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 9, 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.

Introduction

Incorporation of selection into phylogenetic frameworks has already been a long lasting endeavor. Early models focused the influence of selection on the substitution rate between a resident and a mutant [Goldman and Yang, 1994, Muse and Gaut, 1994, Thorne et al., 1996]. These models however, lack site specific equilibrium frequencies. The importance of site specific equilibrium frequencies has long been noted [Felsenstein, 1981, Gojobori, 1983]. Halpern and Bruno [1998] first introduced a framework to incorporate the site specific equilibrium frequencies of amino acids. However, they had to concede that their model was to parameter rich and therefore intractable for biological data sets without simplifying assumptions. More recent models that incorporate site specific equilibrium frequencies still require a large number of parameter to be estimated from the sequence data [Lartillot and Philippe, 2004, Le et al., 2008, HC et al., 2008, Holder et al., 2008, Wu et al., 2013, Tamuri et al., 2014]. Other approaches treat site specific selection as a random effect [Rodrigue et al., 2010, Rodrigue, 2013, Rodrigue and Lartillot, 2014]. A full parameterization 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 the site specific selection on amino acids [Bloom, 2014, Thyagarajan and Bloom, 2014, Bloom, 2017].

Deep mutation scanning (DMS) is often used to generate comprehensive fitness estimates of proteins [Fowler et al., 2014]. The quality of empirical estimates of site specific selection on amino acids from DMS depents on many factors, e.g. the initial library of mutants and the applied selection pressure.

Incorporating empirical estimates of site specific selection on amino acids has some important features. Individual amino acid site along show differences in evolutionary rates strong preferences for amino acids [Halpern and Bruno, 1998, Ashenberg et al., 2013, Echave et al., 2016] The usage of site specific selection acknowledges the heterogeneity in selection along the protein sequence [Hilton et al., 2017]. It reduces the number of parameters that have to be estimated from the data, making it applicable to smaller data sets and allowing

for more complex models. The incorporation of empirical estimates of selection does also have some shortcomings. The need for empirical estimates of selection limits the application to fast growing organisms that can be manipulated under laboratory conditions. This limits the application of experimentally informed models as many organims can not be cultivated under laboratory conditions or have a to long generation time.

Even in the cases were empirical estimates of site specific selection on amino acids can be optained their usefulness for phylogenetic reconstruction is not yet fullly clear. 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 slection on amino acids for the β-lactamase TEM from Stiffler et al. [2016]. We find that experimentally inferred selection does not adequatly reflect evolution in the wild. In contrast, *SelAC* a mechanistical phylogenetic model of stabilizing selection rooted in first princibles with site specific equilibrium frequencies improves model fit, and better reflects evolution in the wild [Beaulieu et al., in review]. *SelAC* does not require extensive laboratory estimates for site specific selection on amino acids. *SelAC* assumes that the distance of two amino acids in physicochemical space affects substitution probabilities and estimates only one discrete parameter per site, the optimal amino acid at a site. Therefore *SelAC* only requires 19 site specific parameters instead of $19 \times L$.

Results

Site Specific Stabilizing Selection on Amino Acids Improves Model Fit

We compared the models *phydms* [Hilton et al., 2017] and *SelAC*, models of stabilizing site specific amino acid selection, to 281 other codon and nucleotide models by fitting them to 49 sequences of the β-lactamase TEM. Models with site specific selection on amino acids improved model fits by 917 to 1483 AICc units over codon or nucleotide models without site

Model	log($\mathcal{L}$)	n	AIC	ΔAIC	AICc	ΔAICc
SelAC	-1498	374	3744	0	3766	6
SelAC +DMS	-1768	111	3758	14	3760	0
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. See full table for all 231 models

specific selection (Table 1). In addition, *SelAC* does outperform *phydms* by 560 to 566 AICc units.

SelAC utilizes a hierarchical model framework and estimates 263 site specific parameters, $\sim 5\%$ of the 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. 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. However, *SelAC* +DMS is favored by AICc. This is solely due to a decrease in the number of parameters estimated, as the log($\mathcal{L}$) decreases from -1498 to -1768 (Table 1). The number of parameter for *SelAC*, however, is reported conservatively as the number of unique site patterns in the TEM alignemnt is only 27 and thus the number of paramters would be 123. This however is likely an under estimate of the degrees of freedom and the true number of parameters remains unclear at this point.

Interestingely, the best codon model (*GY94*) [Goldman and Yang, 1994] is outperformed by a variety of nucleotide model e.g. *SYM* [Zharkikh, 1994]. This indicates that negative frequency dependent selection like it is modeled in *GY94* is not appropriate for TEM [Beaulieu et al., in review]. Figure 1 shows that the estimated phylogenetic trees shift from long terminal branches (*SelAC*) to longer internal branches (*phydms*, *GY94*). All models produce polytomies but their location differs along the phylogeny between models. The

largest polytomies appear in the experimentally informed phylogenies.

