Gent. Res., Camb. (1993), **62**, *pp.* 77–85

Book Reviews

The Causes of Molecular Evolution. By JOHN H. GILLESPIE. Oxford University Press. 1992. 336 pages. Price £25.00. ISBN 0 19 506883 1.

Evolutionary biologists have long struggled to understand the inherited variations through which natural selection causes adaptive change, and to know how far the rates of long-term evolution can be explained by natural selection. For the polygenic traits that shape the whole organism, we are far from solving either question – witness the frustrating debates on punctuated equilibria and on the maintenance of quantitative variation. However, the discovery three decades ago of abundant protein polymorphism and of a more-or-less regular 'molecular clock' gave hope for a simple and definite answer. Kimura's neutral theory of molecular evolution indeed provides an elegant solution, albeit one that makes the bulk of molecular change irrelevant to adaptation and speciation. Those mutations that have no deleterious effects may increase by chance, and contribute to polymorphism as they drift to fixation. The neutral theory is widely accepted as an explanation of both molecular variation and evolution: its dependence on just two parameters (mutation rate and population size) makes it especially attractive. Difficulties in understanding why abundant species are not extremely variable, and why rates of protein evolution are constant per year rather than per generation, led Ohta (1976) to modify strict neutrality by postulating a range of moderately deleterious mutations. Since these are more effectively eliminated from a large population, the rate of production of *effectively* neutral mutations will be lower in abundant creatures with shorter generation time. Kimura brought together the evidence in favour of this view in his 1983 book. Remarkably, there has until now been no similarly coherent exposition of an alternative, in which selection dominates. Ten years on, Gillespie has filled the gap with this masterly work, which should stimulate a fresh debate, and may even go some way towards settling this crucial issue.

The Causes of Molecular Evolution is divided into three parts: the first three chapters describing patterns of protein and DNA evolution, the next two setting out the mathematical basis of an explanation based on fluctuating selection, and the last two chapters assessing the plausibility of neutral and selectionist theories. Gillespie begins by emphasizing evidence for 'microadaptations'. Comparisons between distantly related species are unhelpful, since it is hard to find what fraction of the many amino-acid differences alter enzyme kinetics. Careful studies of close relatives give the strongest evidence that most amino-acid substitutions are adaptive – for example, the adaptation of fish LDH to different temperatures, the increased oxygen affinity of haemoglobins in high-flying vultures, or Watt's study of PGI in *Colias* butterflies. Of course, these are isolated examples, which Gillespie regards as involving unusually strong selection. However, of the first ten enzymes found to be polymorphic in *Drosophila melanogaster*, kinetic differences were found in all six cases examined in detail; this is good evidence that enzyme alleles do generally affect fitness.

Broad surveys of polymorphism have been unhelpful, despite the great efforts expended on them. Some loci are consistently more polymorphic (and tend to evolve more rapidly): but this can be explained either by differences in the fraction of mutations that are neutral (i.e. by 'constraint'), or on the selectionist view, by different exposure to 'environmental challenge'. Perhaps the most striking pattern to emerge is that even the most abundant species have limited polymorphism. While this can be explained under neutrality by mildly deleterious alleles, past bottlenecks, or hitch-hiking, such modifications require some delicate juggling of parameters, and so make the neutral theory less appealing. Similarly, elegant statistical tests based on the distribution of allele frequencies can reject strict neutrality, but cannot distinguish its various modifications.

Rates of molecular evolution are more informative, because the neutral theory makes definite predictions which are insensitive to the various modifications; in particular, the number of substitutions should approximate a Poisson distribution, so that the variance in number of substitutions should be close to the mean. Gillespie's main contribution over the past decade has been to establish that the observed variance in rates is much greater than the neutral prediction, and that this requires that the molecular clock is 'episodic' – that is, substitutions occur in bursts. Here, he presents a new analysis, using data on 20 loci from three groups (man, rodent and artiodactyl). His

https://doi.org/10.1017/S001667230003158X Published online by Cambridge University Press

method uses an unrooted tree, thus avoiding uncertainties in the phylogeny, and separates 'lineage effects' common to all loci from residual fluctuations. He finds that for amino-acid replacements, the latter do vary about 8-fold more than expected, a result incompatible with neutrality.

