asymptoticMK: A web-based tool for the asymptotic McDonald–Kreitman test

Benjamin C. Haller and Philipp W. Messer

Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY 14853, USA

Corresponding author: Philipp W. Messer, 102J Weill Hall, Cornell University, Ithaca, NY 14853, phone: 607-255-3984, email: messer@cornell.edu

Keywords: molecular evolution, positive selection, web service

ABSTRACT

The McDonald–Kreitman (MK) test is a widely used method for quantifying the role of positive selection in molecular evolution. One key shortcoming of this test lies in its sensitivity to the presence of slightly deleterious mutations, which can severely bias its estimates. An asymptotic version of the MK test was recently introduced that addresses this problem by evaluating polymorphism levels for different mutation frequencies separately, and then extrapolating a function fitted to that data. Here we present asymptoticMK, a web-based implementation of this asymptotic McDonald–Kreitman test. Our web service provides a simple R-based interface into which the user can upload the required data (polymorphism and divergence data for the genomic test region and a neutrally evolving reference region). The web service then analyzes the data and provides plots of the test results. This service is free to use, open-source, and available at http://benhaller.com/messerlab/asymptoticMK.html.

INTRODUCTION

The extent to which molecular evolution is driven by positive selection, rather than neutral evolutionary processes such as random genetic drift, is one of the central questions of modern evolutionary biology. This question can be studied quantitatively by estimating the parameter α, which specifies the fraction of nucleotide substitutions in a given genomic region that were driven to fixation by positive selection (Eyre-Walker 2006). Values of α close to one indicate that most substitutions in the region were indeed the result of positive selection, whereas values close to zero indicate neutral evolution.

One of the most widely used approaches for inferring α from polymorphism and divergence data is the McDonald–Kreitman (MK) test (McDonald and Kreitman 1991; Eyre-Walker 2006), which compares levels of divergence between a genomic test region and a neutrally evolving reference region with the levels of polymorphism in the two regions. Early applications of the MK test typically focused on nonsynonymous sites in protein-coding regions as the test region, while synonymous sites were used as the neutral reference. However, the approach can also be applied to arbitrary genomic compartments or classes of mutations (Andolfatto 2005).

The original MK test makes several critical assumptions about the nature of the evolutionary process. First, it assumes that the positively selected mutations that ultimately contribute to divergence in the test region go to fixation quickly, such that they do not contribute noticeably to polymorphism levels. Second, it assumes that deleterious mutations in the test region are sufficiently deleterious to be lost quickly, such that they contribute to neither polymorphism nor divergence. Finally, neutral mutations in the test region are assumed to be subject to drift similar to the mutations in the neutral reference region and can therefore contribute to both polymorphism and divergence. Under these assumptions, it holds that

$$(1) \qquad \alpha = 1 - \frac{d_0}{d}\frac{p}{p_0},$$

where d and d_0 are the levels of divergence in the test region and neutral reference region, respectively, while p and p_0 specify the respective levels of polymorphism in the two regions (Eyre-Walker 2006).

With the growing availability of genome-level polymorphism and divergence datasets, the MK test has become a popular method for inferring positive selection in various organisms (Fay 2011). Several software tools and web services with implementations of the test have also been developed (Egea *et al.* 2008; Librado and Rozas 2009; Eyre-Walker and Keightley 2009; Stoletzki and Eyre-Walker 2011; Vos *et al.* 2013). The estimates of α obtained in these studies range from as high as ~0.5 for nonsynonymous substitutions in *Drosophila* (Sella *et al.* 2009), to close to zero in organisms such as yeast (Elyashiv *et al.* 2010) or many plants (Gossmann *et al.* 2010). Indeed, estimates of α obtained from Equation (1) are often negative, indicating that at least some of the assumptions of the test were likely not met.

One major problem with the original MK test lies in its assumption that deleterious mutations do not contribute to polymorphism in the test region. This stands in contrast to the frequent observation of weakly deleterious mutations in many organisms, and the fact that such mutations can substantially affect the site frequency spectrum (SFS) of polymorphisms in functional genomic regions (Bustamante *et al.* 2005; Eyre-Walker *et al.* 2006). In the presence of weakly deleterious mutations, p will overestimate the rate at which polymorphisms go to fixation in the test region, which will bias estimates of α downwards (providing one possible explanation for the frequent observation of negative α values).