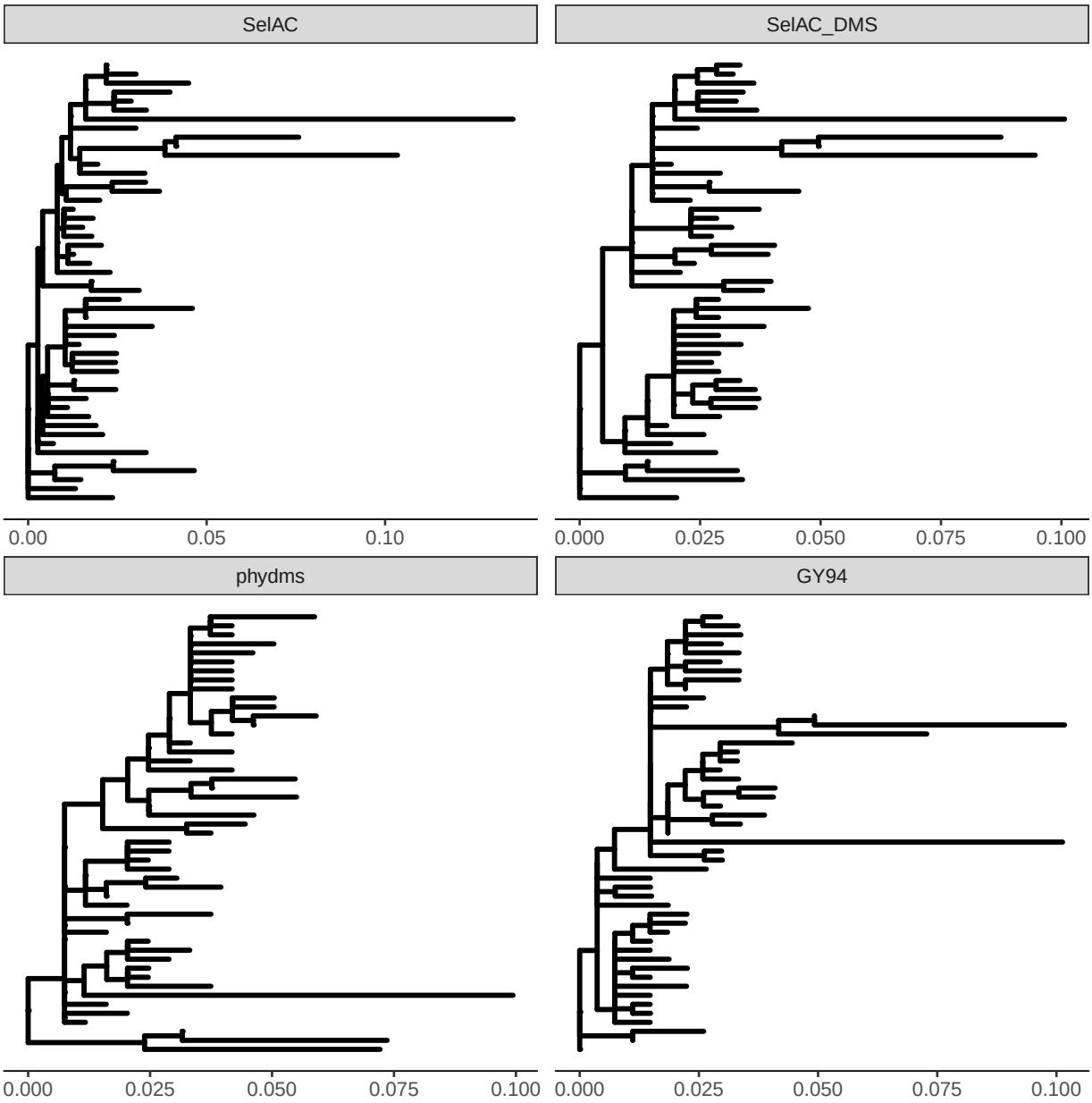

Figure 1: Phylogenies estimated using *SelAC*, *SelAC* +DMS, *phydms*, and *GY94*.

Laboratory Inferences of Selection are inconsistent with Observed Sequences.

Improved model fits with phydms are 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*.

