Letter to the Editor

Distinguishing the Effects of Mutational Biases and Natural Selection on DNA Sequence Variation

Hiroshi Akashi

Section of Evolution and Ecology, University of California, Davis, California 95616

Manuscript received August 4, 1997
Accepted for publication August 15, 1997

EYRE-WALKER (1997) discusses some problems of inferring evolutionary processes when mutation rates and biases vary over time. The letter specifically addresses statistical "tests of neutrality" that compare within- and between-species DNA sequence data between functional classes of mutations. Here, I will attempt to further clarify the role of mutational biases in such tests and will propose some methods to distinguish between the action of natural selection and changes in patterns of mutation.

SAWYER *et al.* (1987) first proposed statistical comparisons of the configurations of classes of mutations (such as replacement and silent) within an aligned set of DNA sequences. The "configuration" of mutations is the proportion of nonancestral nucleotides falling into frequency classes one to n, where n represents the number of sequences examined from a given population. Mutations in frequency classes one to $n - 1$ are "polymorphic" in the population and are the product of evolutionary processes occurring since the most recent common ancestor (MRCA) of within-population variation. If these sequences are compared to a sequence from an outgroup, then mutations "fixed" in the sample, *i.e.*, in frequency class n, can also be identified. Fixations reflect evolution since the split with the outgroup but prior to the MRCA of the polymorphism. SAWYER *et al.* (1987) compared the frequency distributions of segregating polymorphism between two classes of mutations, but the method can be applied to data that capture different aspects of the evolutionary process. MCDONALD and KREITMAN (1991) examined numbers of polymorphic and fixed differences, and TEMPLETON (1996) and AKASHI (1997) examined a combination of the frequency spectrum of segregating mutations and numbers of fixed differences. Although these methods have been applied most often to comparisons between silent and replacement mutations within protein-coding regions, comparisons of configurations can be made between any classes of mutations interspersed within a genetic region.

In a given comparison of within- and between-species DNA sequence data, the expected configurations of two classes of mutations will not differ if four assumptions hold. First, the sequences must have been collected independently of variation in the two classes (*i.e.*, without knowledge of allozyme variation if replacement changes are one of the classes). Second, the classes of mutations must be interspersed randomly with respect to the evolutionary histories of segments within the region. Under this assumption, departures from stationarity, due to either linked selection or population history, will have an equivalent impact on the configurations of the two classes of mutations. Third, although mutation rates may differ *between* the classes of mutations, the ratio of per locus mutation rates between the classes of mutations must have been constant over the time period examined. Finally, the distributions of fitness effects of mutations in the two classes must be equivalent. Neutral evolution for both classes of mutations satisfies the last assumption. If the first three assumptions hold, then departures from the null can be attributed to differences in the magnitude or direction of deterministic forces (natural selection or biased gene conversion) affecting the two categories of genetic variation. Positive directional selection will skew the configuration of mutations toward a larger proportion of observed mutations at high frequencies within the population or fixed in the sample. Negative directional selection will have the opposite effect, a greater proportion of the mutations will be segregating at low frequencies.

Comparisons of within- and between-species configurations of mutations can reveal the role of natural selection in base composition evolution. At "silent" sites in a number of organisms, base composition may be maintained by a balance among mutation pressure, genetic drift, and natural selection in favor of a subset of nucleotides (SHARP and LI 1986; LI 1987; BULMER 1988, 1991). Under such a model, mutations from nonfavored to favored nucleotides, "preferred" mutations, are slightly advantageous. "Unpreferred" mutations in the opposite direction, from favored to nonfavored nucleotides, confer a fitness cost of the same magnitude. If putative favored and nonfavored nucleotides can be identified and if ancestral and derived states at variable

Author e-mail: hakashi@ucdavis.edu

nucleotide positions can be inferred, then mutation-selection-drift can be tested by comparing the configurations of preferred and unpreferred mutations within and between species (AKASHI 1995). The maintenance of base composition bias under weak directional selection predicts polymorphisms segregating at higher frequencies and lower ratios of polymorphism to divergence for preferred than for unpreferred mutations.

BALLARD and KREITMAN (1994) employed this approach and found evidence consistent with the action of selection in mtDNA base composition evolution at the cytochrome *b* locus in Drosophila. Comparisons of preferred and unpreferred mutation have also revealed patterns consistent with selection at silent sites in Drosophila nuclear genes (AKASHI 1995, 1997; AKASHI and SCHAEFFER 1997). The deviations in these studies have been in the direction predicted by selection in favor of preferred mutations, but the possibility that these patterns were caused by departures from the other assumptions of the SAWYER *et al.* (1987) approach had not been previously addressed.

EYRE-WALKER (1997) has shown that particular changes in mutation rates can give rise to patterns mimicking those expected under selection. Changes in mutational biases are most plausible when base composition has changed over the period that fixed differences have accumulated in the sequences. However, inferring selection from comparisons of the evolutionary dynamics of preferred and unpreferred mutations does not require steady-state base composition. Unequal numbers of preferred and unpreferred substitutions within a lineage could reflect changes in either *per site* mutation rates or changes in the magnitude of scaled selection coefficients, $N_e s$. The latter can cause a departure from the third assumption of the null hypothesis by affecting *per locus* mutation rates through changes in the relative numbers of mutable sites over time. However, if fixed differences are observed at a small fraction of sites, then departures from equilibrium caused by a change in $N_e s$ will have a small effect on the relative numbers of mutable sites and thus a negligible impact on the ratio of *per locus* mutation rates. In addition, departures from steady-state can occur in the direction opposite to that expected under selection, *i.e.*, a higher percentage of unpreferred fixations. In this case, the statistical test for a fitness advantage to putative favored nucleotides will be a more conservative one.

