Conflicting signals of adaptive molecular evolution
- Where does the neutral theory stand after 50 years?

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) or S.S. (lssssh@mail.sysu.edu.cn).

Abstract

Measuring positive selection in 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 McDonald-Kreitman (MK) test[1] compares divergence and polymorphism data, while the PAML test[2,3] analyzes multi-species divergence. We used these two methods concurrently 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 more than 100 adaptively evolving genes but the two sets hardly overlap. To rule out false negatives, we merged 20 - 30 genes into "supergenes", 8% - 56% of which yield adaptive signals. Nevertheless, the joint calls still do not overlap. The technical explanations of high false negatives or positives can be rejected. The most likely explanation is the relaxation of negative selection, which results in patterns resembling positive selection and is easily testable by using multi-species polymorphisms. When so tested, *Arabidopsis* (but not *Drosophila*) data fit this hypothesis. PAML and MK may indeed identify distinct classes of genes in *Drosophila*. However, this "both are right" explanation is valid only if positive and negative selection tend to affect the same targets, thus contradicting the (untested) conventional view. In conclusion, the acceptance of adaptive DNA evolution, and hence the rejection of the neutral theory, should be suspended until negative selection is rigorously analyzed.

Introduction

Detecting adaptive evolution in DNA sequences is one of the central tasks of molecular evolutionary studies. The neutral theory[4] holds that such signals are 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[2,3,5–11]. Indeed, the theory may have been rejected as often as new methods are proposed[9,10,12–15]. Such methods fall into two broad classes. One class attempts to

detect positive selection that operates within populations[5,6,8,12,16]. The other focuses on positive selection that operates in the longer term, i.e., the divergence between species[1,10,13,17–22]. Methods of either class may use data of both polymorphism and divergence[1,8,10,13,19–22]. Positive selection signals could be abundant between species but undetectable within populations, or vice versa. Even partial rejection of the neutral theory is informative about molecular evolution[23–25].

In this study, we focus on the between-species tests. 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 changes are under stronger selection than synonymous substitutions. In the absence of selection, R = Ka/Ks ~ 1, the hallmark of neutral evolution[23–27]. In among-species comparisons, genome-wide R ranges mainly between 0.05 and 0.25[27–31], 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 significantly greater than 1[22,23,32,33].

The two prevailing methods that relax the requirement for R > 1 over the entire gene are the MK (McDonald-Kreitman)[10] and the PAML (Phylogenetic analysis by maximum likelihood)[2,3] tests. Each test relies on a different set of assumptions that cannot be easily verified (see below). The purpose of this comparative study is to examine these underlying assumptions and to cross-validate the conclusions. While some previous studies have employed the two tests[33–38], this may be the first study that applies them side-by-side to infer positive selection on the same set of genes and along the same phylogenetic branch. A larger literature on this topic is briefly commented on in the Supplementary Information.

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, when averaged over all sites, is the joint outcome of the two opposing forces. We use the basic population genetic theory to outline the idea:

$$R = Ka/Ks = (1 - p - q) + p\,[\,2N\,f\,(N, s_1)] + q\,[\,2N\,f\,(N, s_2)] \qquad \text{Eq. (1)}$$

where p and q are the proportion of advantageous and deleterious mutations, respectively[24,25,39]; $f(N, s) = (1 - e^{-s})/(1 - e^{-2Ns})$ is the fixation probability of a mutation with a selective coefficient s that can be > 0 (denoted by s_1) or < 0 (s_2) and N is the effective population size. For example, if Ka/Ks = 0.2, then the null hypothesis is the neutrality with q = 0.8, p = 0 and $f(N, s_2) = 0$ (i.e., no fixation of deleterious mutations). The alternative hypothesis would be adaptive evolution with p > 0 and $f(N, s_1) > 0$ (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, marked by red lines in 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[13,17]. Second, the estimation of negative selection is not straightforward. The pA/pS ratio would decrease as the variant frequency increases and may increase again when the mutant frequency approaches 1[11,13,20,40]. 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 (see Methods)[1,10,13,17–22]. 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[20,21].

The other widely used approach to the estimation of adaptive evolution 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. In particular, the proportion of adaptive sites that have a higher non-synonymous than neutral rate is estimated by PAML. There are three sub-models, two of which (the site model and the branch-site model) are used here. The site model identifies sites with an increase or decrease in non-synonymous substitutions in the entire phylogeny[2,3]. The branch-site model compares sites of a pre-selected branch (the foreground) to other sites on all branches as well as the same sites on other branches (the background)[41–43].

Despite the very different approaches, the MK and PAML tests can be used to answer the same question – How much adaptive evolution has happened in the chosen genes on 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 is necessary to compare it to each of the two PAML sub-models. If the MK test identifies genes that are generally prone to evolving adaptively, the proper comparison would be the PAML site model. Alternatively, if the adaptation is specific to a specific branch, then the branch-site model 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 Supplementary Information. 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 determined the distribution of the P values across genes. The MK test P values were obtained from the Fisher's exact test on site count contingency tables. The likelihood ratio test was used to obtain PAML P values. The P value distributions are shown in the four panels of Fig. 2 for two taxa and two tests. The distribution is concentrated above P = 0.8 (the MK test for *Drosophila*) and P = 0.9 (the other panels). This concentration means that a very large percentage of genes show no detectable signal, partly because most genes 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[13,19,20].

Fig. 2 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 from 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 derived from loci 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, reflected by the leftward shift in the P value distribution. The same is true in Fig. 2c-2d where pre-filtering by MK reduces the PAML test P values. The two tests are indeed correlated, but only weakly. This is also true in Extended Data Fig. 1, where the branch-site model of PAML is used.

We now enumerate the overlap between the two tests by comparing the candidate adaptive genes with $P < 0.05$. Given the P value distributions shown in Fig. 2, these genes are merely the most likely candidates proposed 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 between these two sets contains only nine genes. Although the observed overlap is higher than the expected 4.97 ($P < 0.1$, 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 but only 14 genes are called by both tests. Again, the observed overlap is significantly higher than the expected 5.55 ($P < 0.01$, Fisher's exact test) but the actual overlap is minimal. A simple explanation for the non-overlap is a high false-negative rate. In other words, each test may have detected only a small fraction of the true adaptive genes.

