Three into two won't go

SIR—It is becoming hard to believe that after a decade of studies on gene structure and behaviour, observations of interspecific divergences and intraspecific polymorphisms are still interpreted by the constrained assumptions of the hoary old debate between selectionists and neutralists. Greenough and Harvey's recent comment[1] on Hudson, Kreitman and Aguadé's supposed discriminatory test between neutrality and selection[2], based on sequence variability in the *Adh* gene of *Drosophila* species, is but one of several recent discussions written in seeming ignorance of the dynamics of DNA, and which inadvertently disseminates misplaced excitement.

It is highly unlikely that there exist long tracts of sequence that genuinely fall into the catergory of kinetically 'complex' single-copy DNA passively picking up base substitutions and subject only to the vagaries of neutral drift or the whims of selection. DNA is an 'active' hypervariable molecule, in a state of flux, as a consequence of several mechanisms of rearrangement that cause continual gains and losses of large and small tracts of DNA. Much of the coding and non-coding DNA is subject to one or more turnover mechanism (including gene conversion, unequal crossover, slippage and transposition). These can operate at different rates, with different biases and on different units of DNA (for some reviews see refs 3–6). In some intensely examined genes it is clear that turnover mechanisms can operate one on top of another, that they are involved with the generation and dissemination of mutations, and that they make a major contribution to observed levels of divergence and polymorphism in exons and introns (see, for example, refs 7 and 8). In all tests and discussions of molecular evolution, these internal forces of the genome have to be considered in addition to the external phenomena of drift and selection.

How does all this help us with the *Adh* data? The observations are simple to explain. Both the silent site substitutions of the exons and four kilobases of what seems to be junk DNA 5′ to the gene have evolved at comparable rates in the time separating the two sibling species *D. melanogaster* and *D. sechellia* and also the two more distantly related species *D. melanogaster* and *D. pseudoobscura*[9]. But quite unexpectedly, given the assumptions of the neutral theory, the 5′-flanking sequences have approximately fourfold less polymorphism than the gene sequences in *D. melanogaster*. Kreitman and Aguadé[10], and now Greenough and Harvey[1], are happy to conclude that selection is maintaining high levels of balanced polymorphisms in the exons.

The selection hypothesis is *ad hoc* and based on assumptions of passive DNA. Equally plausible is the *ad hoc* hypothesis that turnover mechanisms have contributed considerably to the discrepancy between interspecific divergence and intraspecific polymorphism. It is interesting that the entire 5′-flanking sequence is rich in A+T, has frequent runs of homonucleotides and many small insertions/deletions[10]. These features, all hallmarks of the activities of slippage, are common in flanking DNA in *Drosophila*[10,11] and other species[4], and generate high levels of interspecific divergence[12]. Coupling this observation to the suggestion that gene conversion is operating in a domain that covers at least the 5′-flanking region, it is possible to explain relatively reduced levels of within-species polymorphism in this region.

Conversion domains can begin and end without regard for the functional requirements of the underlying sequences, and can embrace hundreds of kilobases of DNA or involve less than 10 bases[8,13,14]. When conversion is unbiased, high levels of homogeneity ensue within a species for a given DNA region, without disturbing the rate of divergence between species. Homogeneity and conservation are not the same thing. Application of appropriate analytical methods for detecting slippage and conversion to the *Adh* gene and its flanking sequences might go some way to solving our paradox. These mechanisms could even be making a contribution to the non-random distribution of silent substitutions, clustered in the third exon[9].

Whatever the precise causes might be for such evolutionary paradoxes, it is important to uncover the potential answers emanating from the dynamics of DNA behaviour, before squeezing the data into the restricted assumptions of classical population genetics. The same holds true for concepts of evolutionary molecular clocks[15]. It will be hard to escape from under the huge mathematical superstructure traditionally used to interpret the new data but a start has to be made. It is pointless for selection and neutrality to slug it out while the evidence of a third contributing force is passing them in the night. Evolution is a complex beast and we need to grasp the horns of it firmly.