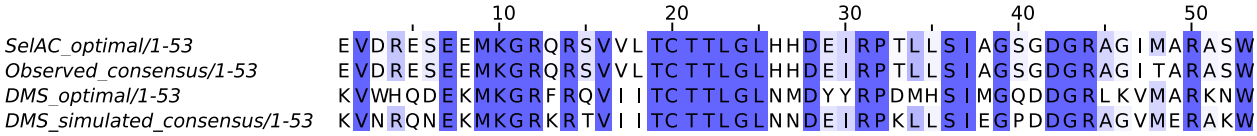

Figure 2: Every 5th residue. DMS and simulation based on DMS do not reflect natural sequences

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. Only when N_e is on the order of 10^0 drift is overpowering the efficacy of selection. With realistic values for $N_e = 10^7$, we find that the simulated sequences to show sequence similarity of 62% with the observed consensus sequence (Figure 3a). This is a higher similarity than the observed consensus sequence shows with the the sequence of selectively favored amino acids estimated using deep mutation scanning. The genetic load of the simulated sequences deacrease slowely with increasing N_e (Figure 3b). At time 1 and $N_e = 10^7$ the simulated sequences show a genetic load of 0.25, which is in contrast to the ~ 8 times higher observed load of 2.1. Thus it appears unlikely that the observed sequences have evolved under the experimentally inferred site specific selection for amino acids.

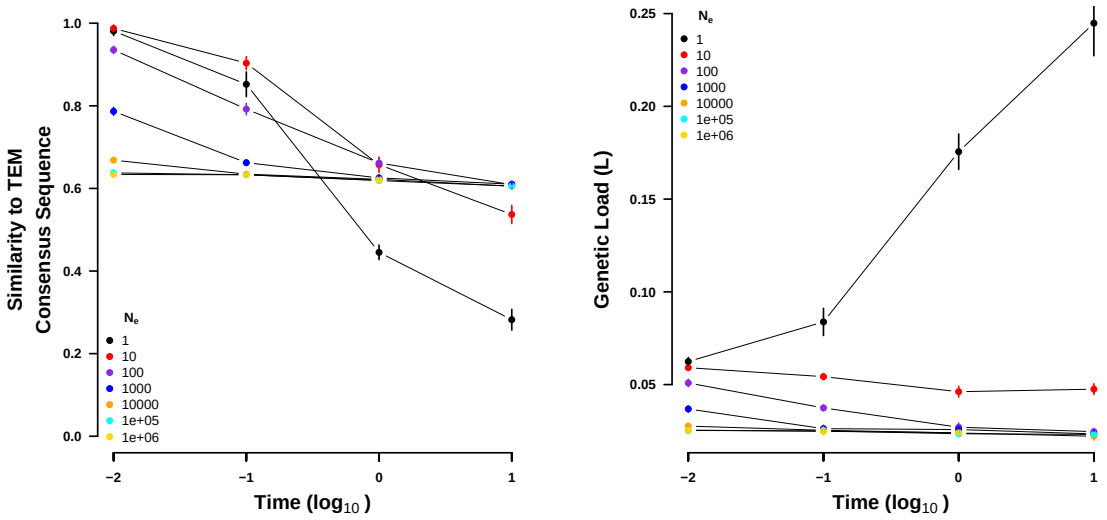

Figure 3: Sequences simulated from the ancestral state under the site specific selection on amino acids estimated using deep mutation scanning. (left) Sequence similarity to the observed consensus sequence at various times for a range on values of N_e. (right) Genetic load of the simulated sequences at various times for a range on values of N_e. Time is given in number of expected mutations. Points indicate sample means and vertical bars indicate standard deviations. Initial sequence is the inferred ancestral state of the TEM variants and not shown.

Stabilizing Selection for Optimal Physicochemical Probabilities increases Model Adequacy

We assessed model adequacy and find that *SelAC* better explains the observed TEM sequences. The observed consensus sequence has a very high sequence similarity with the sequence of selectively favored amino acids estimated by *SelAC* (99 %). Furthermore, assuming the site specific selection estimated by *SelAC*, the observed sequences only show a minimal genetic load (Table 2, Figure 5).

We simulated codon sequences forward in time for various length of time to assess the sequence similarity, assuming the *SelAC* inferred site specific selection for amino acids. 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 state was estimated to be the observed consensus sequence. For small N_e, we find that sequences drift away from the

observed consensus. In turn, the genetic load increases drastically. With increasing $N_e = 10^7$ the simulated sequences reach a sequence similarity at time 1 of 83%, this is in contrast to the observed sequence similarity 98%. We calculated the genetic load at this time of the simulated sequences to be 9.8×10^{-6} (Figure 4b). The genetic load of the observed sequences is estimated 4.2×10^{-5}, one order of magnitute higher. Thus, the simulated sequences show a lower genetic load despite the greater divergence from the observed consensus sequence.