As one strategy to address this problem, it has been proposed to only consider polymorphisms for which the derived allele is above a certain threshold frequency when estimating p and p_0 (Charlesworth and Eyre-Walker 2008). This is because the fraction of weakly deleterious mutations among all polymorphisms should be lower for higher derived-allele frequencies. Ideally, one would wish to set this cutoff high, to minimize the bias due to weakly deleterious mutations; however, the higher this cutoff, the fewer polymorphisms will actually remain in the dataset, thus increasing statistical noise. To circumvent this problematic tradeoff, more sophisticated extensions of the original MK test first attempt to infer the actual distribution of fitness effects among new mutations in the test region from the SFS, and then correct fixation probabilities accordingly (Boyko *et al.* 2008; Eyre-Walker and Keightley 2009). Yet these approaches can still suffer from unknown effects of demography or linked selection that are also expected to affect the shape of the SFS. The most sophisticated extensions of the test therefore additionally incorporate basic demographic models to improve their estimates (Keightley and Eyre-Walker 2007; Boyko *et al.* 2008; Eyre-Walker and Keightley 2009), which requires additional (and often uncertain) assumptions about the demographic history of the population of interest.

In contrast to such model-based approaches, a considerably simpler, heuristic approach was recently proposed by Messer and Petrov (2013). This approach generalizes the frequency-cutoff approach described above, without the need to discard polymorphism data. Instead of setting a specific frequency cutoff, it separately estimates α for each of a set of discrete mutational frequency classes:

$$(2) \qquad \alpha(x) = 1 - \frac{d_0}{d}\frac{p(x)}{p_{0(x)}}.$$

Here $p(x)$ and $p_0(x)$ specify the levels of polymorphism in the test and reference regions, respectively, considering only those polymorphisms for which the derived allele is present at frequency x in the population (estimated from a population sample, for example). In the presence of deleterious mutations, $\alpha(x)$ will underestimate the true value of α for small x, yet should converge to the correct value as x approaches one. The asymptotic estimate of α is then obtained by fitting a function $\alpha_{\text{fit}}(x)$ to the empirical $\alpha(x)$ values and extrapolating this function to $x = 1$:

$$(3) \qquad \alpha_{\text{asymptotic}} = \alpha_{\text{fit}}(x = 1).$$

One key advantage of this approach is that because $\alpha(x)$ does not depend on the individual functions $p(x)$ and $p_0(x)$ but only on their ratio, any biases due to demography or linked selection that affect the SFS in the test and reference regions in the same way will effectively cancel out (Messer and Petrov 2013). Another advantage over model-based approaches is that the asymptotic McDonald–Kreitman approach is much more computationally efficient, as it requires only fitting a simple curve to the data.

In this paper, we present asymptoticMK, a web-based tool for executing the asymptotic McDonald–Kreitman test quickly and easily in any web browser. After the necessary values are entered, asymptoticMK generates analyses and plots that are directly usable in publications. It is based internally on R, but no knowledge of R is needed to use it, nor does the user of asymptoticMK need to have R installed on their computer. For those who do wish to run the test themselves in R, the necessary code is freely available online. The asymptoticMK service can also be run in an automated fashion at the command line, for bulk analysis in script-based workflows.

MATERIALS AND METHODS

Implementation

The asymptoticMK web service is implemented in R (R Development Core Team 2016). It uses the package `FastRWeb` (Urbanek 2008) to parse HTTP requests and generate responses, and uses the package `Rserve` (Urbanek 2003) as the lower-level interface that communicates with the web server through the standard CGI mechanism.

Usage

The web service is free to use, without license restrictions of any kind, and is available at http://benhaller.com/messerlab/asymptoticMK.html. That URL displays an entry page (Figure 1) with an input form in which the user may enter the necessary data for the test: d (the substitution rate in the test region), d_0 (the substitution rate in the neutral reference region), and an uploaded file containing tab-delimited rows of data with values for x (the derived allele frequency), $p(x)$ (the polymorphism level in the test region at that frequency), and $p_0(x)$ (the polymorphism level in the neutral reference region at that frequency). A sample polymorphism file is provided on the website. In practice, it is often advisable to combine polymorphism levels into a smaller number of frequency bins, where x then specifies the central frequency of the bin. This is particularly relevant when the data includes frequencies at which no polymorphisms are actually present in the neutral region, in which case $\alpha(x)$ would be undefined for those particular frequencies according to Equation (2). The input form also allows entry of minimum and maximum values defining a cutoff interval for x, such that the test is run using only the polymorphisms whose frequencies fall within that cutoff interval; this is usually desirable as a means of excluding the lowest- and highest-frequency polymorphisms, where SNP quality issues and polarization errors are generally most pronounced. This frequency cutoff is set to [0.1, 0.9] by default.