The analysis of supergenes and their component genes

False negatives should be common in individual genes harboring few substitutions. To overcome this statistical limitation, we created artificial "supergenes" by merging 20 to 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 due to low substitution numbers, but at the risk of diluting true adaptive signal. We present the results based on the concatenations of neighboring genes in Table 1. In *Drosophila* and *Arabidopsis*, 200 and 500 supergenes are created respectively. The results based on the merger by gene ontology are similar (See Extended Data Table 1).

Our gene merger approach may create biases in the MK test, as pointed out before[20]. 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 used the modified MK test to infer positive selection in merged genes[20]. In *Drosophila*, 112 of the 200 supergenes reject the MK test null hypothesis at the 5% level, and 36 of the 200 significantly deviate from the PAML null (Table 1). The two tests detect far more adaptive supergenes than individual genes: 56% (MK) and 18% (PAML). What is perplexing is that the overlap between the two sets is 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 pass the MK test at the 5% level and 25.6% of supergenes reject the PAML null. The PAML test in *Arabidopsis* 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 examined the component genes within each adaptive supergene. In *Drosophila*, the 112 supergenes passing the MK test contain 3132 component genes (Table 1), among which 158 genes are significant when tested individually. Likewise, 60 out of 1040 component genes are identified by PAML. Between the two subsets of genes (3132 and 1040), 619 genes are common and only three 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 two 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. Comparable analyses using the PAML branch-site model (Extended Data Table 2) yield results similar to those in 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 this relaxed stringency are given in Table 2.

The MK test identifies 824 and PAML 353 genes in *Drosophila*. These sets have 91 loci in common, whereas the expected overlap is 53.6 ($P < 10^{-7}$, Fisher's exact test). In *Arabidopsis*, the two tests yield 1014 and 1172 genes with an overlap of 119 genes, significantly higher than the expected number of 91.6 ($P < 0.002$, Fisher's exact test). Hence, the joint call of adaptive genes accounts for 10.1% (119/1172) to 25.8% (91/353) of the loci identified by each single test. A gene identified by one test as adaptive has a 10% to 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 look at A1, the average number of adaptive sites per gene estimated using the MK test. A1 doubles from 2.84 to 5.71 when genes are pre-screened using PAML in *Drosophila* and increases from 14.98 to 19.94 in loci identified

by both tests compared to just MK. 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. The number of adaptive sites per gene can be calculated using two methods in PAML (A2 and A2' in Table 2[41,44]; 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 the choice of model. The number of A2 sites increases from 5.71 to 10.93 after MK pre-screening in *Drosophila* (Table 2) and from 9.27 to 14.65 when focusing on the loci identified by both PAML and the MK test, compared to PAML alone. The same trend is observed in *Arabidopsis* (Table 2): an increase from 10.15 to 14.24 after MK test pre-screening and 12.74 for PAML only vs 20.36 for genes identified by both tests. Again, a pre-screen by MK helps PAML performance.

The results are similar when we use the PAML branch-site model rather than the site model (see Extended Data Table 3). 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

It is surprising that the two widely used tests are poorly concordant in detecting adaptively evolving genes. We first explore, and reject, methodological explanations for this observation.

Methodological explanations and statistical variations

i) Both tests have high false negative rates: False negatives could be a consequence of the "nearly neutral" evolution proposed by Ohta (1992)[23]. While near-neutrality is often used to indicate slightly deleterious variants, Ohta (1992)[23] and Ohta and Gillespie (1996)[24] have suggested that advantageous mutations may often be nearly neutral as well. However, when the tests are applied to supergenes, which should yield low false negatives given the detection rate up to 56%, the overlap is still no higher than random occurrences.

ii) Both tests have high false positive rates: This does not seem plausible in the comparison of individual genes since neither test identifies more than 3.9% of genes at the nominal cutoff of 5% for either *Drosophila* or *Arabidopsis* (Table 1), suggesting that both tests are conservative.

iii) Both tests have high false positive and false negative rates: This explanation assumes that one of the two tests is entirely unreliable. However, the two tests are comparable in performance. When PAML is done on genes selected by the MK test, the subset of genes yields much stronger signal than the full set. This is also true when the MK test is done on PAML-selected genes.

Biological explanations and the correlation between negative and positive selection

Since the observable evolution rate (R = Ka/Ks) is influenced by both positive and negative selection, any incorrect assumption about one would lead to the mis-estimation of the other. We

suggest that negative selection should be accurately estimated first as it follows relatively simple rules[45–47].

i) Variable negative selection interferes with the assessment of positive selection

All tests assume constant negative selection in the time frame of interest. In the MK test, negative selection is estimated from the polymorphism data and applied to all lineages. However, if negative selection within the species is stronger than the average prevalent in the past, reduced negative selection between species may be mis-interpreted as positive selection. This could happen if the population size increased recently[17,18]. PAML also assumes that proportions of sites under both positive and negative selection are constant across the whole phylogeny.

It is therefore crucial to ascertain the constancy of negative selection by comparing polymorphisms from related species. Between *A. thaliana* and *A. lyrata*, the constancy assumption can be rejected. The polymorphism A/S ratio is 0.400 and 0.605, respectively (see Supplementary Information for the estimation). Since the divergence A/S ratio is 0.493 (Extended Data Table 4), one would reach opposite conclusions depending on the polymorphism data chosen for comparison. Clearly, the variation in the strength of negative selection exceeds the effect of positive selection between *A. thaliana* and *A. lyrata*, rendering both MK and PAML inoperative. Unlike in *Arabidopsis*, polymorphism data in *D. melanogaster* and *D. simulans* do not reject the assumption of constant negative selection. The polymorphism A/S ratio in either species falls between 0.196 and 0.200, both of which being lower than the divergence A/S ratio of 0.320 (Extended Data Table 5).

