Phylogenetic model of stabilizing selection is more informative about site specific selection than extrapolation from laboratory estimates.

CEDRIC LANDERER[1,2,*], BRIAN C. OMEARA[1,2], AND MICHAEL A. GILCHRIST[1,2]

[1]Department of Ecology & Evolutionary Biology, University of Tennessee, Knoxville, TN 37996-1610

[2]National Institute for Mathematical and Biological Synthesis, Knoxville, TN 37996-3410

[*]Corresponding author. E-mail: cedric.landerer@gmail.com

Version dated: October 13, 2018

Abstract

Here we examine the adequacy of experimentally inferred site specific selection for amino acids to inform phylogenetic inferences of sequence evolution. Previous work has shown that laboratory estimates of selection can improve model fit but did not assess their adequacy. We assess the adequacy of experimentally inferred site specific selection using DMS to inform phylogenetic models. We use the β-lactamse TEM for which empirical estimates of site specific selection on amino acids are readily available. We compare our results to *SelAC*, a new phylogenetic model of stabilizing selection. Using simulations to assess model adequacy, we find that experimentally inferred selection does not adequately reflect evolution in the wild. In contrast, *SelAC* improves model fit over models informed by experimentally inferred selection and provides higher model adequacy. We demonstrate the capability of *SelAC* by estimating site specific genetic load of the observed TEM variants.

Introduction

Numerous attempts to incorporat selection into phylogenetic models have been made. Early models focused on the influence of selection on the substitution rate and fixation probability between a resident and a mutant introduced into a population[19, 33, 45]. These models however, lack site specific equilibrium frequencies. The importance of site specific equilibrium frequencies has long been noted [10, 18]. **(author?)** [22] first introduced a model to incorporate site specific equilibrium frequencies of amino acids. However, they had to concede that their model was too parameter rich and therefore intractable for biological data sets without additional simplifying assumptions. More recent models incorporating site specific equilibrium frequencies still require a large number of parameter to be estimated from the sequence data [29, 30, 47, 25, 48, 44]. Other approaches treat site specific selection as a random effect [39, 37, 38]. A full parameterization of site specific equilibrium for amino acids requires $19 \times L$ parameters were L is the length of the sequence. It therefore is an attractive option to utilize laboratory experiments to empirically estimate site specific strength of selection on amino acids and infer their equilibrium frequencies [4, 46, 5].

Deep mutation scanning (DMS) is recently used to generate comprehensive site specific estimates of the strength of selection on amino acids along a protein sequence [15]. The ability to estimate site specific strength of selection on amino acids allows to estimte site specific amino acid preferences and the genetic load a mutation would introduce at a particular site[4, 12, 43]. The quality of empirical estimates from DMS depends on many factors, including the initial library of mutants and the applied selection [13].

Incorporating empirical estimates of site specific strength of selection on amino acids has important advantages. Individual amino acid site along the protein show differences in evolutionary rates, and strong preferences for amino acids [22, 1, 9]. The usage of site specific selection acknowledges the heterogeneity in selection and amino acid preferences along the protein sequence [24]. DMS reduces the number of parameters that have to be estimated from the data, making it applicable for smaller data sets and allowing for the fitting of more

complex models. There are, however, also shortcomings. Estimates of selection can only be obtained for fast growing organisms that can be manipulated under laboratory conditions. Mutation libraries have to be extendsive and prodice a heterogeneous population of competing organsims usually not found in nature. This is a severe limitation of experimentally informed models as many organism can not be cultivated under laboratory conditions or have long generation times.

Even in the cases were empirical estimates of site specific selection on amino acids can be obtained, their applicability for phylogenetic reconstruction is quastionable. In this study, we assess the adequacy of experimentally inferred site specific selection using DMS to inform phylogenetic models. We use site specific estimates of selection on amino acids for the β-lactamase TEM from **(author?)** [43]. We fitted 227 nucleotide and codon models using IQTree and compared their model fits to site specific models of stabilizing selection with (*phydms*, *SelAC*+DMS) and without (*SelAC*) experimentally determined site specific selection coefficients on amino acids [34, 24, 3]. We find that experimentally inferred selection, while improving model fit, does not adequately reflect observed wild type sequences. In contrast, *SelAC* [3] a mechanistic phylogenetic model of stabilizing selection rooted in first principles with site specific equilibrium frequencies improves model fit, and better reflects evolution in the wild. Because *SelAC* assumes that the distance of two amino acids in physicochemical space affects substitution probabilities it is able to infer the optimal amino acid at a site and reduce the number of site specific parameters from $19 \times L$ to L.

Results

Site Specific Stabilizing Selection on Amino Acids Improves Model Fit

We compared the models *phydms* [24] and *SelAC* [3], models of site specific stabilizing selection on amino acids, to 227 other codon and nucleotide models. We fitted all models

to 49 sequences of the β-lactamase TEM. The *phydms* and *SelAC* models with site specific selection on amino acids improved model fits by 917 to 1483 AICc units, respectively over codon or nucleotide models without site specific selection (Table 1). In addition, *SelAC* outperformed the experimentally informed model *phydms* by 560 to 566 AICc units, depending whether site specific selection on amino acids was inferred by *SelAC* or experimentally informed.

SelAC utilizes a hierarchical model and estimates 263 site specific parameters, $\sim 5\%$ of the $19 \times L = 4997$ parameters necessary to fully describe the site specific selection on amino acids. In contrast, *phydms* does not infer any site specific parameters, but utilizes site specific selection on amino acids estimated from deep mutation scanning experiments. We fixed the optimal amino acid at each site to the experimentally determined one in *SelAC* and refitted the model to the 49 TEM sequences (*SelAC*+DMS). Incorporating site specific selection on amino acids estimated from deep mutation scanning experiments into *SelAC* (*SelAC*+DMS) yields a similar AICc value to *SelAC* without that information. We incorporated the experimentally inferred site specific amino acids by estimating the

However, *SelAC*+DMS is favored by AICc. This is solely due to a decrease in the number of parameters estimated, as the model log-likelihood ($\log(\mathcal{L})$) worsens from -1498 to -1768 (Table 1). 263 of the 374 parameters estimated are the discrete optimal amino acid state at each site. However, it is unclear if discrete parameters contribute to the Kullback-Leibler divergence like continous parameters do and have to be penelaized like such. Therefore, the number of parameter for *SelAC* is reported conservatively as the number of unique site patterns in the TEM alignment is only 27 which would yield a total number of 1383 parameters (96 edge length, 15 mutation/selection parameters, and 27 optimal amion acids). This however would likely be an under estimate of the number of parameteres estimated and the true number of parameters remains unclear at this point due to the inherent non-independence of the underlying data and the descrete nature of the optimized parameters.

Model	log($\mathcal{L}$)	n	AIC	ΔAIC	AICc	ΔAICc
SelAC+DMS	-1768	111	3758	14	3760	0
SelAC	-1498	374	3744	0	3766	6
phydms	-2061	102	4326	582	4328	568
SYM+R2	-2230	102	4663	919	4694	934
GY+F1X4+R2	-2243	102	4690	946	4821	1061

Table 1: Model selection, shown are the three models of stabilizing site specific amino acid selection (*SelAC*, *SelAC*+DMS, *phydms*) and the best performing codon and nucleotide model. Reported are the log-likelihood (log($\mathcal{L}$)), the number of parameters estiamted n including edge length, AIC, ΔAIC, AICc, and ΔAICc values. See Table 1 for all models tested.

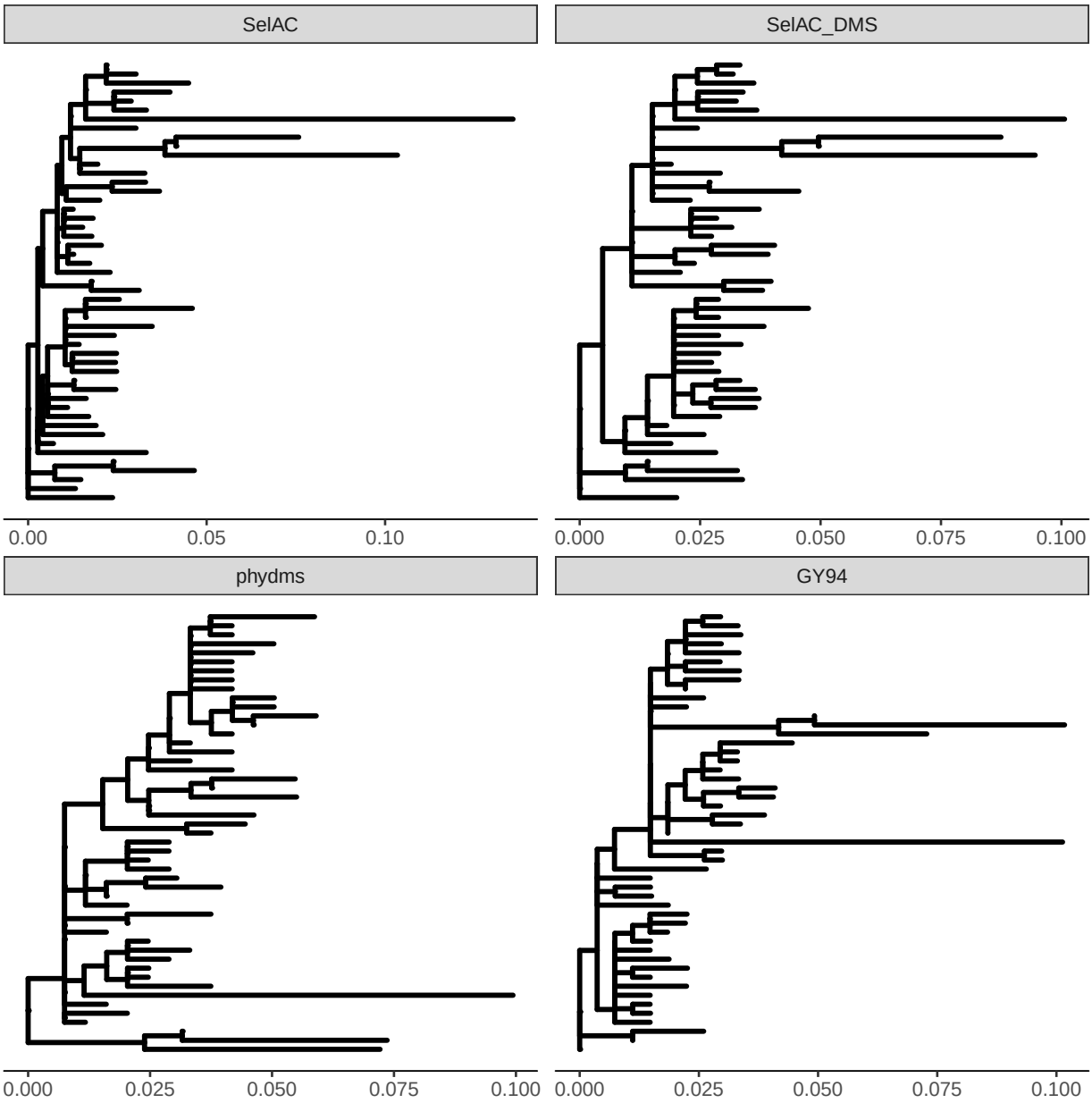

Figure 1: Phylogenies resulting from *SelAC*, *SelAC*+DMS, *phydms*, and *GY94*. As *SelAC* is currently to slow for the inference of tomopogies, the topology for the *SelAC* phylogenies was inferred using the codon model of **(author?)** [28].

