A New Test for Detecting Recent Positive Selection that is Free from the Confounding Impacts of Demography

Haipeng Li*,1

1Laboratory of Evolutionary Genomics, Department of Computational Genomics, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

*Corresponding author: E-mail: lihaipeng@picb.ac.cn.

Associate editor: Carlos Bustamante

Abstract

It has been a long-standing interest in evolutionary biology to search for the traces of recent positive Darwinian selection in organisms. However, such efforts have been severely hindered by the confounding signatures of demography. As a consequence, neutrality tests often lead to false inference of positive selection because they detect the deviation from the standard neutral model. Here, using the maximum frequency of derived mutations (MFDM) to examine the unbalancedness of the tree of a locus, I propose a statistical test that is analytically free from the confounding effects of varying population size and has a high statistical power (up to 90.5%) to detect recent positive selection. When compared with five well-known neutrality tests for detecting selection (i.e., Tajima’s D test, Fu and Li’s D test, Fay and Wu’s H test, the E test, and the joint DH test), the MFDM test is indeed the only one free from the confounding impacts of bottlenecks and size expansions. Simulations based on wide-range parameters demonstrated that the MFDM test is robust to background selection, population subdivision, and admixture (including hidden population structure). Moreover, when two high-frequency mutations are introduced, the MFDM test is robust to the misinference of derived and ancestral variants of segregating sites due to multiple hits. Finally, the sensitivity of the MFDM test in detecting balancing selection is also discussed. In summary, it is demonstrated that summary statistics based on tree topology can be used to detect selection, and this work provides a reliable method that can distinguish selection from demography even when DNA polymorphism data from only one locus is available.

Key words: positive selection, adaptation, demography, population structure, maximum frequency of derived mutations.

Introduction

Positive Darwinian selection is an important mechanism that enables species to adapt to their living environments. These selective events leave their footprint in the genomes of living organisms. Specifically, recent positive selection that has occurred within thousands of years can reduce genetic diversity (Maynard Smith and Haigh 1974; Galtier et al. 2000; Kim and Stephan 2002; Li and Stephan 2005; Nielsen et al. 2005) and alter the pattern of intraspecific polymorphism (Tajima 1989b; Fu and Li 1993; Fay and Wu 2000; Akey et al. 2002; Sabeti et al. 2002; Zeng et al. 2006; Innan and Kim 2008) of a neutral locus that is partially or completely linked to the selected locus. However, demographic effects such as varying population size could create the same patterns in DNA sequence variation. This makes inferred selective events questionable, especially when there is no follow-up functional analysis to confirm the results (MacCallum and Hill 2006). A general approach to distinguishing selection from demography is to use genome-wide polymorphism data (Nielsen et al. 2005; Kelley et al. 2006; Li and Stephan 2006; Voight et al. 2006; Tang et al. 2007) because demographic factors affect the genome-wide pattern of polymorphism while selective events act locally on specific regions of the genome. However, this approach has been questioned in that if selection is much more common than previously assumed, it will affect the entire genome and its genome-wide polymorphism data will not reflect demography alone (Hahn 2008).

Moreover, despite undisputable progress in sequencing technology, functional analysis, and genome-wide approach can only be applied to a few model species because they require considerable effort and expense. Therefore, in order to detect recent positive selection in a wide variety of natural populations, especially in millions of nonmodel species, it is essential to seek a signature of selection that is independent of varying population size and develop a statistical method based on single-locus DNA polymorphism data that is free from the confounding effects of demography.

Due to the effect of hitchhiking, one lineage of a neutral locus partially linked to a selected locus may escape from the selective sweep through recombination and it will not coalesce with any other lineages before the most recent common ancestor (Kaplan et al. 1989; Fay and Wu 2000; Kim and Nielsen 2004). This brings about a particularly long branch linked to the root of the tree, which is then called an unbalanced tree. The unbalanced tree may result in an excess of derived mutations at high frequency (compared with mutations of rare and intermediate frequency) that has been seen as the signature of positive selection (Fay and Wu 2000). However, bottlenecks in population size could also cause the excess of derived mutations at high frequency in the sample; thus, Fay and Wu’s H test may...
lead to false inference of positive selection (Jensen et al.
2005) (supplementary fig. S1A, Supplementary Material online).
Therefore, to develop a statistical test that is analyt-
ically free from the confounding effects of demography, a novel approach is needed.

It is interesting to note that, according to coalescent the-
ory (Hudson 1990), varying population size does not affect
tree topology in a single Wright–Fisher population (supple-
mentary fig. S1B, Supplementary Material online). Thus, the
probability of an unbalanced tree is independent of bottle-
necks and size expansions. In particular, the probability of an
unbalanced tree is usually very small under neutrality, as the
number of possible trees grows very quickly. However, under
a hitchhiking model, this probability will increase substan-
tially when the neutral locus is partially linked to the selected
locus (supplementary fig. S1B, Supplementary Material online).
If an unbalanced tree is inferred at a locus, it indicates
that positive selection may have occurred recently on a
nearby locus; thus, adaptive evolution may be distinguished
drom demography by examining the tree topology of a neutral
locus partially linked to the selected locus.