Given the plausible constancy of negative selection, the MK and PAML tests should be assumed valid, making the non-overlapping results even more intriguing. One possibility is that the adaptive landscape is shifting, with most species continually evolving toward moving fitness peaks that are shifting like sand dunes. In this Red Queen landscape[48–50], positive selection is non-constant and different genes are evolving adaptively at different times. Although the Red Queen hypothesis could explain the results of our comparisons between MK and PAML, it is not testable at present.

ii) *Drosophila* data as the test ground for the “both are right” hypothesis

The last hypothesis we explore is that MK and PAML are complementary; in other words, the two sets of results could both be right but non-overlapping. Complementarity is plausible if the detection of positive selection is influenced by the strength of negative selection. With negative selection, the fixation probability decreases rapidly when the strength (Ns_2) increases (Fig. 3a). This strength reflects the functional importance (x) of the gene, or the part of the gene hit by mutations. For example, mutations in the heme-pocket of hemoglobin would have a high x whereas those in the backbone of the same protein would have a low x. This functional attribute may influence the performance of both MK and PAML.

In particular, PAML does not filter out signals of negative selection; hence, mutations associated with a smaller x could be more readily pushed above Ka/Ks > 1 by positive selection (see Fig. 3b) whereas genes under stringent constraints may require very strong positive selection to appear adaptive. Extended Data Fig. 2 indeed shows that genes with a higher polymorphism A/S ratio (i.e., weaker negative selection) tend to show a stronger signal of positive selection. In contrast, the MK test detects

positive selection by comparing the divergence A/S ratio with the polymorphism, pA/pS. Hence, a smaller pA/pS that reflects stronger negative selection would permit better detection of positive selection by the MK test (see Fig. 3b).

The opposite patterns of Fig. 3b provide a rationale for the "both are right" hypothesis, but the low overlap between MK and PAML may still depend on the abundance of adaptive mutations as a function of x (Fig. 3c vs 3d). Fig. 3c illustrates the conventional view whereby positive selection tends to act on genes where negative selection is weaker (i.e., smaller x). This convention is accepted by both the selectionism and neutralism schools. According to the neutral theory, negative selection is weaker when a gene, or the mutation, is functionally less important (Rules 2 and 3 of neutrality; see p. 103 of Kimura, 1983)[45]. In parallel, in Fisher's geometric model of adaptive evolution[51], positive selection favors mutations of smaller functional changes. In the scheme of Fig. 3c, the MK and PAML results overlap substantially, thus failing to support the "both are right" hypothesis.

In contrast to Fig. 3c, Fig. 3d presents an opposite scenario by postulating that both positive and negative selection should become stronger as x increases. After all, larger functional differences may be more "discernible" to selection, regardless of the direction of selection. The overlaps between MK and PAML would indeed be much smaller in Fig. 3d (47% of the detected genes overlap) than in Fig. 3c (87% overlap). Although the theoretical overlap shown in Fig. 3d is still far larger than the observed value, it is possible to reduce the theoretical overlap with more extreme parameter values. The key message is that the distribution in Fig. 3d, which is far more likely to support the complementarity hypothesis, is not compatible with the conventional view (Fig. 3c). Some recent efforts have challenged this conventional view[50].

Conclusions

In the search for the signals of positive selection, the published literature has investigated negative selection rigorously. Even the central assumption of constant negative selection, easily testable using multi-species polymorphisms, has rarely been affirmed. A more fundamental issue is the correlation between positive and negative selection. The conventional view will not resolve the discordance between the MK and PAML tests (although the convention itself has not been empirically tested). In conclusion, the signals of positive selection and, hence, the rejection of the neutral theory, will remain uncertain until negative selection has been rigorously measured.

Acknowledgements

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).

Online Methods

DNA sequence data

Pre-aligned unique *Drosophila* transcript sequences were downloaded from Flybase[52] (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 2[53]. 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[54]. Genome-wide *A. thaliana* polymorphism data were obtained from the 1001 Genomes Project[55]. 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 PAML site module)[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 on the MK test, we only used common mutations with derived allele frequencies larger than 0.2, as is done previously[20,56]. 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. 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[20]. Hence, we used it in the calculations.

Ideally, the number of A and S polymorphic sites should reflect only neutral variation. However, as can be seen in Fig. 1c-1d, A often includes low-frequency deleterious mutations and, perhaps, high-frequency advantageous mutations[8,18,20,40]. 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[13,19–21] to more accurately measure the polymorphic A's. These methods are mostly ad hoc in nature. Sawyer and Hartl[1] propose a more robust approach to this problem by directly estimating the intensity 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 chi-squared test (df = 2).

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 Supplementary Information.

In the analysis of both models, we also employed Bayes empirical Bayes (BEB)[43] estimates, which are 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.

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. Phylogenetic analysis by maximum likelihood (PAML). *Univ. Coll. London* **44,** 1–37 (2003).
4. Kimura, M. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* **267,** 275–276 (1977).
5. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123,** 585–595 (1989).
6. Fu, Y.-X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147,** 915–925 (1997).
7. Sabeti, P. C. *et al.* Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419,** 832–837 (2002).
8. Fay, J. C. & Wu, C.-I. Hitchhiking under positive Darwinian selection. *Genetics* **155,** 1405–1413 (2000).
9. Hudson, R. R., Kreitman, M. & Aguade, M. A test of neutral molecular evolution based on nucleotide data. *Mol. Biol. Evol.* **32,** 1365–1371 (2015).
10. McDonald, J. & Kreitman, M. Adaptative protein evolution at the Adh locus in Drosophila. *Nature* **354,** 293–295 (1991).
11. 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.* **104,** 6504–6510 (2007).
12. Zeng, K., Shi, S., Fu, Y.-X. & Wu, C.-I. Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics* **174,** 1–33 (2006).
13. Fay, J. C., Wyckoff, G. J. & Wu, C.-I. Testing the neutral theory of molecular evolution with genomic data from Drosophila. *Nature* **415,** 1024–1026 (2002).
14. Kern, A. D. & Hahn, M. W. The Neutral Theory in Light of Natural Selection. *Mol. Biol. Evol.* **35,** 1366–1371 (2018).
15. 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).
16. Charlesworth, B. & Charlesworth, D. Neutral Variation in the Context of Selection. *Mol. Biol. Evol.* **35,** 1359–1361 (2018).
17. Eyre-Walker, A. Changing effective population size and the McDonald-Kreitman test. *Genetics* **162,** 2017–2024 (2002).
18. Fay, J. C. & Wu, C.-I. Sequence divergence, functional constraint and selection in protein evolution. *Annu. Rev. Genomics Hum. Genet.* **4,** 213–235 (2003).
19. Messer, P. W. & Petrov, D. A. Frequent adaptation and the McDonald--Kreitman test. *Proc. Natl. Acad. Sci.* **110,** 8615–8620 (2013).
20. Shapiro, J. A. *et al.* Adaptive genic evolution in the Drosophila genomes. *Proc. Natl. Acad. Sci.* **104,** 2271–2276 (2007).
21. Smith, N. G. C. & Eyre-Walker, A. Adaptive protein evolution in Drosophila. *Nature* **415,** 1022–1024 (2002).
22. Endo, T., Ikeo, K. & Gojobori, T. Large-scale search for genes on which positive selection may operate. *Mol. Biol. Evol.* **13,** 685–690 (1996).
23. Ohta, T. The Nearly Neutral Theory of Molecular Evolution. *Annu. Rev. Ecol. Syst.* **23,** 263–286 (1992).

