Parallel Evolution of Metazoan Mitochondrial Proteins

Galya V. Klink[1] and Georgii A. Bazykin[1,2,*]

[1]Institute for Information Transmission Problems (Kharkevich Institute) of the Russian Academy of Sciences, Moscow, Russia

[2]Skolkovo Institute of Science and Technology, Skolkovo, Russia

*Corresponding author: E-mail: gbazykin@iitp.ru.

Accepted: February 6, 2017

Abstract

Amino acid propensities at amino acid sites change with time due to epistatic interactions or changing environment, affecting the probabilities of fixation of different amino acids. Such changes should lead to an increased rate of homoplasies (reversals, parallelisms, and convergences) at closely related species. Here, we reconstruct the phylogeny of twelve mitochondrial proteins from several thousand metazoan species, and measure the phylogenetic distances between branches at which either the same allele originated repeatedly due to homoplasies, or different alleles originated due to divergent substitutions. The mean phylogenetic distance between parallel substitutions is ~20% lower than the mean phylogenetic distance between divergent substitutions, indicating that a variant fixed in a species is more likely to be deleterious in a more phylogenetically remote species, compared with a more closely related species. These findings are robust to artefacts of phylogenetic reconstruction or of pooling of sites from different conservation classes or functional groups, and imply that single-position fitness landscapes change at rates similar to rates of amino acid changes.

Key words: fitness landscape, epistasis, parallel substitutions, heteropecilly, mitochondria.

Introduction

Amino acid preferences at a site, or single-position fitness landscape (SPFL, Bazykin 2015), change in the course of evolution, so that a variant conferring high fitness in one species may confer low fitness in another, either due to changes at interacting genomic sites or in the environment. These changes can be observed through phylogenetic patterns, in particular, through a non-uniform distribution of amino acid substitutions giving rise to a particular variant (homoplasies) along the phylogeny. Indeed, when a certain amino acid repeatedly arises at a particular site in a certain phylogenetic clade, but is never observed at this site in another clade, this implies that the relative fitness conferred by this variant in the former clade is higher. Different types of homoplasies—reversals, parallelisms, and convergencies—have been found to be clustered on the phylogenies (Rogozin et al. 2008; Povolotskaya and Kondrashov 2010; Naumenko et al. 2012; Goldstein et al. 2015; Zou and Zhang 2015), and an attempt has been made to estimate the rate at which SPFLs change from such data (Usmanova et al. 2015).

SPFL changes in metazoan mitochondrial proteins were previously inferred from amino acid usage patterns (Breen et al. 2012), but this approach has been criticized as sensitive to the underlying assumptions regarding fitness distributions (McCandlish et al. 2013). Here, we develop an approach for the study of phylogenetic clustering of homoplasies at individual protein sites, and apply it to deep alignments of mitochondrial proteins of metazoans (Breen et al. 2012) together with their reconstructed phylogenies. Our approach compares the distributions of distances between parallel and divergent substitutions to infer robustly changes in relative fitness of different variants at a site between branches of the phylogenetic tree.

Materials and Methods

Alignment and Phylogeny

We obtained multiple-species alignments of 12 mitochondrial proteins of metazoans from (Breen et al. 2012), and analyzed alignment columns with <1% gaps (which comprised 77% of all sites). As there is no accepted phylogenetic tree for this large and diverse set of species spanning a wide range of phylogenetic distances, we took a hybrid approach to reconstructing their phylogeny. First, we constrained the tree topology using the curated taxonomy-based phylogeny of the ITOL (Interactive tree of life) project (Letunic and Bork 2007). By requiring the presence of each species in the ITOL database, we were left with >900 metazoan species for each protein

Table 1
Amino Acid Substitutions in Mitochondrial Genes of Metazoans

Gene	Species	Amino Acid Sites	Amino Acids per Site	Substitutions per Site	Amino Acids per Site in Simulation	Substitutions per Site in Simulation
ATP6	2,931	186	9.5	152.1	12.1	154.63
COX1	4,366	404	6.1	63.8	10.6	117.26
COX2	4,131	165	8.9	137.2	12.4	206.83
COX3	2,152	198	9.2	131.6	13.0	167.41
CYTB	5,995	327	9.4	174.2	13.3	252.03
ND1	2,013	253	8.7	92.9	12.7	124.42
ND2	5,765	299	10.2	259.6	13.6	313.94
ND3	2,766	94	9.7	182.3	12.8	194.37
ND4	2,007	392	9.1	127.1	13.1	165.89
ND4L	1,759	82	11.3	139.3	13.6	175.37
ND5	926	516	7.9	57.6	11.3	75.72
ND6	996	119	10.5	76.6	13.2	103.45

(table 1). The resulting topology was not fully resolved, and contained multifurcations. We then used RAxML 8.0.0 (Stamatakis 2014) under the GTR-Γ model to resolve multifurcations and to estimate the branch lengths. Finally, ancestral states were reconstructed using codeml program of the PAML package (Yang 1997) under the substitution matrix and the value of the parameter alpha of the gamma distribution inferred by RAxML.

Independently, using the same methods and parameters, we reconstructed a joint phylogeny of 3,586 chordate, 586 nonchordate, and 178 fungal species (for a total of 4,350 species) based on amino acid sequences of five concatenated mitochondrial genes, each of which was aligned by MUSCLE covering a total of 1,524 amino acid positions.

Transmembrane and nonmembrane sites were obtained from UniProt database (http://www.uniprot.org/).

Clustering of Substitutions on a Phylogeny

Using the inferred states of amino acid sites at each node, we inferred the positions of all substitutions at all protein sites on the phylogenies of the corresponding proteins. For each ancestral amino acid at a site, we defined parallel substitutions as those giving rise to the same derived amino acid, and divergent substitutions, as those giving rise to different derived amino acids (fig. 1*A*). We considered only those pairs of substitutions that happened in phylogenetically independent branches, that is, such that one was not ancestral to the other. For subsequent analyses, we used only homoplasy-informative sites, that is, sites that have at least one pair of parallel substitutions and one pair of divergent substitutions from the same ancestral amino acid. The phylogenetic distance between a pair of substitutions was defined as the distance (measured in the number of amino acid substitutions per amino acid site inferred by RAxML) between the centers of the edges where those substitutions have occurred, i.e. the sum of the distances from the centers of these edges to the last common ancestor of the two substitutions (fig. 1*A*). Alternatively, as a proxy for the site-specific evolutionary distance, we multiplied the phylogenetic distance by the codeml estimate of the site-specific substitution rate.

