SHORT REVIEW

Statistical tests of selective neutrality in the age of genomics

RASMUS NIELSEN

Department of Biometrics, Cornell University, 439 Warren Hall, Ithaca, NY 14853-7801, U.S.A.

Examining genomic data for traces of selection provides a powerful tool for identifying genomic regions of functional importance. Many methods for identifying such regions have focused on conserved sites. However, positive selection may also be an indication of functional importance. This article provides a brief review of some of the statistical methods used to detect selection using DNA sequence data or other molecular data. Statistical tests based on allelic distributions or levels of variability often depend on strong assumptions regarding population demographics. In contrast, tests based on comparisons of the level of variability in nonsynonymous and synonymous sites can be constructed without demographic assumptions. Such tests appear to be useful for identifying specific regions or specific sites targeted by selection.

Keywords: Darwinian selection, neutral evolution, nonsynonymous mutations, statistical tests, synonymous mutations.

Introduction

Since Kimura (1968) and King & Jukes (1969) first suggested that most polymorphisms are selectively neutral, testing the neutral hypothesis has been one of the prime objectives of molecular population genetics. The objective of studies testing neutrality has been to make general inferences about the causes of molecular evolution. However, there has been a shift in focus in the last decade to using the neutral theory as a null model against which specific occurrences of selection can be detected. There has especially been interest in providing evidence for positive selection and selective sweeps. Positive selection occurs when a new selectively advantageous mutation is segregating in a population. This type of selection is of particular interest because it may provide evidence for adaptation at the molecular level and help elucidate genotype–phenotype relationships. Selective sweeps refer to the elimination of variation at neutral sites as a linked positively selected allele goes to fixation in a population. Much of the interest in selective sweeps is spurred by the observation that the rate of recombination is correlated with the level of polymorphism in organisms such as *Drosophila melanogaster* (e.g. Begun & Aquadro, 1992). Since the size of the region affected by a selective sweep is determined by the recombination rate, recurrent selective sweeps provide one possible explanation for this correlation.

The new availability of large genomic data sets has invigorated the field of molecular population genetics and spurred new controversies regarding the causes of molecular evolution. Large samples of Single Nucleotide Polymorphisms (SNPs), microsatellites and DNA sequence data are currently being obtained in humans and other organisms. Using these data and appropriate statistical methodologies, it is in theory possible to identify regions that have undergone selective sweeps or positive selection. By finding genomic regions in which selection has been acting, we can identify the causes for species-specific phenotypic differences. For example, we might be able to identify those parts of the genome that have been undergoing selection in the evolution of humans to their modern form. Likewise, it might be possible to identify regions currently under selection, for example, because of the presence of disease-causing mutations. Tests of neutrality provide us with a powerful tool for developing hypotheses regarding function from genomic data. An important question is therefore how to extract the information regarding natural selection from genomic data and how best to identify regions, loci or specific nucleotide sites which have been targeted by selection.

The problem of testing the neutral hypothesis from molecular data has taken up much of the theoretical literature in population genetics in the last three decades. I will here provide a brief, perhaps opinionated, review of some of this literature. Because of space limitation, the review will not be comprehensive but will focus on some of the classical examples pertinent to the analysis of genomic data and on selected recent developments. I will divide tests of neutrality into two categories: (1) tests based on the allelic distribution and/or level of variability; and (2) tests based on comparisons of divergence/variability between different classes of mutations within a locus, such as nonsynonymous and synonymous mutations. Not all tests naturally fall into one of these categories. For example, tests based on the molecular clock (e.g. Langley & Fitch, 1974) may not belong to either of these categories. However, this categorization is useful for highlighting the following point: despite the fact that much of the literature has concentrated on tests of type (1) they have had very limited success in providing unambiguous evidence for

Correspondence. E-mail: rn28@cornell.edu

selection, mostly because they rely on strong assumptions regarding the demographics of the populations. In contrast, tests of type (2) have been very successful in providing clear evidence for selection.

