HHS Public Access

Author manuscript

Nat Ecol Evol. Author manuscript; available in PMC 2019 January 02.

Published in final edited form as:

Nat Ecol Evol. 2018 August ; 2(8): 1280–1288. doi:10.1038/s41559-018-0584-5.

Multinucleotide mutations cause false inferences of lineage-specific positive selection

Aarti Venkat[1], **Matthew W. Hahn**[2], and **Joseph W. Thornton**[*,1,3]

[(1)]Department of Human Genetics, University of Chicago, Chicago IL 60637, USA

[(2)]Department of Biology and Department of Computer Science, Indiana University, Bloomington IN 47405, USA

[(3)]Department of Ecology & Evolution, University of Chicago, Chicago IL 60637, USA

Abstract

Phylogenetic tests of adaptive evolution, such as the widely used branch-site test, assume that nucleotide substitutions occur singly and independently. But recent research has shown that errors at adjacent sites often occur during DNA replication, and the resulting multinucleotide mutations (MNMs) are overwhelmingly likely to be nonsynonymous. We evaluated whether the branch-site test (BST) might misinterpret sequence patterns produced by MNMs as false support for positive selection. We analyzed two genome-scale datasets– one from mammals and one from flies – and found that codons with multiple differences account for virtually all the support for lineage-specific positive selection in the BST. Simulations under conditions derived from these alignments but without positive selection show that realistic rates of MNMs cause a strong and systematic bias towards false inferences of selection. This bias is sufficient under empirically derived conditions to produce false positive inferences as often as the branch-site test infers positive selection from the empirical data. Although some genes with BST-positive results may have evolved adaptively, the test cannot distinguish sequence patterns produced by authentic positive selection from those caused by neutral fixation of MNMs. Many published inferences of adaptive evolution using this technique may therefore be artifacts of model violation caused by unincorporated neutral mutational processes. We introduce a model that incorporates MNMs and may help to ameliorate this bias.

Keywords

adaptation; adaptive evolution; branch-site test; codon models; transversions

[*]Correspondence:Joseph Thornton, joet1@uchicago.edu.

AUTHOR CONTRIBUTIONS

Analyses were designed by all authors, performed by AV, and interpreted by all authors. The manuscript was written by AV and JWT with contributions from MWH.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests

INTRODUCTION

Identifying genes that evolved under the influence of positive natural selection on phylogenetic time scales is a central goal in studies of molecular evolution. Of many methods developed for this purpose,[1–10] the most widely used is the branch-site test (BST).[5,6] This technique has been the basis for published claims of lineage-specific adaptive evolution in many thousands of genes.[11–15]

The BST uses a likelihood ratio test to compare two probabilistic models of sequence evolution, given an alignment of coding sequences. The null model constrains all codons to evolve with rates of nonsynonymous substitution (d_N) less than or equal to the rate of synonymous substitution (d_S), as expected under purifying selection alone. In the positive selection model, some sites are allowed to have $d_N > d_S$ on one or more branches of interest. If the latter model increases the likelihood more than expected by chance, the null model is rejected and adaptive evolution is inferred. The BST is conservative in the absence of model violation, with a low rate of false positive inferences when sequences are generated according to the null model.[6,16] Although likelihood ratio tests can be biased if the underlying probabilistic model is incorrect,[17] the BST has been found to be reasonably robust to several forms of model violation.[6,18–24]

A recently discovered genetic phenomenon – the propensity of DNA polymerases to produce mutations at neighboring sites – has not been evaluated for its effect on the BST. All current models for identifying positive selection assume that mutations are fixed singly and independently at individual nucleotide sites: codons with multiple differences (CMDs) can be interconverted only by serial single-nucleotide substitutions, the probability of which is the product of the probabilities of each independent event. But molecular studies of replication show that some polymerases are prone to making adjacent mutations.[25–33] In studies of human trios and laboratory organisms, *de novo* mutations often occur in tandem or at nearby sites more frequently than expected if each occurred independently.[25,32–36] The precise frequency at which multinucleotide mutations (MNMs) occur is difficult to estimate, but a recent study concluded that about 0.4% of mutations, polymorphisms, and substitutions in humans are at directly adjacent sites (counting each tandem pair as one event).[34] In *Drosophila melanogaster*, analysis of rare polymorphisms and mutation-accumulation experiments estimated that 1.3% of all mutations are at adjacent sites.[37] Tandem MNMs therefore appear to account for on the order of 1% of mutations.

We hypothesized that MNMs might lead to false signatures of positive selection in the BST and related tests. Because of the structure of the genetic code, virtually all MNMs in coding sequences are nonsynonymous, and most would require multiple nonsynonymous changes if they were to occur by single nucleotide steps (Supplementary Table 1). Further, MNMs tend to be enriched in transversions,[35,38,39] and transversions are more likely than transitions to be nonsynonymous. MNMs are therefore likely to produce CMDs containing an apparent excess of nonsynonymous substitutions, even in the absence of positive selection. When these data are assessed assuming that all substitutions are independent, a model that allows d_N to exceed d_S at some sites may have significantly higher likelihood. CMDs can also be fixed by positive selection,[16,40–42] but current methods may fail to distinguish selected

CMDs from those produced by neutral fixation of MNMs. Simulations suggest that MNMs may increase the rate of positive inference in BST and related selection tests,[43,44] but there has been no comprehensive analysis of the effect of MNMs under realistic, genome-scale conditions.

RESULTS

We analyzed two previously published genome-scale datasets, which represent classic examples of the application of the BST.[12,14,45] The mammalian dataset consists of coding sequences of 16,541 genes from six species; we retained for analysis only the 6,868 genes with complete species coverage. The fly dataset consists of 8,564 genes from six *Drosophila* species, all of which had complete coverage (Supplementary Fig. 1).

We first used the BST to identify genes putatively under positive selection ($P<0.05$) on the human lineage in the mammalian dataset and on each of the six terminal lineages in flies. Eighty-two genes in humans and 3,938 in flies yielded significant tests ($P<0.05$, Supplementary Table 2). Filtering for data quality and correcting for multiple testing (FDR<0.2) yielded 443 fly genes for further analysis. Thirty human genes passed the quality filter, but none survived multiple testing, consistent with previous analyses;[14] nevertheless, we included the 30 initially significant, high-quality genes because this lineage is the object of intense interest and because its short length contrasts with the fly branches, allowing us to examine the performance of the BST under different conditions. These two sets constitute the "BST-significant" genes in flies and humans.

CMDs provide virtually all support for positive selection

We found that CMDs are the primary drivers of BST-positive results. CMDs are dramatically enriched in BST-significant genes compared to non-BST-significant genes (**Figs. 1a,** Supplementary Fig. 2). When CMD-containing sites are excluded from the alignments, the vast majority of genes that were BST-significant lose their signature of selection (**Fig. 1b)**. In virtually all BST-significant genes, >95% of the statistical support for positive selection, defined as the fraction of the total log-likelihood difference between the positive-selection and null models, comes from CMDs; in about 70% of genes, CMDs provide all the support (**Fig. 1c)**. CMDs account for 60% and 90% of sites inferred *a posteriori* to have been positively selected (PP > 0.9) in humans and flies, respectively, although they represent <1% of all codons (Fig. 1d).

Incorporating MNMs eliminates the signature of positive selection in many genes

CMDs could be enriched in BST-positive genes because of an MNM-induced bias or because they were fixed by positive selection. To distinguish between these possibilities, we implemented a version of the BST that is identical to the classic version, but its model allows double-nucleotide changes using an additional parameter δ, which scales the rate of each double-nucleotide substitution relative to single-nucleotide substitutions. We evaluated our implementation of this BS+MNM model using simulations under realistic conditions and found that parameters are estimated with reasonable accuracy (Supplementary Fig. 3). When fit to all empirical mammalian and fly alignments, the BS+MNM null model provides

a statistically significant likelihood increase for 22% of human genes and 57% of fly genes compared to the classic null model without δ; this test has a low rate of false positive inferences **(**Supplementary Table 3**)**. In both datasets, the average estimated value of δ is about twice as high in the subset of BST-significant genes as in BST-nonsignificant genes (**Fig. 2a)**.

We next evaluated the empirical alignments for positive selection using the BS+MNM test, which incorporates MNMs into the null and positive selection model. We found that 94% of the tests on the human lineage that were significant using the classic BST lost significance (**Figs. 2b,** Supplementary Table 4). In flies, 38% of the tests lost significance, and a substantial fraction of the remaining genes were enriched in triple substitutions, a process not accounted for in our model (**Figs. 2b,** Supplementary Table 4).

