

Statistical Tests for Detecting Positive Selection by Utilizing High-Frequency Variants

Kai Zeng,^{*,†,1} Yun-Xin Fu,^{‡,§} Suhua Shi* and Chung-I Wu[†]

^{*}State Key Laboratory of Biocontrol and Key Laboratory of Gene Engineering of the Ministry of Education, Sun Yat-sen University, Guangzhou 510275, China, [†]Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, [‡]Human Genetics Center, School of Public Health, University of Texas, Houston, Texas 77030 and [§]Laboratory for Conservation and Utilization of Bioresources, Yunnan University, Kunming 650091, China

Manuscript received May 30, 2006
Accepted for publication August 16, 2006

ABSTRACT

By comparing the low-, intermediate-, and high-frequency parts of the frequency spectrum, we gain information on the evolutionary forces that influence the pattern of polymorphism in population samples. We emphasize the high-frequency variants on which positive selection and negative (background) selection exhibit different effects. We propose a new estimator of θ (the product of effective population size and neutral mutation rate), θ_L , which is sensitive to the changes in high-frequency variants. The new θ_L allows us to revise Fay and Wu's H -test by normalization. To complement the existing statistics (the H -test and Tajima's D -test), we propose a new test, E , which relies on the difference between θ_L and Watterson's θ_W . We show that this test is most powerful in detecting the recovery phase after the loss of genetic diversity, which includes the postselective sweep phase. The sensitivities of these tests to (or robustness against) background selection and demographic changes are also considered. Overall, D and H in combination can be most effective in detecting positive selection while being insensitive to other perturbations. We thus propose a joint test, referred to as the DH test. Simulations indicate that DH is indeed sensitive primarily to directional selection and no other driving forces.

DETEECTING the footprint of positive selection is an important task in evolutionary genetic studies. Many statistical tests have been proposed for this purpose. Some of them use only divergence data between species (see NEI and KUMAR 2000 and YANG 2003 for reviews) while others use both divergence and polymorphism data (for example, HUDSON *et al.* 1987; McDONALD and KREITMAN 1991; BUSTAMANTE *et al.* 2002). Those that rely only on polymorphism data can be classified into either "haplotype" tests (for example, HUDSON *et al.* 1994; SABETI *et al.* 2002; VOIGHT *et al.* 2006) or the site-by-site "frequency spectrum" tests (for example, TAJIMA 1989; FU and LI 1993).

The frequency spectrum is the distribution of the proportion of sites where the mutant is at frequency x and is the focus of this study. We may divide the spectrum broadly into three parts: the low-, intermediate-, and high-frequency variant classes. Comparing the three parts of the spectrum, we shall have a fuller view of the configuration of polymorphisms. TAJIMA's (1989) D and FU and LI's (1993) D ask whether there are too few or many more rare variants than common ones. FAY and WU's (2000) H takes into consideration the abundance of very high-frequency variants relative to the intermediate-

frequency ones. Thus far, there is no test that addresses the relative abundance of the very high- and very low-frequency classes, although this comparison might be the most informative about new mutations. After all, new mutations are most likely to be found in the very low class and least abundant in the very high class.

In this article, we study the power to detect selection by applying spectrum tests to a linked neutral locus. In the first section, we briefly summarize some classical results of the frequency spectrum theory. In the second section, we introduce a new estimator of θ ($\theta = 4N\mu$, where N is the effective population size and μ is the neutral mutation rate of the gene). On the basis of this new estimator, we revise Fay and Wu's H -test which was not normalized and introduce a new test (E) that contrasts the high- and low-frequency variants. We further propose a joint test, DH , which is a combination of D and H . In the rest of the article, we use computer simulation to compare the powers of the tests to detect positive selection. We also consider their sensitivities to balancing selection, background selection, and demographic factors.

GENERAL BACKGROUND

Let ξ_i be the number of segregating sites where the mutant type occurs i times in the sample. Following FU (1995), we refer to the class of mutations with i

¹Corresponding author: 1101 E. 57th St., Chicago, IL 60637.
E-mail: kzeng@uchicago.edu

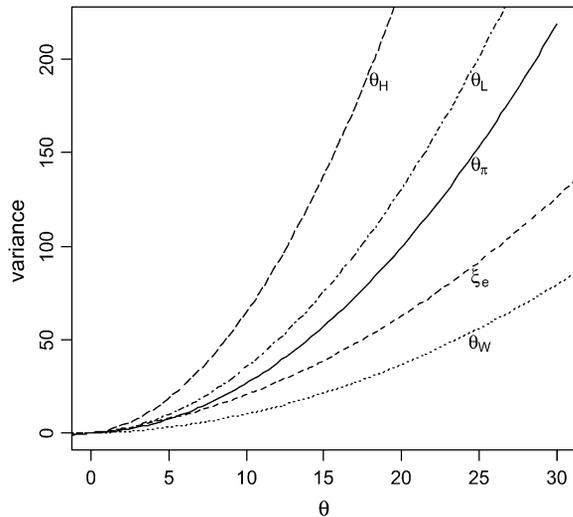


FIGURE 1.—Variance of the five estimators of θ . Sample size (n) is 50.

occurrence as mutations of size i . Given the population mutation rate θ , FU (1995) showed that

$$E(\xi_i) = \theta/i, \quad i = 1, \dots, n-1, \quad (1)$$

for a sample of size n . Since each class of mutations contains information on θ , there can be many linear functions of ξ_i that are unbiased estimators of θ , depending on how each frequency class is weighted. A general form is

$$\eta = \sum_{i=1}^{n-1} c_i \xi_i \quad (2)$$

with $E(\eta) = \theta$, where c_i 's are weight constants. A few well-known examples are WATTERSON'S (1975) θ_W , TAJIMA'S (1983) θ_π , FU and LI'S (1993) ξ_e and FAY and WU'S (2000) θ_H :

$$\begin{aligned} \theta_W &= \frac{1}{a_n} \sum_{i=1}^{n-1} \xi_i, \\ \theta_\pi &= \binom{n}{2}^{-1} \sum_{i=1}^{n-1} i(n-i) \xi_i, \\ \xi_e &= \xi_1, \\ \theta_H &= \binom{n}{2}^{-1} \sum_{i=1}^{n-1} i^2 \xi_i, \end{aligned} \quad (3)$$

where $a_n = \sum_{i=1}^{n-1} (1/i)$. Theoretical variances of these estimators are given in Figure 1. Among them, θ_W has the smallest variance.

