AMER. ZOOL., 24:451–464 (1984)

Science as a Way of Knowing—Molecular Evolution[1]

BRUCE R. LEVIN

Department of Zoology, University of Massachusetts, Amherst, Massachusetts 01003

SYNOPSIS. The teaching of evolution in introductory biology courses is critically discussed and those aspects of molecular evolution believed to be important for inclusion in these courses are briefly reviewed. Consideration is given to the use of biochemical and molecular biological procedures for studies of phylogenetic relationships and genetic variability, "classical molecular evolution and molecular population genetics," as well as the evolutionary basis of molecular biological phenomena, "neoclassical molecular evolution." The major results of classical molecular evolution and molecular population genetics studies are summarized and their implications discussed. Particular consideration is given to the neutralist—selectionist controversy and the processes accounting for differences in rates of organismal and molecular evolution. The scope of neoclassical molecular evolution is outlined and two subjects within this realm considered in some detail: i) experimental studies of the evolution of new genes and functions, and ii) the existence, maintenance and evolutionary roles of parasitic ("selfish") DNA and infectious inheritance.

INTRODUCTION

I have been assigned the task of considering the teaching of molecular evolution and the contributions of molecular biology to evolutionary theory. Thus, ". . . I have to be molecular. Who is not?" (A. Lwoff, 1966). While I shall fulfill my obligation to this charge, I have elected to interpret it rather broadly. I have also succumbed to the temptation of using this essay as a platform to display my biases about the teaching of evolution and the nature and scope of molecular evolution.

While this essay includes elementary reviews of specific areas of molecular evolution and molecular population genetics, it is not intended as a comprehensive review of the subject at large. I have concentrated on those areas, that for subjective as well as objective reasons, I believe should be included in introductory courses.

SOME OBSERVATIONS AND PREJUDICES

While evolution may well be the thread that ties all of biology together, concern about the fabric of the subject seems to have had little play in much of modern biology. There are professional biologists who would be indifferent to the title and substance of Theodosious Dobzhansky's 1973 essay "Nothing in Biology Makes Sense Except in the Light of Evolution." Indeed, as I found the other day, when speaking with a bright, and not-that-young, molecular geneticist, there are biologists out there who have never heard of Professor Dobzhansky. One can be a successful practitioner of many areas of contemporary biology without considering how organisms, molecules or phenomena came to be or their descent relationships. A relative absence of interest in evolution prevails in a number of areas of biology, with high-tech molecular biology being the most prominent of them. There are, of course, two possibilities: evolutionary theory and evolutionary considerations are of little utility in the study and understanding of many biological phenomena, or those of us who believe they are have failed to get our message across. I accept the latter.

It is my feeling that the limited interest in evolution and, in some cases, limited respect for its study as a scientific endeavor, is a reflection of how we teach it at the introductory level and beyond. The past 20 years have witnessed an extraordinary improvement in the quality and scope of introductory biology texts; a phenomenon that can itself be interpreted as a response to an intense selective pressure (Cox and May, 1982). Nevertheless, even in the apparent ancestor of this new wave of introductory texts, *Life* by G. G. Simpson,

[1] From the Symposium on *Science as a Way of Knowing—Evolutionary Biology* presented at the Annual Meeting of the American Society of Zoologists, 27–29 December 1983, at Philadelphia, Pennsylvania.

C. Pittenbrigh and I. H. Tiffany (1957), evolution is treated as a separate subject. Sometimes it is briefly considered in introductions and almost always relegated to final chapters for more comprehensive treatment. If evolution is the thread that ties it all together, and if without considering evolution nothing (or more realistically, little) in biology makes sense, evolutionary considerations should be present thoroughout texts and courses; a point also made by Dr. Moore and Dr. Wake in their essays.

To be sure, treating the evolutionary aspects and implications of all areas of biology covered would add material to the already excessive quantity introductory courses attempt to cover. However, many of these evolutionary considerations could be little more than vignettes, *e.g.*, the evolutionary reasons for anticipating a universal genetic code considered with the presentation of the code itself. By spreading the treatment of evolution throughout, it would be possible to cut down on its separate treatment. Under any condition, if the inclusion of evolutionary considerations throughout results in an increase in the ratio of concepts to facts,[2] then the intellectual yield of these introductory courses would necessarily be augmented.

In the main, the treatment of natural selection and evolution at the elementary level is extremely conservative. The history of a subject and some consideration of the personalities of its practitioners are useful pedagogical tools and, doubtless, increase student interest and appreciation for science as a human endeavor—I could not conceive of teaching Genetics from any but a historical perspective. On the other hand, the use of history and personality in the treatment of evolution at the introductory level sometimes borders on scholasticism, with Darwin playing the role of Aristotle. Do we really need tradition and authority to convince students of the importance and utility of evolution to the study of biology? Perhaps more than any other area of biology (I do not know for sure), evolutionary biology has maintained the school of thought advocacy tradition of earlier eras. Some hypotheses are more often championed than tested. Positions (almost always two for any given issue) are fiercely defended, even when they are not mutually exclusive. New ideas are liable to attack for no reasons other than their real, or even apparent, violation of orthodoxy. Could it be that our pedagogical preoccupation with tradition and personality has played a role in this?

The conservatism in the treatment of evolution at the introductory level is also reflected in the concentration on seemingly solved problems. As practicing scientists, much of what we talk about, think about and occasionally even work on are problems which are not solved. Much of the pleasure of being an evolutionary biologist is speculating on the answers to unanswered questions. Nevertheless, in teaching evolution at the introductory level, we tend to avoid giving the students the opportunity to participate in these speculations or, even worse, suppress speculation by presenting pat nonsolutions—to see what I mean, just consider the treatment of the evolution of sex in most introductory texts. While the display of humility in the face of unanswered questions may tarnish our facade of authority, the recognition of our limitations and the realization that they can make significant contributions to speculative discussions, will certainly increase student interest and self-confidence. Indeed, with their relative absence of preconception and bias, their contributions are likely to be even greater than ours.