MNMs cause false positive inferences on a genome-wide scale

Our finding that incorporating MNMs dramatically eliminates the signature of positive selection from many genes could have several causes, including: 1) the BS+MNM model may have reduced power to identify authentic positive selection compared to the BST, 2) it may ameliorate a false-positive bias in the BST caused by MNMs, or 3) the additional parameter δ may allow the BS+MNM model to fortuitously fit other forms of sequence complexity, potentially reducing a bias in the BST caused by other model violations.

To evaluate these possibilities, we first analyzed the test's power. We simulated sequences under the BST positive selection model, using genome-wide average values for all parameters but varying the strength of positive selection (ω_2) and the proportion of sites under positive selection. We found that the BS+MNM test reliably detects strong positive selection ($\omega_2 > 20$) when it affects ~10% of sites in a typical gene, or moderate positive selection ($10 < \omega_2 < 20$) on a larger fraction of sites (Supplementary Fig. 4a-4b). Its power is similar to that of the classic BST, with a slight reduction under only a few conditions on the fly lineage (Supplementary Fig. 4c). Thus, although some genes may have lost their signature of selection because of reduced power in the BS+MNM test, this is unlikely to be the primary cause of the dramatic reduction in the number of positive results when the test is used.

Next, we used simulations without positive selection to directly evaluate whether realistic rates of multinucleotide mutation increase the BST's propensity to deliver false positive inferences. For every gene in the mammalian and fly datasets, we simulated sequence evolution under the null BS+MNM model using parameters derived from the alignments, including δ, gene length, and selection parameters. The fraction of substitutions that occurred at tandem sites on the branches of interest in the simulations (1.6% in humans and 3.2% in flies) was comparable to or slightly higher than the fraction of tandem substitutions phylogenetically inferred on these branches in the empirical alignments (1.3% in humans and 1.6% in flies), presumably because the BS+MNM model captures some but not all aspects of real sequence evolution (Supplemental Table 5). We then analyzed the simulated alignments using the classic BST. In these experiments, every BST-positive result is false.

The number of genes yielding false positive results was greater than the number of genes that the BST inferred to be under positive selection using the empirical data (**Fig. 3a).** In flies, almost 9 percent of genes were falsely inferred to be under positive selection ($P<0.05$), despite the conservative approach the method uses to calculate P-values,[6,16] compared to just 1 percent under control simulations without MNMs ($\delta = 0$). Over 1,700 of these false positive tests (an average of almost 300 genes per lineage) survived FDR adjustment, compared to a total of just 4 positive tests in the control simulations (Supplementary Table 2**)**. In humans, the fraction of false positive inferences was lower, consistent with the test's reduced power in this dataset, but still about three times greater than in the control simulations. These false inferences were caused specifically by unincorporated MNMs -- not some other form of model violation -- because all other parameters were identical between the model and analysis models.

These findings indicate that MNMs under realistic evolutionary conditions produce a strong and widespread bias in the BST toward false inferences of positive selection. This bias is strong enough to cause the BST to make false inferences of positive selection at about the same rate as it infers selection in the real genomes of humans and flies.

Systematic bias caused by stochastic fixation of neutral MNMs

Only a few percent of mutations are MNMs, and most genes are only several hundred codons long, so on phylogenetic branches of short to moderate length many genes will evolve zero fixed MNMs. If neutral fixation of MNMs is a major cause of bias in the BST, then a gene's propensity to produce a BST-significant result should depend on factors that increase the probability it will contain one or more CMDs by chance, including its length and the gene-specific rate of multinucleotide mutation in that gene.

We first tested for an association between gene length and BST-positive results. As predicted, BST-significant genes were on average 100 and 16 codons longer than non-significant genes in the human and fly datasets, respectively (Fig. 3b). A similar pattern was evident in the null simulations under empirically derived conditions (Supplementary Fig. 5); this finding cannot be attributed to an increase in power to detect true positive selection in longer genes, because no positive selection was present. To directly test the causal relationship between sequence length and false-positive bias in the BST, we simulated multiple replicate alignments under the BS+MNM null model at increasing sequence lengths (L) using evolutionary parameters derived from each BST-significant gene (Supplementary Fig. 6). At L= 5,000 codons, 96% of human genes produced an unacceptable false positive rate (FPR >0.05), with a median FPR of 0.39; doubling the sequence length exacerbated the bias, with every gene now yielding an unacceptable FPR (median 0.56, Fig. 3c). The same pattern was evident in flies, with even higher false positive rates (median FPR=0.74 and 0.90 at L=5,000 and 10,000, respectively). Control simulations under identical conditions but with δ=0 led to very low FPRs, even with very long sequences. Although these experiments involve lengths greater than that of most real genes, they establish that the probability that a gene will yield a false-positive BST result is directly related to the target size it provides for chance fixation of MNMs.

We next evaluated whether the rate of multinucleotide mutation affects a gene's propensity to yield a positive result in the BST. As predicted, BST-significant genes in the empirical datasets had higher estimated δ than nonsignificant genes **(Fig. 2a**). In the null simulations using the BS+MNM model under conditions derived from each empirical alignment, genes producing false positive BST results also tended to have higher δ (Fig 3d). To directly test the causal relationship between the frequency of neutral MNMs and false-positive bias in the BST, we simulated multiple replicate alignments under the BS+MNM null model with empirically derived parameters, but with variable δ. As δ increased, the rate of false positive inferences increased monotonically (Fig 3e), and so too did the inferred value of the parameter ω_2, which represents the inferred intensity of positive selection in the model (Fig. 3f).

We also examined whether BUSTED,[2] a recent method to identify episodic site-specific selection events across an entire tree, was also biased by MNMs. When sequences of length L=5000 were simulated under empirical conditions, BUSTED yielded an unacceptably high false positive rate for every gene in humans and most genes in flies (median FPR 0.29 and 0.50, respectively, Supplementary Fig. 7**)**. In control simulations with δ=0, virtually no genes had a high rate of false positive inferences (FPR <0.03).

Taken together, these data indicate that MNMs under typical evolutionary conditions cause a strong and systemic bias in the BST and related tests. MNMs are rare, however, so whether a specific gene manifests the bias depends on factors that determines the probability of stochastic fixation of MNMs within it. Consistent with this conclusion, fewer genes are BST-positive on the very short human branch – on which substitutions are infrequent and CMDs even more rare– than on the fly phylogeny's longer branches. Many genes with BST-significant results may simply be those that happened to fix multinucleotide substitutions by chance.

Transversion-enrichment in CMDs exacerbates bias in the branch-site test

MNMs tend to produce more transversions than single-site mutational processes, so if CMDs are produced by MNMs, they should be transversion-rich.[35,38,39] As predicted, we found that the transversion:transition ratio is elevated in CMDs relative to that in non-CMDs by factors of three and two in mammals and flies, respectively (Fig. 4a). In the subset of BST-significant genes, CMDs have an even more elevated transversion:transition ratio, as expected if transversion-rich MNMs bias the test (Fig. 4a). These data are consistent with the hypothesis that a transversion-prone MNM process produced many of the CMDs in BST-significant genes, but it is also possible that positive selection could have enriched for transversions.

To directly test whether transversion-enrichment in MNMs exacerbates the BST's bias, we developed an elaboration of the BS+MNM model in which an additional parameter allows MNMs to have a different transversion:transition rate ratio ($\boldsymbol{\kappa_2}$) than single-site substitutions do ($\boldsymbol{\kappa_1}$). We simulated sequence data using this model under empirically derived conditions without positive selection, varying the value of $\boldsymbol{\kappa_2}$, and then analyzed these data using the classic BST. We found that increasing $\boldsymbol{\kappa_2}$ caused a rapid and monotonic increase in the false positive rate. The effect is strong: for example, when κ_2/κ_1 is increased from its baseline

value of 1 to 2, the FPR approximately doubles (Fig. 4b). Thus, realistic rates of MNM generation and transversion-enrichment together cause an even stronger bias in the BST than MNMs alone.

CMDs that invoke multiple nonsynonymous steps drive the signature of positive selection

Finally, we sought further insight into the reasons why CMDs yield a false signature of positive selection in the BST and related tests. We hypothesized that CMDs implying multiple nonsynonymous substitutions under standard models would provide the strongest support for the positive selection model. As predicted, we found that CMDs that imply more than one nonsynonymous step are dramatically enriched in BST-significant genes (Fig 5a). Further, CMDs implying more nonsynonymous single steps provided greater statistical support for the positive selection model (Fig. 5b). CMDs implying one nonsynonymous and one synonymous step typically provide weak to moderate support, but a single CMDs that implies two nonsynonymous steps is often sufficient to yield a statistically significant signature of positive selection for an entire gene (Fig. 5b).

