Comparisons of Site- and Haplotype-Frequency Methods for Detecting Positive Selection

Kai Zeng, Shuhei Mano,† 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, China; †Graduate School of Natural Sciences, Nagoya City University, Nagoya, Japan; and ‡Department of Ecology and Evolution, University of Chicago

In this report, we compare the differences between various site- and haplotype-frequency tests in their power to detect positive selection by doing computer simulations. Our results are the following. 1) Although haplotype-frequency tests that are conditional on the number of haplotypes (K) were developed for nonrecombining haplotypes, these tests are insensitive to recombination. Such tests, including the Ewens–Watterson (EW) test, can therefore be applied to recombining haplotypes. 2) Tests conditional on the number of segregating sites (S) become overly conservative in the presence of recombination. 3) The EW test is usually the most powerful test during the sweep phase, especially when the local recombination rate is high. 4) The "extended haplotype homozygosity" test relies heavily on the prior knowledge of the target of selection. With that knowledge, it is the most powerful test, whereas in the absence of this prior information, the test has little power. We also study the sensitivities of the haplotype-frequency tests to background selection and various demographic forces. We find that these tests are sensitive to some forces other than positive selection. To alleviate the problem of low specificity, compound tests, such as the DH test (Zeng et al. 2006), may be a solution. In the companion paper (Zeng K, Shi S, Wu C-I, in preparation), we use the EW test to devise 2 compound tests, which are more powerful in detecting positive selection than DH, but are also relatively insensitive to demography.

Introduction

DNA sequence data are a rich source for detecting adaptive evolution. Many statistical methods have been proposed for this purpose (see Fay and Wu 2003; Nielsen 2005; Biswas and Akey 2006; Sabeti et al. 2006 for recent reviews). In this study, we focus on methods using within-species polymorphism data. These methods can be loosely classified into 3 categories—site-frequency, haplotype-frequency, and linkage disequilibrium (LD) methods.

Site-Frequency Methods

These methods require only frequencies of variants at polymorphic nucleotide sites. Linkage phase of these variants is neither required nor used. These methods are in general based on the infinite-site model (Kimura 1969; Watterson 1975) and utilize the site-frequency spectrum for detection (Fu 1995). Tajima's D test (1989; referred to as the D test) is the first test of this type. Many other D-like tests have since been proposed, including Fu and Li's tests (1993), Fu's tests (1996, 1997), Fay and Wu's H test (2000), and 2 newly derived ones—E and DH (Zeng et al. 2006). Many papers have examined the properties of these tests (Braverman et al. 1995; Simonsen et al. 1995; Przeworski 2002; Zeng et al. 2006). There are other site-frequency methods which represent a major departure in methodology from the tests mentioned above and will not be considered in this study (e.g., Kim and Stephan 2002; Jensen et al. 2005).

Among these tests, the DH test is of some interest. Its proposal was motivated by the observation that Tajima's D test and Fay and Wu's H test are both powerful in detecting selection, but they are sensitive to different demographic factors and are affected by background selection to different degrees (Zeng et al. 2006). Therefore, by combining these 2 tests, the sensitivity of either test to a certain demographic factor (e.g., population growth) is counterbalanced by the insensitivity of the other test to the same factor. The compound test thus has high specificity to positive selection.

Haplotype-Frequency Methods

Methods in the second category require additional information on the linkage phase among variant sites and score a haplotype as an allele. They examine the level of haplotype polymorphism using simple summary statistics. The 3 widely used statistics are haplotype homozygosity (F), frequency of the most common haplotype (M), and configuration of haplotype frequencies (C). These measures are expressed either as statistics conditional on the number of haplotypes (K) or as statistics conditional on the number of segregating sites (S), elaborated below.

Statistics Conditional on the Number of Haplotypes (Alleles)

The statistics of F, M, or C conditional on K can be expressed in terms of Ewens' sampling distribution, which was derived under the no-recombination infinite-allele model (Ewens 1972; Karlin and McGregor 1972). One of the most remarkable results is the "invariant" property—the distributions of these conditional statistics (i.e., $F|K$, $M|K$, and $C|K$) are independent of the fundamental parameter of population genetics, θ ($=4Nu$, where N is the effective population size of a diploid organism and u is the neutral mutation rate of the locus). This property is useful as the estimation of θ may introduce further uncertainties (Donnelly and Tavaré 1995). We shall refer to $F|K$ (haplotype homozygosity conditional on the number of haplotypes) as the Ewens–Watterson (EW) test statistic (Watterson 1978). The other 2 similar test statistics, $M|K$ and $C|K$, were proposed by Ewens (1973) and Slatkin

Key words: positive selection, haplotype-frequency tests, site-frequency tests.

E-mail: kzeng@uchicago.edu.

Mol. Biol. Evol. 24(7):1562–1574. 2007
doi:10.1093/molbev/msm078
Advance Access publication April 21, 2007

(1994, 1996), respectively. All the 3 test statistics are independent of θ and are in fact highly correlated when positive selection is in operation (see Results). Because the invariant property is true only when there is no intragenic recombination, the sensitivity to (or robustness against) recombination is crucial to the applicability of these 3 test statistics to DNA haplotype data and will be addressed in this report. There are several other test statistics, which use the sampling distribution of K directly (Strobeck 1987; Fu 1996, 1997). The null distributions of these statistics are therefore dependent on θ. Properties of these tests have been studied elsewhere (Fu 1996, 1997) and are not considered here.

Statistics Conditional on the Number of Segregating Sites

These methods are based on the infinite-site model. Similar to the haplotype tests conditional on K mentioned above, these methods usually use F, M, and C to measure levels of variability, but conditional on the number of segregating sites S. Examples include the haplotype partition test of Hudson et al. (1994; referred to as Hudson's test), the haplotype diversity and haplotype number tests (Depaulis and Veuille 1998), and the full configuration test (Innan et al. 2005). Typically, the null distributions of these tests are obtained by doing coalescent simulation with the number of segregating sites fixed (Hudson 1993; Wall and Hudson 2001). Recombination is generally not considered in the null distribution.

LD Methods

Most methods in this category incorporate recombination in their null distributions. They use various statistics to summarize patterns of LD, including length of unrecombined segment (Slatkin and Bertorelle 2001), extent of haplotype sharing (Toomajian et al. 2003), and pairwise measure of LD (Kelly 1997; Wall 1999; Kim and Nielsen 2004). The most popular test in this category is probably the extended haplotype homozygosity (EHH) test (Sabeti et al. 2002) and its extensions (Hanchard et al. 2006; Voight et al. 2006; Wang et al. 2006). The principle underlying the EHH test is that, under neutrality, a new variant requires a long time to reach high frequency in the population. During this period, substantial recombination would have broken down the haplotype on which the mutation occurred. Therefore, long-range LD surrounding a site is considered as the signature of positive selection. Because the information on the target of selection is not easily available and is not used by other tests, our study of the EHH test is limited to the effect of this information on its power.

The main task of this report is to assess the power of haplotype- and site-frequency methods to detect positive selection (also referred to as genetic hitchhiking or selective sweep; Maynard Smith and Haigh 1974). Although several previous studies (Fu 1996, 1997; Depaulis et al. 2003, 2005) have addressed similar issues, there are important differences. First, in this study, we incorporate intragenic recombination and analyze its effects on the applicability and power of the tests. Second, we concentrate on the prefixation (or sweep) phase when the advantageous allele is sweeping through the population. In this phase, the patterns of polymorphism and LD show characteristics that are strongly associated with hitchhiking (Fay and Wu 2000; Stephan et al. 2006; Zeng et al. 2006). In contrast, the postfixation phase is characterized by an excess of low-frequency variants (i.e., accumulation of new mutations). Such an excess is a general characteristic among many processes including population growth, background selection, and recovery from bottleneck (e.g., Fu 1997; Zeng et al. 2006). In practical terms, it may be difficult to distinguish the postfixation phase of sweep from those processes using the single-locus methods considered here. Third, we compare the sensitivities of the tests to other factors including background selection and demographic changes. We conclude that all tests except for the *DH* test are sensitive to some forces other than positive selection. In the companion study (Zeng K, Shi S, Wu C-I, in preparation), we develop new compound tests by using the novel properties of the EW test reported in this paper to circumvent the problem of low specificity.

