Annu. Rev. Ecol. Syst. 1995. 26:403–22

MOLECULAR EVIDENCE FOR NATURAL SELECTION

Martin Kreitman and Hiroshi Akashi
Department of Ecology and Evolutionary Biology, University of Chicago, 1101 East 57th Street, Chicago, Illinois 60637

KEY WORDS: molecular evolution, protein evolution, neutral theory, nearly neutral theory, population genetics

ABSTRACT

Our understanding of the causes of molecular evolution is not as certain as it was a decade ago when Kimura's neutral theory appeared to explain major features of DNA conservation and change. The last ten years have seen the development of empirical approaches and statistical tests for detecting selection in DNA and a proliferation of data that challenge our current understanding of the molecular evolutionary process. We begin this review with a discussion of protein polymorphism and divergence: two major areas of research where the strictly neutral model cannot explain general patterns in the data. We then present a survey of statistical methods for detecting positive selection, which includes tests for balancing selection, for sequence convergence, and for unusually high rates of evolution that cannot be accounted for by neutral models. Finally, we present findings of a number of groups working on within- and between-species variation in *Drosophila*: These highlight the importance of adaptive evolution, purifying selection, and recombination in understanding levels and patterns of nucleotide variation.

INTRODUCTION

The success of the neutral theory of molecular evolution in explaining many patterns of DNA and protein variation in natural populations has created a serious challenge for evolutionary biologists. Under the neutral theory, genetic drift is the predominant force governing change at the molecular level. If adaptive evolution is a regular feature of phenotypic evolution, as it certainly

must be, then the inability to find evidence for natural selection in the genetic material itself (or in the case of allozymes, its proxy) raises far-reaching concerns about our understanding of evolution at the molecular level. Darwinian evolution requires adaptive substitutions in DNA.

To detect natural selection at the DNA level, the statistical analysis of variation within or between species must overcome two problems: identifying features of the process distinguishing genetic drift from natural selection, and detecting that signal when only a subset of mutational changes are under natural selection. Despite these obstacles, the development of methodologies for sequencing DNA and the advance of predictive theories to explain molecular evolutionary data have renewed interest in the statistical analysis of sequence variation. On the theoretical front, Kimura's neutral theory of molecular evolution (52) provides quantitative predictions for levels of variation both within and between species; this theory has laid the modern foundation for most thinking about molecular evolution (see, for example, 63). More recent theoretical advances include weak selection models (reviewed in 79), the reformulation of neutral theory using coalescence theory and gene genealogies (reviewed in 22, 41, 94), and the study of neutral variation linked to sites under directional selection (15, 49) and balancing selection (42, 48, 96). In addition, Gillespie (30) has developed models of selection in fluctuating environments.

The major theories of molecular evolution—Kimura's neutral model, Ohta's slightly deleterious model, and Gillespie's balancing and episodic selection models—are each consistent with at least some aspects of allozyme and DNA data. None of these models, however, can account for all available empirical observations. An understanding of the evidence, we believe, will require a comprehensive theory that emphasizes strong and weak forces acting simultaneously under constraints of genetic linkage and population size.

NEUTRAL THEORY AND PATTERNS OF PROTEIN VARIATION AND EVOLUTION

Allozyme Variation

The modern study of variation at the level of genes began with the development of a methodology for quantifying protein variation at single gene loci, by Hubby & Lewontin (40) and Harris (37), and the subsequent discovery of large amounts of polymorphism in *Drosophila pseudoobscura* (61, 84). The dilemma presented by the findings was quickly established. By 1970, Dobzhansky was satisfied that the large amounts of allozyme polymorphism in *Drosophila* and humans "are clearly in accord with the balance rather than the classical model of genetic population structure" (17, p. 224). However, recognizing arguments put forth by Kimura, Nei, and others, he went on to state

that "The maintenance of abundant polymorphism and heterozygosity in populations demands, however, an explanation. ... The easiest way to cut the Gordian knot is, of course, to assume that a great majority of the polymorphisms observed involve gene variants that are selectively neutral, that is, have no appreciable effects on the fitness of their carriers."

A combination of biochemical studies and natural history data supports a role for natural selection in the maintenance of a number of enzyme polymorphisms (reviewed in 30, 95). Although case studies provide fascinating examples of biochemical adaptation, evidence for the selective maintenance of a few well-characterized enzyme polymorphisms does not help settle the more general issue of the relative contributions of selection and drift in determining observed levels of allozyme variation.

For allozymes, the lack of concrete statistical evidence for positive selection (be it balancing or directional) has led to the widespread belief that adaptive processes are infrequent in protein evolution. Many features of protein polymorphism and divergence are consistent with the neutral theory (52, 63, 74; see 30 for a rebuttal). Estimates of overall heterozygosity (73), the distribution of single locus heterozygosity (26, 75), variance in heterozygosity (32), number of alleles per locus (14), and correlation of single-locus heterozygosity between related species (13) can all be explained by the action of genetic drift on neutral variants. A positive correlation between the amount of allozyme polymorphism and the evolutionary rate is also predicted by neutral theory (87). However, Gillespie (31) has recently shown that such a correlation is also expected to occur at loci under balancing selection. Other authors have expressed caution about the overlap of predictions made by neutral and selection theories, especially as they relate to gene frequency data (21, 53).

ALLOZYME VS. DNA POLYMORPHISM Under the strictly neutral theory, levels of polymorphism (measured by heterozygosity) will be proportional to the product of effective population size and the neutral mutation rate (52). Assuming similar mutation rates across species and populations at equilibrium, neutral theory predicts a positive correlation between genetic variation and population size. However, allozyme heterozygosity does not differ substantially between species. Nei & Graur (76) summarize data for hundreds of organisms in which 20 or more allozyme loci were examined: On average, invertebrates have at most only twice as much polymorphism as vertebrates. Why do polymorphism levels and population size fail to show the correlation predicted by the strictly neutral theory of molecular evolution? One possibility, proposed by Nei & Graur, is the occurrence of population bottlenecks in the recent history of many species.

However, comparison of DNA and protein polymorphism suggests that population bottlenecks do not explain the similarity of allozyme heterozygosity

across species. In *Drosophila melanogaster* and *D. simulans*, the levels of allozyme variation are approximately the same (they may even be lower in *D. simulans*) (16). However, RFLP and DNA sequence studies indicate at least three- to six-fold higher levels of nucleotide polymorphism in *simulans* (reviewed in 4). If nucleotide heterozygosities are a measure of neutrally evolving mutations, and if allozyme variants are also maintained by genetic drift, then protein heterozygosity should show a similar difference between *D. melanogaster* and *D. simulans*, regardless of whether the populations are at equilibrium. The lack of such a pattern for allozyme variation cannot be explained by the strictly neutral theory. Aquadro (4) argues for weak selection against amino acid replacement mutations. Under a slightly deleterious model of protein evolution, selection will be more effective in removing weakly deleterious amino acid replacement mutations in the larger population size species.

