behavioural phenomena like mating and migration, are all contingent on a variable environment, and they cannot be measured once for all. Their pattern of response to environment must be characterized, and then the environmental pattern must be recorded. Fourth, in the case of deducing the past from the present, we would need to know the past environments, an impossible task. In practice, if it were important enough, we might gather the necessary information for one trait, say sickle-cell anaemia in humans, where many of the parameters can be assumed to be constant and where the genetic basis of the trait is simple. But, as a practical programme for any reasonable number of evolutionary pathways, the computational path of deduction is clearly out of the question. Moreover, it is not clear that the computation question is really the one of interest. In 'genetic epidemiology', for example, the question is not 'what pattern of feeble-mindedness in the population would result from a given combination of genotypic and environmental variables' but, on the contrary, 'what are the developmental phenomena actually operating in families to produce the pattern we see?'. In evolutionary population genetics, we are generally more interested in questions like 'Is most of the amino acid substitution that has occurred in the evolution of the species a consequence of natural selection or of random forces?' than in predicting or retrodicting the evolutionary trajectory of a particular enzyme protein. That is, the questions are more often about the forces themselves than about their outcome, because the aesthetics of 'pure' science demands generality rather than specificity.

The consequence of the interest in the forces themselves, the parameters in Eqns (1) and (2), and of the immense practical difficulties of measuring these forces directly is that the theoretical formulations of population genetics have been inverted. Instead of computational devices to predict $x(t)$, they have become inferential structures to estimate π. This is, of course, the classical method of statistical inference: we deduce the observations that *would* be taken, given various hypotheses about the hidden universe, and then use the observations *actually* taken to choose the most likely universe.

Three examples of this method in population genetics will illustrate its operation. In his widely quoted but little-read paper, 'The covariance between relatives on the supposition of Mendelian inheritance', Fisher (1918) showed how the phenotypic variances and covariances of relatives of various degrees could be written in terms of the parameters for two loci with two alleles, each influencing the phenotype and, therefore, by implication, for any number of alleles at any number of loci. For two loci, there are two gene frequencies, two additive effects, two dominance effects, four epistatic

parameters, and an environmental variance (a total of 11 parameters). This can be reduced to ten if one examines only crosses between inbred lines so that segregating gene frequencies will all be $p=q=\frac{1}{2}$. While there is no evolutionary dynamic here, it is still a computational problem analogous to Eqn (1), namely,

$$\sigma_{(R,R^1)_i} = g_i(\pi_1, \pi_2, \ldots, \pi_{11}). \tag{3}$$

That is, the expected phenotypic covariance between relatives R and R^1 of the ith degree can be computed by a specific function of the genetic and environmental parameters. But, of course, no one knows how to measure directly the additive, dominance, and epistatic effects of pairs of loci governing, say, yield in corn, not to speak of higher-order interactions for multiple loci. During the 1940s and 1950s, it became clear that rational choices of breeding methods could be made provided one could estimate these parameters, in particular the degree of dominance and epistasis in comparison with additive effects in a population. The result was a series of estimation equations of the type

$$\pi_i = h_i(\sigma(R, R^1)_1, \sigma(R, R^1)_2, \ldots), \tag{4}$$

a particular example of which is Comstock's estimate of the average degree of dominance over all loci:

$$`a' = \frac{[2(\sigma_f^2 - \sigma_m^2)]^{1/2}}{\sigma_m^2}, \tag{5}$$

where σ_f^2 and σ_m^2 are estimates of the variance among males and females within males in half-sib and full-sib families. The trouble with this estimate is that it is a fictitious 'average dominance' confounded by epistatic interactions and linked genes so that it estimated dominance only under certain simplifying assumptions. As a consequence, experiments of different structure gave different answers to the question of whether there was significant overdominance in corn, a question that was settled only after experiments specially designed to eliminate the effects of linkage were performed.

A less-happy outcome has characterized the second example, estimation of fitness in evolving populations. Beginning with a simple model of differential viability in an organism with discrete generations, we can predict successive generations. This relationship has then been inverted to make the observed ratios as the independent variables and the fitnesses as the dependent variables which can then be estimated from the inverse equations. However, Prout (1965) has shown that this procedure is completely invalid if there is differential fertility and will yield false frequency-dependent fitnesses, or even no evidence of selection when there is strong selection on fertility. Moreover, he has shown that *in principle* no procedure involving the genotypic frequencies in two successive generations will yield correct fitness estimates.