Intraspecific DNA polymorphisms are typically analyzed
using the infinite-site model (Kimura 1969), which assumes
that whenever a mutant appears, it occurs at a previously
homoallelic site. Under the infinite-site model, the max-
um frequency of derived mutations (MFDM) in the sam-
ple can be used to detect the presence of an unbalanced
tree. Following this logic, I have developed a statistical test
(named thereafter as the MFDM test) to detect recent pos-
itive selection in a natural population. I demonstrated an-
alytically and empirically that the MFDM test is free from
the confounding effects of varying population size, includ-
ing bottlenecks and population size expansions. Although
population subdivision and admixture can also lead to an
unbalanced tree due to migration, these migration events
can be easily detected using a phylogenetic method and a
simple sampling scheme (fig. 1). After excluding unbal-
anced trees due to migration, my simulations suggested
that the MFDM test is extremely robust to population
subdivision and admixture, including hidden population
structure. Moreover, simply by using two high-frequency
mutations, the MFDM test becomes robust to the misin-
ference of derived and ancestral variants of segregating sites
due to multiple hits. Simulations also showed that the
MFDM test is robust to background selection and that
its power to detect positive selection is high (up to 90.5%).

Thus, in this study, I demonstrated that summary statistics
related to tree topology is useful to detect recent positive
selection, and the MFDM test can reliably distinguish
selection from varying population size and population
structure even if only a single-locus DNA polymorphism
data are available.

Materials and Methods

Coalescent Simulations

Simulations under neutrality and under the hitchhiking
model of selective sweeps were performed according to
the procedures described previously (Hudson 1990; Kim
and Stephan 2002; Li and Stephan 2006). Positive selection
was assumed to be directional with codominant alleles.

To simulate positive selection on an existing neutral
mutation (Hermisson and Pennings 2005), a previously
described approach was followed (Spencer and Coop 2004).
The trajectory of the neutral mutation backward in time
was simulated from the Wright–Fisher model conditional
on the eventual loss of the mutation from the population.
The structured coalescent before the selective phase was condi-
tioned on the trajectory; the structured coalescent
during the selective phase was performed according to
the procedures described previously (Kim and Stephan
2002; Li and Stephan 2006).

The coalescent simulations under background selection
followed a slightly revised procedure from Fu (1997). I con-
sidered a model of fitness in which a gamete carrying
j dele-
terious mutations has fitness \( w_j = (1 - sh)^j \), where s and h
are the selection and dominance coefficients, respectively.
I assumed that the number of new mutations per individual
per generation in a pair of partially linked genes was a Pois-
son variable with mean \( U \). Each gene spanned 5 kb
and contained 500 equally spaced selected sites.

The coalescent simulations under two-allele balancing
selection followed a procedure modified from one de-
scribed by Hein et al. (2005). I assumed that the frequencies
of two alleles at the selected locus were kept constant over
time in the population by strong balancing selection. Dur-
ing the first phase, the population evolved neutrally. During
the second (recent) phase, the population was under

FIG. 1. Phylogenetic trees for a sample (\( n = 5 \)) collected from one deme (subpopulation) and MDs from other demes (usually, one from each)
under the symmetric island model. For simplicity, the number of demes is two. The unbalanced tree for the sample is presented as solid black
lines and the lineage of the MD is shown as dashed lines. The outgroup used to root the tree is not shown. Derived mutations with frequency
(\( n - 1 \)) in the sample are marked as solid circles. (A) Case with low migration rate. (B) Case with mid-range migration rate. (C) Case with high
migration rate.
balancing selection. The length of the second phase was 10, with the time in units of 2N generations.

The Statistical Test
Assume that the sample consists of n sequences from a random mating population. Let us consider the two basal branches that originate from the root node of the bifurcating genealogy of the sample (i.e., the most recent common ancestor of the sample). Denote the size of a branch in the sample genealogy as the number of sequences in the sample that are descendants of the branch (Fu 1995). Thus, let \( \Psi_1 \) and \( \Psi_2 \) be the size of left- and right-basal branches, respectively, where \( \Psi_1 + \Psi_2 = n \). The genealogical tree is unbalanced if \( \Psi_1 \) is small and \( \Psi_2 \) is large (or vice versa). Furthermore, let \( \Psi = \max\{\Psi_1, \Psi_2\} \). Considering \( \Psi \) as an integer random variable, its probability function (Tajima 1983) is

\[
P(\Psi = x) = \begin{cases} 
2/(n - 1) & \text{if } n/2 < x \leq (n - 1), \\
1/(n - 1) & \text{if } x = n/2.
\end{cases}
\]

(1)

Thus, the probability of having a more or equally unbalanced tree under the null hypothesis is

\[
P(\Psi \geq x) = \begin{cases} 
2(n - x)/(n - 1) & \text{if } n/2 < x \leq (n - 1), \\
1 & \text{if } x = n/2.
\end{cases}
\]

(2)

Although \( \Psi \) can be estimated directly from a rooted tree rebuilt by the unweighted pair group method with arithmetic mean (UPGMA) (Sneath and Sokal 1973) or the neighbor joining (Saitou and Nei 1987), another simple approach is used here. Under the infinite-site model (Kimura 1969), the existence of a derived mutation with absolute frequency \( \xi \) means that there is a branch with size \( \xi \) in the sample genealogy. Derived mutations in the sample are inferred using an outgroup sequence. Here, we do not consider mutations of rare and intermediate frequency \((1 \leq \xi < n/2)\). When \( \xi (\xi \geq n/2) \) increases, \( P(\Psi \geq \xi) \) decreases. Hence, a high-frequency derived mutation provides information about the sizes of basal branches. Therefore, the probability of \( \Psi \geq \xi_{\text{max obs}} \) defined in equation (2) is used to detect recent positive selection in this study, where \( \xi_{\text{max obs}} \) denotes the observed MFDM in the sample. If \( P(\Psi \geq \xi_{\text{max obs}}) \leq \alpha \), the neutral hypothesis is rejected at the significance level \( \alpha \). It is a one-tailed test. Obviously, when \( \xi_{\text{max obs}} = n - 1 \), \( P(\Psi \geq \xi_{\text{max obs}}) \) is minimized \((= 2/(n - 1))\). Thus, to conduct a MFDM test that can reject the neutral hypothesis at the 0.05 significance level, the minimum sample size required by the test is 41 chromosomes (or 21 diploid individuals). For the 0.01 significance level, the minimum sample size required is 201 chromosomes (or 101 diploid individuals).