Over the ten years since Kimura's book was published, a great mass of DNA sequences has accumulated. As Gillespie shows, interpretation of such data requires care: the usual assumptions that substitutions occur independently and uniformly are clearly invalid. In this context, the greatest value of sequence data is that it allows comparison between 'replacement' substitutions that change the amino-acid sequence, and non-coding DNA and 'silent' sites that do not. There are clear differences. Silent sites evolve faster in lineages with shorter generation time (rodents vs. primates, for example), while replacement sites do not. Silent sites also show much less variability in rates (though the tests here are confounded by corrections for multiple substitutions). Thus, they may be largely neutral, and so provide a control against which patterns at replacement sites may be compared. This is the basis of the tests suggested by Hudson, Kreitman & Aguadé (1987) and by McDonald & Kreitman (1991), which suggest that selection is responsible for both protein polymorphism and substitution at ADH in *Drosophila*. This approach is complicated by codon-usage bias and variation in GC content, which show that silent sites must be subject to some selection. This raises the wider issue of how far selection is concentrated on the coding regions, and on simple base-substitutions. Gene duplications must be responsible for the long-term increase in gene number, non-coding sequence must be involved in gene regulation, and transposable elements are responsible for substantial quantitative variation (Mackay & Langley, 1990). However, Gillespie concentrates on proteins, a reasonable approach since it is quite plausible that this accounts for the bulk of adaptive evolution.

Following this description of molecular evolution come two dauntingly mathematical chapters, which make up 40% of the text. These are really a book within a book, giving a detailed introduction to the analysis of stochastic processes. Most readers will be tempted to skip this section. However, the temptation should be resisted, since a rather thorough understanding of the theory is needed to judge the plausibility of Gillespie's selectionist alternative. The difficulty is that any alternative to the neutral theory must predict the general patterns without being sensitive to all the details of natural selection. Gillespie's aim is to find robust results that will be true of a wide range of models; his success depends on how restrictive are his assumptions.

Gillespie argues convincingly that it is absurd to suppose that selection coefficients can remain constant over the very long timescale of molecular evolution; his theory therefore rests on the statistics of fluctuating selection. In a single panmictic haploid population, the allele with the highest geometric mean fitness will go to fixation. However, polymorphism can be preserved by temporal fluctuations in a diploid population, and/or by spatial fluctuations in fitness (provided that the contribution of each patch to the whole population is separately regulated – 'soft selection'). Gillespie shows that when selection is weak, all these schemes reduce to one class of diffusion process, the 'c-haploid model'. This leads to some remarkably general results. For example, provided that the mean selection on each allele is much smaller than the variance in selection, the stationary distribution of allele frequencies approaches a Dirichlet – exactly the same form as under the neutral theory. Combining drift and mutation with fluctuating selection is something of a nightmare. However, Gillespie can again achieve remarkable generality by making the reasonable assumption that selection is strong relative to drift, but mutation is weak ($s \gg 1/2N \gg \mu$, where μ is the nucleotide mutation rate). In this 'SSWM' regime, rare alleles evolve slowly under mutation and drift; as they become common, selection takes over. Thus, the process can be approximated by a Markov chain describing the gain and loss of alleles from the polymorphic state. For a variety of models, this leads naturally to an 'episodic clock', in which substitutions occur in clusters.

This view is appealing – even though the mathematics are considerably more complicated than in the modified neutral theory, and the parameters still less accessible to measurement. One may wonder, however, what actually maintains variation in these models. A polymorphism can only be stable if alleles tend to increase in frequency when rare. Such stability may be due to dominance in diploid populations, or to direct frequency dependence. For both stochastic and deterministic models, temporal fluctuations maintain polymorphism if the geometric mean fitness of the heterozygote is higher than either homozygote: thus, it is the dominance relations rather than the fluctuations per se which maintain variation. In any case, allozyme variation seems not much lower in haploids or predominantly selfing species than in outbred diploids (Nevo *et al.* 1984), though the patterns are confounded with differences in population size and recombination rates. To explain polymorphism in haploids, Gillespie relies on spatial fluctuations and 'soft selection'. Here again, the conditions for polymorphism parallel those in deterministic models; one can argue that variation is maintained because the rarer genotypes exploit less crowded patches, and so gain an advantage. Thus, it is the diversity of separately regulated resources, rather than stochastic fluctuations, which promote variation.

The neutral theory is vulnerable just because it makes simple and general predictions; but conversely, alternatives based on selection find it hard to account

for patterns that are consistent across diverse taxa. It is impressive that rather general stochastic models converge on an episodic clock, and on the Dirichlet distribution of allele frequencies. However, the actual rate of substitution, and the level of variation, both depend on a complicated mix of parameters; it is hard to see why these should be so similar across such different kinds of organism. There are other cases in biology of patterns so general that our varied and arbitrary explanations seem inadequate: for example, the scaling of metabolic rate with (body size)$^{0.75}$, or the prevalence of sexual reproduction.