We observe differences in the topology between model fits. The *SelAC* model is currently too slow to estimate the topology, therefore the topology was estimated using the codon model of **(author?)** [28]. At this point, it is therefore unclear if the difference in topology can be attributes to the experimentally inferred selection on amino acids. We find that the best codon model (*GY94*) [19] is outperformed by several nucleotide model e.g. *SYM* [50]. This could be an indication that negative frequency dependent selection like it is modeled in *GY94* is not appropriate for TEM [19, 3]. Figure 1 shows that the estimated phylogenetic trees shift from long terminal branches (*SelAC*) to longer internal branches (*phydms*). While the *SelAC* model fit shows 84% of all evolution happening at the tips, this reduces to 79% in the *SelAC*+DMS model fit, and 77% in the *phydms* and *GY94* model fits. All models produce polytomies but their location differs between models. The largest polytomies appear in the experimentally informed phylogenies of *phydms*. The position of the sequences with the longest branches differ between *SelAC* and *phydms*.

Laboratory Inferences Inconsistent with Observed Sequences.

The improved model fits with phydms relative to classical nucleotide and codon models are, however, deceiving. The site specific selection inferred by the deep mutation scanning experiment is inconsistent with the observed TEM sequences. We find that the sequence of selectively favored amino acids has only 52% sequence similarity with the observed consensus sequence (Figure 2). This is in contrast to the 99% of sequence similarity with the sequence of selectively favored amino acids estimated by *SelAC*.

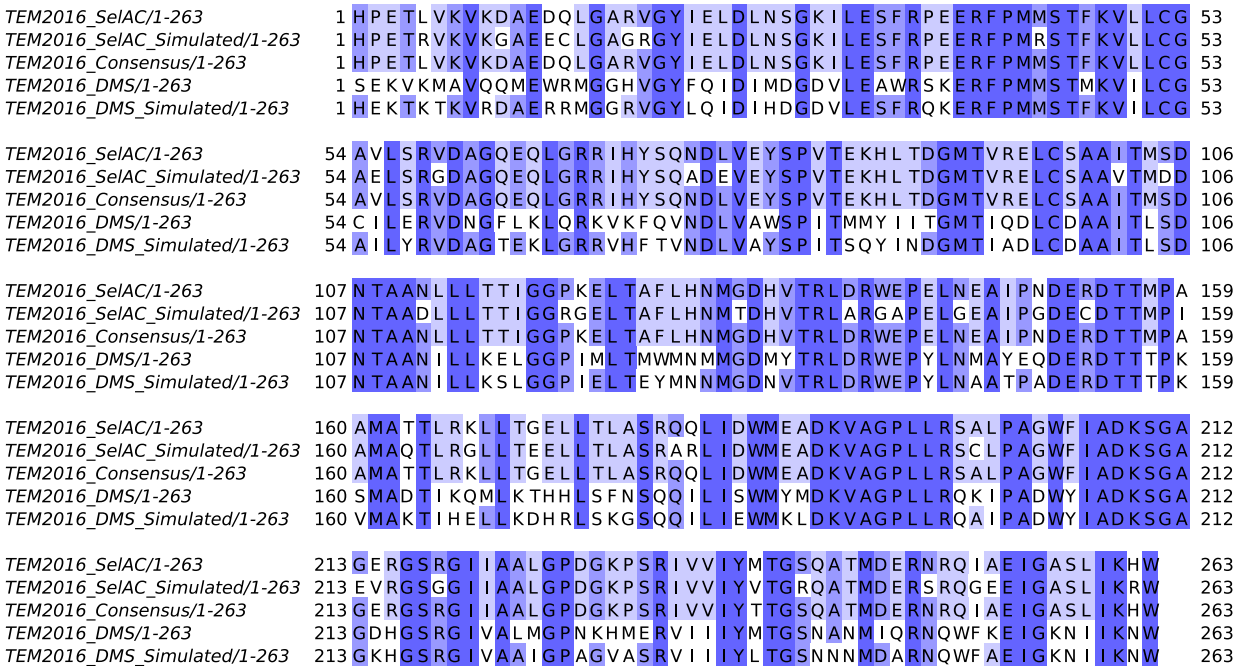

Figure 2: Alignment of TEM optimal and simulated sequences. Indicated is the percentage identity at each site.

Simulations of codon sequences under the experimentally inferred site specific selection for amino acids reveals that we would not expect to see the observed TEM sequences. We simulated under a wide range of effective population sizes N_e, and find that the experimentally inferred site specific selection is very strong. With more realistic values of $N_e = 10^7$, we find that the simulated sequences are 62% similar to the observed consensus sequence (Figure 3a). This is a higher similarity than the observed consensus sequence shows with the sequence of selectively favored amino acids estimated using deep mutation scanning. Only when N_e is reduced to one individual does drift overpower selection (Figure 3b). The genetic load of the simulated sequences decrease slowly with increasing N_e (Figure 3b). After simulating until the sequences reach one expected mutation per site and $N_e = 10^7$ the simulated sequences show a genetic load of 0.25, which is in contrast to the ~ 8 times higher than the estimated observed load of 2.1. Thus it appears unlikely that the observed sequences have evolved under the DMS inferred site specific selection values.

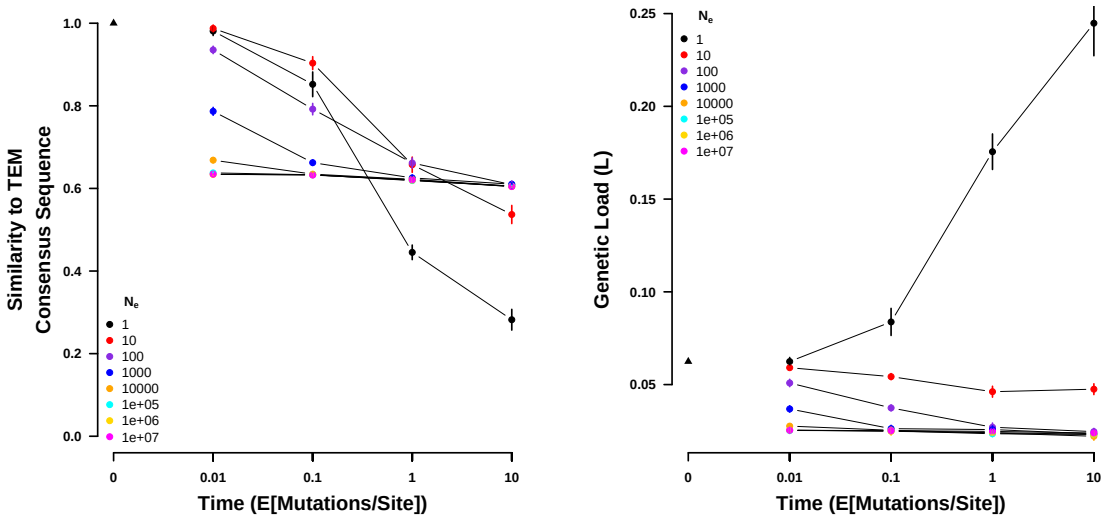

Figure 3: Sequences simulated from the ancestral state under the site specific selection on amino acids estimated using deep mutation scanning. (left) Sequence similarity to the observed consensus sequence at various times for a range on values of N_e. (right) Genetic load of the simulated sequences at various times for a range on values of N_e. Time is given in number of expected mutations per site, which equals the substitution rate of a neutral mutation. Points indicate sample means and vertical bars indicate standard deviations. Initial sequence is the inferred ancestral state of the TEM variants and indicated by a black triangle.

Stabilizing Selection for Optimal Physicochemical Properties Improves Model Adequacy

We assessed model adequacy of *SelAC* and find that *SelAC* better explains the observed TEM sequences. The observed consensus sequence has 99% sequence similarity with the sequence of selectively favored amino acids estimated by *SelAC*. Furthermore, assuming the site specific selection estimated by *SelAC*, the observed sequences represent a very small genetic load on the order of 10^{-6}(Table 2, Figure 5).

We simulated codon sequences forward in time for various length of time, using the *SelAC* inferred site specific selection for amino acids to assess sequence similarity. We simulated the evolution of TEM from the inferred ancestral state using a wide range of effective population sizes N_e (Figure 4a). The ancestral state was estimated to be the

observed consensus sequence. As expected, for small N_e, simulated sequences drift away from the observed consensus sequence. Because of the high similarity between the optimal amino acid sequence estimated by *SelAC* and the observed consensus sequence, the genetic load increases drastically as a result. Increasing N_e to 10^7 the simulated sequences reach a sequence similarity of 83%, this is in contrast to the observed avergage sequence similarity of 98%.

We estimated the total genetic load the sequences represent using the *SelAC* inferred selection on amino acids. The total genetic load of the simulated sequences avergages 9.8×10^{-6} (Figure 4b). The total estimated genetic load of the observed sequences averages 4.2×10^{-5}. Thus, the simulated sequences show a lower genetic load despite the greater divergence from the observed consensus sequence.

