# Compositionally constrained sites drive long branch attraction

Lenard L Szantho, Nicolas Lartillot, Gergely Szöllősi, Dominik Schrempf

# Compositionally constrained sites drive long branch attraction

Lénárd L. Szánthó[1,3,4], Nicolas Lartillot[2], Gergely J. Szöllősi[1,3,4,*,+], and Dominik Schrempf[1,*,+]

[1]Dept. Biological Physics, Eötvös University, Pázmány P. stny. 1A., H-1117 Budapest, Hungary

[2]Laboratoire de Biométrie et Biologie Evolutive UMR 5558, CNRS, Université de Lyon, Villeurbanne, France

[3]ELTE-MTA "Lendület" Evolutionary Genomics Research Group, Pázmány P. stny. 1A., H-1117 Budapest, Hungary

[4]Institute of Evolution, Centre for Ecological Research, Konkoly-Thege M. u 29-33, Budapest, Hungary

[*]Corresponding authors: ssolo@elte.hu and dominik.schrempf@gmail.com

[+]Equal contribution

February 23, 2022

**Abstract**

Accurate phylogenies are fundamental to our understanding of the pattern and process of evolution. Yet, phylogenies at deep evolutionary timescales, with correspondingly long branches, have been fraught with controversy resulting from conflicting estimates from models with varying complexity and goodness of fit. Analyses of historical as well as current empirical datasets, such as alignments including Microsporidia, Nematoda or Platyhelminthes, have demonstrated that inadequate modeling of across-site compositional heterogeneity, which is the result of biochemical constraints that lead to varying patterns of accepted amino acid along sequences, can lead to erroneous topologies that are strongly supported. Unfortunately, models that adequately account for across-site compositional heterogeneity remain computationally challenging or intractable for an increasing fraction of contemporary datasets. Here, we introduce "compositional constraint analysis", a method to investi-

gate the effect of site-specific amino acid diversity on phylogenetic inference, and show that more constrained sites with lower diversity and less constrained sites with higher diversity exhibit ostensibly conflicting signal for models ignoring across-site compositional heterogeneity. We demonstrate that more complex models accounting for across-site compositional heterogeneity can ameliorate this bias. We present CAT-PMSF, a pipeline for diagnosing and resolving phylogenetic bias resulting from inadequate modeling of across-site compositional heterogeneity based on the CAT model. Our analyses indicate that CAT-PMSF is unbiased. We suggest using CAT-PMSF when convergence of the CAT model cannot be assured. We find evidence that compositional constrained sites are driving long branch attraction in two metazoan datasets and recover evidence for Porifera as the sister group to all other animals.

# 1 Introduction

Understanding the biological foundations of contemporary life on Earth requires detailed knowledge of evolutionary history. The history of speciation events informs us about the appearance of advantageous innovations and the loss of dispensable traits in a continuously changing environment. Consequently, development of phylogenetic models inferring the history of speciation events has continued at an impressive pace during the past decades.

Phylogenetic models do not reflect the full complexity of evolution but inevitably present a simplified picture of the generating processes underlying sequence evolution. Unfortunately, overly-simplistic models can lead to model misspecification and long branch attraction (LBA; e.g., Felsenstein, 1978; Hendy and Penny, 1989; Zharkikh and Li, 1993; Tateno et al., 1994; Bergsten, 2005; Brinkmann et al., 2005; Philippe et al., 2011b). LBA is a bias in the inferred topology, as a result of which long branches are more likely to branch together than with short ones, independent of the true evolutionary history. Substitution models (Jukes and Cantor, 1969) account for the possibility of multiple substitutions per site, and reduce LBA (Felsenstein, 1973; but see Farris, 1999) compared to models based on maximum parsimony. Probabilistic substitution models describe the evolution of a site as a series of transitions between states. Every possible transition occurs at a specific

rate which is the expected number of transitions from one state to another per unit time. Standard substitution models perform well on most datasets (Ripplinger and Sullivan, 2010).

However, LBA can be an issue even when using probabilistic substitution models. For example, if variation of evolutionary rate between sites (across-site rate heterogeneity) is ignored. Models accounting for across-site rate heterogeneity have been proposed early on (Yang, 1993), and have been shown to ameliorate LBA in some cases (Kuhner and Felsenstein, 1994; Philippe et al., 2011b). However, across site variation is not restricted to rates. Also, the relative abundance of nucleotide or amino acid characters is variable. For example, sites buried deeply in folded proteins tend to be more constrained than sites that end up on the surface (Koshi and Goldstein, 1995; Yeh et al., 2014; Jimenez et al., 2018), and depending on the protein structure may exhibit stronger preferences in hydrophobicity than other sites. Models ignoring across-site compositional heterogeneity are prone to LBA because they underestimate the probability of convergent substitutions at compositionally constrained sites. In particular, the probability of independent substitutions to the same state depends on the number of acceptable amino-acids, which differs across sites. Models ignoring across-site compositional heterogeneity pool all sites, and ignore the variation of the evolutionary process across sites. Indeed, analyses of a series of datasets exhibiting previously contentious evolutionary relationships provide evidence that ignoring across-site compositional heterogeneity can lead to LBA (Phillips et al., 2004; Brinkmann et al., 2005; Philippe et al., 2005b,a; Delsuc et al., 2006; Lartillot et al., 2007; Philippe et al., 2009, 2011a; Brown et al., 2013; Ryan et al., 2013; Cannon et al., 2016; Simion et al., 2017).

Thus, there is accumulating evidence that accounting for across-site heterogeneities is key to an accurate reconstruction of deep evolutionary relationships. The classic approach to modeling such heterogeneities in the phylogenetic inference process are mixture models that combine substitution models specifically tailored to the evolutionary processes observed in the data. In order to model across-site rate heterogeneity, we use a mixture of substitution models with the same relative but different absolute substitution rates (e.g., Yang, 1993; Kalyaanamoorthy et al., 2017).

Modeling across-site compositional heterogeneity, however, requires construct-

ing a model describing the evolution of sites subject to different compositional constraints. To do so, for time-reversible substitution models, we can leverage the separation of substitution rates into the product of (1) symmetric exchangeabilities describing differences in rates of exchange between pairs of states and (2) stationary frequencies of the target states, or figuratively, the probability of sampling the target states after waiting for a long time. Any time-reversible substitution model is fully specified by a set of symmetric exchangeabilities and the set of stationary frequencies (sometimes also referred to as a profile or a stationary distribution). Assuming that biochemical constraints primarily affect site-specific amino acid preferences in the long term, across-site compositional heterogeneity can be accounted for by composing a number of substitution models sharing a single set of exchangeabilities but differing in their stationary distributions (distribution mixture models; Quang et al., 2008; Schrempf et al., 2020)

We distinguish between general distribution mixture models estimated from curated training databases, and distribution mixture models directly estimated from the datasets at hand. For example, Wang et al. (2008) directly estimate mixture model components using principal component analysis. Susko et al. (2018) use a composite likelihood approach and additional methods such as taxon weighing. In contrast, the rationale behind providing and using general mixture models is the assumption that the underlying evolutionary processes share universal features. Quang et al. (2008) use the expectation maximization algorithm to infer general mixture models consisting of 10, 20, ..., 60 components (C10, C20, ..., C60, collectively CXX models). Schrempf et al. (2020) used a clustering approach together with different compositional transformations to provide a set of general mixture models, termed universal distribution mixtures (UDM), with the number of components ranging from four up to several thousand. They also provide the clustering method EDCluster to infer dataset specific distribution mixture models.

