

## REVIEW

# Revisiting the false positive rate in detecting recent positive selection

Jinggong Xiang-Yu<sup>1</sup>, Zongfeng Yang<sup>1,2</sup>, Kun Tang<sup>1</sup> and Haipeng Li<sup>1,\*</sup>

<sup>1</sup> CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

<sup>2</sup> University of Chinese Academy of Sciences, Beijing 100049, China

\* Correspondence: lihaipeng@picb.ac.cn

Received November 15, 2015; Revised January 28, 2016; Accepted March 20, 2016

**There is increasing interest in studying the molecular mechanisms of recent adaptations caused by positive selection in the genomics era. Such endeavors to detect recent positive selection, however, have been severely handicapped by false positives due to the confounding impact of demography and the population structure. To reduce false positives, it is critical to conduct a functional analysis to identify the true candidate genes/mutations from those that are filtered through neutrality tests. However, the extremely high cost of such functional analysis may restrict studies within a small number of model species. In particular, when the false positive rate of neutrality tests is high, the efficiency of the functional analysis will also be very low. Therefore, although the recent improvements have been made in the (joint) inference of demography and selection, our ultimate goal, which is to understand the mechanism of adaptation generally in a wide variety of natural populations, may not be achieved using the currently available approaches. More attention should thus be spent on the development of more reliable tests that could not only free themselves from the confounding impact of demography and the population structure but also have reasonable power to detect selection.**

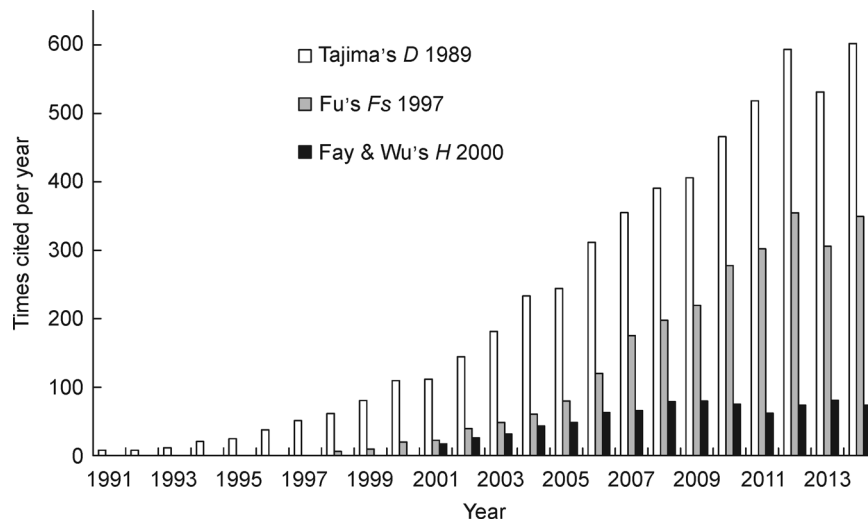
**Keywords:** recent positive selection; selective sweep; demography; population structure; false positive

## INTRODUCTION

In the last 3–4 million years, brain volume within the human lineage has increased from less than 400 cm<sup>3</sup> to 1,400 cm<sup>3</sup>, which sets us apart from other primates. Some evidence suggests that the human brain is still adaptively evolving [1], and that positive selection may also play an essential role in the development of speech [2], and in the enabling of adults to digest the lactose contained in milk [3–6]. Moreover, understanding how and why selection acts can help us to reveal the mechanisms that underlie diseases and microbial resistance [7]. Thus, there is an increasing interest in the articles detecting recent positive Darwinian selection or adaptation. This increasing interest can be seen in the increasing number of citations of three key methodological papers [8–10] (Figure 1).

Positive selection is an evolutionary process that increases the frequency of mutations that confer a fitness advantage to individuals carrying those mutations. These selective events leave their footprints in the genomes of living organisms. Specifically, recent positive selection can reduce genetic diversity [11–15] and alter the pattern of intraspecific polymorphisms [8, 10, 16–18] of a neutral locus that is linked to the selected locus. These features can be used to detect positive selection. However, because demography and the population structure can cause the same patterns in DNA sequence variations, inferred selective events are often questionable, especially when there is no follow-up using a functional analysis to ensure the results [19]. It is a challenging and severe problem. After the integration of the candidates for recent positive selection obtained from nine genome-wide scans in humans, Joshua Akey (2009) found that only 14% of loci (722/5110) are identified in two or more studies. We should be highly aware that the confounding effect of

This article is dedicated to the Special Collection of Recent Advances in Next-Generation Bioinformatics (Ed. Xuegong Zhang).



**Figure 1. The increasing interest in detecting recent positive selection.** Times cited are from the year of the publication of three key methodological papers until the end of 2014. Data were collated from ISI Web of Knowledge.

demography could be one of the major reasons for 86% of the unique candidates. This has consequences. Because the false positive rate is high, a considerably large amount of resources will be wasted in the following functional analysis. This may prevent us from thoroughly understanding the mechanisms of adaptive evolution.

There are a number of reviews about the methodologies for detecting recent positive selection in the genome of living organisms [20–30]. Here we review the advantages and disadvantages of the different techniques used in model species, instead of focusing on the methodologies themselves. We are disappointed to find that, to reliably detect recent positive selection, there are few available methods that are suitable for the mostly untouched non-model species. The rapid advancement of next-generation sequencing technology has made it possible to re-sequence multiple individuals from non-model species. Therefore, it is crucial for us to develop statistical tests that can reliably detect recent positive selection in the absence of knowledge of the demography and population structure.

