

# “REVERSE ECOLOGY” AND THE POWER OF POPULATION GENOMICS

Yong Fuga Li,<sup>1</sup> James C. Costello,<sup>1,2</sup> Alisha K. Holloway,<sup>3</sup> and Matthew W. Hahn<sup>1,2,4</sup>

<sup>1</sup>School of Informatics and <sup>2</sup>Department of Biology, Indiana University, Bloomington, Indiana

<sup>3</sup>Department of Evolution and Ecology, University of California, Davis, California

<sup>4</sup>E-mail: mwh@indiana.edu

Received February 25, 2008

Accepted May 18, 2008

Rapid and inexpensive sequencing technologies are making it possible to collect whole genome sequence data on multiple individuals from a population. This type of data can be used to quickly identify genes that control important ecological and evolutionary phenotypes by finding the targets of adaptive natural selection, and we therefore refer to such approaches as “reverse ecology.” To quantify the power gained in detecting positive selection using population genomic data, we compare three statistical methods for identifying targets of selection: the McDonald–Kreitman test, the mkprf method, and a likelihood implementation for detecting  $d_N/d_S > 1$ . Because the first two methods use polymorphism data we expect them to have more power to detect selection. However, when applied to population genomic datasets from human, fly, and yeast, the tests using polymorphism data were actually weaker in two of the three datasets. We explore reasons why the simpler comparative method has identified more genes under selection, and suggest that the different methods may really be detecting different signals from the same sequence data. Finally, we find several statistical anomalies associated with the mkprf method, including an almost linear dependence between the number of positively selected genes identified and the prior distributions used. We conclude that interpreting the results produced by this method should be done with some caution.

**KEY WORDS:** Adaptation, McDonald-Kreitman test, mkprf, molecular evolution, population genetics, positive selection.

One of the main goals of ecological genomics is to identify the genes and mutations underlying adaptive traits. Although the challenge of identifying individual genes responsible for adaptive phenotypic variation is daunting, there have been a number of recent successes (for review, see Mitchell-Olds et al. 2007). Two of the most successful approaches have used either candidate genes known to affect the trait of interest (e.g., Wittkopp et al. 2002; Nachman et al. 2003), or quantitative trait locus (QTL) mapping techniques to discover regions associated with the adaptive phenotype (Bradshaw et al. 1998; Colosimo et al. 2004). QTL mapping can then be followed by further crossing experiments (Bradshaw and Schemske 2003) or association mapping (Shapiro et al. 2004) to narrow down large genomic regions to single-gene resolution.

However, some systems are refractory to these approaches for two important reasons. First, not all adaptively differentiated

pairs of species can be crossed, precluding genetic analysis (e.g., humans and chimpanzees). Second, for a number of systems the specific trait responsible for ecological adaptation is unknown. For instance, despite the fact that the mosquito vectors of malaria are extremely important to human health, no consistent phenotypic difference has been found between the major subspecies of the African malaria vector, *Anopheles gambiae*, that occupy different environmental niches (White et al. 2007). Without the ability to carry out crosses or to score convenient traits, research on such organisms has turned to modern genomic tools to identify the underlying adaptive genetic differences, as well as the adaptive traits themselves. As these genomic approaches do not rely on prior information about the relevant ecological traits, Matthew Rockman suggested, at the SSE symposium on Ecological Genomics held in Christchurch, New Zealand, that they should be called “reverse

ecology” by analogy with reverse genetics. By finding the genetic targets of selection, the hope is that we can then find the subtle phenotypic differences targeted by selection. One such “reverse ecology” approach measures differences in gene expression between taxa, under the hypothesis that at least some differences in the pattern or level of transcript abundance represent phenotypic traits contributing to adaptation (for review, see Ranz and Machado 2006). A second approach uses population genetic data at a genomic scale (“population genomics”) to find the targets of adaptive natural selection (e.g., Akey et al. 2004; Begun et al. 2007; Williamson et al. 2007). Both of these approaches can of course also be used in systems in which genetic manipulation is possible, and the candidate genes identified can then be confirmed by further experiments (e.g., Schlenke and Begun 2004).

Although population genomic analyses have become more practical as the price of sequencing has dropped, it is still not clear how much sequence needs to be collected or which statistical approaches provide the most information about targets of selection (but see Akashi 1999). These questions are important as they determine the scope of the experiments that must be undertaken to effectively identify candidate genes. If, for example, comparative approaches that only use a single sequence from each species are just as powerful as approaches using a dozen sequences, then researchers may choose not to invest in large-scale resequencing efforts. In this article we take an empirical approach to evaluating these questions by analyzing population genomic datasets from *Homo sapiens*, *Drosophila simulans*, and *Saccharomyces cerevisiae*. Analyses of both the human (Bustamante et al. 2005) and *Drosophila* (Begun et al. 2007) data have previously been published, but little consideration was paid to the power of alternative statistical approaches to identify targets of adaptive natural selection. Here we consider three different approaches, one that uses only a single sequence from each pair of species and two that use both variation within species and divergence between species. Because methods that consider the frequency spectrum of mutations (e.g., Tajima 1989) are affected by both demographic history and natural selection, we focus on statistical tests less prone to false positives. We also examine the statistical properties of these methods, and find a number of unusual results due to one of them.

## Methods

### DATA

*Drosophila simulans* data come from Begun et al. (2007). They include aligned sequences from 11,453 genes among six sequenced strains of *D. simulans* to *D. melanogaster* orthologs. Because there was only light shotgun sequencing conducted on each strain, the number of sequences for each base ranges between  $n = 2$  and  $n = 6$  (average:  $n = 3.9$ ). For analyses of interspecific divergence

only, the *D. simulans* consensus sequence was used. See Begun et al. (2007) for more details.

*Homo sapiens* polymorphism data come from Bustamante et al. (2005). This article provides counts of the number of nonsynonymous and synonymous mutations from 39 individuals present as either polymorphisms or fixed differences for 11,624 human genes relative to the chimpanzee. These are the data included in analyses using polymorphism. For analyses of interspecific divergence only, human–chimpanzee alignments of orthologous coding sequences were taken from the analysis of Nielsen et al. (2005). We collected additional data on the frequency of the human polymorphisms by mapping the frequencies reported in Lohmueller et al. (2008) to the same segregating sites in the same individuals reported in Bustamante et al. (2005). We were able to map the frequencies for 30,876 polymorphisms to the 11,624 genes used in Bustamante et al. (2005) and analyzed here.