Finite mixtures can be used in the context of maximum likelihood (ML) phylogenetic inference. However, statistical analyses of model fit and investigation of known cases of LBA indicate that a large number of components are necessary for robustness against LBA (Schrempf et al., 2020), which is computationally intensive and memory intense.

Bayesian approaches can more easily accommodate richer mixtures. In partic-

ular, nonparametric Bayesian methods do not require explicit specification of the number of mixture components nor their stationary distributions. In particular, the CAT model (Lartillot and Philippe, 2004) uses a Dirichlet process prior to approximate an arbitrary mixture of stationary distributions across sites. The CAT model was shown to be better fitting and less prone to LBA than site-homogeneous models for a series of classic datasets including Nematoda and Platyhelminthes (Lartillot et al., 2007) as well as in phylogenomic analyses of the tree of life (Williams et al., 2020). The impediment of nonparametric Bayesian methods and specifically, the CAT model, is that it compounds two computationally challenging, but separately tractable problems, the non-parametric inference of the underlying distribution across sites and the exploration of tree space. The composition of these two problems is challenging and can lead to convergence problems.

In all cases, however, mixture modeling approaches accounting for across-site compositional heterogeneity are complex and require considerable computational resources (e.g., Whelan and Halanych, 2016). In order to reduce computational cost, Wang et al. (2018) proposed a two-step approximation. First, site-specific stationary distributions are estimated using a reference mixture model and a fixed guide tree. Second, the tree is reconstructed using the fixed stationary distributions obtained in the first step. Thereby, run time is reduced while robustness against LBA is improved compared to using the reference mixture model alone. In particular, the site-specific stationary distributions are set to the posterior mean site frequencies (PMSF) of the reference mixture model. As a result, the phylogenetic accuracy of the PMSF approach is inherently limited by how well the reference mixture model captures across-site compositional heterogeneity. There is no reason to restrict the use of the PMSF approach to empirical mixture models: any random-effect model meant to account for pattern heterogeneity could in principle be used here as a reference mixture model for computing the posterior means of the site-specific stationary distributions.

In this work we follow a multistep procedure similar to the PMSF model and address two points: First, on the computational side, we extend the PMSF approach by using the CAT model instead of an empirical profile mixture model as the reference model for computing the profiles. Importantly, the CAT model is used under a fixed tree topology. Thus, the problem of simultaneous inference of both,

the site-specific stationary distributions and the tree including the tree topology is reduced to a search of site-specific stationary distributions, and tree branch lengths only. We termed our approach CAT-PMSF.

Second, we investigate the contribution of sites with different degrees of compositional constraints to LBA. In particular, we test to what extent more severely constrained sites exhibit bias towards LBA trees under models that do not adequately account for across-site compositional heterogeneity. We employ "compositional constraint" analysis, which examines phylogenetic signal for alternative topologies as a function of per site compositional diversity measured by the effective number of amino acids.

Examining simulated alignments as well as classic empirical datasets including Microsporidia (Brinkmann et al., 2005), Nematoda, and Platyhelminthes (Philippe et al., 2005a), we find conflicting phylogenetic signal across sites with different degrees of compositional constraints. Based on these results we apply compositional constraint analysis to two recent datasets (Ryan et al., 2013; Simion et al., 2017) aiming to resolve the early diversification of animal lineages.

# 2 Results

In the following, we use the terms *site-homogeneous* and *site-heterogeneous* when referring to models ignoring and accounting for across-site compositional heterogeneity, respectively. Further, we use the term *tree* to denote a directed acyclic graph with node labels and branch lengths, in which exactly one branch connects any two nodes. We use the term *topology* to denote a tree without information on branch length. We specify an evolutionary model with exchangeabilities EX, and across-site compositional heterogeneity model ASCH as EX+ASCH. All discussed evolutionary models used for simulations as well as inferences implicitly use discrete gamma rate heterogeneity with four components. We add a flag +PMSF to denote usage of the posterior mean site frequency model (Wang et al., 2018).

In brief, CAT-PMSF comprises three steps: (1) Estimating a guide topology using a site-homogeneous model, (2) Estimating site-specific stationary distributions with the CAT model in PhyloBayes (Lartillot and Philippe, 2004) using the guide topology, and (3) phylogenetic inference in a ML framework with a distri-

bution mixture model sharing one set of exchangeabilities, and using the obtained site-specific stationary distributions (see Methods).

**Simulation study.** We assessed and compared the accuracy of CAT-PMSF with other site-homogeneous and site-heterogeneous models. To this end, we simulated amino acid sequence alignments with a length of 10 000 sites along Felsenstein-type quartet trees (insets of Figure 1; Felsenstein, 1978). We used uniform exchangeabilities (Poisson; Felsenstein, 1973) and an across-site compositional heterogeneity model with site-specific stationary distributions based on a UDM model (see Methods; Schrempf et al., 2020). We set the branch length of the short branch $q$ to 0.1, and varied the length of the long branch $p$ between 0.1 and 2.0.

The true topology (a Felsenstein-type quartet) was not recovered with site-homogeneous models when $p \geq 0.8$ (Figures S7, S8, and Tables S1, S2). Figure 1 shows the results of the compositional constraint analysis for different values of $p$, contrasting the site-wise log-likelihood differences between the maximum likelihood trees constrained to the genuine Felsenstein-type and incorrect Farris-type topologies prone to LBA (see Methods) for $p = 0.2$, 0.8 and 1.2. Figures S7, and S8 show results for other values of $p$. We binned sites according to their effective number of amino acids ($K_{\text{eff}}$, see Methods). Lower values of $K_{\text{eff}}$ correspond to sites under stronger compositional constraint. "Compositional constraint analysis" compares per site phylogenetic signal for two alternative topologies as a function of compositional constraint (i.e., different values of $K_{\text{eff}}$). Here, positive log-likelihood differences indicate support for the true (Felsenstein-type quartet) topology. Conversely, negative values indicate support for the LBA (Farris-type quartet) topology. In the absence of model misspecification we expect consistent phylogenetic signal across sites, and independent of the true value of $K_{\text{eff}}$.

At odds with this expectation, site-homogeneous evolutionary models exhibit conflicting phylogenetic signal between sites with low and high $K_{\text{eff}}$ values (Figures 1, S8 and S7). For the site-homogeneous evolutionary models, more constrained sites with a low value of $K_{\text{eff}}$ exhibit bias towards the incorrect (Farris-type) topology. For $p \geq 0.8$, the bias outweighs the correct signal of less constrained sites with high values of $K_{\text{eff}}$, and the incorrect (Farris-type) topology has higher support than the true (Felsenstein-type) topology. In contrast, the site-heteroge-

neous LG+C60+PMSF model as well as the CAT-PMSF method show consistent support for the true topology across all sites and values of $p$.

