Skepticism toward adaptive signals in DNA sequence comparisons – Is the neutral theory dead yet?

Ziwen He[a,*], Qipian Chen[a,*], Hao Yang[a,*], Qingjian Chen[a], Suhua Shi[a] & Chung-I Wu[a,b,c]

[a] State Key Laboratory of Biocontrol, Guangdong Key Lab of Plant Resources, Key Laboratory of Biodiversity Dynamics and Conservation of Guangdong Higher Education Institutes, School of Life Sciences, Sun Yat-Sen University, Guangdong, China

[b] CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

[c] Department of Ecology and Evolution, University of Chicago, Chicago, Illinois, USA

* These authors contributed equally to this work.

Correspondence should be addressed to C-I.W. (ciwu@uchicago.edu)

Abstract

Measuring positive selection on DNA sequences between species is key to testing the neutral theory of molecular evolution. Here, we compare the two most commonly used tests that rely on very different assumptions. The MK test[1] compares divergence and polymorphism data, while the PAML test[2,3] analyzes multi-species divergence. The two tests are now forced to detect positive selection on the same phylogenetic branch in *Drosophila* and *Arabidopsis* using large-scale genomic data. When applied to individual coding genes, both MK and PAML identify >100 adaptively evolving genes but the two sets hardly overlap. To rule out high false negatives, we merge 20 – 30 genes into "supergenes", 8% - 56% of which yield adaptive signals. Nevertheless, the joint calls still do not overlap. The two tests do show very modest concordance at lower stringencies. There may be several possibilities for the discordance between the two major tests. 1) Selective advantage is weak, falling in the "nearly neutral" range[4]; 2) The adaptive landscape shifts constantly, akin to the Red Queen landscape[5]; 3) Positive selection which accelerates the evolution is confounded by the relaxation of negative selection, resulting in less deceleration. Whether the neutral theory should be rejected depends on which of these factors prevails.

Introduction

Detecting adaptive evolution in DNA sequences is one of the central tasks of molecular evolutionary studies. In the neutral theory[6], such signals are believed to be too infrequent to cause significant deviations from the neutral prediction. Methods have been continually developed to detect positive selection and to test the neutral theory[2,3,7–13]. Indeed, the neutral theory may have been declared dead at least as often as new methods have been proposed[11,12,14–17]. Such methods fall into two broad classes. One class attempts to detect positive selection that operates within populations[7,8,10,14,18]. The other class focuses on positive selection that operates in the longer term, i.e., the divergence between species[1,12,15,19–24]. Methods of either class may use data of both polymorphism and divergence[1,10,12,15,21–24]. The rejection of the neutral theory should be multi-faceted. Positive

selection signals could be abundant between species but undetectable within populations, or vice versa. Either complete or partial rejection of the neutral theory can be informative about the forces driving molecular evolution[4,25,26].

In this study, we focus on the between-species tests that examine the evolutionary rates in coding regions. In such tests, one compares the number of non-synonymous changes per non-synonymous site (Ka or dN) vs. the per site synonymous changes (Ks or dS). The Ka/Ks (or dN/dS) ratio would deviate from 1 if nonsynonymous substitutions are under stronger selection than synonymous changes. In the absence of selection, R = Ka/Ks ~ 1, which has been the hallmark of neutral evolution [4,25–28]. In species comparisons, genome-wide R ranges mainly between 0.05 – 0.25[28–32], thus indicating the prevalence of negative selection. When R > 1, positive selection is evident. Nevertheless, R > 1 is too stringent a criterion as it requires positive selection to overwhelm negative selection. Indeed, few genes in any genome comparison have R's significantly greater than 1[4,24,33,34].

The two prevailing methods that relax the requirement for R > 1 in detecting positive selection are the MK (McDonald-Kreitman)[12] and the PAML (Phylogenetic analysis by maximum likelihood)[2,3] tests. A few studies that employ the two tests generally use one test to complement the other, for example, by using PAML to identify candidate genes and MK to estimate the amount of selection[34–39]. This may be the first study that forces the two tests to infer positive selection on the same set of genes and along the same phylogenetic branch. The concurrent application of the two key tests is important because each test relies on a different set of assumptions, neither of which can be easily verified (see below). The purpose of this comparative study is hence to examine these underlying assumptions and to cross-validate the conclusions.

Theoretical background

While Ka and Ks are the cornerstones for detecting natural selection, they can only inform about either positive or negative selection, but not both. This is because Ka/Ks is the joint outcome of the two opposing forces:

$$R = Ka/Ks = (1 - p - q) + p\,[\,N\,f\,(N, s1)\,] + q\,[\,N\,f\,(N, s2)\,] \qquad \text{Eq. (1)}$$

where p and q are the proportion of advantageous and deleterious mutations, respectively[25,26,40]. Also, $f(N, s) = (1 - e^{-2s})/(1 - e^{-2Ns})$ is the fixation probability of a mutation with a selective coefficient of s where s can be > 0 (denoted by s1) or < 0 (s2) and N is the effective population size. For example, if Ka/Ks = 0.2, then the null hypothesis is the neutrality of q = 0.8, p = 0 and f (N, s2) = 0 (i.e., no fixation of deleterious mutations). The alternative hypothesis would be adaptive evolution with p > 0 and f (N, s1) > 0 (i.e., fixation of advantageous mutations).