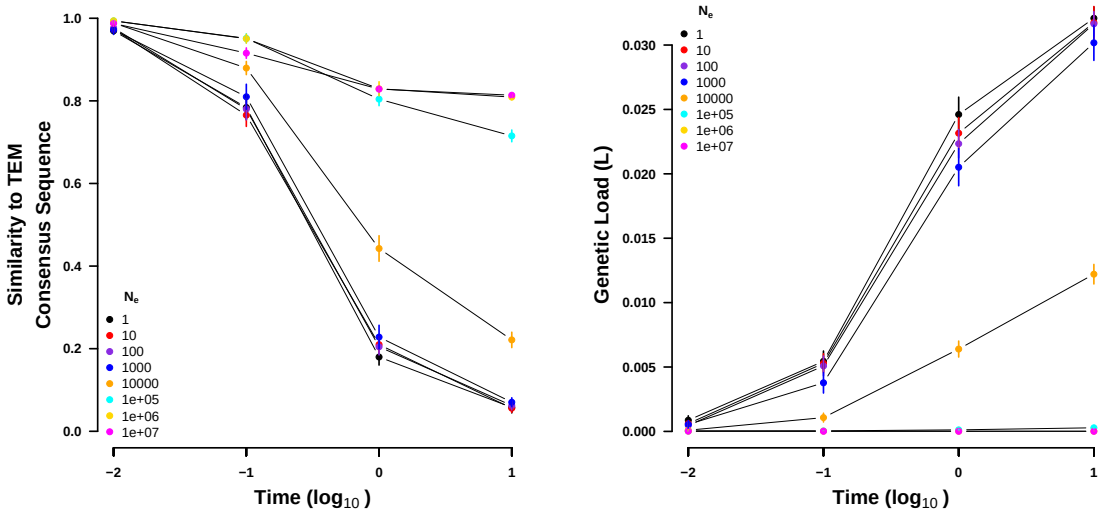

Figure 4: Sequences simulated from the ancestral state under the site specific selection on amino acids estimated using *SelAC*. (left) Sequence similarity to the observed consensus sequence at various times for a range on values of N_e. (right) Genetic load of the simulated sequences at various times for a range on values of N_e. Time is given in number of expected mutations. Points indicate sample means and vertical bars indicate standard deviations. Initial sequence is the inferred ancestral state of the TEM variants and not shown.

To further demonstrate the consistency of *SelAC*, we utilized random codon sequences as starting points. 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 S3a). The same initial sequences under the site specific selection inferred by the deep mutation scanning experiment increase only to $\sim 18\%$ in sequence similarity.

		G		Genetic Load	
Protein	Secondary Structure	Mean	SE	Mean	SE
TEM		219.3	7.5	0.16×10^{-7}	6.5×10^{-8}
	Helix	206.1	12.4	0.18×10^{-7}	0.13×10^{-7}
	Beta Sheet	238.6	15.8	6.8×10^{-8}	2.9×10^{-8}
	Unstructured	224.8	11.4	0.19×10^{-7}	8.1×10^{-8}
	Active Sites	300	0	0	0
SHV		244.9	6.8	4.0×10^{-8}	1.9×10^{-8}
	Helix	234.6	11.5	7.3×10^{-8}	4.8×10^{-8}
	Beta Sheet	253.1	12.8	2.1×10^{-8}	1.1×10^{-8}
	Unstructured	250.3	11.0	1.8×10^{-8}	59×10^{-8}
	Active Sites	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. UPDATE TABLE, MAKE EVERYTHING 10-8

Site Specific estimates of Selection on Amino Acids

SelAC allows for the site specific estimation of 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 distributed along most of the observed TEM sequence with the exception of the region between residue 80 to 120 were three consequtive helices are located (Figure 5). The most noticable increases in genetic load are found in unstructured regions The largest increase in genetic load however, is located at the begining of the last helix. We therefore estimate similar genetic loads for helices and unstructured regions in the observed TEM sequences (Table 2) The highest 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 extraplotate the fitness landscape of related proteins [Bloom, 2014]. 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, and genetic load. We find that site specific efficacy of selection G differes 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). With the expection of the active site, we find that G is increased in SHV (Table 2).

In general, 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 (Figure S4b).