I will here argue that it might be difficult to construct neutrality tests applicable to genomic data based on allelic distributions alone that are robust to the demographic assumptions. In contrast, robust inferences can easily be made by comparing variability in nonsynonymous and synonymous sites or between other categories of mutations, in the same genomic region. In particular, comparisons of the rates and distributions of nonsynonymous and synonymous substitutions are useful for providing robust inferences regarding the presence of selection.

Tests based on the allelic distribution or levels of variability

One locus

One of the milestones of population genetical theory was the discovery of the Ewens sampling formula (Ewens, 1972). This formula provides an analytical expression for the sampling probability under the infinite allele model, whereby every mutation is to a new allelic type, for a sample obtained from a single population of constant size with no population structure. Using Ewens's sampling formula, one of the most famous tests of neutrality, the Ewens–Watterson test (Watterson, 1977) was developed. In this test the expected homozygosity, given the observed number of alleles, is compared to the observed homozygosity. If the difference between the observed and expected homozygosity is larger than some critical value, the neutral null hypothesis can be rejected. This test is applicable to data for which the infinite-alleles model might be reasonable, such as allozyme data.

For nucleotide data, one of the most popular tests is Tajima's D-test (Tajima, 1989). Tajima's D is the scaled difference in the estimate of $\theta = 4N_e\mu$ (N_e = effective population size, μ = mutation rate per generation) based on the number of pairwise differences and the number of segregating sites in a sample of nucleotide sequences. It is defined as

$$D = \frac{\hat{\theta}_\pi - \hat{\theta}_\omega}{S_{\hat{\theta}_\pi - \hat{\theta}_\omega}} \quad (1)$$

where $\hat{\theta}_\pi$ is an estimator of $\hat{\theta}$ based on the average number of pairwise differences, $\hat{\theta}_\omega$ is an estimator of θ based on the number of segregating sites and $S_{\hat{\Theta}_\pi}$ is an estimate of the standard error of the difference of the two estimates. If the value of D is too large or too small the neutral null hypothesis is rejected. The critical values are obtained by simulations if mutational rate variation and recombination are taken into account. There are several similar tests based on slightly different test statistics such as the tests by Fu & Li (1993), Simonsen et al. (1995) and Fay & Wu (2000). A likelihood ratio test of a similar problem was described in Galtier et al. (2000).

These tests have had great success in many applications in testing the neutral equilibrium model. However, the interpretation of significant results is not always clear. The null hypothesis is a composite hypothesis that includes assumptions regarding the demographics of the populations, such as constant population size and no population structure. There is wide awareness in the field of this fact. For example, when examining the power of the Tajima's D-test, Simonsen et al. (1995) examined its power against both demographic and selection alternatives. They found that Tajima's D had a reasonable power to detect population bottlenecks and population subdivision in addition to selective sweeps. The word 'neutrality test' has therefore to some degree become synonymous with tests of the equilibrium neutral population model. Significant deviations from the neutral equilibrium model alone do not provide evidence against selective neutrality.