As so clearly illustrated in Dr. Moore's essay, the study of evolution can be a scientific endeavor, even fulfilling the somewhat narrow "strong inference" criteria of Platt (1964). Nevertheless, in teaching evolution at the introductory level, the message that the study of this subject can be

[2] Perhaps, as minimal rule of pedagogical hygiene, we should not hold students responsible for memorizing facts that we have to relearn and rehearse before presenting. In their evolution, introductory texts have become increasingly encyclopedic. Thus, we can take comfort in knowledge that students have ready access to the facts, and can relax our efforts to cram these facts into their short-term memory.

scientific may well be garbled due to excessive noise. In our efforts to display the facts of evolution, we sometimes give too little consideration to the predictive elements of evolutionary theory and testable hypotheses generated from it. This is particularly true in introductory laboratories.

NATURAL SELECTION AND ITS LIMITATIONS

It is my feeling that the basic elements of natural selection should be presented very early in introductory courses. While the details of the genetic basis of natural selection require some understanding of transmission genetics, the process can be taught and well understood without formal genetics (Darwin certainly did well without it). I also believe that natural selection should be taught in a more general framework than it has been traditionally. While the statement may seem heretical, the pure Darwinian view of natural selection, which commonly prevails in our teaching of the subject at the introductory level, is unnecessarily restrictive. It is a universal process ("law"?), a property of transmitted (inherited) variation of replicating systems: molecules, cells, organisms, populations and communities. It will operate as long as the inherited variation is reflected as differences in rates of replication, be there a struggle for existence or not.

In addition to considering the process and potential of natural selection, we should also treat its limitations early in the course. It is not omnipotent and may well be quite myopic. As eloquently stated by King and Jukes (1969), "natural selection is the editor rather than the composer, of the genetic material. One thing an editor does *not* do is to remove changes which it does not perceive." It is also a process that is constrained by the evolutionary history of the population, the genetic basis of phenotype determination and conflicting selection pressures. Potentially fit variants may not be readily generated or may only be produced through a progression that includes variants that are less adapted than the parental forms. In cases where selection simultaneously operates at more than one level of replication, as it almost invariably will, phenotypes favored at one level, *e.g.*, the population, may be at a disadvantage at another, *e.g.*, the organism.

While the theory of natural selection can be taught and readily understood in a totally hypothetical framework, examples are of course useful. It is, however, critical that at least some of these examples provide direct evidence for the action of natural selection rather than being just *a posteriori* interpretations of phenomena, *i.e.*, "just so stories" (Gould and Lewontin, 1979). For students in introductory courses it would also be helpful to include examples with broad relevance. One example fulfilling these criteria is antibiotic resistance in natural populations of enteric bacteria. There is direct evidence for the existence of antibiotic resistant strains before human use of antibiotics and extensive documentation for the rapid increases in both the frequency of resistant strains and range of resistances borne by individual bacteria, following the clinical and prophylactic use of these substances (reviewed in Anderson, 1969; Falkow, 1975). The antibiotic resistance example also illustrates the various levels of replicating units and levels of action of natural selection. While some antibiotic resistance can be attributed to classical mutation-selection processes, most clinically important antibiotic resistance is coded for by genes borne on plasmids, *i.e.*, autonomously replicating genetic molecules. Furthermore, many of the actual genes responsible for the resistance are themselves parts of semiautonomous genetic molecules, transposons, which can be infectiously transmitted between plasmids and the host chromosome (Falkow, 1976; Cohen, 1976; Kleckner, 1978; Broda, 1979).

Introductory laboratories are not particularly amenable to high-tech molecular evolutionary investigations or long-term experimental studies. However, by using bacteria as a model system, some of the technical and temporal constraints precluding good experimental studies of evolutionary processes can be overcome. One of the most fundamental hypotheses of evolutionary theory, the preadaptivness (randomness) of mutations, can be directly

tested with the fluctuation test experiment (Luria and Delbruck, 1943). Within the course of a single laboratory, it is possible to demonstrate the basic elements of the classical theory of natural selection, exponential growth and the "struggle for existence." By extending these exercises to include a series of laboratory sections, in the course of a week, the process of adaptive evolution through mutation and selection could be demonstrated. I will gladly supply more details about these evolutionary exercises to interested readers.

With a more comprehensive view of natural selection and its limitations, selective neutrality, sexual selection, regressive evolution, parasitic ("selfish") DNA, transposition, and individually deleterious ("altruistic") characters are no longer special cases of evolution, but rather anticipated variations of a common principle. With a more pluralistic view of this process, we may even achieve some success in the campaign to purge the good-of-the species, progressive and Panglossian interpretations which continue to plague professional as well as popular considerations of evolution by natural selection.

CLASSICAL MOLECULAR EVOLUTION AND MOLECULAR POPULATION GENETICS

The majority of that which we currently consider to be Molecular Evolution and Molecular Population Genetics is in fact the application of biochemical and molecular biological technology to traditional evolutionary problems, determinations of phylogenetic relationships and the nature and magnitude of genetic variation in populations. While the evolution of molecules is observed in these studies, interest in molecules themselves and the molecular biological and biochemical implications of this evolution is secondary. For convenience and without pejorative implications, I shall refer to these endeavors as classical molecular evolution and molecular population genetics.

Some technological considerations

Since the turn of the century, immunological procedures have been used in taxonomic and systematic investigations (Nuttall, 1904). The well justified assumptions being that the phylogenetic relationships would be reflected in the extent of antigenic similarity of homologous proteins. The history of these serological studies of taxonomy and systematics is characterized by successive technological improvements increasing the precision of the quantitative estimates of phylogenetic distance. With the use of microcomplement fixation, these methods reached their peak of precision, in some cases being able to distinguish between proteins differing by single amino acids (Cocks and Wilson, 1969). At approximately the same time that the discerning powers of these immunological procedures were approaching their apex, amino acid sequence methods were being developed and comparative amino sequence data were becoming available for evolutionary studies. By the late 1960s, a large number of proteins from many different species were sequenced and catalogued (Dayhoff, 1972). For a review of these protein studies of molecular evolution see Wilson *et al.* (1977).