To ascertain the statistical significance of the compositional constraint analysis we calculated Pearson correlation coefficients and associated p-values between the log-likelihood differences and the site-specific $K_{\text{eff}}$ values (Table S3). Site-homogeneous models exhibit large and significant correlation for $p \geq 0.8$, whereas the log-likelihood differences and $K_{\text{eff}}$ values of site-heterogeneous models are not correlated.

Approximately unbiased (AU) tests (Shimodaira, 2002) of maximum likelihood trees inferred by the GTR+CAT-PMSF model constrained to the two alternative topologies reject the (Farris-type) LBA topology in favor of the true topology for $p < 1.3$ (Table S5). AU tests of the Poisson+CAT-PMSF, and LG+CAT-PMSF models show similar results (Tables S6, and S7).

Finally, we note that site-heterogeneous models with LG exchangeabilities perform well, although the genuine exchangeabilities are uniformly distributed. CAT-PMSF is even more accurate when using the true topology or the true site-specific stationary distributions (Figures S7 and S8).

**Applications to empirical data.** Similar to the simulation study above, for empirical alignments the site-specific stationary distributions obtained in Step 2 of the CAT-PMSF pipeline can be used to quantify the strength of compositional constrains (i.e., to estimate $K_{\text{eff}}$ values per site) and perform compositional constraint analysis. Figure 2 shows results for three datasets exhibiting classic LBA artifacts when we use site-homogeneous models for inference: The placement of Platyhelminthes and Nematoda (Philippe et al., 2005a), as well as the placement of Microsporidia (Brinkmann et al., 2005; Lartillot et al., 2007).

For site-homogeneous models, the site-specific log-likelihood differences between the maximum likelihood trees constrained to the two competing topologies (insets of Figure 2; top vs. bottom) show conflicting phylogenetic signal as a function of the strength of compositional constraint. The bias towards the topologies exhibiting LBA artifacts of more constrained sites outweighs the signal of less constrained sites in all three datasets.

The site-heterogeneous LG+C60+PMSF model shows reduced, but still appar-

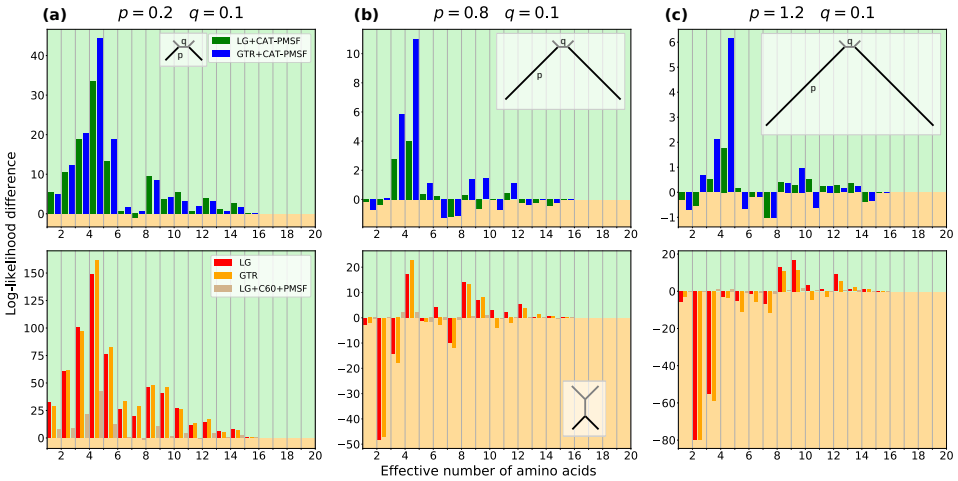


Figure 1: **Highly constrained sites drive long branch attraction artifacts in the Felsenstein zone.** We simulated amino acid alignments with 10 000 sites exhibiting across-site compositional heterogeneity (Schrempf et al., 2020) along Felsenstein-type trees (insets in top row; Felsenstein, 1978) with different branch lengths $q = 0.1$, and $p = 0.3$, 0.8, and 1.2 from (a) to (c). We performed analyses with CAT-PMSF, the LG (Le and Gascuel, 2008) and the GTR (Tavaré, 1986) models constrained to the correct topology as well as the incorrect Farris-type topology (insets in bottom row; Farris, 1999) with IQ-TREE 2 (Minh et al., 2020). The site-wise log-likelihood differences $\Delta \log L$ between the maximum likelihood trees of the two competing topologies binned according to their effective number of amino acids $K_{\text{eff}}$ are shown. A positive value (green background) indicates support for the Felsenstein-type topology, a negative value (orange background) indicates support for the Farris-type topology prone to long branch attraction. We do not expect a uniform distribution across the bins, because they are of different size. The LG, and GTR models incorrectly infer Farris-type trees if $p \geq 0.8$.

ent conflict compared to site-homogeneous models and the LG+C10+PMSF model (Figure S9). For Platyhelminthes, the bias is strong enough to that the total likelihood across all sites is higher for the LBA topology, while for the datasets involving Nematoda and Microsporidia, the LG+C10+PMSF model recover the correct topology, albeit with reduced support. In general, the results of the LG+C10+PMSF and LG+C60+PMSF models are consistent with the observation (Schrempf et al., 2020) that increasing the number of components decreases the bias introduced by more constrained sites. Pearson correlation coefficients are greater for site-homogeneous models than for models LG+C10+PMSF and LG+C60+PMSF models, but significant for each of these (Table S4).

In contrast, CAT-PMSF exhibits consistent signal towards the correct topologies across all sites and datasets with no significant correlation between log-likelihood difference and site-specific $K_{\text{eff}}$ value (Table S4). The maximum likelihood trees inferred by CAT-PMSF are consistent with the accepted phylogenetic relationships and AU tests confirm the rejection of trees with LBA topologies (Tables S8-S10).

**The phylogenetic position of Ctenophora.** Finally, we used CAT-PMSF on two metazoan datasets (Ryan et al., 2013; Simion et al., 2017) to investigate early evolutionary relationships on the animal tree of life. It is currently a matter of intense debate whether sponges (Porifera) or comb jellies (Ctenophora) are the sister group to all other animals (e.g., Kapli and Telford, 2020; Li et al., 2020). We refer to the competing hypotheses as Porifera-sister and Ctenophora-sister, respectively.

Compositional constraint analysis under site-homogeneous models, as well as combinations of PMSF and site-heterogeneous mixture models with 20 and 60 components exhibit patterns of conflicting phylogenetic signal for sites with different degrees of compositional constraints for both the alignments from Simion et al. (2017) and Ryan et al. (2013). The conflicting signal is consistent with LBA driving the placement of Ctenophora as the first animal group to emerge (cf. Table S4).