To compare the distances between parallel and divergent substitutions while circumventing the potential biases associated with pooling sites and amino acids with different properties (see below), we subsampled the pairs of parallel and divergent substitutions, and analyzed the distances in this subset. For this, for each ancestral amino acid at homoplasy-informative sites, we picked randomly min(N_{paral}, N_{diverg}) pairs of parallel substitutions and the same number of pairs of divergent substitutions, where N_{paral} and N_{diverg} are the numbers of parallel and divergent substitutions originating from this amino acid. We repeated this procedure for all ancestral amino acids at all sites, thus obtaining two equal-sized subsamples of parallel and divergent substitutions, and measured all distances in these resulting subsamples. The parallel to divergent ratio (P/D) for each distance bin was calculated by dividing the number of parallel pairs of substitutions by the number of divergent pairs of substitutions that had occurred at a distance from each other falling into this bin. This statistic is closely related to the O-ring statistic widely used in spatial ecology to measure aggregation in communities (Wiegand and Moloney 2004).

To obtain mean values and CIs of each statistic, we bootstrapped sites in 1,000 replicates, each time repeating the entire resampling procedure.

Robustness of Tree Shape and Ancestral States Reconstruction

We identified a set of high-confidence pairs of substitutions, defined as follows. For each branch of the phylogenetic tree, we obtained the bootstrap support value in 100 bootstrap replicates using RAxML. A pair of substitutions was considered high-confidence when 1) at least one node between

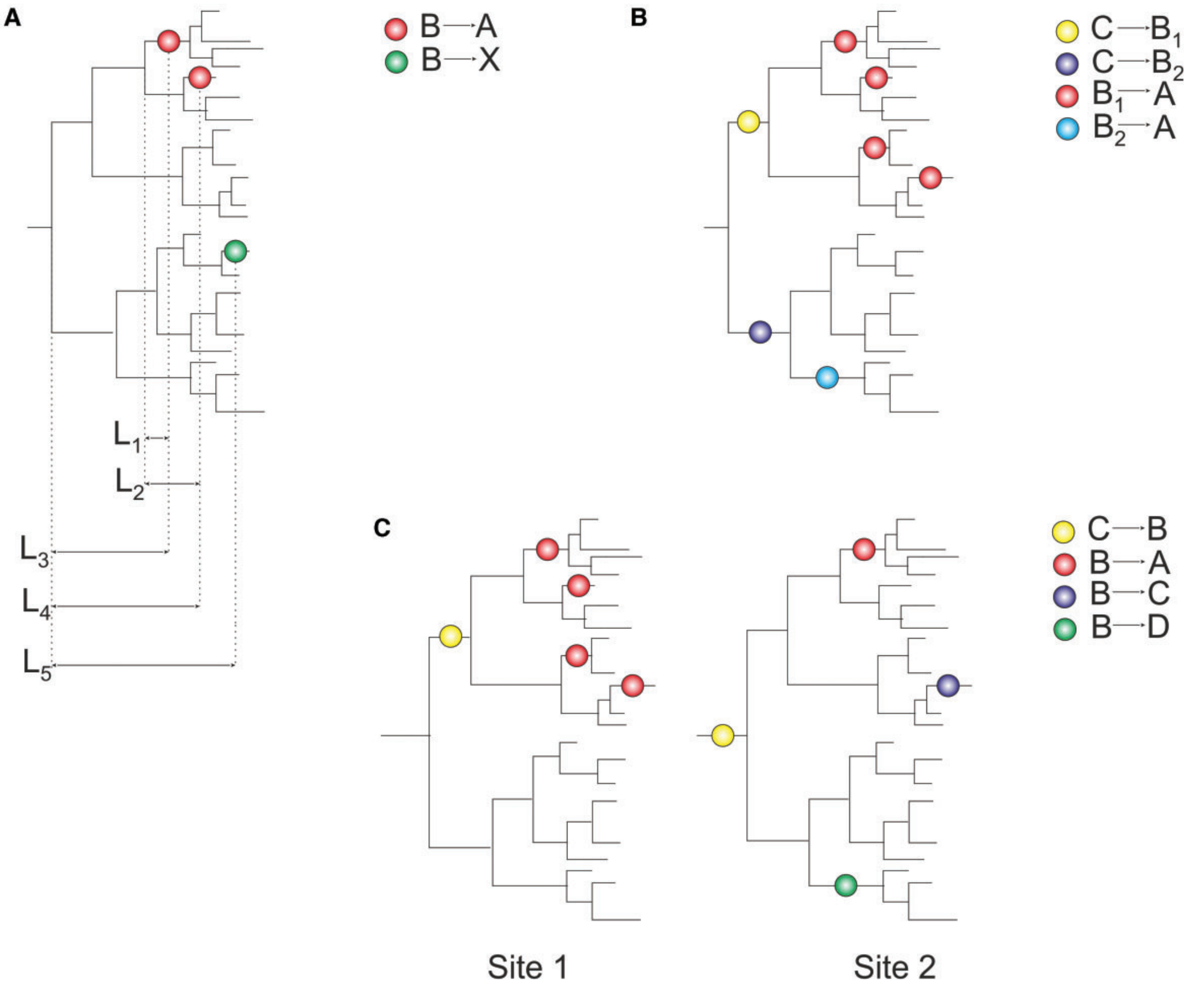

Fig. 1.—Inference of phylogenetic distances between parallel and divergent substitutions. Dots represent substitutions mapped to nodes of a phylogenetic tree. *(A)* For each pair of amino acids (B, A) at a particular amino acid site, we consider the distances between all parallel B→A substitutions (L_1+L_2), and distances between all divergent substitutions B→A and B→X (L_3+L_5, L_4+L_5), where X is any amino acid other than A and B. *(B)* The B_1→A substitution is more frequent than the B_2→A substitution, leading to an excess of homoplasies at small phylogenetic distances when parallel and convergent substitutions are pooled together. *(C)* Pooling of sites with different properties may also lead to an excess of homoplasies at small phylogenetic distances (see text).

substitutions had 100% bootstrap support, ensuring the robustness of these nodes; and 2) for each substitution from the pair, the maximum likelihood estimate for amino acids in ancestral and derived nodes was equal to 1, ensuring the robustness of ancestral state reconstruction.

