Published in final edited form as:
Parasitology. 2015 February ; 142(0 1): S98–S107. doi:10.1017/S0031182014000067.

Recurrent bottlenecks in the malaria life cycle obscure signals of positive selection

Hsiao-Han Chang[1,2] and **Daniel L. Hartl**[1]

[1]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, USA 02138

[2]Center for Communicable Disease Dynamics and Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA 02115

SUMMARY

Detecting signals of selection in the genome of malaria parasites is a key to identify targets for drug and vaccine development. Malaria parasites have a unique life cycle alternating between vector and host organism with a population bottleneck at each transition. These recurrent bottlenecks could influence the patterns of genetic diversity and the power of existing population genetic tools to identify sites under positive selection. We therefore simulated the site-frequency spectrum of a beneficial mutant allele through time under the malaria life cycle. We investigated the power of current population genetic methods to detect positive selection based on the site-frequency spectrum as well as temporal changes in allele frequency. We found that a within-host selective advantage is difficult to detect using these methods. Although a between-host transmission advantage could be detected, the power is decreased when compared with the classical Wright-Fisher population model. Using an adjusted null site-frequency spectrum that takes the malaria life cycle into account, the power of tests based on the site-frequency spectrum to detect positive selection is greatly improved. Our study demonstrates the importance of considering the life cycle in genetic analysis, especially in parasites with complex life cycles.

Keywords

Plasmodium; life history; temporal; site-frequency spectrum

INTRODUCTION

With advances in sequencing technology, malaria parasites from various geographical locations were genotyped or sequenced, and genetic diversity was used to infer demographic history (Joy *et al.* 2003; Chang *et al.* 2012) and population structure (Anderson *et al.* 2000; Mu *et al.* 2005;Manske *et al.* 2012). These same data were examined to reveal genes or nucleotide sites putatively under selection (Volkman *et al.* 2007; Mu *et al.* 2010; Amambua-Ngwa *et al.* 2012). An important component of these analyses is the site-frequency spectrum, which is a histogram depicting the proportion of alleles in different frequency

Hsiao-Han Chang, 677 Huntington Avenue, Kresge Building, Office 506, Boston, Massachusetts 02115, Phone: 617-642-7897, hhchang@hsph.harvard.edu.

classes. Because the site-frequency spectrum can be influenced by changes in population size and natural selection, the site-frequency spectrum of SNPs (single-nucleotide polymorphism) can be used to infer changes in population size and identify selective events. For example, the site-frequency spectrum was used by (Chang *et al.* 2012) to estimate a recent population expansion of 60-fold in Senegal, and Tajima's *D* (one of the summary statistics of the site-frequency spectrum) was used by (Amambua-Ngwa *et al.* 2012) to identify genes under balancing selection in an endemic population in The Gambia.

One alternative to analyzing the site-frequency spectrum is to examine temporal changes in allele frequency. Because allele-frequency fluctuations are largely determined by population size whereas temporal trends depend on selection, population size and selective effects can be estimated based on temporal changes in allele frequency and fluctuations around the trend values. These methods have been developed and implemented in humans, horses, and other species (Waples, 1989; Anderson, 2005; Bollback *et al.* 2008; Malaspinas *et al.* 2012). Temporal samples of malaria parasites are also available and have been used to infer recent parasite dynamics. For instance, Nkhoma *et al.* (2012) used allele frequency fluctuations across 10 years to estimate the effective population size (N_e) on the Thai–Burma border (Nkhoma *et al.* 2012), and Daniels *et al*. (2013) estimated a small N_e of less than 100 following enhanced intervention in Senegal by analysis of molecular barcode genotypes of samples from 2006–2011 (Daniels *et al.* 2013). Moreover, temporal changes in microsatellite allele frequencies were used to estimate N_e of lizard malaria parasites, and a decrease of N_e associated with reduction in prevalence was reported (Schall & St Denis, 2013).

These genetic studies were mostly based on the population genetic tools that assume the Wright-Fisher (WF) population model, which is widely used in organisms including *Drosophila* and humans. A WF population is constant in size, generations are non-overlapping, and each offspring generation is formed from a sample of gametes of the parental generation. We wondered whether the unique and complicated life cycle of malaria could yield allele-frequency dynamics, including the allele-frequency spectrum, allele-frequency changes through time, and variation in allele-frequency changes, when mutations are neutral or under selection, that are qualitatively different from those in the WF model. If so, this would affect inferences from analyses based on such summary statistics as Tajima’s *D* (Tajima, 1989), Fu and Li’s *F** and *D** (Fu & Li, 1993), and Fay and Wu’s *H* (Fay & Wu, 2000). Our initial study (Chang *et al*. 2013) has shown that, in comparison with the WF model, both genetic drift and the efficiency of purifying selection are higher under the malaria life cycle, and that the null distribution of the site- frequency spectrum is skewed toward lower-frequency alleles. These findings suggest that the malaria life cycle leads to biases in estimating demographic history and identifying signals of selection. However, it is as yet unclear how these intrinsic differences between the WF model and the malaria life cycle affect the power to detect sites under selection using typical statistics based on temporal changes in allele frequency or on the site-frequency spectrum.

Here, we examined how the malaria life cycle affects signals of selection identified by temporal changes in allele frequency or by the site-frequency spectrum. We first simulated the changes in frequency of beneficial alleles through time, estimated selection coefficients from temporal changes in allele frequency, and determined the conditions in which sites under positive selection could be detected. We also simulated the site-frequency spectrum for beneficial alleles and compared the power to identify sites under selection when analyzed either the typical null distribution with the WF- model assumption or the adjusted null distribution under the malaria life cycle. Finally, we discuss potential future directions and the importance of considering the life cycle when studying malaria parasites or other species with unconventional life histories.