Different estimators have varying sensitivities to changes in different parts of the frequency spectrum. For example, ξ_e and θ_W are sensitive to changes in low-frequency variants, θ_π to changes in intermediate-frequency variants, and θ_H to high-frequency ones. When the level of variation is influenced by different population genetic

forces, different parts of the spectrum are affected to different extents (SLATKIN and HUDSON 1991; FU 1996, 1997; FAY and WU 2000; GRIFFITHS 2003). The difference between two θ -estimators can thus be informative about such forces. For example, rapid population growth tends to affect low-frequency variants more than it affects high-frequency ones. As a result, θ_W tends to be larger than θ_π . The first test to take advantage of the differences between estimators is Tajima's D -test (TAJIMA 1989), as shown below:

$$D = \frac{\theta_\pi - \theta_W}{\sqrt{\text{Var}(\theta_\pi - \theta_W)}} \quad (4)$$

Others have since been proposed (FU and LI 1993; FU 1996, 1997; FAY and WU 2000). Among the evolutionary forces that may cause the frequency spectrum to deviate from the neutral equilibrium, hitchhiking (MAYNARD SMITH and HAIGH 1974) has attracted much attention. A salient feature of positive selection is the excess of high-frequency variants (FAY and WU 2000).

It should be noted that θ_W and θ_π can also be written as

$$\begin{aligned} \theta_W &= s/a_n, \\ \theta_\pi &= \binom{n}{2}^{-1} \sum_{i < j} d_{ij}, \end{aligned} \quad (5)$$

where s is the number of segregating sites and d_{ij} is the number of differences between the i th and j th sequences. Equation 5 suggests that θ_W and θ_π can be calculated without using an outgroup sequence to determine the mutant alleles. In contrast, ξ_e and θ_H do require knowledge of an outgroup sequence. As consequences, Tajima's D can be calculated without using an outgroup sequence, while the H -test needs one (see FU and LI 1993 for a version of Fu and Li's D -test that uses ξ_e but requires no outgroup sequence in the calculation).

ANALYTICAL RESULTS

Measurements of variation based on very high-frequency variants— θ_H and θ_L : To capture the dynamics of high-frequency variants, we need to put most weight on high-frequency variants in the estimation of θ . FAY and WU (2000) used θ_H (3). Its variance term, however, is not easy to obtain. Here we propose a new estimator, θ_L . The variance of θ_L can be easily obtained and then used to calculate the variance of θ_H .

Consider a genealogy of n genes from a nonrecombining region. The mean number of mutations accumulated in each gene since the most recent common ancestor (MRCA) of the sample can be calculated as

$$\theta'_L = \frac{1}{n} \sum_{i=1}^{n-1} i \xi_i. \quad (6)$$

Since $E(\xi_i) = \theta/i$, it follows that

$$E(\theta'_L) = \frac{n-1}{n}\theta. \tag{7}$$

Although θ'_L is an asymptotically unbiased estimator of θ , it is more convenient to work with its canonical form

$$\theta_L = \frac{1}{n-1} \sum_{i=1}^{n-1} i\xi_i, \tag{8}$$

which has $E(\theta_L) = \theta$. It can be shown that

$$\text{Var}(\theta_L) = \frac{n}{2(n-1)}\theta + \left[2\left(\frac{n}{n-1}\right)^2(b_{n+1}-1) - 1\right]\theta^2, \tag{9}$$

where $b_n = \sum_{i=1}^{n-1} (1/i^2)$. Since $i^2 = n \cdot i - i(n-i)$, it is easy to see that $\theta_H = 2\theta_L - \theta_\pi$. From this property, we obtain

$$\text{Var}(\theta_H) = \theta + \frac{2[36n^2(2n+1)b_{n+1} - 116n^3 + 9n^2 + 2n - 3]}{9n(n-1)^2}\theta^2 \tag{10}$$

(see APPENDIX).

The normalized Fay and Wu's H -statistic—contrasting high- and intermediate-frequency variants: Recall that FAY and WU's (2000) H is defined as $H = \theta_\pi - \theta_H = 2(\theta_\pi - \theta_L)$. Now we can write the normalized H -statistic as

$$H = \frac{\theta_\pi - \theta_L}{\sqrt{\text{Var}(\theta_\pi - \theta_L)}}, \tag{11}$$

where

$$\text{Var}(\theta_\pi - \theta_L) = \frac{n-2}{6(n-1)}\theta + \frac{18n^2(3n+2)b_{n+1} - (88n^3 + 9n^2 - 13n + 6)}{9n(n-1)^2}\theta^2 \tag{12}$$

(see APPENDIX). In practice, θ in (12) can be estimated by θ_W , and θ^2 can be estimated by $s(s-1)/(a_n^2 + b_n)$ (TAJIMA 1989).