Upon submission of the web form, asymptoticMK conducts its analysis and then opens a results page in a new browser tab, presenting a summary of the input data and the results from the analysis. The first plot on this results page shows binned polymorphism counts, $p_0(x)$ and

$p(x)$, for the submitted data; the second plot shows that same data normalized (i.e., the normalized SFS in the test and reference regions). A third plot shows the calculated empirical $\alpha(x)$ as a function of x, estimated from the input data according to Equation (2). The fourth plot shows the same $\alpha(x)$ data, with the best-fitting model and the asymptotic estimate of α superimposed upon the data.

Below these plots, the results of the analysis are presented in two tables. The first table provides the coefficients a, b, and (for exponential fits) c of the model yielding the best fit to the data. The second table provides the estimated $\alpha_{\text{asymptotic}}$ according to Equation (3), and the upper and lower limits of the 95% confidence interval around that estimate, as well as the estimated α from the original non-asymptotic McDonald–Kreitman test (α_{original}) for comparison (also estimated from all polymorphisms falling within the frequency cutoff interval specified on the input page).

For purposes of automation, asymptoticMK can also be run at the command line using the Linux/Unix `curl` command. For example, the command

```
curl -F"d=593" -F"d0=930" -F"xlow=0.1" -F"xhigh=0.9" -
F"datafile=@polymorphisms.txt" -F"reply=table" -o "MK_table.txt"
http://benhaller.com/cgi-bin/R/asymptoticMK_run.html
```

would run asymptoticMK with the given values of d and d_0, the given x cutoff interval, and polymorphism data uploaded from the local file `polymorphisms.txt`, and would output a simple table of results to the file `MK_table.txt`. Further documentation on the use of this feature is provided on the asymptoticMK web page.

Fitting and analysis procedure

The asymptotic McDonald–Kreitman test first involves calculating values of $\alpha(x)$ by applying Equation (2) to each frequency bin provided, as described by Messer and Petrov (2013). The next step involves fitting a function $\alpha_{\text{fit}}(x)$ to these empirical $\alpha(x)$ values. For greater robustness, asymptoticMK fits two functions to the data. The first function is exponential, of the form $\alpha_{\text{fit}}(x) = a + b \exp(-cx)$ and is fitted using the `nls2()` function, from the R package `nls2` (Grothendieck 2013). This fit is done in two steps. First, a brute-force scan for the closest fit is conducted across the likely portion of the three-dimensional parameter space defined by a, b, and c, by exhaustive search. This supplies reasonably good starting values for the second step, which refines those starting values using standard nonlinear least-squares regression. This two-step procedure generally works well, but can occasionally fail to converge if the data is not, in fact, exponential in form.

To address this possibility of nonconvergence of the exponential fit, asymptoticMK also fits a linear function of the form $\alpha_{\text{fit}}(x) = a + bx$, with the `lm()` function that is part of the `stats` package included in R. This fit always converges, and thus provides a backstop that allows the test to complete even when given irregular or extremely noisy data; however, it is always recommended that the results of the analysis be inspected visually to confirm that they are in fact meaningful.

Once these two models have been fitted, asymptoticMK chooses which model will be used for the remainder of the analysis. If the exponential fit failed to converge, then the linear model is chosen; if both fits succeeded, then the better model is chosen using the Akaike information criterion (AIC). Occasionally, in pathological cases, the exponential fit will have the better AIC but will have extremely large coefficient standard error(s); in this case, the linear fit is chosen since predictions from the exponential model would be effectively worthless.

The chosen model is then used to provide an estimate of the value of $\alpha_{\text{asymptotic}}$ according to Equation (3), by evaluating the fitted function $\alpha_{\text{fit}}(x)$ at $x = 1$; this is the primary result of the test, and provides the test's estimate of the true value of α within the test region. A 95% confidence interval around this estimate is also calculated. For the exponential model, this is done using Monte Carlo simulation based upon the fitted model, using the `predictNLS()` function published online by Spiess (2013); for the linear model, it is done using the standard R function `predict()`.