To tease apart positive and negative selection, one often uses DNA sequences from several species, some of which should have polymorphism data. For this study, the data are from *Drosophila* and *Arabidopsis*, as shown in Fig. 1. The hypothesis testing for positive selection by the MK test is done on a particular phylogenetic lineage, such as the ones marked by the red line (Fig. 1). The Ka and Ks values in the red-line lineage are contrasted with the corresponding polymorphisms (pA and pS) in the blue triangle. The value of pA and pS denotes, respectively, the level of nonsynonymous and

synonymous polymorphism (per site) within a species. The rationale of the MK test is that $p \sim 0$ in the polymorphism data thanks to the rapidity with which advantageous mutations are fixed. Thus, Eq. (1) becomes

$$pA/pS \sim (1 - q) + q\varepsilon \quad \text{Eq. (2)}$$

where ε represents the amount of deleterious polymorphism and should be a very small number. In short, the MK test estimates q from Eq. (2) first and then extracts p from Eq. (1).

There are, however, several difficulties in interpreting the MK test results. First, the strength of negative selection is estimated from the recent evolutionary history (the blue triangle in Fig. 1), whereas positive selection is inferred from a different lineage (the red line). As pointed out before, an increase in the effective size of the extant population would lead to the under-estimation of pA/pS and, thus, an over-estimation of positive selection[15,19]. Second, the estimation of negative selection is not as straightforward. The pA/pS ratio would decrease as the variant frequency increases and may increase again when the mutant frequency approaches 1[13,15,22,41]. Both patterns can be seen in *Drosophila* (Fig. 1c). In *Arabidopsis* (Fig. 1d), the pattern is similar at the low frequency end, but not at high frequencies. Given the complex patterns, accurate estimation of negative selection is not straightforward in the MK test[1,12,15,19–24]. In Methods, the estimation of negative selection is described in detail. Third, the MK test is strictly applicable only to sites that share the same genealogy. In the presence of recombination, in particular when unlinked loci are used, biases could be non-trivial, making corrections necessary[22,23].

The other widely used approach to the estimation of adaptive mutations is the PAML method[2,3]. PAML compares the substitution numbers across many lineages to identify positively (or negatively) selected genes on the assumption that unusually high (or low) numbers could be indicative of selection. There are three sub-models, two of which, the site model and the branch-site model, are used here. In the site model, sites are compared for an increase or decrease in non-synonymous substitutions in the entire phylogeny[2,3]. The null model is usually rejected using the likelihood ratio test[42]. In the branch-site model, sites of a pre-selected branch (the foreground) are compared with other sites on all branches as well as the same sites on other branches (the background).

Despite the very different approaches, the MK and PAML tests can be used to answer the same question – How much adaptive evolution has operated on the chosen genes in a given branch (e.g., the red-line branch of Fig. 1a and 1b)? Because the MK test is about positive selection along the red line, it does not offer any information about selection elsewhere in the phylogeny. Therefore, it may be comparable with either of the two sub-models of PAML. First, if the MK results are interpreted to mean the general adaptiveness of a gene, then the gene may also be adaptive along other branches. In that case, the proper comparison would be with the site model of PAML. Alternatively, if the adaptation is specific to the genes on a specific branch, then the branch-site model of PAML would be a more suitable comparison. We will present the site model results in the main text and the branch-site model results in the Supplement. The two sets of comparisons yield qualitatively similar results although the site model appears to be statistically more robust.

Results

Part I – Identifying adaptive genes with high stringency

We first determine the distribution of the P value among individual genes. For the MK test, the P value is obtained from the Fisher's exact test on the contingency table. For the PAML test, the likelihood ratio test was used to obtain the P value. In Fig. 2, the distributions of the P value are shown in the four panels for the two taxa and two tests. The distribution is concentrated above $P > 0.8$ (the MK test for *Drosophila*) or above $P > 0.9$ (the other panels). This concentration means that a very large percentage of genes are uninformative, partly because each gene may experience too few changes to be statistically informative. Furthermore, the null model does not fully incorporate factors that can affect the test. For example, the polymorphism data may not reflect the complete removal of deleterious mutations and the strength of negative selection is often under-estimated[15,21,22].

Fig. 1 suggests that, even if all genes evolve neutrally, far fewer than 5% of them would be detected as adaptive at the 5% cutoff. We therefore compare the observed P values of the MK and PAML tests against each other, rather than against the null model. In each panel of Fig. 2, one line represents the test results on all genes and the other is on genes that have been pre-filtered through the other test. In Fig. 2a-2b, genes pre-filtered through PAML have smaller P values in the MK test, as shown in the leftward shift. The same is true in Fig. 2c-2d where the pre-filtering by MK reduces the P values by the PAML test. The two tests are indeed correlated, but perhaps only weakly. This is also true in Fig. S1, where the branch-site model of PAML is used.

We now enumerate the overlaps between the two tests by comparing the candidate adaptive genes with $P < 0.05$. Given the P value distributions of Fig. 2, these genes are merely the most likely candidates by each test. Hence, significant overlaps would be mutual corroborations. For the "individual genes" analysis in *Drosophila*, we identified 186 from 5425 genes by the MK test and 145 genes by PAML, corresponding to 3.43% and 2.67% of the genome (see Table 1). The overlap of these two sets has only 9 genes. Although the observed overlap is significantly higher than the expected 4.97 ($P < 0.01$, Fisher's exact test), the overlap is too small to be biologically meaningful. The same pattern is true for *Arabidopsis*, in which 145 and 505 genes are called by these two tests, respectively, but only 14 genes are called by both tests. Again, the observed overlap is significantly higher than the expected at 5.55 ($P < 0.01$) but the actual overlap is minimal. A simple explanation for the non-overlap is high false negatives. In other words, each test may have detected only a small fraction of the adaptive genes.