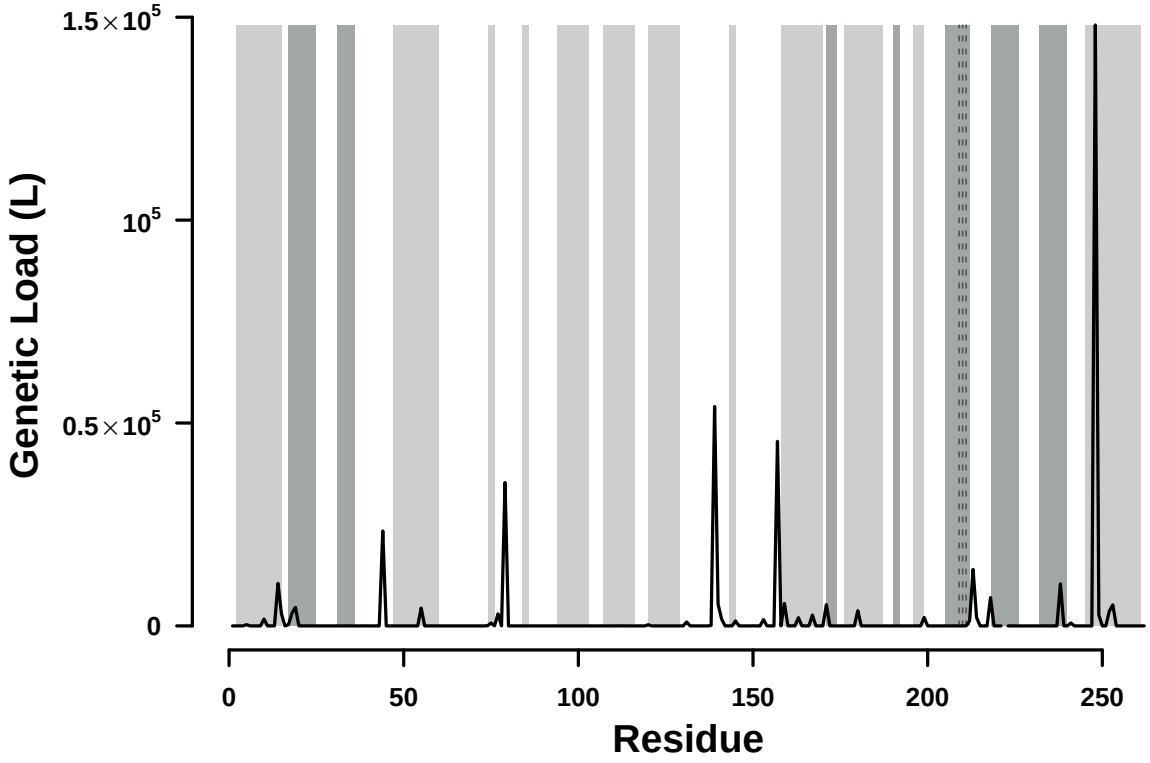

Figure 5: TEM, bars are different secondary structure elements. Dashed dotted line is DMS, solid is SelAC sNe, all lines are means of all sequences. vertical lines are active/binding sites.

The genetic load in SHV is by an order of magnitute lower than in TEM with the exception of residues found in β-sheets and the active site (Table 2). This is consitent with the elevated site specific efficacy of selection G in SHV. As a comparison of site specific efficacy of selection G already indicated, the sites introducing genetic load differ between SHV and TEM (Figure S1). We find the highest genetic load in SHV at the end of the first helix. However, we do find a peak of similar magnitute in the TEM sequence at the end of the first helix.

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 [Bloom, 2014, 2017]. While our study confirms this notion, we identify improtant shortcomings of these laboratory estimates. In contrast, *SelAC* is a more general phylogenetic model of stabilizing selection that does not depend on costly laboratory estimates of selection and is favored by model selection (Table 1).

While previous work showed the advantages of experimentally informed phylogenetics estimates, they did not assess how adequate the estimated selection reflects observed sequences. This becomes abherent in the low sequence similarity between the observed consensus sequence and the sequence of selectively favored amino acids estimated from deep mutation scanning experiments. This begs the question how well the experimental selection coefficients represent evolution in the wild. Deep mutation scanning experiments are performed using a comprehensive library of mutants and a strong artificial selection pressure [Firnberg and Ostermeier, 2012, Jain and Varadarajan, 2014, Fowler and Fields, 2014, Fowler et al., 2014]. This results in a very large selection coefficient s and a competing heterogeneous population.

The induced selection pressure during the deep mutation scanning experiment was limited to ampicillin [Stiffler et al., 2016] and focused on the TEM-1 variant. However, TEM can also confer resistance to a wide range of other antibiotics, like other penicillins, cephalosporins, cefotaxime, ceftazidime, or aztreonam [Sougakoff et al., 1988, 1989, Goussard et al., 1991, Mabilat et al., 1992, Chanal et al., 1992, Brun et al., 1994]. Thus, the inferred selection is biased towards ampicillin and as our simulations show does not reflect the evolution the observed TEM variants have experienced (Figure 3). This may very well be very appropriate to explore the selection on TEM in a modern hospital environment but is unlikely to be

applicable to the selection faced in the wild. We therfore propose to include a variety of selection pressures if the experimental selection estimates are used for phylogenetic inference.