Humans and *Drosophila*, for which there are abundant data, also show little difference in levels of allozyme polymorphism. Li & Sadler (64) addressed this issue by comparing human allozyme and DNA data. They restricted their analysis to alleles that were sequenced in the same lab, some 49 loci in total, to ensure that any differences between two sequences would have been checked. Note that a sequencing error rate of 1% may be 10 or 100 times higher than the naturally occurring polymorphic differences between two alleles, possibly as small as 0.1%–0.01%. What the study revealed was most surprising: The highest level of nucleotide polymorphism was only 0.11% for four-fold degenerate sites, some six-fold lower than estimates in *D. melanogaster* and more than ten-fold lower than in *D. simulans* and *D. pseudoobscura*.

The DNA data support a general intuition: There are (and have been) more flies than humans. But what about the roughly similar protein polymorphism levels in the two species? If the data are correct, and effective population size can be inferred from nucleotide data, the neutral model can again be rejected. Li & Sadler, like Aquadro, propose that slightly deleterious fitness effects for protein variants could account for these patterns. However, the authors do not attempt to evaluate the model quantitatively. Alternatively, pervasive balancing selection can also account for nearly equal levels of protein polymorphism (discussed below).

Protein Divergence

OVERDISPERSION OF THE MOLECULAR CLOCK The clock-like divergence of many proteins was initially considered strong support for the neutral theory of molecular evolution (52). For strictly neutral mutations (with selection coefficients of zero), the rate of divergence will equal the mutation rate to neutral alleles, independent of population size. Variability of protein divergence, however, has been and remains a vexing problem for the neutral theory. The strictly

neutral theory predicts, under a constant mutation rate to neutral alleles, that the expected variance in evolutionary rate will equal the mean rate (52). This is a simple consequence of modeling mutations as a Poisson point process. As early as 1971, Ohta and Kimura recognized that the rate constancy for amino acid replacement changes was violated for β-hemoglobin and cytochrome c (82). But the ratios, R, of the variance: mean ratio were not severely out of line with theoretical predictions ($R\beta$-globin = 2.05; $R\alpha$-globin = 1.37; $R_{\text{cytochrome c}}$ = 1.82). Langley & Fitch (56, 57), in more sophisticated analyses, found that rates of amino acid substitution for four proteins varied among mammalian lineages three-fold more than was predicted by neutral theory. Kimura admitted this apparent discrepancy but criticized detractors for not "seeing the forest for the trees" (52). Several follow-up studies by Gillespie (29, references cited therein) and the identification of an estimation bias (28) yielded a new average estimate of the overdispersion of the molecular clock, $R(t) = 7.75$, and a renewed criticism of neutrality. Takahata (92, see also 24) modeled protein evolution with a changing neutral mutation rate to allow for an increased variance in the rate of substitutions. Gillespie, however, remained adamant, "That replacement substitutions are selected seems almost inescapable" (29, but see also 31).

LINEAGE AND GENERATION-TIME EFFECTS The neutral theory predicts a constant rate of evolution equal to the mutation rate per generation. Given large differences in generation times (even within mammals) the clock-like behavior of protein divergence with absolute time is surprising (52). A generation-time effect observed in DNA evolution was therefore welcomed by neutralists. In a series of papers on DNA evolutionary rate, Li and coworkers (65–67) documented a two-fold slowdown in neutral divergence in humans compared to that of rhesus monkeys, approximately a ten-fold slowdown between humans and mice, and a two- to four-fold difference between humans and artiodactyls. The data are based on synonymous substitutions, noncoding sites (primarily introns) and pseudogenes (α and β-globin). There is also a suggestion of a rate slowdown for protein evolution in primates (humans in particular) compared to rodents (33, 35, 66, 67), but there are not yet enough data available to warrant any conclusion.

The generation-time effect seen for DNA may not hold for proteins. As presented in the now classic 1977 review of the molecular clock hypothesis, Wilson et al (97) rejected a generation-time effect in favor of a rate-constant dependency of protein substitution with absolute (geological) time. As with the allozyme and DNA heterozygosity, there may be an inconsistency between protein evolution and noncoding DNA evolution. Interestingly, Li et al (66) noted the same possible discrepancy in their DNA sequence analysis, but they offered no explanation.

Easteal & Collet (20) recently compared rates of replacement and silent substitution between rodent and primate lineages, using marsupials as an outgroup. At replacement sites, of 14 genes, 12 show greater divergence in the rodent lineage. Many of these loci show variable rates of silent evolution (some appear to be near saturation), but there is no overall lineage effect. Contrary to the findings of Li and coworkers, these results suggest a faster rate of protein evolution in rodents than in primates. Easteal & Collet invoke slightly deleterious mutations going to fixation in the presumably smaller populations of the rodent lineage to explain the higher rates of protein evolution. There is no evidence, however, to support the idea that historical population sizes have been smaller in rodents than in primates. Interestingly, weakly deleterious protein variants are also invoked (77) to explain the lack of a lineage effect in protein evolution found by Li and coworkers; the rate of protein evolution in the presumably smaller populations of primates increases by allowing more slightly deleterious mutations to go to fixation by genetic drift, thereby compensating for the slowdown caused by the intrinsically lower mutation rate per unit of absolute time. Again, models of adaptive evolution cannot be excluded as explanations for the protein data (31).

Slightly Deleterious Protein Evolution

One of the most appealing aspects of the neutral theory is its ability to make quantitative predictions both for expected levels of variation within populations and for divergence between species. Strictly neutral theory fails to explain either the lack of variation in levels of protein polymorphism in different species or the unexpectedly high levels of variation in rates of protein divergence (observed both as overdispersion of the clock and as lineage effects).

Ohta, Kimura, and others (reviewed in 52, 79) have developed the slightly deleterious model of molecular evolution to explain some of these discrepancies. The relative contributions of stochastic and deterministic forces to the evolutionary dynamics of slightly deleterious mutations depends critically on population size. This model posits the existence of a large class of protein variants with selection coefficients in the range $1/N_e$ (the reciprocal of the species effective population size). This allows nonneutral patterns to be explained by the dependence of deleterious selection on population sizes. Unfortunately, the lack of independent estimates of N_e in most natural populations allows great freedom to invoke "near neutrality" to explain many non-neutral patterns of protein evolution. More importantly, although many aspects of the evolutionary dynamics of weakly selected mutations have been investigated theoretically (79), there is little direct evidence that a proportion of amino acid mutations falls within this class of fitness effects. In *Drosophila*, where high levels of silent polymorphism suggest very large evolutionary effective population sizes, the region encompassing:

$|s| < 1/N_e$ may be smaller than $|s| < 10^{-6}$. Such a region could not even be represented by a thin line in the classic chromosome viability histograms (71, 72). We know virtually nothing about the distribution of fitness effects of new mutations around zero.