The third example illustrates a different sort of problem. Finite breeding size in a population results in a predictable probability of identity by descent of two alleles taken at random from a population. A special case of this identity is the frequency of allelism of lethals, which, roughly speaking, will be inversely proportional to the effective breeding size of the population. However, because the number of loci mutating to lethals is limited (about 400–600 per chromosome area in *Drosophila melanogaster*, for example), the observed frequency of allelism of lethals must be corrected for the allelism that arises from this limited number of lethal-bearing loci. That is, the allelism relevant to effective population size is

$$a_{\text{eff}} = \left(a_{\text{total}} - \frac{1}{n}\right),$$

where n is the number of loci mutating to lethals. In turn, the predicted a_{eff} is inversely proportional to effective population size N. That is,

$$a_{\text{eff}} \propto \frac{1}{2N},$$

so that inverting this relationship to estimate N gives us

$$N \propto \frac{1}{\left(a_{\text{total}} - \frac{1}{n}\right)}.$$

That is, the estimate of effective breeding size depends upon the reciprocal of the difference between two very small numbers, each of which has been derived from a very tedious experiment involving the test of allelism of a large number of lethals, and each with a considerable standard error. As a consequence the actual estimate of N may, with high probability, be negative or nearly infinite. An application of this technique to laboratory populations of known size had the result that it was not possible to distinguish between a population of 5000 from an infinite one (Prout, 1954).

These three examples illustrate different problems that arise when a causal relationship is inverted to attempt to estimate the causes. In the first case, the attempt to estimate the degree of dominance of genes suffers from the dependence of the covariances on a very large number of parameters of which the dominance is only one, so that inverting the deductive relationship can only be carried out by assuming values for the unknown parameters. In practice, these parameters can be estimated by yet other experiments or, as was the actual case for the dominance ratio in corn, experiments can be devised that reduce or remove the influence of the other factors such as linkage. In the case of corn, it was necessary to repeat the estimation experiment in various advanced generations of a random-mating population made from the original lines, an experiment that required a long-range plan and many years of experimental work.

In the second case, the possibility of inverting the causal relationship is structurally unstable to perturbations in the underlying model, and no experiments of this class will solve the problem. In the particular case of fitness estimation, this structural instability arises because meiosis and mating are randomizing events that actually destroy information created by the selection process, so that genotypic frequencies at two moments separated in the life cycle by meiosis and mating do not contain the necessary information to reconstruct the selection rules. On the other hand, two moments in the life cycle not separated by meiosis and mating do not contain any information about differential fertility, so the problem is insoluble by this approach.

The third case is the classic one of statistical estimation. Because the causes and effects are related as mathematical reciprocals, the sensitivity of dependent and independent variables to small perturbations are inverted. Large variation in N and n make only small absolute differences in the allelism outcomes, but reciprocally small differences in the observed allelism in an experiment result in immense variation in the estimate N, passing from the positive range through positive infinity into the negative half plane. When estimates are extremely sensitive to random variation in actual parameters or actual observations, the estimation procedure is useless. It is remarkable in population genetics how often a deductive relation has been inverted to create an estimation procedure without asking the question of its sensitivity to errors. For example, models of genetic epidemiology which depend upon path-coefficient analysis or related methods have seldom been subjected to extensive perturbation analysis to see how sensitive the estimates of genetic parameters are to different data sets. Monte Carlo sampling schemes or high-speed computers make this kind of perturbation analysis possible, although not trivial.*

The examples given are not exceptions, but the rule. Because of the large number of causal factors that enter into the determination of the genetic state of a population or an ensemble of populations and because some of these factors are randomizations that actually then destroy the products of information-creating causes, it will almost never be possible to invert the relation that predicts genetic structures from parameters of causal processes, in order to estimate the intensities of these processes. How, then, can population genetics do its business? Is it doomed to be nothing but perpetual number-juggling with no satisfactory closure? The answer lies in two directions, one being a proper understanding of what population

* The recent paper by Greenberg (1984) explicitly attacks this question. After a Monte Carlo simulation study to test how much power the methods have to distinguish between a one-locus and two-locus model with different degrees of dominance, about as simple a choice as one can imagine in genetics, the author notes: 'The studies reported here were extremely time consuming, both in human time and computer time. Yet such studies are an important way of testing the tools that geneticists use'. Right on!

genetics already provides and the second being a reorientation of population genetic research.