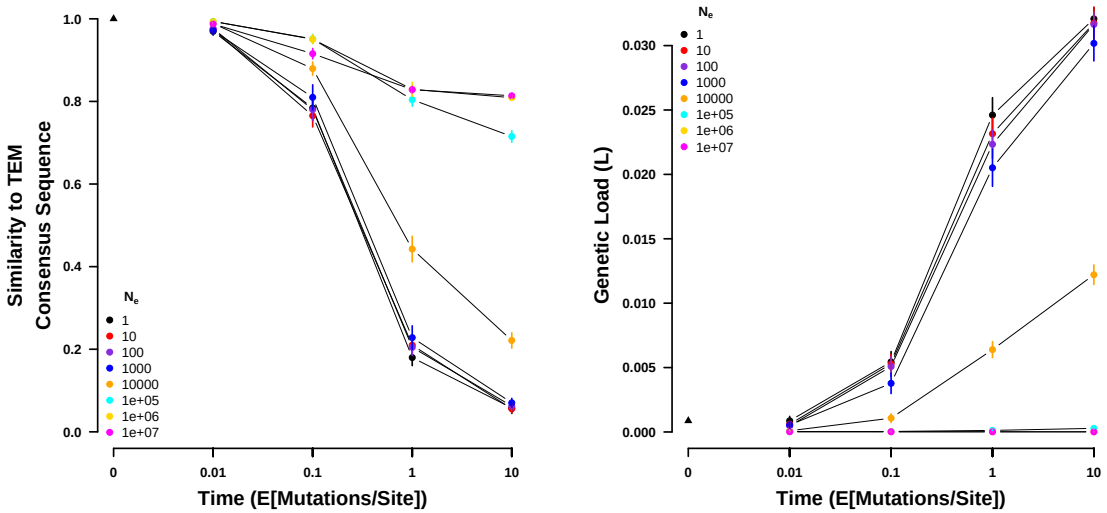

Figure 4: Sequences simulated from the ancestral state under the site specific selection on amino acids estimated using *SelAC*. (left) Sequence similarity to the observed consensus sequence at various times for a range on values of N_e. (right) Genetic load of the simulated sequences at various times for a range on values of N_e. Time is given in number of expected mutations per site, which equals the substitution rate of a neutral mutation. Points indicate sample means and vertical bars indicate standard deviations. Initial sequence is the inferred ancestral state of the TEM variants and indicated by a black triangle.

To further demonstrate the consistency of *SelAC*, we simulated codon sequences over the same period of time using 10 uniform samples codon sequences with 263 sites, the same

length as the observed TEM variants. We find that the sequence similarity increases with effective population size N_e. The random sequences start of with a similarity of $\sim 6\%$ which increases with N_e to $\sim 28\%$ (Figure S2a). The same initial sequences under the site specific selection inferred by the deep mutation scanning experiment increase only to $\sim$ 18% in sequence similarity.

Site Specific estimates of Selection on Amino Acids

SelAC allows for the estimation of site specific selection on amino acids and the genetic load of an observed amino acid relative to the inferred optimal amino acid. We find that the genetic load is s along most of the observed TEM sequence with the exception of the region between residue 80 to 120 were three consecutive helices are located (Figure 5). The most noticeable increases in genetic load are found in unstructured regions The largest increase in genetic load however, is located at the beginning of the last helix. We therefore estimate similar genetic loads for helices and unstructured regions in the observed TEM sequences (Table 2). The highest efficacy of selection G and the lowest genetic load among the TEM secondary structure features is estimated in the β-sheet regions. The Active sites appear to be under the strongest selection, with no accumulated genetic load. This is in concordance with the experimental estimates.

It was previously proposed that experimentally inferred site specific selection for amino acids can be used to extrapolate the fitness landscape of related proteins [4, 5]. We therefore compared the site specific efficacy of selection G, the *SelAC* selection parameters of our *SelAC* TEM model fit to a *SelAC* model fit of SHV, genetic load. We find that site specific efficacy of selection G differs greatly between SHV and TEM ($\rho = 0.12$), despite a similar estimate of the parameter α_G describing the distribution of G values (Figure S4a). We generally find increased G values in SHV, with the exception of the active site (Table 2). However, most *SelAC* selection parameters are very similar between the TEM and the SHV model fit. An exception is the weight for the physicochemical composition property α_c

			G		Genetic Load	
Protein	Secondary Structure	# Residues	Mean	SE	Mean	SE
TEM		263	219.3	7.5	15.9×10^{-8}	6.5×10^{-8}
	Helix	113	206.1	12.4	17.5×10^{-8}	13.1×10^{-8}
	β-Sheet	48	238.6	15.8	6.8×10^{-8}	2.9×10^{-8}
	Unstructured	102	224.8	11.4	18.6×10^{-8}	8.1×10^{-8}
	Active Sites	3	300	0	0	0
SHV		263	244.9	6.8	4.0×10^{-8}	1.9×10^{-8}
	Helix	102	234.6	11.5	7.3×10^{-8}	4.8×10^{-8}
	β-Sheet	66	253.1	12.8	2.1×10^{-8}	1.1×10^{-8}
	Unstructured	95	250.3	11.0	1.8×10^{-8}	0.6×10^{-8}
	Active Sites	3	199.9	100	2.4×10^{-8}	2.4×10^{-8}

Table 2: Efficacy of selection (G) and Genetic Load for TEM and SHV, and separated by secondary structure. G was estimated as a truncated variable with an upper bound of 300.

(Figure S4b). Furthermore, we find that the sequences of selectively favored amino acids estimated by *SelAC* for TEM and SHV only show 68% sequence similarity.

The genetic load in SHV is lower than in TEM with the exception of residues found in β-sheets and the active site (Table 2). This is consistent with the elevated site specific efficacy of selection G in SHV. However, only differences in the genetic load between the TEM and SHV unconstrained regions are significant at the $\alpha = 0.05$ significant level ($p = 0.04$). As a comparison of site specific efficacy of selection G already indicated, the sites introducing genetic load differ between SHV and TEM (Figure S1). In contrast to TEM, we find the highest genetic load in SHV secondary structure features in the helices (Table 2). We find the highest genetic load in SHV at the end of the first helix. However, we do find a peak of similar magnitude in the TEM sequence at the end of the first helix.

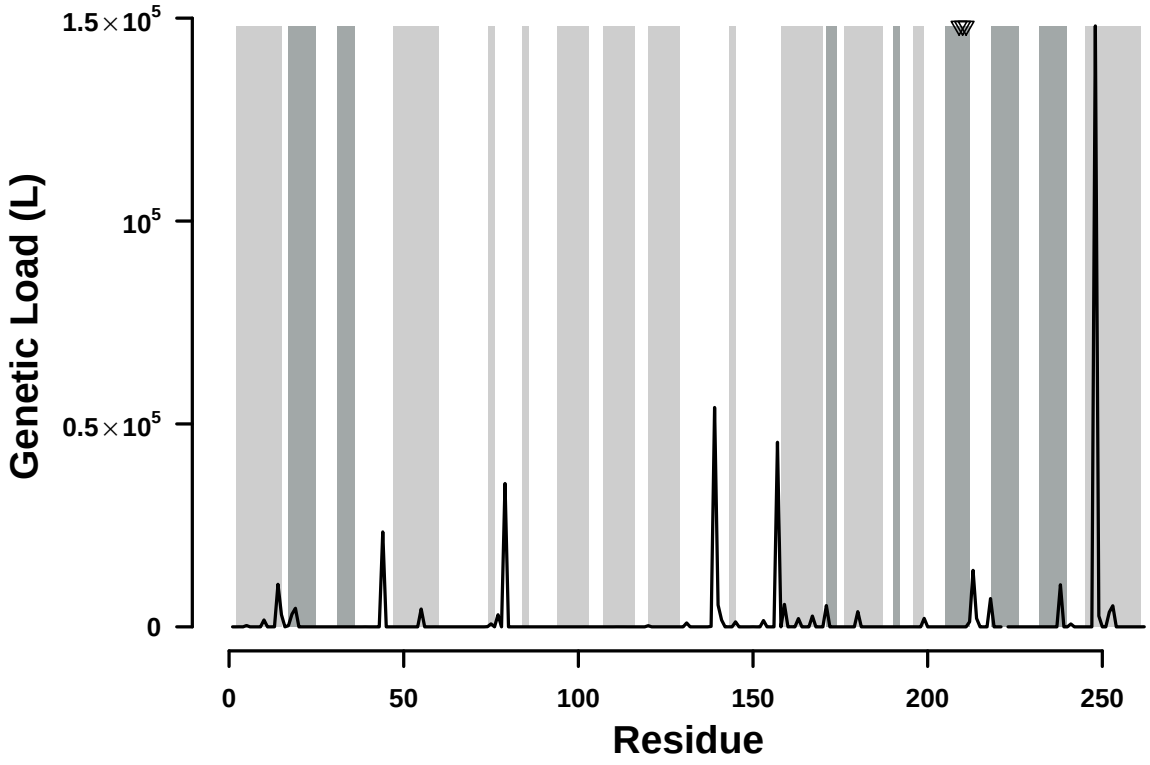

Figure 5: Distribution of genetic load in TEM. Average genetic load over all observed TEM variants is indicated by the black line. Light gray bars indicate where helices are found, and dark gray bars indicate β-sheets. The three residues forming the active sites are indicated by three triangles at the top of the plot.

Discussion

Here we revisited how well experimental selection estimates from laboratory experiments, specifically deep mutation scanning, explain sequence evolution and compared it to *SelAC*, a novel phylogenetic framework. Previous work has shown that laboratory estimates of selection can improve model fit over classical approaches like GY94 [4, 5]. While our study confirms this notion, we identify important shortcomings of these laboratory estimates. In contrast, *SelAC* is a more general phylogenetic model of stabilizing selection that does not

require costly laboratory estimates of selection and is nevertheless favored by model selection (Table 1). *SelAC* does not rely on artificially induced selection in the laboratory but is a mechanistic framework rooted in first principles. It estimates site specific selection on amino acids from the sequence data based on distances between amino acids in physicochemical space [21, 3]. This allows *SelAC* to be applied to any set of protein coding sequences, eliminating the need to extrapolate from one homologous gene family to the next (e.g. from TEM to SHV).