Perhaps the strongest available evidence for weak selection on protein variants comes from Hartl et al's (38) study of replacement and silent DNA polymorphism at the *Escherichia coli* 6-phosphogluconate dehydrogenase (gnd) locus. When compared to that of silent mutations, the frequency distribution of amino acid mutations is significantly skewed toward rare variants, implying the action of purifying selection on protein changes. However, maximum likelihood estimates show that the selection intensity for replacement mutations is probably not more than an order of magnitude greater than for silent changes. It would be very interesting to see whether rates of *gnd* protein evolution vary considerably between lineages with different effective population sizes.

Ohta (80) analyzed data on alcohol dehydrogenase genes in Hawaiian *Drosophila*, the *D. melanogaster* species subgroup, and the *D. obscura* species subgroup, showing that the replacement substitution rate relative to the synonymous rate is 40–50% higher in the (presumably) smaller populations of the Hawaiian species. The data are consistent with a slightly deleterious model, but adaptive alternatives such as the episodic selection model of Gillespie (30) cannot be rejected.

STATISTICAL EVIDENCE FOR ADAPTIVE PROTEIN EVOLUTION

Large-scale patterns of protein polymorphism and divergence allow us to reject the strictly neutral model of molecular evolution. Unfortunately, we cannot distinguish whether nearly neutral or adaptive models better account for these data. Locus-specific molecular evidence for adaptive protein evolution, however, has become increasingly abundant. We briefly review some of the best documented cases; all are based on recent work using DNA comparisons.

Balancing Selection

The theory that balancing selection maintains allozyme polymorphism is an obvious alternative to selection against slightly deleterious mutations (with shifting population size) to explain the lack of variation in allozyme heterozygosities among species. Under this model of selection, levels of polymorphism can be nearly independent of population size, instead being governed by environmental conditions and the rate at which balanced polymorphisms arise. Very few studies have convincingly demonstrated balancing selection in nature, and there are a number of theoretical arguments against a prominent role

for it (52, 60, but see also 88). Single-locus examples, however, exist in both humans and *Drosophila*. Recent work by Berry & Kreitman (9) provides strong evidence for balancing selection maintaining the alcohol dehydrogenase (*Adh*) gene frequency cline along the east coast of the United States in *D. melanogaster*. Of some 20 polymorphic nucleotide sites in the *Adh* locus, only the Fast-Slow amino acid replacement polymorphism and an insertion in an intron (which increases gene expression levels—58) are significantly clinal.

The hypothesis that balancing selection maintains the Fast-Slow polymorphism is also supported by an analysis of silent DNA variation surrounding the single amino acid change. Under balancing selection, alleles can continue segregating in the population longer than would be expected for neutral variants. Neutral mutations will accumulate at sites closely linked to selected alleles, leading to unusually high levels of silent variation between them (42, 48). A conservative statistical test shows an excess of silent variation around the Fast-Slow replacement polymorphism that cannot be explained by genetic drift (43). The HKA test will detect only balanced polymorphisms that have been segregating in a population for a long period of time and only where the recombination rate surrounding the selected site is sufficiently low. The test has been applied to data for several other loci in *D. melanogaster* that are thought to have polymorphism maintained by selection; only *Adh*, and possibly alpha-glycerophosphate dehydrogenase (93) provide evidence for balancing selection. Balancing selection is either rare or short-lived, or else recombination often erases its footprint. In addition, interpretation of the HKA test for signature *Drosophila* data has become complicated by the presence of codon selection (2) as well as selective sweeps and background selection (discussed in later sections).

Hudson et al (44) have recently proposed a test to detect the action of balancing selection on more recently derived alleles. In a DNA sequence study of *D. melanogaster* superoxide dismutase (*Sod*) genes, 19 alleles of the slow allozyme variant share an identical haplotype, whereas 22 alleles of the (presumably older) fast allozyme have far more sequence variation. Hudson et al show that the high frequencies (around 50% in both California and Spain) of the recently derived Sod^s variant are unlikely to have occurred solely through the action of genetic drift. The action of natural selection may be necessary to explain the rapid spread of Sod^s.

A more general claim of overdominant selection maintaining allozyme polymorphism is made by Karl & Avise (51) for the American oyster, *Crassostrea virginica*. Populations of this oyster, which are found along the East coast of the United States, including the Gulf of Mexico, are highly polymorphic for allozymes. There is little genetic differentiation among populations—not surprising for a species with planktonic larvae. However, mitochondrial DNA analysis reveals substantial differentiation between Gulf and East Coast

populations. Subsequent analyses of two random single-copy nuclear DNA polymorphisms appear to confirm differentiation between the two populations. Similar clines for a number of allozymes in the face of strong population subdivision suggests that balancing selection is maintaining the allozymes at relatively constant frequencies across the species's range. But with only two nuclear markers to confirm the mtDNA data, we cannot draw any strong conclusions about a role for selection. Indeed, divergent selection on mtDNA may provide a more parsimonious explanation for the data. Additional nuclear DNA data should be collected for this species.

Accelerated Protein Evolution

The emergence of DNA sequence data for many genes and species reveals a general principle about protein evolution: The vast majority of amino acid replacement mutations are disadvantageous and are eliminated by selection. The rate of amino acid replacement evolution can be as low as zero but is almost always less than the rate of silent evolution. Thus, rates of protein evolution can be explained by a combination of the elimination of deleterious replacement mutations by purifying selection and the fixation of selectively neutral mutations by genetic drift. Natural selection is undoubtedly vigilant in removing deleterious mutations. The question is whether all other (nondeleterious) amino acid changes observed within and between natural populations are neutral, weakly selected, or adaptive.

Gillespie (27, 31) has proposed a model of episodic selection to allow for sporadic bursts of change to account for the larger than expected rate variation in protein evolution. However, there are few, if any, specific examples of episodic selection. He cites several examples of accelerated evolution—baboon hemoglobin, visual pigments and human cytochromes—along with the caveat, "... the causes of most of the accelerations described ... are unknown."

