Testing the Neutral Theory of Molecular Evolution*

Understanding the McDonald-Kreitman Test

Patrick Forber

July 2005

The basic framework

McDonald and Kreitman (1991) propose a test of the neutral mutation-random drift (NM-RD) hypothesis, the central claim of the neutral theory of molecular evolution. The test involves generating predictions from the NM-RD hypothesis about patterns of molecular substitutions. Alternative selection hypotheses predict that the data will deviate from the predictions of the NM-RD hypothesis in specifiable ways. To conduct the test McDonald and Kreitman examine the evolutionary dynamics of the alcohol dehydrogenase (Adh) gene in three species of *Drosophila*. The test compares the number of DNA sequence changes between species and within species. The number of DNA differences is an indicator of the evolutionary rate of the Adh gene. Based on the test they conclude that there is strong evidence for adaptive protein evolution at particular sites in the gene.

Understanding the test requires some basic knowledge about molecular terms and the predictions of neutral theory. The two important terms

*Thanks to Michael Dietrich, Ken Reisman, and Elliott Sober for helpful comments and discussion.

are *fixed differences* and *polymorphisms*. These are determined by comparing DNA sequences made up of thousands of individual nucleotide sites. A site that is unchanged within a species but different from a related species counts as a fixed difference. These are mutations that occur in some common ancestor of the lineage such that all descendants inherit the change. A site that differs within a species counts as a polymorphism. Determining the number of fixed differences and polymorphisms requires placing each individual gene sequence onto a phylogenetic, or *coalescent*, tree. A coalescent tree charts the ancestral relationships for a set of individual gene sequences. Sequences sampled from within a species form a within-species tree. The common ancestors of each within-species tree form a between-species tree. A detected difference counts as a polymorphism or a fixed difference depending on where it occurs in the phylogenetic tree (cf. Table 1). The test uses the numbers of polymorphisms and fixed differences as indicators of evolutionary rates. By directly comparing the numbers of detected differences the test indirectly compares the rate of evolution on the between-species tree with the rate on the three within-species trees. This comparison also assumes that if mutations occur in any lineage then they will occur at different sites.[1]

TABLE 1: Detected sequence differences count as polymorphisms or fixed differences depending on where they occur on the phylogenetic tree.

Part of the tree:	Between-species branch	Within-species branch
Detected difference:	Fixed difference	Polymorphism

The NM-RD hypothesis generates predictions for the number of fixed

[1]More formally, the test assumes the *infinite-sites model*. The infinite-sites model assumes that each mutation occurs at a new site so as to exclude the problem of multiple hits. When dealing with a slice of evolutionary time and sequences of some specifiable length such that the chance of the same site randomly mutates twice in that time is negligible, the infinite-sites model is a good approximation of the evolutionary process.

differences (D) and polymorphisms (P) that occur. These predictions are based on the assumption that all changes in any given molecule are neutral; that is, they have no effect on fitness. Kimura and Ohta (1971) show that if molecular evolution occurs according to the NM-RD hypothesis then the rate of fixation will equal the rate of mutation ($k = \mu$). Mutation rate is usually understood as the rate of nucleotide mutations per site, per allele, per generation yet comparisons are made with respect to sequences. To fix this, suppose a particular sequence can undergo M possible changes. The mutation rate (μ) represents the *total* mutation rate at any given site. Since there are 4 nucleotides (adenine, thymine, cytosine, and guanine), any particular site can change in three possible ways (e.g., adenine can change into thymine, cytosine or guanine). So, assuming all changes are equally probably, the rate of change to any *one* possibility is equal one-third of the total mutation rate ($\frac{\mu}{3}$). The predicted *total number* of fixed differences between two species will thus equal the product of the number of possible changes, one-third the mutation rate, and total evolutionary branch time on the between-species tree ($D = M \cdot \frac{\mu}{3} \cdot T_b$). NM-RD also predicts that neutral mutations will take longer to fix in a population than if the change has an impact on fitness. Neutral changes fix slowly by genetic drift, whereas non-neutral changes will either increase in frequency quickly or be eliminated by selection. So, when sampling sequences from within a species, we are less likely to find an adaptive polymorphism than a neutral one. By a similar argument, the predicted *total number* of polymorphisms within a species will equal the product of the number of possible changes, one-third the mutation rate, and the total branch time on the within-species trees ($P = M \cdot \frac{\mu}{3} \cdot T_w$). Counting the number of fixed differences between species and the number of polymorphisms within species yield estimates of the between-species and the within-species rates of evolution, respectively.