Given \( P(\Psi \geq \xi_{\text{max obs}}) \leq \alpha \), where \( \alpha \) is the significance level of the MFDM test, I demonstrate here that the false-positive rate of the MFDM test, that is, \( P(\xi_{\text{max}} \geq \xi_{\text{max obs}}) \), remains less than \( \alpha \) under arbitrary demographic scenarios.

In a random mating population, the following holds true:

\[
P(\xi_{\text{max}} \geq \xi_{\text{max obs}}) = P(\xi_{\text{max}} \geq \xi_{\text{max obs}}, \Psi \geq \xi_{\text{max obs}}) = P(\xi_{\text{max}} \geq \xi_{\text{max obs}} | \Psi \geq \xi_{\text{max obs}})P(\Psi \geq \xi_{\text{max obs}}).
\]

(3)

Thus, \( P(\xi_{\text{max}} \geq \xi_{\text{max obs}}) < \alpha \) because \( P(\xi_{\text{max}} \geq \xi_{\text{max obs}}, \Psi \geq \xi_{\text{max obs}}) < 1 \). Therefore, the MFDM test is completely free from the confounding effects of varying population size.

In the discussion above, I assume that there is no recombination within the locus. When recombination occurs within the locus, the probability of having an unbalanced tree within the locus increases (supplementary fig. S2, Supplementary Material online); thus, the false-positive rate of the MFDM test may increase. When the recombination parameter \( (p = 4N_{r}) \) is estimated for the locus (Hudson 2001; McVean et al. 2002), the probability of having an unbalanced tree within the locus can always be used as the maximum false-positive rate of the MFDM test as discussed previously. However, this approach requires a substantial amount of simulations, so I have taken another simple approach here.

In practice, a single DNA fragment (i.e., a locus) may be of short length and only contain a few recombination events. If this is the case, the increased false-positive rate of the MFDM test due to recombination can be solved by using the Bonferroni correction. The general version of the MFDM test consists of three steps: 1) estimate the minimum number of recombination events \( (R_{t}) \) implied by the sequence data using the four-gamete test (Hudson and Kaplan 1985), and the sequenced region is broken into \( (R_{t} + 1) \) short segments accordingly; 2) assume that each segment is evolving independently (which makes the test very conservative) and conduct the test described above for each segment; 3) reject the neutral hypothesis when \( P(\Psi \geq \xi_{\text{max obs}} \text{, } i) \leq \alpha/(R_{t} + 1) \) for the \( i \)-th segment, where \( 1 \leq i \leq R_{t} + 1 \), and \( \xi_{\text{max obs}, i} \) is the \( i \)-th segment. I shall note that when there is no recombination within the locus, \( R_{t} = 0 \). In the following, the MFDM test always refers to this general version relying on the estimate of \( R_{t} \).

Moreover, the derived and ancestral variants of segregating sites in quickly evolving loci may be interpreted incorrectly due to the invalidation of the infinite-site model (i.e., multiple hits). The probability of multiple hits on the sites of rare mutations follows a binomial distribution and is given in the supplementary material text, Supplementary Material online. If this probability is high, two high-frequency mutations (i.e., two informative sites) can be used to reliably infer the sizes of the basal branches. Let \( \{\xi_{1}, \xi_{2}, \ldots, \xi_{k}\} \) be an ordered vector of the absolute frequency of derived mutations in the sample, where \( \xi_{ij} \geq \xi_{ij} \ (i > j) \), \( \xi_{i} = \xi_{\text{max obs}} \) and \( k \) is the number of segregating sites in the sample. Next, randomly locate a mutation of size \( \xi_{i} \) in the sample, and let \( S_{i} \) be the set of sequences that carry the derived allele.
Starting from this mutation, locate its neighboring mutation of size \( \zeta_2 \) in the sample, and let \( S_2 \) be the set of sequences that carry the derived allele of the second mutation. If \( S_1 \supseteq S_2, P(\Psi \geq \zeta_2) \) defined in equation (2) is used to determine whether the neutral hypothesis should be rejected. Other cases in which \((S_1 \cap S_2 \neq S_2)\) are simply discarded.

Additionally, the symmetric island model can result in unbalanced trees (Przeworski 2002); thus, the population structure may increase the false-positive rate of the MFDM test. Moreover, it may be difficult to know whether the sample is actually collected from one deme (subpopulation) or multiple demes because knowledge about the population structure of the target species is often lacking. To solve these problems, I propose two sampling schemes: 1) to control the false-positive rate of the MFDM test in the case of hidden population structure, a requirement of the minimum number of samples collected from each sampling location should be satisfied. This issue will be discussed further in the following sections and 2) when sampling locations do not cover the whole species range, a migrant-detector approach can be used to identify unbalanced trees due to migration, lowering the false-positive rate of the MFDM test. The logic of migrant-detector approach (fig. 1) is explained as follows.