DISCUSSION

This work establishes that the branch-site test suffers from a strong and systematic bias toward false positive inferences. The bias is caused by a mismatch between the method's underlying codon model of evolution – which assumes that a codon with multiple differences can be produced only by two or more independent substitution events – and the recently discovered phenomenon of multinucleotide mutation, which produces such codons in a single event. Under the BST's null model, the probability of two fixation events within a codon is extremely small, but it can increase dramatically when d_N/d_S exceeds one, as the positive selection model allows.

As a result, CMDs in real sequences are the primary drivers of positive results by the BST. Virtually all statistical support for positive selection in the genome-scale alignments we studied comes from CMD-containing sites. These CMDs could have been produced by either neutral fixation of MNMs or positive selection, but the BST provides no reliable basis to distinguish between these possibilities.

The BST's bias is strong and pervasive under realistic, genome-scale conditions. In both the human and fly datasets, the number of genes that the BST infers to be positively selected does not exceed the number expected to be produced by MNM-induced bias alone. Further, the empirically based null simulations on which this expectation is based did not include the elevated transversion rate that characterizes MNMs, which exacerbates the test's bias. Taken together, these results suggest that MNM-induced bias may explain many of the BST's inferences of positive selection in these datasets.

We do not contend that the BST is always wrong or that molecular adaptive evolution does not occur. Some of the CMDs in BST-significant genes may have evolved because of authentic positive selection fixing MNMs or serial single-site mutations. The test cannot distinguish sequence data produced by these two scenarios, however, so it provides no reliable evidence that a gene evolved adaptively. It also cannot reliably estimate the fraction

of genes in a large set that evolved under positive selection. There are numerous examples of strongly supported adaptive evolution – particularly involving host-parasite genetic conflicts – in which sequence signatures of positive selection are likely to be authentic [46–51], but the convincing evidence in these cases comes from sources other than the BST. The bias we discovered may help explain why some studies of have found that codons with a high posterior probability of positive selection in the BST have no effect on putatively adaptive functions, whereas those that do confer those functions have low or moderate PPs.[52–54]

Our results are likely to be generalizable. MNMs appear to be a property of all eukaryotic replication processes, and the MNM rates that we observed in mammals and flies are in the same range as those identified in in a variety of eukaryotic species.[25,34,37] We observed strong bias on lineages with divergence levels ranging from very low (on the human terminal branch) to moderate (the fly branches), so this problem does not appear to be unique to highly diverged sequences. The major factors determining whether a gene returns a false positive result in the BST test are those that affect the probability that one or more MNMs will be stochastically fixed – gene length, MNM rate, and overall substitution probability. We must therefore consider the possibility that some – and potentially many -- of the thousands of the genes previously reported to be under positive selection based the BST could simply be those that happened by chance to neutrally fix one or more multinucleotide mutations.

If the BST is so prone to error, what should researchers do? The BS+MNM test may be a promising means to accommodate MNMs, but there are many other forms of evolutionary complexity that are not incorporated into this model.[55–57] More work is therefore required before the BS+MNM test or related techniques[9] can be used with confidence. An alternative strategy—using functional experiments to explicitly test hypotheses about the genes and substitutions that drove molecular adaptation—can produce strongly supported inferences, but it is not clear how to implement such time-consuming bench and field work on a genome-wide scale.[50,58–60] Future research may develop and validate more robust models to detect positive selection, and these may help to identify candidate genes for which specific hypotheses of past molecular adaptation on specific lineages can be formulated and tested. The primary method used for this purpose until now is unreliable.

METHODS

Datasets, quality control, and inference of BST-significant genes.

We analyzed two previously published comprehensive datasets of protein-coding alignments on a genomic scale, one in six mammals, the other in six *Drosophila* species **(**Supplementary Table 2**)** [12,14,45]. We aimed to apply the branch-site test on every terminal lineage in the *Drosophila* dataset, and on the human lineage in the mammal dataset. We only retained gene alignments without gross misalignments, possessing complete coverage in all fly species, and minimally all primate species. We then applied the branch-site test as implemented in CODEML 4.7 to each alignment, assuming the phylogenetic relationships reported in the published studies (Supplementary Fig. 1) [12,14]. Branch lengths and model parameters were estimated for each alignment by maximum likelihood (ML), and the F3×4 model was used for codon frequencies. We tested each gene in mammals for selection on the terminal branch

leading to humans; in flies, each gene was tested separately for selection on each of the six terminal branches, and we express the fraction of positive inferences across genes as the proportion of all tests conducted [6]. As is standard practice, we calculated P-values using a likelihood ratio test with 1 df (χ_1^2) which makes the test conservative under the null hypothesis [6]. Genes were initially identified as having a putative BST signature of selection at $P<0.05$. We then applied a correction for multiple testing to a false discovery rate (FDR) <0.20 using the *q-value* package in R (available at http://github.com/jdstorey/qvalue).

To facilitate unambiguous analysis of CMDs, we removed genes containing CMDs falling in gaps. We also removed genes for which the ML ancestral reconstructions reported by CODEML at the base of the tested branch differed between the null and positive selection models, yielding a set of genes with CMDs that do not depend upon which model is chosen. In flies, 443 genes were retained after these filters and constitute the BST-significant set of genes from this dataset; these genes produce 458 positive tests in the BST, because a gene can yield a significant test on more than one branch. No genes on the human lineage were significant after FDR correction, so we retained as the BST-significant set from this dataset those genes that passed the ancestral reconstruction filter and had $P<0.05$ (Supplementary Table 2). The BST-nonsignificant set of genes comprises all genes that pass the alignment and ancestral reconstruction filter that are not in the BST-significant set (n=6757, humans; n=6883, flies). We also repeated our analysis of CMD enrichment (see below) using a gene set that had not been filtered for reconstruction consistency and found that our conclusions were unchanged (Supplementary Table 6).

We only considered genes where the ancestral codons (both CMD and non-CMD codons) have the same reconstruction under the BST null and BST alternate models. In doing so, we have also excluded CMDs in codons with gaps in the alignment. For example, in the human dataset, of the 82 genes that initially provided support for positive selection, 30 genes consist of unambiguously reconstructed codons under the null and alternate model (the BST-significant gene set). In 49 genes, CMDs fall in gaps. We did not consider the ancestral codon reconstructions at these sites, and excluded these from our analyses due to alignment ambiguities. The remaining 3 genes have CMDs that do not fall in gaps, for which the ancestral codons were reconstructed differently under the null and alternate models. If we re-consider these 3 ‘positively selected’ genes that were excluded, we find 3 additional CMDs, one in each of the genes. Including these genes made little to no difference to our CMD enrichment results.

Support for positive selection.

CMDs were identified in BST-significant and BST-nonsignificant genes as codons with 2 or 3 observed nucleotide differences between the ML states at the ancestral and extant nodes for the branch being tested; non-CMDs are codons with 0 or 1 differences on the branch tested. CMDs were not assessed on branches not tested.

To determine the role of CMDs in significant results from the BST, we excluded codon positions in BST-significant genes containing CMDs, reanalyzed the data using the BST, and calculated the fraction of tests that retained a significant result ($P<0.05$).

We quantified the proportion of statistical support for positive selection in BST-significant genes that comes from CMDs as follows. The site-specific support provided by one codon site in an alignment is the difference between the log-likelihoods of the positive selection model and the null model given the data at that site. Support for positive selection provided by all CMDs in a gene ($support_{CMD}$) is the support summed over all CMD sites in the alignment. The proportion of support provided by CMDs is $support_{CMD} / (support_{CMD} + support_{nonCMD})$. This proportion can be greater than 1 if support by non-CMDs is negative, as occurs if the likelihood of the null model at non-CMD sites is higher than that of the positive selection model, given the parameters of each model estimated by ML over all sites.

Sites were classified *a posteriori* as under positive selection if their Bayes Empirical Bayes posterior probability of being in class 2 ($\omega_2>1$) under the positive selection model in CODEML was >0.5 (moderate support) or >0.9 (strong support).

We categorized observed CMDs by the minimum number of nonsynonymous single-nucleotide steps implied under the Goldman-Yang model between the ancestral and derived states. For each CMD comprising two nucleotide differences, there are two paths by which they can be interconverted by two single nucleotide steps. We determined whether the steps on these paths would be nonsynonymous or synonymous using the standard genetic code and then calculated the mean number of nonsynonymous steps averaged over the two paths. Paths involving stop-codons were not included. We conducted a similar analysis for all possible CMDs in the universal genetic code table.

BS+MNM codon substitution model and test.