GENE DUPLICATIONS The strongest evidence for accelerated protein evolution is that which follows gene duplication (62, 78, 81). The phenomenon was first described for hemoglobins following the split of the α and β families (34, 36). Ohta cites a number of additional examples, including hemoglobin γ (Anthropoidia) and β (goat vs. sheep), stomach lysozyme ruminants, visual pigment and adrenergic receptor (human), histocompatibility antigen (antigen recognition site, human and mouse), immunoglobulins (mouse heavy chain and rat appa chain), and protease inhibitor (inhibitory site, many species). Other examples of accelerated evolution associated with gene duplication are provided by Li (62), including somatostatin (anglerfish and catfish), cytochrome c (*Drosophila*), and growth hormone genes (human and bovine). Rate accelerations have also been convincingly demonstrated for insulin in hystricomorph rodents (11, 30). In this case, however, the acceleration is attributed not to

duplication but to a change in the active protein from a hexamer to a monomer in the guinea pig lineage.

In all the above examples, the evidence for rate acceleration is high value for protein divergence, k_a, relative to silent divergence, k_s, or relative to average k_a values in other lineages. In only one case does k_a exceed k_s (histocompatibility antigen). Given the large expected variance in evolutionary rates, the absence of adequate statistical testing, and a lack of functional information about the consequences of the substitutions, it is unlikely that this approach will be useful for distinguishing between adaptive evolution and relaxed constraints. Perhaps it is not surprising that Li's (62) conclusion about the likely cause of accelerated protein evolution following gene duplication, "... relaxation of selective constraints seems to be a more plausible explanation than advantageous mutation," contrasts sharply with Ohta's (78) conclusion about the same data: "Although it is difficult again to judge which of the two hypotheses is correct, it is likely that natural selection favored those individuals that possessed desired mutations in the duplicated gene copies."

A convincing case of adaptive evolution following gene duplication is the rapidly evolving *jingwei* gene in *D. yakuba* and *D. teissieri* (68). The two species form a monophyletic clade within the *D. melanogaster* species subgroup. A gene duplication arose from the *Adh* gene by retrotransposition in the *yakuba-teissiere* clade, landing on a different chromosome. Not only does k_a exceed k_s between paralogous genes within each species as well as between the homologous *jingwei* genes between species, there is also a significantly higher ratio of amino acid replacement to synonymous changes between species than to those within species. Such comparisons, if synonymous changes are neutral, reveal the action of positive selection for amino acid changes (70, 85, discussed below).

REPLACEMENT VS. SILENT DIVERGENCE As discussed above, some proteins show faster rates of replacement than of synonymous DNA evolution. First, and possibly most dramatic, is the rapid evolution of the antigen recognition sites (ARS) of class I and II genes of the major histocompatibility loci of humans and mice. As Hughes & Nei (46) were able to show, replacement divergence in the ARS region is higher than at silent and noncoding sites in the same species. If silent sites are evolving neutrally, as is generally believed, then the higher rate for the antigenic sites must reflect the contribution of positive natural selection. In addition, nonsynonymous substitutions resulting in side-chain charge changes occur in the binding cleft of the ARS more frequently than predicted by chance (47). These same loci, it is worth noting, offer dramatic evidence for persistent balancing selection: Two highly diverged alleles are shared in mice and rats, suggesting a most recent common ancestry of at least 13 million years ago (23).

Several points should be made about this form of evidence for positive selection. First, requiring amino acid replacement changes to be more frequent than silent changes is an extremely stringent criterion for detecting selection. Because purifying selection is the most prominent form of selection on proteins (52), the rate of amino acid replacement substitution will tend to be much lower than that of synonymous divergence. Probably only in the rarest of instances will positive selection raise the rate to a level exceeding the neutral rate; hence, many instances of adaptive evolution may be missed. Second, in species exhibiting biased codon usage, purifying selection appears to constrain synonymous divergence (reviewed in 86). The validity of the $k_a > k_s$ comparison depends critically on the assumption of neutrality at silent sites, an assumption that must be examined carefully (2, 86). Third, the elevated rate of replacement substitution will likely be restricted to a single, and possibly small, domain of the protein. In the case of MHC, circularity in defining the selected region is avoided because the antigen recognition site has been identified independently of the pattern of replacement divergence.

Elevated protein divergence compared to synonymous divergence is observed for sperm lysins in 20 California abalone species, suggesting positive selection (58, 59). Amazingly, a number of pairwise comparisons show an excess of nonsynonymous divergence in the whole protein. In addition, Lee et al (58) find little bias in codon usage in these genes, suggesting that purifying selection at silent sites does not explain the excess of replacement divergence between species. The selection pressure driving adaptive evolution of abalone sperm lysins, however, remains to be established.

LINEAGE-BASED COMPARISONS The very rapid evolution of antigenic sites in the hemaglutinins of the influenza A virus affecting humans has long been thought to be driven by selection to escape the immune response of their host. Fitch et al (25) provide statistical evidence supporting this contention. The protein can be divided into antigenic and nonantigenic sites (much about the structure and function of the protein is known, largely due to the efforts of Wiley and colleagues, cited in 25). In addition, this virus provides a unique opportunity to study molecular evolution. Because many (now extinct) strains were collected and kept in the laboratory, evolution in the surviving lineages can be contrasted with that of the extinct branches. Fitch et al show a significant excess of antigenic changes on the (surviving) trunk of the phylogenetic tree compared to the dead-end (extinct) branches. Positive selection appears to be driving the rapid evolution of the antigenic site in the influenza A virus.

Hughes (45) has tested the neutral theory prediction that the number of amino acid replacements in a given gene region should be a linear function of that in another region of the same gene. Applying this idea to members of the heat-shock protein 70 gene family, Hughes finds nonlinear relationships between three functional domains of the protein. The results are consistent with adaptive divergence among subfamily members of the gene, but the possibility of changes in functional constraint in one or more domains in some of the lineages cannot be rejected.

Convergent Evolution

Functional convergence through parallel substitutions in different evolutionary lineages demonstrates the action of Darwinian protein evolution. Although protein functional convergence can occur with or without structural or sequence convergence, rigorous evidence for adaptive protein evolution is limited to cases of convergence in both function and amino acid sequence.

To demonstrate sequence convergence, shared changes must be shown to be derived characters rather than conserved ancestral states or chance events; convergence can be discerned only in a phylogenetic context. Gut lysozyme evolution in the cow and langur monkey provides the strongest available evidence for adaptive protein convergence (89, 90). Lysozymes, expressed in macrophages, function as antibacterial enzymes. Two mammalian orders (primates and artiodactyls) have independently recruited lysozyme to digest cellulose in a fermentative foregut. The authors developed a phylogeny-based test for convergence by contrasting silent and replacement changes in four lineages: cow (artiodactyl), mouse (rodent), and rhesus and langur monkeys (primates). The test involves constructing the three possible relationships among the sequences of the lysozymes of the four species. Of 14 silent DNA changes, 13 support the well-established tree (primates together, cow and mouse together) while 6 of the 15 replacement sites support a genealogy linking the langur monkey with the cow. The distributions of silent and replacement changes on the standard tree are significantly different, suggesting that at least some of the 6 replacement changes have occurred in parallel.