*Saccharomyces* genome sequences were obtained for three sequenced *S. cerevisiae* strains, S288c (the common laboratory strain; Goffeau et al. 1996), YJM789 (a clinical strain; Gu et al. 2005), and RM11–1a (a vineyard strain; Ruderfer et al. 2006). We also obtained the sequence of the sister species to *S. cerevisiae*, *S. paradoxus* (Kellis et al. 2003). The *Saccharomyces* Genome Database annotations of the 5851 known ORFs (excluding mitochondrial genes) for the genome of S288C were used as the reference sequences for alignment to the other genomes. Because the YJM789, RM11–1a, and *S. paradoxus* genome sequences have only been assembled into contigs, the S288C ORF annotations are inputs to the program NUCmer (Kurtz et al. 2004) as reference sequences and queried against the other three genomes. For each reference gene from S288C, if the NUCmer output produced a nonoverlapping match to one strains’ contigs covering at least 70% of the reference sequence, and all three query genomes produced a match, then this gene was considered for further analysis. This procedure resulted in 4253 alignable ORFs across all four genomes. Multiple sequence alignments were made using ClustalW (Thompson et al. 1994), and the number of nonsynonymous and synonymous polymorphisms and fixed differences from these alignments were calculated using custom perl scripts.

### ANALYSES OF SELECTION USING ONLY DIVERGENCE DATA

#### PAML

We used a codon-based method to detect adaptive natural selection using only a single sequence from each species (Goldman and Yang 1994). To do this, we ran the “site” models M1a and M2a for each pair of orthologs in PAML (Yang 2007). Model M1a allows for two classes of codons, those with the ratio of nonsynonymous divergence per nonsynonymous site ( $d_N$ ) to synonymous divergence per synonymous site ( $d_S$ ) less than one ( $d_N/d_S < 1$ ), and those with  $d_N/d_S = 1$ . Model M2a has parameters for these

two classes, in addition to a third class, with  $d_N/d_S > 1$ . If  $d_N/d_S$  is significantly greater than 1, then adaptive substitutions are assumed to have occurred to fix nonsynonymous differences between species. If  $d_N/d_S \leq 1$ , adaptive evolution may still have occurred on some fraction of all differences, but cannot be inferred with certainty. Likelihood ratio tests between M1a and M2a provide statistical confidence for evidence of positive selection when the model with an extra parameter allowing for  $d_N/d_S > 1$  (M2a) is significantly better than the one without (M1a). As recommended by Yang (2007), we use two degrees of freedom for the likelihood-ratio test.

## ANALYSES OF SELECTION USING POLYMORPHISM AND DIVERGENCE DATA

### McDonald–Kreitman test

Our second test for adaptive evolution among these three species pairs was suggested by McDonald and Kreitman (1991). The McDonald–Kreitman (MK) test asks whether the counts of nonsynonymous and synonymous polymorphisms and fixed differences are proportional using a  $2 \times 2$  test of independence (carried out here using Fisher's exact test). Under the neutral hypothesis, the ratio of the number of nonsynonymous polymorphisms to fixed differences ( $P_N:D_N$ ) should be equivalent to the ratio of the number of synonymous polymorphisms to fixed differences ( $P_S:D_S$ ) because neutral polymorphisms are expected to fix in proportion to their frequency. An excess of nonsynonymous fixed differences is indicative of positive selection having acted to fix these mutations, whereas an excess of nonsynonymous polymorphism indicates segregating weakly deleterious variation that does not go on to be fixed. Under stringent conditions an excess of nonsynonymous polymorphism may also result from strong balancing selection (Weinreich and Rand 2000).

One method for summarizing the results of the  $2 \times 2$  McDonald–Kreitman table is the neutrality index (NI; Rand and Kann 1996). The neutrality index is equal to  $(P_N/D_N)/(P_S/D_S)$ . For convenience, we use  $-\log_{10}(\text{NI})$ , such that strict neutrality gives a value of 0, positive selection results in positive values, and weak negative selection gives negative values. Because 0-values for some of the counts make NI undefined (even though there may still be enough power to conduct the test of independence), we added a pseudocount of 1 to each cell for plotting the results in figures; pseudocounts were not used in the MK tests themselves.

Templeton (1996) proposed a modification of the standard MK test that considered “singleton” polymorphisms (i.e., those found in only one individual) separately from more common polymorphisms. This modification aims to improve the power of the MK test because segregating deleterious nonsynonymous polymorphisms are most likely to be at low frequency, and hence to be singletons. Fay et al. (2001, 2002) used a similar modification that simply removed all polymorphisms at frequencies  $< 15\%$  from the

MK test. Charlesworth and Eyre-Walker (2008) have shown that this cut-off is theoretically justified, and we have used it here for the modified MK tests.

### mkprf

The third method we used to detect adaptive natural selection is based on Poisson random field (PRF) theory (Sawyer and Hartl 1992). This approach models the frequencies of polymorphisms as being drawn from a PRF, and provides explicit predictions for the frequencies of mutations under neutrality, positive selection, and negative selection. This theory can be used to infer adaptive natural selection when allele frequencies are known (e.g., Bustamante et al. 2001; Williamson et al. 2005), or when the frequencies are simply counted as polymorphic or fixed (Bustamante et al. 2002). This latter approach is used here as it requires only the four values used in the MK test ( $P_N$ ,  $D_N$ ,  $P_S$ , and  $D_S$ ). The mkprf method takes these four values as input and estimates the average value of the selection parameter,  $\gamma$  ( $= 2N_e s$ ), for each gene using PRF theory. This selection parameter is positive ( $\gamma > 0$ ) when there are an excess of nonsynonymous fixed differences, and negative ( $\gamma < 0$ ) when there are an excess of nonsynonymous polymorphisms.

Because of the large number of parameters that must be estimated from only the four cell counts, mkprf implements a Markov Chain Monte Carlo method to estimate  $\gamma$  (Bustamante et al. 2002). Only those loci with  $P_N + D_N \geq 4$  are potentially informative about positive selection using mkprf, so we restricted all of our analyses to genes meeting this criterion (cf. Bustamante et al. 2005). We considered a gene to be under positive selection if the mean posterior probability of  $\gamma$  was positive and the 95% Bayesian credibility intervals did not overlap 0. All mkprf analyses were conducted using the website provided by the authors of this method (<http://cbsuapps.tc.cornell.edu/mkprf.aspx>).

## Results

### POSITIVELY SELECTED GENES FOUND USING $d_N/d_S$

To directly compare results between different methods for detecting positive selection, we used only the set of genes from each dataset that met the criterion  $P_N + D_N \geq 4$ . In total we analyzed 3836 genes between *S. cerevisiae* and *S. paradoxus*, 8887 genes between *D. melanogaster* and *D. simulans*, and 3119 genes between *H. sapiens* and *P. troglodytes* (Table 1). Calculations of  $d_N/d_S$  use only a single sequence from each species and provide significant evidence for positive selection when  $d_N/d_S > 1$  for some fraction of sites in a gene. Using a nominal  $P$ -value of 0.05 we found 46 significant genes in the yeast comparison (1.2%), 393 in the *Drosophila* comparison (4.4%), and 99 in the primate comparison (3.2%) (Table 1). As can be seen, there are fewer than 5% significant comparisons in all three species pairs. Although this result seems to imply that there are no true positives in this dataset,

**Table 1.** Results of tests for selection.

Dataset	Number of genes	PAML <sup>1</sup>	MK <sup>2</sup>	mkprf <sup>3</sup>
<i>H. sapiens</i>	3119	99	10	291
<i>D. simulans</i>	8887	393	1054	3250
<i>S. cerevisiae</i>	3836	46	2	226

<sup>1</sup>Number of genes significant at  $P < 0.05$  using  $\chi^2$  distribution with  $df = 2$ .