Some insights into the problems associated with these tests have been gained by considering the genealogical structure of the data. For example, a complete selective sweep tends to produce genealogies similar to those generated by a severe bottleneck (Fig. 1b). In both cases, the lineages in the genealogy are forced to coalesce at the time of the selective sweep or the bottleneck. The average number of pairwise differences is decreased compared to the number of segregating sites, leading to negative values of Tajima's D. The fundamental problem is that both the demographic process and selection can have very similar effects on the genealogy. It is therefore quite difficult to distinguish these effects when a single locus is considered. For the case of weak selection, it may be even more difficult to use allelic distributions to distinguish selection from demographic processes. Neuhauser & Krone (1997) and Golding (1997) have argued that weak selection may at best have only a slight effect on the genealogy. Neutrality tests based on allelic distribution might therefore
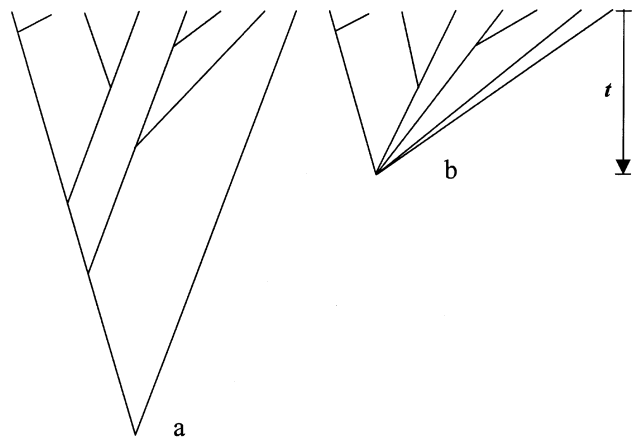
Fig. 1 Genealogies simulated under (a) the standard neutral equilibrium model and (b) a model with a severe bottleneck or a complete selective sweep t generations in the past. The effect of a severe bottleneck or a complete selective sweep is to force all lineages in the genealogy to coalesce at the time of the bottleneck/sweep.

often have much less power against the common models of selection than against demographic deviations from the neutral equilibrium model.

Multiple loci

Several statistical tests have been proposed for employing data from multiple loci. One of the most famous is the Lewontin–Krakauer test (Lewontin & Krakauer, 1973). In its original form, this test considers data at diallelic loci from multiple populations. For each locus,

$$F = \sigma_{\mathrm{p}}^2 / [\bar{p}\,(1 \quad \bar{p})] \qquad (2)$$

is calculated, where $\bar{p}$ and σ_{p}^2 are the mean and variance in allele frequency, respectively, across populations. If the variance in F is too large among loci, the neutral null model can be rejected. The problem with this test is how to determine when the variance in F is too large. In its original form, critical values were calculated assuming independence among populations, a condition that is violated by shared common ancestry or migration between populations (Robertson, 1975). The test relies on very strong, and in many cases arguably unrealistic, demographic assumptions.

The most popular test applicable to DNA sequence data obtained from multiple loci is the HKA test (Hudson *et al.*, 1987). In this test variability within and between species is compared for two or more loci. The idea is that in the absence of selection, the expected number of segregating sites within species (polymorphisms) and the expected number of fixed differences between species (divergence) are both proportional to the mutation rate, and the ratio of the two expectations should be constant among loci. Selection is inferred when the variance among loci of the ratio of divergence to polymorphism is too high. One problem that is often ignored in interpreting results of this test is that the variance in the number of segregating sites depends strongly on the demographic model. For example, we can consider the realistic case in which we have sampled DNA sequences from a population that exchanges migrants with another unobserved population. The coefficient of variation (standard deviation divided by the mean) in the number of segregating sites under this model is in Fig. 2. Notice that the coefficient of variation approaches infinity as the migration rate goes to zero. This implies, paradoxically, that as there is less and less chance of observing evidence for genetic exchange between populations, it is more and more likely that tests based on comparing levels of variability in a single population in different regions will give falsely significant results due to migration. The reason is that for low migration rates, the probability that an ancestral lineage visits the other unobserved population is very small. However, if a lineage happens to visit the other population it will tend to stay there for a very long time. The effect is a very high variance in the coalescence time among different loci.