Designing the test

To test the NM-RD hypothesis against alternative selection hypotheses McDonald and Kreitman examine a functional gene Adh. Because Adh codes for an important protein (alcohol dehydrogenase) we have reason to suspect that some changes in the gene will not be neutral (Kreitman, 1983). Some nucleotide changes will change the amino acid sequence of the protein. But, because the genetic code is degenerate, some changes will not. So McDonald and Kreitman classify all the possible changes as *synonymous* or *replacement*. Synonymous changes do not alter the amino acid sequence of the protein, whereas replacement changes do. Given that there are M possible nucleotide changes, they let M_r equal the number of detectable (adaptive or neutral) replacement changes and M_s equal the number of possible synonymous changes. The remainder $M - (M_r + M_s)$ possible changes are deleterious.

Why is there a remainder when all changes in the coding sequence are either replacement or synonymous? Why not let M_r equal all possible replacement changes, adaptive, neutral, and deleterious? The reason is that strongly deleterious replacement changes can never be detected in a population because selection eliminates them so quickly. Functional analysis of Adh reveals that some possible replacement changes destroy or have strong adverse effects on the protein. These changes will never show up as a fixed difference or as a polymorphism. Given that these deleterious mutations occur, McDonald and Kreitman want to test whether the replacement changes that can be detected in populations (i.e., the non-deleterious ones) are effectively neutral or adaptive.

With M_r and M_s we can now determine the predictions of the NM-RD hypothesis. Let r and s denote the evolutionary dynamics at replacement and synonymous sites respectively. The equations given in Table 2 predict the number of changes at replacement and synonymous sites. The test compares the number of fixed differences and the number of poly-

TABLE 2: The equations for determining the numbers of fixed differences and polymorphisms predicted by the neutral theory for both replacement and synonymous regions of a gene.

	Fixed differences	Polymorphisms
Replacement	$D_r = M_r \cdot \frac{\mu}{3} \cdot T_b$	$P_r = M_r \cdot \frac{\mu}{3} \cdot T_w$
Synonymous	$D_s = M_s \cdot \frac{\mu}{3} \cdot T_b$	$P_s = M_s \cdot \frac{\mu}{3} \cdot T_w$

morphisms as indicators of between-species and within-species rates of evolution (respectively). The comparison involves calculating two ratios: (1) fixed replacement differences to fixed synonymous differences ($D_r : D_s$); and (2) replacement polymorphisms to synonymous polymorphisms ($P_r : P_s$). Using ratios allows mutation rate and time variables to cancel. So both (1) and (2) estimate the ratio of $M_r : M_s$, and this ratio is an indicator of evolutionary rates at the molecular level. The comparison of the between-species ratio with the within-species ratio controls for the difference in time scales. If the NM-RD hypothesis correctly describes evolution of the Adh gene then the observed ratios for both between-species and within-species evolution should be equal. Thus ratio (1) should equal ratio (2). If selection has occurred, the ratios will not be equal. Different selection hypotheses predict that the ratios will deviate from the neutral predictions in different ways. The hypothesis of positive selection predicts more fixed replacement differences than predicted by the NM-RD hypothesis since these are the changes that may have fitness effects. The hypothesis of balancing selection makes a different prediction: a preponderance of replacement polymorphisms maintained by selection.

Applying the test

To apply the test McDonald and Kreitman sampled several sequences from each of three species of *Drosophila*. By comparing these sequences they calculated the number of fixed replacement and synonymous substitutions and the number of replacement and synonymous polymorphisms. Replacement and synonymous changes are determined by reference to the genetic code. The result of their comparisons is as follows (Table 3).

TABLE 3: The results of the McDonald-Kreitman test for the Adh gene in *Drosophila* (McDonald & Kreitman, 1991, 654).