When \( 4Nm < 0.1 \) or \( >1 \), where \( N \) is the island population size and \( m \) the fraction of each subpopulation made up of migrants in each generation, the island model may not increase the probability of an unbalanced tree for a sample collected from one deme because all lineages will coalesce within the same deme (fig. 1A) or demes can be treated as a large random mating population (fig. 1C). However, when \( 0.1 \leq 4Nm \leq 1 \), a lineage may migrate from one deme to another and the lineage may not coalesce with any others before the most recent common ancestor. Thus, it leads to an unbalanced tree (solid black lines in fig. 1B) and it may increase the false-positive rate of the MFDM test. Interestingly, the migrant-detector approach can be used to identify these unbalanced trees caused by migration. When a sample is collected from certain areas within the species range, extra chromosomes should also be collected from other areas, referred to as migrant detectors (MDs). The migrant lineage would first coalesce with one of the lineages of the MDs before coalescing with any others (fig. 1B). Therefore, in practice, if the neutral hypothesis is rejected for the data set of a sample by the MFDM test, the information of the basal branches in the sample genealogy is known. Based on nucleotide differences, a UPGMA tree (Sneath and Sokal 1973) is then reconstructed for the sample and the MDs. The inferred UPGMA tree is rooted by an outgroup. If one of the migrant-detector lineages is linked to the minor basal branches, the unbalanced tree of the sample should be due to migration instead of selection (fig. 1B). I shall note that, with little modification, this approach can be applied to the cases of the island model with multiple demes, the split population model and the admixture model.

In practice, multiple MDs may be sampled from the same deme, which will apparently make the MFDM test more conservative. On the other hand, we may have an incomplete MD-sampling problem (i.e., MDs from certain demes are not available), which may increase the false-positive rate of the MFDM test. This situation will be investigated further in the following sections.

**Results**

In the following sections, I first examined the false-positive rates of the MFDM test together with five other tests based on the frequency spectrum, that is, Tajima’s D test (Tajima 1989b), Fu and Li’s D test (Fu and Li 1993), Fay and Wu’s H test (Fay and Wu 2000), the E test (Zeng et al. 2006), and the joint DH test (Zeng et al. 2006). In this study, I considered the cases of both nonrecombination and recombination loci. Then, I studied the power to detect recent positive selection by applying these tests to a simulated neutral locus linked to a selected locus. In this article, the significance level of all the tests is 0.05.

**Analysis of False-Positive Rates**

I first examined the population size expansion model (supplementary fig. S3A, Supplementary Material online), including recent and old population size expansions. The maximum false-positive rate of Fu and Li’s D test, Tajima’s D test, and the E test under the examined expansion model is about 50–80% (fig. 2A and supplementary fig. S4A, Supplementary Material online). Thus, these three tests are very sensitive to population size expansions, which is consistent with previous findings (Tajima 1989a; Fu and Li 1993; Zeng et al. 2006). In contrast to these tests but similar to Fay and Wu’s H test and the joint DH test, the MFDM test is very robust with respect to these expansion scenarios. The false-positive rate of the MFDM test remains below the significance level of the test (<0.05).

Then, I examined the population size bottleneck model (supplementary fig. S3B, Supplementary Material online), including recent and old bottlenecks. Notably, the false-positive rate of the MFDM test under the bottleneck model remains below the significance level of the test (<0.05), whereas no other tests are robust with respect to these bottleneck scenarios (fig. 2B and supplementary fig. S4B, Supplementary Material online). In the examined cases, the maximum false-positive rates of Fay and Wu’s H test, the joint DH test, Fu and Li’s D test, Tajima’s D test, and the E test are 0.232, 0.140, 0.287, 0.388, and 0.352, respectively. These values are much higher than the significance level of the tests, consistent with the previous finding that Fay and Wu’s H test is sensitive to bottlenecks (Jensen et al. 2005). Additionally, as expected, recombination makes the MFDM test more conservative (fig. 2B and supplementary fig. S4B, Supplementary Material online) as the Bonferroni correction is implemented in the MFDM test. I also investigated the false-positive rate of the MFDM test and Fay and Wu’s H test under an inferred bottleneck scenario for the European Drosophila population (Li and Stephan...
Other parameters for these two tests were the same as those shown in figure 2B, and the false-positive rates are 0.030 and 0.146, respectively. Overall, these simulated results for the MFDM test agree with the analytical results shown in equation (3), and the MFDM test is indeed the only one that is free from the confounding effects of population size expansion and bottleneck.

I further explored the effects of population subdivision under a symmetric island model with different numbers of demes. The maximum false-positive rates of Fay and Wu’s H test, Tajima’s D test, and the joint DH test are 0.514, 0.250, and 0.292, respectively, much higher than 0.05 (fig. 3A and supplementary fig. S5, Supplementary Material online). When the information from MDs is not available, the MFDM test is generally more conservative than Fay and Wu’s H test, Tajima’s D test, and the joint DH test, but its false-positive rate can be higher than 0.05. However, the false-positive rate of the MFDM test remains below the significance level of the test (0.05) if the migrant-detector approach is applied. This suggests that the MFDM test is free from the confounding effects of population subdivision if the simple sampling scheme (i.e., MDs) is applied.

