LETTER

Scanning for the Signatures of Positive Selection for Human-Specific Insertions and Deletions

Chun-Hsi Chen,[*][1] *Trees-Juen Chuang,*[†][1] *Ben-Yang Liao,* and *Feng-Chi Chen*[*][‡][§]

*Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Zhunan, Miaoli County, Taiwan; †Genomics Research Center, Academia Sinica, Taipei, Taiwan; ‡Department of Life Science, National Chiao-Tung University, Hsinchu, Taiwan; and §Department of Dentistry, Chinese Medical University, Taichung, Taiwan

Human-specific small insertions and deletions (HS indels, with lengths <100 bp) are reported to be ubiquitous in the human genome. However, whether these indels contribute to human-specific traits remains unclear. Here we employ a modified McDonald–Kreitman (MK) test and a combinatorial population genetics approach to infer, respectively, the occurrence of positive selection and recent selective sweep events associated with HS indels. We first extract 625,890 HS indels from the human–chimpanzee–macaque–mouse multiple alignments and classify them into nonpolymorphic (41%) and polymorphic (59%) indels with reference to the human indel polymorphism data. The modified MK test is then applied to 100-kb partially overlapped sliding windows across the human genome to scan for the signs of positive selection. After excluding the possibility of biased gene conversion and controlling for false discovery rate, we show that HS indels are potentially positively selected in about 10 Mb of the human genome. Furthermore, the indel-associated positively selected regions overlap with genes more often than expected. However, our result suggests that the potential targets of positive selection are located in noncoding regions. Meanwhile, we also demonstrate that the genomic regions surrounding HS indels are more frequently involved in recent selective sweep than the other regions. In addition, HS indels are associated with distinct recent selective sweep events in different human subpopulations. Our results suggest that HS indels may have been associated with human adaptive changes at both the species level and the subpopulation level.

Surveys of human-specific changes in the genome give the most straightforward clues for what makes us human. Among these genetic changes, human-specific small insertions and deletions (<100 bp; designated as "HS indels") may associate with three possible mechanisms underlying human evolution, namely protein evolution, regulatory evolution (King and Wilson 1975), and "less-is-more" (i.e., the type of evolution in which loss of function increases the fitness of the affected individuals) (Li and Saunders 2005). Indeed, it has been shown that HS indels affect a large number of coding and potential regulatory regions (e.g., 5′ untranslated regions) (Chen et al. 2007). These indels might have been directly subject to positive selection, as mammalian *Catsper1* (Podlaha and Zhang 2003; Podlaha et al. 2005) and fruit fly *Acp26Aa* (Schully and Hellberg 2006) have experienced. Furthermore, indels have recently been suggested to increase the rate of nucleotide substitutions in their surrounding genomic regions (Tian et al. 2008). HS indels, as such, may also increase the number of human-specific substitutions. With the dual potential of disrupting–modifying functional elements and accelerating regional sequence evolution only in the human lineage, HS indels may have significant impacts on human evolution. However, the selection forces that act on these indels have not been systematically studied. We employ two complementary methods aiming to understand whether HS indels contribute to human adaptations. For relatively ancient adaptive events, we propose a new test, which is a modified version of the McDonald–Kreitman (MK) test (McDonald and Kreitman 1991) similar to the method of Podlaha

et al. (2005), to examine whether HS indels are subject to positive selection after the *Homo–Pan* divergence. Because there is clear evidence showing that human subpopulations have genetically adapted to their respective living environments, such as diet (Perry et al. 2007; Tishkoff et al. 2007), we also examined the association of HS indels with recent selective sweep events in three human subpopulations (African, Asian, and European).

Materials and Methods
Data Sources

Multiple alignments were downloaded from the University of California, Santa Cruz Genome Browser (UCSC) (http://hgdownload.cse.ucsc.edu/goldenPath/hg18/multiz28way/maf/). We focused on the genomes of human (hg18), chimpanzee (panTro2), Rhesus macaque (rheMac2), and mouse (mm8). The IDP (3,369,034 events) were integrated from the dbSNP (SNP129, http://hgdownload.cse.ucsc.edu/goldenPath/hg18/database/snp129.txt.gz) and two recently published human genomes (Levy et al. 2007; Wheeler et al. 2008), which accounted for another 581,158 events. To reduce potential sequencing and alignment errors, IDPs located in repeat-masked regions (annotated by RepeatMasker; Jurka et al. 2005) were excluded. Overall, 621,449 IDPs were analyzed (including 60,366 events from the Venter and Watson genomes; Levy et al. 2007; Wheeler et al. 2008).