Under site-homogeneous models, sites with $K_{\text{eff}}$ values up to approximately $10 - 12$ exhibit strong preference for Ctenophora-sister (Figures 3 and S12). Sites with higher $K_{\text{eff}}$ values, however, switch their preference toward Porifera-sister. In contrast, under the CAT-PMSF models the Simion et al. (2017) dataset exhibits

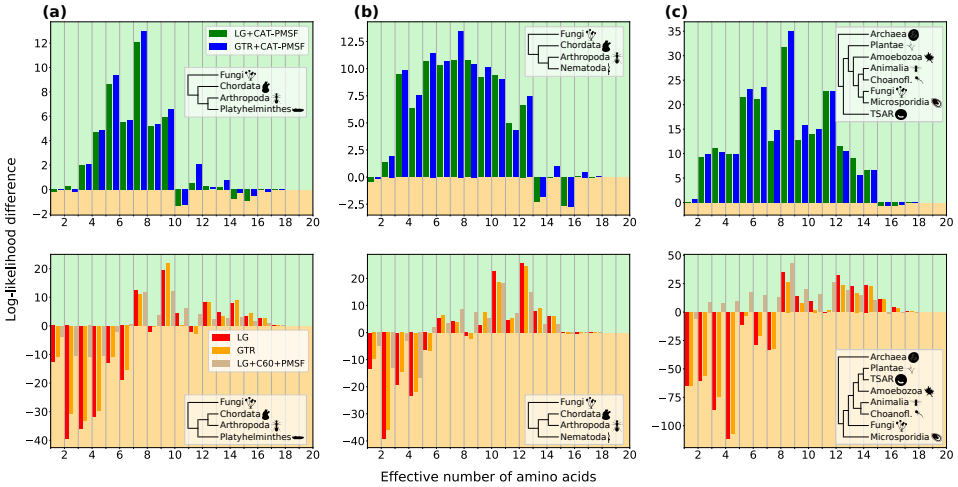


Figure 2: **Highly constrained sites explain classic examples of long branch attraction.** We analyzed three empirical datasets including (a) Platyhelminthes and (b) Nematoda (Philippe et al., 2005a), and (c) Microsporidia (Brinkmann et al., 2005). We performed analyses with CAT-PMSF, the LG (Le and Gascuel, 2008), the GTR (Tavaré, 1986), and the LG+C60+PMSF (Quang et al., 2008; Wang et al., 2018) models constrained to either one of two competing topologies (insets in top versus bottom rows) with IQ-TREE 2 (Minh et al., 2020). The site-specific log-likelihood differences $\Delta \log L$ between the maximum likelihood trees of the two competing topologies binned according to their effective number of amino acids $K_{\text{eff}}$ estimated by PhyloBayes (Lartillot and Philippe, 2004) are shown. A positive value (green background) indicates support for the now accepted topology, a negative value (orange background) indicates support for the topology prone to long branch attraction. We do not expect a uniform distribution across the bins, because they are of different size. Site-homogeneous models infer the wrong topology for all three datasets.

consistent phylogenetic signal (Table S3) favoring a Porifera-sister topology and rejecting the Ctenophora-sister hypothesis (AU test p-values between $3.1 \times 10^{-4}$ and $7.7 \times 10^{-4}$; Table S11) with the closest out group, Choanoflagellatea (Figure S4a, Figure S10). For the Ryan et al. (2013) alignment, the total log-likelihood difference of CAT-PMSF between the two hypotheses is marginal at only 0.8, suggesting a lack of resolution in this dataset. None of the models we investigated exhibit phylogenetic consistent signal across sites with different degrees of compositional constraints.

# 3 Discussion

We introduce CAT-PMSF, a method for phylogenetic inference from alignments exhibiting across-site compositional heterogeneity. The CAT-PMSF pipeline uses the site-specific amino-acid preferences estimated by a non-parametric Bayesian approach in the context of a downstream maximum likelihood analysis. Doing so combines the benefits of both approaches: a more accurate inference of the patterns across sites with a computationally more efficient and more reproducible inference of the tree topology. In addition to phylogenetic inference the CAT-PMSF pipeline can also be used to investigate the consistency of phylogenetic signal for sites under different degrees of compositional constraints. Using compositional constraint analysis, we elucidate inconsistent signal when using site-homogeneous and site-heterogeneous mixture models for phylogenetic inference from simulated data exhibiting across-site compositional heterogeneity as well as empirical data.

**Simulation study and empirical results** In the simulation study (Figure 1), site-homogeneous models favored the LBA (Farris-type) topology when the length $p$ of the terminal branches was long enough. By separating the contribution of sites as a function of compositional constraint, we demonstrated that sites under strong compositional constraints drive the bias leading to the LBA.

The threshold $K_{\text{eff}}$ value separating sites supporting the correct topology and sites supporting the LBA topology depended on the length $p$ of the terminal branches: The longer the terminal branches, the higher the threshold $K_{\text{eff}}$ value. We expect this observation holds more generally.

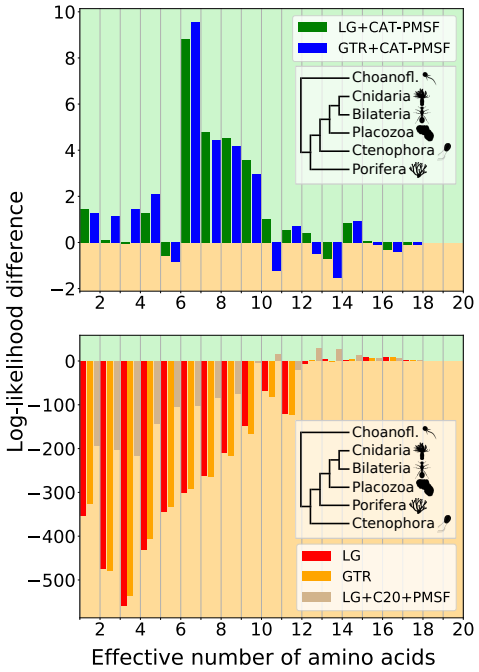


Figure 3: **CAT-PMSF shows consistent signal for Porifera as the sister group to all other animals**. We performed analyses with CAT-PMSF, the LG (Le and Gascuel, 2008), the GTR (Tavaré, 1986), and the LG+C20+PMSF (Quang et al., 2008; Wang et al., 2018) models constrained to either one of two competing topologies (insets in top versus bottom rows) with IQ-TREE 2 (Minh et al., 2020) on the alignment from Simion et al. (2017). The site-specific log-likelihood differences $\Delta \log L$ between the maximum likelihood trees of the two competing topologies binned according to their effective number of amino acids $K_{\text{eff}}$ estimated by PhyloBayes (Lartillot and Philippe, 2004) are shown. We do not expect a uniform distribution across the bins, because they are of different size. Site-homogeneous models and the site-heterogeneous LG+C20+PMSF model show inconsistent signal between more versus less constrained sites and favor Ctenophora at the animal root. CAT-PMSF favors Porifera at the animal root, although this result is only significant when using the closest outgroup exclusively.

In our simulations, total support of site-homogeneous models shifted from a Felsenstein-type topology towards a Farris-type topology when $p \geq 0.8$. In this case, sites with $K_{\text{eff}}$ values above the mentioned separating threshold failed to compensate for the bias introduced by sites with $K_{\text{eff}}$ values below the threshold.