Simulated Evolution

For each gene, we simulated amino acid evolution using the evolver program of the PAML package (Yang 1997), under the empirical_F model and discrete-gamma distributed rates between sites. The phylogenetic tree, substitution matrix, alpha parameter and number of categories for discrete gamma of the gamma-distribution were obtained from the output of RAxML for the corresponding gene. From the amino acids thus simulated for the leaves of the tree (i.e. extant species), we then reconstructed the ancestral states using codeml under the same parameters.

Simulated Evolution under Different Substitution Matrices

To test the effect of differences in substitution matrices between clades due to clade-specific biases, we obtained individual RAxML-generated matrices for each of the three major groups of species in the joint five-gene phylogeny: chordates, nonchordates and fungi, using the same methods as above.

We then used these matrices to simulate evolution of the corresponding groups of species of the joint 4,350-species tree, and used generated sequences for the analysis.

Results

Phylogenetic Clustering as Evidence for SPFL Changes

We devised an approach for analysis of the clustering of parallel substitutions at a site. While conceptually related to the previous methods (Povolotskaya and Kondrashov 2010; Goldstein et al. 2015; Zou and Zhang 2015), it is designed to be robust to other specifics of the phylogenetic distribution of substitutions, and to control for any potential biases that can arise from pooling sites with different properties. Since it is difficult to obtain robust evidence for SPFL changes for an individual amino acid site even using large numbers of species, getting a significant signal of SPFL changes requires pooling different amino acid sites. The problem is that these sites may differ in their properties, and such differences may lead to artefactual evidence for SPFL changes, for the following reasons.

First, pooling of parallel and convergent substitutions giving rise to the same descendant variant, that is, substitutions with the same and different ancestral variants, may provide artefactual evidence for SPFL changes due to reasons such as the structure of the genetic code. For example (fig. 1*B*), an amino acid A within a clade may arise repeatedly from the ancestral amino acid B_1 at a particular clade simply because the $B_1 \rightarrow A$ mutation is frequent. If another amino acid B_2 is more prevalent than B_1 at a different clade, and the $B_2 \rightarrow A$ mutation is less frequent, this will lead to an excess of substitutions giving rise to A in the former clade; this excess; however, is not an evidence for SPFL changes, but instead occurs for nonselective reasons. To control for this, we do not consider convergent substitutions, and separately consider the distributions of parallel and divergent substitutions from each ancestral variant B.

Second, even independent consideration of different ancestral variants still permits clustering of homoplasies without SPFL changes when sites, and amino acids within sites, with diverse properties are pooled together. To illustrate this, assume that we analyze phylogenetic distances between parallel and divergent substitutions in a pooled sample of sites. Consider the hypothetical scenario in figure 1C. At site 1, the amino acid B only resides within a relatively small clade. Therefore, both $B \rightarrow A$ and $B \rightarrow X$ substitutions are, by necessity, phylogenetically close to each other. In contrast, at site 2, the amino acid B is long living, and the distances between substitutions from it may be larger. If such sites also differ systematically in their amino acid propensities, this might lead to artefactual evidence for SPFL changes. For example, if sites where B is short-living (like site 1) also tend to be those where few amino acids confer high fitness (so that $B \rightarrow A$ substitutions are more frequent), while sites where B spans a large clade tend to be promiscuous with respect to the occupied amino acid (so that $B \rightarrow X$ substitutions are more frequent), pooling such sites may result in an excess of homoplasies within short phylogenetic distances.

We circumvent this problem by resampling matched sets of parallel and divergent substitutions. Specifically, at each homoplasy-informative site (see "Materials and Methods" section), we consider different ancestral amino acids separately. For each ancestral amino acid B, we subsample our sets of pairs of substitutions: for each pair of parallel ($B \rightarrow A$, $B \rightarrow A$) substitutions, we randomly pick exactly one pair of divergent ($B \rightarrow A$, $B \rightarrow X$) substitutions from the same site. Finally, we pool together these subsamples from different sites, and analyze distances between parallel and divergent substitutions in this pooled set. This approach controls for any possible biases associated with differences in phylogenetic distributions of different ancestral amino acids. From the resulting subsets of distances, we calculate the ratio of the numbers of parallel to divergent substitutions (P/D) for each distance bin (see "Materials and Methods" section).

Parallel Substitutions in Mitochondrial Proteins Are Phylogenetically Clustered

We applied this approach to the phylogenetic trees of 12 orthologous mitochondrial proteins of metazoans, each including >900 species. At the vast majority of sites, we observe many amino acid variants, in line with (Breen et al. 2012). By reconstructing the ancestral states and substitutions at each site, we observe that most of these variants have originated more than once, allowing us to study the phylogenetic distribution of homoplasies in detail (table 1 and supplementary table S1, Supplementary Material online).

We observe an excess of parallel substitutions for species at small phylogenetic distances from each other (figs. 2 and 3), in line with the previous findings in vertebrates that used a smaller data set (Goldstein et al. 2015). The P/D ratio is ~1.7–2.5 at phylogenetic distances <0.1, but drops to ~1 rapidly for larger distances (fig. 4 and supplementary fig. S1, Supplementary Material online). In simulated data, only a very weak decrease in the P/D ratio was observed, which is possibly attributable to minor biases in phylogenetic reconstruction.