The haplotype information used in the *DH* test (see below) was retrieved from HapMap Release 22. Three human subpopulations, including 60 Utah residents with Northern and Western European ancestry from the Centre d'Etude du Polymorphisme Humain collection, 60 Yoruba in Ibadan, Nigeria and 90 Japanese in Tokyo, Japan + Han Chinese in Beijing, China, were included in the HapMap project (http://hapmap.ncbi.nlm.nih.gov/downloads/phasing/2007-08_rel22/phased/).

Identification of HS Indels

Assuming all IDPs are independent, we generated as many human sequences as the number of IDPs. These human genomic sequences were generated by inserting or deleting the IDP sequences in the reference human genome in the UCSC multiple sequence alignments. The newly added sequences, together with the original multiple-species sequences, were then realigned using the MUSCLE package (Edgar 2004). HS indels were then identified by four-way comparisons of mammalian genomes (human, chimpanzee, rhesus macaque, and mouse) as previously suggested (Chen et al. 2007). Further, HS indels can be divided into nonpolymorphic and polymorphic by recognizing whether the HS indels are observed in the IDP-containing human genome sequences (see Supplementary Material online for more details).

The Modified MK Test

To explore the possibility of positive selection on HS indels, we employed a modified MK test similar to that previously proposed (Podlaha et al. 2005). The classical MK test (McDonald and Kreitman 1991) posits that, under neutral selection, the ratios of nonsynonymous-to-synonymous nucleotide substitutions should be the same for divergence (fixed changes) and diversity (polymorphisms). When a genomic region is subject to positive selection, divergence increases, whereas diversity decreases. As a result, the nonsynonymous-to-synonymous ratio of fixed substitutions should be larger than that of polymorphic substitutions. In the modified MK test, human-specific nucleotide substitutions are used as the neutral reference, whereas indels are used to substitute the nonsynonymous substitutions in the traditional MK test. It may be argued that small and large indels can have different effects on the modified MK test. Nevertheless, although ~91% of the indels are smaller than 10 bp, and ~99% are smaller than 30 bp (supplementary fig. S4, Supplementary Material online), the length variations of indels do not seem to be a major issue. We used 100-kb sliding windows that partially overlapped 50 kb with adjacent windows to perform the modified MK test. Two-by-two contingency tables were then established with nonpolymorphic and polymorphic human-specific indels and nucleotide substitutions. A window is considered as nontestable when any of the expected values of the contingency table is zero, for zero cannot be used as the expected value in the calculation of the χ^2 statistics. We further corrected the χ^2 statistics derived from the contingency tables of which one or more of the expected numbers are smaller than 5 as previously proposed (Hartl and Clark 2007).

Detection of Selective Sweep

The DH test (Zeng, Shi, et al. 2007) is a combination of the Tajima's D test (Tajima 1989) and Fay and Wu's H test (Fay et al. 2001) and is more robust than both of these tests in detection of positively selected regions. The DH test is particularly sensitive to high-frequency–derived SNPs. Therefore, it is suitable for de-tection of the co-occurrence of HS indels and high-frequency SNPs in positively selected regions. To perform the DH test, the test window must be first determined. We used the EHH algorithm (Sabeti et al. 2002) to search for windows that were centered on the target indel. Briefly, EHH calculates the haplotype homozygosity starting from the target site (indel in this study) and extends the calculation to either side (up- or downstream) of the target site. As the number of SNPs increases with the extension, the homozygosity decreases rapidly. In this study, the boundaries of the "EHH window" were set at the farthest SNPs where the haplotype homozygosity decreases to 0.05. In addition, the "EHH windows" were limited in 1-Mb regions to minimize the effects recombination. The number of SNPs in each window must exceed 50 for test accuracy. The DH test was then performed on all the available EHH windows surrounding HS indels using the program kindly provided by Kai Zeng with default parameters.