There are other potential examples of convergent evolution. Yokoyama & Yokoyama (98) compared visual pigments in blind cave fish and humans. The authors argue for the independent evolution of the red visual pigments in both fish and humans from ancestral green pigments. However, the postulated convergence involves only two or three possible amino acid changes out of 15 variable residues, and it remains to be shown that the parallel changes are not merely coincidental events. Functional significance of the putative convergent changes should be investigated.

Other than the lysozymes, few claims of convergence are supported by phylogenetic analyses. An example lacking such analysis is rattlesnake cytochrome c, hypothesized to be convergent with the human homologue (3). A comparison of the amino acid sequence of rattlesnake cytochrome c to that of eight other vertebrates using a difference matrix indicates that rattlesnake cytochrome c is most similar to human cytochrome c. Human and rattlesnake

cytochrome c differ by 14 of 104 amino acids. However, rattlesnake and monitor lizard cytochrome c differ by only 16 amino acids. Without any evidence of functional convergence between human and rattlesnake cytochrome c and in the absence of a phylogenetic analysis, the similarity between these two molecules could be explained as homoplasy resulting from accelerated substitution rather than convergent evolution.

Examples of convergent evolution supported by phylogenetic tests are few in number. Unfortunately, a phylogenetic analysis does not guarantee detection of convergence because the power of the statistical analysis depends on the number of parallel substitutions. Like the comparison of rates of replacement and silent divergence, this test will fail to detect instances where the number of adaptive changes is small.

Ratios of Polymorphism and Divergence

McDonald & Kreitman (70) tested a simple prediction of the strictly neutral theory for polymorphism within species and substitutions between species. According to theory, levels of polymorphism and rates of change are positively correlated, both being governed by the neutral mutation rate (52). A region of a gene with many possible neutral mutations should be more polymorphic and should evolve faster than a similar-sized region under more severe selective constraints. The prediction was tested for amino acid replacement changes and for synonymous changes at the *Adh* locus in three species, *D. melanogaster*, *D. simulans*, and *D. yakuba*. A statistically significant excess of amino acid replacement changes between species compared to synonymous changes was observed, suggesting that a significant fraction of the amino acid replacement changes between species was driven by natural selection.

Several other proteins appear to violate the neutral theory prediction for polymorphism and divergence. In a within- and between-species comparison of the *glucose-6-phosphate dehydrogenase* locus of *D. melanogaster* and *D. simulans* (18), only two amino acid replacement polymorphisms were detected in a total of 44 *G6pd* alleles from the two species. One, like *Adh-F/S*, is likely to be a balanced polymorphism. In contrast, there are 21 replacement differences between the species. The McDonald-Kreitman test is highly significant. Karotam et al's study of esterase 6 (50) showed that, unlike *Adh* and *G6pd*, *Est 6* is highly polymorphic for amino acid replacements. Nevertheless, there is a statistically significant excess of replacement substitutions between species compared to synonymous changes.

There are two possible alternative explanations for a significant departure from the neutral prediction in the direction of "too many" amino acid replacements between species. The first is that synonymous changes, rather than being neutral, are subject to weak negative selection. This is likely to be true for *Adh*, which is highly codon-biased. Compared to the neutral case, polymorphism will be less severely reduced than will substitutions; a departure from neutrality in the direction observed is expected if synonymous changes are negatively selected (2). The second alternative, suggested by Ohta (80), is that the population sizes of the extant species have recently increased compared to their evolutionary sizes. If amino acid replacement mutations are slightly deleterious (selection coefficients of order $1/N_e$), then the increase in population size would allow selection to remove the deleterious mutations from current populations. Under this scenario, a higher level of replacement divergence would be expected compared to polymorphism.

NATURAL SELECTION, RECOMBINATION, AND GENETIC VARIATION IN DROSOPHILA

Genetic Hitchhiking and Selective Sweeps

The final kind of evidence supporting the adaptive evolution hypothesis (but not the balancing selection hypothesis) is genetic hitchhiking in *Drosophila*. First studied theoretically by Maynard Smith & Haigh (69) and by Ohta & Kimura (83), and more recently by Kaplan et al (49), hitchhiking occurs when a neutral mutation changes frequency through genetic linkage to a mutation that is selected. Of particular interest is the effect of a recent adaptive fixation on the level of neutral polymorphism in a region surrounding the beneficial mutation. Depending on a number of factors—neutral mutation rate, recombination rate, population size, strength of selection, and time since the selective substitution—a selective sweep of a favored mutation will not only homogenize the population (or species) for the favored mutation, but it will also homogenize the population for sufficiently tightly linked neutral mutations. Genetic hitchhiking can reduce variation surrounding the site under selection.

Hitchhiking events (i.e. selective substitutions) can be inferred from a relative lack of synonymous or noncoding polymorphism at a locus in a species otherwise known to have relatively high levels of silent polymorphism. In addition, it must be shown that the lack of polymorphism, even if it is noncoding, cannot be attributed to selective constraints on the sites under consideration. This is relatively easily accomplished by comparing evolutionary divergence for the affected sites or region among species; selective constraint, but not hitchhiking, will result in a relatively smaller divergence between species.

In contrast to stringent tests for adaptive evolution described in the previous sections, which can require many changes concentrated in a small region of a gene, reduced variation via hitchhiking can result from only a single selection event. Unfortunately, a loss of precision is the penalty for the higher sensitivity of this test; the selected mutation cannot be localized within the region of

reduced variation, nor can it be classified as a replacement or noncoding change.

The first evidence for a hitchhiking effect in *Drosophila* was a report of low DNA polymorphism levels in the yellow-achaete-scute region of the X chromosome in *D. melanogaster* (1, 6, 7, 19). This is especially so for *D. simulans*, where there is a dramatic lack of polymorphism. Yellow-achaete-scute is at the distal tip of the X chromosome, a region known to have severely reduced recombination, rendering its polymorphism level sensitive to selective sweeps at a distance.

Following the same reasoning, Berry et al (10) sequenced 19 *cubitus-interruptus-Dominant, ci^D*, genes in *D. melanogaster* and *D. simulans*. Only a single polymorphism was found, whereas the two species differed at 10% of sites. Clearly, selective sweeps have occurred in the relatively recent past of both species (they are too distantly related for a single sweep to have occurred in the common ancestor). *ci^D* is located on the fourth chromosome, which does not undergo recombination. These data also suggest that adaptive sweeps may be a regular feature of molecular evolution. There are only approximately 50 known complementation groups on the fourth chromosome (39), and independent sweeps have occurred in both *melanogaster* and *simulans*.

