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 7, 2018

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.

- Incorporating selection into phylogenetic frameworks is already a long lasting endeavor.
  - Phylogenetic inference of sequence relationship was long focused on rates of substitutions.
  - No site specific equilibrium frequencies until (HB98, Bloom2014, ...).
  - Such models however, tend to be unfeasible as they are very parameter rich.
  - The type of selection on a protein is not always clear, or differs between proteins phylogenetic models also have to make generalizing assumptions.
  - Incorporating selection from experimental sources therefore seems like an attractive option.
  - Incorporating empirical fitness has some important features.
    * It allows for site specific amino acid preferences, acknowledging the heterogeneity of selection along the protein sequence.
    * It greatly reduces the number of parameters that have to be estimated from the data.
    * It allows for the fitting more complex models

- However, the incorporation of empirical fitness also has some important shortcomings.
    - Loss of generality.
    - DMS experiments are limited to proteins and organisms that can be manipulated under laboratory conditions.
    - But even in the case of TEM, the applied selection pressure is limited to the defense against a specific antibiotic.
    - TEM, however, has evolved to compete against conspecifics and other microbes using secreted metabolites to gain an advantage.
    - Furthermore, DMS relies on a library of mutants and therefore on a heterogeneous population with competing genotypes.
    - Therefore, it is important to ask how adequate such experiments reflect natural evolution.

- In this study we will assess how adequate DMS inference of site specific selection on amino acids, using TEM and provide an alternative, more generally applicable solution.
    - Simulations using DMS inferred site specific selection on amino acids show that observed TEM variants are unexpected; revealing the inadequacy of DMS.
    - Models fits achieved by the incorporation of DMS experiments can be improved upon using a hierarchical phylogenetic framework of stabilizing selection: SelAC.
    - Extrapolating site specific selection on amino acids between sequences (TEM and SHV) with related function can be inadequate.

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

Results

Site Specific Selection on Amino Acids Improves Model Fit

We compared the models *phydms* [Hilton et al., 2017] and *SelAC* [Beaulieu et al., in review], 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 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.

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.