### TESTS BASED ON SNP OR MICROSATELLITE POLYMORPHISM DATA FROM A SINGLE LOCUS

When positive selection occurs recently, it often creates a star-like tree. That is, all of the lineages coalesce during a very short time period. This results in an excess of rare derived mutations, compared to the standard neutral model. This has been used as the signature of selection in the Tajima's  $D$  test [8], the Fu & Li's  $D$  test [16] and the  $E$  test [18]. These statistical tests all have the form:

$$\frac{\hat{\theta}_1 - \hat{\theta}_2}{\sqrt{\text{Var}(\hat{\theta}_1 - \hat{\theta}_2)}}$$

where  $\hat{\theta}_1$  and  $\hat{\theta}_2$  are different unbiased estimators of the population mutation rate  $\theta$ , and an excess of rare derived mutations has distinct effects on different estimators. Here  $\hat{\theta}_1$  is equal to  $\hat{\theta}_2$  under the standard neutral model, so that these summary statistics tend to be 0.  $\hat{\theta}_1$  tends to be smaller than  $\hat{\theta}_2$  when positive selection occurs. Thus these summary statistics are expected to be negative under selection. However, population size expansion, bottlenecks and population structure can also create a star-like tree and lead to the excess of rare derived mutations. This results in a high false positive rate of the three tests [16,18,31].

Recent positive selection creates not only an excess of rare derived mutations but also an excess of high-frequency-derived mutations. The latter case has been treated as the signature of selection in the Fay & Wu's  $H$  test [10]. However, it has been demonstrated that the Fay & Wu's  $H$  test is sensitive to a number of neutral scenarios, such as bottlenecks and population structure [32]. Moreover, the excess of high-frequency-derived mutations may result from unequal sample sizes between different subpopulations in a hidden structured population [33], which was also confirmed by our coalescent-based simulations [34] using the  $ms$  program [35]. When two subpopulations are partially isolated ( $4N\mu=5$ ,  $4Nm=0.001$ , where  $N$  is the effective population size,  $\mu$  is the mutation rate and  $m$  is the migration rate) and when 80 and 20 chromosomes are sampled from the sub-

populations, respectively, the false positive rate of the Fay & Wu's  $H$  test is 0.976. Therefore, to avoid the false inference of positive selection due to unequal sampling size in the hidden structured population, it is important to equalize the sample sizes from each sampling location.

The tests mentioned above are based on the mutation frequency spectrum, and a generalized form of these tests has been described [36]. Generally, the distributions of the summary statistics in those tests are obtained under the standard neutral model. Thus, the violations of the standard neutral assumption may cause false positives when detecting selection. It has been suggested that to reduce the number of false positives, one should use the Tajima's  $D$  test and the Fay & Wu's  $H$  test jointly [18] because those two tests detect different signals of positive selection.

A statistical test based on tree topology (the MFDM test, named after the maximum frequency of derived mutations) has been proven analytically and empirically under conditions that are free from the confounding impact of demography [37]. This can be achieved because the varying population size does not affect the probability of tree topology while selection does. The probability of having a more or an equally unbalanced tree under the neutral hypothesis [37] is

$$p(\psi \geq x) = \begin{cases} \frac{2(n-x)}{n-1}, & \text{if } \frac{n}{2} < x \leq n-1 \\ 1, & \text{if } x = \frac{n}{2} \end{cases},$$

where  $n$  is the number of leaves of the genealogical tree,  $x = \max\{\psi_1, \psi_2\}$ ,  $\psi_1$  and  $\psi_2$  are the size of left- and right-basal branches.

As an alternative to SNPs, microsatellites or simple/short tandem repeats are useful markers to examine traces of positive selection in genome. To date, there are only three methods based on single microsatellite locus data to detect the departure from the standard neutral model. The first test by Cornuet and Luikart [38], which compares the observed and the expected heterozygosity, is designed to detect population bottlenecks. Another test by Schlötterer *et al.* [39] uses the number of alleles at a microsatellite locus and determines whether an excess of the number of alleles is due to positive selection. However, as the authors have noted, the test needs a reliable locus-specific estimate of the scaled mutation rate,  $\theta = 4N\mu$ , that is usually difficult to obtain. Moreover, its sensitivity to population demography is still unknown. Recently, it has been found that an unbalanced allele frequency spectrum could usually be observed under the hitchhiking model, and thus, positive selection can be detected when comparing the polymorphism pattern within the locus. It avoids the difficulty in estimating the mutation rate of microsatellite loci and is robust with the wide range of demography parameters [40].

## MULTIPLE-LOCUS TESTS TO IDENTIFY SELECTION

Instead of using the DNA polymorphism data from a single locus, polymorphism data may also be collected from the surrounding loci. When a beneficial mutation goes to fixation (i.e., its allele frequency increases to 1.0), the process will not only change the mutation frequency spectrum at the surrounding loci, but will also reduce their polymorphisms. This locally reduced polymorphism due to positive selection is called selective sweep. Two likelihood ratio tests have been proposed to detect local reductions of nucleotide variation along a recombining chromosome [13,15]. By contrasting the standard neutral model and the hitchhiking model for recent positive selection, the likelihood ratio tests calculate the ratio of the maximum likelihoods between the two models (hypotheses) and make a choice between the two models. However, it is important to realize that these tests are based on the standard neutral model. The violation of the assumptions of the standard neutral model, for example bottlenecks, may influence the results and favor the hitchhiking model [32,41].

To distinguish between selective sweep and demography, a goodness-of-fit (GOF) test has been proposed [32] for instances when the standard neutral model is rejected by the KS composite likelihood ratio test [13]. The null hypothesis of the GOF approach is that the data are drawn from a selective sweep model, and the alternative hypothesis claims that the data are not drawn from a selective sweep scenario. However, simulations suggest that the false positive rate of the GOF approach and the value of the cut-off still depend on the demographic scenario [23]. Thus, results obtained using the GOF test should be interpreted carefully. Moreover, the local reductions of nucleotide variation along a chromosome can be caused by a low mutation rate. Therefore, when detecting selective sweeps, it is important to avoid mutation rate mis-specification by estimating the local mutation rate from divergence [42].

Interestingly, a strong linkage disequilibrium (LD) can also be generated in regions at both sides of the location of beneficial mutation, but a lack of LD is expected across the two sides [43–45]. Therefore, besides using the spatial distribution polymorphism, Kim and Nielsen [46] extended the previous KS composite likelihood ratio test to incorporate information regarding LD. However, the improvement made by including the LD is rather small, suggesting that most of the relevant information regarding selective sweeps may be captured by the spatial distribution and the allele frequencies of polymorphisms. A neutrality test based on information regarding the LD has been proposed to detect positive selection [47]. Simulations have shown that the power to detect strong

selection is approximately 50%–60%.