We also measured P/D ratios for phylogenetic distances normalized by site-specific evolutionary rates (see "Materials and Methods" section) and obtained similar plots (supplementary fig. S2, Supplementary Material online). We also asked whether the mean P/D distance is different between sites with different evolutionary rates, but saw no systematic differences (supplementary fig. S3, Supplementary Material online). These findings suggest that the P/D ratios are more sensitive to the evolutionary distance spanned by the species rather than by the individual site. Finally, we observed similar effect in trans- and nonmembrane residues. In most proteins, the effect is slightly weaker in transmembrane residues, but this difference is extremely weak (supplementary fig. S4, Supplementary Material online).

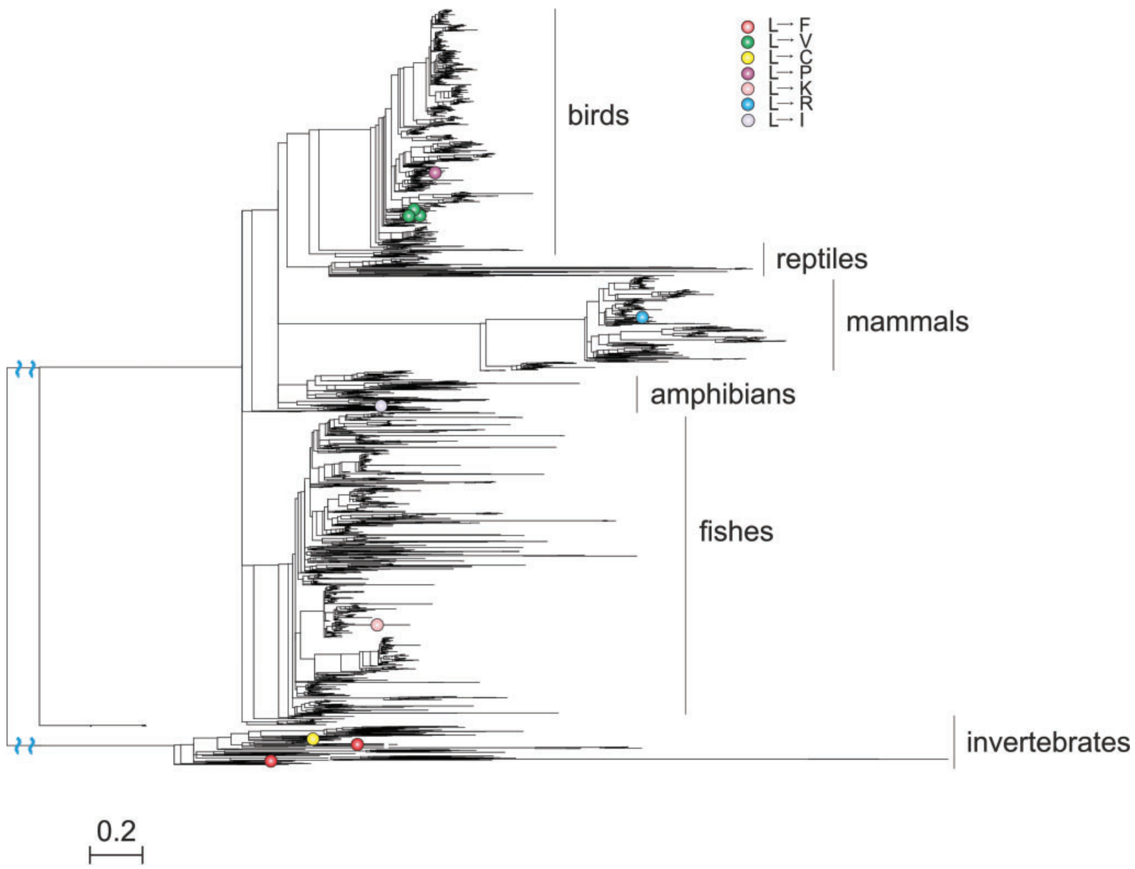

FIG. 2.—Parallel and divergent substitutions at site 202 of ATP6 (NCBI reference sequence numbering for the human sequence). The ancestral variant (L) has experienced multiple substitutions, which are scattered throughout the phylogeny. However, the two parallel L→F substitutions occur in closely related species; the same is true for the three parallel L→V substitutions. Phylogenetic distances are in numbers of amino acid substitutions per site. The branches indicated with the blue waves are shortened by 1.2 distance units.

The rate at which the P/D ratio declines with phylogenetic distance varies a lot between genes. We asked whether this difference has to do with the intrinsic rate of protein evolution, which also varied strongly between genes. We used the number of substitutions between human and *Drosophila* as the proxy for the evolutionary rate of the protein, with higher values corresponding to rapidly evolving genes; and the phylogenetic distance at which the P/D ratio (which is initially always larger than one) reaches one, as the proxy for the rate of the decline of the P/D ratio, with higher values corresponding to a slower decline. Among the 10 analyzed genes for which sequences both for human and *Drosophila* were available, the P/D decline appeared to be somewhat faster in fast-evolving genes (fig. 5), although this trend was not significant (Spearman's test: $R = 0.53$, $P = 0.11$).

Excess of Parallel Substitutions at Small Phylogenetic Distances Is Not an Artefact

Conceivably, the decline of the P/D ratio could be an artefact of erroneous phylogenetic reconstruction. Indeed, if a clade is erroneously split on a phylogeny, synapomorphies (shared derived character states) may be mistaken for parallel substitutions, and this is more likely for closely related species (Mendes et al. 2016). We tested the contribution of such artefacts by performing our analyses only on high-confidence pairs of substitutions (see "Materials and Methods" section). Our definition of this set was conservative, because branches with parallel substitutions are expected to have a reduced bootstrap support as such substitutions cause attraction of the branches where they occur in phylogenetic reconstruction. Indeed, in the Breen et al. data set, this procedure discarded the vast majority of parallel substitutions at very small phylogenetic distances, leaving us with too little data. To circumvent this, we reconstructed a five-gene, 4,350-species phylogeny (see "Materials and Methods" section for details). Similarly to the main analysis, the P/D ratio declined monotonically with distance between substitutions in this data set (fig. 6), implying that the excess of homoplasies at short phylogenetic distances is unlikely to be an artefact of phylogenetic reconstruction.