The analysis of supergenes and their component genes

False negatives should be common in individual genes with few substitutions. To overcome the statistical limitations, we create artificial "supergenes" by merging 20 – 30 genes into a longer sequence. They are either concatenations of neighboring genes (i.e., by physical location) or genes of the same ontology (by function). The merger would reduce false negatives with the risk of diluting the true adaptive signals. We now present the results based on the concatenations of neighboring genes in Table 1. The results based on the merger by gene ontology are similar (See Table S1). In *Drosophila* and *Arabidopsis*, 200 and 500 supergenes are created respectively.

For the MK test, gene merger would create biases as pointed out before[22]. When the level of polymorphism is negatively correlated with the rate of nonsynonymous divergence across loci, false positives would be common in the merger. Hence, we use the modified MK test to infer positive selection in merged genes. In *Drosophila*, 112 of the 200 super genes are significant at the 5% level by the MK test and 36 of the 200 are significant by PAML (Table 1). The two tests detect far more adaptive supergenes than individual genes - 56% and 18% by MK and PAML respectively. What is perplexing is that the overlaps are random (10.0% observed vs. the expected 10.1%) as if the two tests are completely uncorrelated. In *Arabidopsis*, 8.2% of the 500 supergenes are significant by the MK test at the 5% level and, by the PAML test, 25.6% of supergenes are. In *Arabidopsis*, the PAML test detects many more adaptive supergenes than the MK test, in the opposite direction of *Drosophila*. However, the overlap is also random with 2.0% observed vis-à-vis the expected 2.1%. In both taxa, the two tests appear uncorrelated at the level of supergenes.

Because gene merger might dilute the adaptive signal by mixing a few adaptively evolving genes with many other non-adaptive genes, we examine the component genes within each adaptive supergene. In *Drosophila*, the 112 significant supergenes by the MK test contain 3132 component genes (Table 1), among which 158 genes are significant when tested individually. The corresponding numbers by the PAML test are 60 out of 1040 component genes. Between the two subsets of genes (3132 and 1040, respectively), 619 genes overlap and only 3 genes are significant by both tests. The 0.48% overlap of component genes is slightly higher than the expected 0.29%. The observations in *Arabidopsis* are given in the last row of Table 1. The overlap in component genes is also very low, at 2 of the 258 genes, or 0.78%. Clearly, the MK and PAML tests are uncorrelated by the standard statistical criteria, which are relaxed in the next section. In Table S2, the comparable analyses using the branch-site model of PAML yield results similar to those of Table 1.

Part II - Identifying weakly adaptive genes with low stringency

We note in Fig. 2 that genes yielding a P value of 0.25 by either test may be moderately informative about positive selection. Therefore, when carrying out the MK and PAML tests simultaneously, we set the cutoff in each test at $P < 0.224$. By doing so, the expected overlap would be $0.224^2 = 5\%$ if the two tests are completely uncorrelated. The results by the relaxed stringency are given in Table 2.

For *Drosophila*, 824 and 353 genes are now called, respectively, by the MK and PAML tests at $P < 0.224$. These two sets have an overlap of 91 genes whereas the expected overlap is 53.6 ($P < 10^{-6}$ by Fisher's exact test). For *Arabidopsis*, the two tests yield 1014 and 1172 genes with an overlap of 119 genes, significantly higher than the expected number of 91.6 ($P < 0.0025$, Fisher's exact test). Hence, the joint call of adaptive genes accounts for 10.1% (119/1172) to 25.8% (91/353) of the call by a single test. A gene identified by one test as adaptive has a 10% - 25% chance of being called adaptive by the other.

While the overlap between the two tests is at most modest, the performance of one test conditional on the pre-screen by the other indeed suggests some concordance. We first compare A1, the average number of adaptive sites per gene called by the MK test. In the A1 row of *Drosophila* in Table 2, A1

doubles from 2.84 to 5.71 (without vs. with the pre-screen by PAML) and increases from 14.98 to 19.94 (MK vs. [MK + PAML]). The trend is even more pronounced in *Arabidopsis*: 0.84 to 1.97 and 19.36 to 28.98. Thus, the PAML screen can enhance the performance of the MK test.

The procedure is now applied in the reverse direction by pre-screening the genes with the MK test before subjecting them to the PAML test. In PAML, the number of adaptive sites per gene can be calculated in two ways, A2 and A2' in Table 2[43,44]; see Methods). Since the purpose is to compare PAML with MK, we use the A2 numbers, which are closer to A1 from the MK test. The qualitative conclusion, nevertheless, is not affected much by either estimate. In the A2 row of *Drosophila* in Table 2, A2 increases 5.71 to 10.93 (without vs. with the pre-screen by MK) and 9.27 to 14.65 (PAML vs. [PAML + MK]). The same trend is observed in the A2 row of *Arabidopsis* (Table 2) – from 10.15 to 14.24 (without vs. with the pre-screen by MK) and from 12.74 to 20.36 (PAML vs. [PAML + MK]). Again, a pre-screen by MK helps the performance of PAML.

While we use the site model of PAML above, the results are similar when the branch-site model is used (see Table S3). It is clear that the MK and PAML tests are correlated but the correlation is weak. In other words, when one test detects a strong adaptive signal in a gene, the other test would often find a signal in the same gene, albeit a much weaker one.