Test datasets

To provide a test of asymptoticMK using empirical data, we used the same *Drosophila melanogaster* dataset that Messer and Petrov (2013) used in their Figure 3C. This data set consists of SNPs obtained from the genome sequences of 162 inbred fly lines generated by the *Drosophila* genetic reference panel (Mackay *et al.* 2012). Divergence data was obtained from genome alignments between *D. melanogaster* and *D. simulans*, extracted from the 12 *Drosophila* genomes data (Clark *et al.* 2007). The test data in the asymptoticMK analysis (d and p) are genome-wide nonsynonymous mutations, while synonymous sites were used as the neutral reference (d_0 and p_0). The polymorphism data is available online at http://benhaller.com/messerlab/sample_polymorphism_levels.txt, with associated values d = 59570 and d_0 = 159058. The default frequency cutoff interval of [0.1, 0.9] was used in the analysis of this dataset with asymptoticMK.

We also tested asymptoticMK on simulated data, using the forward genetic simulation framework SLiM 2 (Haller and Messer 2017). A population of 1000 diploid individuals was simulated to evolve for 200,000 generations. The simulated chromosome was 10^7 base pairs long. Nucleotide mutations occurred uniformly at a rate of 10^{-9} per base per generation, and recombination occurred uniformly at a rate of 10^{-7} per base per generation. Each new mutation was either of neutral type "m1" (relative proportion of 0.5 of all new mutations), of functional non-beneficial type "m2" (relative proportion of 0.5 of all new mutations), or of functional beneficial type "m3" (a relative proportion of 0.0005 of all new mutations). The neutral m1 mutations always had a selection coefficient of s = 0.0; the selection coefficients of m2 mutations were drawn from a gamma distribution with a mean of $s = -0.02$ and a shape parameter of 0.2; and m3 mutations always had a selection coefficient of s = 0.1. Fitness effects were assumed to be codominant. A burn-in of 10000 generations was run to equilibrate the model. Every 500 generations thenceforth, all polymorphisms were recorded in the population by dividing them according to their frequency into 50 equal-width frequency bins, and then adding them to an ongoing binned tabulation. The SLiM configuration script used for these simulations is provided online at http://benhaller.com/messerlab/asymptoticMK_SLiM.html.

At the end of the model run, we obtained binned values for $p(x)$ and $p_0(x)$, where p_0 was estimated from all mutations of type m1, while p was estimated from the combined mutations of types m2 and m3. Values for d and d_0 were obtained from the set of mutations fixed during the simulation; as with p_0 and p, d_0 was estimated from all mutations of type m1, while d was estimated from the combined mutations of types m2 and m3. These values, output by the model, were used in asymptoticMK with the default x interval of [0.1, 0.9]. The true value of α was also calculated by the SLiM model as the fraction $d_3 / (d_2 + d_3)$, where d_2 is the number of m2 mutations fixed and d_3 is the number of m3 mutations fixed. This value provides a metric for the accuracy of asymptoticMK – a benefit of using simulated data, where the true α can be calculated.

RESULTS AND DISCUSSION

Results from our test of asymptoticMK with the empirical *D. melanogaster* dataset are shown in Figures 2A and 2B. The fitted exponential function is: $\alpha_{\text{fit}}(x) = 0.585 - 0.622 \exp(-3.80x)$. The asymptotic McDonald–Kreitman α estimate provided by this model is 0.571. These results match those obtained by Messer and Petrov (2013) using the same dataset (their Figure 3C), as expected. The α estimate provided by the original McDonald–Kreitman test is 0.407, by comparison (shown in Figure 2B).

The asymptoticMK results from the analysis of the SLiM simulation dataset are shown in Figures 2C and 2D. In this case, asymptoticMK deemed the linear fit to be superior to the exponential fit (and indeed, the data within the cutoff interval looks very linear in shape). The fitted linear function in this case is: $\alpha_{\text{fit}}(x) = 0.143 + 0.234x$. The asymptotic McDonald–Kreitman α estimate provided by this model is 0.377. This may be compared to the true α value provided by the simulation, 0.331. The α value from the original McDonald–Kreitman test within the cutoff interval, on the other hand, is 0.228 (shown in Figure 2D). If the cutoff interval is widened to [0.05, 0.95], the encompassed data then has a more exponential shape, and asymptoticMK then prefers the exponential fit (not shown), providing an improved asymptotic McDonald–Kreitman α estimate of 0.317 (as compared to the same true α value of 0.331, and an original McDonald–Kreitman α estimate of 0.170 within this cutoff interval). The default cutoff interval thus obscured the exponential shape of the data and prevented a good fit; this underlines the importance of choosing a cutoff interval that fits the shape of the data, and of examining the quality of the fit critically rather than simply accepting the default fit. Nevertheless, even the linear fit provided by the default cutoff interval provided a much closer estimation of the true α value than did the original non-asymptotic test.