Given a departure from equivalent configurations of preferred and unpreferred mutations within and between species, how can we distinguish between the action of natural selection and the effect of changing mutational biases? EYRE-WALKER (1997) suggests three approaches. The first is to test equilibrium base composition; changes in mutation rates are a less plausible explanation if base composition is at a steady state. However, even if the numbers of preferred and unpreferred fixations are equal, a change in mutational bias close to the time of the MRCA could cause departures from the null in the direction predicted by selection. An increase in the mutation rate to disfavored nucleotides will result in an excess of rare variants within the population and a higher ratio of polymorphism to divergence for unpreferred mutations. As EYRE-WALKER has pointed out, the time interval over which such changes cause departures from the null without affecting base composition is shorter than the time interval over which mutational changes affect both substitution rates and tests of neutrality. However, without estimates for the frequency or magnitude of changes in mutational biases, we cannot determine whether a mutational explanation for a given configuration of preferred and unpreferred mutations is plausible.

Comparisons of the frequency distributions of preferred and unpreferred polymorphism suffer from the same problem. Patterns consistent with selection could result from a change in the mutational biases around the time of the MRCA. In addition, the power to detect mutation-selection-drift is generally considerably lower for frequency distribution data than for comparisons that include divergence (AKASHI 1997).

Finally, EYRE-WALKER (1997) suggests estimating the mutational changes required to explain a given departure from the null hypothesis. Again, evaluating the likelihood of a given mutational change requires knowledge of the magnitude and frequency of changes in mutational patterns over evolutionary time. Small estimates for changes in mutational parameters that explain a given sample seem plausible, but we do not know what parameter ranges should exclude a mutational explanation.

Two methods may help to distinguish between the action of natural selection and changes in mutational biases when interpreting departures from equivalent configurations of preferred and unpreferred mutations. First, base composition evolution can be compared between putatively selected sites and closely linked, neutrally evolving sites. Changes in mutational biases predict changes in base composition for both classes of changes. For example, in the *Drosophila melanogaster* lineage, the base composition of silent sites in coding regions has undergone a decline in G + C content but the base composition of introns appears to have been constant since the split between *D. melanogaster* and *D. simulans*. The genome-wide reduction in codon bias in *D. melanogaster* appears to be due to a reduction in $N_e s$ rather than changes in mutation pressure (AKASHI 1996). This approach is limited, however, in requiring a class of mutations for which we are confident that neutral evolution is an appropriate model.

Comparing the evolutionary dynamics of preferred and unpreferred mutations in independent lineages may be a more robust method for differentiating mutational biases and natural selection. Although we do not know how often, or to what extent, mutational biases

change over time, they are unlikely to change in the same direction and at similar times, in independent lineages. Changes in mutational biases are not likely to explain why preferred mutations segregate at higher frequencies than unpreferred changes in nuclear genes of both *D. simulans* and *D. pseudoobscura* (AKASHI and SCHAEFFER 1997). EYRE-WALKER's analyses suggest that either selection or changes in mutational biases could have caused the patterns observed by BALLARD and KREITMAN (1994) in *D. simulans* mtDNA. Within- and between-species sequence data from the mitochondial genomes of other Drosophila species will help determine whether natural selection contributes to mtDNA base composition evolution.

I am grateful to DAVE CUTLER and two anonymous reviewers for their comments and suggestions. H.A. is supported by a National Science Foundation/Alfred P. Sloan Foundation Postdoctoral Fellowship in Molecular Evolution.

LITERATURE CITED

AKASHI, H., 1995 Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in Drosophila DNA. Genetics **139:** 1067–1076.

AKASHI, H., 1996 Molecular evolution between *Drosophila melanogaster* and *D. simulans:* reduced codon bias, faster rates of amino acid substitution, and larger proteins in *D. melanogaster.* Genetics **144:** 1297–1307.

AKASHI, H., 1997 Codon bias evolution in Drosophila: population genetics of mutation-selection-drift. Gene (in press).

AKASHI, H., and S. W. SCHAEFFER, 1997 Natural selection and the frequency distributions of "silent" DNA polymorphism in Drosophila. Genetics **146:** 295–307.

BALLARD, J. W., and M. KREITMAN, 1994 Unraveling selection in the mitochondrial genome of Drosophila. Genetics **138:** 757–772.

BULMER, M., 1988 Are codon usage patterns in unicellular organisms determined by selection-mutation balance. J. Evol. Biol. **1:** 15–26.

BULMER, M., 1991 The selection-mutation-drift theory of synonymous codon usage. Genetics **129:** 897–907.

EYRE-WALKER, A., 1997 Differentiating between selection and mutation bias (letter). Genetics **147:** 000–000.

LI, W. H., 1987 Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. J. Mol. Evol. **24:** 337–345.

MCDONALD, J. H., and M. KREITMAN, 1991 Adaptive protein evolution at the Adh locus in Drosophila. Nature **351:** 652–654.

SAWYER, S. A., D. E. DYKHUIZEN and D. L. HARTL, 1987 Confidence interval for the number of selectively neutral amino acid polymorphisms. Proc. Natl. Acad. Sci. USA **84:** 6225–6228.

SHARP, P. M., and W. H. LI, 1986 An evolutionary perspective on synonymous codon usage in unicellular organisms. J. Mol. Evol. **24:** 28–38.

TEMPLETON, A. R., 1996 Contingency tests of neutrality using intra/interspecific gene trees: the rejection of neutrality for the evolution of the mitochondrial cytochroms oxidase II gene in the Hominoid primates. Genetics **144:** 1263–1270.

Communicating editor: A. G. CLARK