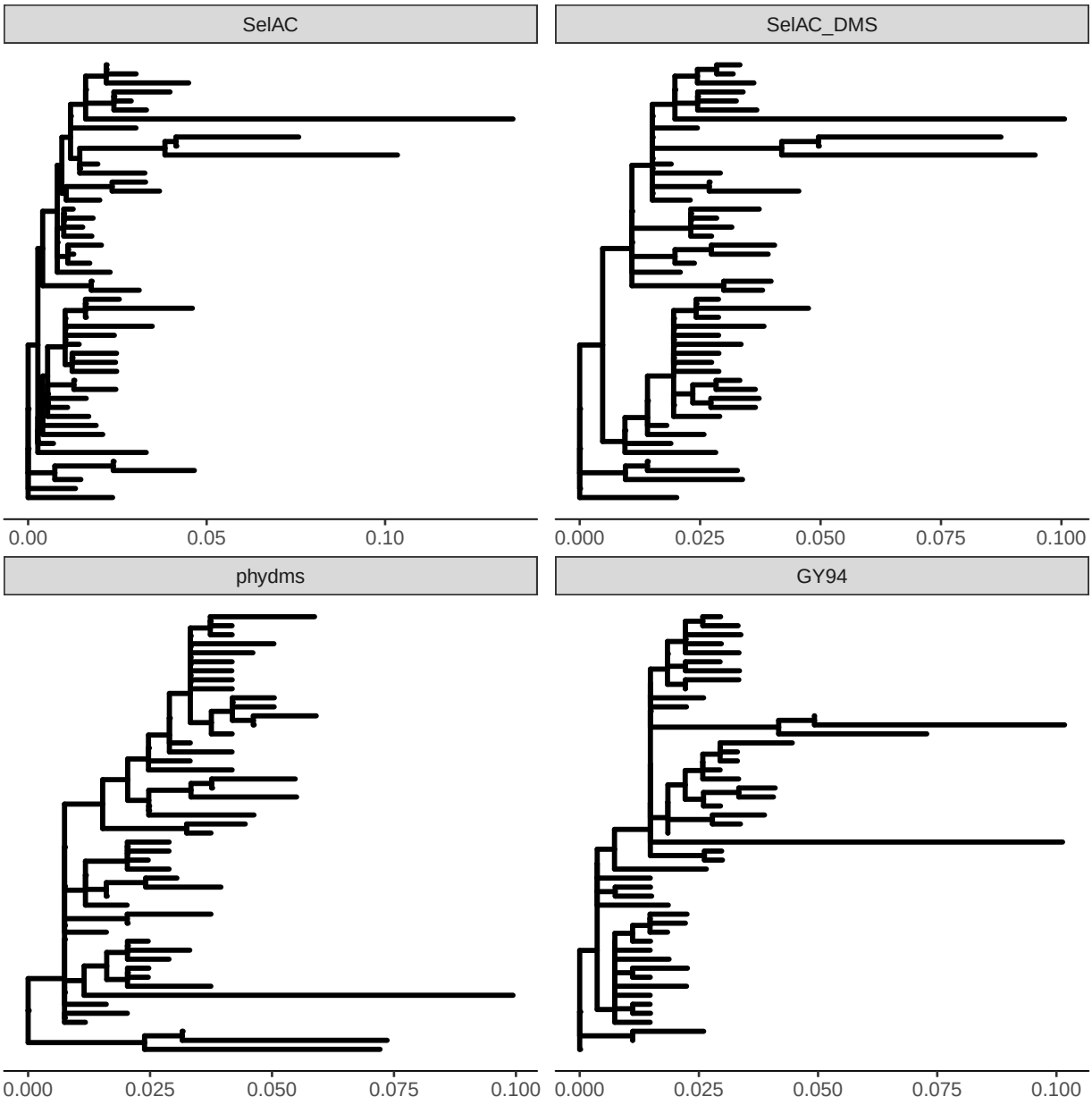

Figure 1: Phylogenies estimated using *SelAC*, *SelAC* +DMS, *phydms*, and *GY94*.

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 49 % 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*.

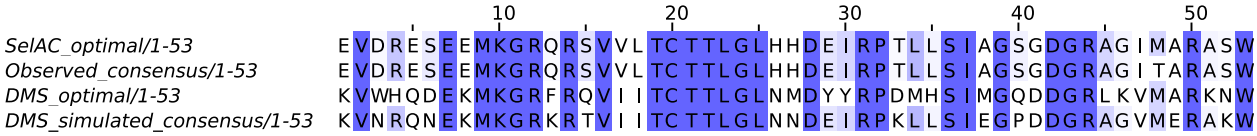

Figure 2: Every 5th residue. DMS and simulation based on DMS do not reflect natural sequences

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.