GABRIEL A. DOVER

Department of Genetics,
University of Cambridge,
Cambridge CB2 3EH, UK.

1. Greenough, J. & Harvey, P.H. *Nature* **329,** 585–586 (1987).
2. Hudson, R.R., Kreitman, M.E. & Aguadé, M. *Genetics* **116,** 153–159 (1987).
3. Jeffreys, A.J., Wilson, V. & Thein, S.L. *Nature* **314,** 67–73 (1985).
4. Tautz, D., Trick, M. & Dover, G.A. *Nature* **322,** 652–656 (1986).
5. Blaisdell, B.E. *J. molec. Evol.* **21,** 278–288 (1985).
6. Levinson, G.L. & Gutman, G.A. *Molec. Biol. Evol.* **4,** 203–221 (1987).
7. Eikebush, T.H. & Burke, W.D. *J. molec. Biol.* **190,** 357–366 (1986).
8. Erhart, M.A., Piller, K. & Weaver, S. *Genetics* **115,** 511–519 (1987).
9. Schaeffer, S.W. & Aquadro, C.F. *Genetics* **117,** 61–73 (1987).
10. Kreitman, M.E. & Aguadé, M. *Genetics* **114,** 93–110 (1986).
11. Tautz, D. & Renz, M. *Nucleic Acids Res.* **12,** 4127–4138 (1984).
12. Tautz, D., Tautz, C., Webb, D.A. & Dover, G.A. *J. molec. Biol.* **195,** 525–542 (1987).
13. Shen, S.H., Slightom, J.L. & Smithies, O. *Cell* **26,** 191–203 (1981).
14. Baltimore, D. *Cell* **24,** 592–594 (1981).
15. Dover, G.A. *J. molec. Evol.* (in the press).

GREENOUGH AND HARVEY REPLY — Dover's suggestion that the "dynamics of DNA behaviour" provide an alternative to natural selection and neutral drift in explaining the *Adh* data of Hudson *et al.*[1] is wrong. His suggestion amounts to neither an alternative explanation nor a viable one.

Hudson *et al.* found that DNA sequence data for the *Adh* locus of *Drosophila* and its 5′-flanking region are inconsistent with the neutral model of molecular evolution: either the amount of intraspecific silent polymorphism in the gene is too high relative to polymorphism in the flanking region, or its interspecific divergence is too low[1]. To account for this inconsistency, Dover suggests that slippage (a mechanism by which nucleotide sequences may gain or lose repeat units) has increased the rate of divergence in the flanking region, while conversion has maintained its low variability.

Dover's specific proposal is wrong for at least three reasons. First, divergence was determined from aligned base sequences[1], so the possibility of any effects due to slippage (which causes length mutations) was eliminated.

Second, as we reported[2], Hudson *et al.* can eliminate the entire class of explanations based on discrepancies in the divergence data: the data show equal rates of divergence in the silent sites of the *Adh* locus and the flanking region, which accords with the expectation that the two regions are equally unconstrained by selection (that is, they are equally free to change). Dover's assumption that the flanking region would have diverged less were it not for the intervention of "the internal forces of the genome" leads to the conclusion that the flanking region is more constrained by selection than the silent sites of the *Adh* locus. And yet, not only does the flanking region have no open reading frames, but insertion/deletion polymorphisms are tolerated in the region, and tests have failed to find any lethal-mutable loci or any large regions affecting *Adh* expression there[3].

Third, again as we reported[2] from Hudson *et al.*[1], the problem in the *Adh* data seems to be too much variability at the *Adh* locus rather than too little in the 5′-flanking region. The proportion of nucleotide sites which are polymorphic in the 3′-flanking region is about one-third that in the *Adh* locus. Furthermore, most of the polymorphism of the *Adh* locus is