We observe no bias when using site-heterogeneous models such as CAT-PMSF (Figure 1). Although we expect such a result, it is satisfying that inferences of CAT-PMSF lack bias even for large values of $p \leq 1.2$ (Figures S7 and S8)

We can discover bias towards one of the topologies in simulation studies because we know the true parameters and trees. Bias is harder to detect in analyses of empirical data. The compositional constraint analyses detect conflicting signal between more and less constrained sites. Detection of such inconsistencies is a strong indicator for bias: Knowing the stationary distribution of a site alone should not provide us with information about the favored evolutionary history. In mathematical terms, the log-likelihood difference of a site between two hypotheses should be conditionally independent given the stationary distribution of that site. In contrast, we expect the signal obtained from more and less constrained sites be consistent up to random statistical error.

In our analyses of empirical data we observed strong inconsistencies between more and less constrained sites for site-homogeneous models and hardly any inconsistencies when using CAT-PMSF (Figure 2). Pearson correlation coefficients and p-values confirm this observation across a wide range of simulated and empirical datasets (Tables S3, and S4).

The results are more nuanced for the alignments involving Ctenophora. In the case of site-homogeneous models, we observe the value of $K_{\text{eff}}$ correlates strongly with the log-likelihood difference between the two competing topologies (Figure 3). Moreover, for the dataset provided by Simion et al., CAT-PMSF supports Porifera-sister — similar to the results reported by the original authors, who applied the CAT model to sub-sampled alignments comprising 100 000 sites. The support of CAT-PMSF for Porifera-sister is significant, when we use the closest outgroups exclusively (Table S11). If we add more distant outgroups, the results are less conclusive (Table S11). Long branch attraction provides an explanation for this observation: more distant outgroups attract the outgroups closer related to the species of interest. In turn, the elongated basal branch of animals increases bias

due to LBA for branches leading to the metazoan root.

We interpret these findings as a confirmation for sponges being the sister group to all other animals (dataset of Simion et al., 2017), and believe that the inconclusive results obtained from the dataset of Ryan et al. (2013) reflect a lack of phylogenetic resolution. Irrespective of the final evolutionary history of Metazoa, our results add important evidence that ignoring across-site compositional heterogeneity leads to LBA (Phillips et al., 2004).

**Further notes** The results of CAT-PMSF are conservative because the CAT model estimates the site-specific stationary distributions using guide topologies prone to LBA artifacts. That is, the guide topologies are obtained with site-homogeneous models. Even so, CAT-PMSF correctly infers the genuine trees in the simulation study (Table S1), and trees that we are convinced to be free from LBA artifacts in the analyses comprising empirical datasets (Figures S3, S2, S1, S6, S5 and S4). This observation justifies the usage of site-homogeneous models in Step 1 of the CAT-PMSF pipeline.

In the simulation study, we observe severe bias when using site-homogeneous models, and no bias or reduced bias when using CAT-PMSF. Further, the absolute values of the log-likelihood differences are greater for site-homogeneous models than for site-heterogeneous models. That is, site-heterogeneous models have reduced power in discriminating between competing hypotheses (bias-variance tradeoff).

In general, site-homogeneous models show conflicting signal between more and less constrained sites, but we observe hardly any such inconsistencies when using CAT-PMSF. In any case, even when the signal across sites is consistent, evidence obtained from highly constrained sites should be examined carefully, especially when highly constrained sites weigh more heavily than less constrained sites. We are convinced that inconsistencies between more and less constrained sites are a strong indicator for the presence of LBA.

Li et al. (2020) argue that only the most parameter rich models favor Porifera-sister, and so Porifera-sister is not a likely scenario. In contrast, Schrempf et al. (2020) report that statistical tests favor models using more stationary distributions. This point is confirmed here, where we see that CXX models, in spite of being generally more robust against LBA than site-homogeneous models, may still be

insufficient and result in conflicting signal (Figures S8, S9 and S10). In practice, each site is different, and we can not expect all sites to share a universal stationary distribution. In fact, we do not even expect stationarity. In our opinion, we should analyze complex models and decide about which parameters are necessary to grasp the complexities of evolution. With CAT-PMSF we further explored this path. The CAT-PMSF method uses site-specific stationary distributions and therefore is a parameter-rich model.

In comparison, the site-specific posterior mean stationary distributions of the classical PMSF approach are a superposition of a finite set of stationary distributions of the underlying mixture model. Consequently, the stationary distribution with the lowest $K_{\text{eff}}$ value constitutes a hard, lower limit. Further, we expect even the richest distribution mixture models do not offer adequate variability of components with stationary distributions exhibiting low $K_{\text{eff}}$ values. For example, there are twenty different stationary distributions with $K_{\text{eff}}$ values close to 1.0, $190 = \binom{20}{2}$ stationary distributions with $K_{\text{eff}}$ values close to 2.0, and so on.

Finally, the speed benefit of CAT-PMSF originates from fixing the topology during the Bayesian analysis with the CAT model. Of course, estimating the site-specific stationary distributions is still by far the most time-consuming step. In the future, we aim to design improved methods estimating site-specific stationary distributions. Specifically, we are thinking about methods based on machine learning such as AlphaFold (Jumper et al., 2021).

In conclusion, our results provide evidence for a potential LBA caused by model misspecification, and thus, an independent qualitative argument for choosing the adequate model for phylogenetic inference.

# 4 Methods

**Effective number of amino acids.** Let $\pi = (\pi_A, \pi_R, \ldots, \pi_V)$ be a distribution of amino acid frequencies, and

$$G(\pi) = \sum_{i \in \{A,R,\ldots V\}} \pi_i^2. \tag{1}$$

$G(\pi)$ is the probability of sampling the same amino acid twice, which is equivalent to a random event of character homoplasy. The effective number of amino acids of distribution $\pi$ is defined as

$$K_{\text{eff}}(\pi) = G(\pi)^{-1}. \tag{2}$$

$K_{\text{eff}}$ is a convenient measure because it ranges from 1.0, when one amino is used exclusively, to 20.0 for a uniform distribution.

**Simulations.** In order to assess the accuracy of CAT-PMSF, we simulated alignments of 10000 amino acids under a distribution mixture model (Schrempf et al., 2020). We used Poisson exchangeabilities (Felsenstein, 1973; Nei, 1987) and a discrete gamma rate model (Yang, 1993) with 4 categories with shape parameter $\alpha = 0.78$. The distribution mixture model has site-specific stationary distributions. For each site, we sampled a random distribution from a universal set of distributions (Schrempf et al., 2020) obtained from the HOGENOM (Dufayard et al., 2005) and HSSP (Schneider et al., 1997) databases.

We used Felsenstein-type topologies with four leaves. The quartet trees had different branch length proportions between short ($q$) and long ($p$) branches (insets of Figure 1). We fixed $q$ to 0.1, and changed $p$ between 0.1 and 2.0. We stored the randomly sampled site-specific stationary distributions used for the simulation. For the simulations we used the ELynx suite (Schrempf, 2021). The scripts and the simulated data are available at `https://github.com/drenal/cat-pmsf-paper`.