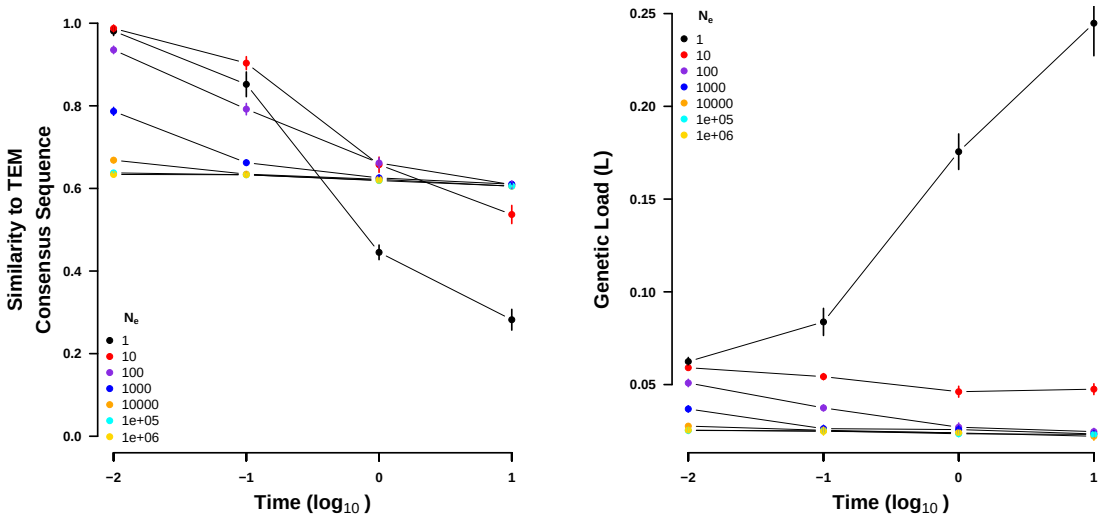

Figure 3: Sequences simulated from the ancestral state under the site specific selection on amino acids estimated using deep mutation scanning. (left) Sequence similarity to the observed consensus sequence at various times for a range on values of N_e. (right) Genetic load of the simulated sequences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

SelAC 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.

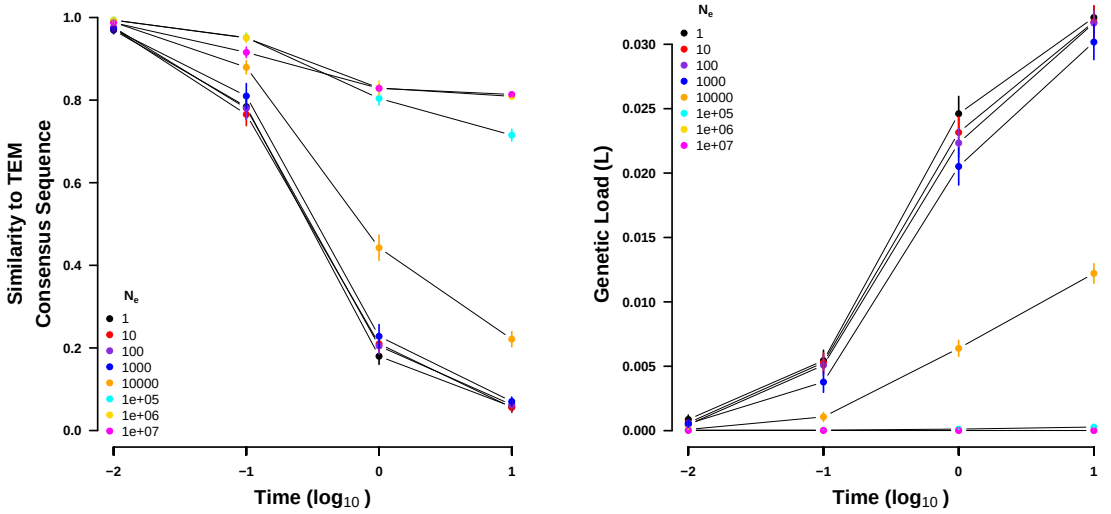

Figure 4: Sequences simulated from the ancestral state under the site specific selection on amino acids estimated using *SelAC*. (left) Sequence similarity to the observed consensus sequence at various times for a range on values of N_e. (right) Genetic load of the simulated sequences at various times for a range on values of N_e. Time is given in number of expected substitutions. 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 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.

Protein	Secondary Structure	Mean G	SE G	Mean Genetic Load	SE Genetic Load
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.

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.