## GENOME-WIDE APPROACHES BASED ON SNP DATA OR MICROSATELLITE POLYMORPHISM DATA

At least twenty-two genome-wide scans for recent and ongoing positive selection have been performed in humans [48–58]; see also [59,60]. These genome data sets have enhanced our understanding of recent positive selection in human evolutionary history. For example, more than 4,000 genes have been found with a signature of positive selection in humans from 16 genome-wide analyses [60]. The whole-genome sequence analyses have suggested that approximately 1%–4% of non-African genomes and approximately 4%–6% of Melanesian genomes are derived from the archaic hominins Neanderthals and Denisovans, respectively [57,58]. *EPAS1*, a hypoxia pathway gene, has the strongest signature of positive selection in Tibetans, which has linked *EPAS1* in Tibetans to an adaptation to a hypoxic environment. In addition, a high-frequency haplotype of *EPAS1* in Tibetans was derived from Denisovans [61].

A number of genomic scans were also carried out to map recent positive selection in several model species [23], such as in mice [62], *Drosophila* [21,42,63–66] and *Arabidopsis* [67,68]. Generally, these methods can be divided into two categories: the model-based approaches (i.e., neutrality tests) and the outlier approaches. Based on the empirical distribution of the summary statistics, outlier approaches will always identify 5% of the surveyed loci/regions as candidates for selection if the significance level of tests is set at 0.05. That means, under neutrality, no matter how the population size varied in the past, there will be only 5% of the loci that are falsely identified as candidates for selection (and all of the candidates are false positives). In contrast, under a recurrent adaptive evolution model (assuming all of the loci are affected by selection), the power of the outlier approaches to detect selection is as low as 5%.

### Tests based on the mutation frequency spectrum and selective sweeps

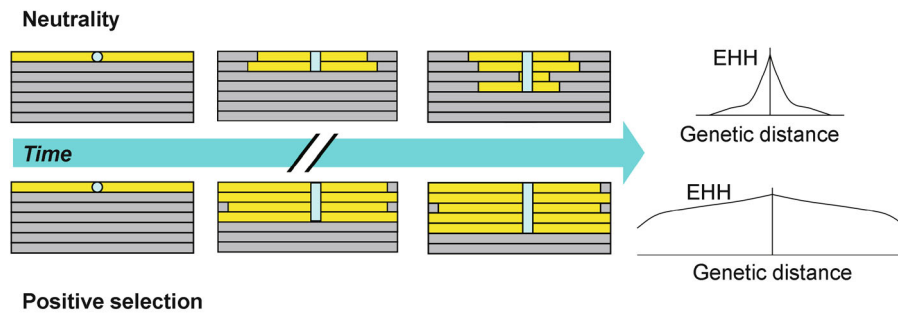
Although there is a confounding effect of demography, the expected mutation frequency spectrum at different loci remains the same because of the shared demographic history experienced at those loci. Selection, by contrast, can alter the mutation frequency spectrum at individual loci. Therefore, by analyzing Tajima's *D* statistic from 151 loci, the effects of demography and selection in human history may be disentangled by the outlier approach [69]. Another application of Tajima's *D* statistic is Carlson *et al.* [52]. Particularly, attention should also be

paid to the ascertainment bias of genotyped data when applying tests based on the mutation frequency spectrum [14]. The ascertainment bias is introduced by current methods of SNP discovery. SNPs identified in a very small sample are further typed into a large sample. Therefore, the frequency spectrum of genotyped SNPs is biased towards common or intermediate frequency variants. Such a bias bears significantly on the Tajima's *D* statistic. Therefore, several methods are proposed to correct this ascertainment bias [14,41]. The preciseness of these corrections may depend on the available documentation describing how each of the SNPs are initially identified. Moreover, it is known that the KS composite likelihood rate test [13] suffers from a high false positive rate due to the demography [32]. To solve this difficulty, Li and Stephan [42] describe a statistical method that detects footprints of selection in chromosome- or genome-wide data while taking fluctuations of the population size into account. After estimating the demography, a sliding window analysis is performed to find genomic regions that are affected by recent positive selection. In their likelihood ratio test, the null hypothesis assumes that the population has suffered an inferred bottleneck scenario, and that the alternative is that there is an inferred bottleneck scenario together with a selective sweep. Later simulations have shown that this strategy does well to control the false positive rate [23,41].

As an alternative, selective sweeps may be detected by a composite likelihood ratio test without inferring demography [14]. A modified method by Kim and Stephan [13] can be considered, which differs from the original method by using the background frequency instead of a standard neutral model to define the test statistic. However, it has been suggested that this approach is not robust for the bottleneck scenarios [23].

### Tests based on long haplotype structure

Another popular approach is known as the long haplotype-based tests, originally proposed by Sabeti *et al.* [17] (see also reviews by [22,24]). These tests make use of the haplotype profiles left by recombination events under partial selective sweeps. Considering that a beneficial mutation is rapidly rising in a population, recombination may not break down the association with nearby alleles (Figure 2). Therefore, the long-range associations (i.e., long haplotype structure) can be used to detect ongoing positive selection. The long-range associations can be simply measured by the extended haplotype homozygosity (EHH) [17], which shows that the information about the length of the haplotype block associated with the beneficial mutation. If the *N* chromosomes in a sample form *G* homozygous groups in the considered range, with each group *i* having



**Figure 2.** Illustration of how haplotype profile changes under neutrality. Under neutrality, the change of allele frequency of the new mutation (blue) takes a long time, so that by the time it reaches a high level, the original haplotype associated with it (the yellow bars) has already decayed substantially. Under positive selection, the integrality of the ancestral haplotype remains at a high level because the allele frequency increases very rapidly.

chromosomes, EHH is defined as [55]:

$$EHH = \frac{\sum_{i=1}^G \binom{n_i}{2}}{\binom{N}{2}}$$

Then, the EHH of the tested allele (with the core haplotype or chromosomes carrying the beneficial mutation) is contrasted against another allele within the same population. This ratio of EHH is then compared to simulated neutral sites or random genomic regions to determine the statistical significance [17].