Discussion

To detect adaptive signals in DNA sequences, different statistical tests make various assumptions, most of which cannot be easily verified. The low concordance between methods demands that we scrutinize these assumptions. Two categories of explanations are considered.

One category is methodological. The simplest explanation is that both tests have high false negatives and miss a large fraction of adaptive genes. However, when the tests are applied to supergenes, the overlap is still no higher than random occurrences. Since the revised test yields an adaptive signal in up to 56% of supergenes, false negatives could not possibly be high. Alternatively, either (but not both) test may have a very high false-positive rate and the overlap would be essentially random noises. This is not a satisfactory explanation because the two tests seem comparable in performance. When PAML is done on genes selected by the MK test, the subset of genes yield much stronger signals than the full set. This is also true when the MK test is done on PAML-selected genes. Overall, both tests have at least some merits in detecting adaptive evolution and the low concordance cannot be solely attributed to methodological reasons.

The other category is biological. First, the discordance may be a consequence of the "nearly neutral" evolution proposed by Ohta (1992)[4]. While near-neutrality is often used to indicate "slightly deleterious" variants, Ohta (1992)[4] and Ohta and Gillespie (1996)[25] have suggested that advantageous mutations may often be nearly neutral as well. When the selective strength is very small, slight variations in the analyses are likely to yield discordant results bordering on random noises. In this sense, this explanation is also about high false negatives, but the reason is biological, rather than methodological.

A second possibility is a shifting adaptive landscape. In an unchanging landscape, a species would climb up the adaptive peak and stay there by fending off deleterious mutations. Once a species is on the adaptive peak, molecular evolution would consist mainly of neutral changes while deleterious mutations are purged. In this view, adaptive molecular evolution should be infrequent and, importantly, the search for adaptive signals could not have been based on the supposition of an unchanging fitness landscape. (Why would one search for something that is believed to be nearly non-existent?) In an alternative view, adaptive evolution is ongoing and most species are continually evolving toward the fitness peak, which may be shifting like a sand dune. In this Red Queen landscape[5,45,46], different genes are evolving adaptively at different times.

Neither explanation above, however, is entirely satisfactory. In the nearly-neutral theory, one might expect the stronger signals from the supergenes to overlap somewhat between the MK and PAML tests but they do not overlap at all. Also, if the fitness landscape is shifting, then both positive and negative selective pressure would be changing from time to time. Using the same argument, the third explanation below would be more parsimonious as it is based on simpler and known mechanisms.

The third explanation is the relaxation of negative selection. Positive selection is inferred when there is a speed-up in the rate of molecular evolution, relative to some background references. In the MK test, the reference is the polymorphism, and in the PAML test, the reference is the larger phylogeny (or other genes in the larger phylogeny). It is generally assumed that negative selection stays constant between the focus lineage and the references. If negative selection is not constant (for example, if the effective population size fluctuates), then the inferred adaptive signals could simply be the result of the relaxation of negative selection. In other words, less negative selection could often speed up the molecular evolution as much as positive selection could.

In the search for the signals of positive selection, we have not paid sufficient attention to the operation of negative selection that enforces evolutionary constraint. Unlike positive selection, negative selection follows some simple and nearly universal rules and can be more confidently measured[47]. At present, adaptive signals in DNA sequences are not distinguished from signals of relaxed negative selection. The discordance between the MK and PAML tests seems to reflect that uncertainty.

Methods

DNA sequence data

Pre-aligned unique *Drosophila* transcript sequences were downloaded from Flybase (r6.16, http://flybase.org). We collected 8560 FASTA alignments of five species (*D. melanogaster, D. simulans, D. sechellia, D. yakuba, and D. erecta*, Fig. 1a). Genome-wide *D. melanogaster* polymorphism data were obtained from the *Drosophila* Population Genomics Project Phase[48]. Genes with high divergence rates, apparently caused by misalignment, were discarded. Only genes with more than 40 codons and 10 samples of polymorphism data were used. The final dataset contains 5245 genes with an average of 50 samples of polymorphism data.

The DNA sequences of *A. thaliana* and *A. lyrata* were obtained from the Phytozome database[49]. Genome-wide *A. thaliana* polymorphism data were obtained from the 1001 Genomes Project[50]. We also obtained DNA sequences from *Capsella grandiflora*, *C. rubella*, and *Boechera stricta* (Fig. 1b) for analyses using the PAML program[3]. We began with 14953 alignments of the five species and then filtered the data as we did for *Drosophila*. Only genes with more than 300 samples of polymorphism data were collected. The final dataset consists of 12975 genes.

Supergene construction

To overcome statistical limitations, we created artificial supergenes by merging genes into longer sequences. We used two concatenation approaches: by physical location and by ontology. The first method involved merging 20 to 30 nearby genes residing on the same chromosome. This resulted in 200 *Drosophila* and 500 *Arabidopsis* supergenes. To apply the ontology approach, we first identified GO (gene ontology) term(s) for each gene. To ensure that every gene was present in only one supergene, we sorted GO terms by the number of genes they comprised and checked the component genes in each supergene. If a gene was previously included in a set, it was not merged again. GO terms with fewer than eight genes in *Drosophila* and 10 in *Arabidopsis* were discarded. The final set comprised 184 *Drosophila* and 454 *Arabidopsis* supergenes.