The codon substitution model of the classic BST is based on the Goldman-Yang (GY) model [5]. Sequence evolution is modeled as a Markov process, where the matrix element q_{ij}, the instantaneous rate of change from ancestral codon i to derived codon j, is defined for four types of changes: synonymous transitions and transversions, and nonsynonymous transitions and transversions (see q_{ij} equation 1). Three parameters are estimated from the data by maximum-likelihood: ω, the ratio of nonsynonymous substitution rate to the synonymous substitution rate (d_N/d_S); π_j, the equilibrium frequency of codon j; and κ, the transversion:transition rate ratio.

$$qij = \begin{cases} \kappa\pi_j & \text{synonymous transversion} \\ \pi_j & \text{synonymous transition} \\ \kappa\omega\pi_j & \text{non-synonymous transversion} \\ \omega\pi_j & \text{non-synonymous transition} \\ 0 & \text{two or more differences} \end{cases} \quad 1$$

Element q_{ij} is zero for substitutions involving more than one difference, so codons with multiple differences can only evolve through intermediate codons that are a single change away. A scaling factor applied to the matrix ensures that branch lengths are interpreted as the expected number of substitutions per codon.

$\kappa_1\pi_j$	synonymous transversion
π_j	synonymous transition
$\omega\kappa_1\pi_j$	non-synonymous transversion
$\omega\pi_j$	non-synonymous transition
$\omega\delta\kappa_2^2\pi_j$	non-synonymous, 2 transversions
$\omega\delta\pi_j$	non-synonymous, 2 transitions
$\omega\delta\kappa_2\pi_j$	non-synonymous, 1 transversion, 1 transition
$\delta\pi_j$	synonymous, 2 transitions
$\delta\kappa_2^2\pi_j$	synonymous, 2 transversions
$\delta\kappa_2\pi_j$	synonymous, 1 transversion, 1 transition
0	otherwise

2

We developed a modification of the GY model that incorporates MNMs using the parameter, δ, which represents the relative instantaneous rate of double substitutions to that of single substitutions (see q_{ij} equation 2). When δ = 0, the BS+MNM model reduces to the classic BST model that does not incorporate MNMs (q_{ij} equation 1). Triple substitutions have an instantaneous rate of zero.

The BS+MNM test of positive selection is identical to the BST, except it utilizes this MNM codon model. We implemented this test by modifying the branch-site test batch file (YangNielsenBranchSite2005.bf) in Hyphy 2.2.6 software by declaring δ a global variable, incorporating it into the codon table, and allowing it to be optimized by ML as it other model parameters are.

We validated the BS+MNM implementation by simulating 50 replicate alignments using the BS+MNM null model in Hyphy under genome-median parameters (see below). We then used the BS+MNM procedure to find the ML estimate of each parameter, including branch lengths, given each alignment and the topology of the phylogeny used to generate the sequences. We compared the distribution of estimates over replicates to the “true” values used to generate the sequences (Supplementary Fig. 3).

To test if there is statistical support in the data for the BS+MNM null model relative to the standard BST null model, we performed an LRT with 1 df, comparing the fit of the BS +MNM null model and the BST null model on our empirical genes. Briefly, for each of the 6868 human genes, we tested if the BS+MNM null model fit the data better than the BST null model at P<0.05 and also applied and adjustment for multiple testing (FDR<0.2). We performed similar LRTs for each of the six terminal lineages in flies. To determine whether this test might be prone to falsely infer support for the BS+MNM model, we simulated control sequences under the null BST model with parameters derived from the empirical sequences and performed the LRT as described above. Only 2 percent of genes in humans and 2.6 percent in flies yielded significant support for BS+MNM at P < 0.05. Zero human

genes and 0.006 percent of fly genes retained significance after multiple testing adjustment (FDR <0.2). (Supplementary Table 3).

Simulations and analysis of false-positive bias.

To characterize bias in the BST and other tests of selection, we conducted sequence simulations in the absence of positive selection under empirically derived conditions. We used the BS+MNM method we implemented in Hyphy to estimate by maximum likelihood (ML) the gene-specific branch lengths and parameters of the null BS+MNM model for every gene in the mammalian and fly datasets. We also calculated the genome-wide median of each parameter over all genes in each dataset (the “genome-average” parameter value). Probability density characterizations for parameters δ and gene length were performed using the *density* function in R.

We simulated sequence evolution under the BS+MNM null model using either gene-specific or genome-median parameters. First, we simulated a “pseudo-genome” without positive selection by simulating one replicate of each of the 6868 and 8564 mammalian and fly alignment, each at its empirical length, using the BS+MNM null model and the ML parameter estimates inferred for that gene from the empirical data. We then ran the BST on these sequences, testing for signatures of positive selection on the human lineage and each terminal fly lineage (Supplementary Table 2). Control simulations were conducted under identical conditions but with δ=0.

To test the effect of gene length on bias in the BST, we focused on genes in the BST-significant set. For each gene’s gene-specific parameters, we simulated 50 replicates alignments of length 5,000 or 10,000 codons. We analyzed these alignments using the BST, assigning the human branch as foreground for mammalian genes or, for flies, the same branch that produced a significant result when the empirical data were analyzed. The false positive rate (FPR) for any gene’s parameters is the fraction of replicates yielding a positive test ($P<0.05$). We also repeated these simulations and analyses using the genome-median value of δ. For control experiments without MNMs, we set δ =0 in the simulations.

To test the effect of the rate at which MNM substitutions are produced on false positive inference rates, we simulated evolution of alignments 5,000 codons long under the BS +MNM null model, using genome-median estimates for all parameters except δ, which we varied. At each value of δ, we simulated 50 replicates. We analyzed each replicate using the BST for selection on the human or *D. simulans* lineages and calculated the proportion of replicates for each value of δ that yielded a false positive inference ($P<0.05$).

We computed the observed proportion of tandem substitutions as a fraction of all substitutions on the human and *D. melanogaster* lineages in both empirical and simulated datasets. For each of the 6868 genes in the curated mammalian dataset, we aligned the human gene to the phylogenetically inferred sequence of the human-chimp ancestor, identified all substitutions as differences between these sequences, and calculated the proportion of tandem substitutions as the number of substitutions at adjacent sites divided by the sum of substitutions at adjacent sites and those at non-adjacent sites across all sites in the dataset. Substitutions at adjacent sites were counted as a single tandem substitution. For each

of the 8564 genes in the fly dataset, we aligned the *D. melanogaster* sequence to the *D. melanogaster/D. simulans* ancestor and followed the procedure described above. For simulated sequences, we repeated this procedure using the corresponding ancestral and terminal sequences simulated under the BS+MNM null model and parameters estimated from each gene in the empirical datasets, including δ.

BUSTED.

To examine the accuracy of BUSTED, we used Hyphy software 2.2.6 (batch files BUSTED.bf and QuickSelectionDetection.bf). We analyzed the 5,000 codon-long alignments simulated under the BS+MNM null model, using parameters estimated by ML for each BST-significant gene, with δ assigned either to its gene-specific estimate, its genome-average, or to zero. We applied BUSTED to the replicate alignments to test for selection ($P<0.05$) on the human lineage or the same fly lineage that was significant for that gene in the BST of the empirical data.

Power analyses.

To characterize the statistical power of the BST and BS+MNM tests, we simulated sequence evolution with positive selection of variable intensity and pervasiveness (Supplementary Fig. 4). Specifically, we used the BS positive model in Hyphy to simulate sequence evolution with the human and *D. simulans* terminal branches as the foreground branches. We used genome-average estimates of all parameters, including gene length (418 and 510 codons for mammals and flies, respectively), but we varied ω_2 and p_2. 20 replicate alignments were simulated under each set of conditions and then analyzed using the BST, the BS+MNM test, or BUSTED. For each set of conditions, the true positive rate was calculated as the fraction of replicates yielding a significant test of positive selection ($P<0.05$ for BST and BS+MNM, FDR<0.20 for at least one site in the alignment for BUSTED).

BS+MNM+ κ_2 model:

We developed the BS+MNM+ κ_2 model, which incorporates into the BS+MNM model (q_{ij} equation 2) two different transversion:transition rate ratio parameters, κ_1 for single-site substitutions and κ_2 for MNMs (see q_{ij} equation 3). All free parameters of the model are estimated by ML given a sequence alignment. This model was implemented by further modifying our BS+MNM batchfile in Hyphy 2.2.6 software by declaring κ_2 a global variable, incorporating it into the codon table, and allowing it to be optimized by ML as other parameters are in the batch file.