Methods

Statistical Tests

Site-Frequency Tests

Two site-frequency tests, Tajima's D test (1989) and the *DH* test (Zeng et al. 2006), were considered in the simulations. Fay and Wu's H test (2000) was not included because its behavior is quite similar to that of the *DH* test, but with lower specificity to positive selection (Zeng et al. 2006). To carry out the *DH* test, we define

$$f_S(p)=P\{d(X_S)\le d_p \text{ and } h(X_S)\le h_p\}, 0\le p\le 1, \quad (1)$$

where X_S is a random variable representing random samples (of size n) under neutrality with S ($S\ge 1$) segregating sites; $d(X)$ is the value of the D statistic of sample X, and $h(X)$ is that of the H statistic; d_p and h_p are critical values such that $P\{d(X_S)\le d_p\} = P\{h(X_S)\le h_p\} = p$. Note that we used the normalized H test here (Zeng et al. 2006). By definition, $f_S(p)$ is the significance level of the compound *DH* test. In practice, for a prespecified significance level β, we solve $f_S(p) = \beta$ numerically. We denote the solution as p^*, and d_p and h_p corresponding to p^* as d^* and h^*, respectively. Then a sample with S segregating sites, denoted X_S^*, is called significant if $d(X_S^*)\le d^*$ and $h(X_S^*)\le h^*$.

Tests Based on the Ewens Sampling Formula

We focused primarily on the EW test in the simulations because this test shows better statistical properties than the other 2 tests in this category (see Results). The test statistic of the EW test is defined as

$$F|K=n^{-2}\sum_{i=1}^{K} f_i^2 \quad (2)$$

where F is the sample homozygosity; K is the number of haplotypes in a sample of size n; f_i is the number of occurrence of the i-th haplotype (Watterson 1978). Similarly, the test statistic of Ewens' test is defined as $M|K$, where M is the

frequency of the most common haplotype (Ewens 1973), and the test statistic of Slatkin's exact test (1994, 1996) is $C|K$, where $C = (f_1, \ldots, f_K)$ is the configuration of haplotype frequencies. Under the no-recombination infinite-allele model, Ewens (1972; see also Karlin and McGregor 1972) showed that, conditional on K, the distribution of C ($C = (f_1, \ldots, f_K)$) is independent of θ. Thus, $F|K$ and $M|K$ are also independent of θ. Note that our implementation of Slatkin's exact test (1996) was different from the original definition in that we defined the P value of an observed sample as the sum of probabilities of occurrence of configurations, which are equal to or larger than that of the observed configuration. This change was made because when using the original definition of the test, we had little power to detect positive selection (Zeng K, unpublished data).

Haplotype Tests Conditional on the Number of Segregating Sites

The Hudson test rejects neutrality by investigating the probability of occurrence of a subset of sequences with low variation, that is, a major "haplotype class" (Hudson et al. 1994). Here, we used the simplified definition of the test statistic given in Innan et al. (2005):

$$M|S = \max\{f_i, i = 1, \ldots, K\} \quad (3)$$

where S and K are, respectively, the numbers of segregating sites and haplotypes in the sample. The haplotype diversity test due to Depaulis and Veuille (1998) is similarly defined, with $1 - F$ in place of M in equation (3). But for consistency, we used F as the test statistic in our simulations. We will refer to this test as the DV test. The haplotype number test ($K \mid S$) proposed by Depaulis and Veuille (1998) did not outperform any tests in our simulations (Zeng K, unpublished data) and therefore is not discussed here. We did not consider the configuration test proposed by Innan et al. (2005) despite its similarity to Slatkin's exact test because its null distribution is difficult to simulate even for samples of moderate sizes.

The EHH Test

We used a single polymorphic site as the core in the EHH analyses. Because the goal was to look for signal of recent positive sweeps, we used the derived allele at the core as the core haplotype. The EHH statistic is defined as the probability that 2 randomly chosen genes carrying the core haplotype of interest are identical by descent for the entire interval from the core single nucleotide polymorphism to the point x (Sabeti et al. 2002). EHH values are therefore directional and can be calculated in both the 3′ and 5′ directions. Here, we adopted the definition of EHH proposed by Hanchard et al. (2006), that is, we calculated EHH value at the variant farthest away from the core in the 3′ direction and that at the farthest variant in the 5′ direction and took the arithmetic mean of the 2 values.

Determining Critical Values

In this study, all tests were one-sided and were conducted at the 5% significance level. The tail of the null distribution, which can maximize a test's power to detect selection or minimize its sensitivity to recombination was used. For example, for tests EW and EHH, values falling into the upper 5% tail were considered significant; for D and H (performed jointly as the DH test), values falling in the lower tail were considered significant.

Site- and Haplotype-Frequency Tests

The null distributions of the EW and related tests were determined by Slatkin's exact enumeration algorithm (1994). Critical values of Tajima's D, the DH test, and haplotype tests conditional on S were obtained by coalescent simulation with the number of polymorphic sites fixed (Hudson 1993; Wall and Hudson 2001).

The EHH Test

For a sample with S segregating sites, to determine the level of significance of an EHH value at the given focal site (core), we need to generate neutral samples conditional on the frequency of the mutant allele at the core. To do this, first, we have to know the distribution of population frequency of the mutant allele at the core. Using the diffusion approximation, Griffiths (2003) has shown that, for a sample of size n, when the number of occurrence of the mutant allele at the core is b, the population frequency x ($0 < x < 1$) of the mutant allele has the following distribution:

$$f(x|n, b) \propto x^{b-1}(1 - x)^{n-b}. \quad (4)$$

Second, we need to simulate frequency trajectories of the derived allele at the core, starting at the birth of the allele to the present. This can be done by utilizing the reversibility argument of the diffusion, that is, the backward diffusion process (starts at present and goes backward to the time when the mutant allele arose) has the same distribution as the usual forward process conditional on absorption at zero (Griffiths 2003; Coop and Griffiths 2004; Ewens 2004). Properties of conditional diffusion were reviewed in Ewens (2004). Here, we used the pseudo-sampling device proposed by Kimura (1980) to simulate the diffusion process directly (see Griffiths [2003]; Coop and Griffiths [2004]; and Spencer and Coop [2004] for another way of simulating the conditional diffusion process).

In each replica of the simulation, we 1) generated a population frequency x by sampling from equation (4); 2) simulated a trajectory which started at x and ended at 0 using the pseudo-sampling device; 3) constructed genealogy by using the structured coalescent method with b lineages in the derived background and the rest in the ancestral background (Braverman et al. 1995; Zeng et al. 2006); 4) put the remaining S-1 segregating sites on the genealogy according to the shape of the local gene trees; and 5) calculated EHH for this random sample. The critical values were then obtained by examining the empirical distribution produced above. In the simulation above, we also assumed (true) local recombination rate and haplotype phase were known.

Simulation Algorithms

We used the coalescent algorithms implemented in the software package "ms" (Hudson 2002) to generate random

Table 1
The Actual Rejection Probabilities of the EW Test, Hudson's Test, and Tajima's *D* in the Presence of Recombination

		ρ/θ						
θ	Tests	0	1	5	7.5	10	12.5	15
1	EW	0.042	0.045	0.042	0.042	0.045	0.044	0.045
	Hudson	0.039	0.034	0.024	0.022	0.020	0.020	0.022
	Tajima	0.047	0.058	0.048	0.038	0.038	0.038	0.039
2.5	EW	0.046	0.046	0.054	0.048	0.053	0.050	0.069
	Hudson	0.050	0.030	0.012	0.009	0.005	0.006	0.006
	Tajima	0.047	0.032	0.022	0.022	0.018	0.020	0.022
5	EW	0.047	0.053	0.057	0.058	0.057	0.058	0.061
	Hudson	0.056	0.019	0.001	0.000	0.000	0.000	0.000
	Tajima	0.051	0.038	0.010	0.008	0.006	0.005	0.002
7.5	EW	0.052	0.054	0.056	0.048	0.069	0.062	0.058
	Hudson	0.050	0.013	0.000	0.000	0.000	0.000	0.000
	Tajima	0.040	0.016	0.003	0.002	0.000	0.002	0.002
10	EW	0.048	0.059	0.055	0.060	0.053	0.051	0.049
	Hudson	0.053	0.007	0.000	0.000	0.000	0.000	0.000
	Tajima	0.048	0.012	0.000	0.000	0.000	0.000	0.000
12.5	EW	0.052	0.058	0.057	0.050	0.056	0.048	0.045
	Hudson	0.049	0.005	0.000	0.000	0.000	0.000	0.000
	Tajima	0.059	0.008	0.001	0.001	0.000	0.000	0.000

NOTE.—θ and ρ are population mutation and recombination rates of the locus, respectively. All tests were one-sided and were conducted at the 5% level. Every rejection rate was based on 2,000 simulations of sample size 90. Denote the actual rejection rate in a cell as γ. The 95% acceptance region where γ is "not" significantly different from its nominal value 0.05 is roughly $\gamma \in (0.04, 0.06)$.

samples under the neutral model with intragenic recombination and under various demographic models. To simulate hitchhiking, the coalescent process with a selective phase was used (Kaplan et al. 1989; Braverman et al. 1995; Kim and Stephan 2002). This model assumes that the fitnesses of the 3 genotypes *AA*, *Aa*, and *aa* at the selected site are $1 + s$, $1 + hs$, and 1, respectively, where *A* is the derived allele, *s* and *h* are the selection and dominance coefficients. The behavior of the selected allele is mainly governed by the scaled selective pressure α ($=2Ns$) and *h*. Our implementation followed the description in Zeng et al. (2006) with the following modification: frequency trajectories of the selected allele were obtained by using the pseudo-sampling device described above rather than the approximate deterministic model. This modification takes into account the random effects in the early stage of a sweep. It also allows us to examine the effects of selective sweeps with arbitrary level of dominance.