Demographic factors affect all loci in the genome of an organism. Selection will in contrast target specific loci or nucleotide sites. Common sense would therefore dictate that
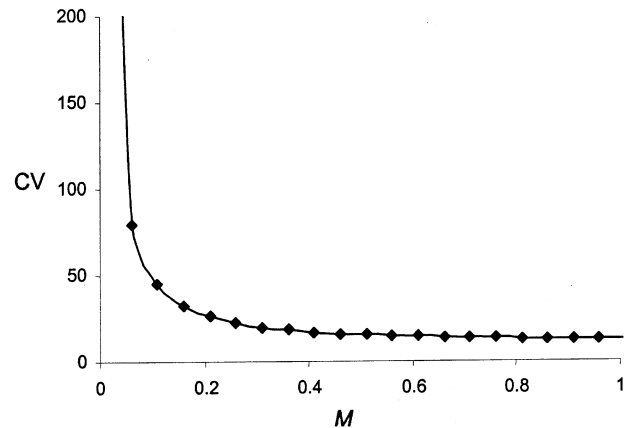
Fig. 2 The coefficient of variation (standard deviation divided by the mean) of the number of segregating sites. The coefficient of variation (CV) was evaluated by simulating samples of 25 genes from a single neutrally evolving population under the infinite sites model (Watterson, 1975). It was assumed that $\theta = 10$ (θ is four times the effective population size times the mutation rate) and that the population on average exchanges M migrants per generation symmetrically with another unobserved population of the same size. Ten thousand simulations were performed for each point in the graph. For more details on coalescence simulations and on the model see pp. 19–24 in Hudson (1990).

it is possible to detect selection by comparing multiple loci. If there is strong statistical evidence against the neutral equilibrium model for a particular locus, but the model fit the data in other loci quite well, this will usually be interpreted as evidence for selection at that locus. For example, one can imagine searching for genomic regions of low variability and/ or small values of Tajima's D as a method for identifying regions that have undergone a recent selective sweep. We readily realize that searches for regions with low levels of variability might be difficult to perform robustly, because the variance in measures of variability is strongly dependent on the demographic models (e.g. Fig. 2). Unfortunately, we face a similar problem when searching for genomic regions with extreme values of Tajima's D or other related statistics; not only the expectation but also the variance of Tajima's D depends on the demographic model. For example, we can consider the previously described demographic model, in which there is a low level of migration between the sampled population and another unobserved population (Fig. 3). In such a model the mean value of Tajima's D is approximately zero, independently of migration, but the variance in Tajima's D is increased. When the average number of migrants per generation is 0.1 it is 6–7 times as likely to observe an extreme value of $|D| > 2$ as when there is no migration. Variation in the observed value of Tajima's D or other similar summary statistics along a chromosome may therefore only in extreme cases be interpreted as evidence for selection.

As more genomic data is collected, there will be an increased demand for robust and general tests for identifying regions
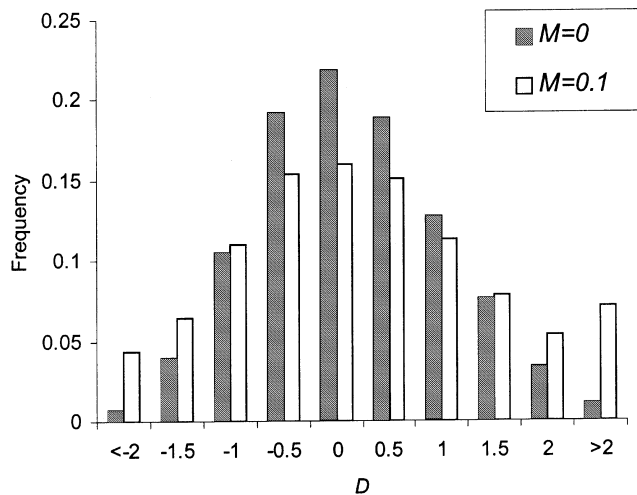
Fig. 3 The distribution of Tajima's *D* evaluated using 10 000 simulations under the same assumptions as in Fig. 2. The case of $M = 0$ corresponds to the standard neutral equilibrium model without migration.

that have experienced selection. In constructing such tests, we face the challenge that most observations based on a single summary statistic easily can be explained by demographic factors. However, it may be possible to construct more robust test by using methods that capture more of the information in the data.