While previous work showed the advantages of experimentally informed phylogenetics, they did not assess how adequate the estimated selection reflects observed wild-type sequences. The low sequence similarity between the observed consensus sequence and the sequence of selectively favored amino acids estimated from deep mutation scanning experiments is evidence for that. This begs the question how well the experimental selection coefficients represent evolution of sequences in nature. Deep mutation scanning experiments are performed using a comprehensive library of mutants and a strong artificial selection pressure [13, 26, 14, 15]. This results in a very large selection coefficient s and a heterogeneous population of competing individuals unlikey to occur in nature.

The selection pressure imposed during the DMS experiment was limited to ampicillin and focused solely on TEM-1 [43]. However, TEM variants can also confer resistance to a wide range of other antibiotics, including penicillins, cephalosporins, cefotaxime, ceftazidime, or aztreonam [41, 42, 20, 31, 7, 6]. Thus, the inferred selection is biased towards ampicillin and is inconsistent with the observed TEM sequences (Figure 3). This may very well be very appropriate to explore the selection on TEM in a hospital environment but is unlikely to be representative of the selection faced by *E. coli* in nature.

If we assume that the DMS selection coefficients underly the evolution of the observed TEM sequences we are left with two possible explanations for the observed sequences. First, the sequences are unable to reach a fitness peak, potentially due to a low selection pressure, or not enough time. Second, the observed TEM sequences are highly maladapted. Both options

seem unlikely. *E. coli* has a large effective population size N_e, estimates are on the order of 10^8 to 10^9 [35, 23]. As new mutations are introduced into a population at a rate proportional to N_e, *E. coli* can effectively explore the sequence space. However, this also raises concerns that the population mutation rate of *E. coli* $\Theta = 4N_e\mu$ exceeds 0.1 and violated *SelAC*'s weak mutation assumption [8]. We therefore expect the observed sequence variants to be near mutation-selection-drift equilibrium. This expectation is supported by our simulations in which we observe a higher sequence similarity with the observed TEM consensus sequence and decreased genetic load even with much smaller N_e (Figure 3). Furthermore, previous work showed that the catalytic reaction performed by TEM of penicillin-class antibiotics is close the diffusion limit, making TEM a so-called perfect enzyme [32].

As experimental selection estimates are not readily available for most organsims and proteins, one solution is to extrapolate the estimates to homologous gene families [4, 5]. When extrapolating the selection estimates from the β-lactamase family TEM to the SHV family, the sequence similarity between the observed consensus sequence and the sequence of selectively favored amino acids estimated from deep mutation scanning experiments drops slightly from 52% to 49%. In contrast, the site specific efficacy of selection (G) revealed large differences in the site specific selection on amino acids between TEM and SHV. The mismatched in physicochemical weights also indicates differences in selection constraints. While the polarity of amino acids is of similar importance in TEM and SHV, amino acid composition plays a much greater role in SHV than in TEM. In contrast to the experimental selection estimates, the *SelAC* selection estimates are consistent with the observed sequences, e.g. the selectively favored amino acids estimated by *SelAC* shows a high sequence similarity with the observed TEM and SHV consensus sequence (99%).

While *SelAC* better explains the observed TEM sequences than the experimental estimates of site specific selection on amino acids, it is not without shortcomings itself. *SelAC* is currently to slow to be used in topology searches, therefore it is unclear if the differences in topology between *phydms* and *SelAC* can be attributed to the same inadequacies of ex-

perimentally inferred selection. As the simulation of TEM evolution from he ancestral state under the *SelAC* inferred site specific selection revealed, the formulation of *SelAC* can and should be improved upon. Starting from the ancestral sequence, the simulated sequences show initial divergence despite stabilizing selection for the optimal amino acid. While *SelAC* allows for site heterogeneity in selection for amino acids, it still ignores epistatis. This however, is a shortcoming shared with experimental estimates by deep mutation scanning, as each mutant typically only carries one mutation [13, 26]. *SelAC* is a model stabilizing selection, however, not every protein is under stabilizing selection. TEM plays a role in chemical warfare with conspecifics and other microbes, therefore some sites may be under negative frequency dependent selection. This potential heterogeneity in selection highlights another shortcoming of *SelAC*. *SelAC* assumes the same distribution for the efficacy of selection (G) and physicochemical sensitivities across the whole protein. However, it is easy to imagine that sites in different secondary structures or at active sites do not share a common distribution.

As *SelAC* assumes that the fitness of an amino acid at a site declines with its distance in physicochemical space to the optimal amino acid, the choice of physicochemical properties becomes important. In this study, we assumed the physicochemical properties estimated by **(author?)** [21] for all sites. However, a wide range of additional physicochemical properties of amino acids have been described [27]. A more optimal choice of physicochemical properties may be possible as well as the a relaxation of the assumptions that the same properties apply to all sites equally. Future work will attempt to address these shortcomings, however, *SelAC*'s hierarchical model structure and the open-source code base allow researchers to easily address these shortcoming if desired.

In conclusion, experimental estimates of site specific selection on amino acids have to be treated with skepticism and their adequacy should be assessed before using them to inform phylogenetic inferences. This study was initiated to assess the quality of *SelAC* with the expectation that *SelAC* could be a faster, cheaper, and more readily available alternative

to experimentally inferred selection; specifically in organisms where these experiments are not feasible. Intuitively one would expect that selection coefficients estimated of mutations in living organisms would provide more information on the evolution of proteins than a model relying on many simplifying assumptions. As we show in this study, not only can *SelAC* estimate site specific selection on amino acids but our approach is a more adequat descripton of selection on amino acids in nature than experimental estimates.

Materials and Methods

Phylogenetic Inference and Model selection

TEM and SHV sequences were obtained from **(author?)** [5] already aligned. We however, separated the TEM and SHV sequences into individual alignments. Experimentally fitness values for TEM were taken from **(author?)** [43]. We followed [5] to convert the experimental fitness values into site specific equilibrium frequencies for *phydms*. *phydms* (version 2.5.1) was fitted using site specific selection on amino acids estimated from deep mutation scanning experiments from **(author?)** [43] and python (version 3.6).

SelAC (version 1.6.1) was fitted to the TEM alignment using R (version 3.4.1) [36] with and without site specific selection on amino acids estimated from deep mutation scanning experiments. We assumed the physicochemical properties estimated by **(author?)** [21]. We chose the contraint free general unrestricted model [49] as mutation model . All other models were fitted using IQTree [34].

We report each model's $\log(\mathcal{L})$, AIC, and AICc. Models were selected based on the AICc values.

Sequence Simulation

Sequences were simulated by stochastic simulations using a Gillespie algorithm [17] that was model independent. The simulation followed **(author?)** [40] to calculate fixation probabil-

ities. The fitness values were estimated using *SelAC* or experimentally inferred. We chose the fitness values of the highest concentration (2500 $\mu g/mL$) treatment of ampicillin for our comparison. We modified the experimental fitness such that the amino acid with the highest fitness at each site has a value of one. Mutation rates were taken from the *SelAC* or *SelAC*+DMS fit. The initial sequences were either a random sample of 263 codons or the ancestral sequence reconstructed using FastML [2] (last accessed: 30.09.2018). Each sequence was simulated 10 times and we report average genetic load and sequence similarity and the corresponding standard error. The sequences were sampled at times 0.01, 0.1, 1, and 10 expected mutations per site.

Estimating site specific efficacy of selection G

SelAC does not by default estimate site specific values for G but assusmes G values follow a gamma distribution [11]. Site specific values for G were optimized by fixing all estimated parameters and perfoming a maximum likelihood search without the usual integration over G. In contrast to *SelAC* that assumes G to be purley positive, we allowed negative values for G and constraint the search to values between -300 and 300.

Estimating site specific fitness values w_i

Following **(author?)** [3] w_i is proportional to

$$w_i \propto \exp(-A_0 \eta \psi) \tag{1}$$

were A_0 describes the decline in fitness with each high energy phosphate bond wasted per unit time, and ψ is the protein's production rate. η is the cost/benefit ratio of a protein (see [3] for details). However, *SelAC* only estimates a composition parameter $\psi' = A_0 \psi N_e$. N_e describes the effective population size. *SelAC* assumes $N_e = 5 \times 10^6$. *SelAC* assumes

$A_0 = 4 \times 10{-}7$ [16]. Thus,

$$\psi = \frac{\psi'}{A_0 N_e q} \tag{2}$$

Model Adequacy

Model adequacy was assessed by comparing the observed sequences and simulations under the site specific selection inferred by the deep mutation scanning experiment or *SelAC*. First, similarity between the sequence of selectively favored amino acids and the observed TEM sequences was assessed. Sequence similarity was measured as the number of differences in the amino acid sequence. Second, the genetic load of the observed and the simulated sequences was calculated using either the site specific selection inferred by the deep mutation scanning experiment or *SelAC*.

Genetic load was calculated as

$$L_i = \frac{w_{max} - w_i}{w_{max}} \tag{3}$$

were w_{max} is the fitness of the sequence of selectively favored amino acids estimated using the site specific selection inferred by the deep mutation scanning experiment or *SelAC*. w_i represents the fitness of the ith residue. This the genetic load L of a sequence is given by $\sum_{i=1}^{n} L_i$ where n is the number of amino acids.

Acknowledgments

This work was supported in part by NSF Award and DEB-1355033 (BCO, MAG, and RZ) with additional support from The University of Tennessee Knoxville. CL received support as a Graduate Student Fellow at the National Institute for Mathematical and Biological Synthesis, an Institute sponsored by the National Science Foundation through NSF Award DBI-1300426, with additional support from UTK. The authors would like to thank Russel

Zaretzki, Jeremy Beaulieu and Alexander Cope for their helpful criticisms and suggestions for this work.

References

[1] O Ashenberg, LI Gong, and JD Bloom. Mutational effects on stability are largely conserved during protein evolution. *Proccedings of the National Academy of Sciences U.S.A*, 110:21071–21076, 2013.

[2] H Ashkenazy, O Penn, A Doron-Faigenboim, O Cohen, G Cannarozzi, O Zomer, and T Pupko. Fastml: a web server for probabilistic reconstruction of ancestral sequences. *Nucleic Acids Research*, 40(Web Server Issue):W580–4, 2012.