Background selection was simulated using the 2-locus model described in Hudson and Kaplan (1994). This model assumes that the deleterious locus is in mutation-selection equilibrium and is not recombining, but recombination can occur between the neutral locus and the deleterious locus. Here, we extended the Hudson and Kaplan model by allowing intragenic recombination within the neutral locus.

Results

Effects of Intragenic Recombination on Haplotype-Frequency Tests

The haplotype-frequency tests considered in this study were derived under the no-recombination neutral model. To apply them to nuclear DNA sequence data, we need to make sure recombination does not inflate type I error. In table 1, we show the rejection probabilities of 3 tests, EW, Hudson's test, and Tajima's *D* test, generated with various combinations of θ and recombination rate (measured by $\rho = 4Nr$, where *r* is the recombination rate of the region every generation). Results of other tests are very similar and are not presented (e.g., the rejection rates of Ewens' test are very similar to those of the EW test). The most striking result is that the type I error rate of the EW test is insensitive to recombination. This conclusion is true over a wide range of combinations of parameter values (see supplementary table S1, Supplementary Material online), except for some rare occasions with extreme θ and ρ values which may be unlikely to cause problems in practice. When the EW test was conducted at other significance levels (e.g., 2%, 1%, and 0.5%), the conclusion still holds (Zeng K, unpublished data). On the contrary, Tajima's *D* and Hudson's test become very conservative in the presence of intragenic recombination, in agreement with previous reports (Wall 1999; Depaulis et al. 2005). The difference between these 2 tests (Tajima's *D* and Hudson's test) and EW is most significant when ρ/θ is large, in which case the rejection rates of Tajima's *D* and Hudson's test are effectively zero. This result may imply that, when the effect of recombination is un-negligible in the history of a sample, the EW test may be more powerful than Tajima's *D* and Hudson's test (see fig. 3*B* below). The difference in susceptibility to recombination between haplotype tests conditional on *K* and tests conditional on *S* may be due to the fact that *K* and *S* contain a very different amount of information about the local recombination rate. Multiple studies have shown that *K* is a good indicator of local recombinational pressure (Wall 2000; Innan et al. 2005). But *S* does not possess this property. In fact, in a simple regression analysis with ρ as the *x* axis and *S* as the *y* axis, no correlation was observed (Zeng K, unpublished data).

There is a difference between haplotype-frequency tests and site-frequency tests that we need to pay special attention. For a site-frequency test, either tail of the null distribution responses to recombination in a very similar

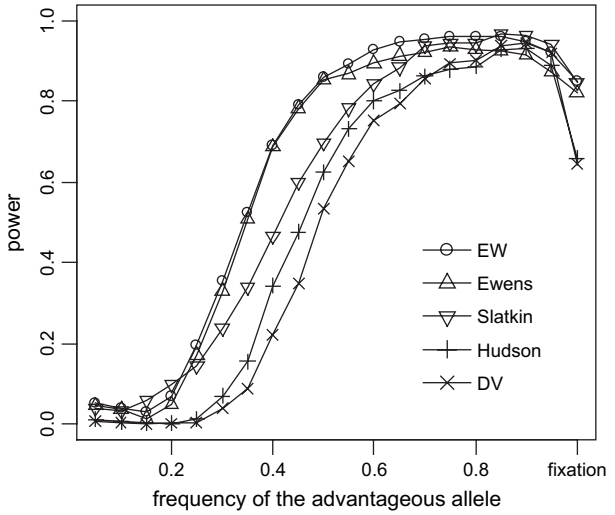

FIG. 1.—Power of haplotype-frequency tests to detect positive selection before fixation. Selective sweep was modeled by the coalescent process described in Methods. We assumed the genic selection model ($h = 0.5$). The scaled selection coefficient α ($=2Ns$, where N is the effective population size and s is the selection coefficient) of the selected allele was 300. Sample size (n) was 90. We assumed that the selected site was in the middle of an otherwise neutral region. The population mutation (θ) and recombination (ρ) rates of the neutral region were both 10.

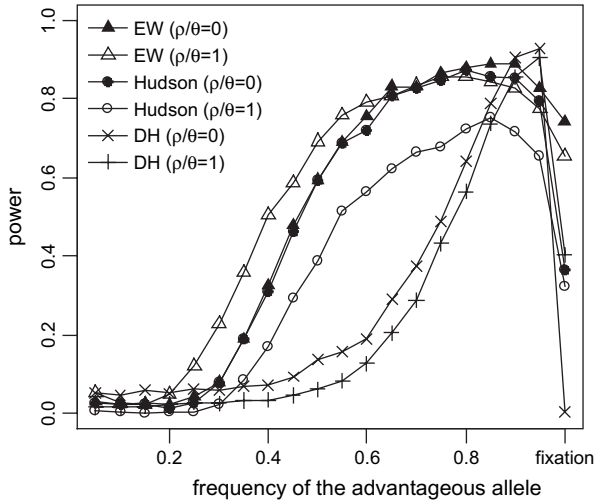

FIG. 2.—The effects of intragenic recombination on the power of the tests to detect positive selection. Positive selection was simulated with $\alpha = 150$, $h = 0.5$. Sample size was 90. The selected site was placed in the middle of a neutral region. The population mutation rate θ of the neutral region was set to 10. Two different levels of intragenic recombination ($\rho/\theta = 0$ or 1) were considered.

manner; however, this is not the case for haplotype tests conditional on S. For example, if we conduct the DV test as a two-sided test, its type I error rate is significantly higher than the nominal significance level in the presence of recombination (Depaulis et al. 2005; supplementary fig. S1, Supplementary Material online).

Power of the Haplotype-Frequency Tests to Detect Positive Selection

In this section, we compare the differences among haplotype-frequency tests in their ability to detect positive selection. The following results allow us to select the more powerful tests to compare with the 2 site-frequency tests in the next section. We only discuss the prefixation phase of a selective sweep for reasons given in the Introduction. An example is given in figure 1. Several features can be observed. First, the power of all tests increases rapidly as a function of the frequency of the advantageous allele. When the frequency of the selected allele reaches 50%, all tests have obtained substantial power. Second, the power curves reach their peaks when the advantageous allele is at very high frequency and then start to decline before fixation. This is probably because at the time around fixation, levels of variation have been significantly reduced and thus, the samples tend to be uninformative. Third, tests conditional on K are generally more powerful than tests conditional on S, both before and after fixation (supplementary fig. S2, Supplementary Material online). This observation is true in most of our simulations (Zeng K, unpublished data; see Discussion).

Lastly, tests based on C, M, and F (full haplotype configuration, frequency of the most common haplotype, and haplotype homozygosity, respectively) yield very similar results. This is true no matter they are conditional on K or S. The similarity is probably due to the presence of a predominant haplotype under hitchhiking (Barton 1998; Depaulis et al. 2003), resulting in a strong correlation among C, M, and F. Hereafter, we choose the EW (i.e., F conditional on K) and the Hudson test (M conditional on S) to represent the 2 kinds haplotype-frequency tests.

Note that the power curves in figure 1 were generated by assuming that the selected site was in the middle of an otherwise neutrally evolving region. It is well known that the strength of hitchhiking depends on the distance between the selected site and the hitchhiking variants (Maynard Smith and Haigh 1974; Stephan et al. 1992; Fay and Wu 2000; Kim and Stephan 2002). Even the LD pattern has been shown to depend on this spatial relationship (Kim and Nielsen 2004; Stephan et al. 2006). In the online supplementary material (supplementary fig. S3), we show that, while the power decreases when the distance between the selected site and the neutral region increases, the relative performances of these tests remain the same as in figure 1.

Contrasting the Power of Haplotype- and Site-Frequency Tests

In this section, we compare the 2 representative haplotype-frequency tests (EW and Hudson's test) with the 2 site-frequency tests, Tajima's D and the compound test, DH (Zeng et al. 2006). We first investigate the effects of intragenic recombination on the power of the tests by generating samples under hitchhiking with or without recombination (fig. 2; Tajima's D is less powerful than DH, and is not shown). The EW test is in fact more robust against recombination than the other 2 tests. On the contrary, Hudson's test tends to suffer the most significant loss of power.