Comparing variability in different classes of mutations

McDonald–Kreitman type tests

Tests based on allelic distribution or variability alone are, as just argued, quite sensitive to the underlying demographic assumptions, mostly because the structure of the gene genealogy is a product of the demographic processes in the populations. However, it is possible to establish tests of neutrality based on statistics with distributions that are independent of the genealogy or only depend on the genealogy through a nuisance parameter that can be eliminated. A famous example is the McDonald–Kreitman test (McDonald & Kreitman, 1991). In this test, the ratio of nonsynonymous to synonymous polymorphisms within species is compared to the ratio of the number of nonsynonymous and synonymous fixed differences between species in a 2×2 contingency table. The justification of this test is very similar to the HKA test. If polymorphism and divergence are driven only by mutation and genetic drift, the ratio of the number of fixations to polymorphisms should be the same for both nonsynonymous and synonymous mutations. In statistics, parameters that are of no interest to the researcher but cannot be ignored are labelled 'nuisance parameters'. A common approach is to eliminate such parameters by conditioning on a sufficient statistic, i.e. a statistic that contains all the relevant information in the data regarding the parameter. In the case of the McDonald–

Kreitman test, the total tree length is the nuisance parameter and the total number of substitutions is a sufficient statistic for this parameter. By conditioning on the total number of substitutions in the 2×2 table, the total tree length parameter is eliminated. In this manner a test of neutrality is established that is valid for any possible demographic model. The McDonald–Kreitman test has been very useful for detecting selection. For example, Eanes *et al.* (1993) found very strong evidence for selection in the *G6pd* gene in *Drosophila melanogaster* and *D. simulans.*

Although the McDonald–Kreitman test does provide unambiguous evidence for selection, it is not always clear which type of selection is acting on the gene. For example, changes in the population size combined with weak selection against slightly deleterious mutations may either increase or decrease the number of nonsynonymous polymorphisms. An increase in the population size will lead to a deficiency of nonsynonymous polymorphisms and a decrease in population size will lead to an excess of nonsynonymous polymorphisms. Significant results from the McDonald–Kreitman can not be interpreted directly as evidence for positive selection.

A related test was applied by Akashi (1994) to examine if there is selection for optimal codon usage in *Drosophila*. In the *Drosophila* genome, some codons occur at a higher frequency than others coding for the same amino acid. The common codons are usually referred to as 'preferred codons' and the rare codons are named 'unpreferred codons'. Akashi (1995) developed a test to examine if the presence of preferred codons could be attributed to selection or, alternatively, to mutational biases. He demonstrated that changes to unpreferred codons showed a significantly higher ratio of polymorphism to divergence than preferred changes in the *Drosophila simulans* lineage, providing evidence for the action of selection at silent sites.

These types of test do not rely on assumptions regarding the demographics of the populations because they are constructed by comparing different types of variability within the same locus, or genomic region. Since nonsynonymous and synonymous sites, for example, are interspersed among each other in a coding region, the effect of the demographic model is the same for both types of site.

Test based on allelic distribution in nonsynonymous and synonymous sites

Other robust tests of neutrality can be constructed by comparing the allelic distribution in different types of sites. For example, differences in the allelic distributions (frequency spectra) between synonymous and nonsynonymous polymorphisms, provide quite unambiguous evidence for selection. Such tests are particularly relevant for genomic data sets in which large numbers of polymorphisms can be obtained. Akashi (1999) suggested comparing the frequency distribution in nonsynonymous sites to the frequency distribution in synonymous sites using a test of homogeneity. If selection is of no importance, the frequency distributions of synonymous and nonsynonymous sites should be the same. For example, Cargill *et al.* (1999) and Sunyaev *et al.* (2000) demonstrated

that the overall frequency spectra in the human genome of nonsynonymous and synonymous mutations differ, providing evidence for selection on segregating mutations. Similar information was used in the test by Nielsen & Weinreich (1999) in which the ages of nonsynonymous and synonymous mutations were estimated. Differences in the average age of nonsynonymous and synonymous mutations provided evidence for selection.