**CAT-PMSF.** Figure 4 shows an overview of the CAT-PMSF pipeline. The input to CAT-PMSF is an alignment. The output of the CAT-PMSF pipeline is a tree robust to LBA.

Step 1: Use a site-homogeneous model to infer a ML tree with IQ-TREE 2 (Minh et al., 2020). Specifically, we used LG exchangeabilities (Le and Gascuel, 2008), the empirical stationary distribution of amino acids, and a discrete gamma rate model with 4 categories (LG+G4 in IQ-TREE 2 terminology).

Step 2: Use the obtained tree, which is prone to LBA artifacts, in a subsequent Bayesian analysis with the CAT model (Lartillot and Philippe, 2004) in PhyloBayes (Lartillot et al., 2013). Fix the topology of the tree for the second

step. Thereby, we reduce the computational requirements of the CAT model. Then, extract the posterior mean site-specific stationary distributions of amino acids. In our analyses, we either used Poisson, LG or GTR (Tavaré, 1986) exchangeabilities, and a discrete gamma rate model with 4 categories. We ran two Markov chains until either the effective sample size of all parameters was above 100, or after visual inspection with Tracer (Rambaut et al., 2018) indicated convergence. For the GTR model, we also extracted the posterior mean exchangeabilities from the results of PhyloBayes. All scripts are available at `https://github.com/drenal/cat-pmsf-paper`.

Step 3: Use the custom site-specific stationary distributions in IQ-TREE 2. To this end, use capabilities of IQ-TREE 2 implemented as part of the PMSF method (Wang et al., 2018). The PMSF method has two steps. First, infer the site-specific stationary distributions. Second, use the inferred site-specific stationary distributions for phylogenetic inference. Here, we use the second step of the PMSF method together with the custom site-specific stationary distributions obtained in Step 2 of the CAT-PMSF pipeline.

**Preparation of figures.** We calculated the site-specific likelihood differences between two analyses constrained to topologies $A$, and $B$, respectively (`-g` flag in IQ-TREE 2). For example, in the simulation study, topology $A$ was of the Farris-type and topology $B$ was of the Felsenstein-type. For each site $i$, we calculated the log-likelihood difference as

$$\Delta \log L_i = \log L_i^{\mathrm{B}} - \log L_i^{\mathrm{A}}. \tag{3}$$

A positive value of $\Delta \log L_i$ indicates that site $i$ supports topology $B$. A negative value indicates support for topology $A$.

We ordered and binned the sites according to their $K_{\mathrm{eff}}$ values. For the simulation study, we used the genuine $K_{\mathrm{eff}}$ values. For the analyses of empirical datasets, we used the $K_{\mathrm{eff}}$ values calculated from the site-specific stationary distribution obtained in Step 2 of the CAT-PMSF pipeline. We performed binning with windows sizes of 1.0 $K_{\mathrm{eff}}$ and summed the site-specific log-likelihood differences within each bin.

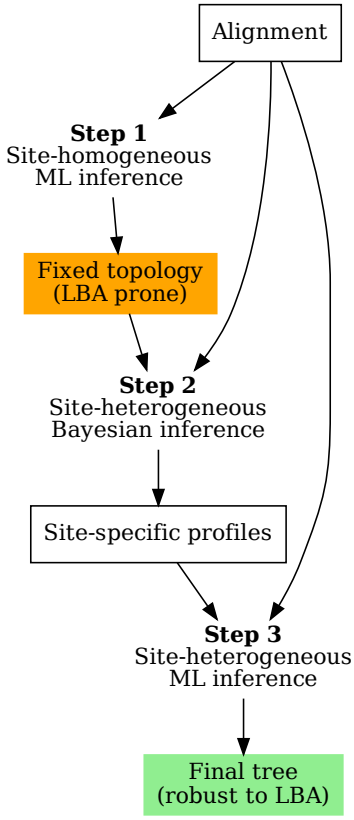


Figure 4: **The CAT-PMSF pipeline.** (1) Apply a site-homogeneous maximum likelihood (ML) model (LG+G4; Le and Gascuel, 2008; Yang, 1993) with IQ-TREE 2 (Minh et al., 2020). The obtained tree may still suffer from long branch attraction (LBA). (2) Fix the topology of this tree in a site-heterogeneous inference with the Bayesian CAT model (Lartillot and Philippe, 2004) and extract the posterior mean site-specific stationary distributions of amino acids. (3) Estimate a tree robust to LBA with the obtained site-specific stationary distributions in IQ-TREE 2.

Exemplary taxa on the inset trees are represented using Phylopic (`http://phylopic.org/`), the silhouette for Microsporidia is based on Tosoni et al. (2002, Figure 7).

**Dataset involving Platyhelminthes and Nematoda.** Philippe et al. (2005a) address a well-known LBA artifact concerning the placement of Platyhelminthes and Nematoda on the tree of Bilateria. Lartillot et al. (2007) revisit the same dataset and provide two reduced, and overlapping alignments which contain 37 species for Nematoda and 32 species for Platyhelminthes, respectively. Both alignments have a length of 35371 amino acids. Figure 2 (a) and (b), S2 and S1 show simplified and complete species trees, respectively.

**Dataset involving Microsporidia.** The dataset provided by Brinkmann et al. (2005) comprises 40 species with 24294 amino acids. It contains an archaean outgroup and eukaryotic taxa. Of particular interest are the Microsporidia, a group of unicellular parasites which lack mitochondria and instead possess mitosomes. Microsporidia evolve fast, and site-homogeneous methods fail to correctly classify them. Application of site-heterogeneous methods confirms that Microsporidia are the closest sister species of Fungi (Brinkmann et al., 2005). For these reasons, the dataset containing Microsporidia is ideal as a proof of concept for CAT-PMSF. Figure 2 (c) and S3 show simplified and complete species trees, respectively.

**Metazoan Datasets.** The placement of Ctenophora on the tree of Metazoa is still a matter of debate. We apply CAT-PMSF to two datasets. First, the alignment provided by Ryan et al. (2013) contains 61 species with 88384 amino acids. Second, the alignment provided by Simion et al. (2017) contains 97 species with 401632 amino acids. The complete set of outgroups comprises 2 Filasterea, 5 Ichthyosporea, and 18 Choanoflagellatea. The Choanoflagellatea are the closest outgroup. Figure 3 shows results obtained from a reduced alignment in which we retained only the Choanoflagellatea. The reduced alignment yields 90 species. Figure 3, and Figure S4 show simplified and complete species trees, respectively.

# 5 Acknowledgments

This work was supported by the Gordon and Betty Moore Foundation through grant GBMF9741 to G.J.Sz and L.L.Sz. D.S. and G.J.Sz. received funding from the European Research Council under the European Union's Horizon 2020 Research and Innovation Program (Grant Agreement No. 714774).

The authors thank Tom A. Williams, Edu Ocaña-Pallarès and László G. Nagy for constructive criticism of the manuscript.