HS Indels in ~10 Mb of the Human Genome Are Positively Selected

To investigate the evolutionary forces imposed on HS indels, we first examined whether these indels are polymorphic in the human population, for positively selected indels are more likely fixed. Accordingly, we integrated the human indel polymorphisms (IDPs) from Single Nucleotide Polymorphism database (dbSNP) (build 129) and two recently published individual human genomes (Levy et al. 2007; Wheeler et al. 2008) into multiple sequence alignments (human, chimpanzee, rhesus macaque, and mouse) to differentiate polymorphic and nonpolymorphic HS indels (see supplementary fig. S1, Supplementary Material online, and Materials and Methods for more detail). To reduce potential alignment or sequencing errors, indels located in repeat-masked regions were excluded. We thus obtained 625,890 HS indels, of which 41.3% were nonpolymorphic (supplementary fig. S2, Supplementary Material online). Note that the percentage of nonpolymorphic HS indels may be overestimated because some polymorphic indels may be misclassified as nonpolymorphic indels due to insufficient sampling. However, the "real" fixed HS indels should be included in the currently identified nonpolymorphic events. Furthermore, as we will discuss later, our estimate of positively selected regions is actually conservative. The nonpolymorphic HS indels were subsequently analyzed for possible association with positive selection.

Because the standard tests for positive selection (such as the d_N/d_S ratio test (Yang and Bielawski 2000) and the MK test; McDonald and Kreitman 1991) cannot be readily applied to the analysis of indels, we modified the MK test to examine whether the ratio of nonpolymorphic to polymorphic HS indels significantly departs from the neutral expectation, assuming that most of the human-specific substitutions are selectively neutral (see Materials and Methods). This assumption is reasonable because most of the genomic regions are noncoding, and more than 99% of the substitutions in our data set are located in noncoding regions. To evaluate the applicability of this

Table 1
The R Values in Different Polymorphism Data Sets

Data source	Nonpolymorphic Indels	Polymorphic Indels	R_{ID}[a]	Nonpolymorphic Substitutions	Polymorphic Substitutions	R_{NT}[a]
dbSNP	119,353	161,470	0.74	962,193	1,107,124	0.87
Seattle + NIEHS	3,466	6,623	0.52	22,520	50,552	0.45
Seattle SNPs	723	2,224	0.33	5,070	12,950	0.39
NIEHS SNPs	2,764	4,451	0.62	17,593	38,044	0.46

Note.—Note that some of the analyzed regions of Seattle and National Institute of Environmental Health Sciences (NIEHS) SNPs overlap with each other. Therefore, the numbers in the row of "Seattle + NIEHS" are smaller than the sums of the two individual data sets. In addition, the R_{ID} and R_{NT} values of Seattle and NIEHS SNPs are obviously different from those of dbSNP because of the specific purposes of the two data sets. The Seattle SNPs data set includes mainly inflammatory response genes, whereas the NIEHS data set includes environmental response genes.

[a] R_{ID} and R_{NT} are the ratios of nonpolymorphic changes to polymorphic changes for indels and nucleotide substitutions, respectively.

approach, we calculated the genome-wide ratio of nonpolymorphic to polymorphic HS indels (R_{ID}) and the same ratio for HS substitutions (R_{NT}). R_{ID} (0.74) is in fact lower than R_{NT} (0.87) ($P \approx 0$, χ^2 test), indicating that the modified MK test tends to report positive selection conservatively. To further confirm that the modified MK test is conservative, we calculated the R_{ID} and R_{NT} values in the introns of two resequenced polymorphism data sets—the National Institute of Environmental Health Sciences (NIEHS) (http://egp.gs.washington.edu/) and Seattle single nucleotide polymorphisms (SNPs) (http://pga.gs.washington.edu/). Not surprisingly, both the R_{ID} and R_{NT} derived from dbSNP are overestimated (table 1). However, it is noteworthy that the overestimation of R_{NT} (93%) is far more serious than that of R_{ID} (42%), again supporting the conservativeness of our test (see supplementary table S1 and Supplementary Material online, for more details). Furthermore, a recent study (Chen et al. 2009) has shown that the ratio of substitutions to indels tends to be higher in more divergent than in less divergent sequences. In this vein, we obtain

$$\frac{S_{fix}}{I_{fix}} > \frac{S_{poly}}{I_{poly}},$$

where S_{fix}, I_{fix}, S_{poly}, and I_{poly} represent the numbers of fixed substitutions, fixed indels, polymorphic substitutions, and polymorphic indels, respectively. We can thus obtain

$$\frac{S_{fix}}{S_{poly}} > \frac{I_{fix}}{I_{poly}}.$$

Accordingly, the ratio of fixed to polymorphic substitutions is intrinsically higher than that of indels in the same region. This finding supports the conservativeness of our modified MK test.