While the protein data provided quantitiative information about the degree of genetic identity of related species, to some extent the analysis of proteins was destined to be a transient period in the history of classical molecular evolution. Ultimately phylogenic relationships had to be based on DNA and, shortly after the appropriate quantitative procedures for DNA manipulation became available, they were applied to evolutionary studies. Initially, this was accomplished through quantitative determinations of the stability of inter- and intraspecific hybrid DNA molecules generated *in vitro* (Kohne, 1970). This was followed by the analysis of the fragments generated by treating defined blocks of DNA with restriction endonucleases (Kan and Dozy, 1978). With the development of procedures for identifying and isolating specific regions of DNA, cloning those DNAs with bacterial host-vector systems, and relatively facile methods for DNA sequencing (Maxam and Gilbert, 1977), it became possible to directly determine the degree of identity of homologous regions of DNA. While this, the seemingly ultimate technology for the molecular study of phylo-

genetic associations, has only recently begun to be applied for this purpose, central sources for the compilation and dissemination of these data are already in place and these data can be acquired on computer tapes from commercial sources.[3]

The development of the technology of Molecular Population Genetics was essentially parallel to that of that of molecular evolution. Serological data, inherited antigenic differences (primarily blood groups), were used for studies of variation within populations. Starting in the 1960s, protein electrophoresis began to be used for these types of investigations. This procedure provided relatively unbiased estimates of variation in structural genes that code for water soluble proteins (Harris, 1966; Lewontin and Hubby, 1966). It also permitted estimates of phylogenetic relationship between closely related taxa (Nei, 1971), which could not be accomplished readily with immunological and protein sequencing procedures. As was the case with empirical studies of molecular phylogeny, these protein based procedures were destined to be a tranisent phase.[4] Shortly after relatively facile procedures for the determination of quantitative differences in DNA became available, they were applied to studies of genetic variation. At first this was through the analysis of restriction enzyme fragments. Very recently, however, studies of within species variation in the DNA sequence of specific genes have been published (Kreitman, 1983; Milkman and Crawford, 1983).

Most of the molecular evolution and population genetic studies in eukaryotes have been concerned with chromosomal genes and chromosomal DNA. However, there have also been a number of studies of evolution and variation in mitochondrial DNA. In most cases these are accomplished through restriction fragment analysis (see reviews by Brown, 1983; and Avise and Lansman, 1983). DNA evolution in these maternally inherited extrachromosomal elements occurs at a relatively more rapid rate than that of chromosomal DNA and, as a result, mitochondrial DNA data permit relatively more precise estimates of genetic distance among closely related taxa (Ferris *et al.*, 1981).

The results and their implications

In many ways, the classical molecular evolution studies provide the single most compelling and readily presented evidence for the fact of evolution. Descent relationships are readily illustrated by comparing the amino acid sequences of the same protein in different species. For some proteins such as cytochrome c, these comparisons can be made for organisms in very distantly related groups, *e.g.*, yeast and humans. While proteins vary considerably in their rates of change, with only minor exceptions, the relative extent of sequence divergence among taxa remains constant and independent of the protein considered. At this juncture, I am unaware of any major differences in the phylogenies generated by adequately comprehensive studies employing both protein and DNA data. Nevertheless, it is clearly prudent to wait until a sufficient number of the returns are really in.

While the molecular evolution data added only additional support to the already overwhelming evidence for the fact of evolution, they had a profound effect on our view of the tempo and mode of the evolutionary process. Regressions between the number of nucleotide differences, estimated from the amino acid sequence data, and the divergence time of the organisms being compared, as determined from fossil evidence, are essentially linear; indicating that the rate of protein evolution was approximately constant in absolute (calendar) time. This observation suggested that the protein data can be used to estimate divergence time of extant organisms, *i.e.*, act as a "molecular clock" (Zuckerkandl and Pauling 1965; Fitch, 1976; Thorpe, 1982). Although in the great

[3] Bolt, Beranck, and Newman. Genebank Research Systems Division, Cambridge, Mass.

[4] While there may well be pressure to use the most up-to-date technology for many classical molecular evolution and molecular population genetic studies, older procedures are adequate and may well be more efficient.

majority of cases, the branching points of phylogenies generated with molecular data are consistant with those based on traditional, morphological criteria, for some groups there were very substantial differences in the apparent rates at which morphological and molecular evolution proceeded (Wilson *et al.*, 1977). Among the most dramatic of these differences in the rates of organismal and molecular evolution was that for the higher primates (for an interesting, and less technical account see Gribbon and Cherfas, 1982). Based on objective morphological criteria, humans and chimpanzees are placed in different families; if, however, these taxa were based solely on molecular data, humans and chimpanzees would be closely related members of the same genus (King and Wilson, 1975).

At the time that enzyme electrophoresis was first being used for population studies, there was considerable controversy about the magnitude of standing genetic variation maintained in natural populations. Morphological and serological procedures for studying genetic variation were necessarily biased; genes had to be polymorphic to be considered (however, see Lewontin, 1967). Although quantitative genetic (artificial selection) studies suggested that there was a substantial amount of standing genetic variation in natural populations of outbreeding organisms, the absolute extent of this variation was unknown. Protein electrophoresis represented the first method to allow for a relatively unbiased estimate of genetic variation. The results of these studies suggested that standing genetic variation could be quite substantial. On the average, outbreeding eukaryotic organisms are polymorphic (rare alleles with frequencies in excess of 5%) for approximately 30% of structural gene loci with the average degree of heterozygosity (genic diversity) being between 10% and 15% (see review by Nevo, 1978). The only prokaryotic species adequately studied in this manner, *E. coli*, has between four and five times as much genic diversity as outbreeding eukaryotes (Selander and Levin, 1980).

The neutralist-selectionist controversy

In many ways the results of these molecular studies ended an "academic"[5] period that population genetics was in at the time these procedures began to be used. The high rate of gene substitution, the relative constancy of this rate, and the substantial quantity of genic variation indicated by the molecular data came as surprises. These observations were not anticipated from the mathematical models of directional and stabilizing selection that were favored at the time or the prevailing interpretations of this formal population genetic theory. To some extent these deviations from theory were an artifact of the procedures used to calculate relative fitness. With different, but not necessarily more realistic, models, the observed frequencies of polymorphic loci could be accounted for by natural selection (Sved *et al.*, 1967; King, 1967; Milkman, 1967; Sved, 1968). On the other hand, these molecular data could also be accounted for by what is effectively a null hypothesis, *i.e.*, recurrent mutation and random genetic drift of *selectively neutral* alleles (Kimura, 1968; King and Jukes, 1969; Kimura and Ohta, 1971).