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

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

26. 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).

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

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

29. Yang, L. & Gaut, B. S. Factors that contribute to variation in evolutionary rate among Arabidopsis genes. *Mol. Biol. Evol.* **28,** 2359–69 (2011).

30. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420,** 520–562 (2002).

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

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

33. Nielsen, R. *et al.* A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* **3,** 0976–0985 (2005).

34. Luisi, P. *et al.* Recent positive selection has acted on genes encoding proteins with more interactions within the whole human interactome. *Genome Biol. Evol.* **7,** 1141–1154 (2015).

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

36. 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.* **15,** 1–11 (2015).

37. Welch, J. J. Estimating the genomewide rate of adaptive protein evolution in drosophila. *Genetics* **173,** 821–837 (2006).

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

39. Hartl, D. L. & Clark, A. G. *Principles of Population Genetics. Biometrics* **46,** (Sinauer associates Sunderland, 1990).

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

41. 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).

42. Yang, Z. & Nielsen, R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* **19,** 908–917 (2002).

43. Yang, Z., Wong, W. S. W. & Nielsen, R. Bayes Empirical Bayes(BEB) Inference of Amino Acid Sites Under Positive Selection. *Mol. Biol. Evol.* **22,** 1107–1118 (2005).

44. 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).

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

46. 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).

47. Xu, S. *et al.* The origin, diversification and adaptation of a major mangrove clade (Rhizophoreae) revealed by whole-genome sequencing. *Natl. Sci. Rev.* **4,** 721–734 (2017).

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

49. Van Valen, L. A new evolutionary law. *Evol. Theory* **1,** 1–30 (1973).

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

51. Wagner, G. P. & Zhang, J. The pleiotropic structure of the genotype--phenotype map: the evolvability of complex organisms. *Nat. Rev. Genet.* **12,** 204 (2011).

52. Clark, A. *et al.* Genomics on a phylogeny: Evolution of genes and genomes in the genus Drosophila. *Nature* **450,** 203–218 (2007).

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

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

55. Weigel, D. & Mott, R. The 1001 Genomes Project for Arabidopsis thaliana. *Genome Biol.* **10,** 107 (2009).

56. Akashi, H. Within- and between-species DNA sequence variation and the 'footprint' of natural selection. *Gene* **238,** 39–51 (1999).

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]	**56.0%** (112/200)	**18.0%** (36/200)	**10.1%**	**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)	**25.6%** (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 concatenations of 20-30 neighboring genes by physical location. See Extended Data Table 1 for supergenes concatenated by gene function.

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

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	/	53.6	/	/
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	/	91.6	/	/
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[20].

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

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

[d] $P < 10^{-7}$ by Fisher's exact test, given 53.6 as the expected value.

[e] $P < 0.002$ by Fisher's exact test, given 91.6 as the expected value.