However, it is expected that an incomplete MD-sampling problem may occur in practice due to sampling difficulties and hidden population structure in some species distributed areas. I examined this situation by assuming that MDs are only available from several randomly chosen demes (fig. 3B). Generally, the more MDs that are available, the lower the false-positive rate of the MFDM test is. Surprisingly, the false-positive rate of the MFDM test remains lower than 0.05 even if MDs are only available from about one-third of the demes. Therefore, the MFDM test is robust to the incomplete MD-sampling problem.

Moreover, the sample may be collected from multiple demes or even from the whole species range because of a lack of knowledge of population structure or because researchers want to identify global selective events. Unbalanced trees may result from unequal sampling sizes between different demes (Wakeley and Aliacar 2001; Przeworski 2002) and thus may lead to false positives. To investigate this situation, the false-positive rates of Fay and Wu’s H test and the MFDM test are presented (fig. 4) when samples are drawn unequally from different demes in the island model. I find that unequal sampling could have a much more severe impact on Fay and Wu’s H test than what was shown previously (Przeworski 2002) when I explore a wide range of parameters. For example, when two subpopulations are partially isolated (4Nm = 0.001) and 80 and 20 chromosomes are sampled from these two subpopulations, the false-positive rate of Fay and Wu’s H test is 0.976. However, the false-positive rate of the MFDM test is only 0.0002 in this situation.

![Fig. 2. False-positive rate of the tests estimated from 10^5 simulated neutral data under exponential population growth (A) and population bottleneck (B). I assumed that there is no recombination within the locus and n = 100. For exponential population growth, I assumed N_0/N_1, where N_0 and N_1 are the current and ancestral effective population size, respectively, and \( \theta = 4N_0\mu = 5 \). For population bottlenecks, I assumed N_0/N_1 = 100 and \( t_1 = 0.01 \), where N_1 is the effective population size during the bottleneck, \( t_1 \) the duration of the bottleneck, and \( \theta = 4N_0\mu = 5 \).](https://doi.org/10.1093/molbev/msq211)

![Fig. 3. False-positive rate of the tests estimated from 10^5 simulated neutral data under the finite island model. \( \theta = 4N_0\mu = 5 \), where N is the effective population size for each deme. n = 100 and the sample is collected from one deme. Some MDs may also be collected from other demes (one from each). I assumed that there is no recombination within the locus. (A) The number of demes is two. (B) The number of demes is ten, and incomplete (the number of MDs <9) and complete (the number of MDs = 9) MD-sampling cases are considered here.](https://doi.org/10.1093/molbev/msq211)
Generally, the MFDM test is more robust with respect to the hidden structured population than Fay and Wu’s $H$ test (fig. 4). Furthermore, from equation (2) and figure 4, it is known that the hidden structured population will not affect the MFDM test if the sample size from each sampling location is larger than $\alpha(n - 1)/2$. Hence, this simple sampling scheme can be used to efficiently avoid the false inference of positive selection due to hidden structured population.

As the island model may be unrealistic for many species, such as humans and Drosophila melanogaster, I further examined a population subdivision model (supplementary fig. S3C, Supplementary Material online). I considered an ancestral population with population size $N$ that has split into two subpopulations, each of which has size $N$. The split time is in units of $2N$ generations. When the information of the MD is not available, the false-positive rate of the MFDM test is a function of split time and migration rate (supplementary fig. S6, Supplementary Material online). In such a situation, the MFDM test generally remains robust to population subdivision if the split time is less than 1.0. When the split event is very old (the split time = 4.0), the MFDM test is slightly more conservative than Fay and Wu’s $H$ test and the joint DH test. However, when the migrant-detector approach is applied, the false-positive rate of the MFDM test always remains below the significance level of the test ($<0.05$) (supplementary fig. S6, Supplementary Material online).

I also considered a population admixture model (supplementary fig. S7A, Supplementary Material online) because both the island model and the population subdivision model assume constant migration rate among subpopulations. In the population admixture model, migration can only occur during a short time period ($t_i$ in fig. S7A, Supplementary Material online). I found that Fay and Wu’s $H$ test is very sensitive to population admixture; the maximum false-positive rate of Fay and Wu’s $H$ test is 0.467 in the examined cases (fig. S7B, Supplementary Material online). Notably, the false-positive rate of the MFDM test remains below the significance level of the test ($<0.05$) when the migrant-detector approach was applied.

Next, I checked the dependence of the false-positive rate on background selection (Charlesworth et al. 1993). I considered a partially linked gene pair that, as a pair, is subject to purifying selection, with a neutral DNA segment placed in between. It is known that background selection is not likely to cause an excess of high-frequency derived mutations (Fu 1997). As expected, the false-positive rate of the MFDM test under background selection remains below the significance level of the test ($<0.05$) (supplementary fig. S8, Supplementary Material online). In the case of strong background selection variability is reduced; therefore, the false-positive rate of the MFDM test is also reduced (supplementary fig. S8B, Supplementary Material online). The sensitivity of the other tests under background selection is consistent with the previous findings (Zeng et al. 2006).