The recognition that the same body of molecular evolution and polymorphism data can be explained by diametrically opposing hypotheses, initiated the "neutralist-selectionist controversy," which has virtually dominated molecular population genetics for the past 15 years. The observed constancy of the rate of protein evolution could be accounted for by the neutral gene hypothesis, but was also consistant with the hypothesis of selection operating at a constant level, if averaged over long periods of time. Although the action of selection could be inferred for some protein polymorphisms, for most, the neutral gene null

[5] The word "academic" is used as in Stent (1969, 1978), as one of the phases in the history of a science; classical, romantic, dogmatic and academic. However, in Stent's view a science proceeds through this sequence only once. It is my impression that evolutionary biology cycles through this progression. Indeed, the same may be true for Molecular Biology, whose obituary Stent wrote in that 1969 essay.

hypothesis could not be rejected. While the neutral gene theory made very specific predictions about the number of alleles and distributions of allelic frequencies for polymorphic loci, the estimates of the parameters needed to test the fit of these distributions were difficult to come by and almost always equivocal. Thus, the controversy raged, techniques were improved, more data were gathered, additional models were developed, but an absolute resolution was not in the offing (for reviews see Lewontin, 1974; Nei, 1975, 1983; Clarke 1979; Nei, 1983; Kimura, 1983*a*, *b*).

It is my feeling that the neutralist-selectionist controversy is more a product of the sociology of science (the two camp-advocacy approach) than its substance. From a very superficial perspective, the neutral gene hypothesis appeared to be a challenge to Darwinian and neodarwinian orthodoxy. In point of fact, it is not and never was. Even its most staunch advocates, present it as a pluralistic hypothesis that accepts the constraints that natural selection imposes on genes and their products and sees natural selection as the unique force of adaptive evolution (see Kimura, 1983*a*, or more extensive, 1983*b*). It differs from naive selection theory only in the fact that it accepts the perceptive limitations of natural selection. Thus, even from the start, the neutral gene-selection question was a quantitative one of the relative contributions of random drift of selectively neutral alleles, and directional and stabilizing selection to molecular evolution and polymorphism. It remains to be seen whether future evidence will support the hypothesis that "the great majority of evolutionary changes at the molecular level are not caused by Darwinian selection but by the random fixation of selectively neutral or nearly neutral mutations" (Kimura, 1983*a*). I personally find the current evidence in support of this hypothesis to be very compelling.

Unequal rates of organismal and molecular evolution

To my mind, the most intriguing result of the molecular phylogeny studies was the observation that organismal and molecular evolution may proceed at different rates. This not only questioned the validity of divergence time estimates obtained from traditional, morphological procedures, but it also posed an additional challenge to the orthodox Darwinian and neodarwinian view of phyletic evolution as a smooth process operating through the gradual accumulation of allelic differences (Gould, 1980). These data also very clearly illustrated just how little is known about the genetic, molecular and developmental basis of morphological variation.

Could it be that changes in structural genes of the sorts being considered in classical molecular evolution studies are *not* those primarily responsible for major morphological differences in closely related organisms? It seems almost certain that the answer to this question is yes. To be sure, it is possible that small modifications in the activities (but not the functions) of enzymes, or in the structures of nonenzymatic proteins, that may result from single nucleotide substitutions, may yield major morphological modifications. However, changes in the genes regulating the time of action and/or level of product formation of these structural genes seem far more reasonable as candidates for the genetic basis of major phenotypic differences (Wilson, 1976; Wilson *et al.*, 1977).

At this time, eukaryotic developmental biology is not yet at a stage where a direct test of this *regulatory gene hypothesis* is possible. However, there is clear evidence for genes with regulatory functions in eukaryotes. One fine example of this is the bithorax complex of 12 or more loci which play a role in the development of the major body segments of *Drosophila melanogaster.* Flies homozygous for mutant alleles at two loci in this complex have four wings, rather than the one pair of wings-one pair of halteres as is the case for wild-type *Drosophila* (Lewis, 1963; and review by Hunkapillar *et al.*, 1982). There is also circumstantial evidence in support of the hypothesis that regulatory gene changes play a major role in the determination of species differences. The concentrations of homologous pro-

teins in the same tissues of different species can differ by a factor of ten or more (Wilson *et al.*, 1977). The results of experimental studies with bacteria also provide evidence for the potentially dominant role of regulatory gene changes in adaptive evolution. When wild-type, lac inducible *E. coli* are grown in lactose limited chemostats, constitutive mutants rapidly evolve (Horiuchi *et al.*, 1962). While these regulatory gene changes may be supplemented by increasing the number of copies of genes in the lac region, to increase β-galactosidase concentration, Horiuchi and his colleagues failed to find any evidence for changes in the structural gene coding for this enzyme. The studies of the evolution of new metabolic functions in bacteria, described in the following section, provide additional evidence for the importance of regulatory mutations in adaptive evolution.

NEOCLASSICAL MOLECULAR EVOLUTION

Early in the classical period of molecular evolution and molecular population genetics it was reasonable, or at least convenient, to give only secondary consideration to the molecular biological and biochemical processes being studied. However, within short order, it was apparent that a deeper consideration of these processes was required for interpreting the results of these evolutionary studies. Part of this more ecumenical perspective came as a direct by-product of the neutralist-selectionist fray. It was clear that more had to be known about the constraints natural selection posed on the products of structural genes. This stimulated more detailed studies of the physiological, biochemical and ecological basis of selection of polymorphic enzymes of known function (see review by Koehn *et al.*, 1983, and the studies by Dykhuizen and Hartl, 1980). As a consequence of exposure to the technology of molecular biology, evolutionary biologists, were seduced into considering the evolutionary basis of some of its phenomenology, see for example the transposon studies of Biel and Hartl (1983) and Chao *et al.* (1983). In addition, and largely independently of the work proceeding in classical molecular evolution and molecular population genetics, biochemists and molecular biologists were studying or speculating on the evolutionary aspects of biochemical and molecular biological phenomena. Included among these endeavors were theoretical considerations of: i) the evolution of the genetic code (Jukes, 1983); ii) the evolution of accessory elements in bacteria (Campbell, 1981, 1983); and iii) the origin and evolutionary role of exons and the phenomenon of split genes (reviewed in Hunkapillar *et al.*, 1982). Also included in this category are various studies of the molecular basis of the origin of life. For convenience, I shall refer to investigations, where primary consideration is the evolutionary basis of molecuar biological phenomena, as neoclassical molecular evolution.[6]

The scope of neoclassical molecular evolution is substantially broader than that of classical. Thus, even in this superficial review, it is necessary to limit my consideration of it. The neoclassical evolutionary problems considered below were chosen because: i) I find them to be particuarly interesting; ii) I believe that they should be considered in introductory courses; and iii) they are not now given adequate (or, in most cases, any) treatment in introductory texts.