<sup>2</sup>Number of genes significant by Fisher's exact test ( $P < 0.05$ ) and with  $-\log(\text{NI}) > 0$ .

<sup>3</sup>Number of genes for which  $\gamma$  is positive and the 95% credibility interval does not overlap 0 using the nonhierarchical MCMC method of mkprf (Bustamante et al. 2005) with  $\sigma = 8$ .

using two degrees of freedom in the likelihood ratio test between models M1a and M2a in PAML can be conservative (Yang 2007). Simulations have shown that only 2% of results at this nominal  $P$ -value are false positives (Wong et al. 2004), implying that at least some fraction of the *Drosophila* and primate genes are true positives. After Bonferroni correction for the number of tests conducted in each species comparison, zero genes are significant in yeast, five in *Drosophila*, and seven in primates.

#### POSITIVELY SELECTED GENES FOUND USING THE MCDONALD–KREITMAN TEST

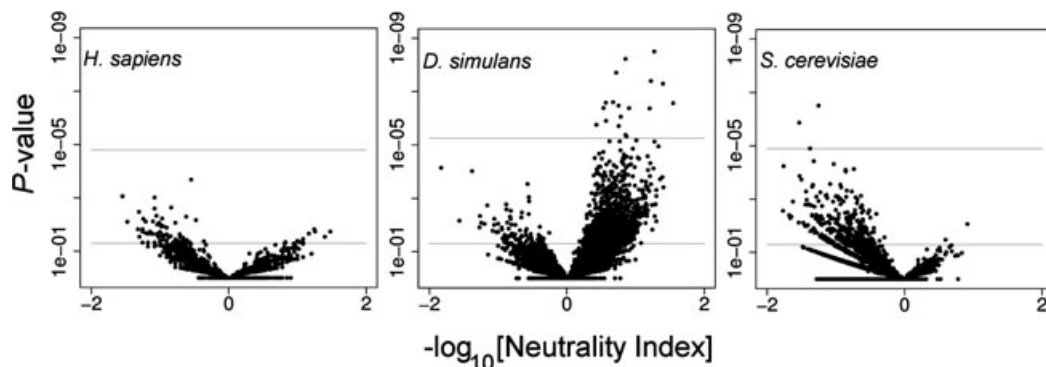
We carried out McDonald–Kreitman tests for the same genes as in the  $d_N/d_S$  analysis. These tests all used polymorphism data from a single species (*S. cerevisiae*, *D. simulans*, and *H. sapiens*) compared to the number of fixed differences with a single sequence from the outgroup. Again using a nominal  $P$ -value of 0.05 and only counting those genes with a pattern suggestive of positive selection, we found two significant genes in the yeast comparison (0.05%), 1054 in the *Drosophila* comparison (11.8%), and 10 in the primate comparison (0.33%) (Table 1). After Bonferroni correction for the number of tests conducted in each species comparison, 0 genes are significant in yeast, 18 in *Drosophila*, and 0 in primates.

Figure 1 shows the values of the Neutrality Index for each gene in each species ( $-\log_{10}[\text{NI}]$ ) plotted against the  $P$ -value for each MK test. As can be seen, both the human and yeast datasets have a majority of points below zero, whereas the *Drosophila* dataset has a majority of points greater than zero. These results are consistent with the much larger number of significant MK tests in the *Drosophila* comparison than the other two datasets, as  $-\log_{10}(\text{NI})$  values greater than zero are consistent with an excess of fixed nonsynonymous differences. Both human and yeast have an excess of nonsynonymous polymorphism, which is generally interpreted as an excess of weakly deleterious mutations segregating in the population. We also observed a relationship between  $-\log_{10}(\text{NI})$  and the  $P$ -value for a gene, with larger absolute values associated with smaller  $P$ -values (Fig. 1). This result is expected, as genes with more extreme configurations in  $2 \times 2$  contingency tables are likely to both have more extreme values of NI and lower  $P$ -values.

Because of the larger number of individual humans sequenced, we were able to use a modified MK test that does not consider either synonymous or nonsynonymous polymorphisms at frequencies  $< 15\%$  (Fay et al. 2001; Charlesworth and Eyre-Walker 2008). Of the 3119 human genes considered above, 1697 met the criterion  $P_N + D_N \geq 4$  after removing all polymorphisms reported at frequencies less than 15% in Lohmueller et al. (2008). Only eight genes (0.47%) were significant in the direction of positive selection by the MK test from this set, indicating that while the statistical power of the test increased after removing low-frequency polymorphisms, the greatly reduced number of genes considered means that fewer targets of selection were identified overall.

#### POSITIVELY SELECTED GENES FOUND USING MKPRF

We carried out mkprf analyses for the same set of genes from each species for the two previous tests (Table 1). Replicating the methods used by Bustamante et al. (2005), we initially set the prior distribution on  $\gamma$ , the selection coefficient for each gene, to have



**Figure 1.** Neutrality indices for all three datasets. Each point represents the neutrality index (NI) for a specific gene and the  $P$ -value associated with the McDonald–Kreitman test for that gene. For plotting purposes and so that each value for NI is defined, NI is calculated as  $(P_N + 1/D_N + 1)/(P_S + 1/D_S + 1)$ . The lower line in each graph represents  $P = 0.05$  whereas the upper line represents  $P = 0.05/n$ .

a mean of 0 and a standard deviation of 8 for each dataset. Using these parameters we found 291 genes under positive selection in humans, 3250 in *Drosophila*, and 226 in yeast (Table 1). The slight discrepancy between our results in humans and the results of Bustamante et al. (304 genes under positive selection) is largely due to the fact that some of the genes used in the original analysis have been removed from the updated human genome annotation; we obtained exactly the same results as this previous paper if we did not filter the dataset based on updated annotations.