In this paper, we presented asymptoticMK, a new web-based tool for executing the asymptotic McDonald–Kreitman test. To demonstrate its functionality, we analyzed both empirical data and a simulation-generated dataset. Results from both of these datasets illustrate the greater power of the asymptotic McDonald–Kreitman test to estimate the true value of α, compared to the estimates provided by the original non-asymptotic test. The asymptoticMK service presented here allows the user to obtain these results quickly and easily through any web browser.

ACKNOWLEDGMENTS

The authors thank G. Grothendieck, J. Horner, and S. Urbanek for their contributions to the R packages used here. We also thank S. Urbanek for his unstinting help with the installation and use of `FastRWeb`, A.-N. Spiess for his `predictNLS()` function, and most of all the R Core Team for R itself. asymptoticMK is only possible because of the free software developed by these and other contributors. This work was supported by funds from the College of Agriculture and Life Sciences at Cornell University to PWM.

LITERATURE CITED

Andolfatto, P., 2005 Adaptive evolution of non-coding DNA in *Drosophila*. Nature 437: 1149–1152.

Boyko, A. R., S. H. Williamson, A. R. Indap, J. D. Degenhardt, R. D. Hernandez *et al.*, 2008 Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS Genet. 4: e1000083.

Bustamante, C. D., A. Fledel-Alon, S. Williamson, R. Nielsen, M. T. Hubisz *et al.*, 2005 Natural selection on protein-coding genes in the human genome. Nature 437: 1153–1157.

Charlesworth, J., and A. Eyre-Walker, 2008 The McDonald-Kreitman test and slightly deleterious mutations. Mol. Biol. Evol. 25: 1007–15.

Clark, A. G., M. B. Eisen, D. R. Smith, C. M. Bergman, B. Oliver *et al.*, 2007 Evolution of genes and genomes on the *Drosophila* phylogeny. Nature 450: 203–218.

Egea, R., S. Casillas, and A. Barbadilla, 2008 Standard and generalized McDonald-Kreitman test: a website to detect selection by comparing different classes of DNA sites. Nucleic Acids Res. 36: W157-62.

Elyashiv, E., K. Bullaughey, S. Sattath, Y. Rinott, M. Przeworski *et al.*, 2010 Shifts in the intensity of purifying selection: an analysis of genome-wide polymorphism data from two closely related yeast species. Genome Res. 20: 1558–73.

Eyre-Walker, A., 2006 The genomic rate of adaptive evolution. Trends Ecol. Evol. 21: 569–575.

Eyre-Walker, A., and P. D. Keightley, 2009 Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. Mol. Biol. Evol. 26: 2097–108.

Eyre-Walker, A., M. Woolfit, and T. Phelps, 2006 The distribution of fitness effects of new deleterious amino acid mutations in humans. Genetics 173: 891–900.

Fay, J. C., 2011 Weighing the evidence for adaptation at the molecular level. Trends Genet. 27: 343–9.

Grothendieck, G., 2013 nls2: Non-linear regression with brute force. Available at: https://CRAN.R-project.org/package=nls2. Accessed: December 14, 2016.

Gossmann, T. I., B.-H. Song, A. J. Windsor, T. Mitchell-Olds, C. J. Dixon *et al.*, 2010 Genome wide analyses reveal little evidence for adaptive evolution in many plant species. Mol. Biol. Evol. 27: 1822–1832.

Haller, B. C., and P. W. Messer, 2017 SLiM 2: Flexible, interactive forward genetic simulations. Mol. Biol. Evol. 34: 230–240.

Keightley, P. D., and A. Eyre-Walker, 2007 Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. Genetics 177: 2251–2261.

Librado, P., and J. Rozas, 2009 DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451–1452.

Mackay, T. F., S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles *et al.*, 2012 The *Drosophila melanogaster* genetic reference panel. Nature 482: 173–178.