The evolution of new genes and phenotypes

For the most part, the mechanisms by which new genes and new phenotypes evolve have been restricted to speculation supplemented by primarily *a posteriori* interpretations of evolutionary processes (see for example Horowitz, 1965; or Jensen, 1976). There have, however, been a number of elegant experimental studies of gene evolution using bacteria. The results of these are intriguing in their own right, have important implications for some observations of classical molecular evolution, and illustrate the potential for exper-

[6] In the Stent (1969, 1978) interpretation, the classical period would be followed by a romantic. Maybe so!

imental studies of molecular evolutionary processes.

These studies of gene evolution in bacteria have been primarily concerned with the acquisition of new metabolic functions. Most commonly, they take the form of experimental sequences intiated by challenging populations of bacteria with substances that they are incapable of metabolizing, "novel" substrates, as unique resources and letting mutation (sometimes at augmented rates) and selection play out their roles. Occasionally, but certainly not invariably, mutants capable of growth on the novel substrate are isolated. In most cases, the capacity for growth on the new substrate by the initial mutant is modest and open to improvement. The experimental sequence then proceeds with additional rounds of mutation and selection to augment the rate of growth on the novel resource. These "Directed Evolution" studies have been reviewed by Clarke, (1978); Mortlock (1982); and Hall (1983).

In these experimental studies, new functions are acquired and improved on by both regulatory and structural gene modifications. This is illustrated in the pioneering investigations of this type by E. C. Lin and his colleagues on Xylitol utilization in a strain of *Klebsilla aerogenes.* The parental strain used in this experimental sequence could take up Xylitol by diffusion, but could not grow on it. The first mutant capable of growth on Xylitol acquired this capacity through the constitutive production of the hydrolytic enzyme necessary for the fermentation of Ribitol, ribitoldehydrogenase, RDH. While RDH normally was capable of oxidizing Xylitol as part of the D-arabitol pathway, Xylitol does not induce the D-arabitol operon. By starting the second cycle of mutation and selection with these first order xyl+ mutants, Lin and his colleagues were able to obtain Xylitol fermenting mutants with 2.5 fold higher growth rates on Xylitol than the parental mutant. These second order mutants were the result of a mutation at the RDH structural gene which increased the activity of RDH enzyme on the Xylitol, but not on the Ribitol substrate. Mutations leading to further increases in growth rate on Xylitol were generated from these second order xyl+ strains. In this case, the phenotype change was acquired by regulatory mutation at the D-arabitol operon leading to the constitutive production of the D-arabitol permease, which was also capable of transporting Xylitol.

While the details of the scenarios by which the capacities to utilize novel substrates are acquired vary among species and substrate, the combined role of regulatory and structural gene changes are essentially typical. This also seems to be the case when the substrate is not novel to the organism, but where the structural gene required for fermentation is deleted. One example of this is lactose metabolism in *E. coli* via the ebg operon. Mutants of this type were first isolated by Campbell *et al.* (1973) and the evolution at the ebg operon has been extensively studied by B. Hall and his colleagues. In this situation, the original strain was deleted for the lacZ (β-galactosidase) locus and the acquisition of lactose fermenting phenotype required mutations at two loci, i) a structural gene, ebgA, that was responsible for the synthesis of the hydrolytic enzyme EBG (for evolved β-galactosidase), and ii) a regulatory mutation, ebgR that was necessary to increase the level of EBG synthesis. The ebgA and ebgR loci are two of three genes in an operon that is genetically very distant from that of the classical Lac operon (66 min as opposed to 7 min the *E. coli* K-12 chromosome) and the EBG hydrolytic enzyme has no apparent homology with the lacZ β-galactosidase.

By further rounds of mutation, recombination and selection, additional ebgA and ebgR alleles were selected using a variety of different β-galactosides as challenging substrates. In this manner, Hall and his colleagues obtained mutants of regulatory as well as structural genes with products that varied in their substrate affinities. One of the selected ebgA mutants was capable of converting lactose into a substrate capable of inducing the classical Lac operon, something the original ebgA could not do, but classical lacZ β-galactosidase could.

I see two major messages to be learned

from these studies of directed evolution in bacteria. First, they suggest that regulatory genes play a major role in the process leading to the evolutionary modification of structural genes. Second, they suggest that the affinities of enzymatic and regulatory proteins can be quite malleable. By changes in the concentrations of these enzymes or by modest changes in their structure, such as those acquired by single nucleotide substitutions, the role of these structual gene products can be dramatically changed.

Molecular perversities: Parasitic DNA and infectious inheritance

Until very recently, most evolutionary theory was restricted to chromosomal, stay-at-home, genes that are transmitted vertically between generations. This is, indeed, a restrictive view of the genome and the evolutionary process. The bacteria abound with autonomously replicating DNA molecules, plasmids and phage, as well as a variety of semiautonomous DNAs, insertion sequences, transposons and vestigial prophage. Through either their own devices or by hitchiking, these autonomous and semiautonomous genetic molecules can be infectiously transmitted between cells. While some of these "accessory" genetic elements may not determine specific host-manifest phenotypes, it is clear that as a class they play a significant role in the adaptation and evolution of bacteria. Some do code for potentially adaptive host phenotypes (antibiotic resistance is only one of many extrachromosomally determined characters) and conjugative plasmids and phage serve as vehicles for the infectious transfer of chromosomal DNA (Falkow, 1975; Reanny, 1976; Campbell, 1980).