A new E -test—contrasting high- and low-frequency variants: As mentioned before, there is no test contrasting the low- and high-frequency parts of the spectrum. Both D and H use intermediate-frequency variants as a benchmark for comparison. Nevertheless, there are occasions when it is informative to contrast high- and low-frequency variants. For example, when most variants are lost (e.g., due to selective sweep), the recovery of the neutral equilibrium is most rapid for the low-frequency ones and slowest for the high-frequency ones (e.g., Figure 2B). Taking advantage of the newly derived θ_L , we propose a new test statistic

$$E = \frac{\theta_L - \theta_W}{\sqrt{\text{Var}(\theta_L - \theta_W)}}, \tag{13}$$

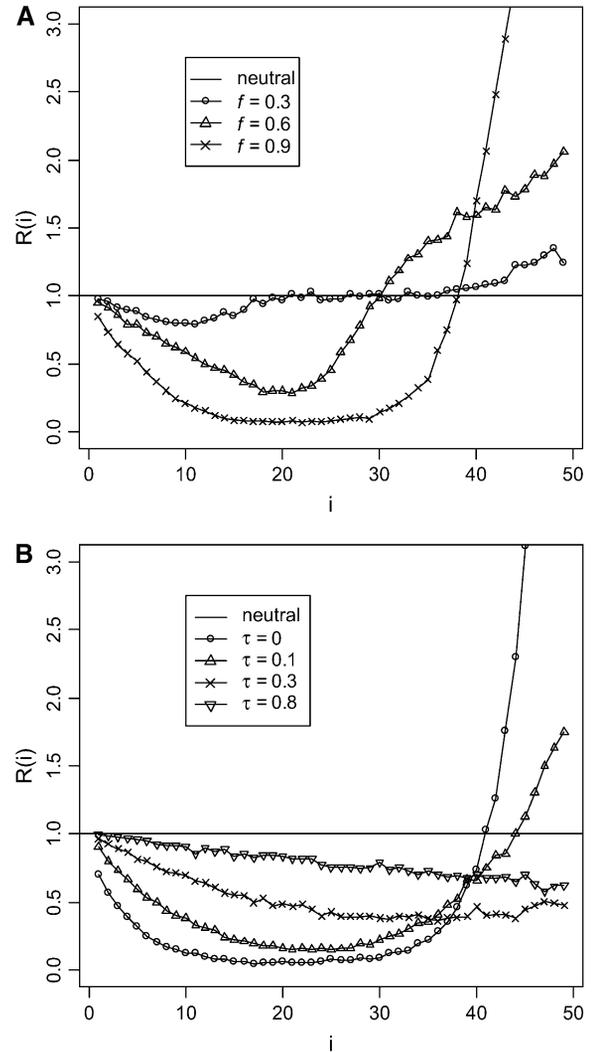


FIGURE 2.—(A) Changes in $R(i) = E(\xi'_i)/E(\xi_i)$ at the linked neutral locus as the advantageous mutation increases in frequency (f). (B) Changes in $R(i)$ at different times τ (measured in units of $4N$ generations) after fixation of the advantageous mutation. In all simulations, the parameters are defined as follows: $\theta = 4N\mu$, where μ is the mutation rate for the linked neutral locus; s is the selective coefficient of the advantageous mutation and c is the recombination distance (between the neutral variation under investigation and the advantageous mutation nearby), which is usually scaled by the selective coefficient. The parameter values are $\theta = 5$, $s = 0.001$, $c/s = 0.02$, and sample size (n) is 50. In the simulation for hitchhiking, we also incorporated intragenic recombination among the neutral variants under investigation. The intragenic recombination rate of the neutral locus, multiplied by $4N$, is 25 here and in Figure 3. The values of θ and intragenic recombination rate were chosen to reflect the reality of *D. melanogaster*; i.e., the scaled local recombination rate is about fivefold as large as the local population mutation rate. Intragenic recombination in other cases has a negligible effect on the results and was not incorporated.

TABLE 1
Critical values of the *D*- and *H*-tests (d^* and h^*) when performed jointly as the *DH* test

<i>s</i>	$\alpha = f_s(p^*) = 5\%$			$\alpha = 1\%$			$\alpha = 0.1\%$		
	d^*	h^*	p^*	d^*	h^*	p^*	d^*	h^*	p^*
10	-0.971	-0.535	0.176	-1.468	-1.721	0.053	-1.872	-3.302	0.010
20	-1.054	-0.766	0.145	-1.536	-2.022	0.039	-1.982	-3.537	0.006
30	-1.084	-0.820	0.135	-1.568	-2.061	0.034	-1.995	-3.644	0.005
40	-1.109	-0.867	0.128	-1.562	-2.071	0.033	-2.019	-3.489	0.005
50	-1.107	-0.885	0.125	-1.579	-2.133	0.030	-2.062	-3.656	0.003

Sample size is 50. *s* is the number of segregating sites in the sample. α is the significance level of the *DH* test. p^* is the solution of the equation $f_s(p) = \alpha$; $f_s(\cdot)$ is defined by (15). d^* and h^* are the critical values such that $P\{d(X_s) \leq d^*\} = p^*$ and $P\{h(X_s) \leq h^*\} = p^*$, respectively; X_s is random samples under neutrality with *s* ($s \geq 1$) segregating sites; $d(\cdot)$ is the value of the *D*-statistic of a sample and $h(\cdot)$ is that of the normalized *H*-statistic.

where

$$\begin{aligned} & \text{Var}(\theta_L - \theta_W) \\ &= \left[\frac{n}{2(n-1)} - \frac{1}{a_n} \right] \theta + \left[\frac{b_n}{a_n^2} + 2 \left(\frac{n}{n-1} \right)^2 b_n \right. \\ & \quad \left. - \frac{2(nb_n - n + 1)}{(n-1)a_n} - \frac{3n+1}{n-1} \right] \theta^2 \end{aligned} \tag{14}$$

(see APPENDIX). In practice, θ and θ^2 in (14) can be estimated by methods described after (12). Thus, the three tests, *D*, *H*, and *E*, contrast the three parts of the frequency spectrum, low-, intermediate-, and high-frequency variants, in a pairwise manner. By properly choosing a test statistic, we can have a better chance to detect a particular driving force (*e.g.*, Table 3).