Tests based on the dN/dS ratio

The most direct method for showing the presence of positive selection is to demonstrate that the number of nonsynonymous substitutions per nonsynonymous site (d_N) is significantly larger than the number of synonymous substitutions per synonymous site (d_S). For example, Hughes & Nei (1988) showed that $d_N > d_S$ in the antigen binding cleft of the Major Histocompatibility Complex. This observation provided un-ambiguous evidence for positive selection in the region, presumably overdominant or frequency dependent selection. A value of $d_N > d_S$ implies that nonsynonymous mutations are fixed with a higher probability than neutral ones due to positive selection.

A statistical framework for making inferences regarding d_N and d_S was developed by Goldman & Yang (1994) and Muse & Gaut (1994). In this framework the evolution of a nucleotide sequence is modelled as a continuous-time Markov chain with state space on the 61 possible codons in the universal genetic code. In one parameterization, the instantaneous rate matrix of the process $Q = \{q_{ij}\}$, is given by

$$
q_{ij} = \begin{cases}
0, & \text{if the two codons differ at} > 1 \text{ position,} \\
\pi_j, & \text{for synonymous transversion,} \\
\kappa\pi_j, & \text{for synonymous transition,} \\
\omega\pi_j, & \text{for nonsynonymous transversion,} \\
\omega\kappa\pi_j, & \text{for nonsynonymous transition}
\end{cases}
\tag{3}
$$

where π_j is the stationary frequency of codon j, κ is the transition/transversion rate ratio and ω ($= d_N/d_S$) is the nonsynonymous/synonymous rate ratio. Using this model, it is possible to calculate the likelihood function for ω and for other parameters using the general algorithm of Felsenstein (1981). It is thereby possible to obtain maximum likelihood estimates of these parameters, and hypotheses such as H_0: $\omega \leq 1$ can be tested using likelihood ratio tests. This maximum likelihood method has several advantages over previous methods in that it correctly accounts for the structure of the genetic code, it can incorporate complex mutational models and it is applicable directly to multiple sequences, taking the structure of the underlying genealogical tree into account.

In general, testing if $\omega \leq 1$ ($d_N < d_S$) for an entire gene is a very conservative test of neutrality. Purifying selection must occur quite frequently in functional genes to preserve function. For this reason, the average d_N is expected to be much less than the average d_S for most genes, even if positive selection is occurring in some sites quite frequently. However, when multiple divergent sequences are available it is possible to detect the presence of positively selected sites, even when most

sites are under negative selection, by allowing variation in ω among codon sites. Nielsen & Yang (1998) developed a model in which there are three categories of sites: invariable sites ($\omega = 0$), neutral sites ($\omega = 1$) and positively selected sites ($\omega > 1$). By comparing the maximum likelihood calculated under a constrained model in which the frequency of positively selected sites is set to zero (neutral model), to the maximum likelihood calculated under the general model (positive selection model), a likelihood ratio test of the hypothesis H_0: $\omega \leq 1$, for all sites, can be performed. In other words, we can test if all of the sites in the sequence have values of $\omega \leq 1$. Tests based on more realistic models for the distribution of ω were also considered in Yang et al. (2000a). These tests have reasonable power, even when the majority of sites are constrained or are evolving neutrally. In fact, it has been possible in several cases to detect selection even when the majority of sites were constrained and only a few percent of sites were evolving under positive selection (Yang et al., 2000a). The test has provided evidence for positive selection in many viral systems including HIV-1 (Nielsen & Yang, 1998; Zanotto et al., 1999), in reproductive proteins (Swanson et al., 2000), in abalone sperm lysin (Yang et al., 2000b), plant chitinases (Bishop et al., 2000) and for a variety of other genes including beta-globin (Yang et al., 2000a).

When positive selection has been detected, sites undergoing positive selection can be identified using an empirical Bayes method. Swanson et al. (2000) showed that this method correctly identifies the positively selected sites in known test cases. It is therefore in many cases possible to identify the exact location of sites targeted by selection.

It is also possible to detect selection occurring on a particular lineage of a phylogeny using similar methods. By allowing ω to vary among lineages, hypotheses such as H_0: $\omega^{(j)} \leq 1$ can be tested, where $\omega^{(j)}$ is the value of ω on a particular lineage of a phylogeny (Yang, 1998). This type of test has been used in detecting selection, for example, in the human *BRCA1* gene (Huttley et al., 2000).