The significance of autonomous and semiautonomous DNA and infectious inheritance in the adaptation and evolution of eukaryotes is just beginning to be evaluated. It is clear that analogous genetic elements are present in eukaryotic cells and that multicelled higher organisms have substantially more DNA than they appear to need. Some evidence and compelling arguments suggest that much of this accessory DNA is nonfunctional, essentially parasitic ("selfish"), and is present as an artifact of evolution at the subcellular level (Doolittle and Sapienza, 1980; and Orgel and Crick, 1980).[7]

Although the phenomena of parasitic DNA and infectious inheritance have yet to work their way into the classical evolutionary literature, it is my feeling that some consideration of them should be included in introductory biology courses. These DNAs serve as examples of molecular evolution at the subcellular level and, in spite of their selfish nature, they and their capacity for infectious (horizontal) transmission may play an important role in organismal evolution.

Natural selection and the mainienance of parasitic genetic molecules

It seems reasonable to assume that as a consequence of various kinds of replication errors, renegade copies of DNA (as well as RNA) will be continuously generated. While some of these additions to the genome may be selected for at the organism level, *e.g.*, by gene amplification, others may not be under positive host selection. However, if the rate of replication of these analogs of selfish DNA exceeds their rate of loss, they can become established and will be maintained in that organism. There are essentially two non-exclusive mechanisms by which this "overreplication" (Campbell, 1981) can obtain: i) by coding for some phenotype which enhances the fitness of their host; or ii) by infectious transmission. If the former mechanisms are involved in the maintenance of these DNAs, then from the perspective of their host cell, they are not "selfish." While innocuous (selectively neutral) parasitic DNA may persist for extended periods of time, as long as there is some finite rate of loss, they would require continuous de novo production for their persistence.

---

[7] While generally not presented in that light, all DNA not coding for phenotypes that are under direct selection at the level of the organism or the population, fulfill this parasitic-selfish criteria. This includes the DNA produced by various kinds of duplication processes as well as the "pseudogenes" that most likely arise by relaxation of selection pressures (Ohno, 1970; Li, 1983).

Natural selection would, of course, favor mechanisms for overreplication in parasitic genetic molecules. This was recognized by Doolittle and Sapienza (1980) and Orgel and Crick (1980) and was fundamental to their "selfish DNA" intepretation (also see Dawkins, 1982). On the other hand, the fact that overreplication would be favored is not a sufficient condition for the maintenance of these elements. Natural selection in this situation is necessarily a multilateral process involving the immediate host cell, the whole organism (if multicellular) and, to a lesser extent, the population. In cases where the cell is the whole organism, as it is for bacteria, this coevolutionary process is a seemingly simple matter and, on *a priori* grounds, there are conditions where these elements can be maintained solely by infectious transfer (reviewed in Campbell, 1981 and Levin and Lenski, 1983). There have not been many formal considerations of the conditions for the maintenance of selfish genetic molecules in eukaryotes. It is, however, conceivable that some become established and are maintained by distorting segregation ratios in their favor (Hickey, 1982), or by infectious transmission as viruses. It is also possible that these elements are sufficiently innocuous to become established in populations solely by recurrent de novo synthesis in a manner analogous to that of selectively neutral alleles (Kimura, 1968). In the latter case, they would accumulate at a rate equal to their rate of production.

Selfish DNA, infectious inheritance and adaptive evolution

From the perspective of organismal evolution, parasitic DNA may have absolutely no role. It could be no more than artifact of natural selection operating at the molecular level and the constraints and myopia of natural selection operating at higher levels. Thus, as suggested by Doolittle and Sapainza (1980) and Orgel and Crick (1980), quests to ascertain the phenotype of these DNAs could be in vain. On the other hand, the fact that these DNAs do not determine specific host phenotypes does not necessarily imply that they have no role in the adaptive evolution of their hosts. Parasitic DNA is the raw material of genome expansion and, in this light, adaptive evolution operating through genome expansion is a byproduct of the subcellular evolutionary processes leading to the establishment and maintenance of parasitic DNA (also see Cavalier-Smith, 1978, 1980). There are compelling arguments and some evidence that transposible parasitic DNAs play a role in hybrid dysgenesis and thus may be important in the speciation process (see reviews by Rose and Doolittle, 1983*a*, *b*). Rose and Doolittle (1983*a*) have also suggested that selection for horizontal transmission by parasitic DNAs could have resulted in the evolution of sex in eukaryotes (perhaps, the major unsolved evolutionary problem). This, is almost certainly the case for plasmid and phage mediated gene transfer in bacteria, where chromosomal gene recombination (sex) is an accident of the infectious transmission of these elements (Levin and Lenski, 1983).

One of the most impressive aspects of biological evolution is the production of the extraordinary amount of diversity and complexity of organisms in the relatively short span of time available. Even accepting the possibility that the processes leading to punctuated branching in phylogenies also augments the rate of evolution, it seems worth questioning whether the diversity and complexity of organisms has been generated solely through evolution along species lines. It is clear that the rate of evolution would be accelerated if vertical genetic transmission was supplemented by horizontal (infectious) gene transfer between lineages. In this way, phenotypes that evolve in one lineage could be precipitiously acquired by members of other lineages. In the bacteria, it is clear that this process does operate (Reanney, 1976). In one or more steps, plasmids and phage can be transferred between genetically very different hosts. In the course of their travels, they can pick up transposons and chromosomal genes from one host and transmit them to another. With appropriate selection pressures, these acquired genes may become incorporated in the recipient genome.

At this time, we know relatively little

about the role of infectious inheritance in the evolution of extant higher eukaryotes. While this process may be important for unicelluar eukaryotes, on the face of it, it seems unlikely that it plays a very significant role in the adaptation and evolution of multicellular eukaryotes. Although there are viruses that could conceivably serve as vectors for gene transfer, if pathological viruses are typical of viruses in general, the host range of these vectors would be limited. Furthermore, even if gene transfers did occur, for these genes to become incorporated into the host genome, they would have to work their way from the infected somatic cells to the gametic genomes. While there would clearly be selection pressures for autonomous and semiautonomous DNAs to work their way from somatic cell lines to gametic, for the reasons considered earlier, the mere existence of these selection pressures is not sufficient. On the other hand, there is no evidence to suggest that viral mediated interspecific gene transfer does not occur and that the moved DNAs can not become incorporated into the recipient genome. I personally would not be at all surprised, if the results of DNA sequence studies provide evidence for a significant role of infectious inheritance in evolution and adaptation of multicelled eukaryotes.