MATERIALS AND METHODS

Simulations of allele frequency changes and the site-frequency spectrum

Simulations for allele frequency changes and the site-frequency spectrum were performed as described in Chang *et al*. (2013). Three kinds of models were examined. In one, selection is assumed to take place only in the host organism (the *host selection* model); in another, the selection is effected through transmission from host to vector (the *transmission selection* model); and in the third, selection takes place both in the host as well as through transmission (the *host/transmission selection* model). Drug resistance genes and genes that function in hemoglobin breakdown and metabolism are examples of host selection; genes for sporozoite development and migration to the salivary glands and genes related to the development of the sexual forms of malaria parasites are examples of transmission selection. The quantity h is the selection coefficient measuring the advantage or disadvantage of a mutant allele within the human host per asexual cycle, and t is the selection coefficient measuring the advantage or disadvantage of a mutant allele in host-to-vector transmission. In the host selection model $h \neq 0$ but $t = 0$, and in the transmission selection model $h = 0$ but $\mathrm{t} \neq 0$. The host/transmission selection model with $h = t \neq 0$ assumes equality of the selection coefficients for the sake of simplicity and to compare the relative importance of h and t. In what follows, the usual symbol of the selection coefficient (s) in the WF model, when used in reference to malaria, refers to h in the host selection model, t in the transmission selection model, and $h = t$ in the host/transmission selection model. The default values of parameters from Chang *et al*. (2013) were used in the simulations unless stated otherwise. We assume that population sizes of human host and mosquito vector are both 1000, and the number of parasites transmitted between the vector and the human host is 10. If population sizes of human host and mosquito vector are different, there is another level of bottleneck, and it is likely that variation in allele- frequency changes is higher and the estimated selection coefficient less accurate than shown here. The site-frequency spectrum was obtained by sampling one allele within each of n randomly chosen (without replacement) human hosts, with $n = 50$. The simulation model does not allow for mixed infections, hence it assumes complete inbreeding. The level of mixed infection affects the rate of recombination in the population, and a model that allows mixed infection would be of great interest (see Chang *et al.* 2013).

Estimating the selection coefficient (s)

Selection coefficients were estimated using allele frequency changes through time by the program *sel2ns* (Bollback *et al.* 2008). Allele frequencies at three time points separated by intervals of 5 generations were used for estimation. Sample allele frequencies were obtained by sampling one allele from each of 1000 human hosts. One generation under the malaria model is defined as one complete life cycle that includes a vector-host transmission bottleneck, population expansion within the host, a host-vector bottleneck, and population expansion within the vector. When estimating the selection coefficient *(s)* from allele frequency changes in the initial stage of the host selection model, exponential grid spacing was used (spacing ratio = 100 and $\lambda = 0.5$) because allele frequency is close to 0; equal grid spacing was used for all other estimations. The number of grid points was set to 500. Only mutations that eventually become fixed in the population were used for estimating *s* under the malaria model (*condfix* = 1). The number of replicates for each selection coefficient of each model was 10. Effective population size under the malaria model was estimated from simulated frequency changes of neutral alleles using the program *CoNe* (Anderson, 2005). The effective population size, 1000, was used when estimating *s* in the malaria model because the estimate of effective population size from neutral alleles is in the order of 1000. The significance of a selection coefficient is determined by a likelihood ratio test. The estimated *s* is significantly different from 0 if the difference in log-likelihoods of the estimated *s* and 0 is greater than 1.92, which is half of the 95th quantile of a chi-square distribution with 1 degree of freedom.

Site-frequency spectrum

Tajima's *D* statistic was used to summarize site-frequency spectrum (Tajima, 1989). The null distributions of Tajima's *D* under the Wright-Fisher model and the malaria model were generated by randomly sampling *n* alleles from the site-frequency spectra 10,000 times, with $n = 50$. The probability that Tajima's *D* of selected alleles is significant was obtained by comparing 100 simulated Tajima's *D* values of selected alleles with standard (WF) or adjusted (malaria) null distributions and calculating the proportion of Tajima's *D* values greater than the 95th quantile of the null distribution (one-sided test). When comparing two Tajima's *D* distributions, a one-sided Mann- Whitney test was used (with the alternative hypothesis that Tajima's *D* of beneficial alleles is larger than Tajima's *D* of neutral alleles).

RESULTS

Allele frequency changes through time

We first compared the changes in frequency of beneficial alleles with $s = 0.1$ in three malaria models: *host selection* model (selection only in red-cell stages), *transmission selection* model (selection only via transmission), and *host/transmission selection* model (selection in both red-cell stages and transmission). The results are summarized in Figs. 1 and 2. Note that these are cases when beneficial alleles are eventually fixed in the population. Fig. 1 shows that frequency of beneficial alleles increases faster in the initial generation if there is a selective advantage within the human host (that is, in the host selection and host/transmission selection models). Afterwards, the transmission advantage gradually overtakes, and the allele frequency in the transmission selection and host/transmission selection models increases more rapidly (Fig. 2). Fig. 2 also shows that the fixation time is more sensitive to selection affecting transmission. The fixation time under the host/transmission selection and transmission selection models is much shorter than the host selection model. Without a transmission advantage, the frequency of the beneficial allele in the host selection model fluctuates stochastically before becoming fixed in the population. The reason that the fixation time in the host/transmission selection model is shorter than that in the transmission selection model is because the mutant allele has both a within-host selective advantage and a between-host transmission advantages.

Estimation of selection coefficients

We then estimated *s* using temporal allele-frequency changes (Fig. 3) and tested whether estimated values are significantly different from 0 (Table 1). We found that, under the Wright-Fisher (WF) model with population size 10,000 and sample size 1000, the median of the estimated *s* is highly correlated with the true *s* (*P*-value = 5×10^{-6}), although the absolute values of the estimates are not exactly correct (Fig. 3). Moreover, under the WF model, the estimated values of *s* within the simulated range of selection coefficients (*s* = 0.05 ~ 0.1) are all significantly different from 0. Under the malaria life cycle, on the other hand, the allele-frequency trajectories and the estimates of *s* are highly dependent on the model (host selection, transmission selection, or host/transmission selection).

In the host selection model, we estimated *s* using allele-frequency changes in both the initial stage when the allele frequency is still low and at an intermediate stage when the allele frequency is greater than 20%. The selective advantage appears almost immediately in the host selection model, but once the beneficial allele becomes fixed within a human host, allele-frequency changes are dominated by the neutral process and any selective advantage is unlikely to be detected because there is no transmission advantage between hosts. This is the rationale for the estimates of *s* using both the initial and intermediate stages of allele-frequency change. The results show that the medians of estimates from both stages are not correlated with the true *s* [*P*-values = 0.68 (initial) and 0.64 (intermediate), Fig. 3] and are unlikely to be significantly different from zero [15% (initial) and 8% (intermediate)] (Table 1). Although the estimates of *s* from the initial frequency changes are slightly more likely to be significantly different from 0 than those from the intermediate stages, only 15% of the estimates are significantly different from zero, demonstrating that the selective advantage within the human host is difficult to detect. In fact, Fig. 1 shows that selective advantage is only obvious during the first two generations when *h* is 0.1. Moreover, it is unlikely that a random sample includes alleles of low frequency.