Changes in the P/D ratio with increasing phylogenetic distance imply changes in the rate of the B→A substitution

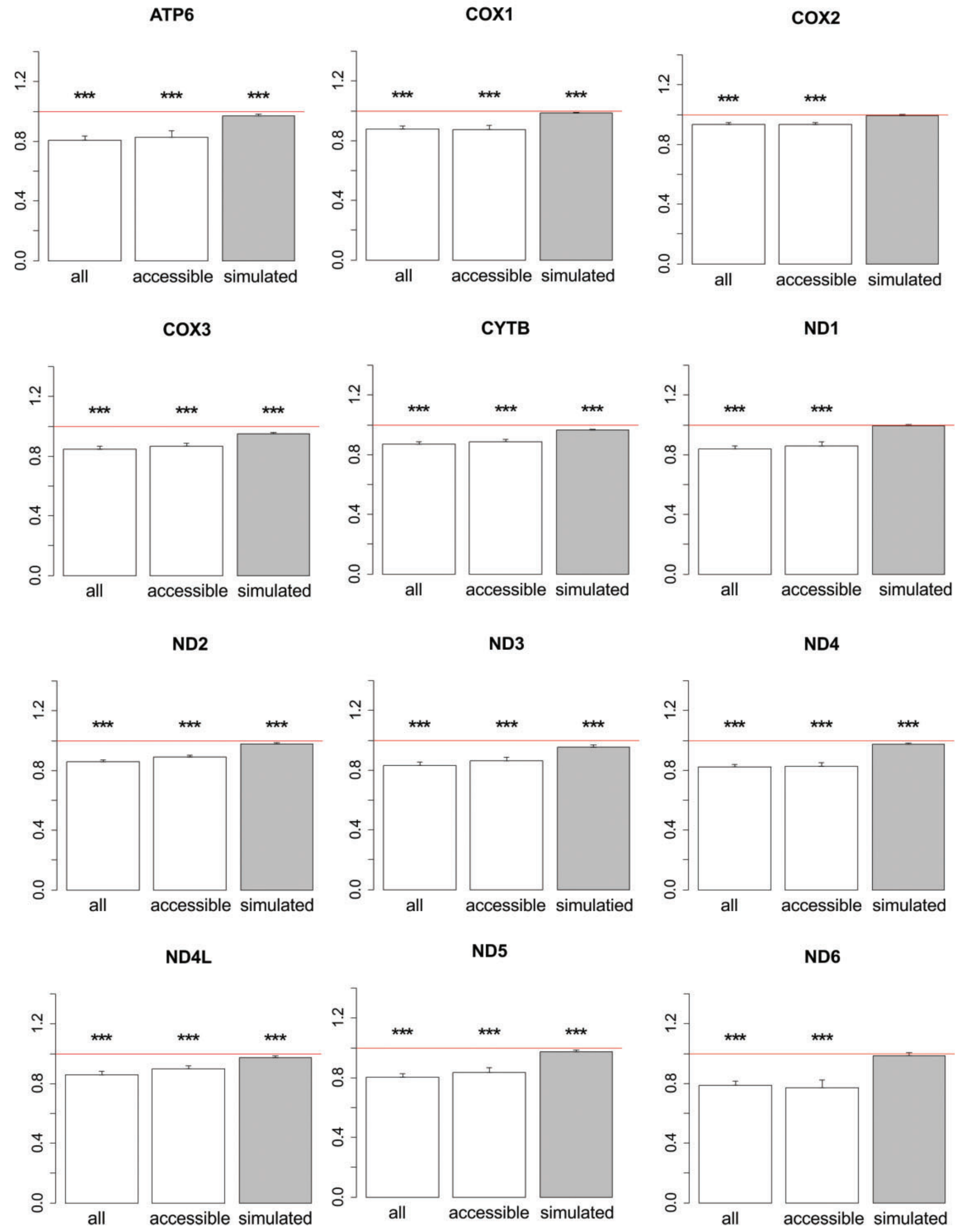

FIG. 3.—Ratios of phylogenetic distances between parallel and divergent substitutions in metazoan phylogenies. Values below 1 imply that the parallel substitutions are closer at the phylogeny to each other, compared with divergent substitutions. The bar height and the error bars represent respectively the median and the 95% CIs obtained from 1,000 bootstrap replicates, and asterisks show the significance of difference from the one-to-one ratio (red line; ***$P <$ 0.001; no asterisk, $P >$ 0.05). all, real data; accessible, real data only for substitutions from accessible amino acid pairs (see text); simulated, simulated data.