In their 2005 paper, Bustamante et al. introduced a modified MCMC method used to estimate  $\gamma$  relative to the method used in a previous publication (Bustamante et al. 2002). This modification makes it so that a single Gaussian prior standard deviation on  $\gamma$ -values is set for all loci, rather than the hierarchical method used previously. To see the effect of this updated method, we reran analyses for all of the genes from all three species using the older hierarchical MCMC method ( $\sigma^2 = 10$ ; note that the hierarchical method takes the variance on  $\gamma$  as a prior, rather than the standard deviation in the updated method). Using this method we found seven genes under positive selection in humans, 3310 in *Drosophila*, and zero in yeast (Table 2). Although the results for *Drosophila* are largely unchanged, the human and yeast results appear to be highly sensitive to the method used to estimate  $\gamma$  values.

To further explore the effects of the exact method used to specify the prior distribution on  $\gamma$ , we reran our analyses with a series of values for the standard deviation of the prior. In addition to the value of the standard deviation used in the original paper ( $\sigma = 8$ ; Bustamante et al. 2005), we used five additional values of  $\sigma$  ranging from 0.1 to 16 for the newer MCMC method and three additional values for  $\sigma^2$  for the older MCMC method. Our results show that the number of positively selected genes for each species is highly dependent on the value for the standard deviation of the prior distribution on  $\gamma$ -values (Fig. 2). For each dataset there is a clear positive relationship between the variance of the prior on  $\gamma$  and the number of positively selected genes. For instance, the number of positively selected genes in humans ranges from 1 to 322 depending on whether a standard deviation of 0.1 or 16 is used (Table 2). This positive correlation makes sense, as larger standard deviations allow individual values of  $\gamma$  to be further from 0. This in turn makes it more likely that the credibility intervals around each mean value of  $\gamma$  do not overlap 0, allowing for there to be more genes inferred to be under positive selection. However, the fact that the prior distribution has such a strong effect on the results of analyses using mkprf is not ideal. It is also not clear which value for the standard deviation should be used, or how many genes are truly under positive selection in each of the three datasets.

Even if we accept the number of genes under positive selection with  $\sigma = 8$  for each dataset, it is unclear how to correct

**Table 2.** Effect of prior distribution and MCMC method on mkprf results.

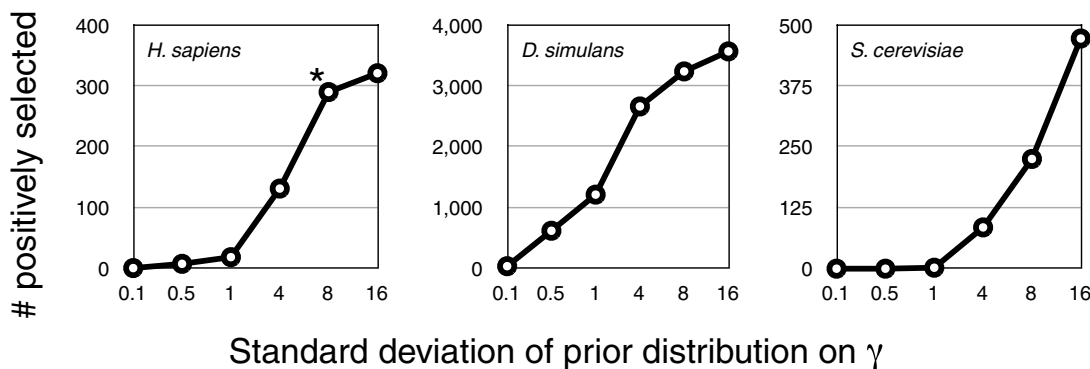
Dataset	Method	Value of prior	$P(\gamma > 0   \text{Data})^3$
<i>H. sapiens</i>	mkprf <sup>1</sup> ( $\sigma$ )	0.1	1
		0.5	8
		1	19
		4	132
		8	291
	mkprf <sup>2</sup> ( $\sigma^2$ )	16	322
		1	7
		10	7
		100	7
		1000	10
<i>D. simulans</i>	mkprf ( $\sigma$ )	0.1	41
		0.5	625
		1	1220
		4	2673
		8	3250
	mkprf ( $\sigma^2$ )	16	3580
		1	3316
		10	3310
		100	3312
		1000	3319
<i>S. cerevisiae</i>	mkprf ( $\sigma$ )	0.1	0
		0.5	0
		1	2
		4	85
		8	226
	mkprf ( $\sigma^2$ )	16	474
		1	0
		10	0
		100	0
		1000	0

<sup>1</sup>Results from nonhierarchical MCMC method (Bustamante et al. 2005).

<sup>2</sup>Results from hierarchical MCMC method (Bustamante et al. 2002).

<sup>3</sup>Number of genes for which  $\gamma$  is positive and the 95% credibility interval does not overlap 0.

for multiple tests using mkprf. The problem is that mkprf is a Bayesian method, and posterior probabilities are not expected to be uniformly distributed as  $P$ -values are. For instance, the use of 95% credibility intervals on the posterior probabilities of  $\gamma$  to identify targets of selection does not imply that 5% of genes would be in the tails of these distributions under the null model of no selection. Therefore, typical  $P$ -value corrections for multiple hypothesis tests—such as the Bonferroni correction (Rice 1989)—cannot be applied to the results from mkprf, and were not applied by Bustamante et al. (2005). This also presents a problem when comparing results between methods, as the genes identified as positively selected by mkprf are not “significant” in the same way as genes identified either by PAML or the MK test. We were confused by the statement in Bustamante et al. (2005) that



**Figure 2.** Effect of prior distribution on results from mkprf. Relationship between the standard deviation on the prior distribution chosen for  $\gamma$  and the number of genes found to be under positive selection using the nonhierarchical MCMC. The asterisk (\*) indicates the value used in Bustamante et al. (2005). Note the different scales of the y-axis in each panel and the nonuniform values on the x-axis for all panels.

“the chief advantage of the mkprf method is an increase in power to detect selection without an increase in type I error,” as the concepts of “power” and “type I error” are not usually applied to Bayesian analyses and it is not clear how one would evaluate these properties in a Bayesian context. To provide confidence in genes identified as targets of positive selection in humans, Bustamante et al. (2005) conduct simulations to show that there is a slight excess of posterior probabilities in the upper 1% tail, although there is a larger deficit of data in the 5% tail relative to the simulated datasets. Although we have not conducted simulations here, the issue of how many genes are actually in the tail (i.e., which value of  $\sigma$  to use) confounds any possible simulation results.

Carlos Bustamante (pers. comm.) has suggested that all of the values for the standard deviation on  $\gamma$  used here (and in Bustamante et al. 2005) are informative, and that truly uninformative priors would have to be very large. This implies that the true number of genes under positive selection in all three species is larger than any value estimated here, and lies to the right of even the largest values seen in Figure 2. This suggestion is supported by the fact that the number of positively selected genes appears to be leveling off for both humans and flies (Fig. 2). However, if the true number is indeed this large we would expect a great excess of data in the tails of the distribution relative to neutral simulations, which is not seen (Bustamante et al. 2005).