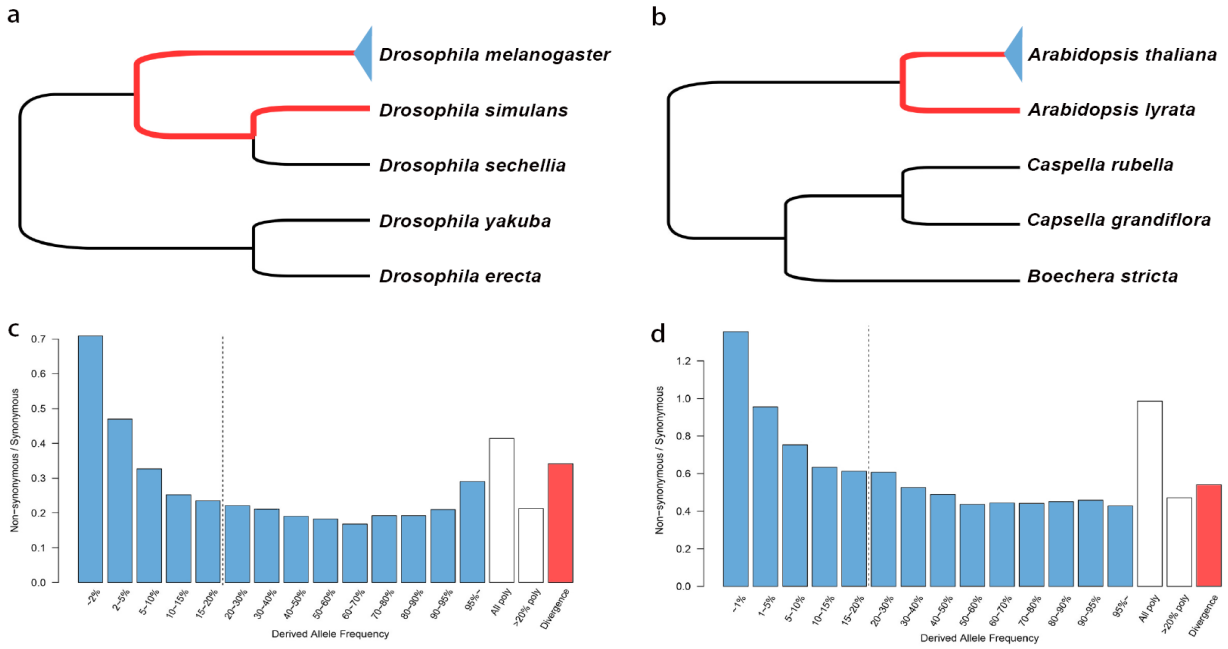


Figure 1. Between-species divergence and within-species polymorphism for detecting positive selection. a and **b**, Phylogeny of *Drosophila* and *Arabidopsis* species. Both the MK and PAML tests are forced to detect positive selection along the branches marked by red. The MK test uses polymorphisms (indicated by the blue triangle) for reference. The reference for PAML is 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.

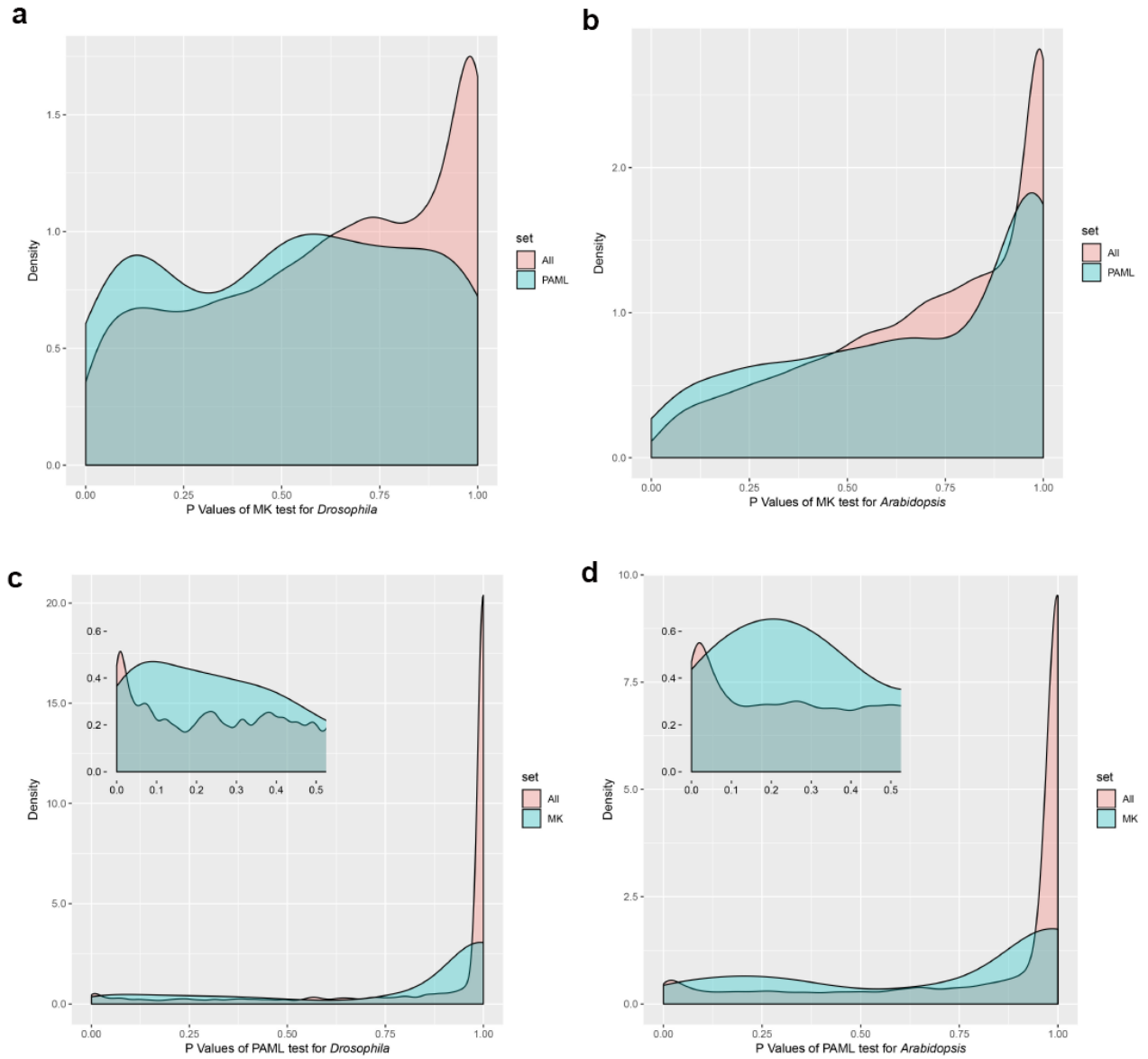


Figure 2. P value distributions of the MK and PAML test. a and **b**, P values of the MK test for *Drosophila* and *Arabidopsis*. The distribution for all genes is shown in red and the distribution for genes pre-filtered by the PAML test is shown in blue. **c** and **d**, P values of the PAML test. Results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