[3] JM Beaulieu, BC O'Meara, R Zaretzki, C Landerer, JJ Chai, and MA Gilchrist. Population genetics based phylogenetics under stabilizing selection for an optimal amino acid sequence: A nested modeling approach. *Molecular Biology and Evolution*, X:NA, in review.

[4] JD Bloom. An experimentally informed evolutionary model improves phylogenetic fit to divergent lactamase homologs. *Molecular Biology and Evolution*, 31(10):2753–2769, 2014.

[5] JD Bloom. Identification of positive selection in genes is greatly improved by using experimentally infromed site-specific models. *Biology Direct*, 12:1, 2017.

[6] T Brun, J Peduzzi, MM Canica, G Paul, P Nevot, M Barthelemy, and R Labia. Characterization and amino acid sequence of irt-4, a novel tem-type enzyme with a decreased susceptibility to beta-lactamase inhibitors. *FEMS Microbiology Letters*, 120:111–117, 1994.

[7] C Chanal, MC Poupart, D Sirot, R Labia, J Sirot, and R Cluzel. Nucleotide sequences of

caz-2, caz-6, and caz-7 beta-lactamase genes. *Antimicrob. Agents Chemother.*, 36:1817–1820, 1992.

[8] APJ de Koning and BD De Sanctis. The rate of molecular evolution when mutation may not be weak. *bioRxiv*, 2018.

[9] J Echave, SJ Spielman, and CO Wilke. Causes of evolutionary rate variation among protein sites. *Nature Reviews Genetics*, 17:109–121, 2016.

[10] J Felsenstein. Evolutionary trees from dna sequences: a maximum likelihood approach. *Journal of Molecular Evolution*, 17:368–376, 1981.

[11] J Felsenstein. Taking variation of evolutionary rates between sites into account in inferring phylogenies. *Journal of Molecular Evolution*, 53(4):447–455, 2001.

[12] E Firnberg, JW Labonte, JJ Gray, and M Ostermeier. A comprehensive, high-resolution map of a gene's fitness landscape. *Molecular Biology and Evolution*, 31(6):1581–1592, 2014.

[13] E Firnberg and M Ostermeier. Pfunkel: Efficient, expansive, user-defined mutagenesis. *PLOS ONE*, 7(12):e52031, 2012.

[14] DM Fowler and S Fields. Deep mutational scanning: a new style of protein science. *Nature Methods*, 11:801–807, 2014.

[15] DM Fowler, JJ Stephany, and S Fields. Measuring the activity of protein variants on a large scale using deep mutational scanning. *Nature Protocols*, 9:2267–2284, 2014.

[16] MA Gilchrist. Combining models of protein translation and population genetics to predict protein production rates from codon usage patterns. *Molecular Biology and Evolution*, 24(11):2362–2372, 2007.

[17] DT Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, 1976.

[18] T Gojobori. Codon substitution in evolution and the "saturation" of synonymous changes. *Genetics*, 105:1011–1027, 1983.

[19] N. Goldman and Z. H. Yang. Codon-based model of nucleotide substitution for protein-coding DNA-sequences. *Molecular Biology and Evolution*, 11:725–736, 1994.

[20] S Goussard, W Sougakoff, C Mabilat, A Bauernfeind, and P Courvalin. An is1-like element is responsible for high-level synthesis of extended-spectrum beta-lactamase tem-6 in enterobacteriaceae. *J. Gen. Microbiol.*, 137:2681–2687, 1991.

[21] R Grantham. Amino acid differences formula to help explain protein evolution. *Science*, 185(4154):862–864, 1974.

[22] AL Halpern and WJ Bruno. Evolutionary distances for protein-coding sequences: Modeling site-specific residue frequencies. *Molecular Biology and Evolution*, 15(7):910–917, 1998.

[23] DL Hartl, EN Moriyama, and SA Sawyer. Selection intensity for codon bias. *Genetics*, 138:227–234, 1994.

[24] SK Hilton, MB Doud, and JD Bloom. phydms: software for phylogenetic analyses informed by deep mutation scanning. *PeerJ*, 5:e3657, 2017.

[25] MT Holder, DJ Zwickl, and C Dessimoz. Evaluating the robustness of phylogenetic methods to among-site variability in substitution processes. *Philos Trans R Soc Lond B*, 363:4013–4021, 2008.

[26] PC Jain and R Varadarajan. A rapid, efficient, and economical inverse polymerase chain reaction-based method for generating a site saturation mutant library. *Analytical Biochemistry*, 449:90–981, 2014.

[27] S Kawashima, P Pokarowski, M Pokarowska, A Kolinski, T Katayama, and M Kanehisa.

Aaindex: amino acid index database, progress report 2008. *Nucleic Acids Research*, 36:D202–D205, 2008.

[28] Carolin Kosiol, Ian Holmes, and Nick Goldman. An empirical codon model for protein sequence evolution. *Molecular Biology and Evolution*, 24(7):1464–1479, Jul 2007.

[29] N Lartillot and H Philippe. A bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution*, 21:1095–1109, 2004.

[30] SQ Le, N Lartillot, and Gascuel O. Phylogenetic mixture models for proteins. *Philos Trans R Soc Lond B Biol Sci*, 363:3965–3976, 2008.

[31] C Mabilat, J Lourencao-Vital, S Goussard, and P Courvalin. A new example of physical linkage between tn1 and tn21: the antibiotic multiple-resistance region of plasmid pcff04 encoding extended-spectrum beta-lactamase tem-3. *Mol Gen Genet*, 235:113–121, 1992.

[32] A Matagne, J Lamotte-Brasseur, and JM Frere. Catalytic properties of class a beta-lactamases: efficiency and diversity. *Biochemistry Journal*, 300:581–598, 1998.

[33] SV Muse and BS Gaut. A likelihood approach for comparing synonymous and non-synonymous nucleotide substitution rates, with application to the chloroplast genome. *Molecular Biology and Evolution*, 11(5):715–724, 1994.

[34] LT Nguyen, HA Schmidt, A von Haeseler, and BQ Minh. Iq-tree: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1):268–274, 2015.

[35] H Ochman and AC Wilson. *Evolutionary history of enteric bacterian*, pages 1649–1654. ASM Press, 1987.

[36] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.

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

[38] N Rodrigue and N Lartillot. Site-heterogeneous mutation-selection models within the phylobayes-mpi package. *Bioinformatics*, 30:1020–1021, 2014.

[39] N Rodrigue, H Philippe, and N Lartillot. Mutation-selection models of coding sequence evolution with site-heterogeneous amino acid fitness profiles. *Proccedings of the National Academy of Sciences U.S.A*, 107:4629–4634, 2010.

[40] G Sella and AE Hirsh. The application of statistical physics to evolutionary biology. *Proceedings of the National Academy of Sciences of the United States of America*, 102:9541–9546, 2005.

[41] W Sougakoff, S Goussard, and P Courvalin. The tem-3 beta-lactamase, which hydrolyzes broad-spectrum cephalosporins, is derived from the tem-2 penicillinase by two amino acid substitutions. *FEMS Microbiology Letters*, 56:343–348, 1988.

[42] W Sougakoff, A Petit, S Goussard, D Sirot, A Bure, and P Courvalin. Characterization of the plasmid genes blat-4 and blat-5 which encode the broad-spectrum beta-lactamases tem-4 and tem-5 in enterobacteriaceae. *Gene*, 78:339–348, 1989.

[43] MA Stiffler, DR Hekstra, and Ranganathan R. Evolvability as a function of purifying selection in tem-1 β-lactamase. *Cell*, 160:882–892, 2016.

[44] AU Tamuri, N Goldman, and M dos Reis. A penalized likelihood method for estimating the distribution of selection coefficients from phylogenetic data. *Genetics*, 197:257–271, 2014.

[45] JL Thorne, N Goldman, and DT Jones. Combinng protein evolution and secondary structure. *Molecular Biology and Evolution*, 13:666–673, 1996.

[46] B Thyagarajan and JD Bloom. The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. *eLife*, 3:e03300, 2014.

[47] HC Wang, K Li, E Susko, and AJ Roger. A class frequency mixture model that adjusts for site-specific amino acid frequencies and improves inference of protein phylogeny. *BMC Evolutionary Biology*, 8:331, 2008.

[48] CH Wu, MA Suchard, and AJ Drummond. Bayesian selection of nucleotide substitution models and their site assignments. *Molecular Biology and Evolution*, 30:669–688, 2013.

[49] ZH Yang. Maximum-likelihood phylogenetic estimation from DNA-sequences with variable rates over sites - approximate methods. *Journal of Molecular Evolution*, 39:306–314, 1994.

[50] A Zharkikh. Estimation of evolutionary distances between nucleotide sequences. *Journal of Molecular Evolution*, 39(3):315–329, 1994.