It is now becoming apparent that many if not all regions of (severely) reduced recombination in *Drosophila* exhibit reduced levels of polymorphism (5). Begun & Aquadro (8) have elevated the selective sweep process to prominence as a force in molecular evolution, suggesting the possibility of a high density of sweeps throughout the genome. Comparing recombination rates and polymorphism levels for 17 loci in *D. melanogaster*, they find a surprisingly strong correlation ($r^2 = 0.42$). Correctly noting the large expected variances for RFLP-based polymorphism estimates, the correlation suggests that many of the regions have sustained relatively recent selective sweeps. For this kind of comparison to be meaningful, the analysis must take into account variation in levels of selective constraint among loci, which can be estimated by quantifying between-species divergence levels. Unfortunately, such a correction complicates the development of an appropriate sampling theory and statistical test.

Background Selection

The recent development of an alternative model to explain the correlation between recombination rate and polymorphism level further complicates this picture. Charlesworth et al (15) show that selection against newly arising deleterious mutations, "background selection," can substantially reduce the level of linked neutral variation if deleterious mutations arise at sufficiently high rates and in tightly linked blocks. The mechanism for the reduction in polymorphism can be attributed to the reduction in the number of nondeleterious chromosomes

in the population. According to the authors' calculations, this mechanism will reduce polymorphism levels by as much as 78% at the base of chromosomes two and three. It cannot account for the severe reduction in polymorphism observed for the fourth chromosome and for the tip of the X. Thus, at least for these regions, positive selective sweeps are likely to have occurred.

The frequency distribution of segregating mutations may provide information to distinguish between selective sweeps and background selection as explanations for reduced variation. A recent adaptive fixation will cause an excess of rare variants as well as a reduction in the nucleotide heterozygosity in the affected region. The effect is equivalent to the recovery of variation following a population bottleneck (12). Background selection, however, appears to have little effect on the expected frequency spectra of mutations (15). Tajima (91) has developed a statistical test to determine if frequency spectra deviate from that expected for neutral mutations at equilibrium. Braverman et al (12) suggest that this test should have sufficient power, in available data sets, to detect the excess of rare variants predicted by selective sweeps. They conclude that other forces (in addition to selective sweeps) must be invoked to explain the lack of evidence for skewed frequency distributions in regions of reduced variation in *Drosophila*. Aquadro & Begun's hypothesis remains highly contentious—it will certainly remain a major focus of attention in *Drosophila* population genetics.

As a final comment, we note that the reduction in variation observed for all regions of reduced recombination in *D. melanogaster* violates the balanced polymorphism hypothesis for the maintenance of variation. If a single balanced polymorphism was maintained on the fourth chromosome, for example, linked neutral polymorphism would be expected to accumulate throughout the two selected chromosomes (96). Recall, this was the explanation for the two highly diverged MHC alleles shared between mouse and rat. The lack of high levels of polymorphism in regions of reduced recombination allows us to reject balancing selection as a general explanation for the maintenance of genetic variation. Balancing selection, if it occurs with appreciable frequency, may not last long enough for the accumulation of linked neutral mutations.

CONCLUSIONS

A detailed picture of DNA polymorphism in *Drosophila* is emerging. Unfortunately there is no simple explanation for the complexity of the observed patterns. Levels of variation at a locus may depend on selection at the locus, selection (both positive and negative) in the chromosomal region of the locus, and the population dynamics of the species. Although none of the "standard" models of population genetics adequately explains all the molecular data, individual features of the data can be explained in terms of simple processes such as genetic drift or background selection against deleterious mutations.

Neither the strictly neutral model nor any model of molecular evolution can account for major features of protein evolution. Although genetic drift may play a major role in DNA evolution, the new data resurrect the question of what causes protein polymorphism and divergence. Curiously, we cannot distinguish between deleterious and adaptive models for much of the data. Whether adaptive evolution at the molecular level achieves the hegemony it enjoys in phenotypic evolution is debatable, but recent evidence suggests it deserves a new level of appreciation.

ACKNOWLEDGMENTS

Special thanks to Jennifer Hess for literature research and discussion of convergent evolution. We are also grateful to Peter Andolfatto, Eli Stahl, and Ling-Wen Zeng for their help in improving this manuscript. H Akashi is a Howard Hughes Medical Institute Predoctoral Fellow.