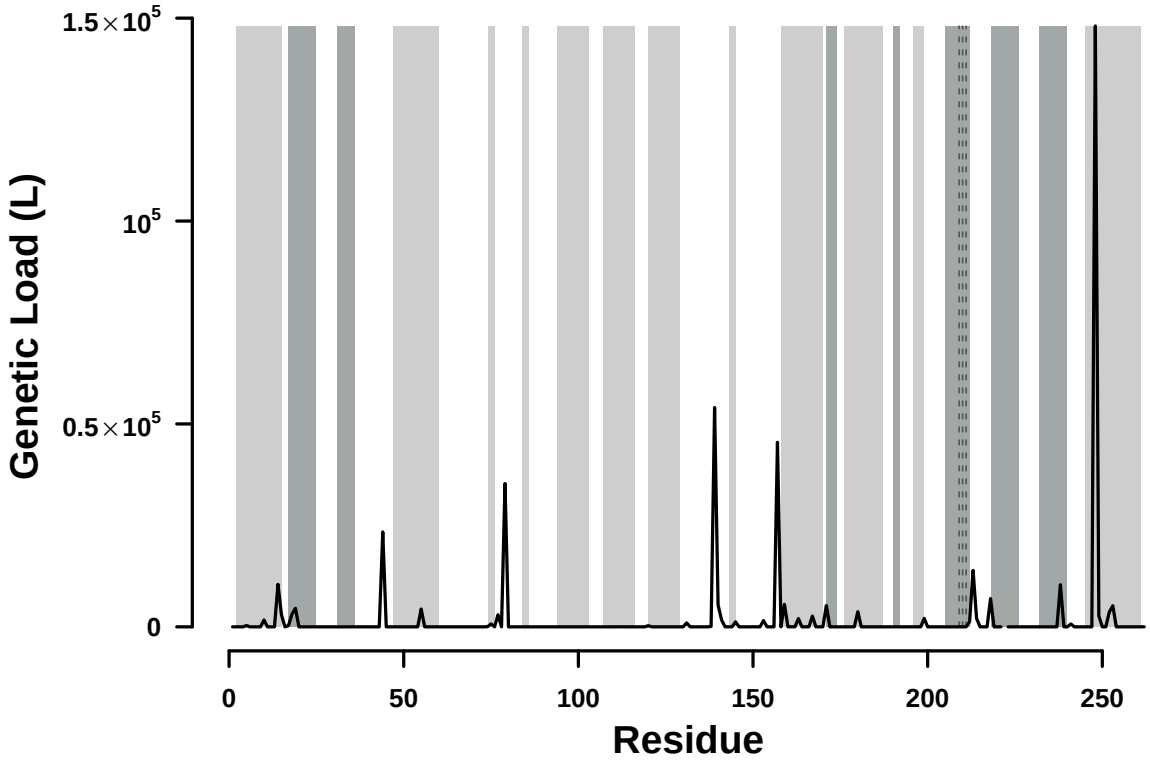

Figure 5: TEM, bars are different secondary structure elements. Dashed dotted line is DMS, solid is SelAC sNe, all lines are means of all sequences. vertical lines are active/binding sites.

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 ($\rho = 0.17$), despite a similar estimate of the parameter α_G describing the distribution of G values (Figure S3a). 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 S3b).

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

We evaluated 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 also raise awarness for shortcomings of these laboratory estimates and propose a more general applicable alternative. *SelAC*, in contrast, is a more general alternative that does not depend on costly laboratory estimates of selection and is favored by model selectin (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. In contrast, the selectively favored amino acids estimated by *SelAC* shows a high sequence similarity with the observed consensus sequence. This begs the question how well the experimental selection coefficients represent evolution. 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. Thus, the inferred selection is biased towards ampicillin and is therefore unlikely to reflect the evolution the observed TEM variants have experienced. We therfore propose to include a variety of selection pressures if the experimental selection estimates are used for phylogenetic inference.

TODO: Lack of repeatability between labs introduces further problems (Firnberg et al 2014 vs. Stifler et al. 2016).

- We evaluated how well experimental selection estimates from DMS experiments explain natural sequence evolution and compared it to a novel phylogenetic framework, SelAC.
  - Previous work has shown that DMS selection estimates can improve model fit over classical approaches like GY94 and our work confirms this.
  - Model selection favored the SelAC model fit and the corresponding fitness estimates over the DMS estimates using both, SelAC and phyDMS (Table 1).
- Adequacy of the DMS selection has previously not been assessed.
  - The amino acid with the cumulative highest fitness experimentally estimated with DMS only has 49% concordance with the observed alignment.
  - In contrast, the SelAC estimate has 99% concordance (Figure 2).
  - Estimates of selection coefficients do not represent evolution.
    - Due to artificial selection environment; Heterogeneous population, very large s.