Overall, Gillespie presents a convincing case that for proteins at least, both variation and evolution are dominated by selection. This selection may not be very strong – the frequencies of null alleles in natural *Drosophila* suggest selection coefficients of no more than 10^{-3} (Langley *et al.* 1981) – but will nevertheless dominate random sampling drift in all but the rarest species. One of the earliest objections to this view, and one which stimulated the development of the neutral theory, was that if the effects of different loci combine multiplicatively, then the fittest genotype would produce absurdly many offspring. Gillespie argues cogently that since such ideal genotypes do not exist, we should instead consider measurable quantities such as the genetic variance in fitness; this is consistent with moderate selection at thousands of loci. This amounts to assuming a concave relation between fitness and heterozygosity, such that the ideal genotype is not excessively fit (cf. Kondrashov, 1988); it would be interesting to know whether there is any reason why gene interactions should evolve so as to reduce the genetic load in this way. In any case, there are other constraints on the amount of variation that can be maintained by selection: the selectionist view implies substantial random perturbations due to hitch-hiking, and maintenance of variation by frequency-dependent selection requires many independently regulated resources.

The Causes of Molecular Evolution should be read by all students of the evolutionary process: it provides an admirably critical summary of our current knowledge (and lack of knowledge) about how DNA and protein sequences evolve, and as a bonus, includes an excellent introduction to the mathematics needed to understand the consequences of fluctuating selection. To me, Gillespie's view that selection acts on many thousands of linked loci is attractive, if only because it makes population genetics much more interesting than the neutral alternative. I hope that it will not be too long before we know whether he is right.

References

Hudson, R. R., Kreitman, M. & Aguadé, M. (1987). A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press.

Kondrashov, A. (1988). Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**, 435–441.

Langley, C. H., Voeolker, R. A., Leigh Brown, A. J., Ohnishi, S., Dickson, B. & Montgomery, E. (1981). Null allele frequencies at allozyme loci in natural populations of *Drosophila melanogaster*. *Genetics* **99**, 151–156.

Mackay, T. F. C. & Langley, C. H. (1990). Molecular and phenotypic variation in the achate-scute region of *Drosophila melanogaster*. *Nature* **348**, 64–66.

McDonald, J. H. & Kreitman, M. (1991). Adaptive protein evolution at the *Adh* locus in Drosophila. *Nature* **351**, 652–654.

Nevo, E., Beiles, A. & Ben-Shlomo, R. (1984). The evolutionary significance of genetic diversity: ecological, demographic and life-history correlates. In *Evolutionary Dynamics of Genetic Diversity* (ed. G. S. Mani), pp. 13–213. Berlin: Springer-Verlag.

Ohta, T. (1976). Role of very slightly deleterious mutations in molecular evolution and polymorphism. *Theor. Pop. Biol.* **10**, 254–275.

NICK BARTON
Institute of Cell, Animal and Population Biology
University of Edinburgh
West Mains Road
Edinburgh EH9 3JT, Scotland

Aging, Sex and DNA Repair. By CAROL BERNSTEIN and HARNI BERNSTEIN. Academic Press. 1991. 382 pages. Price $44. ISBN 0 12 092840 4.

Genetic Effects on Aging II. Edited by D. E. HARRISON. CRC Press. 1990. 573 pages. Price £50. ISBN 0 936923 318.

Sex and death are topics of absorbing interest beyond mere academic circles. Understanding their causes is an important challenge, and the level of research activity devoted to them is attested by a recent plethora of texts on both subjects. Part of the fascination for academics comes from the very varied scientific cultural backgrounds and goals of the participants. Death is often a consequence of ageing or senescence, which brings with it a baleful catalogue of biological malfunctions responsible for the increased likelihood of mortality with age. Gerontologists in general view the process mechanistically, and want to know what goes wrong, how it goes wrong, and how to fix it. Evolutionary biologists, on the other hand, are more interested in the ultimate reasons for the presence of ageing: is it a late life cost of processes that were beneficial in youth, or is it a reflection of the failure of the weak natural selection on the late part of the life history to eliminate from populations mutations with nasty effects then? In theory either community could continue to work in blissful ignorance of the activities of the other, because gerontologists do not need to worry about population biology and the evolution of ageing can be studied, for a time at least, without any knowledge of the mechanisms at work within individuals. However, each approach can help the other to make sense of its findings. For instance, the evolutionary findings

https://doi.org/10.1017/S001667230003158X Published online by Cambridge University Press