	Fixed differences	Polymorphisms
Replacement	$D_r = 7$	$P_r = 2$
Synonymous	$D_s = 17$	$P_s = 42$

As you can see, ratio (1) is much greater than ratio (2), and the deviation fits the positive selection hypothesis.[2] The test provides strong evidence for positive selection and against neutral mutation-random drift by controlling for confounding evolutionary factors. Since both ratio estimates are obtained from the same lineages both are products of the same evolutionary processes. Positive selection occurring between species for different variants of the Adh protein can explain why we observe a relatively large number of fixed replacement differences. Positive selection can speed up the rate of evolution for sites that make a difference to the structure of the protein. Hence we would expect to see a faster rate of

[2]McDonald and Kreitman (1991, 654) apply the G-test for statistical independence to determine if the proportion of replacement changes is independent of whether the changes are fixed (occurring between species) or polymorphic (occurring within species). According to the statistical test, the data does not support this null hypothesis ($P = 0.006$)—there is a significant preponderance of fixed replacement changes.

evolution between species than within species. Because the data diverge significantly from the predictions of the NM-RD hypothesis yet fit the predictions of the positive selection hypothesis, McDonald and Kreitman argue that this provides strong evidence for adaptive protein evolution in Adh.

Another alternative hypothesis

McDonald and Kreitman admit that the discrepancy between the ratios can be explained if we appeal to a different set of evolutionary factors. If all three species experienced severe bottlenecks then we would observe a higher number of fixed replacements than predicted on the strict NM-RD hypothesis. A population experiences a bottleneck when it decreases significantly in size then rapidly expands again. The decrease in population size allows several slightly deleterious mutations to fix, yielding a higher number of replacement fixed differences (D_r). The populations would then have to quickly expand so that some of these slightly deleterious mutations would be selected against within each species in order to explain the low number of replacement polymorphisms (P_r). This appeal to major population size fluctuations and the occurrence of slightly deleterious mutations can also explain the pattern of molecular substitutions observed. Based on the fragility of the assumptions made about effective population size and the magnitude of selection coefficients McDonald and Kreitman discard this alternative. Also, there is independent evidence that the effective population size of *Drosophila* has been quite large over the relevant time span (Kreitman, 1983). This leaves positive selection as the confirmed alternative.

Why count this as a test of the NM-RD hypothesis?

McDonald and Kreitman tested whether the same evolutionary processes occurred within and between species by comparing the two ratios. But a more direct test is possible. The direct test involves generating quantitative predictions of what the expected ratio would be. A neutral model would be used to predict a specific $M_r : M_s$ ratio. The same kind of predictions could be generated by different selection models as well. Then the two ratios estimated from the data could provide direct contrastive tests of the models. So there are two possible kinds of tests: (i) testing whether the same processes occurred in two trees; and (ii) directly testing for a predicted $M_r : M_s$ ratio.

McDonald and Kreitman opt for (i) for practical reasons. Because the Adh gene codes for a functional protein there is some degree of constraint on the evolution of the gene. Some replacement changes will be deleterious and eliminated by purifying selection. Recall that M_r represents the number of detectable (non-deleterious) replacement changes. If we knew the exact number of possible deleterious changes then it would be possible to calculate a specific $M_r : M_s$ ratio based on the different hypotheses. Due to practical limitations this information is not available. But, given the comparison involves the same gene, the degree of constraint for both the between-species tree and the within-species tree is constant. So McDonald and Kreitman propose a comparison of the two estimates of the $M_r : M_s$ ratio to control for this unknown degree of constraint.

So why call this comparison of evolutionary rates between and within species a test of the NM-RD hypothesis against alternative selection hypotheses? As discussed earlier, using ratios controls for time scale differences. Comparing the ratios controls for the degree of constraint. Given this experimental design and the assumption that mutation rates are relatively constant over evolutionary time, the NM-RD hypothesis predicts that the ratios obtained from the data should be same. Different selection

hypotheses predict that the ratios will deviate in specifiable ways. Positive selection for changes in the Adh protein will yield more fixed replacement differences than expected on neutral theory. This is precisely the result McDonald and Kreitman obtain. Therefore their study tests the NM-RD model against alternative selection models, and the deviation observed provides strong evidence for positive selection.

References

Kimura, M., & Ohta, T. (1971). Protein polymorphisms as a phase of molecular evolution. *Nature*, *229*, 467–469.

Kreitman, M. (1983). Nucleotide polymorphism at the alcohol dehydrogenase locus of drosophila melanogaster. *Nature*, *304*, 412–417.

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