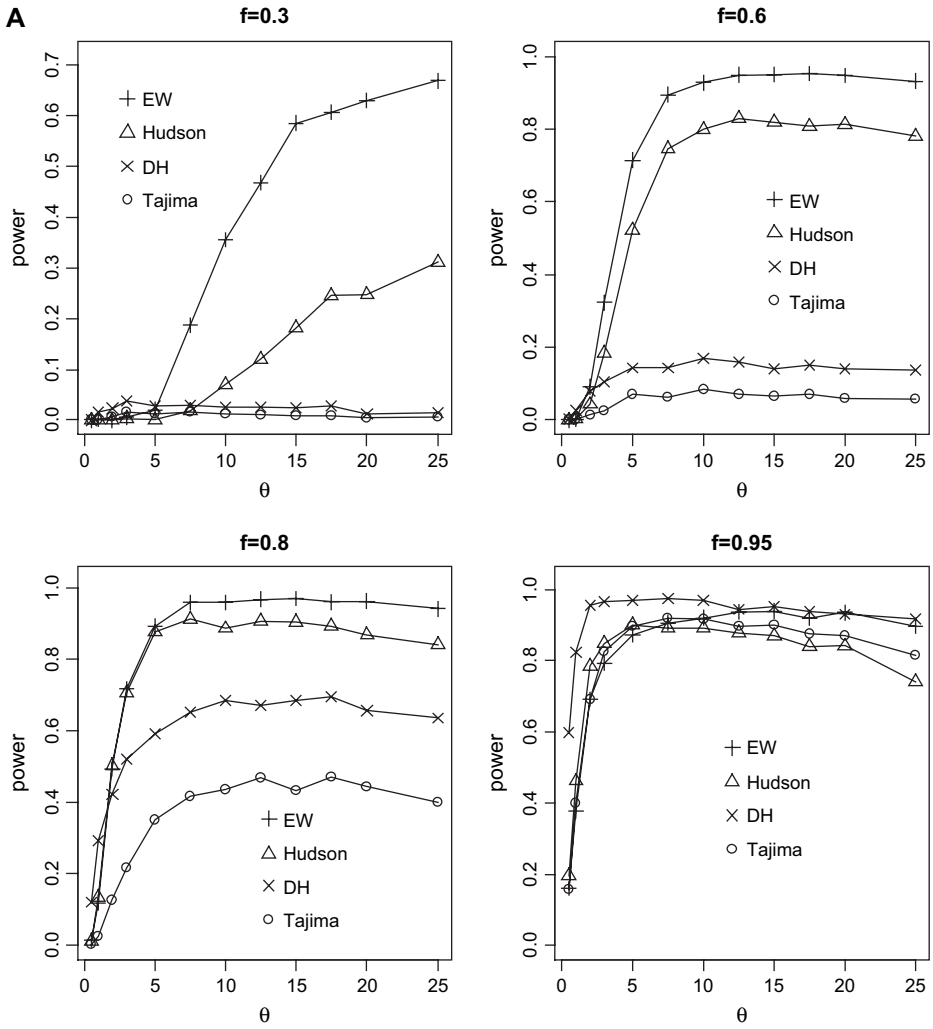

FIG. 3.—Power of the tests to detect positive selection as a function of the size of the neutral region when the advantageous allele was at various frequencies (0.3, 0.6, 0.8, and 0.95 for each panel). Positive selection was simulated with $\alpha = 300$, $h = 0.5$. Sample size was 90. Size of the neutral region was measured by θ, the population mutation rate. The selected site was placed in the middle of the neutral region. Intragenic recombination was included with 2 different intensities: (*A*) $\rho/\theta = 1$ and (*B*) $\rho/\theta = 5$, where ρ is the population recombination rate of the neutral region. Note that the scales on the *y* axis are different.

Figure 3 shows the power properties of the tests when the advantageous allele is at different frequencies. In general, haplotype-frequency tests are much more powerful than site-frequency tests before fixation. The EW test is usually the most powerful test, especially when recombination rate is much higher than mutation rate (fig. 3*A* vs.3*B*). Site-frequency tests can outperform haplotype tests only when the advantageous allele is very close to fixation (e.g., the fourth panel in fig. 3*A*). In agreement with the previous report (Zeng et al. 2006), the *DH* test is more powerful than Tajima's *D* before fixation. In contrast, Hudson's test (representing haplotype tests conditional on *S*) is rarely the best test.

In figure 3*A*, all tests have good power to detect positive selection when *f* (frequency of the selected allele) is 0.95 (fourth panel). For this case, we further explore the effects of the distance between the selected site and the neutral region on the power (fig. 4). As expected, the power decreases when the scaled distance C_{bet}/s increases, where C_{bet} is the recombination distance between the site under selection and the left end of the neutral region (we assume the selected locus is on the left-hand side of the neutral region). Most interestingly, the 2 haplotype tests, especially the EW test, are more powerful than the 2 site-frequency tests over a large range of C_{bet}/s values. In practical terms, the EW test can detect positive selection further away from the site of selection than the other tests. This conclusion is generally true before fixation (Zeng K, unpublished data). The factor underlying the difference between haplotype- and site-frequency tests in the prefixation phase is likely LD. We know that the loss of diversity and the increase in LD are characteristic for this stage of hitchhiking (e.g., Stephan et al. 2006; Zeng et al. 2006). Site-frequency tests hence lose power because they do not consider the increase in level of LD.

After fixation, the patterns of polymorphism and LD caused by hitchhiking will linger on for a short period of time, and what follows is a prolonged period when the

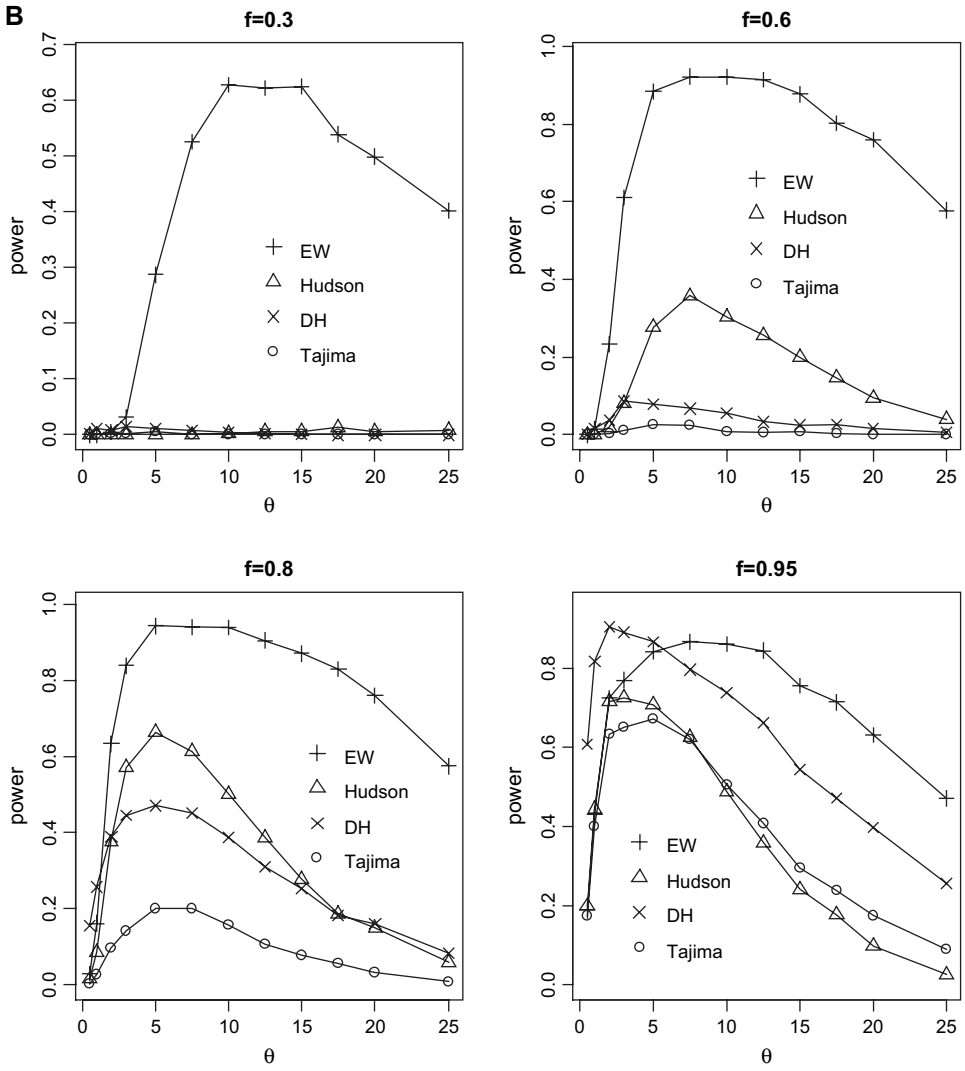

FIG. 3—Continued

population shows a deficit of intermediate- to high-frequency variants (Kim and Stephan 2002; Przeworski 2002; Kim and Nielsen 2004; Zeng et al. 2006). Tajima's *D* and the newly derived *E* test may be most powerful during this latter period (e.g., Zeng et al. 2006). Simulation results indicate a distinct point of transition between the 2 postfixation phases. In supplementary figure S4 (Supplementary Material online), it is around $\tau = 0.05$ (τ is time after fixation, measured in units of $4N$ generations). However, this recovery phase when *D* and *E* are most powerful is in fact hardly distinguishable from the recovery phase of some demographic changes (e.g., population growth) or background selection (e.g., Fu 1997; Zeng et al. 2006). Effectively, the transition point may be the limit of our ability to detect the unique influence of selective sweep.