Based on similar principle several tests were developed and applied to human genome-wide scans [53,70]. It has also been demonstrated that the integrated haplotype score (iHS) test is robust to multiple demographic scenarios [53]:

$$iHS = \ln \left( \frac{iHH_A}{iHH_D} \right),$$

where  $iHH_A$  stands for the integrated EHH of the ancestral allele, and  $iHH_D$  for that of derived allele. Nonetheless, iHS has a markedly low power for fixed or near-fixed selective sweeps because iHS relies on the comparison between the alternative alleles of a SNP site in the same population.

When the beneficial allele goes to fixation, the information required for the LRH and iHS tests vanishes together with the alternative allele. To solve this problem, two groups independently developed the lnRsb test [54] and the XP-EHH test [55], which are very similar in nature. In these new approaches, the EHH values for a certain SNP are compared between two populations [54,55]. The population under positive selection should

have a much higher EHH than another neutral evolving population. Therefore, both methods exhibit high power to detect selection when the beneficial allele is fixed or near fixation.

However, there are still several concerns for the long haplotype based tests. First, these tests are based on a heuristic idea rather than on stringent mathematical derivation. The EHH related statistics are functions of the recombination rate, the mutation rate and demography. There is however a lack of effort to fully understand the EHH related statistics. Secondly, as for many other neutrality tests, the EHH related tests encounter the multiple-test problem of large-scale data. For example, 51 human populations have been surveyed and each population has been genotyped for ~650k SNPs [56]. In such cases, it is difficult to determine the cutoffs. Thirdly, most of the EHH related tests use an empirical  $p$ -value, which derives from ranking the value of the candidate site against the empirical distribution (i.e., the outlier approach) because of the difficulty associated with obtaining the distribution of the EHH related statistics under actual demography. Like other outlier approaches, it only tells us how “extreme” a SNP is compared to the other SNPs across the genome. Regardless, this approach successfully reveals a number of well-known adaptive evolving genes, such as the *LCT* gene that affects lactose tolerance in adulthood, the *SLC24A5* gene that is a major determinant of skin pigmentation, and the *CYP3A* gene that is a major drug-metabolizer in the liver [5,71,72]. Finally, compared to the selective sweep approaches and the tests based on the mutation frequency spectrum, the long haplotype based tests are relatively less affected by the ascertainment bias because the haplotype structure may not strongly depend on the allele frequency. However, the long haplotype based tests require densely distributed SNPs across a whole genome or chromosome,

which may not be easily achieved in non-model species.

### Tests based on population differentiation, *Fst*

Adaptation can also be identified from high differences in allele frequencies among different populations, summarized by an estimate of *Fst* [73,74]. A Bayesian approach has been developed that can identify loci subject to positive selection when the selection coefficient is at least five times the migration rate [75]. However, the problem of identifying statistically significant departures from neutrality is complicated by the confounding effects of the distribution of *Fst* estimates [76]. Moreover, because background selection in a structured population influences *Fst* [77], it is hard to distinguish recent hitchhiking events from background selection [78]. Thus, attention should be paid to whether the neutral assumption has been severely violated. To address this problem, an outlier approach is proposed to use loci with extremely high *Fst* as candidates for selection [48,50]. Another concern about the *Fst* approach is the substantial heterogeneity among *Fst* values. It has been argued that *Fst* estimates of individual SNPs are too variable to be used as indicators of positive selection [79]. Thus, it has been suggested that using 5-MB windows to smooth out the very high variation in *Fst* at individual SNPs in human genomic scans [79].

### lnRV and lnRH tests

Based on microsatellite polymorphism data, lnRV and lnRH tests are suitable methods to detect selective sweeps. Given that the microsatellite mutation rates differ substantially among loci [80,81], it is always difficult to directly compare genetic diversity levels among loci. This problem can be circumvented by calculating the ratio of variability in two populations, which is believed to be independent of the mutation rate [21,82]. The ratio of variability in both populations (RV or RH) provides an estimator with the same expectation for neutral loci. Here, if the microsatellite variability is estimated from the variance in repeat number, the ratio of variability is named as RV. If the microsatellite variability is measured by the expected heterozygosity, the ratio of variability is called RH. Then, the logarithm of the estimators (lnRV and lnRH) will be highly positive or negative when the locus evolves adaptively in one population but neutrally in another. By further using a normalized empirical distribution (i.e., outlier-like approach) of lnRV or lnRH, the false positive rate due to demography can be controlled. The candidates for selection will be the outliers that have high variability in one population, but low variability in another. To use these approaches, multiple or genome-wide independent microsatellite

polymorphism data are needed to calculate the standard deviation of lnRV and lnRH among loci. To analyze the polymorphism data from partially linked microsatellite loci, an extended lnRH has been developed [83].

### Joint analysis of multiple summary statistics

To reduce false positives, it is straight-forward to use multiple signals to detect selection. In their pioneering work, Zeng and his colleagues [18] proposed the joint use of the Tajima's *D* and the Fay & Wu's *H* because those two summary statistics rely on different signatures of selection. This work has been extended to analyze the human genome-wide polymorphism data [84,85]. However, when considering joint use of the Tajima's *D* and the Fay & Wu's *H*, this strategy does work although it cannot completely avoid a high false positive rate [37]. We further examined whether, to detect selection, combining more signals is helpful or not. After combining four different signals (i.e., the excess of rare derived mutations, the excess of high-frequency derived mutations, selective sweep, and the long-haplotype structure), the false positive rate is still high in some bottlenecked cases if the demography history has been ignored in the analysis [86]. This provides a warning sign that the issue of high false positives may not be completely solved by a combination of different summary statistics, although the usage of different summary statistics has been seen in recent publications [85,87].

### SOME CONFOUNDING EFFECTS

The confounding effect of mutation rate heterogeneity has been discussed extensively in the literature. It has been suggested that the estimate of the time to the most recent common ancestor of the sample is highly dependent on the mutation model used [88], and that the heterogeneity of mutation rates appears to have the same effect on the distribution of observed pairwise differences (or mismatch distribution) as a population size expansion [89]. A method detecting recent selective sweeps in the context of the mutation rate and background selection was developed recently [90].