Furthermore, we found one further issue with respect to the signature of “positive selection” detected by mkprf that requires some caution when interpreting its results. In our initial comparison of  $\gamma$  to the NI values from the MK tests, one of the genes originally identified as under positive selection in humans by the mkprf method (*NEB*) had a positive value for  $\gamma$  ( $= 1.5$ ), but a significantly negative value of NI ( $-\log[\text{NI}] = -1.8$ ). This result is counterintuitive as it means the two methods are giving opposite results: mkprf infers the gene to have a history of positive selection, whereas the McDonald–Kreitman test shows a history

of weak negative selection. As there are a number of genes like *NEB* for which the values of  $\gamma$  and NI are in opposite directions, we examined the estimation of  $\gamma$  by the mkprf method in greater detail.

The original paper outlining the PRF model says that  $\gamma$  can be estimated from just the ratio of  $P_N/D_N$  (equation 22 in Sawyer and Hartl 1992). In this case  $\gamma$  is inversely proportional to this ratio, with smaller values of  $P_N/D_N$  leading to higher values of  $\gamma$ . Because the mkprf method appears to be using the same estimator of  $\gamma$  (Bustamante et al. 2005), it is unsurprising that  $\gamma$  is in fact highly correlated with  $P_N/D_N$  in the human dataset (correlation between  $\gamma$  and  $\log(P_N+1/D_N+1)$ :  $r = -0.97$ ,  $P < 0.0001$ ). The somewhat surprising implication of the estimator being used is that the values of  $D_S$  and  $P_S$  do not determine the relative values of  $\gamma$  among genes, although they may affect the absolute values because they are used in estimating divergence times and mutation rates.

To unambiguously demonstrate that  $D_S$  and  $P_S$  do not determine  $\gamma$ , for all three datasets we permuted the values of  $D_S$  and  $P_S$  together for each gene while holding the values of  $D_N$  and  $P_N$  the same for the given gene and reestimated  $\gamma$ . In other words, this permutation procedure leaves  $D_N$  and  $P_N$  constant for each gene but uses the  $D_S$  and  $P_S$  counts from another gene. For each of the three datasets we generated two new permuted datasets and found a very strong and significant correlation between  $\gamma$ -values estimated from the original data and the permuted data each time (all  $r > 0.99$ ; Table 3). This correlation between  $\gamma$ -values was present whether we used hierarchical MCMC or nonhierarchical MCMC methods (Table 3). Our permutation results explain why the results from the mkprf and MK analyses can give such different results: mkprf is using only  $D_N$  and  $P_N$  to estimate the strength of selection, whereas the MK test is using  $D_N$ ,  $P_N$ ,  $D_S$ , and  $P_S$ . However, it is still not clear from these analyses which of the two methods provides the “correct” history of selection for each gene (see Discussion).

**Table 3.** Correlation of  $\gamma$ -values between original and permuted datasets using mkprf.

Dataset	Method	Shuffle 1 <sup>3</sup>	Shuffle 2
<i>H. sapiens</i>	mkprf <sup>1</sup> ( $\sigma=8$ )	0.999 <sup>4</sup>	0.999
	mkprf <sup>2</sup> ( $\sigma^2=10$ )	0.999	0.999
<i>D. simulans</i>	mkprf ( $\sigma=8$ )	0.999	0.999
	mkprf ( $\sigma^2=10$ )	0.999	0.999
<i>S. cerevisiae</i>	mkprf ( $\sigma=8$ )	0.998	0.998
	mkprf ( $\sigma^2=10$ )	0.999	0.999

<sup>1</sup>Results from nonhierarchical MCMC method (Bustamante et al. 2005).

<sup>2</sup>Results from hierarchical MCMC method (Bustamante et al. 2002).

<sup>3</sup>Shuffle 1 and Shuffle 2 represent the two permuted datasets made separately for each species comparison.

<sup>4</sup>Pearson correlation between  $\gamma$ -values for each gene estimated using the same MCMC method on both the original and permuted data in each row.

### CORRELATION AMONG METHODS FOR DETECTING SELECTION

The above results demonstrate that the three methods used to detect positive selection appear to differ either in their statistical power or in the data being used to identify significant genes. However, these differences in power do not immediately tell us about the similarities among the genes identified. To examine the correlations between methods, we looked at both overlap among the positively selected genes identified by each method and at the correlations in summary statistics of selection from each method.

The overlap in genes identified by each method as being under positive selection was fairly good across all three datasets. Taking the results from mkprf to be those generated using the nonhierarchical MCMC with  $\sigma = 8$ , in all three datasets there was a significant excess of genes also identified as under positive selection by both the MK test and PAML included within this superset ( $P < 0.01$  for all except yeast genes identified by PAML). For instance, of the 10 genes significant at  $P < 0.05$  by the MK test, six were also among the 292 genes identified by mkprf as being under positive selection whereas two were among the 99 genes identified by PAML (both  $P < 0.01$  by binomial sampling).

This overlap among significant genes suggests that the summary statistics of selection from each method should also be correlated. We used the value of  $\gamma$  from mkprf, the value of  $-\log_{10}(\text{NI})$  from the MK test, and the value of  $\log_{10}(d_N/d_S)$  from PAML as summary statistics of selection. We found significant correlations between these three measures in all three datasets, although the correlations were of different magnitude between the various tests (Table 4). For instance, the correlations between  $\gamma$  and  $-\log_{10}(\text{NI})$  were often higher than the correlations between either  $\gamma$  and  $\log_{10}(d_N/d_S)$  or  $-\log_{10}(\text{NI})$  and  $\log_{10}(d_N/d_S)$ . This is most likely due to the fact that both mkprf and the MK test are

**Table 4.** Correlations among tests for selection.

Dataset	PAML <sup>1</sup>	MK <sup>2</sup>	mkprf <sup>3</sup>
PAML	–	0.47/0.17/0.33 <sup>4</sup>	0.25/0.22/0.03 <sup>4</sup>
MK	**/**/**5	–	0.70/0.61/0.32 <sup>4</sup>
mkprf	**/**/5	**/**/**5	–

<sup>1</sup>Values of  $\log(d_N/d_S)$ .

<sup>2</sup>Values of  $-\log(\text{NI})$ .

<sup>3</sup>Values of  $\gamma$ .

<sup>4</sup>Pearson correlation coefficients from human/*Drosophila*/yeast.

<sup>5</sup> $P$ -values associated with correlation coefficients: \* $P < 1 \times 10^{-4}$ , \*\* $P < 1 \times 10^{-50}$ .

using polymorphism data, whereas PAML is not. Nevertheless, significant correlations between the statistics indicate that they are at least partly in agreement with respect to the direction and magnitude of natural selection.