Power of the EHH Test

The main difference between the EHH test and the other tests mentioned above is that EHH requires the prior definition of the focal site (and core haplotype) and knowledge of the local recombination rate. In figure 5, we show power curves of the EHH test conducted in 2 different ways (i.e., with or without using the site under selection as the core). The power of the EHH test is indeed very high when the precise site under selection is used as the core (the top graph in fig. 5), but its power drops precipitously when the next and closest segregating site is used as the core (the bottom graph). The power of the EW test is in between those of the EHH test and is much closer to the one with the selected site as the core (fig. 5). Considering that the EW test does not directly model recombination and does not require the knowledge of the location of the selected site, its performance is indeed quite impressive. Two factors may result in the "core-dependent" property of the EHH test. First, the site under selection is the defining site for the special pattern of LD caused by the rapid fixation of the advantageous allele (Kim and Nielsen 2004; Stephan et al. 2006). That is LD between 2 neutral loci can be eliminated by a sweep if they are separated by the selected site, whereas strong LD

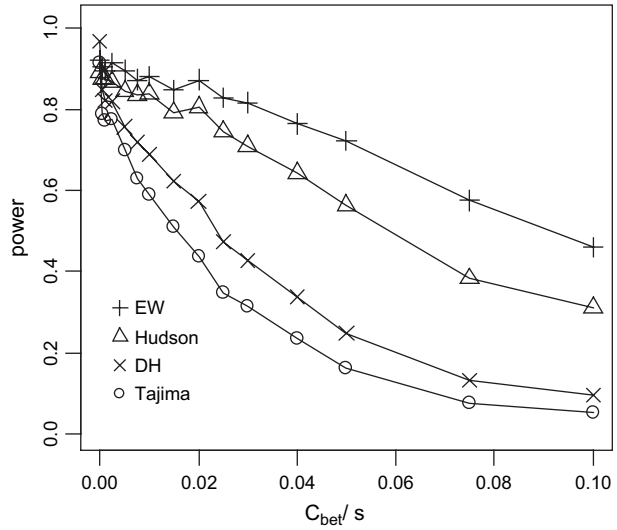

FIG. 4.—Power to detect positive selection as a function of the scaled distance between the selected site and the neutral region (C_{bet}/s) when the frequency of the advantageous allele was 95%. Positive selection was simulated with $\alpha = 300$, $h = 0.5$. Sample size was 90. The population mutation (θ) and recombination (ρ) rates of the neutral region were both set to 10. We assumed the selected site was on the left-hand side of the neutral region. C_{bet} is the recombination distance between the left end of the neutral region and the selected site. Note that $C_{bet}/s = 0$ means the selected site is in the middle of the neutral region.

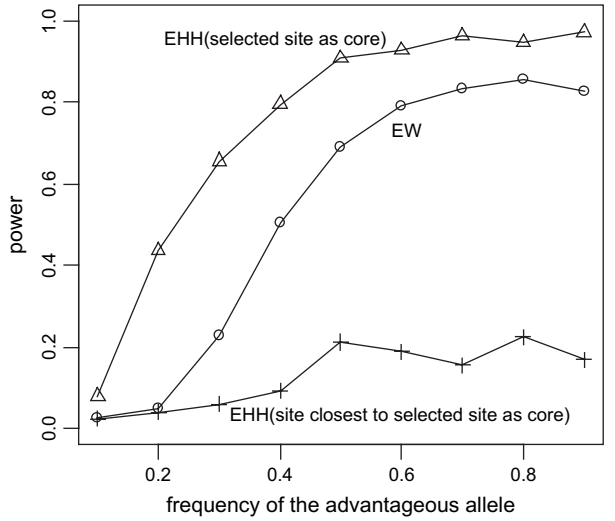

FIG. 5.—Power of the EHH test to detect positive selection before fixation. Positive selection was simulated with $\alpha = 150$, $h = 0.5$. Sample size was 90. The selected site was placed in the middle of the neutral region whose population mutation and recombination rates were both 10. The EHH test was conducted in 2 different ways: EHH(selected site as core) meant the site under selection was used as the core and EHH(site closest to selected site as core) meant the polymorphic site closest to the selected site was used as the core. We used the true population recombination rate ($\rho = 10$) to obtain the critical values of the EHH test (see Methods).

can be observed if the neutral loci are on the same side of the selected site. Hence, when the selected site is not used as the core, the signal of selection may be weakened. Second, ancestral alleles are much more common in the population than derived ones when a sweep starts. Thus, at the closest polymorphic site, it is more likely for the ancestral allele to hitchhike with the advantageous allele, that is, the use of derived allele at this site causes further loss in power. But, in the search of recent sweeps, there is no point to use ancestral alleles as the core haplotype. Recently, Voight et al. (2006) proposed an EHH-like test, the integrated haplotype score (*IHS*) test. They showed that this new test is not dependent on the precise knowledge of location of the selected site. The study of other properties of the EHH test and the *IHS* test is beyond the scope of this paper.

Sensitivities of the Tests to Other Driving Forces

All tests considered so far use the standard neutral model as the null hypothesis. This model makes a number of assumptions (for instance, constant population size). Violation of different assumptions can usually lead to similar test results (e.g., Zeng et al. 2006). In this section, we study the sensitivities of the haplotype- and site-frequency tests to background selection and various demographic factors. We will still use the EW test and Hudson's test to represent haplotype tests conditional on K and S, respectively.

Background Selection

Selection against linked deleterious mutations maintained by recurrent mutation is referred to as background selection (Charlesworth et al. 1993). Typically, background selection results in a reduction in effective population size; it also reduces the levels of polymorphism at linked neutral loci (Charlesworth et al. 1993; Hudson and Kaplan 1994). Previous studies suggest that positive selection and background selection can leave similar patterns of polymorphism at linked neutral loci (Charlesworth et al. 1995; Fu 1997). Thus, discrimination between these 2 very different modes of selection is important.

Here, we used Hudson and Kaplan's 2-locus model (1994) to simulate background selection (see Methods). The recombination distance between the deleterious region and the left end of neutral locus is denoted as C_{bet}. (We assume the deleterious region is on the left-hand side of the neutral region.) The conventional value $sh = 0.02$ was assumed in the simulations, where s and h are, respectively, selection and dominance coefficients for mutant alleles. The results are shown in table 2. There are 2 general trends that apply to all tests except *DH*, which is essentially not affected by background selection. First, the sensitivity of the tests is monotonely increasing as U (deleterious mutation rate per diploid genome) increases. When U is small (say, $U < 0.05$), the tests are not affected even if the population size is relatively small (say, $N \leq 5{,}000$). However, as U gets larger, the tests exhibit inflated type I error rates. In this case, larger populations are affected to a lesser extent. The second trend is that, when deleterious mutation rate is high, we can observe very large haplotype blocks with reduced variability. For example, when N is 10,000, U is 0.1, and C_{bet} is 5×10^{-4}, the rejection rates of the EW test and Hudson's test are still 18.1% and 10.6%, respectively. More significant blocky haplotype structure can be seen when N is smaller (Zeng K, unpublished data).