The sensitivity of the tests to balancing selection was also examined. I assumed that the selected locus had two alleles, $A$ and $a$ with frequencies $f(A)$ and $f(a)$ and $f(A) \geq f(a)$. Generally, all tests are robust to balancing
Detecting Recent Positive Selection · doi:10.1093/molbev/msq211

FIG. 5. False-positive rate of the MFDM test due to falsely identifying the derived and ancestral variants of segregating sites. (A) The possible matrix representing three segregating sites of a sample ($n = 9$). The ancestral state is coded with a zero, and the mutant, or the derived state, is indicated with a one. Open circles represent that the ancestral and derived status of sites has been correctly identified, and solid circles represent that their status has been falsely interpreted. As a consequence of the false inferences, two high-frequency derived mutations are carried by two different sets of chromosomes. (B) False-positive rate of the MFDM test estimated from $10^5$ simulated neutral data when one or two informative sites are used. I assumed that there is no recombination within the locus. $n = 100$ and $\theta = 20$ for the quickly evolving locus.

selection if the $a$ allele is very rare ($f(a) = 0.001$) (supplementary fig. S9A, Supplementary Material online). When $f(a) = 0.01$, most of the examined tests, including Fay and Wu’s $H$ test, the joint $DH$ test, and the MFDM test, are sensitive to balancing selection (supplementary fig. S9B, Supplementary Material online). In that case, balancing selection and positive selection may be distinguished by examining the genetic diversity of neighboring loci because genetic diversity at the selected site is increased under balancing selection and reduced under positive selection (Stephan et al. 1992). When $f(a) = 0.1$, Fay and Wu’s $H$ test remains very sensitive to balancing selection, whereas the MFDM test is very robust (supplementary fig. S9C, Supplementary Material online). When $f(a)$ increases to 0.5, Fay and Wu’s $H$ test also becomes robust (supplementary fig. S9D, Supplementary Material online).

Finally, I examined the sensitivity of the MFDM test to falsely identify the derived and ancestral variants of segregating sites at quickly evolving loci (fig. 5), which could be due to the context-dependent mutation rate (Hernandez et al. 2007). I simulated data sets in the presence of misinference by switching the frequency of the derived and ancestral variants, with a probability given as the misinference parameter for each segregating site (Baudry and Depaulis 2003). Indeed, the false-positive rate of the MFDM test can increase when only one informative site (i.e., one high-frequency mutation) is used (fig. 5B). However, when two informative sites are used to infer the unbalancedness of the tree, the MFDM test always remains robust to the false inference of the derived and ancestral variants of polymorphic sites (even if the ancestral status of 10% segregating sites were mis-labeled).

Power Analysis
My simulations suggested that the power of the MFDM test can be slightly higher than that of Fay and Wu’s $H$ test (81.5% vs. 78.50%) (fig. 6A). When the number of sampled diploid individuals is larger than 50 (i.e., $n > 100$), the power difference between the MFDM test and Fay and Wu’s $H$ test becomes small. Generally, the power of the MFDM test remains high enough to detect recent selection although it is relatively lower than that of Tajima’s $D$ test and the joint $DH$ test (but higher than that of Fu and Li’s $D$ test, and the $E$ test). Moreover, my simulations show that the $E$ test is not sensitive to recent selection, which agrees with previous findings (Zeng et al. 2006) but the power of the $E$ test does increase when selective events are relatively old (results not shown).

I also studied the power of the tests in a varying size population. Generally, the MFDM test has high power to detect recent selective events in a bottlenecked population (fig. 6B). In an expanding population, Tajima’s $D$ test and Fu and Li’s $D$ test can have the highest power to detect selection (fig. 6C) because the size expansion also causes an excess of rare mutation; thus, the size expansion itself increases the probability that the standard neutral model is rejected by these two tests. Moreover, the power of the MFDM test (ranging from 47% to 89%) is generally higher than that of Fay and Wu’s $H$ test but lower than that of the joint $DH$ test.

Interestingly, when recombination occurs within the locus, the power of the tests (supplementary fig. S10, Supplementary Material online) remains very similar to that when there is no recombination within the locus (fig. 6). This suggests that recombination may not lower the power of the MFDM test because selection reduces the population recombination parameter ($4N_e\theta$). For example, when $s = 0.005$, where $s$ is the selection coefficient, the population recombination parameter is reduced from 40 to 0 (the median).

I then examined the power of the MFDM test in detail. Because the frequency spectrum of selection coefficients is estimated for two wild Drosophila populations (Li and Stephan 2006), I chose the values of selection coefficients accordingly. Even if the selection strength is very weak ($s = 0.0005$) the probability to detect selection is still 32.8% (supplementary fig. S11A, Supplementary Material online). Generally, the stronger the selection, the higher the power of the MFDM test. When $s = 0.05$, the probability to detect these strong selective events is 81.5%.

The MFDM test can only detect recent selective sweeps (the time back to the completion of the selective
to increase the power while the effect of recombination may remain small.

Because positive selection could occur on an existing neutral mutation (Falconer and Mackay 1996; Hermisson and Pennings 2005) instead of a new beneficial mutation, the power of the MFDM test was examined in this situation (supplementary fig. S11D, Supplementary Material online). When the frequency of standing genetic variation is low in the population, we have reasonable power to detect the selection. However, when the frequency of standing genetic variation is relatively high in the population, we have less chance to detect it by the MFDM test. This problem may be partially overcome by sequencing both the flanking regions of the standing genetic variation.