One result that is slightly puzzling is the strong correlation between the results of mkprf and the MK test for all three datasets (Table 4). As discussed in the previous section, results from these two methods can sometimes be directly conflicting because mkprf uses only the ratio of  $P_N/D_N$  to estimate the strength of selection. The correlation may be explained by the fact that both  $\gamma$  and  $-\log_{10}(\text{NI})$  are significantly correlated with a third variable,  $P_N/D_N$  ( $r = -0.97$  and  $r = -0.70$ , respectively, for the human dataset). In fact, the partial correlation between  $\gamma$  and  $-\log_{10}(\text{NI})$  after controlling for the correlations with  $P_N/D_N$  is 0.08 ( $P < 0.01$ ). Although the relationship is still significant, it is obvious that the apparent strong correlation between these two methods ( $r = 0.70$ ) is actually due to their shared correlation with a third explanatory variable.

### Discussion

As DNA sequencing technologies have come down in price, it has become easier to sequence a large number of loci from even nonmodel organisms. Data from these sequencing projects allow researchers to quickly scan for evidence of adaptive evolution using a variety of statistical techniques. In this article we have compared three of these tests for selection on three independent datasets with the aim of determining their relative power and cost-effectiveness. Although the utility and effectiveness of each test appears to vary with the dataset considered, several general conclusions can be drawn from our analyses.

Our main expectation coming into this analysis was that the population genetic approaches (i.e., MK and mkprf) would be more informative about adaptive natural selection than approaches that just use divergence data (i.e., PAML). However, this was not always the case. Comparing only the number of genes found to be under positive selection by the McDonald–Kreitman test with the number from PAML, we actually found fewer significant genes

with the MK method in two of the three datasets (Table 1). This result is nonintuitive, as it can be easily shown that there are cases of positive selection in which the excess of nonsynonymous differences over synonymous differences are not great enough to be significant by  $d_N/d_S$ , but are significant by the MK test. Even in cases that conformed to our expectations—as in *Drosophila*—differences in statistical power did not appear to wholly explain our results: the set of genes identified as under positive selection by PAML was not wholly a subset of those by the MK test. In fact, only 66 of the 393 *Drosophila* genes significant by  $d_N/d_S$  were also significant by the MK test. Although this overlap is statistically significant we would expect a much larger proportion to overlap if the only difference between the methods was statistical power. There are cases in which the MK test will not reject neutrality but  $d_N/d_S$  will, as with very strong balancing selection (Hughes and Nei 1988), but these are thought to be rare.

The explanation for the observed patterns may lie in the differences among the three datasets and in the specific data used by each method. In the *Drosophila* data there is generally an excess of nonsynonymous fixed differences, whereas in both the human and yeast data there are generally an excess of nonsynonymous polymorphisms (Fig. 1). The difference in patterns of polymorphism and divergence between these species parallels the difference in apparent power of the MK test, with more power compared to  $d_N/d_S$  in *Drosophila* and less in human and yeast. In other words, *Drosophila* shows the expected pattern and human and yeast do not. Why might the differences in polymorphism and divergence explain the patterns of positive selection? One must remember that although the MK test explicitly considers fixed differences between species,  $d_N/d_S$  only compares a single sequence from each species and therefore incorporates both polymorphism and fixed differences into its measure of divergence. This means that higher ratios of nonsynonymous to synonymous polymorphism are used by  $d_N/d_S$  to detect positive selection, but are not used by the MK test. As mentioned earlier, an excess of nonsynonymous polymorphism in the MK test can be interpreted as either weakly deleterious variation or evidence for strong balancing selection, although the conditions necessary for balancing selection to cause such a pattern are much more restrictive (Weinreich and Rand 2000). Two nonexclusive explanations for our results are therefore: (1) there is much more strong balancing selection occurring in nature than previously thought, or (2)  $d_N/d_S$  is positively misleading about signatures of positive selection because it interprets segregating deleterious polymorphisms as evidence for adaptive divergence. A third possibility is that although the MK test may have more power to detect natural selection all things being equal, all things are not equal. The human and yeast datasets have many fewer polymorphisms per gene than does the *D. simulans* dataset: an average of 5.55 polymorphisms for humans and 6.77 for yeast versus 27.04 for *D. simulans*. It may therefore be that—given the

number of polymorphisms and fixed differences considered for each species—the  $d_N/d_S$  comparison has power to detect natural selection in human and yeast whereas the MK test does not have that power. Although the MK test clearly has enough statistical power in these species to detect an excess of nonsynonymous polymorphism (see Fig. 1), the difference in diversity may at least explain some of the differences among species.

A strict comparison of the statistical power of the methods considered here does not provide a complete picture of the effectiveness of each method in detecting targets of adaptive natural selection. This is partly due to the fact that many loci in each species comparison were removed from our dataset because they could not be analyzed by each method with sufficient statistical power. Although PAML still has the power to detect positive selection at a locus with no or little variation, both the MK test and mkprf require a minimum number of segregating sites. For example, of the 11,453 genes sequenced from *D. simulans* that could be aligned to *D. melanogaster*, only 8887 were considered here because  $P_N + D_N$  was required to sum to  $\geq 4$ ; an even greater proportion of human genes were excluded. There still may be many genes under positive selection among the excluded set, and in those cases only PAML can be used to detect the signature of selection. In the *D. simulans* dataset considered here the MK test identified almost three times as many genes under positive selection as did  $d_N/d_S$  (Table 1), although this ratio will be lower if there are genes under positive selection that did not meet our criteria for the number of polymorphisms present. This means that methods that require less data may still provide an equivalent (or close to equivalent) number of candidate adaptation genes, even if their statistical power is reduced when compared to population genetic methods. Our results also do not say whether researchers will generally gain more power by sequencing more individuals within a population or by sequencing additional species. The addition of new species to analyses of  $d_N/d_S$  can improve the power to detect selection, both because more observations are made for each codon and because more powerful “branch-site” methods can be used to detect selection along individual lineages (Yang 2007). Further analyses will be needed among more species and population genomic datasets to determine the relative value and cost-effectiveness of sequencing projects.

One of the most surprising findings of our study was the heavy reliance of results from the mkprf method on the prior distributions used. In particular, it was clear from using multiple values for the prior on the standard deviation of  $\gamma$  that the number of genes identified as being under positive selection is almost linearly dependent on the breadth of this distribution (Fig. 2). As far as we are aware the dependence of the results from mkprf on the prior distributions has not previously been discussed in the literature on this method (Bustamante et al. 2002, 2005). We also have not found any previous analyses comparing the number of



positively selected genes using the hierarchical MCMC implemented by mkprf (Bustamante et al. 2002) with the number found using from the newer, nonhierarchical method (Bustamante et al. 2005). The results presented here strongly suggest that caution should be used when interpreting any particular set of results from mkprf, and that any reported mkprf results should explicitly include all values for model input parameters.