# References

Johannes Bergsten. A review of long-branch attraction. *Cladistics*, 21(2):163–193, 4 2005. doi: 10.1111/j.1096-0031.2005.00059.x.

Henner Brinkmann, Mark Van Der Giezen, Yan Zhou, Gaëtan Poncelin De Raucourt, and Hervé Philippe. An empirical assessment of long-branch attraction artefacts in deep eukaryotic phylogenomics. *Systematic Biology*, 54(5):743–757, 2005. doi: 10.1080/10635150500234609.

Matthew W. Brown, Susan C. Sharpe, Jeffrey D. Silberman, Aaron A. Heiss, B. Franz Lang, Alastair G. B. Simpson, and Andrew J. Roger. Phylogenomics demonstrates that breviate flagellates are related to opisthokonts and apusomonads. *Proceedings of the Royal Society B: Biological Sciences*, 280(1769):20131755, 10 2013. doi: 10.1098/rspb.2013.1755.

Johanna Taylor Cannon, Bruno Cossermelli Vellutini, Julian Smith, Fredrik Ronquist, Ulf Jondelius, and Andreas Hejnol. Xenacoelomorpha is the sister group to nephrozoa. *Nature*, 530(7588):89–93, 2 2016. doi: 10.1038/nature16520.

Frédéric Delsuc, Henner Brinkmann, Daniel Chourrout, and Hervé Philippe. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature*, 439(7079):965–968, 2 2006. doi: 10.1038/nature04336.

Jean-François Dufayard, Laurent Duret, Simon Penel, Manolo Gouy, François Rechenmann, and Guy Perrière. Tree pattern matching in phylogenetic trees: automatic search for orthologs or paralogs in homologous gene sequence databases. 21(11):2596–2603, 2005. doi: 10.1093/bioinformatics/bti325.

James S. Farris. Likelihood and inconsistency. *Cladistics*, 15(2):199–204, 6 1999. doi: 10.1111/j.1096-0031.1999.tb00262.x.

Joseph Felsenstein. Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Systematic Biology*, 22 (3):240–249, 9 1973. doi: 10.1093/sysbio/22.3.240.

Joseph Felsenstein. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Biology*, 27(4):401–410, 12 1978. doi: 10.1093/sysbio/27.4.401.

Michael D. Hendy and David Penny. A framework for the quantitative study of evolutionary trees. *Systematic Zoology*, 38(4):297, 12 1989. doi: 10.2307/2992396.

María José Jimenez, Miguel Arenas, and Ugo Bastolla. Substitution rates predicted by stability-constrained models of protein evolution are not consistent with empirical data. *Molecular Biology and Evolution*, 35(March):743–755, 2018. doi: 10.1093/molbev/msx327.

Thomas H. Jukes and Charles R. Cantor. Evolution of protein molecules. In H. N. Munro, editor, *Mammalian Protein Metabolism*, pages 21–132. Academic Press, 1969.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Kavukcuoglu Koray, Kohli Pushmeet, and Hassabis Demis. Highly accurate protein structure prediction with alphafold. 596(7873):583–589, 2021. doi: 10.1038/s41586-021-03819-2.

Subha Kalyaanamoorthy, Bui Qyuang Minh, Thomas K. F. Wong, Arndt von Haeseler, and Lars S. Jermiin. Modelfinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14:587–589, 2017. doi: doi:10.1038/nmeth.4285.

549 Paschalia Kapli and Maximilian J. Telford. Topology-dependent asymmetry in sys-
550 tematic errors affects phylogenetic placement of ctenophora and xenacoelomor-
551 pha. *Science Advances*, 6(50):eabc5162, 12 2020. doi: 10.1126/sciadv.abc5162.

552 Jeffrey M. Koshi and Richard A. Goldstein. Context-dependent optimal substitu-
553 tion matrices. *Protein Engineering Design and Selection*, 8(7):641–645, 7 1995.
554 doi: 10.1093/protein/8.7.641.

555 Mary K. Kuhner and Joseph Felsenstein. A simulation comparison of phylogeny
556 algorithms under equal and unequal evolutionary rates. *Molecular biology and*
557 *evolution*, 11(3):459–468, 1994.

558 Nicolas Lartillot and Hervé Philippe. A bayesian mixture model for across-site
559 heterogeneities in the amino-acid replacement process. *Molecular Biology and*
560 *Evolution*, 21(6):1095–1109, 2004. doi: 10.1093/molbev/msh112.

561 Nicolas Lartillot, Henner Brinkmann, and Hervé Philippe. Suppression of long-
562 branch attraction artefacts in the animal phylogeny using a site-heterogeneous
563 model. *BMC Evolutionary Biology*, 7(SUPPL. 1):1–14, 2007. doi: 10.1186/
564 1471-2148-7-S1-S4.

565 Nicolas Lartillot, Nicolas Rodrigue, Daniel Stubbs, and Jacques Richer. PhyloBayes
566 MPI: Phylogenetic Reconstruction with Infinite Mixtures of Profiles in a Parallel
567 Environment. *Systematic Biology*, 62(4):611–615, 04 2013. doi: 10.1093/sysbio/
568 syt022.

569 Si Quang Le and Olivier Gascuel. An improved general amino acid replacement
570 matrix. *Molecular Biology and Evolution*, 25(7):1307–1320, 4 2008. doi: 10.1093/
571 molbev/msn067.

572 Yuanning Li, Xing-Xing Shen, Benjamin Evans, Casey W. Dunn, and Antonis
573 Rokas. Rooting the animal tree of life. 2020. doi: https://doi.org/10.1101/2020.
574 10.27.357798.

575 Bui Quang Minh, Heiko A. Schmidt, Olga Chernomor, Dominik Schrempf,
576 Michael D. Woodhams, Arndt von Haeseler, and Robert Lanfear. Iq-tree

2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37(5):1530–1534, 2 2020. doi: 10.1093/molbev/msaa015.

Masatoshi Nei. *Molecular Evolutionary Genetics*. 1987.

Hervé Philippe, Nicolas Lartillot, and Henner Brinkmann. Multigene analyses of bilaterian animals corroborate the monophyly of ecdysozoa, lophotrochozoa, and protostomia. *Molecular Biology and Evolution*, 22(5):1246–1253, 2005a. doi: 10.1093/molbev/msi111.

Hervé Philippe, Yan Zhou, Henner Brinkmann, Nicolas Rodrigue, and Frédéric Delsuc. Heterotachy and long-branch attraction in phylogenetics. *BMC Evolutionary Biology*, 5:1–8, 2005b. doi: 10.1186/1471-2148-5-50.

Hervé Philippe, Romain Derelle, Philippe Lopez, Kerstin Pick, Carole Borchiellini, Nicole Boury-Esnault, Jean Vacelet, Emmanuelle Renard, Evelyn Houliston, Eric Quéinnec, Corinne Da Silva, Patrick Wincker, Hervé Le Guyader, Sally Leys, Daniel J. Jackson, Fabian Schreiber, Dirk Erpenbeck, Burkhard Morgenstern, Gert Wörheide, and Michaël Manuel. Phylogenomics revives traditional views on deep animal relationships. *Current Biology*, 19(8):706–712, 4 2009. doi: 10.1016/j.cub.2009.02.052.