ACKNOWLEDGMENTS

I wish to thank Sandy Collins and Susan Hattingh for comments on this manuscript and members of the EcLF luncheon group for stimulating discussions. Computer time for processing these words was provided by the P. & S. Apple Foundation and incidental costs by an NIH Grant GM19848.

REFERENCES

Anderson, E. S. 1968. The ecology of transferable drug resistance in the Enteriobacteria. Ann. Rev. Microbiol. 22:131–181.

Avise, J. C. and R. A. Lansman. 1983. Polymorphisms of mitochondrial DNA in populations of higher animals. *In* M. Nei and R. K. Koehn (eds.), *Evolution of genes and proteins*, pp. 147–165. Sinauer Assoc., Sunderland, Mass.

Biel, S. W. and D. L. Hartl. 1983. Evolution of transposons: natural selection for Tn5 in *Escherichia coli* K12. Genetics 163:581–592.

Broda, P. 1969. *Plasmids* W. H. Freeman, San Francisco.

Brown, W. M. 1983. Evolution of animal mitochondrial DNA. *In* M. Nei and R. K. Koehn (eds.), *Evolution of genes and proteins*, pp. 62–88. Sinauer Assoc., Sunderland, Mass.

Campbell, A. 1981. Evolutionary significance of accessory DNA elements in bacteria. Ann. Rev. Microbiol. 35:55–83.

Campbell, A. 1983. Transposons and their evolutionary significance. *In* M. Nei and R. K. Koehn (eds.), *Evolution of genes and proteins*, pp. 258–279. Sinauer Assoc., Sunderland, Mass.

Campbell, J. H., J. A. Lengyel, and J. Langridge. 1973. Evolution of a second gene for β-galactosidase synthesis in *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. 70:1841–1875.

Cavalier-Smith, T. 1978. Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the DNA C-value paradox. J. Cell Sci. 34:247–278.

Cavalier-Smith, T. 1980. How selfish is DNA? Nature 285:617–618.

Chao, L., C. Vargas, B. B. Spear, and E. C. Cox. 1983. Transposable genetic elements as mutator genes in evolution. Nature 303:633–635.

Clarke, B. C. 1979. The evolution of genetic diversity. Proc. Roy. Soc. of London B 205:453–474.

Clarke, P. H. 1978. Experiments in micorobial evolution. *In* L. H. Ornston and J. R. Sokatch (eds.), *The bacteria*, pp. 137–218. Academic Press, New York.

Cocks, G. T. and A. C. Wilson. 1969. Immunological detection of single amino acid substitutions in alkaline phosphatase. Science 164:188–189.

Cohen S. N. 1976. Transposable genetic elements and plasmic evolution. Nature 263:731–738.

Cox, E. C. and R. M. May. 1982. Rating in biological science. Nature 302:166–167.

Dawkins, R. 1982. *The extended phenotype* Freeman, Oxford.

Dayhoff, M. O. (ed.). 1972. *Atlas of protein sequence and structure*, Vol. 5. National Biomed. Res. Foundat., Silver Springs, Md.

Dobzhansky, Th. 1973. Nothing in biology makes sense except in the light of evolution. Amer. Biol. Teacher 35:125–129.

Doolittle, W. F. and C. Sapienza. 1980. Selfish genes, the phenotype paradigm and genome evolution. Nature 284:601–603.

Dykhuizen, D. and D. Hartl. 1980. Selective neutrality of 6PGD allozymes in *E. coli* and the effects of genetic background. Genetics 96:801–817.

Falkow, S. 1975. *Infectious multiple drug resistance.* Pion, London.

Ferris, S. D., A. C. Wilson, and W. M. Brown. 1981. Evolutionary tree for apes and humans based on cleavage maps of mitochondrial DNA. Proc. Nat. Acad. Sci. U.S.A. 78:2432–2436.

Fitch, W. 1976. The evolutionary clock. *In* F. J. Ayala (ed.), *Molecular evolution*, pp. 160–178. Sinauer Assoc., Sunderland, Mass.

Gould, S. J. 1980. Is a new general theory of evolution emerging? Paleobiology 6:119–130.

Gould, S. J. and R. C. Lewontin. 1979. The spandrels

of San Marco and the Panglossian paradigm. A critique of the adaptionist programme. Proc. Roy. Soc. London B 205:581–598.

Gribbon, J. and J. Cherfas. 1982. *The monkey puzzle: Reshaping the evolutionary tree* Pantheon, New York.

Hall, B. G. 1983. Evolution of new metabolic functions in laboratory organisms. *In* M. Nei and R. K. Koehn (eds.), *Evolution of genes and proteins*, pp. 234–257. Sinauer Assoc., Sunderland, Mass.

Harris, H. 1966. Enzyme polymorphisms in man. Proc. Roy. Soc. London B 164:298–310.

Hickey, D. 1982. Selfish DNA: A sexually transmitted nuclear parasite. Genetics 101:519–531.

Horiuchi, T., J. Tomozawa, and A. Novick. 1962. Isolation and properties of bacteria capable of high rates of β-galactosidase synthesis. Biochem. Biophys. Acta 48:164–171.

Horowitz, N. H. 1965. The evolution of biochemical syntheses—retrospect and prospect. *In* V. Bryson and H. Vogel (eds.), *Evolving genes and proteins*, pp. 15–32. Academic Press, New York.

Hunkapiller, T., H. Huang, L. Hood., and J. H. Campbell. 1981. The impact of modern genetics on evolutionary theory. In Perspectives on evolution, pp. 164–189. Sinauer Assoc., Sunderland, Mass.

Jensen, R. A. 1976. Enzyme recruitment in evolution of new functions. Annu. Rev. Microbiol. 30:409–425.

Jukes, T. H. 1983. Evolution of the amino acid code *In* M. Nei and R. K. Koehn (eds.), *Evolution of genes and proteins*, pp. 191–208. Sinauer Assoc., Sunderland, Mass.

Kan, Y. W. and A. M. Dozy. 1981. Polymorphism of DNA sequence adjacent to the human globin structural gene: relationship to sickle cell mutation. Proc. Nat. Acad. Sci. U.S.A. 75:5631–5635.