Our analyses also showed that the mkprf method is actually using slightly different data from the classical McDonald–Kreitman test to identify targets of selection. Whereas the MK test is explicitly comparing the counts of nonsynonymous polymorphism and divergence to synonymous polymorphism and divergence, mkprf is “informed only by the nonsynonymous cell entries in a conventional McDonald–Kreitman test (that is,  $P_N$  and  $D_N$ )” (Bustamante et al. 2005). Because the selection parameter,  $\gamma$ , in mkprf is being estimated solely from the ratio of  $P_N/D_N$ , this method appears to be closer in concept to the HKA test (Hudson et al. 1987) than the MK test: the genes with the strongest signatures of positive selection are those with the highest ratios of divergence to polymorphism. Such a genome-wide HKA test could be very informative about selection (e.g., Begun et al. 2007), although the interpretation of  $\gamma$  would be very different. For instance, part of the appeal of the classical MK test is that because it is comparing nonsynonymous and synonymous mutations, selective events linked to the locus of interest will either increase or decrease the number of both types of polymorphisms, and the test will therefore not incorrectly identify linked neutral loci as the targets of selection. But if only  $P_N/D_N$  is used, then the effect of linked selection on another gene or noncoding element will be to lower  $P_N$ , making it appear as though the gene of interest has a history of strong positive selection when none existed. For this reason, results using this instantiation of the PRF method should be viewed with caution (see Desai and Plotkin 2007 for problems with other applications of PRF theory).

Our study was motivated by the desire to identify those genes and mutations that underlie ecological adaptation. This “reverse ecology” approach has identified a number of significant genes in each of the three datasets examined, some of which have obvious ecological and evolutionary consequences. For instance, among the seven genes identified as being under positive selection in humans by PAML after Bonferroni correction, one is the gene *Neuralized-2* (*Neur2*). This gene is involved in neurogenesis and is known to interact with the colorfully named *Mind Bomb-1* in mice (Song et al. 2006). This association with neurogenesis and brain development has obvious implications for human evolution. Many of the functions associated with positively selected genes in *D. simulans* have been discussed at length (Begun et al. 2007) and we will simply reiterate here the high frequency of reproduction-related and chromosomal architecture-related genes. Because there were so few significant genes in the

yeast dataset, it is difficult to identify any particular gene or function as having played a role in recent yeast evolution. Previous studies have found increases in the number of flocculation genes in yeast (Hahn et al. 2005) and in the number of intragenic repeats within flocculin genes (Verstrepen et al. 2005), although neither pattern would be detected by methods that are simply comparing nonsynonymous to synonymous mutations. Flocculation in the brewer’s yeast is a known important adaptation to the brewing process (Jin and Speers 1998).

Our results also demonstrate that differences in population size can have very important effects on patterns of molecular variation. Figure 1 clearly shows that the two species with small population sizes—human and yeast—have an excess of segregating nonsynonymous polymorphism and a deficit of nonsynonymous fixed differences. Because natural selection is less efficacious in smaller populations, we would expect that such species would be less able to either remove deleterious mutations or fix advantageous ones. A similar pattern of polymorphism and divergence has been found in the primarily selfing species, *Arabidopsis thaliana* (Weinreich and Rand 2000; Bustamante et al. 2002). In yeast it has been shown that strains that have been “domesticated” to the laboratory for longer periods of time show more relaxed selection and therefore greater rates of evolution, presumably due to the constant bottlenecks that are imposed (Gu et al. 2005). This interpretation of the yeast data was challenged by Ronald and colleagues (2006), who found even greater rates of nucleotide substitution in the “wild” vineyard strain. However, it is now clear that this strain was actually domesticated for use in viticulture thousands of years ago (Fay and Benavides 2005), and therefore shows exactly the pattern of molecular evolution expected with reduced population size.

Overall, the results presented in this article do not make a clear case for the power of population genomic datasets. In two of the three datasets examined there were actually more significant results using simple comparative methods, although these results may not be due to positive directional selection. It appears that the differences in power arise from differences in the patterns of polymorphism and divergence among species, so that a priori information about there being either an excess of nonsynonymous polymorphism or divergence could help researchers determine whether the investment in sequencing a population genetic dataset is worthwhile. Finally, there is obviously more information in population genetics data than has been considered here. Tests of natural selection utilizing either the frequency spectrum of mutations (e.g., Tajima’s  $D$ ) or that compare levels of polymorphism to divergence without regard to the types of each mutation (e.g., Hudson et al. 1987) can both be applied to population genomic datasets, although the interpretation of such tests is much more ambiguous. Population sequencing data also provides a plethora of molecular markers for crossing experiments and

other genetic studies, prerequisites for studies of local adaptation. For these reasons it is clear that population genomic datasets will continue to accumulate and contribute to our understanding of the genetic basis of ecological divergence, regardless of whether the sequence data itself are ever used to test for natural selection.

### ACKNOWLEDGMENTS

We thank C. Bustamante for discussion and A. Eyre-Walker for constructive comments on the manuscript. This work was supported by a grant from the National Institutes of Health to S. V. Nuzhdin and MWH (R01-GM076643A). AKH was funded by a grant from the National Institutes of Health (GM071926) to D. J. Begun.