The *DH* test—a joint *D* and *H* test: Simulation results (*e.g.*, Table 3) suggest that both *D* and *H* are powerful in detecting selection, but they are sensitive to different demographic factors and are affected by background selection to different degrees. Therefore, one may conjecture that *D* and *H* in combination may be sensitive only to selection and less sensitive to other forces. For example, the sensitivity of *D* to population expansion may be counterbalanced by the insensitivity of *H* to the same factor in the joint *DH* test. To carry out the joint *DH* test, we first define

$$f_s(p) = P\{d(X_s) \leq d_p \text{ and } h(X_s) \leq h_p\} \quad 0 < p < 1, \tag{15}$$

where X_s is random samples (of size *n*) under neutrality with *s* ($s \geq 1$) segregating sites, $d(X)$ is the value of the *D*-statistic of sample *X* and $h(\cdot)$ is that of the normalized *H*-statistic, and d_p and h_p are critical values such that $P\{d(X_s) \leq d_p\} = P\{h(X_s) \leq h_p\} = p$. In other words, for a given *s*, when the significance levels of both *D* and *H* are *p*, that of the joint *DH* test is $f_s(p)$. That is, for samples with *s* segregating sites, the rejection region of the *DH* test at the significance level of $f_s(p)$ is $\{X_s \mid d(X_s) \leq d_p \text{ and } h(X_s) \leq h_p\}$. In practice, for a given value of *p*, $f_s(p)$ can be evaluated using neutral simulations with the number of segregating sites *s* fixed (HUDSON 1993). In our imple-

mentation, at least 50,000 rounds of simulations were carried out. To determine a value of *p* (and consequently d_p and h_p) such that the significance level of *DH* is, for example, 5%, we solved the equation $f_s(p) = 0.05$ using the bisection algorithm (PRESS *et al.* 1992). We denote the solution as p^* . We also denote d_p and h_p corresponding to p^* as d^* and h^* , respectively. In Table 1, we give the p^* -, d^* -, and h^* -values for three significance levels of *DH* [*i.e.*, $f_s(p^*)$], given various observed values of *s*. In general, p^* is close to three times the value of $f_s(p^*)$.

APPLICATIONS

Here we apply the tests discussed previously to data simulated under a variety of conditions for the purpose of evaluating their statistical powers. Such simulations have been done extensively in the past for studying existing tests (BRAVERMAN *et al.* 1995; SIMONSEN *et al.* 1995; FU 1997; PRZEWORSKI 2002). Here all tests were one-sided and values falling into the lower 5% tail of the null distribution were considered significant. The reason for the one-sided test is that positive selection predictably shifts the pattern to one side of the null distribution. We report only the results of the normalized *H*-test, because it is always more powerful than the original one (varying from ~1 to 8%). Details of the simulation algorithms and other related issues are described in the APPENDIX.

The power to detect positive selection before and after fixation: For the linked neutral locus, we define $R(i) = E(\xi'_i)/E(\xi_i)$, where $E(\xi'_i)$ and $E(\xi_i)$ are the expected numbers of segregating sites of size *i* under selection and neutrality, respectively (FU 1997). Figure 2 shows the dynamics of $R(i)$ before and after the fixation of an advantageous mutation. The reduction in intermediate-frequency variants and the accumulation of high-frequency variants start at the early stage of the sweep (Figure 2A, $f = 0.3$; *f* is the population frequency of the advantageous mutation). When the frequency of the advantageous mutation reaches 60% in the population, the spectrum is already highly deviant from neutrality, losing many of the intermediate-frequency

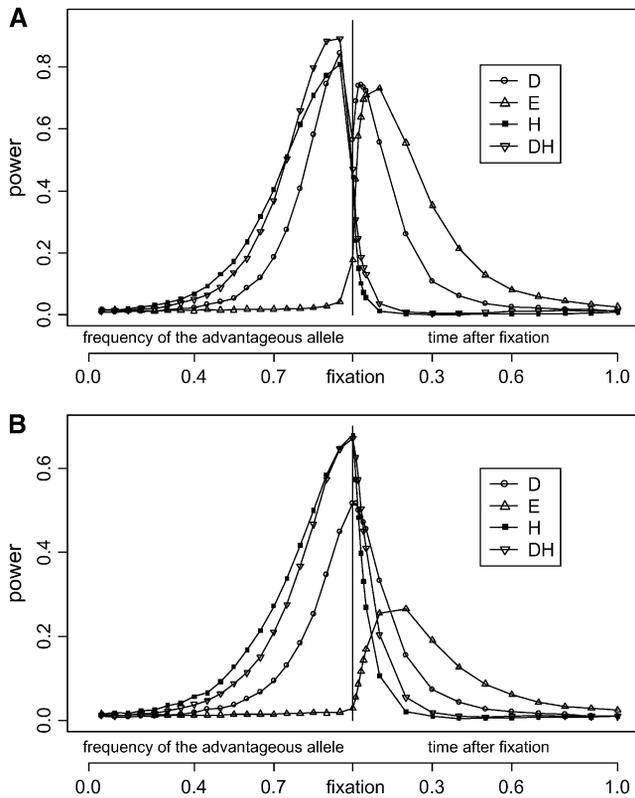


FIGURE 3.—Power of the tests before and after hitchhiking is completed. The x -axis on the left represents the increase in the frequency of the advantageous mutation; on the right is the time after fixation (measured in units of $4N$ generations). All parameter values are the same as those of Figure 2. All tests were one-sided; values falling into the lower 5% tail of the null distribution were considered significant. Results shown in Figures 4–6 were produced by the same method. (A) $c/s = 2 \times 10^{-5}$; (B) $c/s = 0.02$.

variants (Figure 2A, $f = 0.6$). By the time of fixation, the population has lost virtually all the intermediate-frequency variants, and mutations appear at either very high or very low frequency (Figure 2B, $\tau = 0$; τ is time after fixation of the advantageous mutation and is measured in units of $4N$ generations). In neutral equilibrium, there should be 22.4 mutations in the sample on average for $\theta = 5$ and $n = 50$. However, only 11.4 mutations, on average, are seen in the samples at $\tau = 0$ (Figure 2B), among which 57% have a frequency $<10\%$, and 31% have a frequency $>80\%$. Strikingly, the mean number of mutations of size 49 is 7.2 times that of the neutral expectation. When $\tau > 0$, these high-frequency variants quickly drift to fixation and no longer contribute to polymorphism. For a long period after fixation, fewer-than-expected numbers of intermediate- and high-frequency mutations are observed (Figure 2B).