193 * Only one antibiotic used, maybe a mixture of antibiotics would better reflect
194 natural evolution.
195 * Lack of repeatability between labs introduces further problems (Firnberg et
196 al 2014 vs. Stifler et al. 2016).

197 • Assuming that the DMS selection inference adequately reflects natural evolution, the
198 observed TEM sequences are either mal-adapted or where unable to reach a fitness
199 peak.

200 – *E. coli* has a large effective population size, estimates are on the order of 10^8 to
201 10^9 (Ochman and Wilson 1987, Hartl et al 1994).
202 – The large N_e would allow *E. coli* to effectively "explore" the sequence space,
203 thus suggesting that the TEM sequences are mal-adapted according to the DMS
204 estimates.
205 – Our simulations of sequence evolution with various N_e values and the DMS fitness
206 values in contrast show that we would expect higher adaptation even with much
207 smaller N_e (Figure 3).

208 • Estimates of selection coefficients do not represent evolution.

209 – Due to artificial selection environment; Heterogeneous population, very large s.
210 – Only one antibiotic used, maybe a mixture of antibiotics would better reflect
211 natural evolution.
212 – Lack of repeatability between labs introduces further problems (Firnberg et al
213 2014 vs. Stifler et al. 2016).
214 – Still better than models without site specific equilibrium frequencies.

215 • DMS estimates of the observed TEM variants predict them to be mal-adapted while
216 SelAC predicts most TEM variants to be well adapted.

- Given *E. coli*'s large effective population size, the efficacy of selection should be very large.
- We therefore expect the observed sequence variants to be at the selection-mutation-drift barrier, which in turn can expected to be near the optimum.
- We find the majority of sequences near the optimum, therefore the SelAC estimates are consistent with theoretical population genetics results.
- In contrast, finding strong selection against the observed TEM variants indicates that DMS is not consistent with theoretical population genetics expectations.
- This is consistent when thinking about that DMS only reflects the selection on the TEM sequence with regards to one antibiotic, which seems appropriate to model selection in modern hospital environments but not when the interest lies in the natural evolution of TEM.

- We find that SelAC produces similar selection against the observed TEM variants if we assume the fitness peaks (optimal AA) that are estimated by DMS.
  - This shows that DMS and SelAC can provide consistent estimates of selection against amino acids.
  - SelAC has the advantage that it can be applied to any protein coding sequence alignment.
  - This removes the need for extrapolation e.g. from TEM to SHV.

- SelAC has the advantage that it can be applied to any protein coding sequence alignment.
  - This removes the need for extrapolation e.g. from TEM to SHV.

- Difference in selection parameters between TEM and SHV indicate that extrapolation is not a good idea.

- The difference in the site specific strength of selection shows that TEM and SHV are facing different selection pressures.
- this is also highlighted by the differences in physicochemical weightings between the two proteins.

- SelAC outperforms DMS, but is not without flaws itself
  - Like DMS and most phylogenetic models, SelAC assumes site independence.
  - SelAC is a model of stabilizing selection, in contrast to e.g. GY94 which is a model of frequency dependent selection.
    * Since TEM plays a role in the chemical warfare with conspecifics and other microbes, some sites may be under negative frequency dependent selection.
  - SelAC assumes the same G distribution across all sites.
    * Different G distribution for each type of secondary structure
    * active sites may not follow distribution.
  - SelAC assumes that selection is proportional to distance in physicochemical space.
    * We used Grantham (1974) properties, however many other distances are available which may an even better model fit.