In the transmission selection model, the median of estimates of the selection coefficient is significantly correlated with the true value (*P*-value = 0.003) and the median of estimates from the WF model (*P*-value = 0.003) (Fig. 3), and the estimates are likely to be significantly different from zero (69%, Table 1). In the host/transmission selection model, while the estimated coefficients are likely to be significantly different from 0 (69%, Table 1), the median of the estimates is not significantly correlated with the true value (*P*-value = 0.1) or the median of estimates from the WF model (*P*-value = 0.1). These results suggest that a transmission advantage is more likely to be detected as statistically significant than a selective advantage within the human host.

Site-frequency spectrum of selected alleles

Tests based on the site-frequency spectrum are commonly used for detecting signals of selection. Here, we investigated the power to detect signals of selection by Tajima's *D*, a test based on the site-frequency spectrum, under the malaria life cycle. The site-frequency spectra of selected alleles under three malaria models were simulated and are shown in Fig. 4, along with the malaria neutral site-frequency spectrum and than the WF neutral site-frequency spectrum. Under the host selection model, the spectra of alleles with s = 0.1 and 0.01 are more skewed toward high frequency alleles than the malaria neutral site-frequency spectrum, but less skewed than the WF neutral site-frequency spectrum. This comparison indicates that uncritical use of the WF neutral site-frequency spectrum can sometimes lead one to interpret a statistical signal of selection as having the wrong sign. Under the transmission selection and host/transmission selection models, the spectra of alleles with s = 0.1 and 0.01 are more skewed toward high frequency alleles than either neutral frequency spectra, while the spectra of alleles with s = 0.001 are more skewed than the malaria neutral frequency spectrum but not that of WF.

We then compared Tajima's *D* values of selected alleles to the null Tajima's *D* distributions generated from both the WF model (standard null) and the malaria model (adjusted null) (Table 2). Under the host selection model, Tajima's *D* values of selected alleles are significantly higher than Tajima's *D* under the neutral malaria model, but not significantly different from Tajima's *D* of the neutral WF model (Table 2, left column). In the transmission selection and host/transmission selection model, Tajima's *D* values of selected alleles are significantly higher than Tajima's *D* from both neutral models if the number of segregating sites (S) is larger and/or s = 0.01 (Table 2, left column). Tajima's *D* values are more likely to be significantly greater than 0 when using the null distribution of the malaria model, and the improvement is more obvious under the transmission selection and host/transmission selection models when the number of segregating sites is larger and s = 0.1 or 0.01 (Table 2, right column). Note that Tajima's *D* is less likely to be significantly positive when s = 0.1 than when s = 0.01 because there are more high-frequency derived alleles when s = 0.1 and the higher proportion of alleles with frequency close to 1 decreases Tajima's *D* value. In this situation, Fay and Wu's test (Fay & Wu, 2000) might be more powerful to detect positive selection because it is sensitive to the excess of high-frequency derived alleles.

DISCUSSION

Malaria is one of the leading causes of morbidity and mortality worldwide, and efforts have been made to detect signals of selection in the genome of malaria parasites in order to identify targets for intervention. In this study, we investigated the power to detect signals of selection under the malaria life cycle.

We showed that allele-frequency changes through time are more sensitive to transmission advantage than to selective advantage within the human host. Similarly, transmission advantage also affects the site-frequency spectrum to a greater extent. Transmission advantage predominates because within-host selection is important only when mutations are still segregating within a host. After a beneficial mutation becomes fixed within a host, it is transmission selection that determines the changes of allele frequency in the overall parasite population. Because of multiple asexual generations within a host, a mutation under positive selection within the human host reaches high frequency or becomes fixed within a host quickly and therefore, with single infections, transmission selection usually predominates over host selection. However, even though allele-frequency changes are more sensitive to transmission advantage, the power to detect selection by following temporal changes in allele frequency is smaller under the malaria life cycle than under the WF model (Table 1). The reduced power occurs because the changes in allele frequency depend on serial passages of alleles within and between hosts, and during these passages genetic drift is enhanced by the bottlenecks and the selective and transmission effects of a mutation may differ. For example, the genes related to the development of the sexual forms of the malaria parasite are expected to be under transmission selection but not host selection. These serial passages enhance the variance in allele-frequency changes and the difficulty to detect selection.

The power to detect selection could be potentially increased when the number of hosts is large because signals of selection are expected to be clearer with decreased level of genetic drift. The power of tests based on the site-frequency spectrum is also decreased by the malaria life cycle unless one adjusts the null distribution in accordance with the malaria life cycle. When using the standard WF null distribution of Tajima's *D*, the highest rate of detecting positive selection in our simulation is only 21% (Table 2). When the null distribution is adjusted for the malaria life cycle, the power of detecting positive selection is greatly improved (the highest rate is 78%, Table 2). Because the true demographic parameters (such as the numbers of human hosts and mosquito vectors) of a real population are unknown, they need to be estimated when attempting to obtain an adjusted null distribution by simulations. However, demographic parameters are unlikely to be estimated accurately because, even if we only consider neutral sites, both the demographic history and the malaria life cycle act on patterns of polymorphism, and no current method is known that can separate them. An alternative approach to obtaining an adjusted null distribution of Tajima's *D* is by randomly sampling alleles from empirical site-frequency spectrum of a group of SNPs that are thought to be neutral or nearly neutral, such as synonymous sites. In this way, the adjusted null distribution of Tajima's *D* is influenced by both demographic history and the malaria life cycle and could be used to improve the power of detecting positive selection or balancing selection in the genome of malaria parasites. This is similar to the analysis in Chang *et al.* (2012), in which demographic parameters were estimated by fitting a demographic model to the synonymous site-frequency spectrum (Gutenkunst *et al.* 2009), and Tajima's *D* null distribution was generated by the estimated demographic model. This approach reduces the extent to which the malaria life cycle reduces the power to identify genes under positive or balancing selection using positive Tajima's *D*. In contrast, Amambua-Ngwa *et al.* (2012) only examined genes with positive Tajima's *D*, whereas even negative Tajima's *D* could be "effectively positive" if demographic history and the malaria life cycle were taken into account.