Figure 3 shows how the tests behave before and after the fixation of the advantageous mutation. Before fixation (Figure 3, A and B, left side), the power of D , H , and DH increases rapidly as the frequency of the advantageous mutation increases. H is generally more powerful than D in this stage. The behavior of DH

resembles that of H , but note that it is often the most powerful test when the advantageous allele is in high frequency. We should note again that the purpose of the DH test is not to increase the power in detecting positive selection over either D or H but to decrease sensitivity to demographical factors and background selection. This is apparent in later sections. In contrast, the E -test has little power. This is expected as the reduction of moderate-frequency variants and the accumulation of high-frequency variants characterize this stage (Figure 2A). Sometimes, H , D , and DH reach their peak of power before fixation. This happens when too much variation is removed by the selective sweep, as is the case in Figure 3A. After fixation (Figure 3, A and B, right side) the E -test quickly becomes the most powerful test ($\tau \approx 0.1$ in Figure 3A; $\tau \approx 0.15$ in Figure 3B). From Figure 2B, we can see that after fixation the low-frequency part of the spectrum recovers first, and the high-frequency part returns to its equilibrium level last. The result of the E -test fits this observation. Furthermore, since the recovery phase is much longer than the selective phase when selection is strong, the E test can be useful for detecting sweep. The D -test performs reasonably well in the recovery stage. The power of the H -test, however, decreases quickly. As a consequence, the power of DH also decreases but at a much lower rate than that of H .

The power to detect balancing selection: Intuition would suggest that the results of an incomplete sweep (Figure 3) may also apply to the initial stage of balancing selection (*i.e.*, a relatively short time period after the birth of the selected allele) under which the selected allele would reach an equilibrium frequency that is high, but $<100\%$. Furthermore, as the selected allele does not reach fixation, the hitchhiking variants also may not have extreme frequencies. As a result, the signature of positive selection and hitchhiking may linger longer under balancing selection.

Simulation results (see supplemental Figure S1 at <http://www.genetics.org/supplemental/>) indeed suggest that in the phase when the selected allele is approaching the equilibrium frequency (75% in this case), the tests behave very much like the patterns in the left side of Figure 3. After the equilibrium is reached, the powers of H and DH also remain higher than in the right side of Figure 3 for a longer period of time. Nevertheless, the gain is modest as the power to detect selection generally diminishes when the selected allele stops increasing in frequency.

Sensitivity of the tests to other driving forces: A good test of a particular population genetic force, say, positive selection, should be sensitive to that force and to that force only. Although a test that is sensitive to many different factors can be a useful general tool, it is ultimately uninformative about the true underlying force. Hence, power can in fact be a blessing in disguise. We examine the sensitivity of these tests to forces other than positive selection and balancing selection below.

TABLE 2

Powers of D , E , H , and DH to detect background selection

U	N	θ	D	E	H	DH
0.01	2,500	1	0.06	0.05	0.04	0.04
		10	0.06	0.06	0.05	0.05
	50,000	1	0.05	0.04	0.05	0.04
0.1	2,500	1	0.08	0.02	0.04	0.01
		10	0.34	0.34	0.02	0.09
	5,000	1	0.06	0.01	0.04	0.01
0.2	2,500	1	0.09	0.02	0.02	0.01
		10	0.60	0.55	0.02	0.07
	5,000	1	0.07	0.00	0.02	0.00
0.1	5,000	1	0.06	0.01	0.04	0.01
		10	0.23	0.23	0.03	0.06
	10,000	1	0.04	0.00	0.04	0.01
0.1	10,000	1	0.04	0.00	0.04	0.01
		10	0.15	0.14	0.03	0.06
	20,000	1	0.02	0.00	0.04	0.00
0.1	20,000	1	0.02	0.00	0.04	0.00
		10	0.11	0.10	0.04	0.05
	50,000	1	0.02	0.00	0.05	0.00
0.1	50,000	1	0.02	0.00	0.05	0.00
		10	0.07	0.06	0.05	0.04
	0.2	2,500	1	0.09	0.02	0.02
0.2	2,500	1	0.09	0.02	0.02	0.01
		10	0.60	0.55	0.02	0.07
	5,000	1	0.07	0.00	0.02	0.00
0.2	5,000	1	0.07	0.00	0.02	0.00
		10	0.47	0.35	0.03	0.05
	10,000	1	0.04	0.00	0.03	0.00
0.2	10,000	1	0.04	0.00	0.03	0.00
		10	0.35	0.18	0.02	0.03
	20,000	1	0.02	0.00	0.03	0.00
0.2	20,000	1	0.02	0.00	0.03	0.00
		10	0.21	0.08	0.03	0.02
	50,000	1	0.02	0.00	0.01	0.00
0.2	50,000	1	0.02	0.00	0.01	0.00
		10	0.11	0.02	0.03	0.01

Sample size is 50. A total of 5000 samples were simulated for each parameter set. All runs employed values of h (dominance coefficient) and s (selection coefficient) of 0.2 and 0.1; *i.e.*, $sh = 0.02$. U is the deleterious mutation rate per diploid genome. N is the effective population size.

Background selection: Selection against linked deleterious mutations maintained by recurrent mutation, often referred to as background selection, can have effects on the level of genetic diversity similar to those of a selective sweep (CHARLESWORTH *et al.* 1993). The distinction between selective sweep and background selection is therefore crucial. Fortunately, the two modes of selection often have very different effects on the frequency spectrum (FU 1997). For example, background selection is not likely to have any effect on high-frequency variants (FU 1997) and its effect on the low-frequency variants depends on U and N , where U is the deleterious mutation rate per diploid genome and N is the effective population size.

Table 2 summarizes the power of the tests to detect background selection. Both D and E are sensitive, but D is slightly more powerful than E . In general, the power increases as U increases, but this increase also depends on N . For a given U , the power of the tests decreases as N increases. H is not affected in all cases and hence is discriminatory between the two modes of selection. Significantly, DH is also not sensitive to background selection. It seems that the sensitivity of D is well counterbalanced by the insensitivity of H .