It has been suggested that the genomic variants subject to recent positive selection have had less of an opportunity to be affected by recombination. Thus, the two processes have an intimately related impact on genetic variation, and inference of either may be vulnerable to confounding by the other [91]. This argument has been further supported by the observation that genome-wide scans that detect loci under recent selection in humans have tended to highlight loci in regions of low recombination [91]. Moreover, because the pattern of LD at the selected locus may sometimes be

complex [44,45], simulations have shown that recent positive selection can create false hotspots of recombination [92].

The EHH based methods rely on the measurement of the LD, therefore the recombination rate variation across a genome may introduce bias. In view of this, most available methods take specific steps to remove the impact of variable recombination rates. This is usually conducted in two ways: the first is to contrast the EHH values of the same locus between two alleles, thereby canceling out the factor of local recombination rate [53–55]; the second is to use genomic distance instead of physical distance [53,55]. Such corrections rely on the assumption that individuals or populations share the same recombination rate across the genome. It should be noted that the EHH based method is not sensitive to fine-scale recombination rate variations because the extension of a haplotype is required to be up to hundreds of kilo-bases to be considered indicative of positive selection.

## CONCLUSIONS

Selective events often leave their footprints behind them. However, it is always difficult to distinguish these footprints from the confounding signatures of demography, such as population size expansion and bottleneck [93,94]. According to our work, the most reliable approach to detect selection, besides functional analysis, may be a genome-wide approach. However, a genome-wide approach and a functional analysis may not be easily applied to millions of non-model species. Specifically, even if the genome-wide polymorphism data are available, it does not guarantee that researchers can reliably detect the recent selective events with the available methodologies. For example, twenty-two genome-wide scans have been performed to detect recent positive selection in humans. However, the lists of candidates for selection are incompatible within these studies. The map of positive selection across nine genome-wide scans is integrated [59]. In total, 5,110 distinct regions were identified in one or more studies. These regions encompass 409 Mb of the sequence (14% of the human genome). Strikingly, only 722 regions (14.1%) were identified in two or more studies. Therefore, this suggests that attention should now be turned towards understanding the biological relevance and adaptive significance of the regions identified as being subject to recent positive selection [19,59,95].

The issue of false acceptance of the adaptation hypothesis cannot be over-emphasized because we still have little knowledge about the demography and the molecular mechanism of adaptation in most of species. Thus, how can we detect recent positive selection in millions of natural species? How can we study the

molecular mechanism of adaptation to recent environmental changes? Although we are in the age of genomics, most of the available DNA polymorphism data from those species are still from a single locus or a small number of loci. However, it is known that, based on the single locus data, the available methods, including the Tajima's  $D$  test and the Fay & Wu's  $H$  test, cannot reliably detect recent positive selection in varying size populations. To solve this problem, one approach is to utilize a *priori* information about the demography of the species to help researchers choosing an appropriate test [18].

It is clear that a large amount of data is needed to estimate the parameters of demography, and the estimation procedure demands the computational power. For example, based on the genome-wide joint mutation frequency spectrum (or multi-population allele frequency spectrum) between the African *Drosophila* population and the European *Drosophila* population, the parameters of the bottleneck for the European population have been estimated [42]. Later studies have demonstrated that it is efficient to fit the demographic models to the joint mutation frequency spectrum [96–98].

Therefore, considering the low efficiency of combining neutrality tests with functional analysis, we would suggest that priority should be given to develop more reliable tests that detect selection. It should be noted that, when a new test is proposed, it is always difficult or even impossible to investigate all of the possible neutral scenarios by simulations because of the intricate nature of evolutionary history in natural populations. This may be one of the reasons why several papers have claimed that their tests are specific for detecting recent positive selection, but later studies with wide range simulations demonstrate that these tests are still sensitive to certain neutral scenarios. Here, to help investigate the robustness of new methods more efficiently in the future, we would summarize two neutral scenarios that often lead to false positives when detecting selection. First, the bottleneck scenarios with severity approximately 1.0, where the severity is equal to  $t_1 N_0 / N_1$ ,  $N_0$  is the current effective population size,  $N_1$  is the size during the bottleneck, and  $t_1$  is the duration of the bottleneck. The time is scaled such that one unit represents  $2N_0$  generations. Under bottlenecks, the variance of the summary statistics, such as Tajima's  $D$ , Watterson's  $\theta$  and the mean number of nucleotide differences between two sequences ( $\pi$ ), can be increased, especially when the severity of the bottleneck is approximately 1.0. Second, the population structure with different sampling schemes should be examined carefully. A reasonable range of the migration parameter  $Nm$  is 0.001–10.

Future studies should be aimed at improving the robustness of tests with respect to demography. Ideally, new statistical tests should not only be free from the

confounding impact of both demography and population structure, but should also have reasonable power to detect selection. Due to obvious practical reasons, they should rely on single locus DNA polymorphism data. However, it would be a great challenge for all who are working in this field to develop such tests. Recently, the (joint) analysis of demography and selection [25,26,42,90, 93,99,100] has become a central issue, and it is believed that we are on the right track. However, researchers should be made aware that simulations have shown that the popularly used *SweepFinder* is not robust to certain demographics, for example, severe and recent bottlenecks [23].

## ACKNOWLEDGEMENTS

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB13040800), the National Natural Science Foundation of China (No. 91531306) and the 973 project (No. 2012CB316505). We thank three anonymous reviewers for their insightful suggestions to improve the manuscript.

## COMPLIANCE WITH ETHICS GUIDELINES

The authors Jinggong Xiang-Yu, Zongfeng Yang, Kun Tang and Haipeng Li declare that they have no conflict of interests.

This article does not contain any studies with human or animal subjects performed by any of the authors.