The importance of considering life history in studying population dynamics and evolution has been discussed and reviewed (Barrett *et al.* 2008). The factors such as epidemiological dynamics and host life history or behavior could potentially affect transmission of parasites between hosts, effective population size, and patterns of genetic diversity. For example, it was shown that *Microbotryum violaceum* from plant host species with different life histories have different patterns of genetic diversity (Bucheli *et al.* 2001). Unless life history is taken into proper consideration, inferences with regard to parasite evolution and effectiveness of intervention could possibly be misleading.

Future work

This study underlines the importance of developing modeling and methods that take life histories into account on studying malaria parasites or other species with unconventional life histories. Ignoring life histories could potentially lead to misinterpretation of selection signals and mislead our understanding of how selection has shaped the *P. falciparum* genome. Questions including which features of life history are important for studying parasites dynamics and/or parasite evolution, how they are important, and methods to incorporate them into analytical tools, remain to be studied. More specifically, among the population genetic tools that need to be investigated and reexamined under the malaria life cycle are tests comparing both within-species diversity and between-species divergence (the McDonald–Kreitman test, or MK test) (McDonald & Kreitman, 1991), approaches based on the relationship between phenotype (such as drug resistance) and genotype (genome-wide association study, GWAS), linkage disequilibrium-based tests (EHH test and iHS test) (Sabeti *et al.* 2002; Voight *et al.* 2006), and tools based on between-population differentiation (F_{st}-based test).

In population genetics, the MK test is thought to be robust to demographic changes because synonymous and nonsynonymus mutations are both affected by the same demography (Nielsen, 2001). However, the malaria life cycle might influence fixation and polymorphism in different ways than in the WF model, and so the robustness of the MK test could be compromised. Even more importantly, different species that used to calculate divergence (such as *Plasmodium falciparum* and *Plasmodium reichenowi*) may have different hosts and life histories, and this adds another complexity and potential bias in tests in which between-species divergence is required. Genome-wide association tests and linkage disequilibrium-based tests have been used in malaria parasites to identify loci related to drug resistance (Van Tyne *et al.* 2011; Chang *et al.* 2012; Park *et al.* 2012), yet the power of these tools under the malaria life cycle and their relationship with effective recombination rate and transmission intensity have not been investigated. It remains to be determined how recombination and enhanced genetic drift under the malaria life cycle interact to affect the decay of linkage disequilibrium through time. With genomic sequences of samples from different geographical locations available, F_{st}-based tests could be useful to detect local adaptation, but again the interaction of migration between populations, drift, and selection under the malaria life cycle need to be studied. In the meantime, researchers should be cautious and are encouraged to consider potential biases introduced by life history when interpreting patterns of genetic diversity and/or drawing inferences from genetic analysis of malaria parasites.

Acknowledgments

The authors thank Daniel Park for helpful discussion on the method of estimating selection coefficients.

FINANCIAL SUPPORT

This study is supported by NIH grant AI099105. H.H.C is supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number U54GM088558. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Amambua-Ngwa A, Tetteh KK, Manske M, Gomez-Escobar N, Stewart LB, Deerhake ME, Cheeseman IH, Newbold CI, Holder AA, Knuepfer E, Janha O, Jallow M, Campino S, Macinnis B, Kwiatkowski DP, Conway DJ. Population genomic scan for candidate signatures of balancing selection to guide antigen characterization in malaria parasites. PLoS Genetics. 2012; 8:e1002992.10.1371/journal.pgen.1002992 [PubMed: 23133397]

Anderson EC. An efficient Monte Carlo method for estimating Ne from temporally spaced samples using a coalescent-based likelihood. Genetics. 2005; 170:955–967.10.1534/genetics.104.038349 [PubMed: 15834143]

Anderson TJ, Haubold B, Williams JT, Estrada-Franco JG, Richardson L, Mollinedo R, Bockarie M, Mokili J, Mharakurwa S, French N, Whitworth J, Velez ID, Brockman AH, Nosten F, Ferreira MU, Day KP. Microsatellite markers reveal a spectrum of population structures in the malaria parasite Plasmodium falciparum. Molecular Biology and Evolution. 2000; 17:1467–1482. [PubMed: 11018154]

Barrett LG, Thrall PH, Burdon JJ, Linde CC. Life history determines genetic structure and evolutionary potential of host-parasite interactions. Trends in Ecology and Evolution. 2008; 23:678–685.10.1016/j.tree.2008.06.017 [PubMed: 18947899]

Bollback JP, York TL, Nielsen R. Estimation of 2Nes from temporal allele frequency data. Genetics. 2008; 179:497–502.10.1534/genetics.107.085019 [PubMed: 18493066]

Bucheli E, Gautschi B, Shykoff JA. Differences in population structure of the anther smut fungus Microbotryum violaceum on two closely related host species, Silene latifolia and S. dioica. Molecular Ecology. 2001; 10:285–294. [PubMed: 11298945]

Chang HH, Moss EL, Park DJ, Ndiaye D, Mboup S, Volkman SK, Sabeti PC, Wirth DF, Neafsey DE, Hartl DL. The malaria life cycle intensifies both natural selection and random genetic drift. Proceedings of the National Academy of Sciences of the United States of America. 2013; 110:20129– 20134.10.1073/pnas.1319857110 [PubMed: 24259712]

Chang HH, Park DJ, Galinsky KJ, Schaffner SF, Ndiaye D, Ndir O, Mboup S, Wiegand RC, Volkman SK, Sabeti PC, Wirth DF, Neafsey DE, Hartl DL. Genomic sequencing of Plasmodium falciparum malaria parasites from Senegal reveals the demographic history of the population. Molecular Biology and Evolution. 2012; 29:3427–3439.10.1093/molbev/mss161 [PubMed: 22734050]

Daniels R, Chang HH, Sene PD, Park DC, Neafsey DE, Schaffner SF, Hamilton EJ, Lukens AK, Van Tyne D, Mboup S, Sabeti PC, Ndiaye D, Wirth DF, Hartl DL, Volkman SK. Genetic surveillance detects both clonal and epidemic transmission of malaria following enhanced intervention in Senegal. PLoS One. 2013; 8:e60780.10.1371/journal.pone.0060780 [PubMed: 23593309]

Fay JC, Wu CI. Hitchhiking under positive Darwinian selection. Genetics. 2000; 155:1405–1413. [PubMed: 10880498]

Fu YX, Li WH. Statistical tests of neutrality of mutations. Genetics. 1993; 133:693–709. [PubMed: 8454210]

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genetics. 2009; 5:e1000695.10.1371/journal.pgen.1000695 [PubMed: 19851460]