$\kappa_1\pi_j$	synonymous transversion	
π_j	synonymous transition	
$\omega\kappa_1\pi_j$	non-synonymous transversion	
$\omega\pi_j$	non-synonymous transition	
$\omega\delta\kappa_2^2\pi_j$	non-synonymous, 2 transversions	
$\omega\delta\pi_j$	non-synonymous, 2 transitions	3
$\omega\delta\kappa_2\pi_j$	non-synonymous, 1 transversion, 1 transition	
$\delta\pi_j$	synonymous, 2 transitions	
$\delta\kappa_2^2\pi_j$	synonymous, 2 transversions	
$\delta\kappa_2\pi_j$	synonymous, 1 transversion, 1 transition	
0	otherwise	

For validation, we estimated the parameters of the BS+MNM+ κ_2 null model by ML for every alignment in each dataset and calculated the genome-average median estimate of each parameter (**Supplementary Fig. 8**). We then simulated 50 replicate alignments of length 418 and 510 codons in the mammalian and fly datasets respectively, under the BS+MNM+ κ_2 null model with all model parameters set to their genome-wide median. We then estimated each parameter by ML under the null model given each alignment and compared the distribution of estimates to the parameters used to generate the alignments. We found that most parameters were estimated accurately, but estimates of κ_2 had high variance (Supplementary Figs. 8a-8b), presumably because the quantity of data in a single gene, in which CMDs are typically rare, is inadequate to support a robust estimate of this parameter. We therefore limited our use of this model to generating sequences by simulation rather than making inferences from sequence data.

To determine the effect of the MNM-specific transversion:transition rate on false-positive bias in the BST, we simulated sequences under the BS+MNM+κ_2 null model, using genome-median parameters except κ_2, which we varied. Sequences of length 10,000 codons long were used, because simulating shorter sequences resulted in a high variance in the realized transversion:transition ratio. For each value of κ_2, we simulated 50 replicates, applied the BST, and calculated the FPR as the fraction of replicates yielding a positive inference ($P<0.05$).

Data availability.

The empirical alignments reanalyzed in this study are available in the supplementary information of the original publications that generated and analyzed these data.[12,14,45]

Code availability.

The custom HYPHY batch codes for the BS+MNM and BS+MNM+κ_2 tests are available as supplementary files and at https://github.com/JoeThorntonLab/MNM_SelectionTests.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

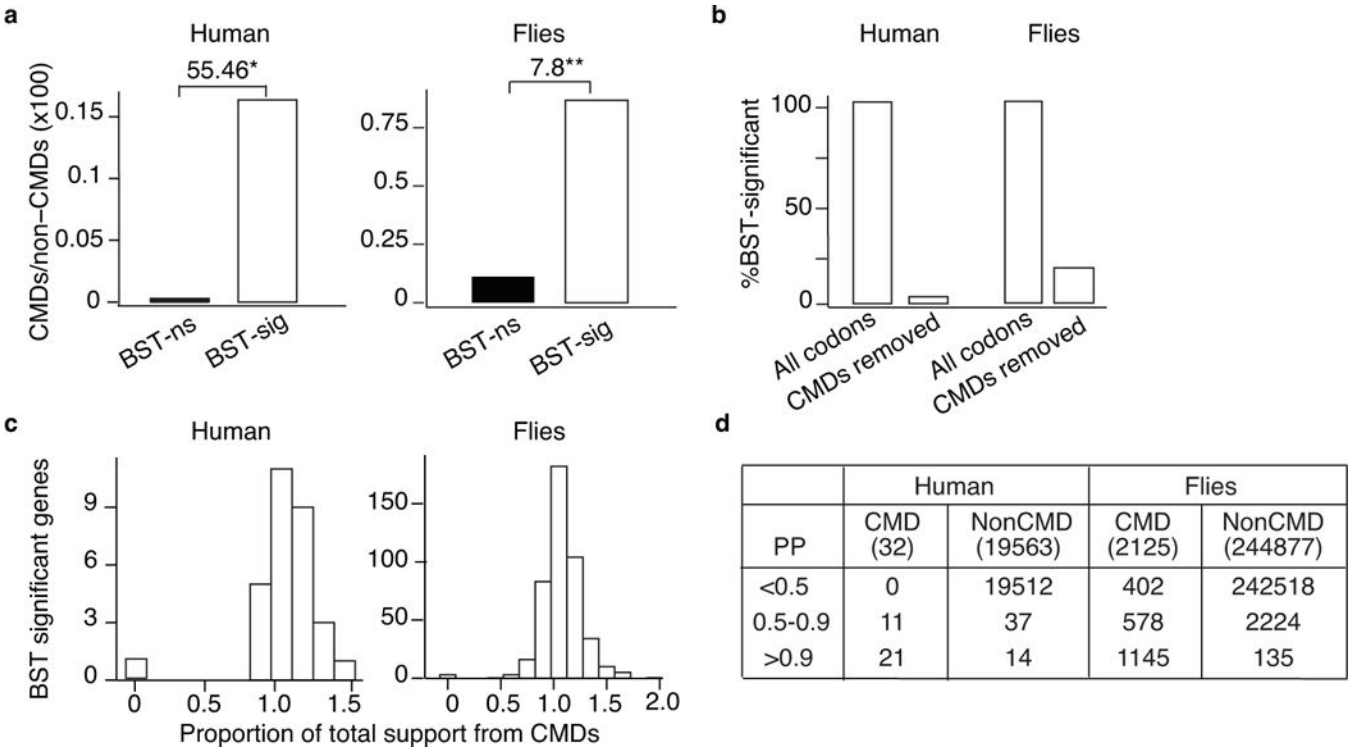
We are grateful to the members of the Thornton lab for discussion and helpful comments. We thank the Beagle2, Midway2, and Tarbell supercomputing clusters at the University of Chicago. We also thank the developers of HyPhy for presenting an open source platform that allows limitless customization of standard analyses. Funding was provided by NIH R01GM104397 and R01GM121931 (JWT), NSF DEB-1601781 (JWT and AV), NSF DBI-1564611 (MWH), and the Precision Health Initiative of Indiana University (MWH).

REFERENCES CITED

1. Goldman N & Yang Z A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11, 725–736 (1994). [PubMed: 7968486]
2. Murrell B et al. Gene-wide identification of episodic selection. Mol Biol Evol 32, 1365–1371 (2015). [PubMed: 25701167]
3. Murrell B et al. Detecting individual sites subject to episodic diversifying selection. PLoS Genet 8, e1002764 (2012). [PubMed: 22807683]
4. Smith MD et al. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Mol Biol Evol 32, 1342–1353 (2015). [PubMed: 25697341]
5. Yang Z & Nielsen R Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol 19, 908–917 (2002). [PubMed: 12032247]
6. 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). [PubMed: 16107592]
7. Pond SL , Frost SD & Muse SV HyPhy: hypothesis testing using phylogenies. Bioinformatics 21, 676–679 (2005). [PubMed: 15509596]
8. Kosiol C , Holmes I & Goldman N An empirical codon model for protein sequence evolution. Mol Biol Evol 24, 1464–1479 (2007). [PubMed: 17400572]
9. Whelan S & Goldman N Estimating the frequency of events that cause multiple-nucleotide changes. Genetics 167, 2027–2043 (2004). [PubMed: 15342538]
10. Muse SV & Gaut BS A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Mol Biol Evol 11, 715–724 (1994). [PubMed: 7968485]
11. Han MV , Demuth JP , McGrath CL , Casola C & Hahn MW Adaptive evolution of young gene duplicates in mammals. Genome Res 19, 859–867 (2009). [PubMed: 19411603]
12. Consortium DG et al. Evolution of genes and genomes on the Drosophila phylogeny. Nature 450, 203–218 (2007). [PubMed: 17994087]
13. Foote AD et al. Convergent evolution of the genomes of marine mammals. Nat Genet 47, 272–275 (2015). [PubMed: 25621460]
14. Kosiol C et al. Patterns of positive selection in six Mammalian genomes. PLoS Genet 4, e1000144 (2008). [PubMed: 18670650]
15. Roux J et al. Patterns of positive selection in seven ant genomes. Mol Biol Evol 31, 1661–1685 (2014). [PubMed: 24782441]
16. Yang Z & dos Reis M Statistical properties of the branch-site test of positive selection. Mol Biol Evol 28, 1217–1228 (2011). [PubMed: 21087944]
17. Zhang J Performance of likelihood ratio tests of evolutionary hypotheses under inadequate substitution models. Mol Biol Evol 16, 868–875 (1999). [PubMed: 10368963]
18. Gharib WH & Robinson-Rechavi M The branch-site test of positive selection is surprisingly robust but lacks power under synonymous substitution saturation and variation in GC. Mol Biol Evol 30, 1675–1686 (2013). [PubMed: 23558341]