It may be argued that the data set used in Chen et al. (2009) was different from the one used in this study. We thus examined whether our data set has the property that the frequencies of indels and substitutions are positively correlated, a premise on which the study of Chen et al. (2009) was based. As shown in supplementary figure S3 (Supplementary Material online), the positive correlation between HS indels and HS substitutions is highly significant. Therefore, it is reasonable to apply the results of Chen

et al. (2009) results in support of the validity of our modified MK test.

The modified MK tests were then performed across the human genome on 100-kb sliding windows that overlapped with each other by 50 kb. A total of 53,241 windows that contain HS indels and substitutions were examined. HS indels in 2,174 (~4.1%) windows, comprising ~179 Mb of the human genome, are found to be positively selected (designated as "PSWs," $P < 0.05$), whereas those in 46,092 (86.6%) windows are selectively neutral, and the rest (4,975 windows; 9.3%) appear to be negatively selected (table 2). If we set the false discovery rate (Storey 2002) to be smaller than 0.05 (which decreases the P value threshold to 0.000824), the number of PSWs becomes 417 (supplementary table S2, Supplementary Material online).

Meanwhile, because positive selection can be falsely identified because of biased gene conversion (BGC) (Galtier and Duret 2007; Duret and Galtier 2009), we examined whether the three primary features of BGC-prone regions occurred in PSWs: 1) being located in subtelomeric regions (here defined as the 5% termini of each chromosome); 2) being located at recombination hotspots; and 3) a high proportion of AT to GC substitutions. We find that 13.2% (286/2,174) of the PSWs are located in subtelomeric regions, 65.3% (1,405/2,153) of the autosomal PSWs overlap with HapMap-annotated recombination hotspots (Frazer et al. 2007), and 44% (957/2,174) contain >40% AT to GC substitutions (supplementary table S2, Supplementary Material online). If we remove the PSWs that satisfy any of the above three conditions, the number becomes 364 (or 116 [0.2% of the tested windows] with the false discovery rate $Q < 0.05$, comprising 9.7 Mb of the human genome) (supplementary table S2,

Table 2
Results of the Modified MK Test

Summary	PSW[a]	NSW[a]	Neutral	Total
No. of windows (A)	2,174	4,975	46,092	53,241
No. of gene-overlapping windows (B)	1,563	3,263	29,527	34,353
Percentage (B/A)	71.9	65.6	64.1	64.5

Note.—In the modified MK test, the numbers of nonsynonymous substitutions are replaced by those of HS indels. That is, the test examines whether R_{ID} is significantly larger than R_{NT}. See Materials and Methods for more details.

[a] "PSW" and "NSW" represent positively and negatively selected windows, respectively.

Table 3
Results of the *DH* Test in the EHH Windows with or without HS Indels

Subpopulation	No. of Windows	#SSRs	Ratio[a]	Q Value[b]
With HS indels				
African	195,513	5 (1[c])	2.6×10^{-5} (7.2×10^{-6c})	0.720
European	168,525	175 (154)	**1.0×10^{-3}** (**9.1×10^{-4}**)[d]	0.122
East Asian	171,907	324 (292)	**1.9×10^{-3}** (**1.7×10^{-3}**)[d]	0.098
Without HS indels				
African	498,491	6 (2)	1.2×10^{-5} (3.6×10^{-6})	0.699
European	352,576	256 (224)	7.3×10^{-4} (6.3×10^{-4})	0.127
East Asian	369,998	440 (394)	1.2×10^{-3} (1.1×10^{-3})	0.105

[a] Number of SSRs divided by number of windows.
[b] The false discovery rate (Storey 2002).
[c] The number (or ratio) of SSRs corrected according to the Q value.
[d] Significantly higher in the regions with HS indels than those without HS indels (boldfaced, P values < 0.007, χ^2 test).

Supplementary Material online). Therefore, at least some of the HS indels are indeed positively selected, rather than falsely identified because of BGC.