Joy DA, Feng X, Mu J, Furuya T, Chotivanich K, Krettli AU, Ho M, Wang A, White NJ, Suh E, Beerli P, Su XZ. Early origin and recent expansion of Plasmodium falciparum. Science. 2003; 300:318–321.10.1126/science.1081449 [PubMed: 12690197]

Malaspinas AS, Malaspinas O, Evans SN, Slatkin M. Estimating allele age and selection coefficient from time-serial data. Genetics. 2012; 192:599–607.10.1534/genetics.112.140939 [PubMed: 22851647]

Manske M, Miotto O, Campino S, Auburn S, Almagro-Garcia J, Maslen G, O'Brien J, Djimde A, Doumbo O, Zongo I, Ouedraogo JB, Michon P, Mueller I, Siba P, Nzila A, Borrmann S, Kiara SM, Marsh K, Jiang H, Su XZ, Amaratunga C, Fairhurst R, Socheat D, Nosten F, Imwong M, White NJ, Sanders M, Anastasi E, Alcock D, Drury E, Oyola S, Quail MA, Turner DJ, Ruano-Rubio V, Jyothi D, Amenga-Etego L, Hubbart C, Jeffreys A, Rowlands K, Sutherland C, Roper C, Mangano V, Modiano D, Tan JC, Ferdig MT, Amambua-Ngwa A, Conway DJ, Takala-Harrison S, Plowe CV, Rayner JC, Rockett KA, Clark TG, Newbold CI, Berriman M, MacInnis B, Kwiatkowski DP. Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. Nature. 2012; 487:375–379.10.1038/nature11174 [PubMed: 22722859]

McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in Drosophila. Nature. 1991; 351:652–654.10.1038/351652a0 [PubMed: 1904993]

Mu J, Awadalla P, Duan J, McGee KM, Joy DA, McVean GA, Su XZ. Recombination hotspots and population structure in Plasmodium falciparum. PLoS Biology. 2005; 3:e335.10.1371/journal.pbio.0030335 [PubMed: 16144426]

Mu J, Myers RA, Jiang H, Liu S, Ricklefs S, Waisberg M, Chotivanich K, Wilairatana P, Krudsood S, White NJ, Udomsangpetch R, Cui L, Ho M, Ou F, Li H, Song J, Li G, Wang X, Seila S, Sokunthea S, Socheat D, Sturdevant DE, Porcella SF, Fairhurst RM, Wellems TE, Awadalla P, Su XZ. Plasmodium falciparum genome-wide scans for positive selection, recombination hot spots and resistance to antimalarial drugs. Nature Genetics. 2010; 42:268–271.10.1038/ng.528 [PubMed: 20101240]

Nielsen R. Statistical tests of selective neutrality in the age of genomics. Heredity. 2001; 86:641–647. [PubMed: 11595044]

Nkhoma SC, Nair S, Al-Saai S, Ashley E, McGready R, Phyo AP, Nosten F, Anderson TJ. Population genetic correlates of declining transmission in a human pathogen. Molecular Ecology. 2012; 22:273–285.10.1111/mec.12099 [PubMed: 23121253]

Park DJ, Lukens AK, Neafsey DE, Schaffner SF, Chang HH, Valim C, Ribacke U, Van Tyne D, Galinsky K, Galligan M, Becker JS, Ndiaye D, Mboup S, Wiegand RC, Hartl DL, Sabeti PC, Wirth DF, Volkman SK. Sequence-based association and selection scans identify drug resistance loci in the Plasmodium falciparum malaria parasite. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109:13052–13057.10.1073/pnas.1210585109 [PubMed: 22826220]

Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES. Detecting recent positive selection in the human genome from haplotype structure. Nature. 2002; 419:832–837.10.1038/nature01140 [PubMed: 12397357]

Schall JJ, St Denis KM. Microsatellite loci over a thirty-three year period for a malaria parasite (Plasmodium mexicanum): bottleneck in effective population size and effect on allele frequencies. Parasitology. 2013; 140:21–28.10.1017/S0031182012001217 [PubMed: 22948096]

Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989; 123:585–595. [PubMed: 2513255]

Van Tyne D, Park DJ, Schaffner SF, Neafsey DE, Angelino E, Cortese JF, Barnes KG, Rosen DM, Lukens AK, Daniels RF, Milner DA Jr, Johnson CA, Shlyakhter I, Grossman SR, Becker JS, Yamins D, Karlsson EK, Ndiaye D, Sarr O, Mboup S, Happi C, Furlotte NA, Eskin E, Kang HM, Hartl DL, Birren BW, Wiegand RC, Lander ES, Wirth DF, Volkman SK, Sabeti PC. Identification and functional validation of the novel antimalarial resistance locus PF10_0355 in Plasmodium falciparum. PLoS Genetics. 2011; 7:e1001383.10.1371/journal.pgen.1001383 [PubMed: 21533027]

Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. PLoS Biology. 2006; 4:e72.10.1371/journal.pbio.0040072 [PubMed: 16494531]

Volkman SK, Sabeti PC, DeCaprio D, Neafsey DE, Schaffner SF, Milner DA Jr, Daily JP, Sarr O, Ndiaye D, Ndir O, Mboup S, Duraisingh MT, Lukens A, Derr A, Stange-Thomann N, Waggoner S, Onofrio R, Ziaugra L, Mauceli E, Gnerre S, Jaffe DB, Zainoun J, Wiegand RC, Birren BW, Hartl DL, Galagan JE, Lander ES, Wirth DF. A genome- wide map of diversity in Plasmodium falciparum. Nature Genetics. 2007; 39:113–119.10.1038/ng1930 [PubMed: 17159979]

Waples RS. A generalized approach for estimating effective population size from temporal changes in allele frequency. Genetics. 1989; 121:379–391. [PubMed: 2731727]