If we assume that the experimental selection estimates 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, potntially due to a lack of selection of not enough time. Second, the observed TEM sequences are mal-adapted. 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 [Ochman and Wilson, 1987, Hartl et al., 1994]. As new mutations are introduced into a population proportional to N_e, *E. coli* can effectively explore the sequence space. We therefore expect the observed sequence variants to be near mutation-selection-drift equilibrium. This is confirmed by our simulations as we would expect to observe a higher sequence similarity and decreased genetic load even with much smaller N_e (Figure 3). Previous work showed that TEM the catalytic reaction of penicillin-class antibiotics is close the diffusion limit, making TEM a so-called perfect enzyme [Matagne et al., 1998].

As experimental selection estimates are not readily available, one solution is to extrapolate the estimates to homologeous gene families [Bloom, 2014, 2017]. When extrapolating the selection estimates from the β-lactamase family TEM to SHV, the sequence similarity between the observed consensus sequence and the sequence of selectively favored amino acids estimated from deep mutation scanning experiments drops rom 52% to 49%. Comparrison of the site specific efficacy of selection (G) revealed large differences in the site specific selection on amino acids between TEM and SHV. The missmatched in physicochemical weights also indicates differences in selection constrains. While the polarity of amino aicds 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 consensus sequence (99%).

SelAC does not rely on artifically induced selection in the laboratory but is a mechanistical framework rooted in first princibles. It estimates site specific selection on amino acids from the seqeuence data based on distances between amino acids in physicochemical space [Grantham, 1974, Beaulieu et al., in review]. This allows *SelAC* to be applied to any set of protein coding sequences, eliminating the need to extrapolate from one homologeous gene family to the next (e.g. from TEM to SHV).

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. While *SelAC* allows for site heterogeneity in selection for amino acids, it still assumes multiplicative fitness across all sites and therfore ignores epistatis. This however, is a shortcoming shared with experimental estimates by deep mutation scanning, as each mutation typically only carries one mutation [Firnberg and Ostermeier, 2012, Jain and Varadarajan, 2014]. *SelAC* is a model stabilizing selection, however, not every protein is under stabilizing selection. TEM playes a role in chemical warfare with conspecifics and other microbes, therfore 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) across the whole proteins. 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 Grantham [1974] for all sites. However, a wide range of physicochemical properties of amino acids have been assessed. 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. The hierarchical model structure allows to easilly address these shortcoming as needed.

In conclusion, experimental estimates of site specific selection on amino acids have to be

treated with great care and their adequacy should be assessed before informing phylogenetic studies. We also show that information on site specific selection on amino acids can be extracted from sequence data with mechanistical models rooted in first princibles.

Materials and Methods

Phylogenetic Inference and Model selection

TEM and SHV sequences were obtained from Bloom [2017] already aligned. We however, separated the TEM and SHV sequences into individual alignemnts. Experimentally fitness values for TEM were taken from Stiffler et al. [2016]. We followed [Bloom, 2017] to convert the experimental fitness values into site specific equilibrium frequencies for *phydms*.

SelAC (version 1.6.1) was fitted to the TEM alignment using R (version 3.4.1) [R Core Team, 2013] with and without site specific selection on amino acids estimated from deep mutation scanning experiments. We assumed the physicochemical properties estimated by Grantham [1974]. *phydms* (version 2.5.1) was fitted using site specific selection on amino acids estimated from deep mutation scanning experiments from Stiffler et al. [2016] and python (version 3.6). All other models were fitted using IQTree [Nguyen et al., 2015].

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 [Gillespie, 1976] that was model independent. The simulation followed Sella and Hirsh [2005] to calculate fixation probabilities. The fitness values were estimated using *SelAC* or experimentally inferred. We chose the fitnest 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 [Ashkenazy et al., 2012] (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 G

Estimating site specific fitness values w_i

Following Beaulieu et al. [in review] w_i is proportional to

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

were A_0 decribes 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 [Beaulieu et al., in review] 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 $A_0 = 4 \times 10{-}7$ [Gilchrist, 2007]. 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.

References

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.

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

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

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

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

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.

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

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

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

311 MT Holder, DJ Zwickl, and C Dessimoz. Evaluating the robustness of phylogenetic methods
312 to among-site variability in substitution processes. *Philos Trans R Soc Lond B*, 363:4013–
313 4021, 2008.

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

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

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

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

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

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

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

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

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.

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.

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

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

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

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.

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

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

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.

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

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.

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.

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.

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.

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.

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.

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

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

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

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

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL `http://www.R-project.org/`.

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.

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.

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.

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.

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.