Literature Cited

1. Aguadé M, Miyashita N, Langley CH. 1989. Reduced variation in the yellow-achaete-scute region in natural populations of *Drosophila melanogaster*. *Genetics* 122:607–15
2. Akashi H. 1995. Inferring weak selection from patterns of polymorphism and divergence at 'silent' sites in *Drosophila* DNA. *Genetics* 139:1067–76
3. Ambler RP, Daniel M. 1991. Rattlesnake cytochrome c: a reappraisal of the reported amino acid sequence. *Biochem. J.* 274:825–31
4. Aquadro CF. 1991. Molecular population genetics of *Drosophila*. In *Molecular Approaches to Fundamental and Applied Entomology*, ed. J Oakeshott, MJ Whitten, pp. 222–66. New York: Springer
5. Aquadro CF, Begun DJ. 1993. Evidence for and implications of genetic hitchhiking in the *Drosophila* genome. In *Mechanisms of Molecular Evolution*, ed. N Takahata, AG Clark, pp. 159–78. MA: Sinauer
6. Beech RN, Leigh Brown AJ. 1989. Insertion-deletion variation at the yellow-achaete-scute region in two natural populations of *Drosophila melanogaster*. *Genet. Res.* 53:7–15
7. Begun DJ, Aquadro CF. 1991. Molecular population genetics of the distal portion of the X chromosome in *Dro- sophila*: evidence for genetic hitchhiking of the yellow-achaete region. *Genetics* 129:1147–58
8. Begun DJ, Aquadro CF. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* 356: 519–20
9. Berry AJ, Kreitman M. 1993. Molecular analysis of an allozyme cline: alcohol dehydrogenase in *Drosophila melanogaster* on the East coast of North America. *Genetics* 134:869–93
10. Berry AJ, Ajioka JW, Kreitman M. 1991. Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. *Genetics* 129:1111–17
11. Blundell TL, Wood SP. 1975. Is the evolution of insulin Darwinian or due to selectively neutral mutations? *Nature* 257:197–203
12. Braverman JM, Hudson RR, Kaplan NL, Langley CH, Stephan W. 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphism. *Genetics*. 140:783–95
13. Chakraborty R, Fuerst PA, Nei M. 1978. Statistical studies on protein polymorphism in natural populations. *Genetics* 88:367–90
14. Chakraborty R, Fuerst PA, Nei M. 1980. Statistical studies on protein polymorphism in natural populations. III. Distribution of allele frequencies and the number of alleles per locus. *Genetics* 94:1039–63
15. Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* 134:1289–303
16. Choudhary M, Singh R. 1987. A comprehensive study of genetic variation in natural populations of *Drosophila melanogaster*. III. Variations in genetic structure and their causes between *Drosophila melanogaster* and its sibling species *Drosophila simulans*. *Genetics* 117:697–710
17. Dobzhansky T. 1970. *Genetics of the Evolutionary Process*. New York: Columbia Univ. Press
18. Eanes WF, Kirchner M, Yoon J. 1993. Evidence for adaptive evolution of the G6PD gene in the *Drosophila melanogaster* and *D. simulans* lineages. *Proc. Natl. Acad. Sci. USA* 90:7475–79
19. Eanes WF, Labate J, Ajioka JW. 1989. Restriction-map variation in the yellow-achaete-scute region in five populations of *Drosophila melanogaster*. *Mol. Biol. Evol.* 6:492–502
20. Easteal S, Collet C. 1994. Consistent variation in amino-acid substitution rate, despite uniformity of mutation rate: protein evolution in mammals is not neutral. *Mol. Biol. Evol.* 11:643–47
21. Ewens WJ. 1979. *Mathematical Population Genetics. Biomathematics*, Vol. 9. New York: Springer-Verlag
22. Ewens WJ. 1989. Population genetics theory—the past and the future. In *Mathematical and Statistical Problems of Evolutionary Theory*, ed. S Lessard, pp. 177–227. Dordrecht: Kluwer Acad.
23. Figueroa F, Günther E, Klein J. 1988. MHC polymorphism pre-dating speciation. *Nature* 355:265–67
24. Fitch WM. 1971. Rate of change of concomitantly variable codons. *J. Mol. Evol.* 1:84–96
25. Fitch WF, Leiter JME, Li X, Palese P. 1991. Positive Darwinian evolution in human influenza A viruses. *Proc. Natl. Acad. Sci. USA* 88:4270–74
26. Fuerst PA, Chakraborty R, Nei M. 1977. Statistical studies on protein polymorphism in natural populations. I. Distribution of single locus heterozygosity. *Genetics* 86:455–83
27. Gillespie JH. 1984. The molecular clock may be an episodic clock. *Proc. Natl. Acad. Sci. USA* 81:8009–13
28. Gillespie JH. 1986. Variability of evolutionary rates of DNA. *Genetics* 113: 1077–91
29. Gillespie JH. 1989. Lineage effects and the index of dispersion of molecular evolution. *Mol. Biol. Evol.* 6:636–47
30. Gillespie JH. 1991. *The Causes of Molecular Evolution*. New York: Oxford Univ. Press
31. Gillespie JH. 1994. Substitution processes in molecular evolution. II. Exchangeable models from population genetics. *Evolution* 48:1101–13
32. Gojobori T. 1982. Means and variances of heterozygosity and protein function. In *Molecular Evolution, Protein Polymorphism and the Neutral Theory*, ed. M Kimura, pp. 137–50. New York: Springer Verlag
33. Goodman M. 1961. The role of immunochemical differences in the phyletic development of human behavior. *Human Biol.* 33:131–62
34. Goodman M. 1976. Protein sequences in phylogeny. In *Molecular Evolution*, ed. FJ Ayala, pp. 141–59. Sunderland, MA: Sinauer
35. Goodman M, Barnabas J, Matsuda G, Moore GW. 1971. Molecular evolution in the descent of man. *Nature* 233:604–13
36. Goodman M, Czelusniak J, Koop BF, Tagle DA, Slightom JL. 1987. Globins: a case study in molecular phylogeny. *Proc. Cold Spring Harbor Symp. Quant. Biol.* 52:875–90
37. Harris H. 1966. Enzyme polymorphisms in man. *Proc. Roy. Soc. B.* 164:298–310
38. Hartl DL, Moriyama EN, Sawyer S. 1994. Selection intensity for codon bias. *Genetics* 138:227–34
39. Hochman B. 1976. The fourth chromosome of *Drosophila melanogaster*. In *The Genetics and Biology of Drosophila*, Vol. 1b, ed. M Ashburner, E Novitski, pp. 903–28. New York: Academic
40. Hubby JL, Lewontin RC. 1966. A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics* 54:577–94
41. Hudson RR. 1992. Gene genealogies and the coalescent process. In *Oxford Series in Ecology and Evolution*, vol. 7, ed. D Futuyma, J Antonovics, pp. 1–44. Oxford: Oxford Univ. Press
42. Hudson RR, Kaplan NL. 1988. The coalescent process in models with selection and recombination. *Genetics* 120:831–40
43. Hudson RR, Kreitman M, Aguadé M. 1987. A test of neutral molecular evo-