Kimura, M. 1968. Evolutionary rate at the molecular level. Nature 217:624–626.

Kimura, M. 1983*a*. The neutral theory of molecular evolution. *In* M. Nei and R. K. Koehn (eds.), *Evolution of genes and proteins*, pp. 208–233. Sinauer Assoc., Sunderland, Mass.

Kimura, M. 1983*b*. *The neutral theory of molecular evolution.* Cambridge Univ. Press, Cambridge.

Kimura, M. and T. Ohta. 1971. Protein polymorphism as a phase of molecular evolution. Nature 229:467–469.

King, J. L. 1967. Continuously distributed factors affecting evolution. Genetics 49:561–576.

King, J. L. and T. H. Jukes. 1969. Non-Darwinian evolution. Science 164:788–798.

King, M.-C. and A. C. Wilson. 1975. Evolution at two levels in humans and chimpanzees. Science 188:107–116.

Kleckner, N. 1981. Transposable genetic elements in prokaryotes. Ann. Rev. Genetics 15:341–404.

Koehn, R. K., A. J. Zera, and J. G. Hall. 1983. Enzyme polymorphisms and natural selection. *In* M. Nei and R. K. Koehn (eds.), *Evolution of genes and proteins*, pp. 115–136. Sinauer Assoc., Sunderland, Mass.

Kohne, D. E. 1970. Evolution of higher organism DNA. Quart. Rev. Biophys. 3:327–375.

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

Levin, B. R. and R. E. Lenski. 1983. Coevolution in bacteria and their viruses and plasmids. *In* D. J. Futuyma and M. Slatkin (eds.), *Coevolution*, pp. 99–127. Sinauer Assoc., Sunderland, Mass.

Lewis, E. B. 1963. Genes and developmental pathways. Amer. Zool. 3:33–56.

Lewontin, R. C. 1967. An estimate of average heterozygosity in man. Amer. Jour. Hum. Genet. 19:681–685.

Lewontin, R. C. *The genetic basis of evolutionary change.* Columbia Univ. Press, New York.

Lewontin, R. C. and J. L. Hubby. 1966. A molecular approach to the study of genic heterozygosity in natural populations: Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura.* Genetics 54:595–609.

Li, W.-H. 1983. Evolution of duplicate genes and pseudogenes. *In* M. Nei and R. K. Koehn (eds.), *Evolution of genes and proteins*, pp. 14–37. Sinauer Assoc., Sunderland, Mass.

Luria S. E. and M. Delbruck. 1943. Mutations of bacteria from virus sensitivity to virus resistance. Genetics 28:491–511.

Lwoff, A. 1968. The prophage and I. *In* J. Cairns, G. S. Stent, and J. D. Watson (eds.), *Phage and the origins of molecular biology*, pp. 88–99. Cold Springs Harbor, New York.

Maxam, A. M. and W. Gilbert. 1977. A new method for sequencing DNA. Proc. Nat. Acad. Sci. U.S.A. 74:560–564.

Milkman, R. 1967. Heterosis as a major cause of heterozygosity in nature. Genetics 55:493–495.

Milkman, R. and I. P. Crawford. 1983. Clustered third-base substitutions among wild strains of *Escherichia coli.* Science 221:378–380.

Mortlock, R. P. 1982. Metabolic acquisition through laboratory selection. Ann. Rev. Micro. 36:259–284.

Nei, M. 1971. Interspecific gene differences and evolutionary time estimated from electrophoretic data on protein identity. Amer. Nat. 105:385–396.

Nei, M. 1975. *Molecular population genetics and evolution.* North Holland Publ., Amsterdam and New York.

Nei, M. 1983. Genetic polymorphisms and the role of mutation in evolution. *In* M. Nei and R. K. Koehn (eds.), *Evolution of genes and proteins*, pp. 165–190. Sinauer Assoc., Sunderland, Mass.

Nevo, E. 1978. Genetic variation in natural populations: patterns and theory. Theor. Pop. Biol. 13:121–177.

Nuttall, G. H. F. 1904. *Blood immunity and blood relationships.* Macmillan, New York.

Ohno, S. 1970. *Evolution by gene duplication.* Springer-Verlag, Berlin.

Orgel, L. E. and F. H. C. Crick. 1980. Selfish DNA: The ultimate parasite. Nature 284:604–607.

Platt, J. R. 1964. Strong inference. Science 164:347–353.

Reanney, D. 1976. Extrachromsomal elements as

possible agents of adaptation and development. Bacterol. Revs. 40:552–590.

Rose, M. and F. Doolittle. 1983. Parasitic DNA—the origin of species and sex. New Scientist, June 16, 1983, pp. 787–789.

Rose M. and W. F. Doolittle. 1983. Molecular biological mechanisms of speciation. Science 220: 157–161.

Selander, R. K. and B. R. Levin. 1980. Genetic variability and structure in *Escherichia coli* populations. Science 210:545–547.

Simpson, G. G., C. S. Pittenbrigh, and I. H. Tiffany. 1957. *Life: An introduction to Biology.* Harcourt Brace, New York.

Stent, G. S. 1969. *The coming of the golden age: A view of the end of progress.* The Natural History Press, Garden City.

Stent, G. S. 1978. *Paradoxes of progress.* Freeman, San Francisco.

Sved, J. A. 1968. Possible rates of gene substitution in evolution. Amer. Nat. 102:283–293.

Sved, J. A., T. E. Reed, and W. F. Bodmer. 1967. The number of balanced polymorphisms that can be maintained in a natural population. Genetics 55:469–481.

Thorpe, J. P. 1982. The molecular clock hypothesis: Biochemical evolution, genetic differentiation and systematics. Ann. Rev. Ecol. and Systm. 13:139–168.

Wilson, A. C. 1976. Evolutionary significance of gene regulation. *In* F. J. Ayala (ed.), *Molecular evolution,* pp. 225–234. Sinauer Assoc., Sunderland, Mass.

Wilson, A. C., S. S. Carlson, and T. J. White. 1977. Biochemical evolution. Ann. Rev. Biochem. 46: 573–639.

Zukerkandl, E. and L. Pauling. 1965. Evolutionary divergence and convergence in proteins. *In* V. Bryson and H. J. Vogel (eds.), *Evolving genes and proteins,* pp. 97–166. Academic Press, New York.