Supplementary Material

No.	Model	LnL	n	AIC	ΔAIC	AICc	ΔAICc
1	*SelAC*+DMS +G4	-1768	111	3758	14	3760	0
2	*SelAC* +G4	-1498	374	3744	0	3766	6
3	*phydms*	-2060.85	102	4326	582	4328	568
4	SYM+R2	-2229.616	102	4663.232	919.232	4693.862	933.862
5	TIMe+R2	-2232.406	100	4664.811	920.811	4694.172	934.172
6	TVMe+R2	-2232.838	101	4667.677	923.677	4697.668	937.668
7	TIM3e+R2	-2234.332	100	4668.664	924.664	4698.024	938.024
8	TIM2e+R2	-2234.381	100	4668.763	924.763	4698.123	938.123
9	K3P+R2	-2235.777	99	4669.553	925.553	4698.291	938.291
10	TNe+R2	-2236.078	99	4670.155	926.155	4698.892	938.892
11	SYM+R3	-2229.616	104	4667.232	923.232	4699.162	939.162
12	TIM+F+R2	-2230.958	103	4667.915	923.915	4699.191	939.191
13	TIMe+R3	-2232.404	102	4668.808	924.808	4699.437	939.437
14	GTR+F+R2	-2228.537	105	4667.073	923.073	4699.665	939.665
15	K3Pu+F+R2	-2232.617	102	4669.234	925.234	4699.864	939.864
16	TVM+F+R2	-2230.105	104	4668.21	924.21	4700.14	940.14
17	TVMe+R3	-2232.838	103	4671.676	927.676	4702.952	942.952
18	K2P+R2	-2239.424	98	4674.847	930.847	4702.969	942.969
19	TIM3e+R3	-2234.332	102	4672.664	928.664	4703.293	943.293
20	TIM2e+R3	-2234.381	102	4672.762	928.762	4703.391	943.391
21	TIM3+F+R2	-2233.064	103	4672.127	928.127	4703.403	943.403
22	TIM2+F+R2	-2233.114	103	4672.227	928.227	4703.503	943.503
23	K3P+R3	-2235.777	101	4673.553	929.553	4703.545	943.545
24	TN+F+R2	-2234.624	102	4673.249	929.249	4703.878	943.878
25	TPM3u+F+R2	-2234.673	102	4673.347	929.347	4703.977	943.977
26	TPM3+F+R2	-2234.674	102	4673.348	929.348	4703.978	943.978
27	TPM2u+F+R2	-2234.681	102	4673.363	929.363	4703.993	943.993
28	TPM2+F+R2	-2234.683	102	4673.365	929.365	4703.995	943.995
29	TNe+R3	-2236.077	101	4674.155	930.155	4704.146	944.146
30	TIM+F+R3	-2230.958	105	4671.915	927.915	4704.507	944.507
31	HKY+F+R2	-2236.266	101	4674.531	930.531	4704.522	944.522
32	GTR+F+R3	-2228.536	107	4671.073	927.073	4705.011	945.011
33	K3Pu+F+R3	-2232.617	104	4673.234	929.234	4705.163	945.163
34	TVM+F+R3	-2230.105	106	4672.21	928.21	4705.471	945.471
35	K2P+R3	-2239.192	100	4678.384	934.384	4707.745	947.745
36	TIM3+F+R3	-2233.063	105	4676.127	932.127	4708.718	948.718
37	TIM2+F+R3	-2233.113	105	4676.227	932.227	4708.818	948.818
38	TN+F+R3	-2234.624	104	4677.249	933.249	4709.178	949.178
39	TPM3u+F+R3	-2234.673	104	4677.347	933.347	4709.277	949.277

40	TPM3+F+R3	-2234.674	104	4677.348	933.348	4709.277	949.277
41	TPM2u+F+R3	-2234.681	104	4677.363	933.363	4709.293	949.293
42	TPM2+F+R3	-2234.682	104	4677.364	933.364	4709.294	949.294
43	HKY+F+R3	-2236.074	103	4678.148	934.148	4709.424	949.424
44	SYM+I+G4	-2243.212	102	4690.424	946.424	4721.054	961.054
45	TVMe+I+G4	-2244.533	101	4691.066	947.066	4721.057	961.057
46	TIMe+I+G4	-2246.457	100	4692.914	948.914	4722.275	962.275
47	K3P+I+G4	-2248.166	99	4694.332	950.332	4723.069	963.069
48	TVM+F+I+G4	-2241.853	104	4691.707	947.707	4723.636	963.636
49	TIM3e+I+G4	-2247.379	100	4694.758	950.758	4724.119	964.119
50	K3Pu+F+I+G4	-2245.156	102	4694.311	950.311	4724.941	964.941
51	GTR+F+I+G4	-2241.484	105	4692.968	948.968	4725.559	965.559
52	TIM+F+I+G4	-2244.418	103	4694.836	950.836	4726.112	966.112
53	TPM3u+F+I+G4	-2246.03	102	4696.06	952.06	4726.69	966.69
54	TPM3+F+I+G4	-2246.069	102	4696.138	952.138	4726.768	966.768
55	TIM2e+I+G4	-2248.934	100	4697.868	953.868	4727.228	967.228
56	TNe+I+G4	-2250.587	99	4699.174	955.174	4727.911	967.911
57	TIM3+F+I+G4	-2245.534	103	4697.068	953.068	4728.344	968.344
58	K2P+I+G4	-2252.181	98	4700.362	956.362	4728.484	968.484
59	TPM2u+F+I+G4	-2247.579	102	4699.158	955.158	4729.788	969.788
60	TPM2+F+I+G4	-2247.685	102	4699.371	955.371	4730	970
61	HKY+F+I+G4	-2249.065	101	4700.13	956.13	4730.121	970.121
62	TIM2+F+I+G4	-2247.009	103	4700.018	956.018	4731.294	971.294
63	TN+F+I+G4	-2248.511	102	4701.023	957.023	4731.652	971.652
64	TVMe+I	-2254.804	100	4709.608	965.608	4738.968	978.968
65	K3P+I	-2257.72	98	4711.439	967.439	4739.561	979.561
66	SYM+I	-2254.11	101	4710.221	966.220	4740.212	980.212
67	TIMe+I	-2257.074	99	4712.149	968.149	4740.886	980.886
68	TVM+F+I	-2252.157	103	4710.315	966.315	4741.591	981.591
69	K3Pu+F+I	-2254.856	101	4711.712	967.712	4741.704	981.704
70	TIM3e+I	-2257.796	99	4713.592	969.592	4742.33	982.33
71	TPM3+F+I	-2255.771	101	4713.543	969.543	4743.534	983.534
72	TPM3u+F+I	-2255.771	101	4713.543	969.543	4743.534	983.534
73	K2P+I	-2261.218	97	4716.436	972.436	4743.949	983.949
74	GTR+F+I	-2252.067	104	4712.133	968.133	4744.063	984.063
75	TIM+F+I	-2254.783	102	4713.566	969.566	4744.195	984.195
76	TNe+I	-2260.579	98	4717.158	973.158	4745.28	985.28
77	TIM3+F+I	-2255.684	102	4715.368	971.368	4745.998	985.998
78	HKY+F+I	-2258.352	100	4716.703	972.703	4746.064	986.064
79	TIM2e+I	-2259.878	99	4717.757	973.757	4746.494	986.494
80	TVMe+G4	-2258.853	100	4717.705	973.705	4747.066	987.066
81	SYM+G4	-2257.573	101	4717.146	973.146	4747.137	987.137
82	TPM2+F+I	-2257.712	101	4717.423	973.423	4747.415	987.415
83	TPM2u+F+I	-2257.712	101	4717.423	973.423	4747.415	987.415
84	K3P+G4	-2261.922	98	4719.844	975.844	4747.966	987.966

85	TIMe+G4	-2260.683	99	4719.365	975.365	4748.103	988.103
86	TN+F+I	-2258.28	101	4718.561	974.561	4748.552	988.552
87	TIM3e+G4	-2261.255	99	4720.51	976.51	4749.247	989.247
88	TVM+F+G4	-2256.108	103	4718.216	974.216	4749.492	989.492
89	TIM2+F+I	-2257.643	102	4719.286	975.286	4749.915	989.915
90	K3Pu+F+G4	-2258.971	101	4719.941	975.941	4749.933	989.933
91	TPM3u+F+G4	-2259.716	101	4721.433	977.433	4751.424	991.424
92	TPM3+F+G4	-2259.717	101	4721.434	977.434	4751.425	991.425
93	GTR+F+G4	-2255.75	104	4719.5	975.5	4751.43	991.43
94	TIM+F+G4	-2258.638	102	4721.276	977.276	4751.906	991.906
95	K2P+G4	-2265.454	97	4724.907	980.907	4752.421	992.421
96	TNe+G4	-2264.219	98	4724.437	980.437	4752.559	992.559
97	TIM3+F+G4	-2259.366	102	4722.732	978.732	4753.361	993.361
98	TIM2e+G4	-2263.57	99	4725.141	981.141	4753.878	993.878
99	JC+R2	-2266.233	97	4726.466	982.466	4753.98	993.98
100	F81+F+R2	-2262.327	100	4724.654	980.654	4754.015	994.015
101	HKY+F+G4	-2262.499	100	4724.999	980.999	4754.359	994.359
102	TPM2+F+G4	-2261.915	101	4725.829	981.829	4755.82	995.82
103	TPM2u+F+G4	-2261.915	101	4725.829	981.829	4755.82	995.82
104	TN+F+G4	-2262.169	101	4726.338	982.338	4756.329	996.329
105	TIM2+F+G4	-2261.585	102	4727.17	983.17	4757.8	997.8
106	F81+F+R3	-2262.028	102	4728.056	984.056	4758.685	998.685
107	JC+R3	-2265.997	99	4729.994	985.994	4758.731	998.731
108	F81+F+I+G4	-2274.845	100	4749.69	1005.69	4779.05	1019.05
109	JC+I+G4	-2279.318	97	4752.636	1008.636	4780.149	1020.149
110	F81+F+I	-2283.56	99	4765.119	1021.119	4793.857	1033.857
111	JC+I	-2287.984	96	4767.968	1023.968	4794.881	1034.881
112	F81+F+G4	-2287.834	99	4773.669	1029.669	4802.406	1042.406
113	JC+G4	-2292.095	96	4776.19	1032.19	4803.103	1043.103
114	*GY94* +F1X4+R2	-2242.963	102	4689.926	945.926	4821.251	1061.251
115	MGK+F1X4+R2	-2243.111	102	4690.221	946.221	4821.546	1061.546
116	*GY94* +F1X4+R3	-2238.022	104	4684.043	940.043	4822.271	1062.271
117	MGK+F3X4+R2	-2229.923	108	4675.846	931.846	4828.729	1068.729
118	*GY94* +F1X4+I+G4	-2247.179	102	4698.359	954.359	4829.684	1069.684
119	MGK+F1X4+I+G4	-2247.292	102	4698.583	954.583	4829.908	1069.908
120	MGK+F1X4+R3	-2241.989	104	4691.978	947.978	4830.206	1070.206
121	MGK+F3X4+R3	-2224.78	110	4669.559	925.559	4830.217	1070.217
122	*GY94* +F1X4+G4	-2251.144	101	4704.287	960.287	4832.263	1072.263
123	MGK+F1X4+G4	-2251.472	101	4704.944	960.944	4832.919	1072.919
124	*GY94* +F3X4+R3	-2227.048	110	4674.096	930.096	4834.754	1074.754
125	*GY94* +F3X4+R2	-2233.068	108	4682.136	938.136	4835.019	1075.019
126	MGK+F3X4+I+G4	-2233.539	108	4683.078	939.0781	4835.962	1075.962
127	MGK+F3X4+G4	-2237.512	107	4689.024	945.024	4838.134	1078.134
128	*GY94* +F3X4+I+G4	-2238.243	108	4692.485	948.485	4845.368	1085.368
129	*GY94* +F3X4+R4	-2227.106	112	4678.213	934.213	4846.96	1086.96