PSWs Tend to Overlap Annotated Genes

To investigate whether the identified PSWs are functional, we examined whether the tested windows overlapped with Ensembl-annotated genes. Interestingly, the proportion of gene-overlapping PSWs (71.9%) is significantly higher than the average of 64.5% ($P \approx 0$, χ^2 test, table 2). Even though these indels are not directly occurring at coding sequences, the fact that they tend to overlap with coding regions in the tested windows demonstrate that the selected indels tend to occur in the vicinity of coding sequences. However, we speculate that the target of positive selection is not the coding sequences per se. Rather, the target of selection is likely on the *cis*-noncoding sequences regulating gene expression elements because noncoding regions comprise >97% of the tested windows in terms of length and the vast majority (>94%) of the tested indels are within intergenic and intronic regions. In fact, most of the coding regions by themselves cannot be tested for lack of information (see Materials and Methods for the definition of testability). And none of the testable coding regions passes the modified MK test. One obvious reason is that most of the indels are negatively selected in coding regions (Chen et al. 2007). For a genomic region to be designated as positively selected by the modified MK test, fixed HS indels must occur repeatedly in this specific region. As we know, the less-is-more evolution can be simply induced by a frameshift mutation without any indel events (not to mention repetitive indel events). Adaptive indels associated with less-is-more evolution thus cannot be identified. Unless selection strongly favors active functional elements with dynamic sequence length alterations, the HS indel–affected regions may not pass the modified MK test. An alternative explanation for the presence of positively selected HS indels in noncoding regions is that the indels could change the relative positions or functional motifs of regulatory elements, thus conveying selective advantages by changing the expression patterns or transcriptional–translational regulations of the neighbor-

ing genes. This scenario is consistent with recent findings that although *cis*-elements are evolutionarily relevant (Wray 2007), their architectures are extremely dynamic (Brown et al. 2007). In addition, a recent study indicates that local DNA topology can be altered by minor genetic changes, thus leading to functional changes (Parker et al. 2009). The small number of potential PSWs (116 out of 53,241, or 0.2%) are therefore of great interest.

Human-Specific Indels Are Associated with Recent Selective Sweep Events

We have demonstrated that ~4% (or 0.2%, strictly speaking) of the HS indel–affected regions are positively selected. However, the modified MK test has two limitations. First, the test is not sensitive to recent selective sweep events. The modified MK test considers overrepresentation of "fixed" genetic changes. In recent selective sweep events, however, the positively selected changes may not be completely fixed in the population yet. Second, the modified MK test cannot detect the effects of single indel events because multiple indel events are a prerequisite for a region to pass the test. To compensate for the limitations, we employed a combinatorial test to examine the possibility of recent selective sweep associated with HS indels. We first used the extended haplotype homozygosity (EHH) algorithm (Sabeti et al. 2002) to define the potential "linkage windows" to minimize the effects of recombination (see Supplementary Material online). For comparison, two types of EHH windows were analyzed: the windows that extended from a nearest upstream SNP and a nearest downstream SNP that flanked 1) an HS indel and 2) no HS indels. We then used the *DH* test (Zeng, Shi, et al. 2007), which is a combination of the Tajima's *D* test (Tajima 1989) and Fay and Wu's *H* test (Fay et al. 2001), to search these EHH windows for signatures of recent selective sweep events. We further assessed the false discovery rates (Q values; Storey 2002) of the *DH* test in each test group. As shown in table 3 (detailed information in supplementary table S3, Supplementary Material online), the ratios of selectively swept regions (SSRs) of Europeans and East Asians in HS indel–encompassing windows (group (1)) are significantly higher than the background values (group (2); P values < 0.007, χ^2 test). Furthermore, the proportions of SSRs of the non-African subpopulations are significantly higher than that of the African subpopulation ($P \approx 0$, χ^2 test).

Two questions then ensue. First, what drives the increases of HS indel–associated selective sweep events in Europeans and East Asians? Previous analyses of genome-wide variation patterns have provided support for the "out-of-Africa" hypothesis of recent human evolution (Jakobsson et al. 2008; Li et al. 2008). Therefore, the founder effect could have increased the number of high-frequency–derived alleles in the non-African subpopulations (Keinan et al. 2007). Nevertheless, the *DH* test has been shown to be robust against population bottlenecks and subdivisions (Zeng, Mano, et al. 2007). Therefore, the larger number of recent selective sweep events in Europeans and Asians may not be the result of population history. Rather, it can be associated with subpopulation-specific

adaptations, which is consistent with previous findings (Storz et al. 2004).

Second, why HS indel–affected regions and the other regions have experienced differential selective sweeps? HS indels could have been the drivers of these sweep events. However, because HS indels and substitutions are linked and selected together, we cannot rule out the possibility that these HS indels are in fact hitchhikers in the sweep process. Meanwhile, it is also likely that the HS indels, in combination with the surrounding derived SNPs, constitute the target of recent positive selection. Recall that the nucleotide substitution rates tend to increase in the vicinity of indels (Tian et al. 2008), which may lead to an increased number of HS substitutions around HS indels. Even if most of the HS indels and substitutions are selectively neutral, the increased occurrences of genomic alterations can extend the reaches of the "neutral network" (Wagner 2008) of the affected regions, thus potentially facilitating phenotype changes.