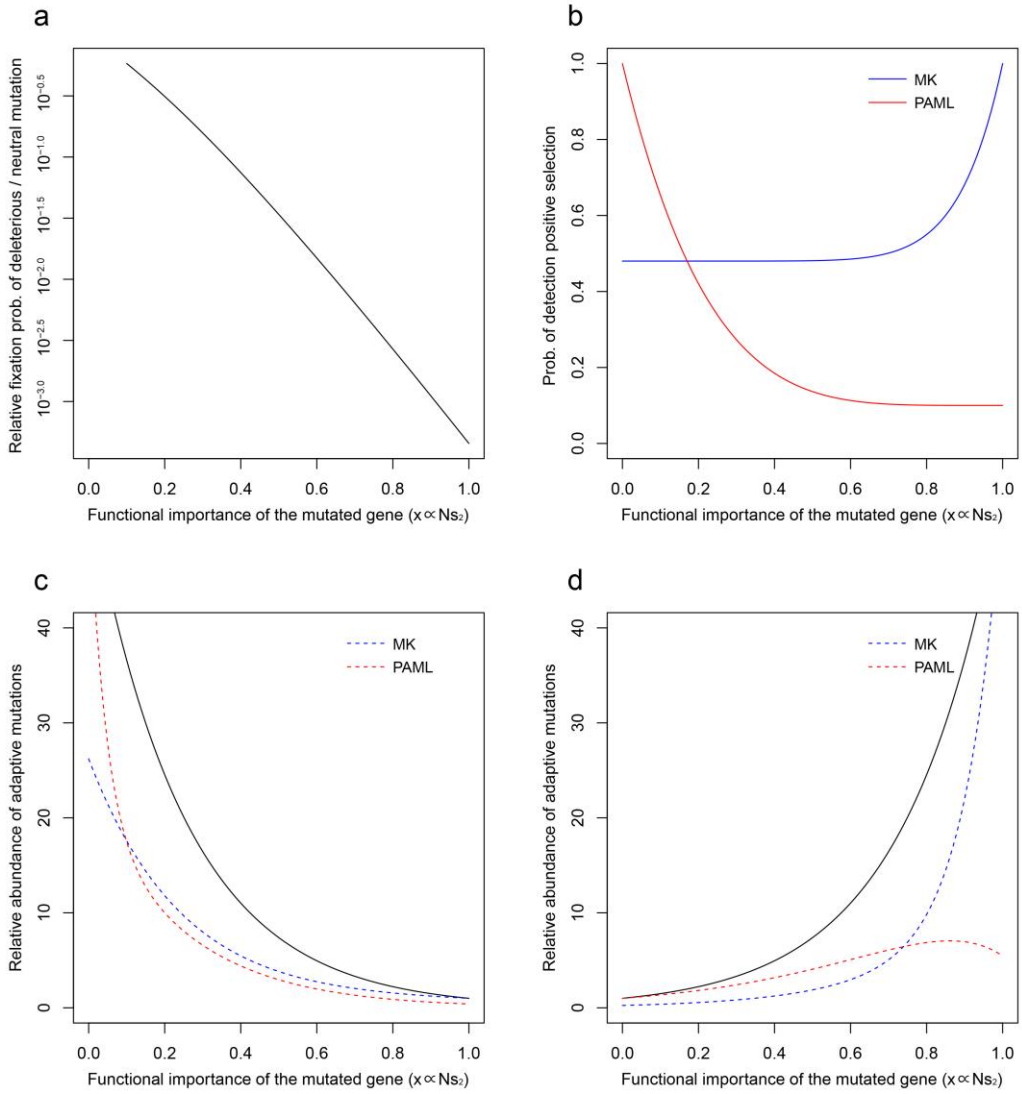


Figure 3. Models for detecting positive selection under different strength of negative selection. x, an arbitrary scale for the functional importance of the mutated gene, determines the strength of negative selection. **a**, The fixation probability of deleterious mutations based on Eq. (2). **b**, The probability of detecting adaptive mutations (P^+) by MK or PAML. For MK, the blue line follows the equation $P_{mk}(x) = (1-c) + cx^i$ and, for PAML, the red line follows the equation $P_{paml}(x) = (1-c) + c(1-x)^i$. The assumption is that MK and PAML has the maximal efficiency at x = 1 and x = 0, respectively (see text). **c**, The relative number of adaptive mutations, $M = e^{4(1-x)}$, as a function of x (black line). The blue line, the product of $M P^+ = M [(1-c) + cx^i]$, denotes the relative number of genes detected by MK. The red line, $M [(1-c) + c(1-x)^i]$, denotes the corresponding number detected by PAML. It is assumed that both tests detect 50% of the adaptive genes. Note that the high detection rate is based on the number of adaptively-evolving genes whereas the empirical observations in Table 1 are based on the number of all genes. The overlap in detection [= Min ($P_{mk}(x)$, $P_{paml}(x)$)] between the two tests is 87% of the detected genes. **d**, same as c but $M = e^{4x}$. The MK test still detects 50% but the PAML test can only detect 30% due to the higher density of genes with a larger x. The overlap accounts for 44% of the MK detection. In short, the overlap between the two tests depends on the distribution of adaptive mutations as a function of x; hence, the overlap in Fig. 3d can be much smaller than that in Fig. 3c.

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

(Same as Table 1 but genes were merged into supergenes by ontology)

Gene Category	MK	PAML (site model)	Expected overlap	Observed overlap
Drosophila				
Supergenes[a]	**31.52%** (58/184)	**14.67%** (27/184)	**4.62%**	**8.70%** (16/184)
Component genes[b]	**6.01%** (51/849)	**7.54%** (36/477)	**0.45%**	**0.00%** (0/306)
Arabidopsis				
Supergenes[a]	**10.57%** (48/454)	**19.38%** (88/454)	**2.05%**	**2.42%** (11/454)
Component genes[b]	**4.46%** (45/1008)	**7.19%** (184/2556)	**0.32%**	**0.29%** (1/341)

[a] Supergenes are the concatenations of genes of the same ontology.

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

Extended Data Table 2 - Proportion of adaptively evolving genes identified by two tests (P < 0.05)

(Same as Table 1 but using the PAML branch-site model)

Gene Category	MK	PAML (branch -site)	Expected overlap	Observed overlap
Drosophila				
Individual genes	**3.43%** (186/5425)	**5.40%** (293/5425)	**0.19%** (10.05/5425)	**0.35%** (19/5425)
Supergenes[a]	**56.00%** (112/200)	**18.00%** (36/200)	**10.08%**	**8.00%** (16/200)
Component genes[b]	**5.04%** (158/3132)	**9.41%** (92/978)	**0.47%**	**1.76%** (8/455)
Arabidopsis				
Individual genes	**1.12%** (145/12975)	**10.02%** (1300/12975)	**0.12%** (14.53/12975)	**0.16%** (21/12975)
Supergenes	**8.20%** (41/500)	**47.20%** (236/500)	**3.87%**	**2.20%** (11/500)
Component genes	**3.62%** (38/1048)	**12.24%** (750/6129)	**0.44%**	**1.36%** (4/295)

[a] Supergenes are concatenations of 20-30 neighboring genes by physical location.