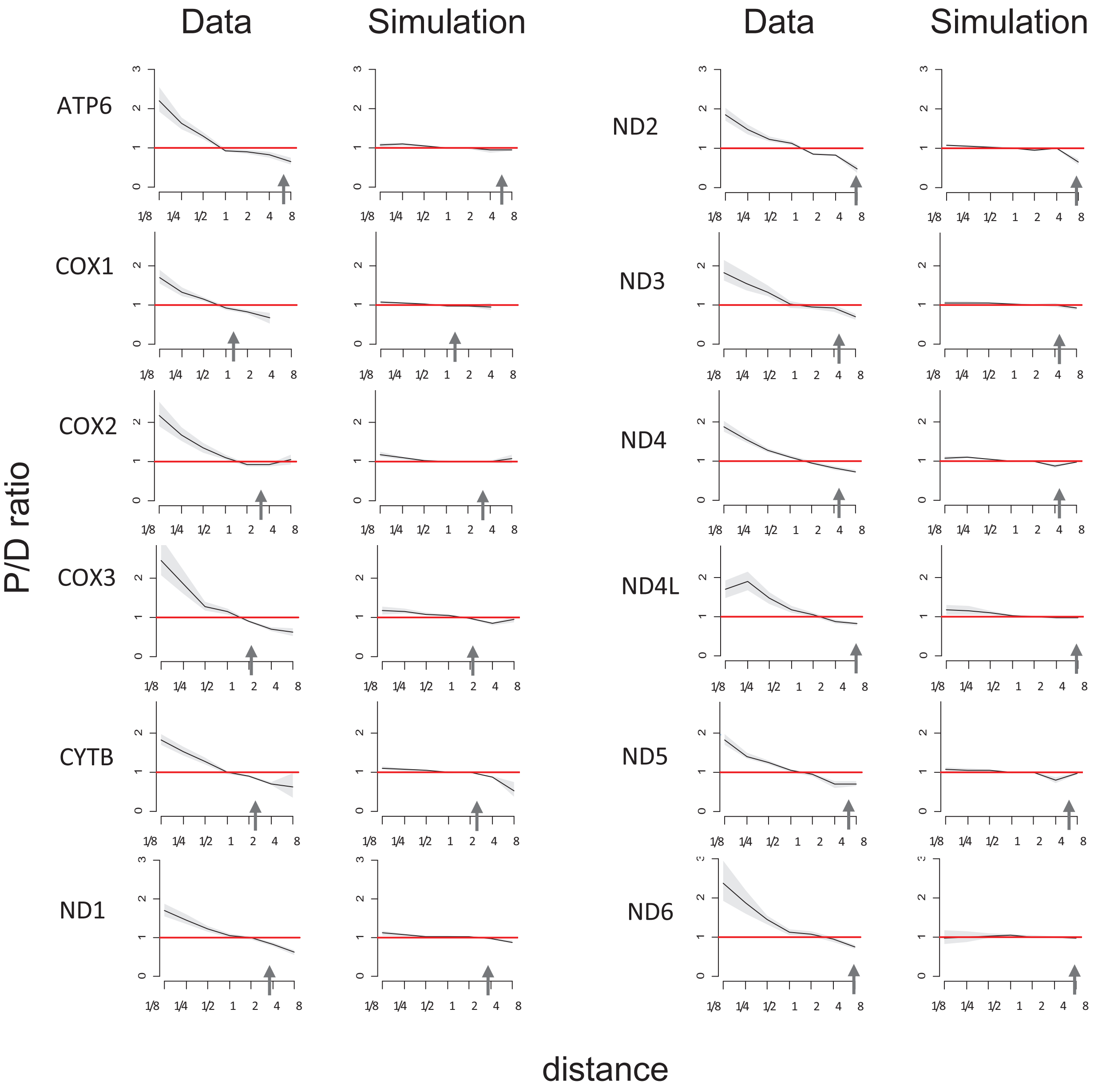

Fig. 4.—Higher fraction of parallel substitutions between closely related species. Horizontal axis, distance between branches carrying the substitutions, measured in numbers of amino acid substitutions per site (split into bins by $\log_2$(distance)). Vertical axis, P/D ratios for substitutions at this distance. Black line, mean; grey confidence band, 95% CI obtained from 1,000 bootstrapping replicates. The red line shows the expected P/D ratio of 1. Arrows represent the distance between human and *Drosophila*.

relative to other substitutions. The rate of a substitution is the product of the mutation and fixation probabilities, and changes in the substitution rate may arise from differences in either of these processes between clades.

Can the changes in substitution rates with phylogenetic distance be explained by changes in mutation rates? There can be two scenarios for this. First, the mitochondrial mutational spectra could differ between clades, potentially leading to differences in the rate at which a particular mutation occurs. If the mutation rate corresponding to a particular substitution is much higher in a particular clade, compared with the rest of the phylogeny, this may lead to an excess of homoplasies falling into this clade. To ask whether this mechanism contributes, we simulated evolution of the three major clades of the 4,350-species phylogeny using an independent substitution matrix for each clade (see "Materials and Methods" section for details) and performed our analysis for the whole phylogeny using simulated

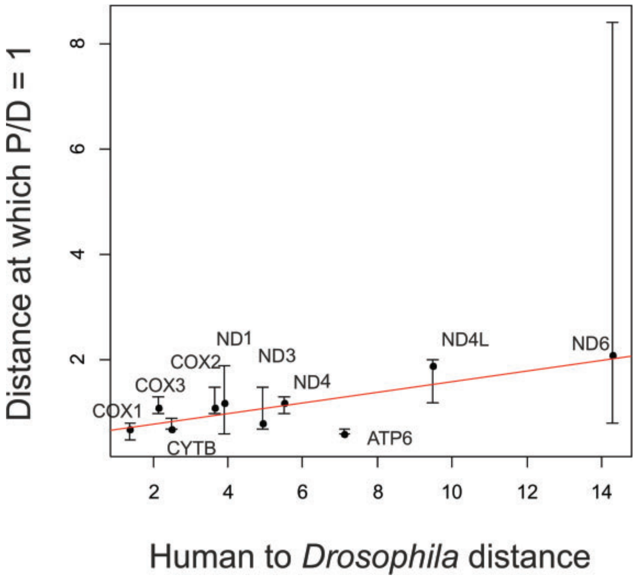

Fig. 5.—Faster decline of the P/D ratio for rapidly evolving genes. Horizontal axis, gene-specific phylogenetic distance between *Homo sapiens* and *Drosophila simulans*. Vertical axis, phylogenetic distance at which the P/D ratio reaches 1. Each dot represents one gene, and the line represents the linear trend. Only the ten genes with available *D. simulans* sequences were used. Error bars correspond to the 95% CI for the distance at which the P/D ratio reaches 1, obtained by bootstrapping sites 1,000 times.

sequences. Results for this simulation are indistinguishable from the simulation constructed with one matrix for all clades (fig. 6), implying that this mechanism is unlikely to cause the observed pattern. Moreover, most of the change in the P/D ratio occurs at very small phylogenetic distances (fig. 4), where the mutation matrices are very similar, and unlikely to contribute to our effect.

Second, even if the changes in the P/D ratio are not due to changes in the overall mutation matrix, they may still arise from differences in codon usage. This may be observed if amino acid B tends to be encoded by different codons in the two clades, and the rate of the parallel B→A substitution is higher in the clade where B is encoded by a codon that predisposes to this mutation. To test this, we defined "accessible" amino acid pairs as those (B, A) pairs where A can be reached through a single nucleotide substitution from any B codon, and considered such accessible pairs independently. In this subset, the excess of parallel changes at small phylogenetic distances was also observed (fig. 3), which means that it is not caused by the structure of the genetic code.