Supplementary Material

Supplementary figures S1–S4 and tables S1–S3 are available at *Genome Biology and Evolution* online (http://www.oxfordjournals.org/our_journals/gbe/).

Funding

Acknowledgments

Literature Cited

Brown CD, Johnson DS, Sidow A. 2007. Functional architecture and evolution of transcriptional elements that drive gene coexpression. Science. 317:1557–1560.

Chen FC, Chen CJ, Li WH, Chuang TJ. 2007. Human-specific insertions and deletions inferred from mammalian genome sequences. Genome Res. 17:16–22.

Chen JQ, et al. 2009. Variation in the ratio of nucleotide substitution and indel rates across genomes in mammals and bacteria. Mol Biol Evol. 26:1523–1531.

Duret L, Galtier N. 2009. Comment on "Human-specific gain of function in a developmental enhancer". Science. 323:714; author reply 714.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Fay JC, Wyckoff GJ, Wu CI. 2001. Positive and negative selection on the human genome. Genetics. 158:1227–1234.

Frazer KA, et al. 2007. A second generation human haplotype map of over 3.1 million SNPs. Nature. 449:851–861.

Galtier N, Duret L. 2007. Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. Trends Genet. 23:273–277.

Hartl DL, Clark AG. 2007. Principles of population genetics. Sunderland (MA): Sinauer Associates, Inc.

Jakobsson M, et al. 2008. Genotype, haplotype and copy-number variation in worldwide human populations. Nature. 451:998–1003.

Jurka J, et al. 2005. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res. 110:462–467.

Keinan A, Mullikin JC, Patterson N, Reich D. 2007. Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. Nat Genet. 39:1251–1255.

King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. Science. 188:107–116.

Levy S, et al. 2007. The diploid genome sequence of an individual human. PLoS Biol. 5:e254.

Li JZ, et al. 2008. Worldwide human relationships inferred from genome-wide patterns of variation. Science. 319:1100–1104.

Li W-H, Saunders MA. 2005. The chimpanzee and us. Nature. 437:50–51.

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature. 351:652–654.

Parker SC, et al. 2009. Local DNA topography correlates with functional noncoding regions of the human genome. Science. 324:389–392.

Perry GH, et al. 2007. Diet and the evolution of human amylase gene copy number variation. Nat Genet. 39:1256–1260.

Podlaha O, Webb DM, Tucker PK, Zhang J. 2005. Positive selection for indel substitutions in the rodent sperm protein catsper1. Mol Biol Evol. 22:1845–1852.

Podlaha O, Zhang J. 2003. Positive selection on protein-length in the evolution of a primate sperm ion channel. Proc Natl Acad Sci U S A. 100:12241–12246.

Sabeti PC, et al. 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature. 419:832–837.

Schully SD, Hellberg ME. 2006. Positive selection on nucleotide substitutions and indels in accessory gland proteins of the Drosophila pseudoobscura subgroup. J Mol Evol. 62:793–802.

Storey JD. 2002. A direct approach to false discovery rates. J R Stat Soc B. 64:479–498.

Storz JF, Payseur BA, Nachman MW. 2004. Genome scans of DNA variability in humans reveal evidence for selective sweeps outside of Africa. Mol Biol Evol. 21:1800–1811.

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

Tian D, et al. 2008. Single-nucleotide mutation rate increases close to insertions/deletions in eukaryotes. Nature. 455:105–108.

Tishkoff SA, et al. 2007. Convergent adaptation of human lactase persistence in Africa and Europe. Nat Genet. 39:31–40.

Wagner A. 2008. Neutralism and selectionism: a network-based reconciliation. Nat Rev Genet. 9:965–974.

Wheeler DA, et al. 2008. The complete genome of an individual by massively parallel DNA sequencing. Nature. 452:872–876.

Wray GA. 2007. The evolutionary significance of cis-regulatory mutations. Nat Rev Genet. 8:206–216.

Yang Z, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. Trends Ecol Evol. 15:496–503.

Zeng K, Mano S, Shi S, Wu CI. 2007. Comparisons of site- and haplotype-frequency methods for detecting positive selection. Mol Biol Evol. 24:1562–1574.

Zeng K, Shi S, Wu CI. 2007. Compound tests for the detection of hitchhiking under positive selection. Mol Biol Evol. 24:1898–1908.