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

Extended Data Table 3 - Proportion of adaptively evolving sites identified by two tests ($P^2 < 0.05$, i.e. $P < 0.224$)

(Same as Table 2 but using the PAML branch-site model)

	MK	MK-PAML overlap	PAML (branch-site)	Total
Drosophila				
No. of genes	824	127[d]	530	5425
Expected overlap	/	80.50	/	/
Proportion of adaptive changes by MK[a]	0.69	0.65	0.31	0.26
No. of adaptive sites per gene by MK (A1)	14.98	20.22	6.24	2.84
No. of adaptive sites per gene by PAML (A2)[b]	8.33	23.96	22.40	5.02
No. of adaptive sites per gene by PAML (A2')[c]	1.95	9.69	9.53	1.23
Arabidopsis				
No. of genes	1014	233[e]	1172	12975
Expected overlap	/	193.89	/	/
Proportion of adaptive changes by MK	0.69	0.69	0.06	0.04
No. of adaptive sites per gene by MK (A1)	19.36	24.07	1.50	0.84
No. of adaptive sites per gene by PAML (A2)	4.48	9.94	7.78	3.33
No. of adaptive sites per gene by PAML (A2')	2.72	9.00	7.69	2.06

[a] Proportion of adaptive changes is done using Shapiro *et al.*'s method of correction[20].

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

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

[d] $P < 10^{-7}$ by Fisher's exact test, given 80.5 as the expected value.

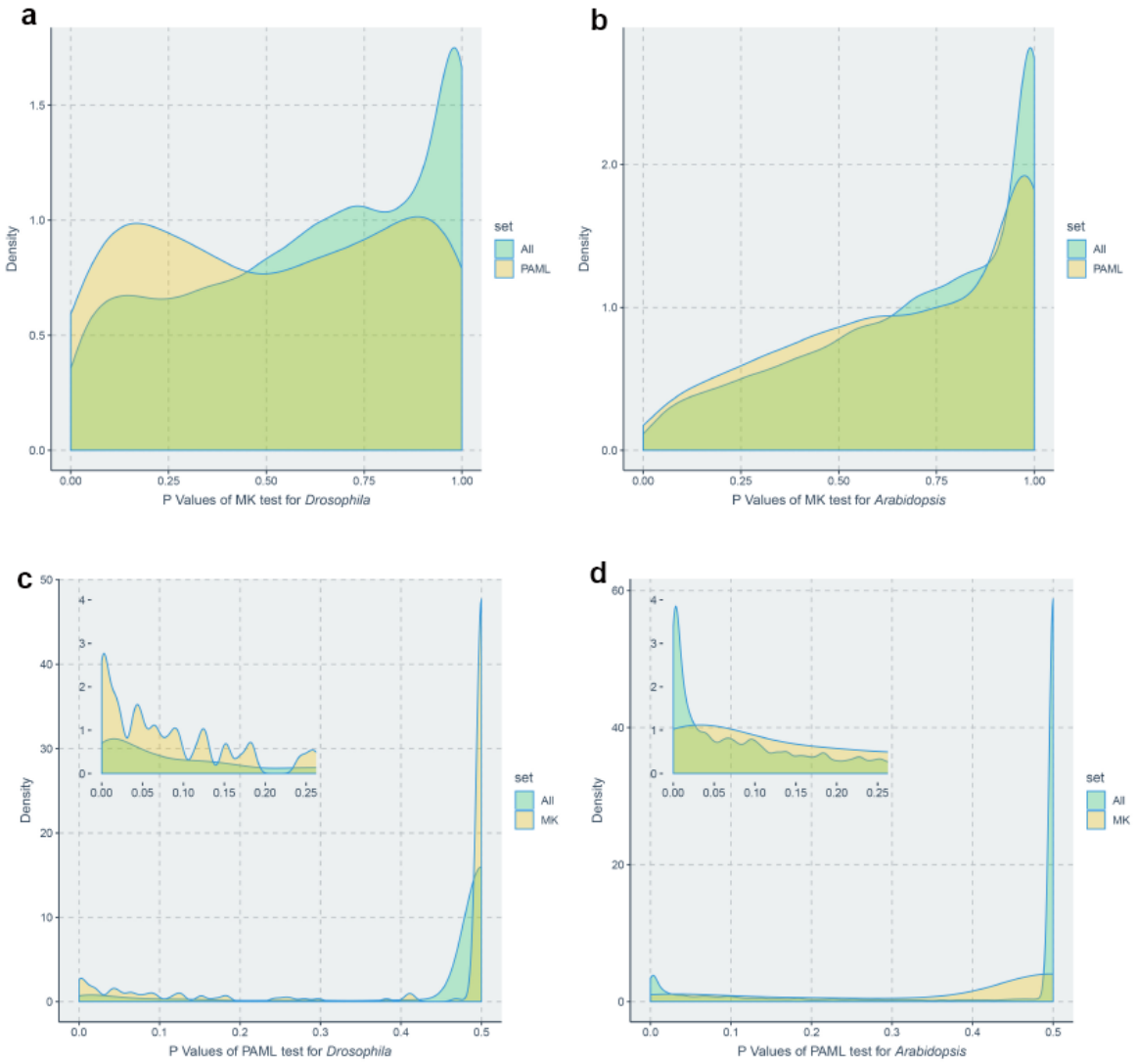
[e] $P < 10^{-10}$ by Fisher's exact test, given 193.9 as the expected value.