Last, I examined the power of the MFDM test to detect selection based on the polymorphism data of a fast-evolving locus. I assumed that \( \theta = 50 \) and \( n = 500 \) and other parameters were the same as in figure 6. When there are no misinferences of derived states, the power of the MFDM test (the form using two informative sites) is still high at 81.3%. However, if the probability of misinference of derived state is 0.01 or 0.1, the power of the MFDM test (the form using two informative sites) is reduced to 75.3% and 42.4%, respectively. The reduced power here is expected because the signature of recent positive selection is hidden in the noise of the misinferences of derived states.

**Discussion**

The issue of false acceptance of the selection hypothesis cannot be overemphasized because we still have little knowledge about demography and the molecular mechanism of recent adaptations for most species. Therefore, the priority of this study is to provide a highly specific statistical method that is suitable for detecting recent positive selection in species that may have varying population size and/or population structure. To do so, I proposed a test based on the MFDM that does not require an estimate of \( q \) or any explicit knowledge of demography. High specificity can be achieved because the unbalanceness of the tree is independent of varying population size, and two simple but efficient sampling schemes can eliminate the confounding impact of population structure. These principles differ from that of the commonly used neutrality tests, such as Tajima’s \( D \) test (Tajima 1989b), Fu and Li’s \( D \) test (Fu and Li 1993), and Fay and Wu’s \( H \) test (Fay and Wu 2000), the \( E \) test (Zeng et al. 2006), and the joint DH test (Zeng et al. 2006). These existing neutrality tests are based on the mutation frequency spectrum, and a recent work (Achaz 2009) suggests that Tajima’s \( D \) test, Fu and Li’s \( D \) test, and Fay and Wu’s \( H \) test can be encompassed by a general framework.

In this study, I demonstrate that the MFDM test is not only analytically and empirically free from the confounding effects of varying population size, including bottlenecks and expansions, but also very robust with respect to population subdivision and admixture. Under the examined neutral cases, the maximum false-positive rate of the

---

**Fig. 6.** Power of the tests from \( 10^5 \) simulated data under positive selection. \( N_0 = 10^6, s = 0.005, \) and \( \tau = 0.0001, \) unless noted otherwise, and \( \tau \) is the time back to the completion of the selective substitution (in units of \( 2N_e \) generations). The locus is 1 kb away from the selected site, and I assumed that there is no recombination within the locus. \( n = 100 \) and \( \theta = 4N_e\mu = 5. \) (A) Power of the six tests under the constant size model. (b) Power of the six tests under bottlenecks, where \( N_0/N_1 = 10, \) \( t_0 = 0.1, \) and \( t_1 = 0.01. \) (C) Power of the six tests under a size expansion scenario, where \( N_0/N_1 = 10, \) \( t_0 = 0.1, \) and \( 4N_e\mu = 5. \)
The unbalanceness of the tree can be assessed with certainty because the existence of a derived mutation with absolute frequency $\zeta$ always implicates the existence of a branch with size $\zeta$ in the sample genealogy under the infinite-site model (Kimura 1969). However, the infinite model could be invalidated occasionally due to multiple hits. For example, due to the presence of multiple hits, the derived and ancestral variants of a segregating site at a quickly evolving locus may be falsely identified; thus, a rare mutation with frequency $\epsilon_{\text{rare}}$ may be falsely referred to as a high-frequency mutation with frequency $(n - \epsilon_{\text{rare}})$. It has been found that those misoriented sites may influence Fay and Wu’s $H$ test (Baudry and Depaulis 2003), but this problem may be solved using multiple outgroups to identify such homoplasy (supplementary fig. S12, Supplementary Material online) (Enard et al. 2002; Hernandez et al. 2007; Sjödin et al. 2008).

Alternatively, it is easy to obtain the probability that the derived and ancestral variants of one or more segregating sites are misoriented due to multiple hits, given that the misoriented sites may affect the MFDM test (eq. S1 in supplementary material, Supplementary Material online). The summation of this probability and the probability of an unbalanced tree will be the maximum false-positive rate of the MFDM test. Usually, the probability of multiple hits should be small if the region is conserved, the sequenced region is not very long and the outgroup is one of the closely related species.

When the probability of multiple hits is high, the derived and ancestral variants may be misoriented in which case I recommend using two high-frequency mutations (see Materials and Methods) in the MFDM test. Because the majority of derived mutations in the sample are rare mutations, it is likely that rare mutation is misoriented as a high-frequency mutation (Baudry and Depaulis 2003; Hernandez et al. 2007). These high-frequency mutations due to misorientation are often carried by different sets of chromosomes because rare mutations tend to be scattered among chromosomes (fig. 5A). However, if multiple high-frequency mutations occur on an unbalanced tree, it is likely that they are carried by the same set of chromosomes. I should note that this two informative site approach may not be limited to quickly evolving genes because multiple high-frequency mutations can also be found even if the gene is very conserved as in the example of FOXP2 (Enard et al. 2002). Therefore, recent positive selection can be distinguished from the misinference of derived and ancestral variants (supplementary fig. S8, Supplementary Material online).

Moreover, several multiple-locus or genome-wide approaches are proposed to detect positive selection by identifying selective sweeps (in particular, valleys of reduced polymorphism) (Kim and Stephan 2002; Li and Stephan 2005; Nielsen et al. 2005), long-range haplotypes in the genome (Sabeti et al. 2007; Tang et al. 2007), or Fst-based approaches (Akey et al. 2002; Beaumont and Balding 2004) (see a review by Pavlidis et al. 2008). By analyzing DNA polymorphism data from multiple loci, the effect of demography on observed polymorphic pattern

**Fig. 7.** Maximum false-positive rate of the tests under the neutral cases that were examined. Parameters are given in figures 2–3 and supplementary figures S4–S7, Supplementary Material online. The significance level of the tests (0.05) is also shown.