The first point is that population genetic theory is not designed to choose among competing hypotheses about causal forces. As I have explained, the attempt to use it for that purpose is destined, nearly always, to failure. Rather, population genetic theory is a descriptive theory that provides the mapping of causal processes as genetic outcomes. It says, '*if* mutation rates are such and such, *if* the mating pattern is such a one, *if* there are five genes affecting the character with the following norms of reaction, *then* the trajectory of the population in time, or the equilibrium state, or the steady state distribution of gene frequencies will be such and such'. This mapping serves two related functions. One is to provide qualitative prohibitions that enable us to exclude certain explanations of the observed genetic structure. So, for example, if we observe long-term stability and universal polymorphism for the Rh blood group in *Homo sapiens*, we cannot explain it by selection against the offspring of incompatible matings, because the mathematical analysis of such a selection process shows that it will lead to fixation of alleles. Or, if all the genetic variance for a trait is dominance variance, no argument about the relative advantage of increasing or decreasing the trait is relevant, because selection cannot be effective in the absence of additive variance. Nor are we allowed to explain a stable polymorphism as a consequence of random uncorrelated variation in selection coefficients.

More important, population genetic theory, precisely because of its weak inferential power, shows that claims for unambiguous explanations of observations are usually wrong. Claims that a certain pattern of genetic differentiation within and between populations are due to natural selection can almost always be shown to be too strong, because the same pattern can be produced by random drift and migration. So, for example, the claim of Prakash & Lewontin (1968) that the association between allozyme alleles and inversions in *Drosophila pseudo-obscura* and *D. persimilis* could only be the result of selection was contradicted in a stochastic analysis by Nei & Li (1975). In general, claims that differences in characters between species are unequivocal evidence that selection has been differential in the species are contradicted by the stochastic theory of population genetics of multiple locus traits. The theory of selection in linked multilocus systems which has become highly developed in the last 20 years does not have as its purpose the explanation of observed polymorphisms, but the addition of another mechanism which, under appropriate conditions of the parameters, must be taken into account in explanation. The theory of Evolutionarily Stable Strategies (ESS) must be seen in the same light. It is extremely unlikely that the necessary fitness parameters will be determined so as to show that particular behavioural repertoires are, in fact, evolutionarily stable. Rather,

ESS theory shows what is possible and what is impossible given various ranges of parameters. It tells us to be surprised at some outcomes but not surprised at others.

The delineation of the prohibited and the possible is the function of population genetic theory. The revelation of the actual is the task of population genetic experiments, a task that such experiments can accomplish provided they are freed of their strong dependence on the quantitative and statistical relations predicted by theoretical formulations and instead are constructed to provide unambiguous qualitative information. The possibility of the reorientation of experimental work, and the way such a change can give unambiguous evidence is shown by the history of studies of molecular polymorphism in natural populations.

Since 1966, when gel electrophoresis of proteins was introduced as a technique for studying genetic variation in natural populations, a very large portion of the work in experimental population genetics has been concerned with the study of enzyme polymorphisms. A general result has emerged that in a typical species about one third of structural gene loci studied show some polymorphism and that a typical individual is heterozygous at about 10% of its loci. These figures vary from species to species: mammals are somewhat less polymorphic than the average, insects rather more; an occasional species is reputed to be nearly totally without variation; one species of *Drosophila* is reported to have 85% of its loci polymorphic, but the modal figures are a reasonable characterization of vascular and non-vascular plants, bacteria, vertebrates and invertebrates of all sorts. The monotonous repeatability of this observation in so many different organisms has resulted in the major preoccupation of population genetics with an explanation of the polymorphism. Three approaches to explanation have been taken. One has been to compare gross statistics like the average heterozygosity or the total number of alleles segregating per locus, or the proportion of loci which are polymorphic, with theoretical predictions generated from selective and stochastic hypotheses. A second, more fine-grained, statistical approach has been to use the distribution of allelic frequencies within populations and the comparison of the distribution between populations to compare with predictions generated by competing hypotheses. A third, non-statistical, approach has been to attempt the direct measurement of selective differences or at least physiological differences between various genotypes segregating at structural gene loci. None of these attempts has been satisfactory. The statistical approaches have lacked the necessary discriminatory power for the reasons I have discussed above. The physiological approach has sometimes demonstrated selection but most often has failed because the necessary experimental power needed to detect the expected small selective differences between the genotypes has not been practical (Lewontin, 1974).