The McDonald-Kreitman (MK) test

Let A and S be the number of nonsynonymous and synonymous changes per gene (or per genome). In Fig. 1c-1d, A/S ratio is given for the number of polymorphic changes at a defined frequency range. (The frequencies were inferred by the free-ratio model of the CODEML program)[2,3]. The A/S ratio becomes lower when the mutant frequency becomes higher. Apparently, the A/S ratio at the low frequency range is boosted by deleterious mutations that have not been removed by negative selection. To avoid the confounding effect of negative selection in the MK test, we only used common mutations with derived allele frequencies >0.2, as is done previously[22,51]. Note that the A/S ratio reaches a steady level at around 0.2 in Fig. 1c-1d.

Now, we let A and S designate the total number of *common* polymorphic mutations (frequency > 0.2) in the MK test. The corresponding numbers of changes between species are designated A' and S'. These four numbers are gathered in a 2×2 contingency table. Fractions of amino acid substitutions which are adaptive (α) can be estimated as $\alpha = 1 - (S' A) / (A' S)$. We used Fisher's exact test on 2×2 contingency tables to estimate statistical significance[52]. Shapiro *et al*. pointed out the possibility of false positives when the MK test is applied across genes and proposed a procedure to correct the bias. This correction, as well as an alternative method[23], is implemented in our calculations.

Ideally, the A and S numbers in polymorphism should reflect only the neutral level of variation. However, it can be seen in Fig. 1c-1d that A often includes deleterious mutations in low frequency and, perhaps, advantageous mutations in high frequency[10,20,22,41]. The inclusion of both kinds of mutations would bias the polymorphic A/S ratio upward, hence reducing the excess of A'/S' over A/S and compromising the power of the MK test. Various solutions have been proposed[15,21–23] to more accurately measure the polymorphic A's. These methods are mostly ad hoc in nature. Sawyer and Hartl[1] propose a more approach to this problem by directly estimating the intensities of negative

selection. While the theory outpaced the data at that time, the approach is feasible now given the large amount of polymorphism data.

If the distribution of the strength of negative selection is known[1], the neutral A/S ratio as reflected in the polymorphism can be accurately estimated. While the estimation of positive selection is indeed different from the conventional numbers, the MK results obtained by various procedures do show the same qualitative pattern of limited overlap with the PAML test. The overall patterns suggest that the discordance between the MK and PAML tests is biological, rather than technical, as presented in Discussion.

The PAML test

We used both the site model and the branch-site model in PAML. The site model, allowing the ω ratio (dN/dS) to vary among sites (codons or amino acids in protein), detected positive selection across the five chosen species. A likelihood ratio test (LRT) was used to compare the alternative model M2a (selection model allowing an additional category of positively selected sites with $\omega>1$ by setting: model = 0, NSsites = 2, fix_omega = 0, omega = 2) with the null model M1a (neutral model allowing only two categories of sites with $\omega<1$ and $\omega=1$ by setting: model = 0, NSsites = 1, fix_omega = 0, omega = 2). Significance was determined using the χ^2 (df = 2) test.

The branch-site model, allowing dN/dS to vary both among sites and across lineages, was used to detect positive selection along specified branches. We compared the likelihood of the alternative model A (positive selection, model = 2, NSsites = 2, fix_omega = 0), to the null model A1 (model = 2, NSsites = 2, fix_omega = 1, omega = 1). *D. melanogaster* and *A. thaliana* were designated as the foreground branches for the test. Significance was calculated using LRT as above. The site model results are presented in the text and the branch-site results are given in the Supplement.

In the analysis of both models, we also employed Bayes empirical Bayes (BEB)[53] estimates, which is available for calculating the posterior probabilities for site classes and can be used to identify sites under positive selection if the likelihood ratio test is significant.

Acknowledgement

This study was supported by the National Natural Science Foundation of China (91731301 and 31600182); the National Key Research and Development Plan (2017FY100705); the 985 Project (33000-18831107) and the Fundamental Research Funds for the Central Universities (17lgpy99).

References

1. Sawyer, S. A. & Hartl, D. L. Population genetics of polymorphism and divergence. *Genetics* **132,** 1161–1176 (1992).
2. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Bioinformatics* **13,** 555–556 (1997).
3. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24,** 1586–1591 (2007).
4. Ohta, T. The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol. Syst.* **23,** 263–286 (1992).

5. van Valen, L. A new evolutionary law. *Evolutionary Theory* **1,** 1–30 (1973).

6. Kimura, M. *The neutral theory of molecular evolution*. (Cambridge University Press, 1983).

7. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123,** 585–595 (1989).

8. Fu, Y.-X. X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147,** 915–925 (1997).

9. Sabeti, P. C. *et al.* Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419,** 832 (2002).

10. Fay, J. C. & Wu, C. I. Hitchhiking under positive Darwinian selection. *Genetics* **155,** 1405–1413 (2000).

11. Hudson, R. R., Kreitman, M. & Aguad??, M. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116,** 153–159 (1987).

12. McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the Adh locus in Drosophila. *Nature* **351,** 652 (1991).

13. Sawyer, S. A., Parsch, J., Zhang, Z. & Hartl, D. L. Prevalence of positive selection among nearly neutral amino acid replacements in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 6504–6510 (2007).

14. Zeng, K., Shi, S., Fu, Y.-X. & Wu, C.-I. Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics* (2006).

15. Fay, J. C., Wyckoff, G. J. & Wu, C.-I. Testing the neutral theory of molecular evolution with genomic data from Drosophila. *Nature* **415,** 1024 (2002).

16. Kern, A. D. & Hahn, M. W. The Neutral Theory in Light of Natural Selection. *Mol. Biol. Evol.* **35,** 1366–1371 (2018).