## REFERENCES

- Pulvers, J. N., Journiac, N., Arai, Y., and Nardelli, J. (2015) MCPH1: a window into brain development and evolution. *Front. Cell. Neurosci.*, 10.3389/fncel.2015.00092
- Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S. L., Wiebe, V., Kitano, T., Monaco, A. P. and Pääbo, S. (2002) Molecular evolution of *FOXP2*, a gene involved in speech and language. *Nature*, 418, 869–872
- Swallow, D. M. (2003) Genetics of lactase persistence and lactose intolerance. *Annu. Rev. Genet.*, 37, 197–219
- Poulter, M., Hollox, E., Harvey, C. B., Mulcare, C., Peuhkuri, K., Kajander, K., Sarner, M., Korpela, R. and Swallow, D. M. (2003) The causal element for the lactase persistence/non-persistence polymorphism is located in a 1 Mb region of linkage disequilibrium in Europeans. *Ann. Hum. Genet.*, 67, 298–311
- Bersaglieri, T., Sabeti, P. C., Patterson, N., Vanderploeg, T., Schaffner, S. F., Drake, J. A., Rhodes, M., Reich, D. E. and Hirschhorn, J. N. (2004) Genetic signatures of strong recent positive selection at the lactase gene. *Am. J. Hum. Genet.*, 74, 1111–1120
- Nielsen, R. (2009) Adaptionism-30 years after Gould and Lewontin. *Evolution*, 63, 2487–2490
- Hurst, L. D. (2009) Fundamental concepts in genetics: genetics and the understanding of selection. *Nat. Rev. Genet.*, 10, 83–93
- Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123, 585–595
- Fu, Y.-X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, 147, 915–925
- Fay, J. C. and Wu, C.-I. (2000) Hitchhiking under positive Darwinian selection. *Genetics*, 155, 1405–1413
- Smith, J. M. and Haigh, J. (1974) The hitch-hiking effect of a favourable gene. *Genet. Res.*, 23, 23–35
- Galtier, N., Depaulis, F. and Barton, N. H. (2000) Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. *Genetics*, 155, 981–987
- Kim, Y. and Stephan, W. (2002) Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics*, 160, 765–777
- Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G. and Bustamante, C. (2005) Genomic scans for selective sweeps using SNP data. *Genome Res.*, 15, 1566–1575
- Li, H. and Stephan, W. (2005) Maximum-likelihood methods for detecting recent positive selection and localizing the selected site in the genome. *Genetics*, 171, 377–384
- Fu, Y.-X. and Li, W.-H. (1993) Statistical tests of neutrality of mutations. *Genetics*, 133, 693–709
- Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z. P., Richter, D. J., Schaffner, S. F., Gabriel, S. B., Platko, J. V., Patterson, N. J., McDonald, G. J., et al. (2002) Detecting recent positive selection in the human genome from haplotype structure. *Nature*, 419, 832–837
- Zeng, K., Fu, Y.-X., Shi, S. and Wu, C.-I. (2006) Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics*, 174, 1431–1439
- MacCallum, C. and Hill, E. (2006) Being positive about selection. *PLoS Biol.*, 4, e87
- Bamshad, M. and Wooding, S. P. (2003) Signatures of natural selection in the human genome. *Nat. Rev. Genet.*, 4, 99–111
- Kauer, M. O., Dieringer, D. and Schlötterer, C. (2003) A microsatellite variability screen for positive selection associated with the “out of Africa” habitat expansion of *Drosophila melanogaster*. *Genetics*, 165, 1137–1148
- Sabeti, P. C., Schaffner, S. F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., Palma, A., Mikkelsen, T. S., Altshuler, D. and Lander, E. S. (2006) Positive natural selection in the human lineage. *Science*, 312, 1614–1620
- Pavlidis, P., Hutter, S. and Stephan, W. (2008) A population genomic approach to map recent positive selection in model species. *Mol. Ecol.*, 17, 3585–3598
- Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C. and Clark, A. G. (2007) Recent and ongoing selection in the human genome. *Nat. Rev. Genet.*, 8, 857–868
- Vitti, J. J., Grossman, S. R. and Sabeti, P. C. (2013) Detecting natural selection in genomic data. *Annu. Rev. Genet.*, 47, 97–120
- Bank, C., Ewing, G. B., Ferrer-Admetlla, A., Foll, M. and Jensen, J. D. (2014) Thinking too positive? Revisiting current methods of population genetic selection inference. *Trends Genet.*, 30, 540–546
- Pool, J. E., Hellmann, I., Jensen, J. D. and Nielsen, R. (2010) Population genetic inference from genomic sequence variation. *Genome Res.*, 20, 291–300
- Chen, H., Patterson, N. and Reich, D. (2010) Population differentiation as a test for selective sweeps. *Genome Res.*, 20, 393–402
- Karlsson, E. K., Kwiatkowski, D. P. and Sabeti, P. C. (2014) Natural selection and infectious disease in human populations. *Nat. Rev. Genet.*, 15, 379–393
- Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., Harney, E., Stewardson, K., Fernandes, D.,