130	*GY94* +F3X4+G4	-2242.394	107	4698.789	954.789	4847.899	1087.899
131	*GY94* +F1X4+I	-2260.085	101	4722.169	978.169	4850.144	1090.144
132	MGK+F1X4+I	-2260.345	101	4722.69	978.69	4850.665	1090.665
133	MGK+F3X4+I	-2246.112	107	4706.225	962.225	4855.335	1095.335
134	MG+F1X4+R2	-2268.482	101	4738.963	994.963	4866.938	1106.938
135	*GY94* +F3X4+I	-2252.532	107	4719.064	975.064	4868.174	1108.174
136	MG+F3X4+R2	-2254.453	107	4722.906	978.906	4872.015	1112.015
137	MG+F1X4+I+G4	-2272.057	101	4746.113	1002.113	4874.089	1114.089
138	MG+F1X4+R3	-2267.523	103	4741.047	997.047	4875.789	1115.789
139	MG+F1X4+G4	-2276.171	100	4752.342	1008.342	4877.033	1117.033
140	MG+F3X4+I+G4	-2257.945	107	4729.891	985.891	4879.001	1119.001
141	MG+F3X4+G4	-2261.949	106	4735.898	991.898	4881.309	1121.309
142	MG+F3X4+R3	-2253.514	109	4725.027	981.027	4881.759	1121.759
143	SYM	-2329.878	100	4859.756	1115.756	4889.116	1129.116
144	TIMe	-2333.105	98	4862.21	1118.21	4890.332	1130.332
145	TIM3e	-2333.481	98	4862.961	1118.961	4891.083	1131.083
146	TVMe	-2333.164	99	4864.328	1120.328	4893.065	1133.065
147	GTR+F	-2328.404	103	4862.809	1118.809	4894.085	1134.085
148	K3P	-2336.391	97	4866.783	1122.783	4894.297	1134.297
149	MG+F1X4+I	-2284.946	100	4769.892	1025.892	4894.583	1134.583
150	TVM+F	-2330.086	102	4864.172	1120.172	4894.802	1134.802
151	TIM+F	-2331.48	101	4864.96	1120.96	4894.952	1134.952
152	TNe	-2336.729	97	4867.458	1123.458	4894.972	1134.972
153	K3Pu+F	-2333.162	100	4866.323	1122.323	4895.684	1135.684
154	TIM3+F	-2331.971	101	4865.942	1121.942	4895.934	1135.934
155	TPM3+F	-2333.648	100	4867.297	1123.297	4896.657	1136.657
156	TPM3u+F	-2333.648	100	4867.297	1123.297	4896.657	1136.657
157	TIM2e	-2336.292	98	4868.584	1124.584	4896.706	1136.706
158	MG+F3X4+I	-2270.442	106	4752.885	1008.885	4898.295	1138.295
159	K2P	-2340.015	96	4872.03	1128.03	4898.943	1138.943
160	TN+F	-2335.102	100	4870.204	1126.204	4899.565	1139.565
161	HKY+F	-2336.783	99	4871.566	1127.566	4900.303	1140.303
162	TIM2+F	-2334.7	101	4871.401	1127.401	4901.392	1141.392
163	TPM2u+F	-2336.381	100	4872.761	1128.761	4902.122	1142.122
164	TPM2+F	-2336.381	100	4872.762	1128.762	4902.123	1142.123
165	JC	-2366.286	95	4922.571	1178.571	4948.892	1188.892
166	F81+F	-2362.554	98	4921.108	1177.108	4949.229	1189.229
167	*GY94* +F1X4	-2315.788	100	4831.575	1087.575	4956.267	1196.267
168	KOSI07+FU+R2	-2325.725	97	4845.45	1101.45	4960.675	1200.675
169	MGK+F1X4	-2318.048	100	4836.095	1092.095	4960.787	1200.787
170	KOSI07+FU+R3	-2323.063	99	4844.126	1100.126	4965.599	1205.599
171	MGK+F3X4	-2304.357	106	4820.713	1076.713	4966.124	1206.124
172	*GY94* +F3X4	-2306.17	106	4824.339	1080.339	4969.749	1209.749
173	KOSI07+FU+I+G4	-2335.554	97	4865.108	1121.108	4980.332	1220.332
174	KOSI07+FU+G4	-2339.513	96	4871.026	1127.026	4983.218	1223.218

175	KOSI07+F3X4+R2	-2315.814	106	4843.627	1099.627	4989.038	1229.038
176	KOSI07+F3X4+R3	-2310.509	108	4837.018	1093.018	4989.901	1229.901
177	KOSI07+F1X4+R2	-2333.491	100	4866.983	1122.983	4991.674	1231.674
178	KOSI07+F1X4+R3	-2328.692	102	4861.383	1117.383	4992.708	1232.708
179	SCHN05+FU+R2	-2344.705	97	4883.411	1139.411	4998.635	1238.635
180	KOSI07+F1X4+I+G4	-2337.965	100	4875.93	1131.93	5000.621	1240.621
181	KOSI07+F1X4+G4	-2341.156	99	4880.312	1136.312	5001.784	1241.784
182	SCHN05+FU+R3	-2341.179	99	4880.358	1136.358	5001.831	1241.831
183	KOSI07+FU+I	-2349.617	96	4891.233	1147.233	5003.426	1243.426
184	KOSI07+F3X4+I+G4	-2323.767	106	4859.534	1115.534	5004.944	1244.944
185	MG+F1X4	-2342.797	99	4883.593	1139.593	5005.065	1245.065
186	KOSI07+F3X4+G4	-2327.376	105	4864.751	1120.751	5006.534	1246.534
187	MG+F3X4	-2328.539	105	4867.078	1123.078	5008.861	1248.861
188	SCHN05+F1X4+R3	-2340.927	102	4885.854	1141.854	5017.179	1257.179
189	KOSI07+F1X4+I	-2349.1	99	4896.2	1152.2	5017.672	1257.672
190	SCHN05+F3X4+R3	-2324.472	108	4864.944	1120.944	5017.827	1257.827
191	SCHN05+FU+I+G4	-2354.523	97	4903.046	1159.046	5018.27	1258.27
192	SCHN05+F1X4+R2	-2348.226	100	4896.452	1152.452	5021.143	1261.143
193	SCHN05+F3X4+R2	-2331.916	106	4875.833	1131.833	5021.243	1261.243
194	SCHN05+FU+G4	-2358.682	96	4909.365	1165.365	5021.558	1261.558
195	KOSI07+F3X4+I	-2336.826	105	4883.653	1139.653	5025.436	1265.436
196	SCHN05+F1X4+I+G4	-2351.096	100	4902.192	1158.192	5026.883	1266.883
197	SCHN05+F1X4+G4	-2353.895	99	4905.79	1161.79	5027.263	1267.263
198	SCHN05+F1X4+R4	-2340.593	104	4889.187	1145.187	5027.414	1267.414
199	SCHN05+F3X4+R4	-2324.102	110	4868.203	1124.203	5028.861	1268.861
200	SCHN05+F3X4+I+G4	-2338.345	106	4888.69	1144.69	5034.101	1274.101
201	SCHN05+F3X4+G4	-2341.811	105	4893.621	1149.621	5035.404	1275.404
202	SCHN05+FU+I	-2370.471	96	4932.943	1188.943	5045.135	1285.135
203	SCHN05+F1X4+I	-2363.696	99	4925.391	1181.391	5046.864	1286.864
204	SCHN05+F3X4+I	-2352.81	105	4915.621	1171.621	5057.404	1297.404
205	KOSI07+FU	-2394.782	95	4979.563	1235.563	5088.785	1328.785
206	KOSI07+F1X4	-2398.44	98	4992.88	1248.88	5111.197	1351.197
207	KOSI07+F3X4	-2383.159	104	4974.318	1230.318	5112.546	1352.546
208	SCHN05+FU	-2419.333	95	5028.665	1284.665	5137.887	1377.887
209	SCHN05+F1X4	-2416.544	98	5029.088	1285.088	5147.405	1387.405
210	SCHN05+F3X4	-2402.838	104	5013.675	1269.675	5151.903	1391.903
211	*GY94* +F+R2	-2208.59	159	4735.181	991.181	5229.161	1469.161
212	*GY94* +F+G4	-2217.694	158	4751.388	1007.388	5234.504	1474.504
213	*GY94* +F+I+G4	-2213.659	159	4745.319	1001.319	5239.299	1479.299
214	*GY94* +F+R3	-2202.599	161	4727.198	983.198	5243.673	1483.673
215	*GY94* +F+I	-2228.346	158	4772.691	1028.691	5255.807	1495.807
216	*GY94* +F+R4	-2202.61	163	4731.219	987.219	5271.26	1511.26
217	*GY94* +F	-2282.254	157	4878.509	1134.509	5351.004	1591.004
218	KOSI07+F+R2	-2291.643	157	4897.286	1153.286	5369.781	1609.781
219	KOSI07+F+G4	-2301.662	156	4915.325	1171.325	5377.438	1617.438

220	KOSI07+F+I+G4	-2298.418	157	4910.835	1166.835	5383.33	1623.33
221	KOSI07+F+R3	-2286.723	159	4891.446	1147.446	5385.426	1625.426
222	KOSI07+F+I	-2311.78	156	4935.559	1191.559	5397.672	1637.672
223	SCHN05+F+R2	-2310.015	157	4934.03	1190.03	5406.525	1646.525
224	SCHN05+F+G4	-2316.684	156	4945.369	1201.369	5407.482	1647.482
225	SCHN05+F+I+G4	-2313.733	157	4941.467	1197.467	5413.962	1653.962
226	SCHN05+F+R3	-2303.732	159	4925.463	1181.463	5419.444	1659.444
227	SCHN05+F+I	-2327.127	156	4966.254	1222.254	5428.367	1668.367
228	SCHN05+F+R4	-2303.45	161	4928.9	1184.9	5445.375	1685.375
229	KOSI07+F	-2357.579	155	5025.157	1281.157	5477.12	1717.12
230	SCHN05+F	-2379.264	155	5068.528	1324.528	5520.491	1760.491

Table S1: Model selection of 230 models of nucleotide and codon evolution.