Hervé Philippe, Henner Brinkmann, Richard R. Copley, Leonid L. Moroz, Hiroaki Nakano, Albert J. Poustka, Andreas Wallberg, Kevin J. Peterson, and Maximilian J. Telford. Acoelomorph flatworms are deuterostomes related to xenoturbella. *Nature*, 470(7333):255–258, 2 2011a. doi: 10.1038/nature09676.

Hervé Philippe, Henner Brinkmann, Dennis V. Lavrov, D. Timothy J. Littlewood, Michael Manuel, Gert Wörheide, and Denis Baurain. Resolving difficult phylogenetic questions: Why more sequences are not enough. *PLoS Biology*, 9(3): e1000602, 3 2011b. doi: 10.1371/journal.pbio.1000602.

Matthew J. Phillips, Frédéric Delsuc, and David Penny. Genome-scale phylogeny and the detection of systematic biases. *Molecular Biology and Evolution*, 21(7): 1455–1458, 7 2004. doi: 10.1093/molbev/msh137.

606 Le Si Quang, Olivier Gascuel, and Nicolas Lartillot. Empirical profile mixture mod-
607 els for phylogenetic reconstruction. *Bioinformatics*, 24(20):2317–2323, 8 2008.
608 doi: 10.1093/bioinformatics/btn445.

609 Andrew Rambaut, Alexei J Drummond, Dong Xie, Guy Baele, and Marc A
610 Suchard. Posterior summarization in bayesian phylogenetics using tracer 1.7.
611 67(5):901–904, 2018. doi: 10.1093/sysbio/syy032.

612 Jennifer Ripplinger and Jack Sullivan. Assessment of substitution model adequacy
613 using frequentist and bayesian methods. *Molecular Biology and Evolution*, 27
614 (12):2790–2803, 7 2010. doi: 10.1093/molbev/msq168.

615 Joseph F. Ryan, Kevin Pang, Christine E. Schnitzler, Anh-Dao Nguyen, R. Travis
616 Moreland, David K. Simmons, Bernard J. Koch, Warren R. Francis, Paul Havlak,
617 Stephen A. Smith, Nicholas H. Putnam, Steven H. D. Haddock, Casey W. Dunn,
618 Tyra G. Wolfsberg, James C. Mullikin, Mark Q. Martindale, and Andreas D.
619 Baxevanis. The genome of the ctenophore mnemiopsis leidyi and its implications
620 for cell type evolution. *Science*, 342(6164):1242592–1242592, 2013. doi: 10.1126/
621 science.1242592.

622 Reinhard Schneider, Antoine de Daruvar, and Chris Sander. The HSSP database
623 of protein structure-sequence alignments. 25(1):226–230, 1997. doi: 10.1093/
624 nar/25.1.226.

625 Dominik Schrempf. The elynx suite — a step towards reproducible research in
626 phylogenetics. *Manuscript in preparation*, 2021.

627 Dominik Schrempf, Nicolas Lartillot, and Gergely Szöllősi. Scalable empirical mix-
628 ture models that account for across-site compositional heterogeneity. *Molecular*
629 *Biology and Evolution*, 9 2020. doi: 10.1093/molbev/msaa145.

630 Hidetoshi Shimodaira. An approximately unbiased test of phylogenetic tree
631 selection. *Systematic Biology*, 51(3):492–508, May 2002. ISSN 1063-
632 5157. doi: 10.1080/10635150290069913. URL http://dx.doi.org/10.1080/
633 10635150290069913.

634 Paul Simion, Hervé Philippe, Denis Baurain, Muriel Jager, Daniel J. Richter,
635 Arnaud Di Franco, Béatrice Roure, Nori Satoh, Éric Quéinnec, Alexander

Ereskovsky, Pascal Lapébie, Erwan Corre, Frédéric Delsuc, Nicole King, Gert Wörheide, and Michaël Manuel. A large and consistent phylogenomic dataset supports sponges as the sister group to all other animals. *Current Biology*, 27 (7):958–967, 2017. doi: 10.1016/j.cub.2017.02.031.

Edward Susko, Léa Lincker, and Andrew J Roger. Accelerated estimation of frequency classes in site-heterogeneous profile mixture models. *Molecular Biology and Evolution*, 35(February 2018):1–53, 2018. doi: 10.1093/molbev/msy026.

Yoshio Tateno, Naoko Takezaki, and Masatoshi Nei. Relative efficiencies of the maximum-likelihood, neighbor-joining, and maximum-parsimony methods when substitution rate varies with site. *Molecular Biology and Evolution*, 11(2):261–277, 1994.

Simon Tavaré. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences*, 17:57–86, 1986.

Antonella Tosoni, Manuela Nebuloni, Angelita Ferri, Sara Bonetto, Spinello Antinori, Massimo Scaglia, Lihua Xiao, Hercules Moura, Govinda S. Visvesvara, Luca Vago, and Giulio Costanzi. Disseminated microsporidiosis caused by encephalitozoon cuniculi iii (dog type) in an italian aids patient: a retrospective study. *Modern pathology*, 15(5):577–583, 2002.

Huai Chun Wang, Karen Li, Edward Susko, and Andrew J. Roger. A class frequency mixture model that adjusts for site-specific amino acid frequencies and improves inference of protein phylogeny. *BMC Evolutionary Biology*, 8(1):1–13, 2008. doi: 10.1186/1471-2148-8-331.

Huai Chun Wang, Bui Quang Minh, Edward Susko, and Andrew J. Roger. Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Systematic Biology*, 67(2):216–235, 3 2018. doi: 10.1093/sysbio/syx068.

Nathan V. Whelan and Keneth M. Halanych. Who let the cat out of the bag? accurately dealing with substitutional heterogeneity in phylogenomic analyses. *Systematic Biology*, 66(2):232–255, 9 2016. doi: 10.1093/sysbio/syw084.

Tom A. Williams, Cymon J. Cox, Peter G. Foster, Gergely J. Szöllősi, and T. Martin Embley. Phylogenomics provides robust support for a two-domains tree of life. 4:138–147, 2020. doi: 10.1038/s41559-019-1040-x.

Ziheng Yang. Maximum-likelihood estimation of phylogeny from dna sequences when substitution rates differ over sites. *Molecular Biology and Evolution*, 11 1993. doi: 10.1093/oxfordjournals.molbev.a040082.

So Wei Yeh, Jen Wei Liu, Sung Huan Yu, Chien Hua Shih, Jenn Kang Hwang, and Julian Echave. Site-specific structural constraints on protein sequence evolutionary divergence: Local packing density versus solvent exposure. *Molecular Biology and Evolution*, 31(1):135–139, 2014. doi: 10.1093/molbev/mst178.

Andrey Zharkikh and Wen-Hsiung Li. Inconsistency of the maximum-parsimony method: the case of five taxa with a molecular clock. *Systematic Biology*, 42(2): 113–125, 6 1993. doi: 10.1093/sysbio/42.2.113.