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature 351: 652–4.

Messer, P. W., and D. A. Petrov, 2013 Frequent adaptation and the McDonald-Kreitman test. Proc. Natl. Acad. Sci. U. S. A. 110: 8615–20.

R Development Core Team, 2016 *R: A language and for statistical computing*. Vienna, Austria.

Sella, G., D. A. Petrov, M. Przeworski, and P. Andolfatto, 2009 Pervasive natural selection in the *Drosophila* genome? PLoS Genet. 5: e1000495.

Spiess, A.-N. predictNLS (Part 1, Monte Carlo simulation): confidence intervals for “nls” models. R-bloggers. Available at: https://www.r-bloggers.com/predictnls-part-1-monte-carlo-simulation-confidence-intervals-for-nls-models/. Accessed: December 14, 2016.

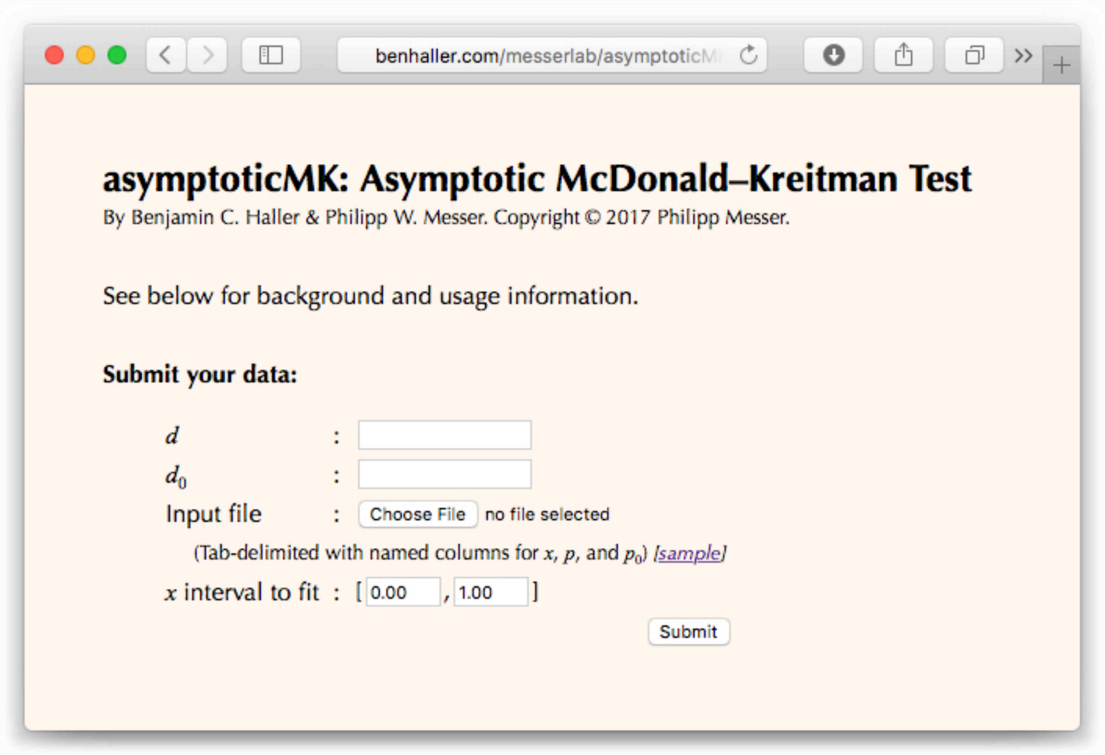
Stoletzki, N., and A. Eyre-Walker, 2011 Estimation of the neutrality index. Mol. Biol. Evol. 28: 63–70.

Urbanek, S., 2008 FastRWeb: Fast interactive web framework for data mining using R, in *IASC 2008 World Congress*. Available at: https://rforge.net/FastRWeb/. Accessed: December 14, 2016.

Urbanek, S., 2003 Rserve - A fast way to provide R functionality to applications. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*. Available at: https://rforge.net/Rserve/. Accessed: December 14, 2016.

Vos, M., T. A. H. te Beek, M. A. van Driel, M. A. Huynen, A. Eyre-Walker *et al.*, 2013 ODoSE: a webserver for genome-wide calculation of adaptive divergence in prokaryotes. PLoS One 8: e62447.