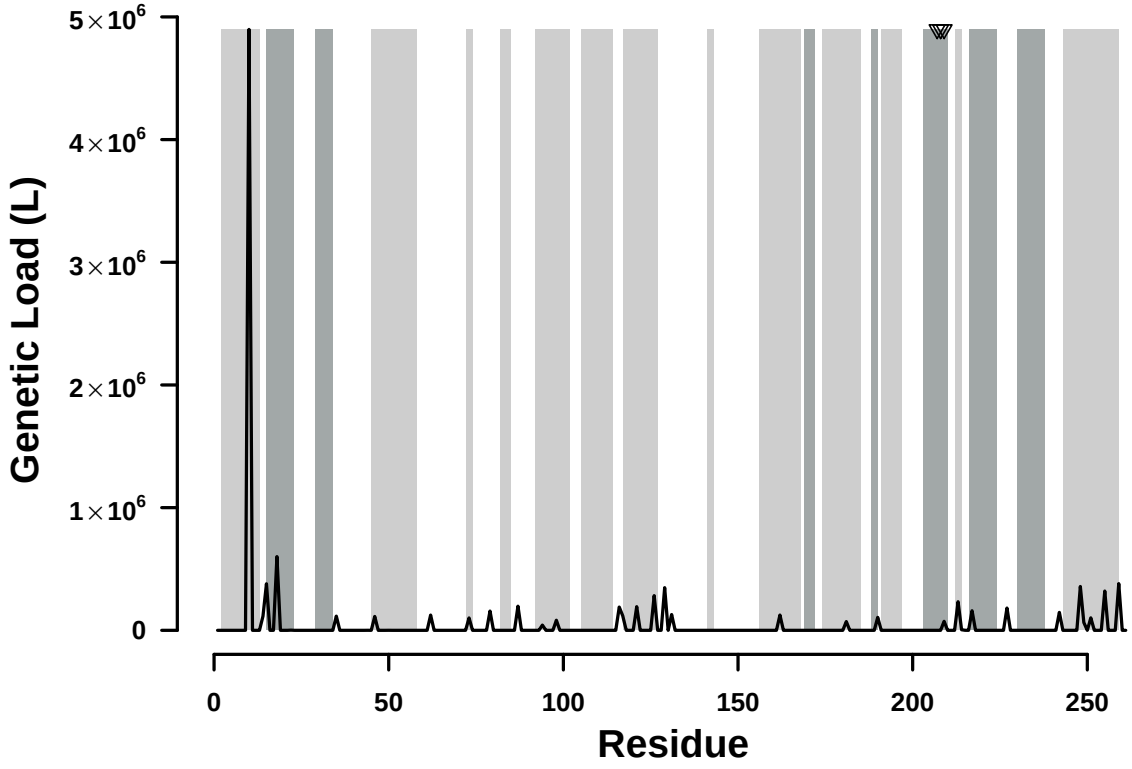

Figure S1: Distribution of genetic load in SHV. Average genetic load over all observed SHV variants is indicated by the black line. Light gray bars indicate where helices are found, and dark gray bars indicate β-sheets. The three residues forming the active sites are indicated by three triangles at the top of the plot.

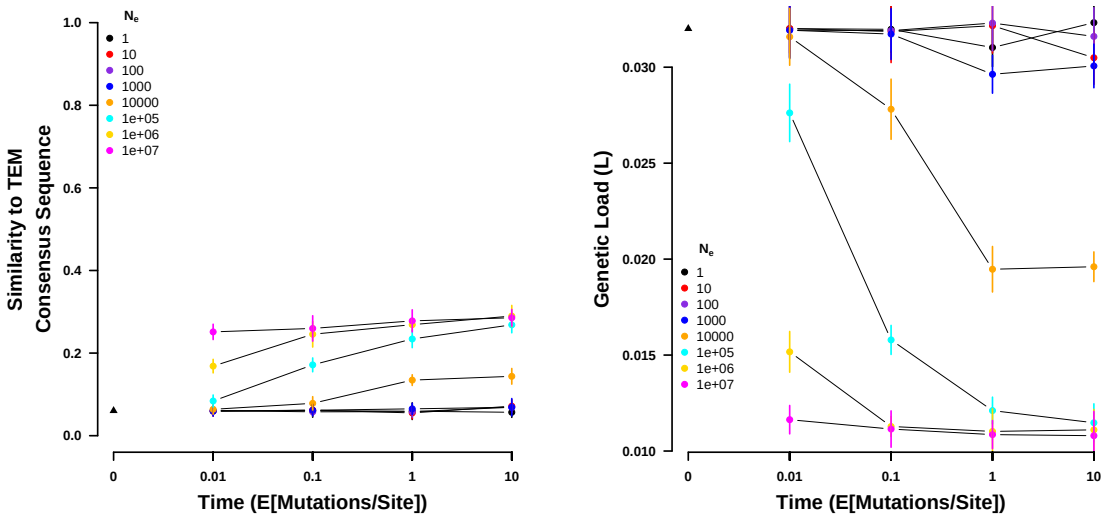

Figure S2: Sequences simulated from a random codon sequence under the site specific selection on amino acids estimated using *SelAC*. (left) Sequence similarity to the observed consensus sequence at various times for a range on values of N_e. (right) Genetic load of the simulated sequences at various times for a range on values of N_e. Time is given in number of expected mutations per site, which equals the substitution rate of a neutral mutation. Points indicate sample means and vertical bars indicate standard deviations. Initial sequence is the inferred ancestral state of the TEM variants and indicated by a black triangle.

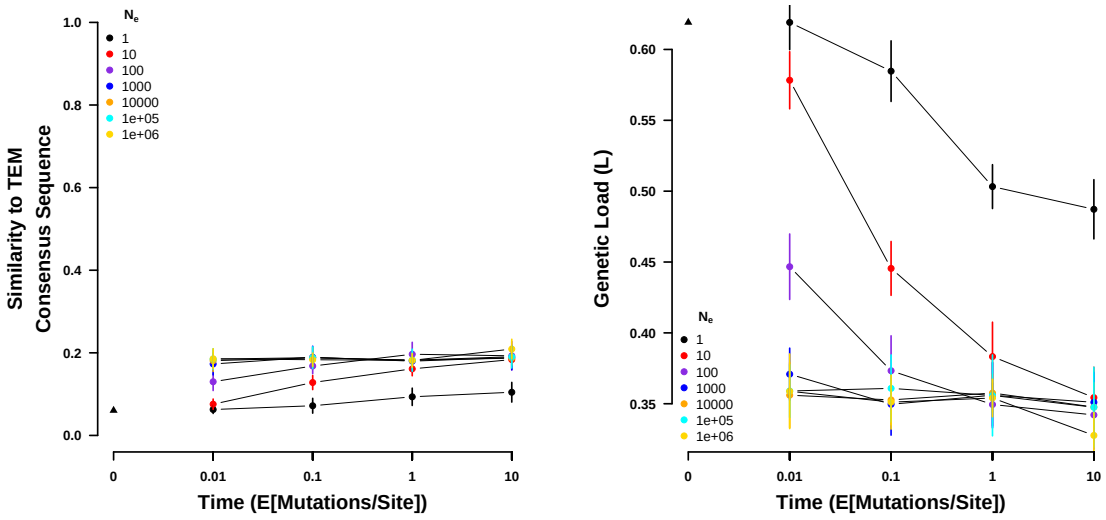

Figure S3: Sequences simulated from a random codon sequence under the site specific selection on amino acids estimated using deep mutation scanning. (left) Sequence similarity to the observed consensus sequence at various times for a range on values of N_e. (right) Genetic load of the simulated sequences at various times for a range on values of N_e. Time is given in number of expected mutations per site, which equals the substitution rate of a neutral mutation. Points indicate sample means and vertical bars indicate standard deviations. Initial sequence is the inferred ancestral state of the TEM variants and indicated by a black triangle.

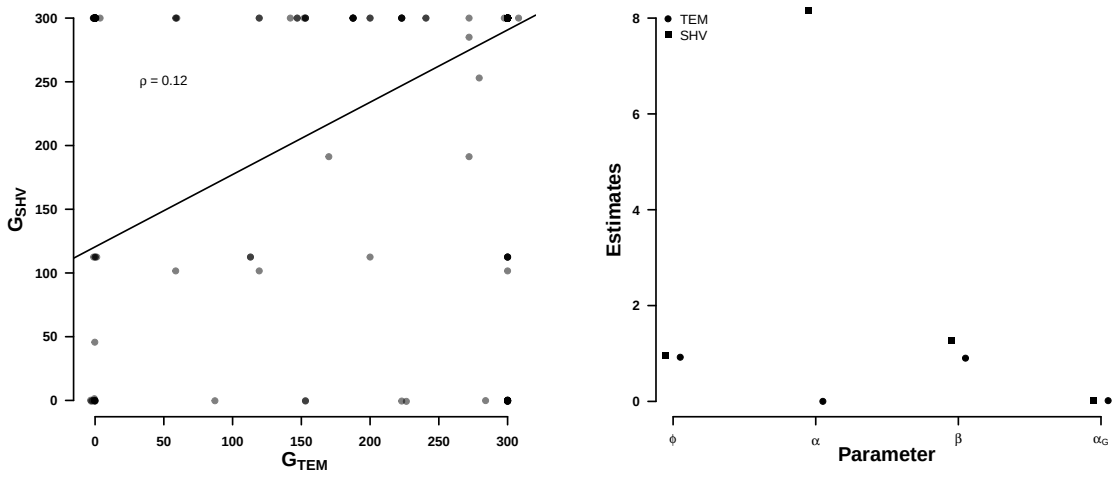

Figure S4: Comparison of selection related parameters between TEM and SHV. (left) Estimated site specific efficacy of selection G. (right) Selection related parameter estimates. Protein functionality production rate ψ, physicochemical weight for amino acid composition α_c, physicochemical weight for amino acid polarity α_p, and the parameter describing the distribution of G, α_G estimated by *SelAC*.