Finally, we analyzed the distribution of the pairs across the phylogenetic tree. Both in data and in simulations, substitutions with high (or low) P/D are clustered on the tree. The distribution of parallel (as well as divergent) substitutions over the tree is nonuniform mainly because of differences in branch lengths: longer branches carry more substitutions of all kinds. To test if there is an excess of branches with more than expected parallel substitutions, we plotted the distribution of pairs of branches by the number of parallel substitutions that had occurred in them (supplementary fig. S5, Supplementary Material online). We see no systematic differences between the distributions obtained from the data and from simulations, implying that our results are not driven by clustering of substitutions in some of the branch pairs. Moreover, in branch pairs that carried many substitution pairs in them, both branches tended to be long, leading to large distances between the two substitutions in a pair; while most of the effect is observed at small distances (fig. 4).

For three of the genes, we also randomly picked and examined visually 20 pairs of parallel and 20 pairs of divergent substitutions among those with distances <0.1 between them. Parallel and divergent pairs in data and in simulations were clustered in the same parts of the tree, demonstrating that this effect is likely to be determined by the shape of the tree. Specifically, they were clustered in the regions of the tree with many short branches, which are the only ones contributing to the phylogenetically close pairs of substitutions in figure 4.

Discussion

The rate at which a specific substitution occurs is a monotonic function of fitness differences between the descendant and the ancestral variants, and changes in the substitution rates in the course of evolution indicate that these fitness differences, and thus the SPFL, change. Obtaining the entire substitution matrix for individual amino acid sites is problematic even given hundreds of species (Rodrigue 2013); still, the changes in the relative substitution rates can be inferred using some summary statistics. The change in the extent of parallelism in the course of evolution is one such convenient statistic: as a substitution becomes more deleterious, its rate decreases, and it becomes more sparsely distributed on the phylogeny.

We observe that parallel substitutions in the evolution of metazoan mitochondrial proteins are phylogenetically clustered; that is, that such substitutions are more likely to occur in the phylogenetic vicinity of each other, compared with divergent substitutions. As a result, the distance between two parallel substitutions on the phylogenetic tree is, on average, ~20% lower than the distance between divergent substitutions, or than expected if their rate was constant across the tree (fig. 3). We show that these results are unlikely to be artefacts of phylogenetic reconstruction or of pooling together sites and amino acids with different properties. Our results cannot be explained by a simple covarion model, in which a site alternates between neutral and constrained (Fitch and Markowitz 1970; Fitch 1971), as the changes we observe are not associated with changes in the overall substitution rates. For the same reason, they also cannot be

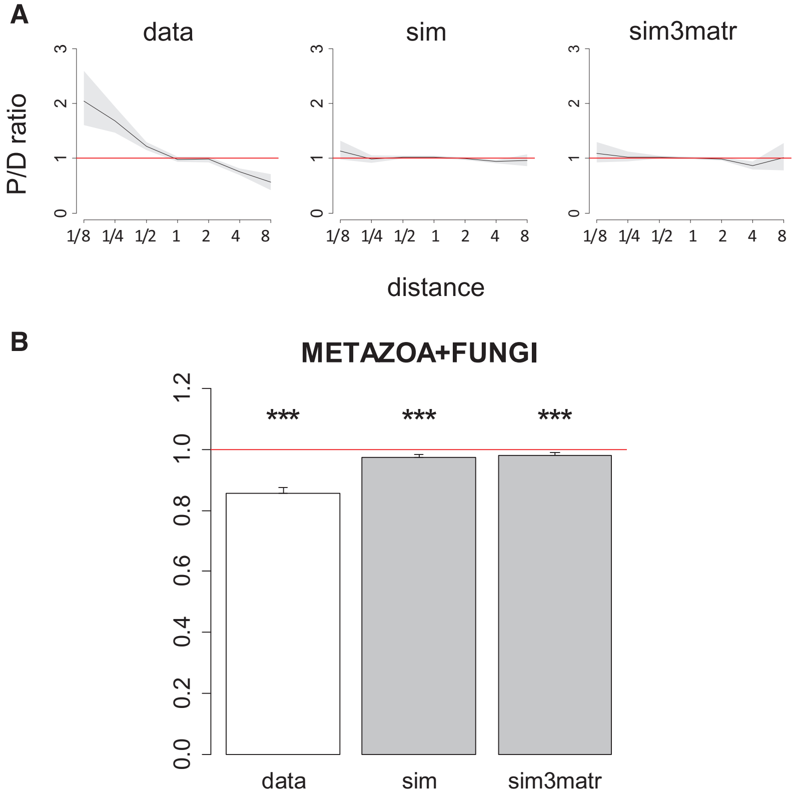

Fig. 6.—Higher fraction of parallel substitutions between closely related species *(A)* and ratios of phylogenetic distances between parallel and divergent substitutions *(B)* in the 4,350-species phylogeny, for high-confidence pairs of substitutions. sim, simulation; sim3matr, simulation with independent substitution matrices for each clade (see text).

explained by a broader class of heterotachy models in which the overall rates of evolution of a site vary with time (Lopez et al. 2002; Yang and Nielsen 2002; Murrell et al. 2012), but require heteropecilly (Tamuri et al. 2009; Roure and Philippe 2011), that is, variation with time of rates of individual substitutions. We show that these differences are unlikely to be caused by systematic gene- or genome-wide differences in substitution matrices between clades, which may result from differences in mutation patterns, selection for nucleotide or amino acid usage, or gene conversion.

Instead, they most likely reflect changes in single-position fitness landscapes (Mustonen and Lässig 2007; Bazykin 2015) that accumulate in the course of evolution. Indeed, site-specific differences in the rate of a substitution leading to a particular amino acid imply that the relative fitness of this amino acid relative to other amino acids at this site changes with time. Decrease in this frequency with phylogenetic distance may be caused by a decline in the fitness of this allele, and/or by an increase in the fitness of other alleles; it is hard to distinguish between these possibilities with the available data, although both factors likely play a role (Naumenko et al. 2012). Our single-site approach also prevents us from distinguishing between the possible causes of the SPFL changes: evolution of other sites (of the same or other proteins) involved in epistatic interactions with the focal site, environmental changes, or perhaps a combination of both.