19. Zhai W , Nielsen R , Goldman N & Yang Z Looking for Darwin in genomic sequences--validity and success of statistical methods. Mol Biol Evol 29, 2889–2893 (2012). [PubMed: 22490825]
20. Nozawa M , Suzuki Y & Nei M Reliabilities of identifying positive selection by the branch-site and the site-prediction methods. Proc Natl Acad Sci U S A 106, 6700–6705 (2009). [PubMed: 19339501]
21. Casola C & Hahn MW Gene conversion among paralogs results in moderate false detection of positive selection using likelihood methods. J Mol Evol 68, 679–687 (2009). [PubMed: 19471989]
22. Anisimova M & Yang Z Multiple hypothesis testing to detect lineages under positive selection that affects only a few sites. Mol Biol Evol 24, 1219–1228 (2007). [PubMed: 17339634]
23. Kosakovsky Pond SL et al. A random effects branch-site model for detecting episodic diversifying selection. Mol Biol Evol 28, 3033–3043 (2011). [PubMed: 21670087]
24. Zhang J Frequent false detection of positive selection by the likelihood method with branch-site models. Mol Biol Evol 21, 1332–1339 (2004). [PubMed: 15014150]
25. Schrider DR , Hourmozdi JN & Hahn MW Pervasive multinucleotide mutational events in eukaryotes. Curr Biol 21, 1051–1054 (2011). [PubMed: 21636278]
26. Saribasak H et al. DNA polymerase ζ generates tandem mutations in immunoglobulin variable regions. J Exp Med 209, 1075–1081 (2012). [PubMed: 22615128]
27. Loeb LA & Monnat RJ DNA polymerases and human disease. Nat Rev Genet 9, 594–604 (2008). [PubMed: 18626473]
28. Matsuda T , Bebenek K , Masutani C , Hanaoka F & Kunkel TA Low fidelity DNA synthesis by human DNA polymerase-eta. Nature 404, 1011–1013 (2000). [PubMed: 10801132]
29. Seplyarskiy VB , Bazykin GA & Soldatov RA Polymerase ζ Activity Is Linked to Replication Timing in Humans: Evidence from Mutational Signatures. Mol Biol Evol 32, 3158–3172 (2015). [PubMed: 26376651]
30. Stone JE , Lujan SA , Kunkel TA & Kunkel TA DNA polymerase zeta generates clustered mutations during bypass of endogenous DNA lesions in Saccharomyces cerevisiae. Environ Mol Mutagen 53, 777–786 (2012). [PubMed: 22965922]
31. Arana ME , Seki M , Wood RD , Rogozin IB & Kunkel TA Low-fidelity DNA synthesis by human DNA polymerase theta. Nucleic Acids Res 36, 3847–3856 (2008). [PubMed: 18503084]
32. Besenbacher S et al. Multi-nucleotide de novo Mutations in Humans. PLoS Genet 12, e1006315 (2016). [PubMed: 27846220]
33. Chen JM , Férec C & Cooper DN Complex Multiple-Nucleotide Substitution Mutations Causing Human Inherited Disease Reveal Novel Insights into the Action of Translesion Synthesis DNA Polymerases. Hum Mutat 36, 1034–1038 (2015). [PubMed: 26172832]
34. Chen JM , Cooper DN & Férec C A new and more accurate estimate of the rate of concurrent tandem-base substitution mutations in the human germline: ~0.4% of the single-nucleotide substitution mutation rate. Hum Mutat 35, 392–394 (2014). [PubMed: 24375656]
35. Harris K & Nielsen R Error-prone polymerase activity causes multinucleotide mutations in humans. Genome Res 24, 1445–1454 (2014). [PubMed: 25079859]
36. Hodgkinson A & Eyre-Walker A Variation in the mutation rate across mammalian genomes. Nat Rev Genet 12, 756–766 (2011). [PubMed: 21969038]
37. Assaf ZJ , Tilk S , Park J , Siegal ML & Petrov DA Deep sequencing of natural and experimental populations of Drosophila melanogaster reveals biases in the spectrum of new mutations. Genome Res 27, 1988–2000 (2017). [PubMed: 29079675]
38. Francioli LC et al. Genome-wide patterns and properties of de novo mutations in humans. Nat Genet 47, 822–826 (2015). [PubMed: 25985141]
39. Zhu W et al. Concurrent nucleotide substitution mutations in the human genome are characterized by a significantly decreased transition/transversion ratio. Hum Mutat 36, 333–341 (2015). [PubMed: 25546635]
40. Averof M , Rokas A , Wolfe KH & Sharp PM Evidence for a high frequency of simultaneous double-nucleotide substitutions. Science 287, 1283–1286 (2000). [PubMed: 10678838]

41. Bazykin GA , Kondrashov FA , Ogurtsov AY , Sunyaev S & Kondrashov AS Positive selection at sites of multiple amino acid replacements since rat-mouse divergence. Nature 429, 558–562 (2004). [PubMed: 15175752]
42. Rogozin IB et al. Evolutionary switches between two serine codon sets are driven by selection. Proc Natl Acad Sci U S A 113, 13109–13113 (2016). [PubMed: 27799560]
43. De Maio N , Holmes I , Schlötterer C & Kosiol C Estimating empirical codon hidden Markov models. Mol Biol Evol 30, 725–736 (2013). [PubMed: 23188590]
44. Suzuki Y False-positive results obtained from the branch-site test of positive selection. Genes Genet Syst 83, 331–338 (2008). [PubMed: 18931458]
45. Larracuente AM et al. Evolution of protein-coding genes in Drosophila. Trends Genet 24, 114–123 (2008). [PubMed: 18249460]
46. Sironi M , Cagliani R , Forni D & Clerici M Evolutionary insights into host-pathogen interactions from mammalian sequence data. Nat Rev Genet 16, 224–236 (2015). [PubMed: 25783448]
47. Elde NC , Child SJ , Geballe AP & Malik HS Protein kinase R reveals an evolutionary model for defeating viral mimicry. Nature 457, 485–489 (2009). [PubMed: 19043403]
48. Patel MR , Loo YM , Horner SM , Gale M & Malik HS Convergent evolution of escape from hepaciviral antagonism in primates. PLoS Biol 10, e1001282 (2012). [PubMed: 22427742]
49. Demogines A , Abraham J , Choe H , Farzan M & Sawyer SL Dual host-virus arms races shape an essential housekeeping protein. PLoS Biol 11, e1001571 (2013). [PubMed: 23723737]
50. Barber MF & Elde NC Nutritional immunity. Escape from bacterial iron piracy through rapid evolution of transferrin. Science 346, 1362–1366 (2014). [PubMed: 25504720]
51. Machkovech HM , Bedford T , Suchard MA & Bloom JD Positive Selection in CD8+ T-Cell Epitopes of Influenza Virus Nucleoprotein Revealed by a Comparative Analysis of Human and Swine Viral Lineages. J Virol 89, 11275–11283 (2015). [PubMed: 26311880]
52. Field SF , Bulina MY , Kelmanson IV , Bielawski JP & Matz MV Adaptive evolution of multicolored fluorescent proteins in reef-building corals. J Mol Evol 62, 332–339 (2006). [PubMed: 16474984]
53. Yokoyama S , Tada T , Zhang H & Britt L Elucidation of phenotypic adaptations: Molecular analyses of dim-light vision proteins in vertebrates. Proc Natl Acad Sci U S A 105, 13480–13485 (2008). [PubMed: 18768804]
54. Zhuang H , Chien MS & Matsunami H Dynamic functional evolution of an odorant receptor for sex-steroid-derived odors in primates. Proc Natl Acad Sci U S A 106, 21247–21251 (2009). [PubMed: 19955411]
55. Bloom JD An experimentally determined evolutionary model dramatically improves phylogenetic fit. Mol Biol Evol 31, 1956–1978 (2014). [PubMed: 24859245]
56. Lopez P , Casane D & Philippe H Heterotachy, an important process of protein evolution. Mol Biol Evol 19, 1–7 (2002). [PubMed: 11752184]
57. Pond SK & Muse SV Site-to-site variation of synonymous substitution rates. Mol Biol Evol 22, 2375–2385 (2005). [PubMed: 16107593]
58. Chan YF et al. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science 327, 302–305 (2010). [PubMed: 20007865]
59. Barrett RD & Hoekstra HE Molecular spandrels: tests of adaptation at the genetic level. Nat Rev Genet 12, 767–780 (2011). [PubMed: 22005986]
60. Siddiq MA , Loehlin DW , Montooth KL & Thornton JW Experimental test and refutation of a classic case of molecular adaptation in Drosophila melanogaster. Nat Ecol Evol 1, 25 (2017). [PubMed: 28812605]



	Human		Flies	
PP	CMD (32)	NonCMD (19563)	CMD (2125)	NonCMD (244877)
<0.5	0	19512	402	242518
0.5-0.9	11	37	578	2224
>0.9	21	14	1145	135

Figure 1.