- Novak, M., *et al.* (2015) Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*, 528, 499–503
31. Tajima, F. (1989) The effect of change in population size on DNA polymorphism. *Genetics*, 123, 597–601
  32. Jensen, J. D., Kim, Y., DuMont, V. B., Aquadro, C. F. and Bustamante, C. D. (2005) Distinguishing between selective sweeps and demography using DNA polymorphism data. *Genetics*, 170, 1401–1410
  33. Przeworski, M. (2002) The signature of positive selection at randomly chosen loci. *Genetics*, 160, 1179–1189
  34. Hudson, R. R. (1990) Gene genealogies and the coalescent process. In *Oxford Surveys in Evolutionary Biology*. Vol. 7, D. Futuyma and J. Antonovics, Editors. 1–44 New York: Oxford University Press
  35. Hudson, R. R. (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*, 18, 337–338
  36. Achaz, G. (2009) Frequency spectrum neutrality tests: one for all and all for one. *Genetics*, 183, 249–258
  37. Li, H. (2011) A new test for detecting recent positive selection that is free from the confounding impacts of demography. *Mol. Biol. Evol.*, 28, 365–375
  38. Cornuet, J. M. and Luikart, G. (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, 144, 2001–2014
  39. Schlötterer, C., Kauer, M. and Dieringer, D. (2004) Allele excess at neutrally evolving microsatellites and the implications for tests of neutrality. *Proc. Biol. Sci.*, 271, 869–874
  40. Li, H. and Wiehe, T. (2013) Coalescent tree imbalance and a simple test for selective sweeps based on microsatellite variation. *PLoS Comput. Biol.*, 9, e1003060
  41. Thornton, K. R. and Jensen, J. D. (2007) Controlling the false-positive rate in multilocus genome scans for selection. *Genetics*, 175, 737–750
  42. Li, H. and Stephan, W. (2006) Inferring the demographic history and rate of adaptive substitution in *Drosophila*. *PLoS Genet.*, 2, e166
  43. Parsch, J., Meiklejohn, C. D. and Hartl, D. L. (2001) Patterns of DNA sequence variation suggest the recent action of positive selection in the janus-ocnus region of *Drosophila simulans*. *Genetics*, 159, 647–657
  44. Stephan, W., Song, Y. S. and Langley, C. H. (2006) The hitchhiking effect on linkage disequilibrium between linked neutral loci. *Genetics*, 172, 2647–2663
  45. McVean, G. (2007) The structure of linkage disequilibrium around a selective sweep. *Genetics*, 175, 1395–1406
  46. Kim, Y. and Nielsen, R. (2004) Linkage disequilibrium as a signature of selective sweeps. *Genetics*, 167, 1513–1524
  47. Jensen, J. D., Thornton, K. R., Bustamante, C. D. and Aquadro, C. F. (2007) On the utility of linkage disequilibrium as a statistic for identifying targets of positive selection in nonequilibrium populations. *Genetics*, 176, 2371–2379
  48. Akey, J. M., Zhang, G., Zhang, K., Jin, L. and Shriver, M. D. (2002) Interrogating a high-density SNP map for signatures of natural selection. *Genome Res.*, 12, 1805–1814
  49. Pickrell, J. K., Coop, G., Novembre, J., Kudaravalli, S., Li, J. Z., Absher, D., Srinivasan, B. S., Barsh, G. S., Myers, R. M., Feldman, M. W., *et al.* (2009) Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.*, 19, 826–837
  50. Kayser, M., Brauer, S. and Stoneking, M. (2003) A genome scan to detect candidate regions influenced by local natural selection in human populations. *Mol. Biol. Evol.*, 20, 893–900
  51. Storz, J. F., Payseur, B. A. and Nachman, M. W. (2004) Genome scans of DNA variability in humans reveal evidence for selective sweeps outside of Africa. *Mol. Biol. Evol.*, 21, 1800–1811
  52. Carlson, C. S., Thomas, D. J., Eberle, M. A., Swanson, J. E., Livingston, R. J., Rieder, M. J. and Nickerson, D. A. (2005) Genomic regions exhibiting positive selection identified from dense genotype data. *Genome Res.*, 15, 1553–1565
  53. Voight, B. F., Kudravalli, S., Wen, X. and Pritchard, J. K. (2006) A map of recent positive selection in the human genome. *PLoS Biol.*, 4, e72
  54. Tang, K., Thornton, K. R. and Stoneking, M. (2007) A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol.*, 5, e171
  55. Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E. H., McCarroll, S. A., Gaudet, R., *et al.*, (2007) Genome-wide detection and characterization of positive selection in human populations. *Nature*, 449, 913–918
  56. Li, J. Z., Absher, D. M., Tang, H., Southwick, A. M., Casto, A. M., Ramachandran, S., Cann, H. M., Barsh, G. S., Feldman, M., Cavalli-Sforza, L. L., *et al.* (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science*, 319, 1100–1104
  57. Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H. Y., *et al.* (2010) A draft sequence of the Neandertal genome. *Science*, 328, 710–722
  58. Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., Viola, B., Briggs, A. W., Stenzel, U., Johnson, P. L. F., *et al.* (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, 468, 1053–1060
  59. Akey, J. M. (2009) Constructing genomic maps of positive selection in humans: where do we go from here? *Genome Res.*, 19, 711–722
  60. Fu, W. Q. and Akey, J. M. (2013) Selection and adaptation in the human genome. *Annu. Rev. Genom. Hum. G.*, 14, 467–489
  61. Huerta-Sánchez, E., Jin, X., Asan, , Bianba, Z., Peter, B. M., Vinckenbosch, N., Liang, Y., Yi, X., He, M., Somel, M., *et al.* (2014) Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*, 512, 194–197
  62. Teschke, M., Mukabayire, O., Wiehe, T. and Tautz, D. (2008) Identification of selective sweeps in closely related populations of the house mouse based on microsatellite scans. *Genetics*, 180, 1537–1545
  63. Glinka, S., Ometto, L., Mousset, S., Stephan, W. and De Lorenzo, D. (2003) Demography and natural selection have shaped genetic variation in *Drosophila melanogaster*: a multi-locus approach. *Genetics*, 165, 1269–1278
  64. Ometto, L., Glinka, S., De Lorenzo, D. and Stephan, W. (2005) Inferring the effects of demography and selection on *Drosophila melanogaster* populations from a chromosome-wide scan of DNA variation. *Mol. Biol. Evol.*, 22, 2119–2130
  65. Emerson, J. J., Cardoso-Moreira, M., Borevitz, J. O. and Long, M. (2008) Natural selection shapes genome-wide patterns of copy-number polymorphism in *Drosophila melanogaster*. *Science*, 320, 1629–1631
  66. Pavlidis, P., Jensen, J. D., Stephan, W. and Stamatakis, A. (2012) A critical assessment of storytelling: gene ontology categories and the importance of validating genomic scans. *Mol. Biol. Evol.*, 29, 3237–3248
  67. Schmid, K. J., Ramos-Onsins, S., Ringys-Beckstein, H., Weisshaar, B. and Mitchell-Olds, T. (2005) A multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure from a neutral