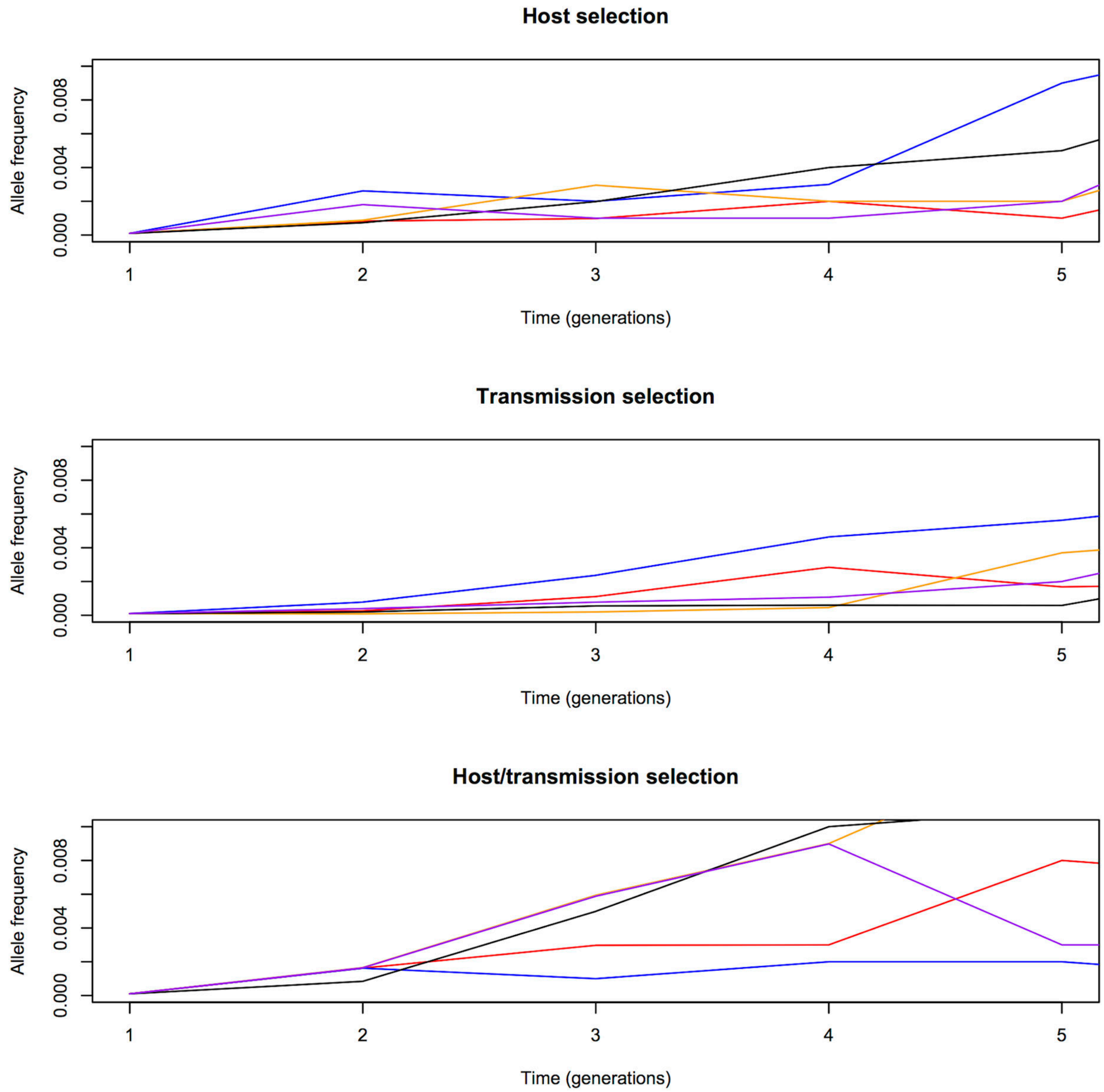

Fig. 1.

Allele frequency changes in the first 5 generations. The trajectories of beneficial mutations with *s* = 0.1 in three models showing that the frequency of beneficial alleles increases faster from generation 1 to generation 2 if there is a selective advantage in the red-blood-cell stages in the human host (i.e., the host selection and host/transmission selection models). The 5 differently colored lines represent independent replicates with the same initial frequency.