Extended Data Table 4 - The polymorphism A/S ratios of related species in *Arabidopsis*

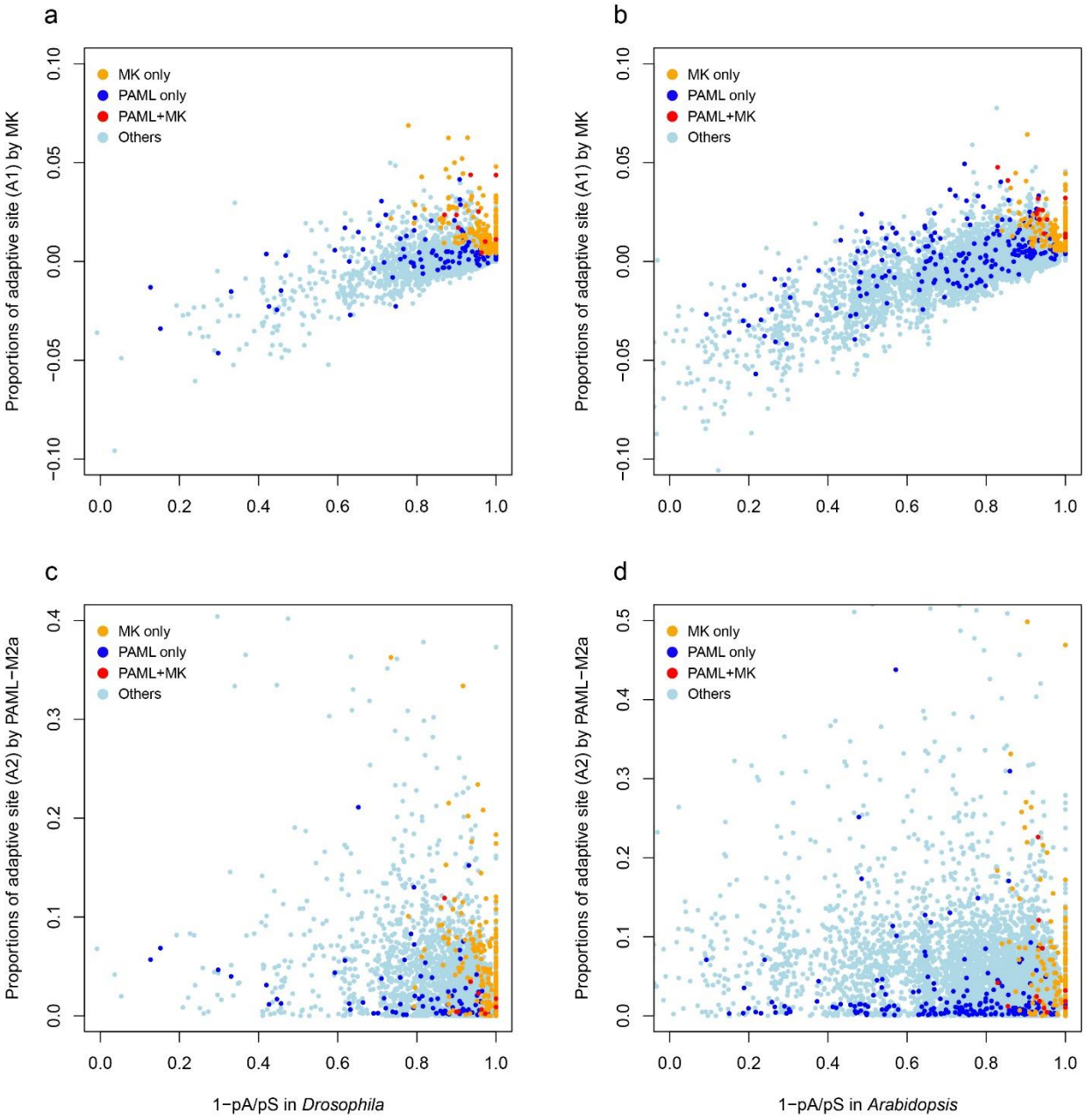
A/S Ratios	~0.1	0.1~0.2	0.2~0.3	0.3~0.5	0.5~0.8	0.8~0.9	0.9~	Poly	Fixed
Ancestral state inferred by dataset of five species: *A. thaliana*, *A. lyrata*, *Capsella grandiflora*, *C. rubella* and *Boechera stricta*									
A. thaliana	1.233	0.632	0.589	0.505	0.440	0.457	0.430	1.029	0.504
A. lyrata	1.082	0.950	0.854	0.760	0.671	0.637	0.599	0.773	0.479
Ancestral state inferred by the other species: *A. thaliana* and *A. lyrata*									
A. thaliana	1.275	0.661	0.610	0.513	0.428	0.425	0.400	1.029	0.493
A. lyrata	1.079	0.952	0.874	0.758	0.668	0.639	0.605	0.773	0.493

Extended Data Table 5 -The polymorphism A/S ratios of related species in *Drosophila*

A/S Ratios	~0.1	0.1~0.2	0.2~0.3	0.3~0.5	0.5~0.8	0.8~0.9	0.9~	Poly	Fixed
Ancestral state inferred by dataset of five species: *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba*, and *D. erecta*									
D. melanogaster	0.561	0.260	0.217	0.193	0.172	0.177	0.194	0.434	0.289
D. simulans	0.307	0.169	0.162	0.166	0.184	0.219	0.285	0.275	0.360
Ancestral state inferred by the other species: *D. melanogaster* and *D. simulans*									
D. melanogaster	0.570	0.266	0.220	0.192	0.171	0.168	0.200	0.434	0.320
D. simulans	0.313	0.171	0.164	0.169	0.176	0.193	0.196	0.275	0.320



Extended Data Figure 1. P value distributions of MK and PAML tests (the branch-site model). In panels (a) and (b), the P values of the MK test for all genes are shown in green, and that for genes selected by PAML test are shown in yellow. In panels (c) and (d), the green distributions represent the P values of the PAML test for all genes, and the yellow distributions the genes selected by MK test.



Extended Data Figure 2. Scatter plots of the proportion of average adaptive sites identified by MK and PAML tests (the site model). The orange dots represent genes chosen by the MK test as candidate genes under positive selection, but not chosen by the PAML test. Genes with more than eight polymorphisms are shown. Genes identified by the PAML but not the MK test, are depicted by dark blue dots. Red dots are genes called by both tests. The remaining genes are represented by light blue dots.