Codons with multiple nucleotide differences (CMDs) drive branch-site signatures of selection. **(a)** CMDs are enriched in genes with a signature of positive selection. Codons were classified by the number of nucleotide differences between the ancestral and terminal states on branches tested for positive selection. CMDs have ≥2 differences; non-CMDs have ≤1 difference. The CMD/non-CMD ratio is shown for genes with a significant signature of selection in the BST (BST-sig) and those without (BST-ns). Fold-enrichment is shown as the odds ratio. *, P=4e-4 by χ^2 test; **, P=1e-41 by Fisher's exact test. **(b)** Percentage of genes that retain a signature of positive selection when CMDs are excluded from the branch-site test analysis. **(c)** Distribution across BST-significant genes of the proportion of total support for the positive selection model that is provided by CMDs. Total support is the difference in log-likelihood between the positive selection and null models, summed over all codons in the alignment. Support from CMDs is summed over codons with multiple differences. The proportion of support from CMDs can be greater than 1 if the log-likelihood difference between models is negative at non-CMDs. **(d)** Most codons classified as positively selected are CMDs. The number of CMDs and non-CMDs in BST-significant genes are grouped by the Bayes Empirical Bayes posterior probability (PP) that they are in the positively selected class.

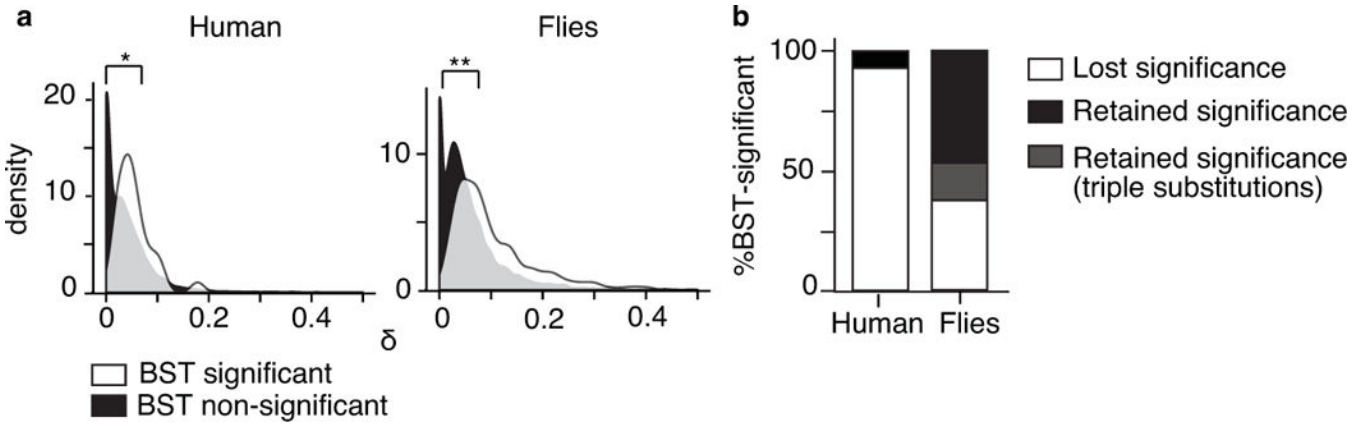


Figure 2.

Incorporating MNMs into the branch-site model eliminates the signature of positive selection in many genes. The mammalian and fly datasets were reanalyzed using a version of the BST that allows MNMs (BS+MNM) by including a parameter δ, a multiplier on the rate of each double substitution relative to single substitutions. **(a)** The distribution of ML estimates of δ across genes with (white) and without (black) a significant result in the classic BST is shown for empirical alignments. Median estimates of δ in BST-significant and BST-nonsignificant genes are 0.047 and 0.026 in humans, respectively, and 0.107 and 0.062 in flies. *, P=6.7e-4; **, P=1e-8 by Mann-Whitney U-test. **(b)** Proportion of tests with a significant result in the BST that lose or retain that signature using the BS+MNM test. Genes with tests that remain significant but contain CMDs with three differences, which are not incorporated into BS+MNM, are also shown.

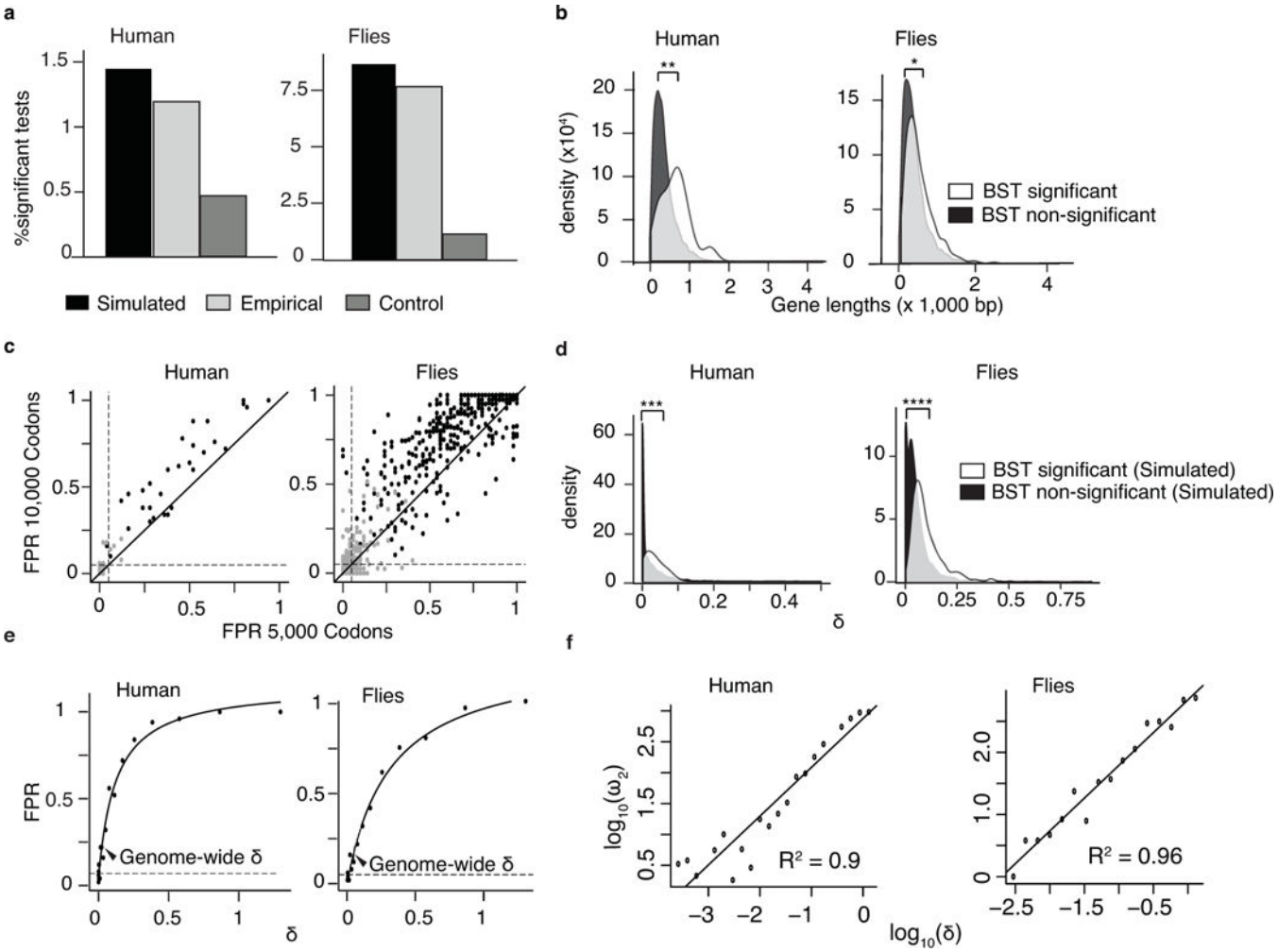


Figure 3.

MNMs cause a strong bias in the BST under realistic conditions. For each gene in the mammalian and fly datasets, the parameters of the BS+MNM null model were estimated by maximum likelihood. Sequences of the original gene length were then simulated under these parameters and analyzed using the classic BST. **(a)** The fraction of all tests that are BST-significant (P<0.05) is shown for the data simulated under the BS+MNM null model, the empirical data, and a control dataset simulated with $\delta = 0$. **(b)** BST-significant genes are longer than BST non-significant genes. The probability density of gene lengths in the two categories is shown for the empirical datasets. Median lengths were 642 and 343 bp in humans; in flies, 448 and 399 bp. Mann-Whitney U test for differences in the distributions: *, P=8e-4; **, P=8e-5. **(c)** Systematic bias in the BST. For each empirical BST-significant gene, we simulated 50 replicate alignments using the BS+MNM null model parameters at lengths 5,000 and 10,000 codons, then analyzed them using the BST. The false positive rate (FPR) for any gene's simulation (black points) is the proportion of replicates significant (P<0.05). Gray points, FPR for control simulations with $\delta = 0$. Dashed lines, FPR of 0.05. Solid diagonal line has a slope of 1. **(d)** Distribution of ML estimates of δ across genes with (white) and without (black) a signature of positive selection in the classic BST is shown for

data simulated under the BS+MNM null model. Median δ in BST-significant and BST-nonsignificant genes was 0.03 and 0.0009 in humans, 0.08 and 0.04 in flies. Mann-Whitney U test for difference between distributioins, *** P=1e-12; **** P=1e-199. **(e)** Increasing the MNM rate exacerbates bias in the BST. Sequences 5,000 codons long were simulated using the BS+MNM null model and the median value of each model parameter and branch length across all genes in each dataset, with variable δ. False positive rate (FPR, $P<0.05$) in 50 replicates at each value of δ is shown. Solid line, hyperbolic fit to the data; dotted line, FPR=5%. Arrowhead, median δ across all genes. **(f)** Relationship between δ and inferred ω_2. Sequences simulated in (e) were used to infer the ω_2 under the BST positive selection model. Best-fit linear regression line and coefficient of determination are shown.