- Low sequence variation in the TEM may be cause for concern as it could be misinterpreted by the model as stabilizing selection because of the short branches.
  - However, provided our simulations support that TEM is actually under stabilizing selection

- In conclusion, DMS experiments have been proposed to supplement information on selection on amino acids in phylogenetic studies.
  - This study shows that information on selection can be extracted from alignments of protein coding sequences.

– This highlights the limitations of DMS to explain natural evolution.

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. *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.

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, 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 $\psi' = A_0 \times \psi \times N_e \times q$. N_e describes the effective population size and is assumed by *SelAC* to be 5×10^6. q is the XXX and is assumed by *SelAC* to be 4×10^{-7}. A_0 assumed to be 4. 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.

Spencer V Muse and Brandon S 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.

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

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.

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

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

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.

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.

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

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.

Figures

Supplementary Figures

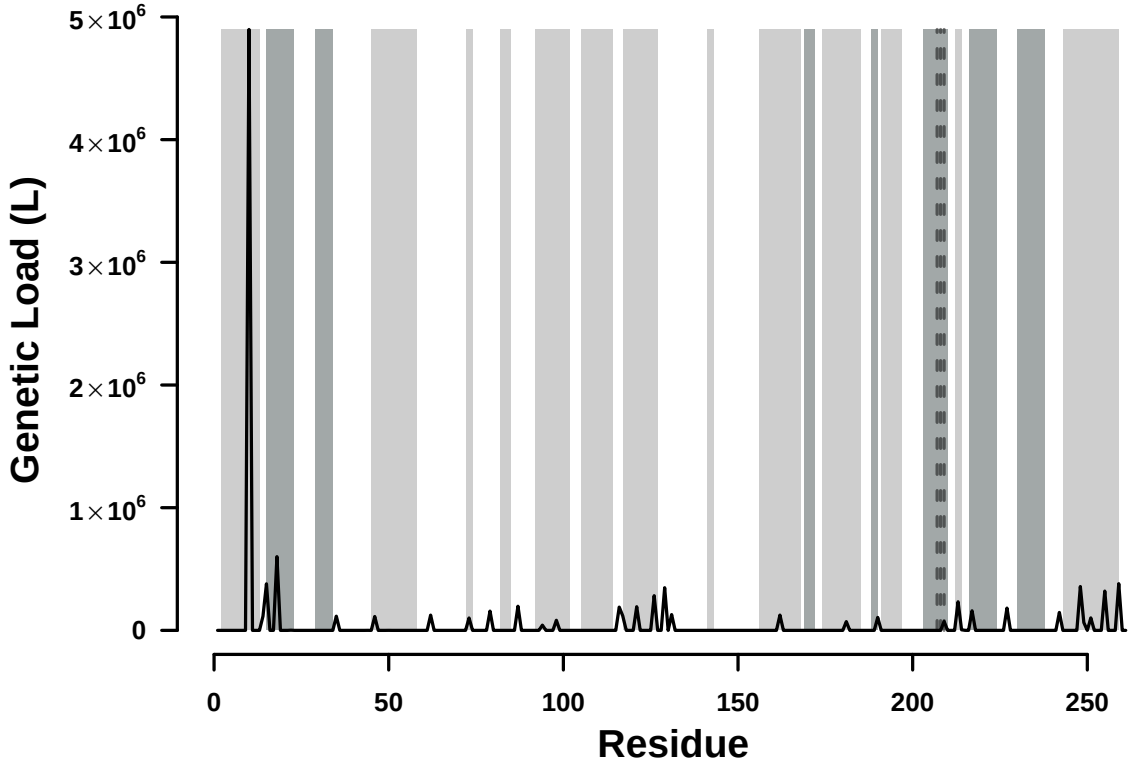

Figure S1: SHV, bars are different secondary structure elements. Dashed dotted line is DMS, solid is SelAC sNe, all lines are means of all sequences. vertical lines are active/binding sites.