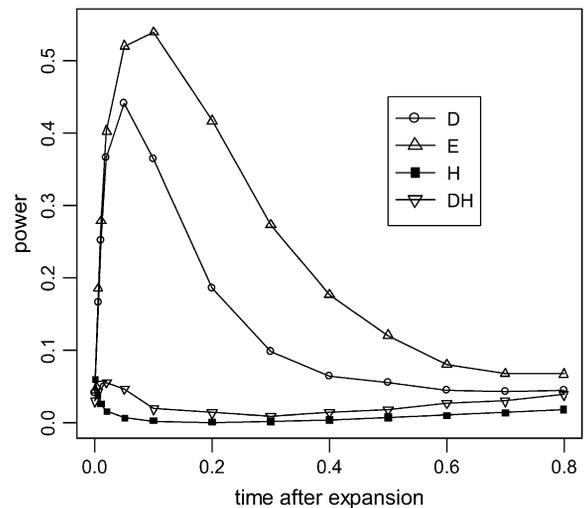


FIGURE 4.—Sensitivity (or power) of the tests to population expansion. We assume that the effective population size increases 10-fold instantaneously at time 0 to $\theta = 5$. Sample size (n) is 50. Time is measured in units of $4N$ generations.

Population growth: When a population increases in size, it tends to have an excess of low-frequency variants (SLATKIN and HUDSON 1991; FU 1997; GRIFFITHS 2003). Both D and E are sensitive to this type of deviation (Figure 4). E is the most sensitive test because high-frequency variants are the last to reach the new equilibrium after expansion. In contrast, H is unaffected. Strikingly, DH has little power in this case even though D exhibits tremendous power.

Population shrinkage: When the population decreases in size, the number of low-frequency variants tends to be smaller than that of the intermediate- and high-frequency ones (FU 1996). Thus, H can be sensitive to population shrinkage, whereas D , E , and DH are largely unaffected (Figure 5).

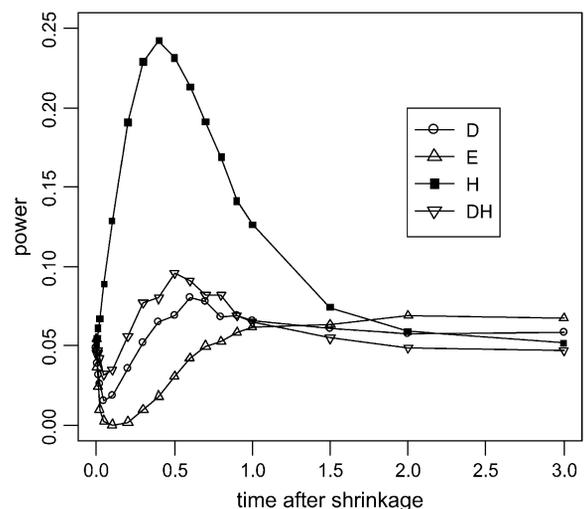


FIGURE 5.—Sensitivity (or power) of the tests to population shrinkage. We assume that the effective population size decreases 10-fold instantaneously at time 0 to $\theta = 2$. Sample size (n) is 50. Time is measured in units of $4N$ generations.

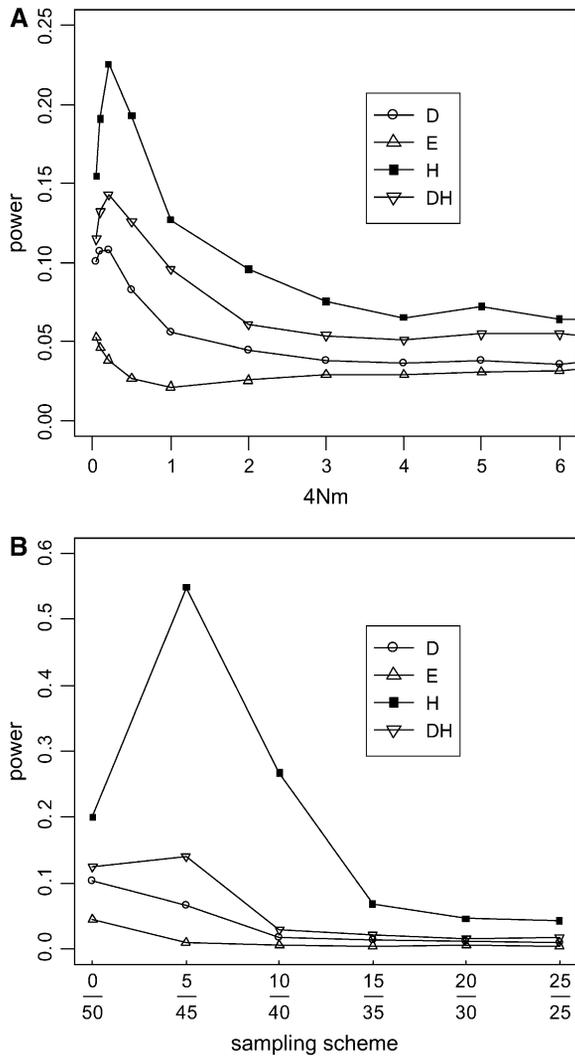


FIGURE 6.—Sensitivity (or power) of the tests to population subdivision. A symmetric two-deme model with $\theta = 2$ per deme ($2N$ genes per deme) was simulated. Populations are assumed to be in drift–migration equilibrium with symmetric migration at a rate of m , which is the fraction of new migrants each generation. Sample size (n) is 50. (A) Sensitivity as a function of the degree of population subdivision, expressed as $4Nm$ on the x -axis. All genes were sampled from one subpopulation. (B) Sensitivity as a function of the sampling skewness; for example, 5/45 means 5 genes are sampled from one subpopulation and 45 from the other. In this case, $4Nm = 0.1$, a value at which the tests show sensitivity to population subdivision in A.