The numbers of substitutions to the same or to another amino acid, that is, convergent and divergent substitutions at different phylogenetic distances, have been used previously to characterize evolution. In ancient proteins, the rate of convergence monotonically decreases with phylogenetic distance, and half of the reversals were estimated to become forbidden after 10% protein divergence (Povolotskaya and Kondrashov 2010). The ratio of the rates of convergent and divergent substitutions drops by more than twofold with phylogenetic distance within vertebrates (Goldstein et al. 2015). Similarly, the rate of convergence decreases with phylogenetic distance in mammals and fruit flies (Zou and Zhang 2015). Our analysis spans larger phylogenetic distances than that of Goldstein et al. (2015); still, most of the observed effect is local (fig. 4). On the basis of the data from different sources, and assuming a two-state fitness space such that each amino acid variant at a particular amino acid site can be either "prohibited" or "permitted", the rate at which a particular variant switches between these two states has been estimated as ~5 such switches per unit time required for a single amino acid substitution to occur at this site (Usmanova et al. 2015). In our data, the rate of SPFL change appears to vary widely between proteins, as the time necessary for the P/D ratio to reach 1 varies between 0.6 for ATP6 and 2.1 for ND6. It also is strongly dependent on the size and the shape of the phylogeny. Still, in our data, the rate of SPFL changes has roughly the same scale as the rate of amino acid evolution (fig. 4), which is consistent with the results of (Mustonen and Lässig 2007) who have shown that fluctuations in SPFLs of *Drosophila* proteins occur with rates comparable with neutral mutation rates.

In summary, our results allow to suggest that the fitness landscapes of amino acid sites of mitochondrial proteins change with time, supporting previous conjectures that such landscapes are dynamic in this data set (Breen et al. 2012, 2013). Whether these changes are driven by changes in the intra-protein or inter-protein genomic context between species, or by environmental changes, remains a subject for future research.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

This work was supported by the Russian Science Foundation grant no.14-50-00150. We thank Shamil Sunyaev, Alexey Kondrashov, Alexander Favorov, Andrey Mironov, Dmitry Pervouchine, Vladimir Seplyarskiy, Sergey Naumenko, Elena Nabieva, Ivan Cytovich, David McCandlish, Joshua Plotkin, Richard Goldstein, and Dinara Usmanova for valuable comments.

Literature Cited

Bazykin GA. 2015. Changing preferences: deformation of single position amino acid fitness landscapes and evolution of proteins. Biol Lett. 11(10):1–7.

Breen MS, Kemena C, Vlasov PK, Notredame C, Kondrashov FA. 2012. Epistasis as the primary factor in molecular evolution. Nature 490:535–538.

Breen MS, Kemena C, Vlasov PK, Notredame C, Kondrashov FA. 2013. Reply to McCandlish et al. Nature 497(7451):E1–E2; discussion E2–E3.

Fitch WM. 1971. Rate of change of concomitantly variable codons. J Mol. Evol. 1:84–96.

Fitch WM, Markowitz E. 1970. An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. Biochem Genet. 4:579–593.

Goldstein RA, Pollard ST, Shah SD, Pollock DD. 2015. Nonadaptive amino acid convergence rates decrease over time. Mol Biol Evol. 32:1373–1381.

Letunic I, Bork P. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinforma Oxf Engl 23:127–128.

Lopez P, Casane D, Philippe H. 2002. Heterotachy, an important process of protein evolution. Mol Biol Evol. 19:1–7.

McCandlish DM, Rajon E, Shah P, Ding Y, Plotkin JB. 2013. The role of epistasis in protein evolution. Nature 497(7451):E1–E2; discussion E2–E3.

Mendes FK, Hahn Y, Hahn MW. 2016. Gene tree discordance can generate patterns of diminishing convergence over time. Mol Biol Evol. 33(12): 3299–3307.

Murrell B, et al. 2012. Detecting individual sites subject to episodic diversifying selection. PLoS Genet. 8:e1002764.

Mustonen V, Lässig M. 2007. Adaptations to fluctuating selection in Drosophila. Proc Natl Acad Sci U S A. 104:2277–2282.

Naumenko SA, Kondrashov AS, Bazykin GA. 2012. Fitness conferred by replaced amino acids declines with time. Biol Lett. 8:825–828.

Povolotskaya IS, Kondrashov FA. 2010. Sequence space and the ongoing expansion of the protein universe. Nature 465:922–926.

Rodrigue N. 2013. On the statistical interpretation of site-specific variables in phylogeny-based substitution models. Genetics 193:557–564.

Rogozin IB, Thomson K, Csürös M, Carmel L, Koonin EV. 2008. Homoplasy in genome-wide analysis of rare amino acid replacements: the molecular-evolutionary basis for Vavilov's law of homologous series. Biol Direct. 3:7.

Roure B, Philippe H. 2011. Site-specific time heterogeneity of the substitution process and its impact on phylogenetic inference. BMC Evol Biol. 11:17.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinforma Oxf Engl 30:1312–1313.

Tamuri AU, Dos Reis M, Hay AJ, Goldstein RA. 2009. Identifying changes in selective constraints: host shifts in influenza. PLoS Comput Biol. 5:e1000564.

Usmanova DR, Ferretti L, Povolotskaya IS, Vlasov PK, Kondrashov FA. 2015. A model of substitution trajectories in sequence space and long-term protein evolution. Mol Biol Evol. 32:542–554.

Wiegand T, A. Moloney K. 2004. Rings, circles, and null-models for point pattern analysis in ecology. Oikos 104:209–229.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci CABIOS 13:555–556.

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

Zou Z, Zhang J. 2015. Are Convergent and Parallel Amino Acid Substitutions in Protein Evolution More Prevalent Than Neutral Expectations?. Mol Biol Evol. 32:2085–2096.

Associate editor: David Bryant