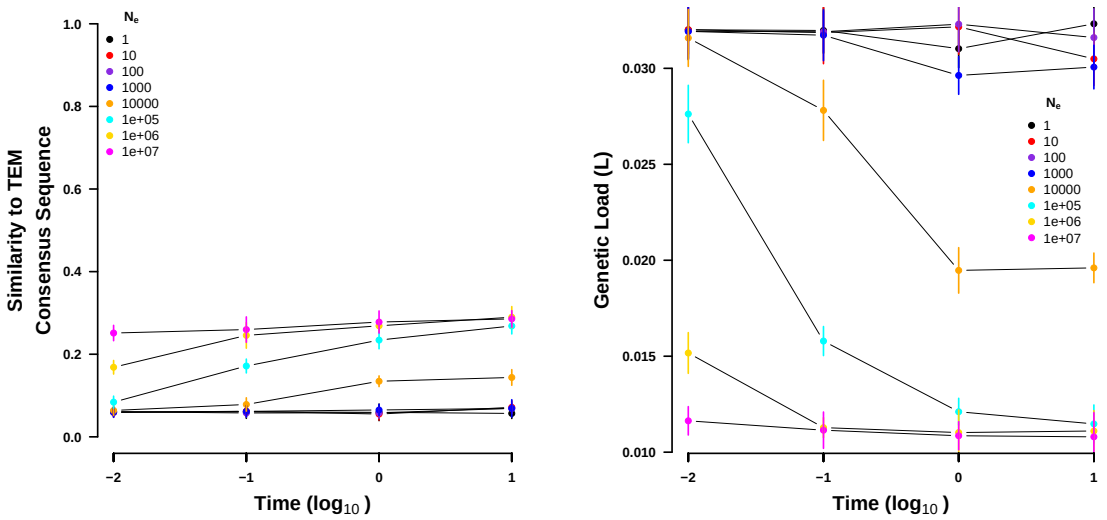

Figure S2: Sequences simulated from a random codon sequence under the site specific selection on amino acids estimated using *SelAC*. (left) Sequence similarity to the observed consensus sequence at various times for a range on values of N_e. (right) Genetic load of the simulated sequences at various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

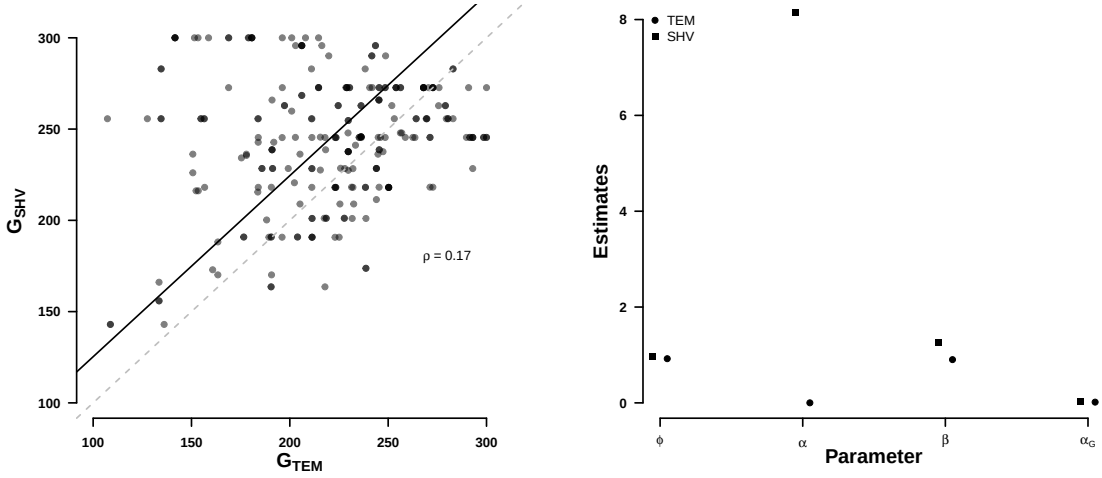

Figure S3: Comparisson of selection related parameters between TEM and SHV.