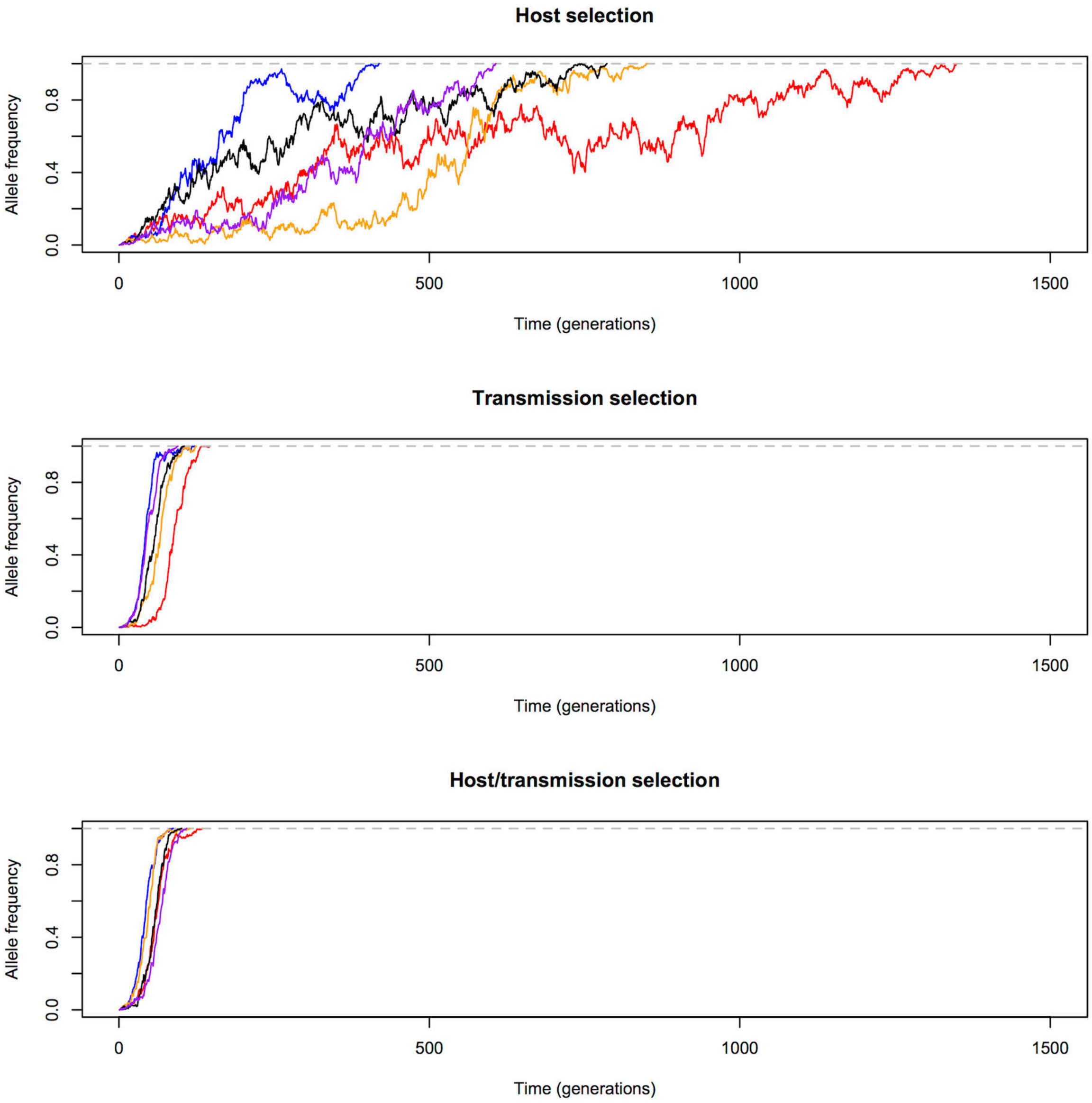

Fig. 2.

Allele frequency changes until fixation. Transmission advantage influences the trajectories of beneficial alleles ($s = 0.1$) more than selective advantage in red-cell stages. Allele frequency in the transmission selection and host/transmission selection models increases much faster than in the host selection model. The 5 differently colored lines represent independent replicates with the same initial frequency.

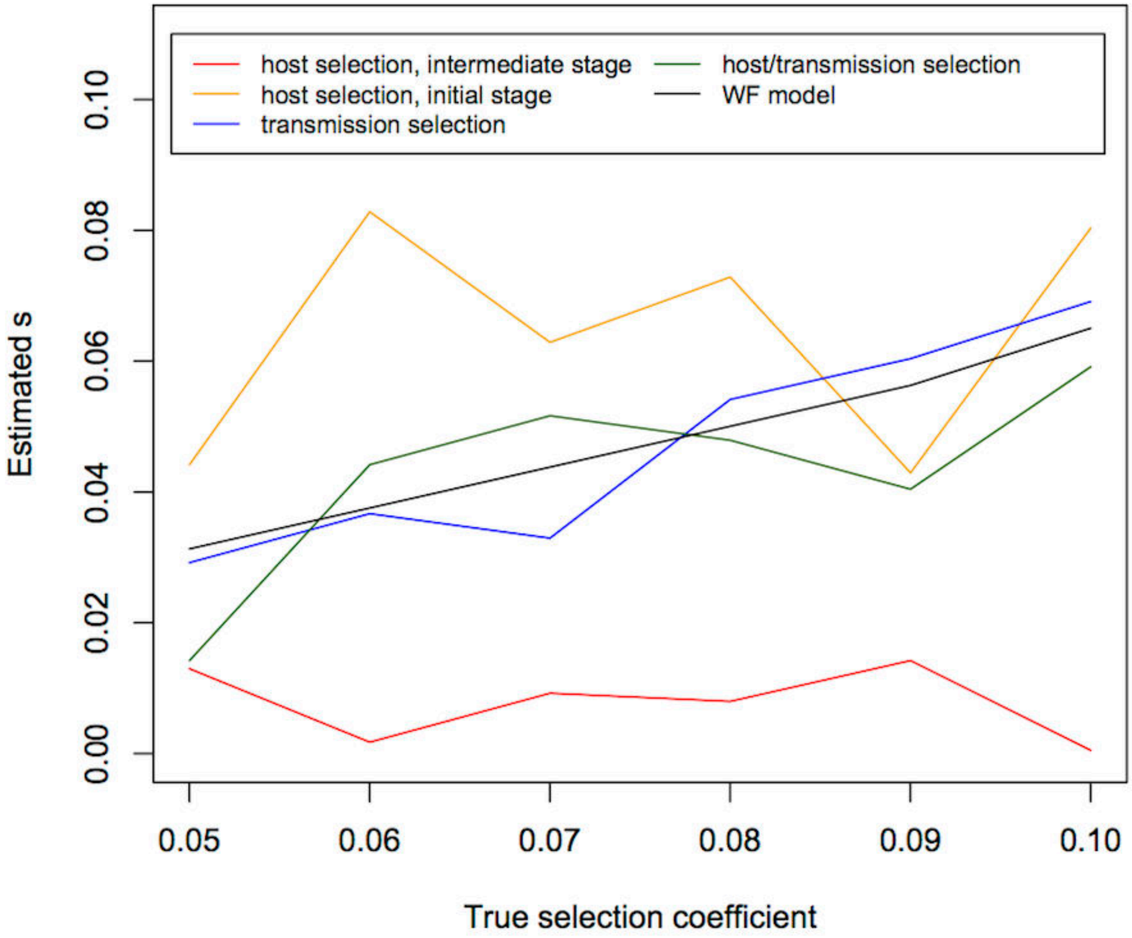

Fig. 3.

The estimates of the selection coefficient (*s*) under the WF and malaria models. The medians of estimates of the selection coefficients under the WF and transmission selection models are correlated with the true values, while the median of estimates in the host selection and host/transmission selection models are not. The medians of 10 replicates are shown here.