Population subdivision: When the population is structured, one commonly uses the fixation index F_{ST} as a measure of genetic differences among subpopulations. In the symmetric two-deme model, $F_{ST} = 1/(1 + 16Nm)$, where m is the fraction of new migrants in the population and N is the population size of a deme (NORDBORG 1997). When there is strong population subdivision ($F_{ST} > 0.2$ or $4Nm < 1$, for example) and samples do not come equally from all subpopulations, the frequency spectrum is deviant from the neutral equilibrium (results not shown). Figure 6A gives an example where all

TABLE 3

A qualitative summary of the powers of the four tests to detect various population genetic forces

Driving force	<i>D</i>	<i>H</i>	<i>E</i>	<i>DH</i>
Positive selection (before fixation)	+++	+++	–	+++
Positive selection (after fixation)	+++	++	+++	++
Balancing selection (initial stage)	++	+++	–	+++
Background selection	++	–	++	–
Population growth	+++	–	+++	–
Population shrinkage	+	++	–	+
Subdivision	+	++	–	+

The results were drawn on the basis of Figures 3–6 and Table 2. All tests were one-sided; values falling into the lower 5% tail of the null distribution were considered significant. – denotes the lack of power under the specified condition. + denotes very weak power. ++ denotes moderately powerful. +++ denotes very powerful.

genes are taken from one subpopulation. In this case, *H* is the most sensitive test. The *D*-test is sensitive only when the subdivision is strong ($4Nm < 0.5$ or $F_{ST} > 0.33$). *DH* has some sensitivity. Its power lies in between those of *D* and *H*, but never rises above 15%. Note that, for relatively mild population subdivisions with $4Nm > 2$, none of these four tests is notably affected. Interestingly, *E* is completely insensitive to population subdivision. Figure 6B shows the effect of sampling on the power of the tests ($4Nm = 0.1$). The *E*-test again shows no sensitivity. Although the power of *H* can be as high as 55%, that of the *DH* test never goes above 14%. This result shows the effectiveness of using *D* and *H* in combination. However, when sampling becomes less and less biased, all tests become progressively less sensitive to population structure.

CONCLUSION

We summarize our results in Table 3. It is clear that, among the four tests studied, *DH* is sensitive primarily to directional selection. In contrast, other tests are sensitive to two or more other driving forces. *D* is indeed a “general purpose” test as it is sensitive, to various extents, to all the processes we considered (Table 3). Being exclusive is a desirable property because it minimizes false positive results in the search for selectively favored mutations. However, no test is powerful in every stage of the selection process. Therefore, in practice, we should utilize *a priori* information (if available) from other sources to help us choose an appropriate test.

We thank J. Braverman and R. R. Hudson for their helpful discussions about the simulation algorithms. We also thank two anonymous reviewers for their constructive comments. K.Z. is supported by Sun Yat-sen University and the Kaisi Fund. S.S. is supported by grants

from the National Natural Science Foundation of China (30230030, 30470119, 30300033, and 30500049). Y.-X. Fu is supported by National Institutes of Health (NIH) grants (GM 60777 and GM50428) and funding from Yunnan University, China. C.-I Wu is supported by NIH grants and an OOCs grant from the Chinese Academy of Sciences.

LITERATURE CITED

- BRAVERMAN, J. M., R. R. HUDSON, N. L. KAPLAN, C. H. LANGLEY and W. STEPHAN, 1995 The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics* **140**: 783–796.
- BUSTAMANTE, C. D., R. NIELSEN, S. A. SAWYER, K. M. OLSEN, M. D. PURUGGANAN *et al.*, 2002 The cost of inbreeding in *Arabidopsis*. *Nature* **416**: 531–534.
- CHARLESWORTH, B., M. T. MORGAN and D. CHARLESWORTH, 1993 The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**: 1289–1303.
- CHARLESWORTH, D., B. CHARLESWORTH and M. T. MORGAN, 1995 The pattern of neutral molecular variation under the background selection model. *Genetics* **141**: 1619–1632.
- FAY, J. C., and C.-I. WU, 2000 Hitchhiking under positive Darwinian selection. *Genetics* **155**: 1405–1413.
- FU, Y. X., 1995 Statistical properties of segregating sites. *Theor. Popul. Biol.* **48**: 172–197.
- FU, Y. X., 1996 New statistical tests of neutrality for DNA samples from a population. *Genetics* **143**: 557–570.
- FU, Y. X., 1997 Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**: 915–925.
- FU, Y. X., and W. H. LI, 1993 Statistical tests of neutrality of mutations. *Genetics* **133**: 693–709.
- GRIFFITHS, R. C., 2003 The frequency spectrum of a mutation, and its age, in a general diffusion model. *Theor. Popul. Biol.* **64**: 241–251.
- HUDSON, R. R., 1993 The how and why of generating gene genealogies, pp. 23–36 in *Mechanisms of Molecular Evolution*, edited by N. TAKAHATA and A. G. CLARK. Sinauer, Sunderland, MA.
- HUDSON, R. R., 2002 Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**: 337–338.
- HUDSON, R. R., and N. L. KAPLAN, 1994 Gene trees with background selection, pp. 140–153 in *Non-Neutral Evolution: Theories and Molecular Data*, edited by B. GOLDING. Chapman & Hall, London.
- HUDSON, R. R., M. KREITMAN and M. AGUADÉ, 1987 A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**: 153–159.
- HUDSON, R. R., K. BAILEY, D. SKARECKY, J. KWIATOWSKI and F. J. AYALA, 1994 Evidence for positive selection in the superoxide dismutase (*sod*) region of *Drosophila melanogaster*. *Genetics* **136**: 1329–1340.
- KAPLAN, N. L., R. R. HUDSON and C. H. LANGLEY, 1989 The “hitchhiking effect” revisited. *Genetics* **123**: 887–899.
- MARKOVTSOVA, L., P. MARJORAM and S. TAVARÉ, 2001 On a test of Depaulis and Veuille. *Mol. Biol. Evol.* **18**: 1132–1133.
- MAYNARD SMITH, J., and J. HAIGH, 1974 The hitchhiking effect of a favourable gene. *Genet. Res.* **23**: 23–35.
- MCDONALD, J. H., and M. KREITMAN, 1991 Adaptive protein evolution at the *adh* locus in *Drosophila*. *Nature* **351**: 652–654.
- NEI, M., and S. KUMAR, 2000 *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.
- NORDBORG, M., 1997 Structured coalescent processes on different time scales. *Genetics* **146**: 1501–1514.
- PRESS, W. H., S. A. TEUKOLSKY, W. T. VETTERLING and B. P. FLANNERY, 1992 *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, UK.
- PRZEWORSKI, M., 2002 The signature of positive selection at randomly chosen loci. *Genetics* **160**: 1179–1189.
- SABETI, P. C., D. E. REICH, J. M. HIGGINS, H. Z. LEVINE, D. J. RICHTER *et al.*, 2002 Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**: 832–837.
- SIMONSEN, K. L., G. A. CHURCHILL and C. F. AQUADRO, 1995 Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* **141**: 413–429.
- SLATKIN, M., and R. R. HUDSON, 1991 Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129**: 555–562.
- STEPHAN, W., T. H. E. WIEHE and M. W. LENZ, 1992 The effect of strongly selected substitutions on neutral polymorphism - analytical results based on diffusion-theory. *Theor. Popul. Biol.* **41**: 237–254.
- TAJIMA, F., 1983 Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**: 437–460.
- TAJIMA, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
- VOIGHT, B. F., S. KUDARAVALLI, X. WEN and J. K. PRITCHARD, 2006 A map of recent positive selection in the human genome. *PLoS Biol.* **4**: e72.
- WALL, J. D., and R. R. HUDSON, 2001 Coalescent simulations and statistical tests of neutrality. *Mol. Biol. Evol.* **18**: 1134–1135.
- WATTERSON, G. A., 1975 On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**: 256–276.
- YANG, Z., 2003 Adaptive molecular evolution, pp. 229–254 in *Handbook of Statistical Genetics*, edited by D. BALDING, M. BISHOP and C. CANNINGS. Wiley, New York.