Table 2
Sensitivity of Various Tests to Background Selection

			Rejection Rate			
N	U	C_{bet}	Tajima	Hudson	EW	DH
5,000	0.01	5×10^{-5}	0.022	0.008	0.069	0.030
		2.5×10^{-4}	0.027	0.009	0.072	0.030
		5×10^{-4}	0.022	0.010	0.063	0.024
	0.05	5×10^{-5}	0.084	0.041	0.134	0.066
		2.5×10^{-4}	0.092	0.049	0.145	0.054
		5×10^{-4}	0.083	0.045	0.148	0.048
	0.1	5×10^{-5}	0.302	0.180	0.299	0.092
		2.5×10^{-4}	0.271	0.154	0.265	0.103
		5×10^{-4}	0.285	0.180	0.299	0.094
10,000	0.01	5×10^{-5}	0.022	0.011	0.059	0.023
		2.5×10^{-4}	0.025	0.011	0.050	0.019
		5×10^{-4}	0.014	0.018	0.061	0.022
	0.05	5×10^{-5}	0.062	0.028	0.085	0.038
		2.5×10^{-4}	0.053	0.030	0.088	0.044
		5×10^{-4}	0.059	0.035	0.082	0.047
	0.1	5×10^{-5}	0.203	0.136	0.204	0.083
		2.5×10^{-4}	0.180	0.119	0.181	0.086
		5×10^{-4}	0.168	0.106	0.181	0.073
25,000	0.01	5×10^{-5}	0.013	0.005	0.048	0.021
		2.5×10^{-4}	0.018	0.006	0.063	0.030
		5×10^{-4}	0.019	0.012	0.065	0.028
	0.05	5×10^{-5}	0.047	0.033	0.072	0.043
		2.5×10^{-4}	0.040	0.020	0.057	0.048
		5×10^{-4}	0.036	0.029	0.067	0.042
	0.1	5×10^{-5}	0.121	0.073	0.121	0.057
		2.5×10^{-4}	0.105	0.074	0.100	0.055
		5×10^{-4}	0.121	0.072	0.114	0.061

NOTE.—Background selection was simulated using the Hudson and Kaplan model with some modifications (see Methods). N is the population size, U is the deleterious mutation rate per generation per diploid genome, and C_{bet} is the recombination distance between the deleterious region and the left end of neutral locus (we assumed that the deleterious region was on the left-hand side of the neutral region). The product of selection and dominance coefficients of the mutants (sh) was set to the conventional value of 0.02. The population mutation rate of the neutral locus was set to $\theta = 10$. Intragenic recombination at the neutral locus was incorporated and was assumed to occur at the same rate as mutation ($\theta = \rho = 10$). Sample size was 90.

Among the tests, the *DH* test has the lowest sensitivity. Hudson's test is more sensitive than *DH*, especially when deleterious mutation rate is high. Tajima's *D* and the EW test are remarkably similar. The sensitivity of these 2 tests is due to the excess of low-frequency variants in the population (Fu 1997).

Population Bottleneck

Population bottleneck is believed to be an important aspect of evolution for species, which may have dispersed from their ancestral range (e.g., the "out-of-Africa" hypothesis of human evolution). Here, we model bottleneck in the following way. The population size at present is N. Going backward in time, the population size reduces exponentially to βN. The length of this size-reducing period is t_b (in units of $4N$ generations). Then the population size restores to N instantaneously (i.e., at time t_b in the past). The effects of mild bottlenecks (say, $\beta > 0.25$) are not detectable by the tests considered (Zeng K, unpublished data). In figure 6, we present results for $\beta = 0.05$. The 4 tests can be divided into 2 groups according to their dynamics—*DH* and Hudson's test form the first group and the other 2 tests form the second group. Tests in the first group are considerably more conservative than those in the second group. The *DH* test is the most conservative test. Its rejection rate never goes above 12% (the peak is 11.7% at $t_b = 0.08$). But the highest rejection rates of other tests are all above 19%. The main difference between Tajima's *D* and the EW test is that the EW test tends to be more powerful in detecting more recent size reduction, whereas Tajima's *D* is better at detecting older ones.

Population Subdivision

In this section, we discuss simulation results obtained by using the symmetric finite island model (Wright 1931) with 2 or 3 subpopulations (demes). The level of differentiation is measured by the F_{ST} statistic, which can be calculated as

$$F_{ST} = \frac{1}{1 + 4N_s md^2/(d-1)^2}, \quad (5)$$

where N_s is the number of breeding individuals per deme, m is the probability that each gene is an emigrant, and d is the number of demes ($d = 2$ or 3) (e.g., Slatkin 1991). In figure 7, we show the rejection rates of the tests as functions of F_{ST}. In both plots, we assume that all genes are sampled from one deme. In general, Tajima's *D* and the EW test behave in a similar manner and are rarely affected. The *DH* test is slightly less conservative than Tajima's *D* and the EW test, but its rejection rate rarely goes above 10%. Hudson's test has similar behavior as the *DH* test when

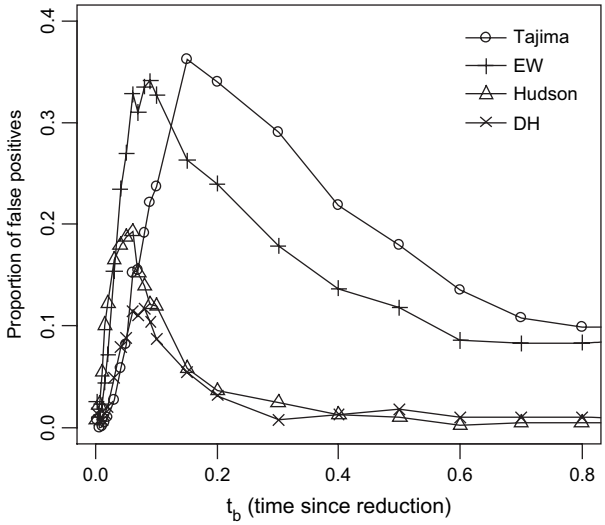

FIG. 6.—The sensitivity of the tests to population bottleneck. We assumed that the population mutation rate θ at present was 10. Going backwards in time, θ decreased exponentially to $\theta_1 = 0.5$ (i.e., $\beta = \theta_1/\theta = 0.05$). The length of this size-reducing period is denoted as t_b (measured in $4N$ generations, where N is population size at present). Intragenic recombination was considered ($\rho = \theta = 10$ at present). Sample size was 90.

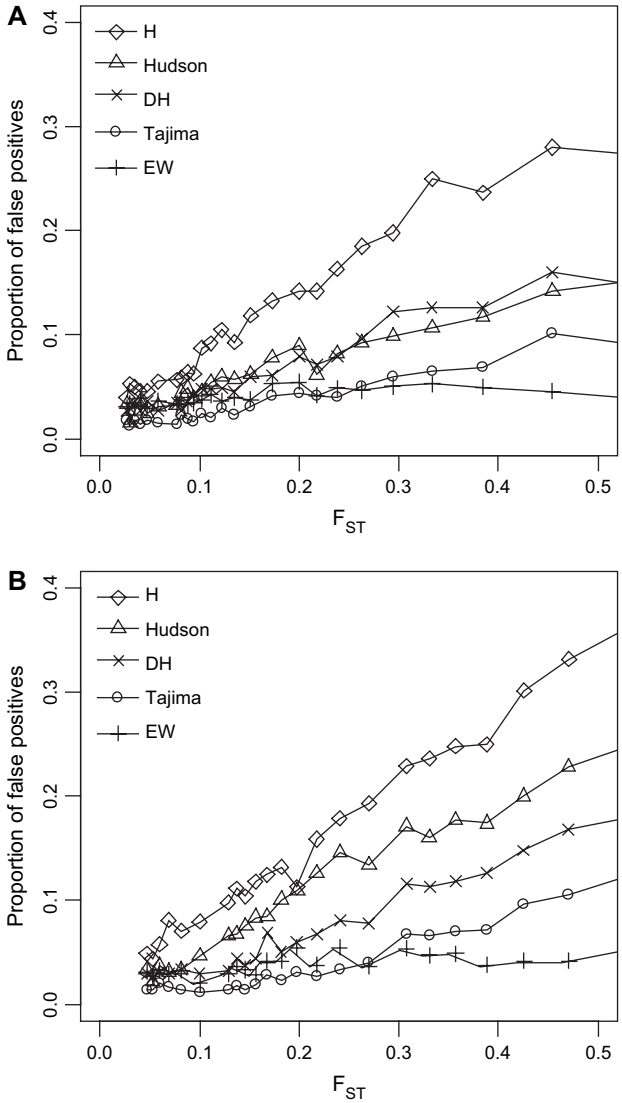

FIG. 7.—Rejection rates of the tests as functions of F_{ST} when all sequences were sampled from one deme in a subdivided population with 2 or 3 demes. We used the symmetric finite island model with 2 (*A*) or 3 (*B*) demes. The population mutation rate of each deme ($\theta_s = 4N_s u$, where N_s is the population size of a deme) was 2.5 for plot *A* (2 demes) or 1.67 for plot *B* (3 demes). That is we fixed the total population mutation rate $\theta_T = 4N_s ud = 5$. Recombination occurred at rate $\rho_s = \theta_s$ in both plots. Sample size was 90.

$d = 2$ (fig. 7*A*), but when $d = 3$ it is not as conservative, and its rejection rate is higher than 10% for $F_{ST} \geq 0.2$ (fig. 7*B*). Among the tests, Fay and Wu's *H* has the highest false-positive rate.