Supplementary Material

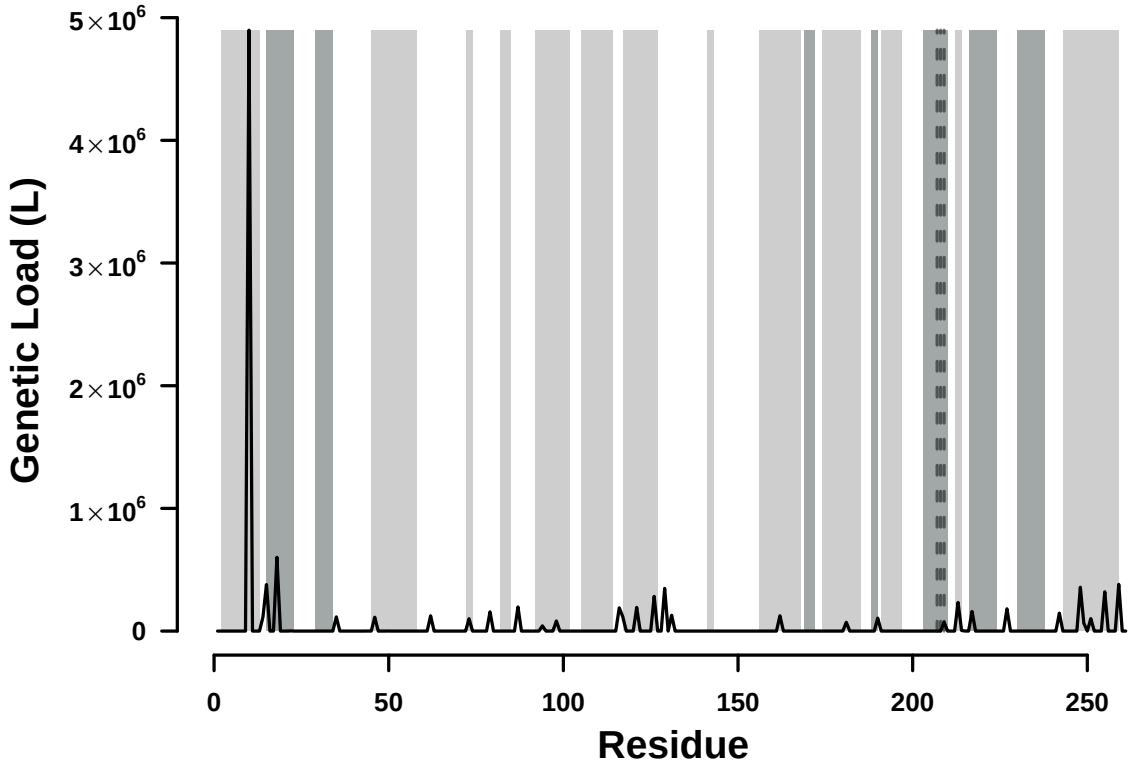

Figure S1: SHV, bars are different secondary structure elements. Dashed dotted line is DMS, solid is SelAC sNe, all lines are means of all sequences. vertical lines are active/binding sites.

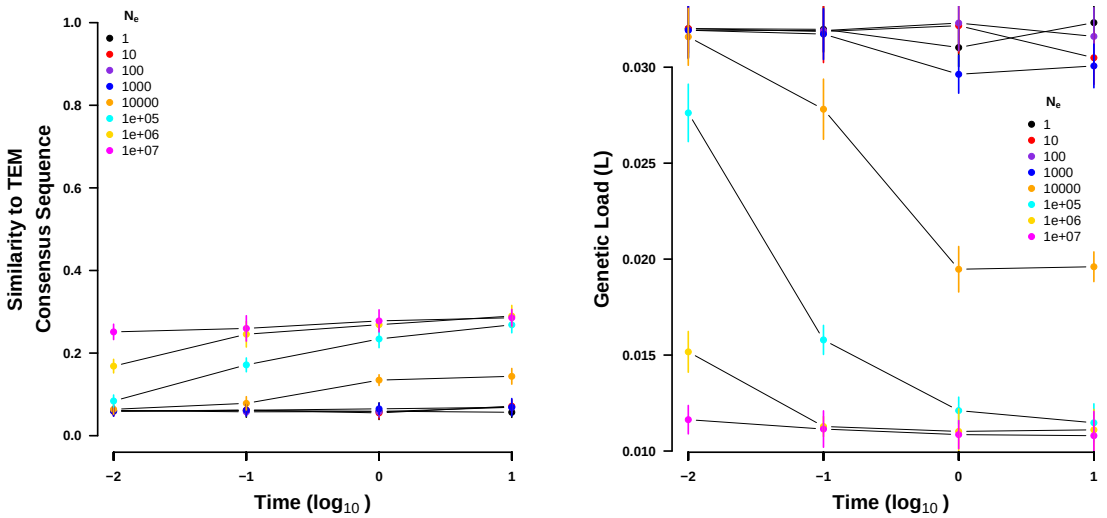

Figure S2: Sequences simulated from a random codon sequence under the site specific selection on amino acids estimated using *SelAC*. (left) Sequence similarity to the observed consensus sequence at various times for a range on values of N_e. (right) Genetic load of the simulated sequences at various times for a range on values of N_e. Time is given in number of expected mutations. Points indicate sample means and vertical bars indicate standard deviations. Initial sequence is the inferred ancestral state of the TEM variants and not shown.