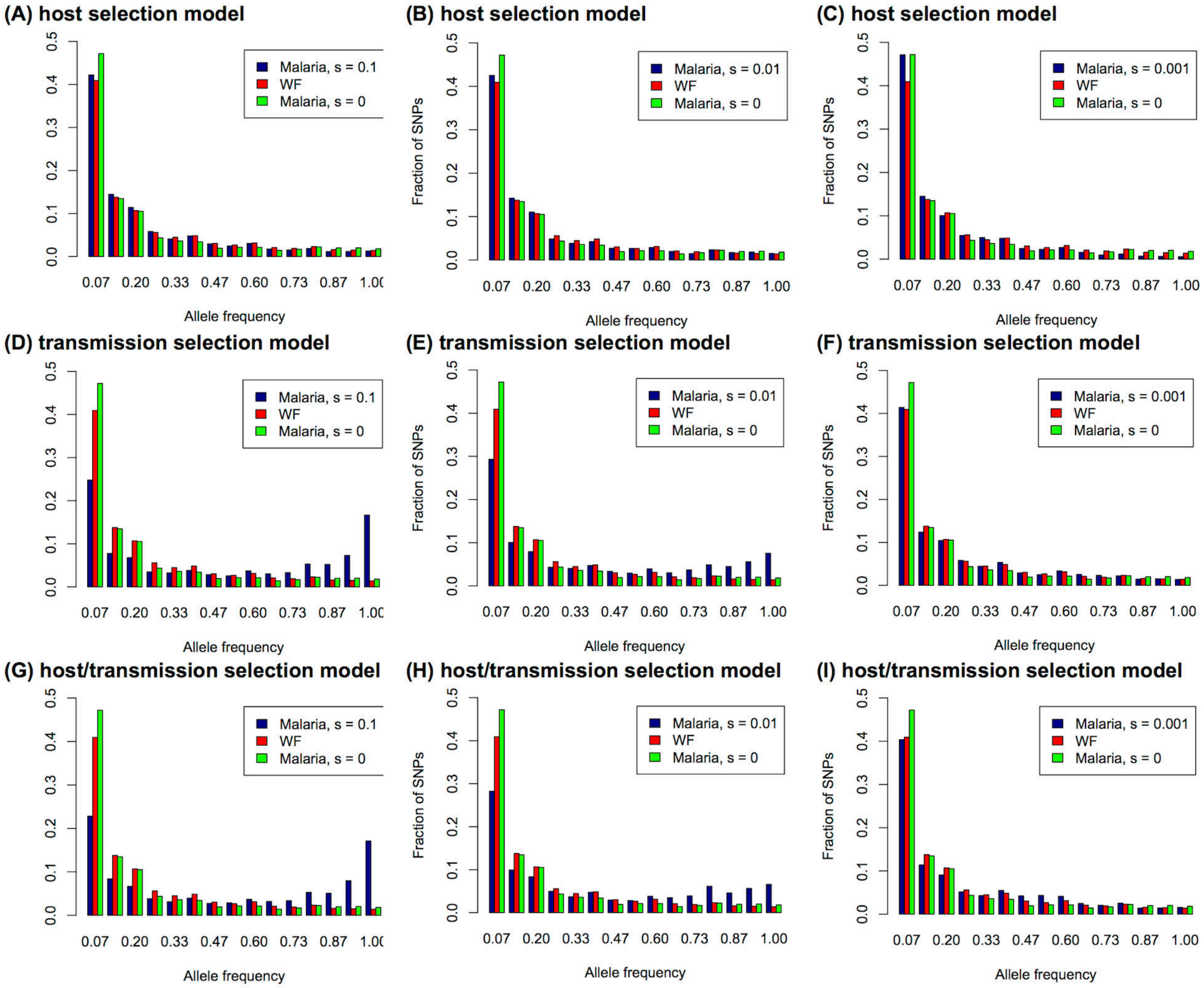

Fig. 4.

Allele frequency spectra of selected alleles under three malaria models. Under the host selection model, the spectra of alleles with *s* = 0.1 and 0.01 are more skewed toward high frequency alleles than the malaria neutral site-frequency spectrum, but less skewed than the WF neutral site-frequency spectrum. Under the transmission selection and host/transmission selection models, the spectra of alleles with *s* = 0.1 and 0.01 are more skewed toward high frequency alleles than both neutral frequency spectra, while the spectra of alleles with *s* = 0.001 are only more skewed than the malaria neutral frequency spectrum.

Table 1

Proportion of estimated selection coefficients significantly different from 0

	Selection coefficient *s*						
	0.05	**0.06**	**0.07**	**0.08**	**0.09**	**0.1**	**All**
Wright-Fisher model	10/10*	10/10	10/10	10/10	10/10	10/10	60/60
Host selection model, initial stage	2/10	2/10	1/10	2/10	1/10	1/10	9/60
Host selection model, intermediate stage	1/10	0/10	1/10	1/10	0/10	2/10	5/60
Transmission selection model	5/10	5/10	5/10	8/10	10/10	8/9	41/59
Host/transmission selection model	2/10	8/10	8/10	8/10	7/10	7/8	40/58

*: The number preceding the virgule (“/”) represents the number of cases with the estimated selection coefficient significantly different from 0; the number following it is the sample size.

Table 2

Comparing Tajima's *D* of selected alleles with null distributions

		***P*-value of Mann-Whitney test in Tajima's *D* distributions**			**Tajima's *D* rejection rate**		
		***s*[¶] = 0.001**	***s* = 0.01**	***s* = 0.1**	***s* = 0.001**	***s* = 0.01**	***s* = 0.1**
***S*[§] = 5**							
Host selection model	Standard null	0.82	0.81	0.87	3 %	3 %	4 %
	Adjusted null	0.001	0.002	0.002	9 %	10 %	9 %
transmission selection model	Standard null	0.61	0.05	0.47	4 %	5 %	5 %
	Adjusted null	0.0001	3.4×10^{-8}	2.9×10^{-5}	13 %	15 %	9 %
host/transmission selection model	Standard null	0.14	0.009	0.26	7 %	7 %	7 %
	Adjusted null	3.7×10^{-7}	2.3×10^{-10}	3.9×10^{-6}	14 %	15 %	15 %
***S* = 10**							
Host selection model	Standard null	0.95	0.98	0.98	4 %	5 %	3 %
	Adjusted null	6.9×10^{-5}	0.0003	0.0004	14 %	11 %	11 %
transmission selection model	Standard null	0.74	0.0001	0.17	5 %	8 %	6 %
	Adjusted null	6.4×10^{-7}	$< 2 \times 10^{-16}$	6×10^{-11}	16 %	22 %	13 %
host/transmission selection model	Standard null	0.02	8.9×10^{-5}	0.03	9 %	15 %	8 %
	Adjusted null	1.4×10^{-12}	$< 2 \times 10^{-16}$	1.2×10^{-13}	24 %	26 %	18 %
***S* = 50**							
Host selection model	Standard null	1	1	1	0 %	3 %	2 %
	Adjusted null	2.5×10^{-7}	5.6×10^{-16}	$< 2 \times 10^{-16}$	21 %	23 %	27 %
transmission selection model	Standard null	0.22	6.9×10^{-11}	0.42	4 %	25 %	6 %
	Adjusted null	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	45 %	63 %	40 %
host/transmission selection model	Standard null	1.3×10^{-7}	$< 2 \times 10^{-16}$	0.008	16 %	21 %	8 %
	Adjusted null	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	59 %	78 %	46 %

[§] *S* is the number of segregating sites that are sampled from the site-frequency spectrum to calculate Tajima's *D*.

[¶] *s* is selection coefficient, referring to refers to *h* in the host selection model, *t* in the transmission selection model, and *h* = *t* in the host/transmission selection model.