When sequences sampled from different demes are pooled together in the analysis, the picture is somewhat more complicated. Summarizing our simulation results, we find the following. 1) When sampling is extremely biased (i.e., the vast majority of the sequences are taken from one deme), the behaviors of the tests resemble those shown in figure 7. The tests, except for the *H* test, usually have rejection rates lower than 10%, if $F_{ST} \leq 0.2$. 2) If the sequences are sampled more or less uniformly across the demes, pooling samples does not have visible effects on the type I error rates of the tests (e.g., supplemental fig. S5, Supplementary Material online).

Joint Effects of Positive Selection and Demography

The results presented above were obtained by considering positive selection and demography separately. In reality, these 2 driving forces may have occurred concurrently. It is thus desirable to understand the effects of 2 interfering processes on the power of the tests. As an example, we assume that a selective sweep occurs in a population, which experienced a recent size expansion (fig. 8). The joint effects of sweep and growth on the EW test are quite complex (fig. 8*A*). The power curve of the EW test obtained under the interfering model (i.e., positive selection and growth together) often falls below those obtained under simpler models (i.e., only positive selection or only growth). In particular, there is a dip in power at $t_g \approx 0.03$. This is because frequency of the selected allele at this time is moderately high (about 40%), and thus, linked variants are likely to have hitchhiked to intermediate frequency (due to size expansion, most of the variants initially linked to the selected allele are at low frequency). Note that the EW test is sensitive to both population growth and positive selection. Therefore, when EW rejects neutrality, it is hard to tell which factor causes the rejection. Similar results were obtained when other simple tests such as Tajima's *D* were considered (Zeng K, unpublished data). On the other hand, the *DH* test, which is sensitive to selection but insensitive to growth, is much less affected when both factors are present (fig. 8*B*). Based on these results and results obtained by considering different interfering models (Zeng K, unpublished data), we conclude that *DH* is usually able to distinguish between positive selection and demography.

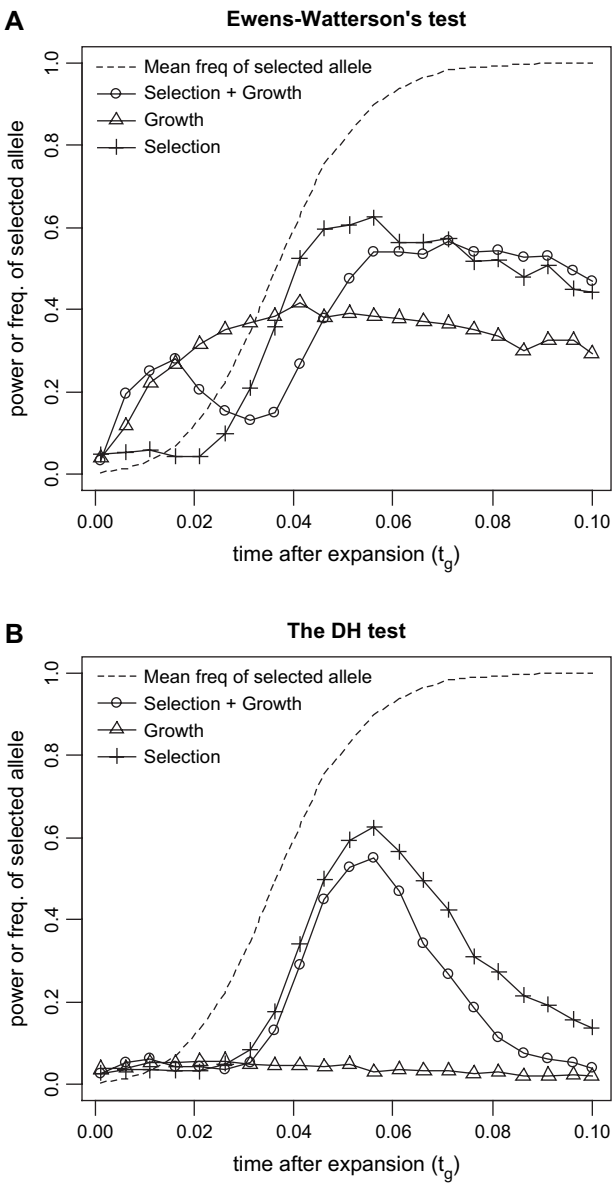

FIG. 8.—Power of the EW test (*A*) and the *DH* test (*B*) to detect positive selection when the population has experienced a recent size expansion. We assumed that the size of the population increased 10-fold instantly at time zero ($t_g = 0$; in units of $4N$ generations, where N is the population size after growth) and that immediately after the expansion, an advantageous allele appeared (i.e., $t_g = 0$). We showed the power of EW or *DH* to reject the neutral model, conditional on ultimate fixation of this advantageous allele (labeled "Selection + Growth"). For comparison, the power of the tests when only population expansion (labeled "Growth") or only positive selection (labeled "Selection") was in operation was also shown. In this figure, selective sweep was simulated by using $\alpha = 150$, $h = 0.5$. The dotted line showed the mean frequency of the selected allele t_b time units after its birth. We assumed that the site under selection was in the middle of a neutral region where $\theta = \rho = 5$ (scaled by using $4N$).

Discussion

The EW Test in the Presence of Recombination

In this study, we have showed that the EW test, albeit derived under the no-recombination neutral model, is applicable to DNA sequences with recombination. Previously, Hudson (1983) found that the conditional distributions of given values of K depend on θ and ρ in a complex way, but in general the effects of recombination do not seem to be very significant. Our results suggest that, averaging over all possible values of K, recombination does not affect the type I error rate of the EW test appreciably. Because K contains information on the local mutation and recombination rates, it may not be surprising that haplotype tests conditional on K have higher power to detect positive selection than those conditional on S. In particular, results of figure 3 *B* indicate that the EW and related tests may be useful for analyzing data from organism like Drosophila whose recombination rate is several fold higher than the mutation rate.

Site-Frequency Tests versus Haplotype-Frequency Tests

By summarizing simulation results obtained under various combinations of parameter values ($10 \leq n \leq 120$, $0.5 \leq \theta \leq 30$, $0 \leq \rho/\theta \leq 15$, and $50 \leq \alpha \leq 1000$), we find that the most important factor determining the relative performance between the *DH* test (representing site-frequency tests) and the EW test (representing haplotype-frequency tests) is the absolute level of polymorphism of the sample (i.e., values of S and n). For example, when there is only one polymorphic site in the sample whose derived allele occurs in 48 of the 50 samples, the P value of the *DH* test is 0.005, but that of the EW test is 0.344. In other words, when S and/or n are small (say, $S < 10$ and/or $n \leq 10$), the EW test may not be the most powerful test among the tests considered. In those cases, we recommend the *DH* test for 2 reasons. First, the *DH* test does not require the identification of linkage phase. Second, the *DH* test is relatively insensitive to forces other than directional selection.

The Effects of Other Factors—Dominance and Haplotype Inference

So far, we have assumed the genic selection model for the advantageous allele (i.e., $h = 0.5$). Recent study has showed that the level of dominance can have important impact on the effect of hitchhiking (Teshima and Przeworski 2006). Based on some exploratory simulation studies with $h = 0.1$ or 0.9, we find that recessive advantageous alleles are indeed more difficult to detect than dominant ones, in agreement with the results of Teshima and Przeworski. Nevertheless, the relative performance of these tests reported above remains true. Furthermore, when the neutral region is very close to the selected site, the effect of dominance tends to be weak (Zeng K, unpublished data). Our analysis also assumes that the haplotypes are known. But, in many nonmodel organisms, they can only be inferred. When the phase inference is incorrect, type I error rates of the haplotype tests may be inflated (Zeng K, unpublished data).

The Issue of Specificity

We summarize qualitatively the power (or sensitivity) of the tests to detect positive selection, background selection, and various demographic scenarios in table 3. It is

Table 3
A Qualitative Summary of the Power or Sensitivity of the Tests to Various Driving Forces

Driving Forces	EW	Hudson	Tajima	*DH*	*H*
Positive selection (before and around fixation)	+++	++	++	++	++
Background selection	++	+	++	−	−
Bottleneck	++	+	++	−	+
Subdivision	−	+	−	−	+

NOTE.—The conclusions were drawn on the basis of table 2 and figures 1–7. For easy comparison, we included properties of the *H* test of Fay and Wu in this table. +++: the power of the test frequently goes above 50%. ++: the power of the test is usually higher than 20%, but often lower than 50%. +: the power of the test in most cases falls in the interval [10%, 20%]. −: the power of the test rarely goes above 10%.

clear that all simple tests can be sensitive to more than one driving forces. In comparison, the compound test, *DH*, is affected by factors other than positive selection to a much lesser extent. This is because the 2 component tests of *DH*, Tajima's *D* and Fay and Wu's *H*, are both powerful in detecting selection but are sensitive to other driving forces in a somewhat mutually exclusive way (table 3; Zeng et al. 2006). Furthermore, figure 8*B* indicate that the *DH* test is reasonably powerful even when positive selection occurs in a nonequilibrium background. These results suggest that the compound test *DH* provides a novel way of separating positive selection from demography. On the other hand, we note that the EW test and Tajima's *D* are similar in many respects and that the EW test is very powerful in detecting positive selection. These observations suggest that, by combining the EW test and *H* or *DH*, we may be able to construct new compound tests with improved performance. We will explore this possibility in the companion paper (Zeng K, Shi S, Wu C-I, in preparation).