17. Yang, Z. & Swanson, W. J. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Mol. Biol. Evol.* **19,** 49–57 (2002).

18. Charlesworth, B. & Charlesworth, D. Neutral Variation in the Context of Selection. *Mol. Biol. Evol.* **35,** 1359–1361 (2018).

19. Eyre-Walker, A. Changing effective population size and the McDonald-Kreitman test. *Genetics* **162,** 2017–2024 (2002).

20. Fay, J. C. & Chung-IWu. *Sequence divergence,functional constraint, and selection in protein evolution*. *Annu. Rev. Genomics Hum. Genet. 2* **4,** (2003).

21. Messer, P. W. & Petrov, D. A. The McDonald-Kreitman test and its extensions under frequent adaptation: Problems and solutions. *arXiv Prepr. arXiv1211.0060* **2013,** (2012).

22. Shapiro, J. A. *et al.* Adaptive genic evolution in the Drosophila genomes. *Proc. Natl. Acad. Sci.* **104,** 2271–2276 (2007).

23. Smith, N. G. C. & Eyre-Walker, a. Adaptive protein evolution in Drosophila . *Nature* **415,** 1022–1024 (2002).

24. Endo, T., Ikeo, K. & Gojobori, T. Large-scale search for genes on which positive selection may operate. *Mol. Biol. Evol.* **13,** 685–690 (1996).

25. Ohta, T. & Gillespie, J. H. Development of neutral nearly neutral theories. *Theor. Popul. Biol.* **49,** 128–142 (1996).

26. Li, W. & others. *Molecular evolution.* (Sinauer associates incorporated, 1997).

27. Lu, J. & Wu, C.-I. Weak selection revealed by the whole-genome comparison of the X chromosome and autosomes of human and chimpanzee. *Proc. Natl. Acad. Sci.* **102,** 4063–4067 (2005).

28. Chen, B. *et al.* Quasi-neutral molecular evolution — When positive and negative selection cancel out. *bioRxiv* 330811 (2018). doi:10.1101/330811

29. Clark, A. G. *et al.* Evolution of genes and genomes on the Drosophila phylogeny. *Nature* **450,** 203–218 (2007).

30. Yang, L. & Gaut, B. S. Factors that contribute to variation in evolutionary rate among Arabidopsis genes. *Mol. Biol. Evol.* **28,** 2359–2369 (2011).

31. Consortium, M. G. S. & others. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420,** 520 (2002).

32. Jordan, I. K., Rogozin, I. B., Wolf, Y. I. & Koonin, E. V. Microevolutionary genomics of bacteria. *Theor. Popul. Biol.* **61,** 435–447 (2002).

33. Yang, Z. & Bielawski, J. P. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15,** 496–503 (2000).

34. Nielsen, R. *et al.* PLoS BIOLOGY A Scan for Positively Selected Genes in the Genomes of Humans and Chimpanzees. *PLoS Biol.* **3,** e170 (2005).

35. Laayouni, H. *et al.* Recent Positive Selection Has Acted on Genes Encoding. *Genome Biol. Evol.* **7,** 1141–1154 (2015).

36. Gayà-Vidal, M. & Albà, M. M. Uncovering adaptive evolution in the human lineage. *BMC Genomics* **15,** 599 (2014).

37. Feldmeyer, B., Greshake, B., Funke, E., Ebersberger, I. & Pfenninger, M. Positive selection in development and growth rate regulation genes involved in species divergence of the genus Radix. *BMC Evol. Biol.* 1–11 (2015). doi:10.1186/s12862-015-0434-x

38. Welch, J. J. Estimating the genome-wide rate of adaptive protein evolution in Drosophila. *Genetics* **837,** 821–837 (2006).

39. Biswas, S. & Akey, J. M. Genomic insights into positive selection. *TRENDS Genet.* **22,** 437–446 (2006).

40. Hartl, D. L., Clark, A. G. & Clark, A. G. *Principles of population genetics*. **116,** (Sinauer associates Sunderland, 1997).

41. Wang, H.-Y. *et al.* Is the evolution in tumors Darwinian or non-Darwinian? *Natl. Sci. Rev.* **5,** 15–17 (2017).

42. Yang, Z. Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. *J. Mol. Evol.* **51,** 423–432 (2000).

43. Nielsen, R. & Yang, Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148,** 929–936 (1998).

44. Zhang, J., Nielsen, R. & Yang, Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* **22,** 2472–2479 (2005).

45. Liow, L. H., Van Valen, L. & Stenseth, N. C. Red Queen: From populations to taxa and communities. *Trends Ecol. Evol.* **26,** 349–358 (2011).

46. Chen, Q. & Wu, C. I. A nearly-universal rule of small-step evolution driven by negative selection on codons. *unpublished*

47. Tang, H., Wyckoff, G. J., Lu, J. & Wu, C.-I. A universal evolutionary index for amino acid changes. *Mol. Biol. Evol.* **21,** 1548–1556 (2004).

48. Pool, J. E. *et al.* Population Genomics of Sub-Saharan Drosophila melanogaster: African Diversity and Non-African Admixture. *PLoS Genet.* **8,** (2012).

49. Goodstein, D. M. *et al.* Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res.* **40,** (2012).

50. Weigel, D. & Mott, R. The 1001 genomes project for Arabidopsis thaliana. *Genome Biol.* **10,** 107 (2009).

51. Fay, J. C., Wyckoff, G. J. & Wu, C. I. Positive and negative selection on the human genome. *Genetics* **158,** 1227–34 (2001).