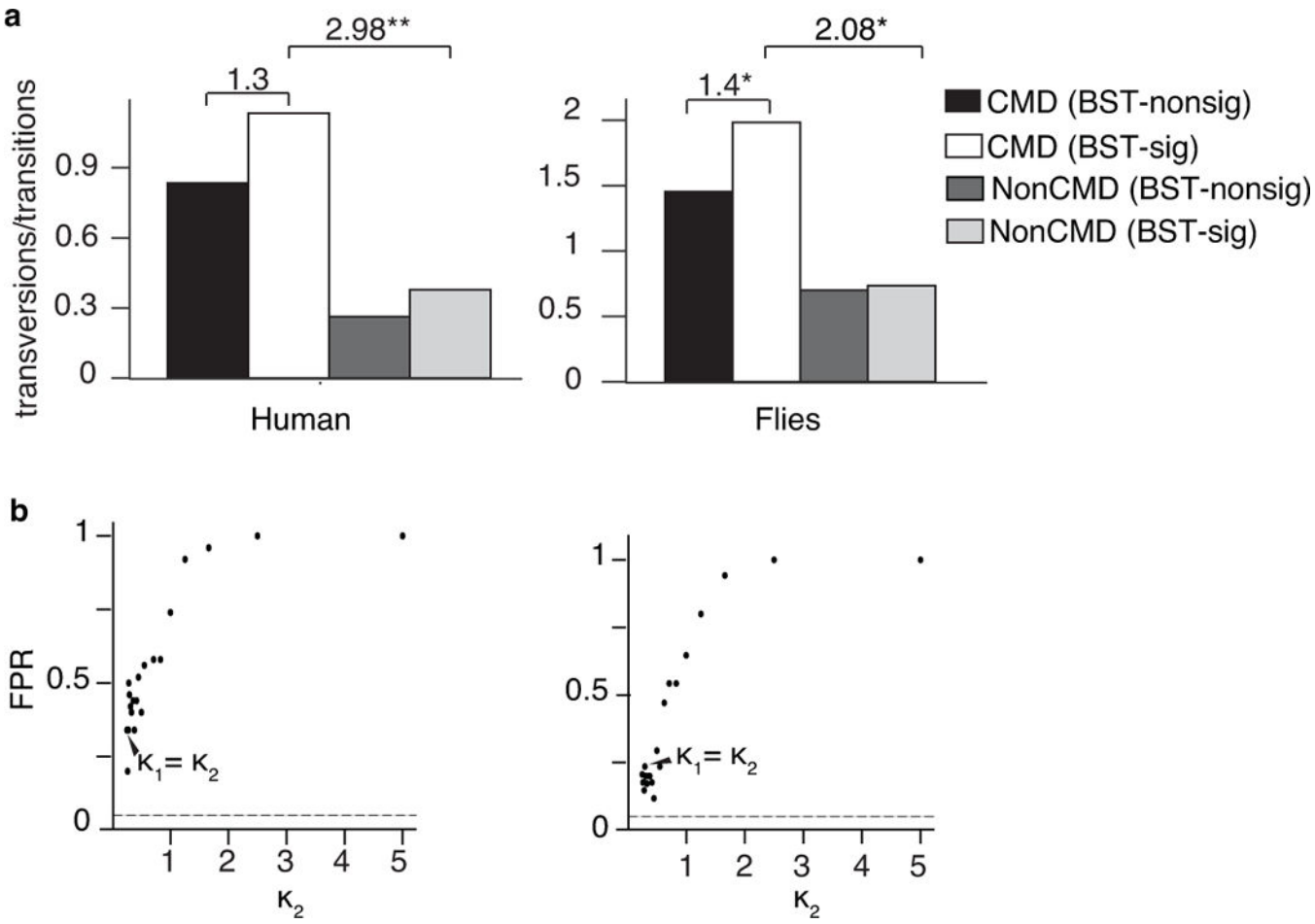


Figure 4.

Transversion enrichment in CMDs biases the BST. **(a)** The ratio of transversions:transitions observed in CMDs and in non-CMDs for BST-significant and BST-nonsignificant genes. Fold-enrichment is shown as the odds ratio. *, P=5e-4; **, P=3e-25 by Fisher's exact test. **(b)** Increasing the transversion rate in MNMs increases bias of the BST. Sequences 10,000 codons long were simulated using an elaboration of the BS+MNM model that allows MNMs to have a transversion:transition rate (κ_2) different from that in single-nucleotide substitutions (κ_1). 50 replicate alignments were simulated under the null model using the average value of every model parameter and branch length across all genes in each dataset, except κ_2 was allowed to vary. The rate of false positives (P<0.05) at each value of κ_2 is shown. Arrowheads show the false positive rate when sequences were simulated with κ_2 equal to κ_1. Dotted line, FPR of 5%.

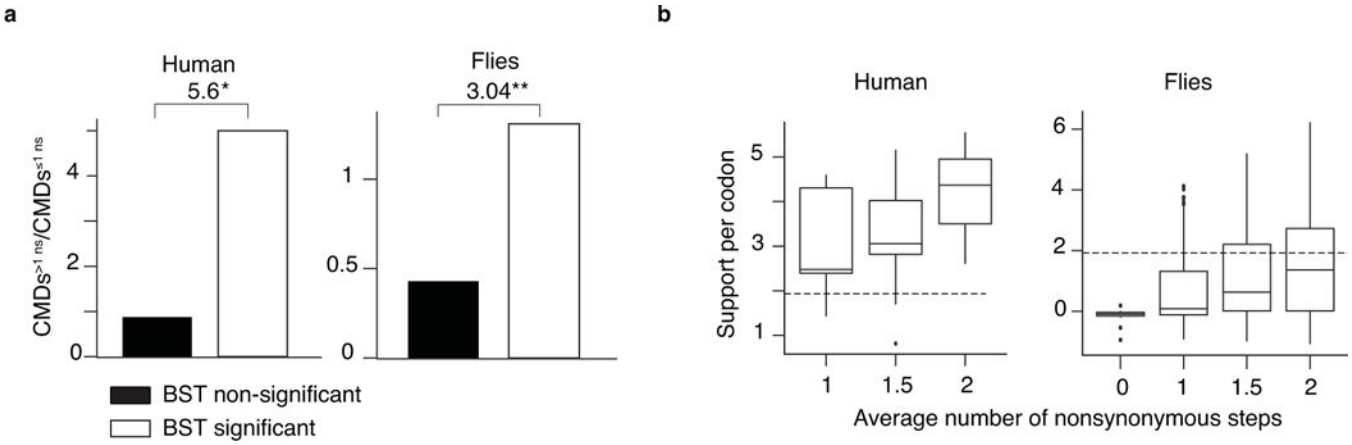


Figure 5.

CMDs implying multiple nonsynonymous steps drive the BST. **(a)** For every CMD in every gene, the mean of the number of nonsynonymous single-nucleotide steps on the two direct paths between the ancestral and derived sequence states on the branch of interest was calculated. In BST-significant and BST-nonsignificant genes, the ratio of CMDs invoking more than one nonsynonymous step to those invoking one or fewer such steps is shown. Fold-enrichment is shown as the odds ratio. *, P=9e-04; **P= 1.6e-67 by Fisher's exact test. **(b)** Support for the positive selection model provided by CMDs depends on the number of implied nonsynonymous single-nucleotide steps. Support is the log-likelihood difference between the positive selection and null models of the BST given the data at a single codon site. Box plots show the distribution of support by CMDs in BST significant genes categorized according to the mean number of implied nonsynonymous steps. Dotted line, support of 1.92, at which the BST yields a significant result for an entire gene (P<0.05). In human BST-significant genes, there are no CMDs that imply zero non-synonymous changes.