The tests of neutrality based on testing H_0: $\omega \leq 1$, differ from other neutrality tests, such as the McDonald–Kreitman test, by providing direct evidence for positive selection. Detecting values of $\omega > 1$ is to date the only direct method available for detecting positive selection from DNA sequence data. However, the tests also have some limitations. In particular, they assume no recombination between the sequences and are therefore in many cases not applicable to intraspecific data. Also, the effect of a strong codon bias on these methods has not been systematically explored.

Tests of neutrality in the genomic future

We have argued that robust tests of neutrality based solely on simple summary statistics of allelic distributions and/or levels of variability are difficult to establish. The reason is that the distribution of genealogies is highly dependent on the demographics of the populations. To detect selection, more information is needed than a single summary statistic evaluated along a sequence. Tests based on comparing the pattern of synonymous and nonsynonymous mutations, in

contrast, are relatively robust because parameters relating to the genealogy can be eliminated as nuisance parameters.

Several genomic sequencing projects have been recently completed or are close to completion (e.g. humans, mouse and *Drosophila*). Assuming that the sequencing projects do not stop here, there will soon be an abundance of comparative data. Such data are perfectly suited for scanning the genome for sites at which positive selection has occurred. Several authors have argued that positive selection might be frequent in the genomes of humans and other organisms (Kreitman & Akashi, 1995; Schmid *et al.*, 1999). If this is true, we have the necessary statistical methods for identifying which sites have undergone selection based on comparative data. It will be possible to make systematic searches for genes that have undergone positive selection in the lineage leading to humans and identify the adaptive changes at the molecular level that were important in the evolution of modern humans. Identifying selection in the genome might very well become one of our most powerful tools for identifying causes for species-specific differences and for identifying genomic regions of functional, and perhaps, medical importance.

References

AKASHI, H. 1994. Synonymous codon usage in *Drosophila melanogaster*: Natural selection and translational accuracy. *Genetics*, **136**, 927–935.

AKASHI, H. 1999. Detecting the 'footprint' of natural selection in within and between species DNA sequence data. *Gene*, **238**, 39–51.

BEGUN, D. J. AND AQUADRO, C. F. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature*, **356**, 519–520.

BISHOP, J. G., DEAN, A. M. AND MITCHELL-OLDS, T. 2000. Rapid evolution in plant chitinases: Molecular targets of selection in plant–pathogen coevolution. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 5322–5327.

CARGILL, M., ALTSHULER, D., IRELAND, J., SKLAR, P. *ET AL.* 1999. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.*, **22**, 231–238.

EANES, W. F., KIRCHNER, M. AND YOON, J. 1993. Evidence for adaptive evolution of the *G6pd* gene in the *Drosophila melanogaster* and *Drosophila simulans* lineages. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 7475–7479.

EWENS, W. J. 1972. The sampling theory of selectively neutral alleles. *Theor. Pop. Biol.*, **3**, 87–112.

FAY, C. F. AND WU, C.-I. 2000. Hitchhiking under positive Darwinian selection. *Genetics*, **155**, 1405–1418.

FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.*, **17**, 368–376.

FU, Y. X. AND LI, W. H. 1993. Statistical tests of neutrality of mutations. *Genetics*, **133**, 693–709.

GALTIER, N., DEPAULIS, F. AND BARTON, N. H. 2000. Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. *Genetics*, **155**, 981–987.

GOLDING, G. B. 1997. The effect of purifying selection on genealogies. In: Donnelly, P. and Tavare, S. (eds) *Progress in Population Genetics and Human Evolution, IMA Volumes in Mathematics and its Applications*, vol. 87, pp. 271–285. Springer, New York.

GOLDMAN, N. AND YANG, Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.*, **11**, 725–736.