52. Upton, C., Mossman, K. & McFadden, G. Encoding of a homolog of the IFN-gamma receptor by myxoma virus. *Science (80-.).* **258,** 1369–1372 (1992).

53. Yang, Z., Wong, W. S. W. & Nielsen, R. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **22,** 1107–1118 (2005).

Table 1 - Proportion of adaptively evolving genes identified by two tests (P < 0.05)

Gene Category	MK	PAML	Expected overlap	Observed overlap
Drosophila				
Individual genes	**3.43%** (186/5425)	**2.67%** (145/5425)	**0.09%** (4.97/5425)	**0.17%** (9/5425)
Supergenes[a]	**56.00%** (112/200)	**18.00%** (36/200)	**10.10%**	**10.0%** (20/200)
Component genes[b]	**5.04%** (158/3132)	**5.77%** (60/1040)	**0.29%**	**0.48%** (3/619)
Arabidopsis				
Individual genes	**1.12%** (145/12975)	**3.89%** (505/12975)	**0.04%** (5.55/12975)	**0.11%** (14/12975)
Supergenes	**8.20%** (41/500)	**25.60%** (128/500)	**2.10%**	**2.00%** (10/500)
Component genes	**3.63%** (38/1048)	**7.44%** (246/3306)	**0.27%**	**0.78%** (2/258)

[a] Supergenes are the concatenations of 20-30 neighboring genes by physical location. See Table S1 for supergenes concatenated by gene function.

[b] Component genes are individual genes within supergenes that have passed the MK and/or PAML tests.

Table 2 - Proportion of adaptively evolving genes identified by two tests ($P^2 < 0.05$, i.e. $P < 0.224$)

	MK	MK-PAML overlap	PAML	Total
Drosophila				
No. of genes	**824**	**91**[d]	**353**	**5425**
Expected overlap	/	**53**	/	/
Proportion of adaptive changes by MK[a]	**0.69**	**0.67**	**0.32**	**0.26**
No. of adaptive sites per gene by MK (A1)	**14.98**	**19.94**	**5.71**	**2.84**
No. of adaptive sites per gene by PAML (A2)[b]	**10.93**	**14.65**	**9.27**	**5.71**
No. of adaptive sites per gene by PAML (A2')[c]	**3.19**	**8.62**	**6.24**	**1.79**
Arabidopsis				
No. of genes	**1014**	**119**[e]	**1172**	**12975**
Expected overlap	/	**91.60**	/	/
Proportion of adaptive changes by MK	**0.69**	**0.67**	**0.06**	**0.04**
No. of adaptive sites per gene by MK (A1)	**19.36**	**28.98**	**1.97**	**0.84**
No. of adaptive sites per gene by PAML (A2)	**14.24**	**20.36**	**12.74**	**10.15**
No. of adaptive sites per gene by PAML (A2')	**3.59**	**11.51**	**8.13**	**2.44**

[a] Proportion of adaptive changes is done using Shapiro *et al.*'s method of correction[22].

[b] A2 is based on PAML-M2a model.

[c] A2' is based on PAML-BEB model.

[d] $P < 10^{-6}$, given 53.6 as the expected value.

[e] $P < 0.0025$, given 91.6 as the expected value.