Supplementary Material

Supplementary table S1 and figures S1, S2, S3, S4, and S5 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Drs R. Hudson, T. Nagylaki, J. Pritchard, and 2 anonymous reviewers for their helpful 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). C.-I.W. is supported by National Institutes of Health grants and an OOCS grant from the Chinese Academy of Sciences.

Literature Cited

Barton NH. 1998. The effect of hitch-hiking on neutral genealogies. Genet Res. 72:123–133.

Biswas S, Akey JM. 2006. Genomic insights into positive selection. Trends Genet. 22:437–446.

Braverman JM, Hudson RR, Kaplan NL, Langley CH, Stephan W. 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics. 140:783–796.

Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics. 134:1289–1303.

Charlesworth D, Charlesworth B, Morgan MT. 1995. The pattern of neutral molecular variation under the background selection model. Genetics. 141:1619–1632.

Coop G, Griffiths RC. 2004. Ancestral inference on gene trees under selection. Theor Popul Biol. 66:219–232.

Depaulis F, Mousset S, Veuille M. 2003. Power of neutrality tests to detect bottlenecks and hitchhiking. J Mol Evol. 57(Suppl 1):S190–S200.

Depaulis F, Mousset S, Veuille M. 2005. Detecting selective sweeps with haplotype tests. In: Nurminsky D, editor. Selective sweep. Georgetown (DC): Landes Bioscience. p. 33–54.

Depaulis F, Veuille M. 1998. Neutrality tests based on the distribution of haplotypes under an infinite-site model. Mol Biol Evol. 15:1788–1790.

Donnelly P, Tavaré S. 1995. Coalescents and genealogical structure under neutrality. Annu Rev Genet. 29:401–421.

Ewens WJ. 1972. The sampling theory of selectively neutral alleles. Theor Popul Biol. 3:87–112.

Ewens WJ. 1973. Testing for increased mutation rate for neutral alleles. Theor Popul Biol. 4:251–258.

Ewens WJ. 2004. Mathematical population genetics. Berlin (Germany): Springer-Verlag.

Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. Genetics. 155:1405–1413.

Fay JC, Wu CI. 2003. Sequence divergence, functional constraint, and selection in protein evolution. Annu Rev Genomics Hum Genet. 4:213–235.

Fu YX. 1995. Statistical properties of segregating sites. Theor Popul Biol. 48:172–197.

Fu YX. 1996. New statistical tests of neutrality for DNA samples from a population. Genetics. 143:557–570.

Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics. 147:915–925.

Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. Genetics. 133:693–709.

Griffiths RC. 2003. The frequency spectrum of a mutation, and its age, in a general diffusion model. Theor Popul Biol. 64:241–251.

Hanchard NA, Rockett KA, Spencer C, Coop G, Pinder M, Jallow M, Kimber M, McVean G, Mott R, Kwiatkowski DP. 2006. Screening for recently selected alleles by analysis of human haplotype similarity. Am J Hum Genet. 78:153–159.

Hudson RR. 1983. Properties of a neutral allele model with intragenic recombination. Theor Popul Biol. 23:183–201.

Hudson RR. 1993. The how and why of generating gene genealogies. In: Takahata N, Clark AG, editors. Mechanisms of molecular evolution. Sunderland (MA): Sinauer Associates. p. 23–36.

Hudson RR. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics. 18:337–338.

Hudson RR, Bailey K, Skarecky D, Kwiatowski J, Ayala FJ. 1994. Evidence for positive selection in the superoxide dismutase (*Sod*) region of *Drosophila melanogaster*. Genetics. 136:1329–1340.

Hudson RR, Kaplan NL. 1994. Gene trees with background selection. In: Golding B, editor. Non-neutral evolution: theories and molecular data. London: Chapman & Hall. p. 140–153.

Innan H, Zhang K, Marjoram P, Tavaré S, Rosenberg NA. 2005. Statistical tests of the coalescent model based on the haplotype frequency distribution and the number of segregating sites. Genetics. 169:1763–1777.

Jensen JD, Kim Y, DuMont VB, Aquadro CF, Bustamante CD. 2005. Distinguishing between selective sweeps and demography using DNA polymorphism data. Genetics. 170:1401–1410.
Kaplan NL, Hudson RR, Langley CH. 1989. The "hitchhiking effect" revisited. Genetics. 123:887–899.
Karlin S, McGregor J. 1972. Addendum to a paper of W. Ewens. Theor Popul Biol. 3:113–116.
Kelly JK. 1997. A test of neutrality based on interlocus associations. Genetics. 146:1197–1206.
Kim Y, Nielsen R. 2004. Linkage disequilibrium as a signature of selective sweeps. Genetics. 167:1513–1524.
Kim Y, Stephan W. 2002. Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics. 160:765–777.
Kimura M. 1969. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. Genetics. 61:893–903.
Kimura M. 1980. Average time until fixation of a mutant allele in a finite population under continued mutation pressure: studies by analytical, numerical, and pseudo-sampling methods. Proc Natl Acad Sci USA. 77:522–526.
Maynard Smith J, Haigh J. 1974. The hitch-hiking effect of a favourable gene. Genet Res. 23:23–35.
Nielsen R. 2005. Molecular signatures of natural selection. Annu Rev Genet. 39:197–218.
Przeworski M. 2002. The signature of positive selection at randomly chosen loci. Genetics. 160:1179–1189.
Sabeti PC, Reich DE, Higgins JM, et al. (17 co-authors). 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature. 419:832–837.
Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES. 2006. Positive natural selection in the human lineage. Science. 312:1614–1620.
Simonsen KL, Churchill GA, Aquadro CF. 1995. Properties of statistical tests of neutrality for DNA polymorphism data. Genetics. 141:413–429.
Slatkin M. 1991. Inbreeding coefficients and coalescence times. Genet Res. 58:167–175.
Slatkin M. 1994. An exact test for neutrality based on the Ewens sampling distribution. Genet Res. 64:71–74.
Slatkin M. 1996. A correction to the exact test based on the Ewens sampling distribution. Genet Res. 68:259–260.
Slatkin M, Bertorelle G. 2001. The use of intraallelic variability for testing neutrality and estimating population growth rate. Genetics. 158:865–874.
Spencer CC, Coop G. 2004. SelSim: a program to simulate population genetic data with natural selection and recombination. Bioinformatics. 20:3673–3675.
Stephan W, Song YS, Langley CH. 2006. The hitchhiking effect on linkage disequilibrium between linked neutral loci. Genetics. 172:2647–2663.
Stephan W, Wiehe THE, Lenz MW. 1992. The effect of strongly selected substitutions on neutral polymorphism—analytical results based on diffusion-theory. Theor Popul Biol. 41:237–254.
Strobeck C. 1987. Average number of nucleotide differences in a sample from a single subpopulation: a test for population subdivision. Genetics. 117:149–153.
Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123:585–595.
Teshima KM, Przeworski M. 2006. Directional positive selection on an allele of arbitrary dominance. Genetics. 172:713–718.
Toomajian C, Ajioka RS, Jorde LB, Kushner JP, Kreitman M. 2003. A method for detecting recent selection in the human genome from allele age estimates. Genetics. 165:287–297.
Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. PLoS Biol. 4:e72.
Wall JD. 1999. Recombination and the power of statistical tests of neutrality. Genet Res. 74:65–79.
Wall JD. 2000. A comparison of estimators of the population recombination rate. Mol Biol Evol. 17:156–163.
Wall JD, Hudson RR. 2001. Coalescent simulations and statistical tests of neutrality. Mol Biol Evol. 18:1134–1135.
Wang ET, Kodama G, Baldi P, Moyzis RK. 2006. Global landscape of recent inferred Darwinian selection for Homo sapiens. Proc Natl Acad Sci USA. 103:135–140.
Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. Theor Popul Biol. 7:256–276.
Watterson GA. 1978. The homozygosity test of neutrality. Genetics. 88:405–417.
Wright S. 1931. Evolution in Mendelian populations. Genetics. 16:97–159.
Zeng K, Fu YX, Shi S, Wu CI. 2006. Statistical tests for detecting positive selection by utilizing high-frequency variants. Genetics. 174:1431–1439.

Jianzhi Zhang, Associate Editor

Accepted April 14, 2007