- model of DNA sequence polymorphism. *Genetics*, 169, 1601–1615
68. Borevitz, J. O., Hazen, S. P., Michael, T. P., Morris, G. P., Baxter, I. R., Hu, T. T., Chen, H., Werner, J. D., Nordborg, M., Salt, D. E., *et al.* (2007) Genome-wide patterns of single-feature polymorphism in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA*, 104, 12057–12062
  69. Stajich, J. E. and Hahn, M. W. (2005) Disentangling the effects of demography and selection in human history. *Mol. Biol. Evol.*, 22, 63–73
  70. Wang, E. T., Kodama, G., Baldi, P. and Moyzis, R. K. (2006) Global landscape of recent inferred Darwinian selection for *Homo sapiens*. *Proc. Natl. Acad. Sci. USA*, 103, 135–140
  71. Kuehl, P., Zhang, J., Lin, Y., Lamba, J., Assem, M., Schuetz, J., Watkins, P. B., Daly, A., Wrighton, S. A., Hall, S. D., *et al.* (2001) Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A5 expression. *Nat. Genet.*, 27, 383–391
  72. Lamason, R. L., Mohideen, M. A., Mest, J. R., Wong, A. C., Norton, H. L., Aros, M. C., Jurynec, M. J., Mao, X., Humphreville, V. R., Humbert, J. E., *et al.* (2005) SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science*, 310, 1782–1786
  73. Lewontin, R. C. and Krakauer, J. (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, 74, 175–195
  74. Beaumont, M. and Nichols, R. A. (1996) Evaluating loci for use in the genetic analysis of population structure. *Proc. R. Soc. Lond. B Biol. Sci.*, 263, 1619–1626
  75. Beaumont, M. A. and Balding, D. J. (2004) Identifying adaptive genetic divergence among populations from genome scans. *Mol. Ecol.*, 13, 969–980
  76. Nei, M. and Maruyama, T. (1975) Letters to the editors: Lewontin-Krakauer test for neutral genes. *Genetics*, 80, 395
  77. Charlesworth, B., Nordborg, M. and Charlesworth, D. (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet. Res.*, 70, 155–174
  78. Stephan, W., Xing, L., Kirby, D. A. and Braverman, J. M. (1998) A test of the background selection hypothesis based on nucleotide data from *Drosophila ananassae*. *Proc. Natl. Acad. Sci. USA*, 95, 5649–5654
  79. Weir, B. S., Cardon, L. R., Anderson, A. D., Nielsen, D. M. and Hill, W. G. (2005) Measures of human population structure show heterogeneity among genomic regions. *Genome Res.*, 15, 1468–1476
  80. Di Rienzo, A., Donnelly, P., Toomajian, C., Sisk, B., Hill, A., Petzl-Erler, M. L., Haines, G. K. and Barch, D. H. (1998) Heterogeneity of microsatellite mutations within and between loci, and implications for human demographic histories. *Genetics*, 148, 1269–1284
  81. Harr, B., Zangerl, B., Brem, G. and Schlötterer, C. (1998) Conservation of locus-specific microsatellite variability across species: a comparison of two *Drosophila* sibling species, *D. melanogaster* and *D. simulans*. *Mol. Biol. Evol.*, 15, 176–184
  82. Schlötterer, C. (2002) A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics*, 160, 753–763
  83. Wiehe, T., Nolte, V., Zivkovic, D. and Schlötterer, C. (2007) Identification of selective sweeps using a dynamically adjusted number of linked microsatellites. *Genetics*, 175, 207–218
  84. Grossman, S. R., Shlyakhter, I., Karlsson, E. K., Byrne, E. H., Morales, S., Frieden, G., Hostetter, E., Angelino, E., Garber, M., Zuk, O., *et al.* (2010) A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science*, 327, 883–886
  85. Grossman, S. R., Andersen, K. G., Shlyakhter, I., Tabrizi, S., Winnicki, S., Yen, A., Park, D. J., Griesemer, D., Karlsson, E. K., Wong, S. H., *et al.* (2013) Identifying recent adaptations in large-scale genomic data. *Cell*, 152, 703–713
  86. Lin, K., Li, H., Schlötterer, C. and Futschik, A. (2011) Distinguishing positive selection from neutral evolution: boosting the performance of summary statistics. *Genetics*, 187, 229–244
  87. Pybus, M., Luisi, P., Dall'Olio, G., Uzkudun, M., Laayouni, H., Bertranpetit, J. and Engelken, J. (2015) Hierarchical boosting: a machine-learning framework to detect and classify hard selective sweeps in human populations. *Bioinformatics*, 31, 3946–3952.
  88. Markovtsova, L., Marjoram, P. and Tavaré, S. (2000) The effects of rate variation on ancestral inference in the coalescent. *Genetics*, 156, 1427–1436
  89. Aris-Brosou, S. and Excoffier, L. (1996) The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. *Mol. Biol. Evol.*, 13, 494–504
  90. Huber, C. D., DeGiorgio, M., Hellmann, I. and Nielsen, R. (2016) Detecting recent selective sweeps while controlling for mutation rate and background selection. *Mol. Ecol.*, 25, 142–156
  91. O'Reilly, P. F., Birney, E. and Balding, D. J. (2008) Confounding between recombination and selection, and the Ped/Pop method for detecting selection. *Genome Res.*, 18, 1304–1313
  92. Reed, F. A. and Tishkoff, S. A. (2006) Positive selection can create false hotspots of recombination. *Genetics*, 172, 2011–2014
  93. Li, J., Li, H., Jakobsson, M., Li, S., Sjödin, P. and Lascoux, M. (2012) Joint analysis of demography and selection in population genetics: where do we stand and where could we go? *Mol. Ecol.*, 21, 28–44
  94. Stephan, W. (2016) Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. *Mol. Ecol.*, 25, 79–88
  95. Kelley, J. L. and Swanson, W. J. (2008) Positive selection in the human genome: from genome scans to biological significance. *Annu. Rev. Genomics Hum. Genet.*, 9, 143–160
  96. Zhai, W., Nielsen, R. and Slatkin, M. (2009) An investigation of the statistical power of neutrality tests based on comparative and population genetic data. *Mol. Biol. Evol.*, 26, 273–283
  97. Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. and Bustamante, C. D. (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.*, 5, e1000695
  98. Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C. and Foll, M. (2013) Robust demographic inference from genomic and SNP data. *PLoS Genet.*, 9, e1003905
  99. Nielsen, R., Hubisz, M. J., Hellmann, I., Torgerson, D., Andrés, A. M., Albrechtsen, A., Gutenkunst, R., Adams, M. D., Cargill, M., Boyko, A., *et al.* (2009) Darwinian and demographic forces affecting human protein coding genes. *Genome Res.*, 19, 838–849
  100. Fijarczyk, A. and Babik, W. (2015) Detecting balancing selection in genomes: limits and prospects. *Mol. Ecol.*, 24, 3529–3545