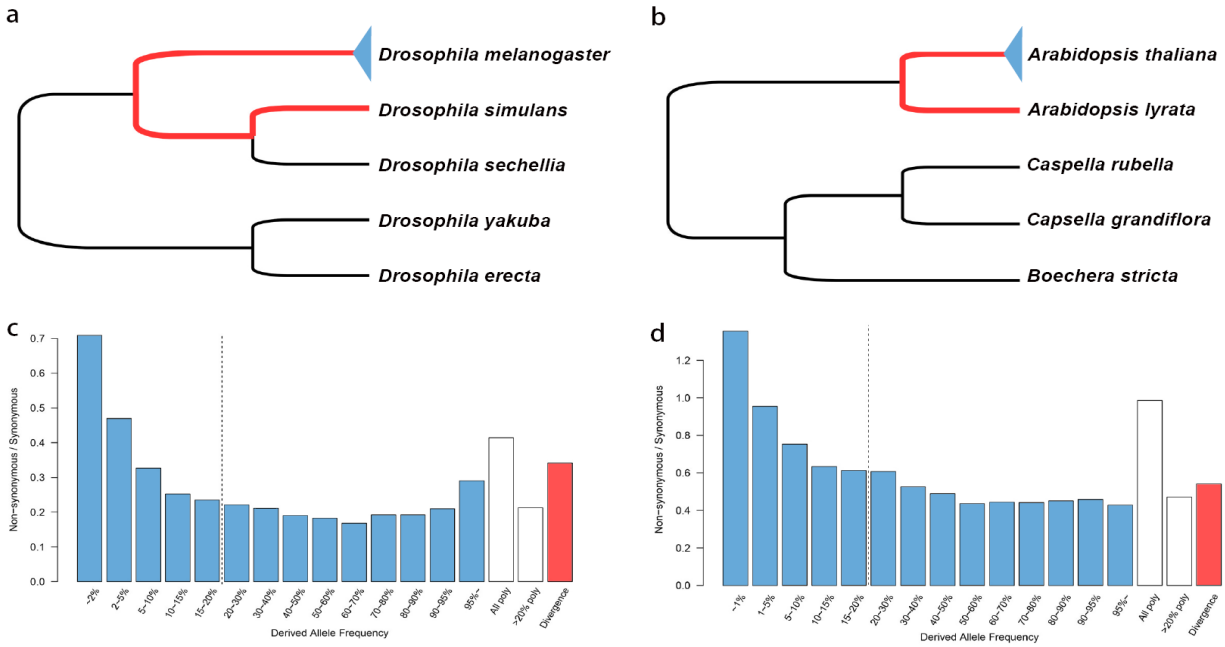

Figure 1. Between-species divergence and within-species polymorphism for detecting positive selection. a and **b**, Phylogeny of *Drosophila* and *Arabidopsis* species. Both the MK and PAML tests are forced to detect positive selection along the red-lined branch. The MK test uses the polymorphism (indicated by the blue triangle) for reference. The references for PAML are described in Methods. **c** and **d**, The A/S ratio as a function of the mutant frequency in *D. melanogaster* and *A. thaliana*, where A is non-synonymous, and S is synonymous polymorphism. The dashed line, separating low and high frequency bins, is placed where the A/S ratio reaches a steady level. Open bars on the right are, respectively, A/S ratios for all bins and for high-frequency bins. The divergence A/S ratio is shown as the red bar.

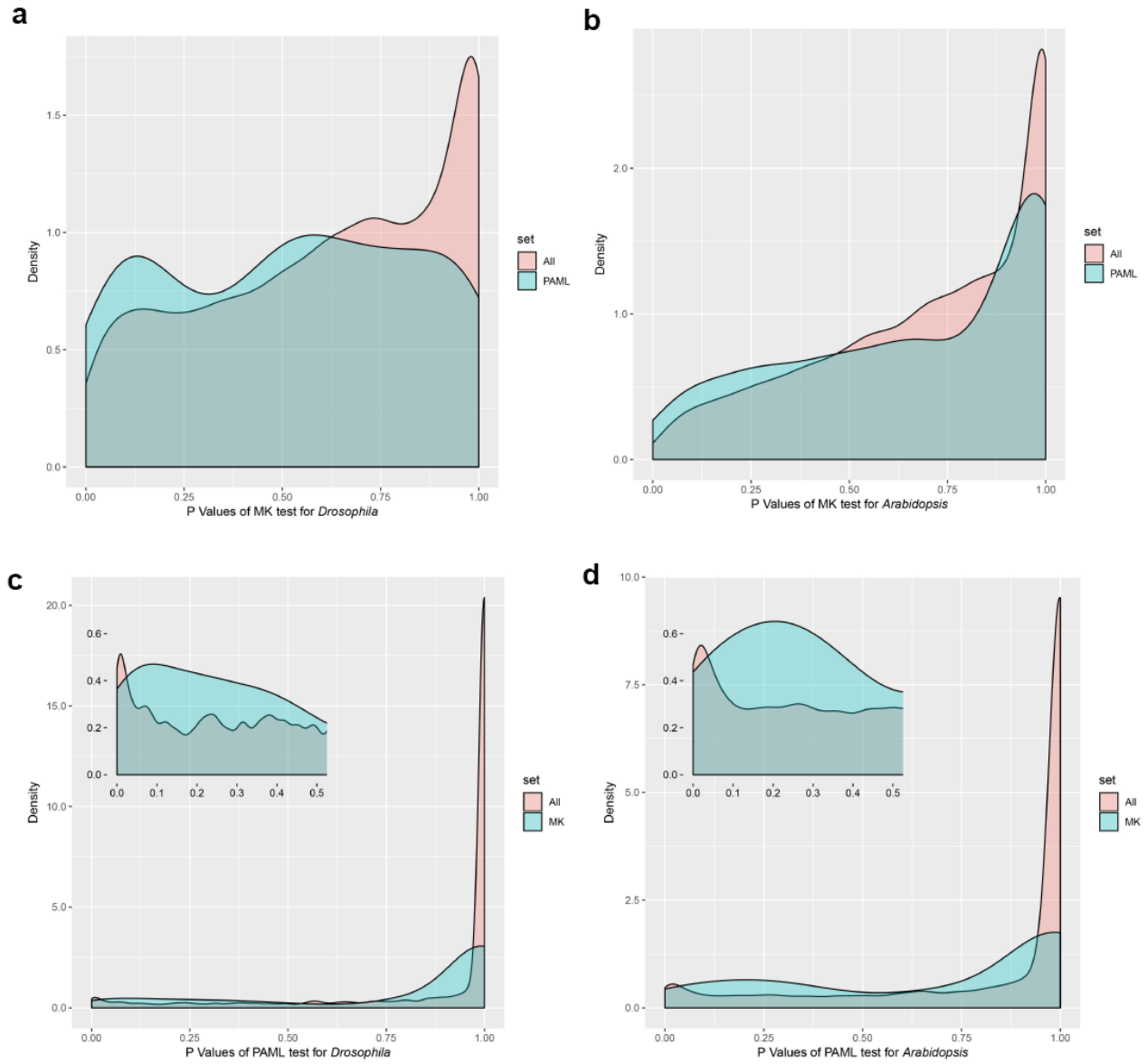

Figure 2. P value distributions of the MK and PAML test. a and **b**, P values of the MK test for *Drosophila* and *Arabidopsis*. The distribution for all genes is shown in red and the distribution for genes pre-filtered by the PAML test is shown in blue. **c** and **d**, P values of the PAML test. These two panels are the mirror images of panels (a-b) with MK and PAML switched.