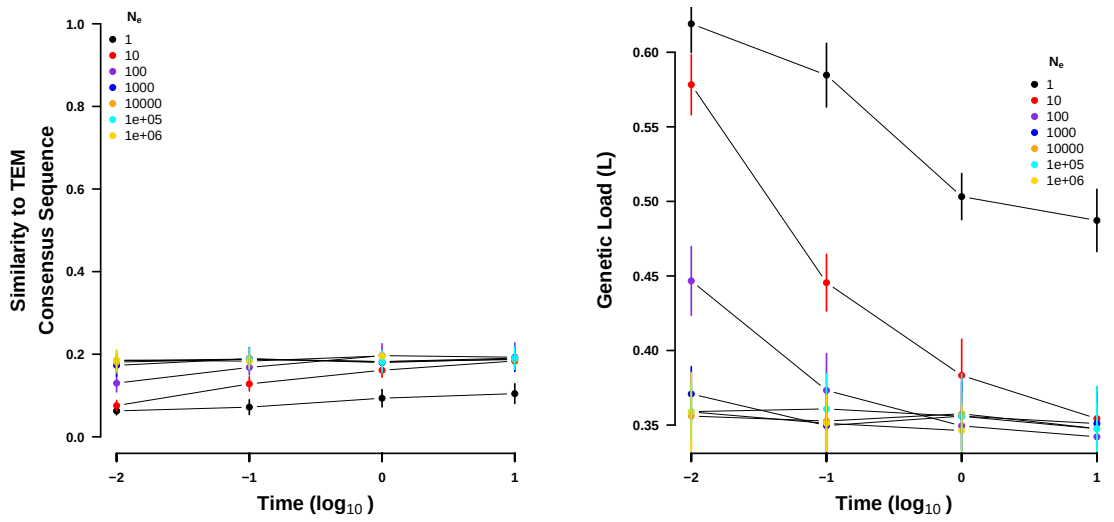

Figure S3: Sequences simulated from a random codon sequence under the site specific selection on amino acids estimated using deep mutation scanning. (left) Sequence similarity to the observed consensus sequence at various times for a range on values of N_e. (right) Genetic load of the simulated sequences at various times for a range on values of N_e. Time is given in number of expected mutations. Points indicate sample means and vertical bars indicate standard deviations. Initial sequence is the inferred ancestral state of the TEM variants and not shown.

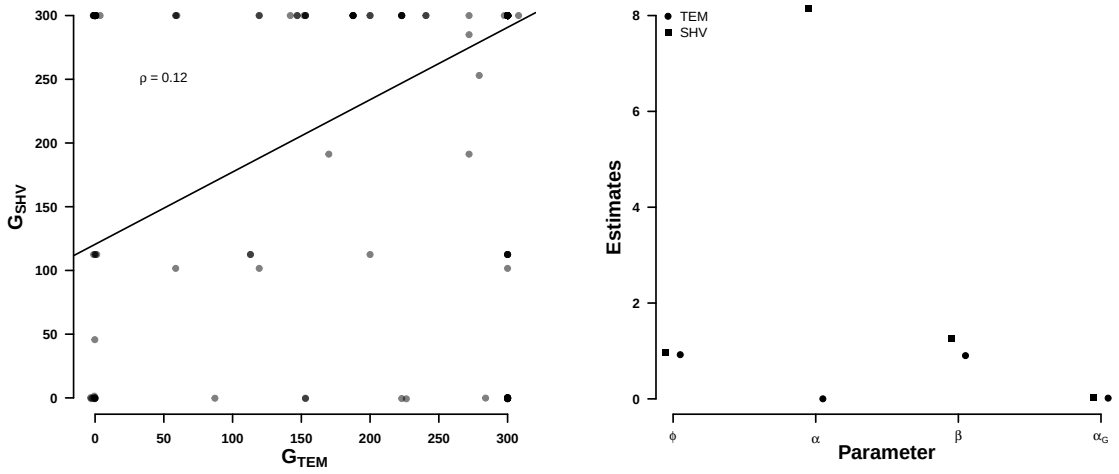

Figure S4: Comparisson of selection related parameters between TEM and SHV.