Communicating editor: N. TAKAHATA

APPENDIX

Analytical results: We list some useful properties of the θ -estimators (see Y.-X. Fu, unpublished data, for mathematical details):

$$\begin{aligned} \text{Cov}(\theta_{\pi}, \theta_L) &= \frac{n+1}{3(n-1)}\theta + \frac{7n^2 + 3n - 2 - 4n(n+1)b_{n+1}}{2(n-1)^2}\theta^2, \\ \text{Cov}(\theta_L, \theta_W) &= \frac{\theta}{a_n} + \frac{nb_n - n + 1}{a_n(n-1)}\theta^2. \end{aligned} \quad (\text{A1})$$

Since $\theta_H = 2\theta_L - \theta_{\pi}$, using (A1), we have

$$\text{Var}(\theta_H) = \theta + \frac{2[36n^2(2n+1)b_{n+1} - 116n^3 + 9n^2 + 2n - 3]}{9n(n-1)^2}\theta^2. \quad (\text{A2})$$

$\text{Var}(\theta_{\pi} - \theta_L)$ and $\text{Var}(\theta_L - \theta_W)$ can be calculated using the above formulas.

Simulation algorithms: *Positive selection:* We used the coalescent process with a selective phase (KAPLAN *et al.* 1989; STEPHAN *et al.* 1992; BRAVERMAN *et al.* 1995) to simulate random samples and estimate the power of various tests. The model assumes that selection at the favored locus is additive and the selective pressure is strong ($2Ns \gg 1$). In this case, the trajectory of the frequency of the advantageous allele in the interval with endpoints ϵ and $1 - \epsilon$ is approximately deterministic (STEPHAN *et al.* 1992). We used $\epsilon = 1/(2N)$ in this study. Note that the choice of ϵ has little effect on the results (BRAVERMAN *et al.* 1995). In fact, the results were identical when we ran simulations with $\epsilon = 100/(2N)$ or $\epsilon = 1/(4N)$. Our implementation of the algorithm followed the description in BRAVERMAN *et al.* (1995), except for the following features: (1) we used a given value of the population mutation rate ($\theta = 4N\mu$), rather than a fixed number of segregating sites, (2) we allowed intragenic recombination within the neutral locus to happen through the whole process (*i.e.*, during both neutral and selective phases), (3) we assumed that only one round of selective sweep had occurred in the history of a sample rather than recurrent sweeps, and (4) the

recombination distance between the selective locus and the neutral locus (c) was a given value rather than a random variable.

We extended the coalescent simulation algorithm described above to incorporate the case of incomplete hitchhiking. Suppose that the frequency of the favored allele is f . By solving the deterministic equation given by STEPHAN *et al.* (1992; Equation 1 in BRAVERMAN *et al.* 1995), one may obtain the time since the birth of this advantageous mutation and the corresponding frequency trajectory. In each replicate of simulation, we assigned the number of genes linked to the advantageous allele by binomial sampling with parameters n and f . Then the genealogy was constructed following the same procedure as described above.

Background selection and demographic scenarios: We simulated background selection using the coalescent method (HUDSON and KAPLAN 1994; CHARLESWORTH *et al.* 1995). The software package kindly provided by HUDSON (2002) was used to simulate all demographic scenarios.

Critical values of the tests: The tests, D , H , and E , were one-sided; values falling into the lower 5% tail were considered significant. The critical values of these three tests were determined by the Monte Carlo method used by FU (1997). For a sample of size n , we first obtained an estimate $\hat{\theta}$ of θ using Watterson's estimator. Then we ran 100,000 rounds of coalescent simulations with no recombination and $\theta = \hat{\theta}$. The critical values were determined by examining the empirical distribution obtained. The supplemental table (at <http://www.genetics.org/supplemental/>) lists the critical values of the E -test for a number of θ -values. Another method to determine critical values was proposed by HUDSON (1993). The differences between these two methods have been discussed extensively (MARKOVTSOVA *et al.* 2001; WALL and HUDSON 2001). The results produced by both methods are quite similar; therefore, we report only those generated by the first method. The critical values of the DH test were generated using the second method.