lution based on nucleotide data. *Genetics* 116:153–59
44. Hudson RR, Bailey K, Skarecky D, Kwiatowsky J, Ayala FJ. 1994. Evidence for positive selection in the superoxide dismutase (*Sod*) region of *Drosophila melanogaster*. *Genetics* 136:1329–40
45. Hughes AL. 1993. Nonlinear relationship among evolutionary rates identify regions of functional divergence in heat-shock protein 70 genes. *Mol. Biol. Evol.* 10:243–55
46. Hughes AL, Nei M. 1989. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc. Natl. Acad. Sci. USA* 86:958–62
47. Hughes AL, Ota T, Nei M. 1990. Positive Darwinian selection promotes diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. *Mol. Biol. Evol.* 7:515–24
48. Kaplan NL, Darden T, Hudson RR. 1988. The coalescent process in models with selection. *Genetics* 120:819–29
49. Kaplan NL, Hudson RR, Langley CH. 1989. The "hitchhiking effect" revisited. *Genetics* 123:887–99
50. Karotam J, Boyce TM, Oakeshott J. 1993. Nucleotide variation at the hypervariable Esterase 6 isozyme locus of *Drosophila simulans*. *Mol. Biol. Evol.* 12:113–22
51. Karl SA, Avise JC. 1992. Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. *Science* 256:100–2
52. Kimura M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge Univ. Press
53. Kingman JFC. 1980. *Mathematics of Genetic Diversity. CBMS-NSF Regional Conf. Ser. in Appl. Math.*, vol. 34. Philadelphia: Soc. Indust. Appl. Math.
54. Kreitman M, Aguadé M. 1986. Genetic uniformity in two populations of *Drosophila melanogaster* as revealed by filter hybridization of four-nucleotide-recognizing restriction enzyme digests. *Proc. Natl. Acad. Sci. USA* 83:3562–66
55. Kreitman M, Hudson RR. 1991. Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. *Genetics* 127: 565–82
56. Langley CH, Fitch WM. 1973. The constancy of evolution: a statistical analysis of the a and b hemoglobins, cytochrome c, and fibrinopeptide A. In *Genetic Structure of Populations*, ed. NE Morton, pp. 246–62. Honolulu: Univ. Hawaii Press
57. Langley CH, Fitch WM. 1974. An estimation of the constancy of the rate of molecular evolution. *J. Mol. Evol.* 3: 161–77
57a. Laurie CC, Stam LF. 1994. The effect of an intronic polymorphism on alcohol dehydrogenase expression in *Drosophila melanogaster*. *Genetics*.138:379–85
58. Lee Y-H, Ota T, Vacquier VD. 1995. Positive selection is a general phenomenon in the evolution of abalone sperm lysin. *Mol. Biol. Evol.* 12:213–38
59. Lee Y-H, Vacquier VD. 1992. The divergence of species-specific abalone sperm lysins is promoted by positive Darwinian selection. *Biol. Bull.* 182:97–104
60. Lewontin RC, Ginzburg L, Tuljapurkar S. 1978. Heterosis as an explanation for large amounts of genic polymorphism. *Genetics* 88:149–70
61. Lewontin RC, Hubby JL. 1966. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* 54:595–609
62. Li W-H. 1985. Accelerated evolution following gene duplication and its implication for the neutralist-selectionist controversy. In *Population Genetics and Molecular Evolution*, ed. T Ohta, K Aoki, pp. 333–52. Tokyo: Japan Sci. Soc. Press
63. Li W-H, Graur D. 1991. *Fundamentals of Molecular Evolution*. Sunderland, MA: Sinauer
64. Li W-H, Sadler LA. 1991. Low nucleotide diversity in man. *Genetics* 129:513–23
65. Li W-H, Tanimura M. 1987. The molecular clock runs more slowly in man than in apes and monkeys. *Nature* 326: 93–96
66. Li W-H, Tanimura M, Sharp PM. 1987. An evaluation of the molecular clock hypothesis using mammalian DNA sequences. *J. Mol. Evol.* 25:330–42
67. Li W-H, Wu C-I. 1987. Rates of nucleotide substitution are evidently higher in rodents than in man. *Mol. Biol. Evol.* 4:74–77
68. Long M, Langley CH. 1993. Natural selection and the origin of *jingwei*, a processed functional gene in *Drosophila*. *Science* 260:91–95
69. Maynard-Smith J, Haigh J. 1974. The hitchhiking effect of a favorable gene. *Genet. Res.* 23:23–35
70. McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351:652–54
71. Mukai T. 1964. The genetic structure of natural populations of *Drosophila melanogaster*. I. Spontaneous mutation rate of polygenes controlling viability. *Genetics* 50:1–19
72. Mukai T, Chigusa SI, Mettler LE, Crow JF. 1972. Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. *Genetics* 72:335–55
73. Nei M. 1983. Genetic polymorphism and the role of mutation in evolution. In *Evolution of Genes and Proteins*, ed. M Nei, RK Koehn, pp. 165–90. Sunderland, MA: Sinauer
74. Nei M. 1987. *Molecular Evolutionary Genetics*. New York: Columbia Univ. Press
75. Nei M, Fuerst PA, Chakraborty R. 1976. Testing the neutral mutation hypothesis by distribution of single locus heterozygosity. *Nature* 262:491–93
76. Nei M, Graur D. 1984. Extent of protein polymorphism and the neutral mutation theory. *Evol. Biol.* 17:73–118
77. Ohta T. 1987. Very slightly deleterious mutations and the molecular clock. *J. Mol. Evol.* 26:1–6
78. Ohta T. 1991. Multigene families and the evolution of complexity. *J. Mol. Evol.* 33:34–41
79. Ohta T. 1992. The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol. Sys.* 23:263–86
80. Ohta T. 1993. Amino acid substitution at the *Adh* locus of *Drosophila* is facilitated by small population size. *Proc. Natl Acad. Sci. USA* 90:4548–51
81. Ohta T. 1994. Further examples of evolution by gene duplication revealed through DNA sequence comparisons. *Genetics* 138:1331–37
82. Ohta T, Kimura M. 1971. On the constancy of the evolutionary rate of cistrons. *J. Mol. Evol.* 1:18–25
83. Ohta T, Kimura M. 1975. The effect of a selected linked locus on heterozygosity of neutral alleles (the hitchhiking effect). *Genet. Res.* 28:307–8
84. Prakash S, Lewontin RC, Hubby JL. 1969. A molecular approach to the study of genic heterozygosity in natural populations. IV. Patterns of genic variation in central, marginal and isolated populations of *Drosophila pseudoobscura*. *Genetics* 61:841–58

85. Sawyer SA, Hartl DL. 1992. Population genetics of polymorphism and divergence. *Genetics* 132:1161–76
86. Sharp PM. 1989. Evolution at 'silent' sites in DNA. In *Evolution and Animal Breeding: Reviews in Molecular and Quantitative Approaches in Honour of Alan Robertson*, ed. WG Hill, TFC Mackay, pp. 23–32. Wallingford, UK: CAB Int.
87. Skibinski DO, Woodwark M, Ward RD. 1993. A quantitative test of the neutral theory using pooled allozyme data. *Genetics* 135:233–48
88. Spencer HG, Marks RW. 1992. The maintenance of single-locus polymorphism. IV. Models with mutation from existing alleles. *Genetics* 130:211–21
89. Stewart C-B, Wilson AC. 1987. Sequence convergence and functional adaptation of stomach lysozymes from foregut fermenters. *Cold Spring Harbor Symp. Quant. Biol.* 52:891–99
90. Swanson KW, Irwin DM, Wilson AC. 1991. Stomach lysozyme gene of the langur monkey: tests for convergence and positive selection. *J. Mol. Evol.* 33:418–25
91. Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–95
92. Takahata N. 1987. On the overdispersed molecular clock. *Genetics* 116:169–79
93. Takano TS, Kusakabe S, Mukai T. 1993. DNA polymorphism and the origin of protein polymorphism at the *Gpdh* locus of *Drosophila melanogaster*. In *Mechanisms of Molecular Evolution*, ed. N Takahata, AG Clark, pp. 179–90. Sunderland, MA: Sinauer
94. Tavaré S. 1984. Line-of-descent and genealogical processes, and their applications in population genetic models. *Theor. Pop. Biol.* 26:119–64
95. Watt WB. 1994. Allozymes in evolutionary genetics: self-imposed burden or extraordinary tool? *Genetics* 136:11–16
96. Watterson GA. 1982. Mutant substitutions at linked nucleotide sites. *Adv. Appl. Prob.* 14:206–24
97. Wilson AC, Carlson SS, White TJ. 1977. Biochemical evolution. *Annu. Rev. Biochem.* 46:573–639
98. Yokoyama R, Yokoyama S. 1990. Convergent evolution of the red- and green-like visual pigment genes in fish, *Astyanax fasciatus*, and human. *Proc. Natl. Acad. Sci USA* 87:9315–18