HUDSON, R. R. 1990. Gene genealogies and the coalescent process. In: Harvey, P. H. and Partridge, L. (eds) *Oxford Surveys in Evolutionary Biology*, vol. 7, pp. 1–44. Oxford University Press, New York.

HUDSON, R. R., KREITMAN, M. AND AGUADÉ, M. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics*, **116**, 153–159.

HUGHES, A. L. AND NEI, M. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature*, **335**, 167–170.

HUTTLEY, G. A. S., EASTEAL, M. C., SOUTHEY, A., TESORIERO GILES, G. G. *ET AL.* 2000. Adaptive evolution of the tumour suppressor *BRCA1* in humans and chimpanzees. *Nature Genet.*, **25**, 410–413.

KIMURA, M. 1968. Evolutionary rate at the molecular level. *Nature*, **217**, 624–626.

KING, J. L. AND JUKES, T. H. 1969. Non-Darwinian evolution. *Science*, **164**, 788–798.

KREITMAN, M. AND AKASHI, H. 1995. Molecular evidence for natural selection. *Ann. Rev. Ecol. Syst.*, **26**, 403–422.

LANGLEY, C. H. AND FITCH, W. M. 1974. An examination of the constancy of the rate of molecular evolution. *J. Mol. Evol.*, **3**, 161–177.

LEWONTIN, R. C. AND KRAKAUER, J. 1973. Distribution of gene frequency as a test of the theory of selective neutrality of polymorphisms. *Genetics*, **74**, 175–195.

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

MUSE, S. V. AND GAUT, B. S. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to chloroplast genome. *Mol. Biol. Evol.*, **11**, 715–724.

NEUHAUSER, C. AND KRONE, S. 1997. The genealogy of samples in models with selection. *Genetics*, **145**, 519–534.

NIELSEN, R. AND WEINREICH, D. M. 1999. The age of nonsynonymous and synonymous mutations in mtDNA and implications for the mildly deleterious theory. *Genetics*, **153**, 497–506.

NIELSEN, R. AND YANG, Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics*, **148**, 929–936.

ROBERTSON, A. 1975. Remarks on the Lewontin–Krakauer test. *Genetics*, **80**, 396.

SCHMID, K. J., NIGRO, L., AQUADRO, C. F. AND TAUTZ, D. 1999. Large number of replacement polymorphisms in rapidly evolving genes of *Drosophila*. Implications for genome-wide surveys of DNA polymorphism. *Genetics*, **153**, 1717–1729.

SIMONSEN, K. L., CHURCHILL, G. A. AND AQUADRO, C. F. 1995. Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics*, **141**, 413–429.

SUNYAEV, S. R., LATHE, W. C. 3RD, RAMENSKY, V. E. AND BORK, I. 2000. SNP frequencies in human genes an excess of rare alleles and differing modes of selection. *Trends Genet.*, **16**, 335–337.

SWANSON, W. J., YANG, Z., WOLFNER, M. F. AND AQUADRO, C. F. 2000. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 2509–2514.

TAJIMA, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

WATTERSON, G. A. 1975. On the number of segregating sites in genetical models without recombination. *Theor. Pop. Biol.*, **7**, 256–276.

WATTERSON, G. A. 1977. Heterosis or neutrality? *Genetics*, **85**, 789–814.

YANG, Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lyzosyme evolution. *Mol. Biol. Evol.*, **15**, 568–573.

YANG, Z., NIELSEN, R., GOLDMAN, N. AND PEDERSEN, A.-M. K. 2000a. Codon-substitution models for variable selection pressure at amino acid Sites. *Genetics*, **155**, 431–449.

YANG, Z., SWANSON, W. J. AND VACQUIER, V. D. 2000b. Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. *Mol. Biol. Evol.*, **17**, 1446–1455.

ZANOTTO, P. M., KALLAS, E. G., SOUZA, R. F. AND HOLMES, E. C. 1999. Genealogical evidence for positive selection in the *nef* gene of HIV-1. *Genetics*, **153**, 1077–1089.