MFDM test is strikingly low at only 0.039 (fig. 7). However, under the same circumstances, the maximum false-positive rate of the existing five tests is much higher than the significance level of the tests. Thus, the MFDM test is indeed the only one that can reliably detect recent positive selection under wide-range neutral cases.

The low false-positive rate of the MFDM test under population subdivision and admixture has been achieved because migrants resulting in an unbalanced tree are identified by a phylogenetic method and a simple sampling scheme (i.e., migrant-detector approach) (fig. 1). Surprisingly, the migrant-detector approach works well for different values of migration rate in different population structure models, and the test is also robust to the incomplete MD-sampling problem; thus, the application of the migrant-detector approach may not be limited by the knowledge of population structure of species. Moreover, the UPGMA method (Sneath and Sokal 1973) is used here to determine whether one of the migrant-detector lineages is linked to the minor basal branches instead of whether the genealogy is unbalanced. Thus, the long-branch attraction phenomenon (Felsenstein 1978) may not affect the test. In principle, other phylogenetic methods such as the neighbor joining (Saitou and Nei 1987) and maximum-likelihood methods could also be used here.

I also found that the MFDM test is free from the confounding impact of hidden structured population if another simple sampling scheme is followed. For example, given 50 diploid individuals ($n = 100$) sampled in total, the false inference of positive selection due to hidden structured population can be avoided if the minimum sampled diploid individuals from each sampling location is at least two.

It is worth noting that these two sampling schemes work well for different ranges of migration rate, and a random mating population can be seen as a specific case of population structure with a large amount of migration between demes. Thus, when there is no knowledge about population structure, these two sampling schemes should always be used to avoid the potential impact of population structure.
may be measured and (partially) excluded when detecting positive selection (Nielsen et al. 2005; Li and Stephan 2006). However, I demonstrate in this study that the explicit knowledge of demography is not always needed to detect recent positive selection because varying population size can never affect tree topology. Therefore, by examining the tree topology of a locus, the MFDM test can reliably detect recent positive selection using DNA polymorphism data from a single locus without inferring demography.

Interestingly, if demographic parameters of the population are estimated from the polymorphism data of multiple loci (see a review by Marjoram and Tavaré 2006), \( P(\xi_{\text{max}} \geq \xi_{\text{max obs}} | \Psi \geq \xi_{\text{max obs}}) \) can be derived or estimated (see eq. 3). I assume that mutations on branches follow the Poisson distribution (Hudson 1990). Then \( P(\xi_{\text{max}} \geq \xi_{\text{max obs}} | \Psi \geq \xi_{\text{max obs}}) = E(1 - e^{-\Psi l/2}) \), where \( l \) is the total length of branches with size \( \geq \xi_{\text{max obs}} \). The branch length is scaled such that one unit represents 2N0 generations, where \( N_0 \) is the current effective population size, \( \Theta = 4N_0\mu \), and \( \mu \) is the mutation rate of the locus. Thus, if demography is known, \( E(1 - e^{-\Psi l/2}) \) can be calculated or estimated; \( P(\xi_{\text{max}} \geq \xi_{\text{max obs}}) \) can be used to reject the neutral hypothesis in the future, and its power to detect selection should be higher than that of the MFDM test.

When considering a recombining region under neutrality, recombination events break the region into a number of nonrecombined fragments. Because those nonrecombined fragments are assumed to be independent of each other in the MFDM test (i.e., the Bonferroni correction), recombination itself actually makes the MFDM test more conservative. However, when positive selection has occurred nearby, recombination within the locus may not reduce the power of the MFDM test considerably because of the reduced recombination parameter due to selection.

Generally, the power of the MFDM test increases when the sample size increases (fig. 6). Thus, I would suggest that the number of sampled diploid individuals should be larger than 50 (i.e., \( n > 100 \)). With the advances in DNA sequencing technology, even genome-wide polymorphism data from a large sample can be available soon, for example, the 1,000 human genomes project (http://www.1000genomes.org/); thus, I expect that the condition of sample size may be easily satisfied for most species.

When the false-positive rate of the MFDM test remains lower than the significance level (<0.05) in wide-range neutral cases, power to detect recent positive selection remains high (up to 0.91) in populations of both constant and varying size. Therefore, for researchers studying recent adaptation of candidate genes in natural populations, the MFDM test may provide the best trade-off between low false-positive rates and high power to detect recent adaptations. Moreover, the unbalancedness of the tree, a summary statistic independent of bottleneck and size expansion, may be useful to detect selection using genome-wide polymorphism data in the future.

### Supplementary Material

Supplementary figures S1–S12 and text are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

### Acknowledgments

I thank Casey Bergman, Ryan Hernandez, Hideki Innan, Li Jin, Yuseob Kim, Martin Lascoux, Jonathan Pritchard, Kai Zeng, and Lili Zhang for discussion and comments on the manuscript. I also thank Joachim Hermisson and Pleun Pennings for useful instructions about how to simulate the trajectory of a neutral mutation backward in time. I would especially thank Thomas Wiehe since a study of microsatellites (together with him) provided the original inspiration of using the unbalanced tree for this work. This work was financially supported by DFG-SFB680, the Shanghai Pujiang Program (08PJ14104), and the Bairen Program.

### References


Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst Zool. 27:401–410.