### LITERATURE CITED

- Akashi, H. 1999. Inferring the fitness effects of DNA mutations from polymorphism and divergence data: statistical power to detect directional selection under stationarity and free recombination. *Genetics* 151:221–238.
- Akey, J. M., M. A. Eberle, M. J. Rieder, C. S. Carlson, M. D. Shriver, D. A. Nickerson, and L. Kruglyak. 2004. Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS Biol.* 2:1591–1599.
- Begun, D. J., A. K. Holloway, K. Stephens, L. W. Hillier, Y.-P. Poh, M. W. Hahn, P. M. Nista, C. D. Jones, A. D. Kern, C. Dewey, et al. 2007. Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol.* 5:e310.
- Bradshaw, H. D., and D. W. Schemske. 2003. Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature* 426:176–178.
- Bradshaw, H. D., K. G. Otto, B. E. Frewen, J. K. McKay, and D. W. Schemske. 1998. Quantitative trait loci affecting differences in floral morphology between two species of monkeyflower (*Mimulus*). *Genetics* 149:367–382.
- Bustamante, C. D., J. Wakeley, S. Sawyer, and D. L. Hartl. 2001. Directional selection and the site-frequency spectrum. *Genetics* 159:1779–1788.
- Bustamante, C. D., R. Nielsen, S. A. Sawyer, K. M. Olsen, M. D. Purugganan, and D. L. Hartl. 2002. The cost of inbreeding in *Arabidopsis*. *Nature* 416:531–534.
- Bustamante, C. D., A. Fedel-Alon, S. Williamson, R. Nielsen, M. T. Hubisz, S. Glanowski, D. M. Tanenbaum, T. J. White, J. J. Sninsky, R. D. Hernandez, et al. 2005. Natural selection on protein-coding genes in the human genome. *Nature* 437:1153–1157.
- Charlesworth, J., and A. Eyre-Walker. 2008. The McDonald–Kreitman test and slightly deleterious mutations. *Mol. Biol. Evol.* 25:1007–1015.
- Colosimo, P. F., C. L. Peichel, K. Nereng, B. K. Blackman, M. D. Shapiro, D. Schluter, and D. M. Kingsley. 2004. The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS Biol.* 2:635–641.
- Desai, M. M., and J. B. Plotkin. 2007. Detecting directional selection from the polymorphism frequency spectrum. *ArXiv e-prints: 0707.2428*.
- Fay, J. C., and J. A. Benavides. 2005. Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. *PLoS Genet.* 1:e5.
- Fay, J. C., G. J. Wyckoff, and C.-I. Wu. 2001. Positive and negative selection on the human genome. *Genetics* 158:1227–1234.
- . 2002. Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature* 415:1024–1026.
- Goffeau, A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, et al. 1996. Life with 6000 genes. *Science* 274:563–567.
- Goldman, N., and Z. Yang. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* 11:725–736.
- Gu, Z. L., L. David, D. Petrov, T. Jones, R. W. Davis, and L. M. Steinmetz. 2005. Elevated evolutionary rates in the laboratory strain of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 102:1092–1097.
- Hahn, M. W., T. De Bie, J. E. Stajich, C. Nguyen, and N. Cristianini. 2005. Estimating the tempo and mode of gene family evolution from comparative genomic data. *Genome Res.* 15:1153–1160.
- Hudson, R. R., M. Kreitman, and M. Aguade. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* 116:153–159.
- Hughes, A. L., and M. Nei. 1988. Pattern of nucleotide substitution at major histocompatibility complex class-I loci reveals overdominant selection. *Nature* 335:167–170.
- Jin, Y. L., and R. A. Speers. 1998. Flocculation of *Saccharomyces cerevisiae*. *Food Res. Int.* 31:421–440.
- Kellis, M., N. Patterson, M. Endrizzi, B. Birren, and E. S. Lander. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 423:241–254.
- Kurtz, S., A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway, C. Antonescu, and S. L. Salzberg. 2004. Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12.
- Lohmueller, K. E., A. R. Indap, S. Schmidt, A. R. Boyko, R. D. Hernandez, M. J. Hubisz, J. J. Sninsky, T. J. White, S. R. Sunyaev, R. Nielsen, et al. 2008. Proportionally more deleterious genetic variation in European than in African populations. *Nature* 451:994–997.
- McDonald, J. H., and M. Kreitman. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351:652–654.
- Mitchell-Olds, T., J. H. Willis, and D. B. Goldstein. 2007. Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nat. Rev. Genet.* 8:845–856.
- Nachman, M. W., H. E. Hoekstra, and S. L. D’Agostino. 2003. The genetic basis of adaptive melanism in pocket mice. *Proc. Natl. Acad. Sci. USA* 100:5268–5273.
- Nielsen, R., C. Bustamante, A. G. Clark, S. Glanowski, T. B. Sackton, M. J. Hubisz, A. Fedel-Alon, D. M. Tanenbaum, D. Civello, T. J. White, et al. 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* 3:e170.
- Rand, D. M., and L. M. Kann. 1996. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. *Mol. Biol. Evol.* 13:735–748.
- Ranz, J. M., and C. A. Machado. 2006. Uncovering evolutionary patterns of gene expression using microarrays. *Trends Ecol. Evol.* 21:29–37.
- Rice, W. R. 1989. Analyzing tables of statistical tests. *Evolution* 43:223–225.
- Ronald, J., H. Tang, and R. B. Brem. 2006. Genomewide evolutionary rates in laboratory and wild yeast. *Genetics* 174:541–544.
- Ruderfer, D. M., S. C. Pratt, H. S. Seidel, and L. Kruglyak. 2006. Population genomic analysis of outcrossing and recombination in yeast. *Nat. Genet.* 38:1077–1081.
- Sawyer, S. A., and D. L. Hartl. 1992. Population genetics of polymorphism and divergence. *Genetics* 132:1161–1176.
- Schlenke, T. A., and D. J. Begun. 2004. Strong selective sweep associated with a transposon insertion in *Drosophila simulans*. *Proc. Natl. Acad. Sci. USA* 101:1626–1631.
- Shapiro, M. D., M. E. Marks, C. L. Peichel, B. K. Blackman, K. S. Nereng, B. Jonsson, D. Schluter, and D. M. Kingsley. 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* 428:717–723.
- Song, R., B.-K. Koo, K.-J. Yoon, M.-J. Yoon, K.-W. Yoo, H.-T. Kim, H.-J. Oh, Y.-Y. Kim, J.-K. Han, C.-H. Kim, et al. 2006. Neuralized-2 regulates a Notch ligand in cooperation with Mind Bomb-1. *J. Biol. Chem.* 281:36391–36400.

- Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
- Templeton, A. R. 1996. Contingency tests of neutrality using intra/interspecific gene trees: the rejection of neutrality for the evolution of the mitochondrial cytochrome oxidase II gene in the hominoid primates. *Genetics* 144:1263–1270.
- Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. Clustal-W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673–4680.
- Verstrepen, K. J., A. Jansen, F. Lewitter, and G. R. Fink. 2005. Intragenic tandem repeats generate functional variability. *Nature Genet.* 37:986–990.
- Weinreich, D. M., and D. M. Rand. 2000. Contrasting patterns of nonneutral evolution in proteins encoded in nuclear and mitochondrial genomes. *Genetics* 156:385–399.
- White, B. J., M. W. Hahn, M. Pombi, B. J. Cassone, N. F. Lobo, F. Simard, and N. Besansky. 2007. Localization of candidate regions maintaining a common polymorphic inversion (2La) in *Anopheles gambiae*. *PLoS Genet.* 3:e217.
- Williamson, S. H., R. Hernandez, A. Fledel-Alon, L. Zhu, R. Nielsen, and C. D. Bustamante. 2005. Simultaneous inference of selection and population growth from patterns of variation in the human genome. *Proc. Natl. Acad. Sci. USA.* 102:7882–7887.
- Williamson, S. H., M. J. Hubisz, A. G. Clark, B. A. Payseur, C. D. Bustamante, and R. Nielsen. 2007. Localizing recent adaptive evolution in the human genome. *PLoS Genet.* 3:901–915.
- Wittkopp, P. J., K. Vaccaro, and S. B. Carroll. 2002. Evolution of yellow gene regulation and pigmentation in *Drosophila*. *Current Biol.* 12:1547–1556.
- Wong, W. S. W., Z. Yang, N. Goldman, and R. Nielsen. 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* 168:1041–1051.
- Yang, Z. H. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24:1586–1591.

Associate Editor: M. Rausher