FIGURES



asymptoticMK: Asymptotic McDonald–Kreitman Test

By Benjamin C. Haller & Philipp W. Messer. Copyright © 2017 Philipp Messer.

See below for background and usage information.

Submit your data:

d	:	
d_0	:	
Input file	:	Choose File no file selected
		(Tab-delimited with named columns for x, p, and p_0) *[sample]*
x interval to fit	:	[0.00 , 1.00]

Figure 1. A screenshot of the web page for asymptoticMK. After entering values for d and d_0, choosing an input file with binned values for x, p, and p_0, and choosing the x interval to fit, the user can click the Submit button and asymptoticMK will provide its results in a new browser window or tab.

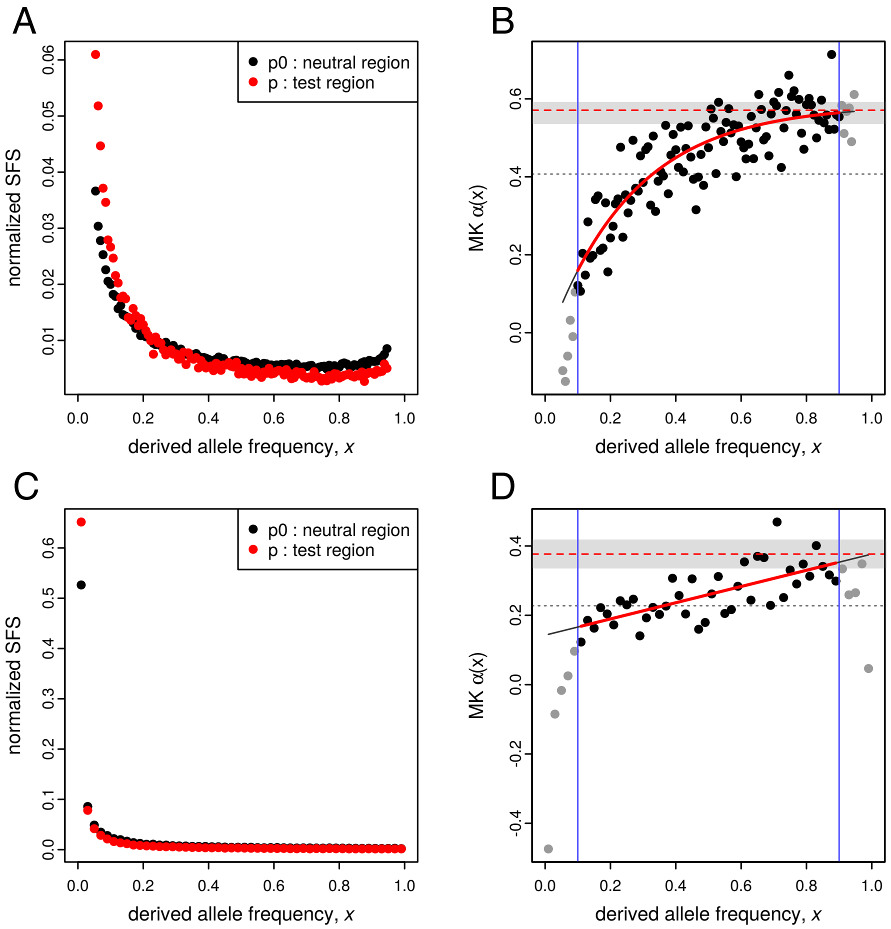


Figure 2. Results from asymptoticMK for two test datasets. A and B show results from the *Drosophila* dataset of Messer and Petrov (2013); C and D show results from a SLiM simulation run (see Materials & Methods). A and C show the normalized site frequency spectra (SFS) for their respective datasets. B and D show the results of the asymptotic MK test. In B and D, the two vertical blue lines show the limits of the frequency cutoff interval used for fitting. Points indicate binned values of $\alpha(x)$, estimated according to Equation 2; points are gray if they are outside the cutoff interval (and thus not used in fitting). The solid red curves show the fitted functions (exponential in B and linear in D). The dashed red lines show the estimates of $\alpha_{\text{asymptotic}}$, obtained from the fitted function according to Equation 3; this is the main result of the asymptotic MK test. The gray bands indicate the 95% confidence intervals around the $\alpha_{\text{asymptotic}}$ estimates. The dotted gray lines show α_{original}, the estimates of α from the original (non-asymptotic) MK test, for comparison (also calculated using only the data within the cutoff interval).