**Supplementary data for**

# The genetic code constrains yet facilitates Darwinian evolution

Elad Firnberg and Marc Ostermeier

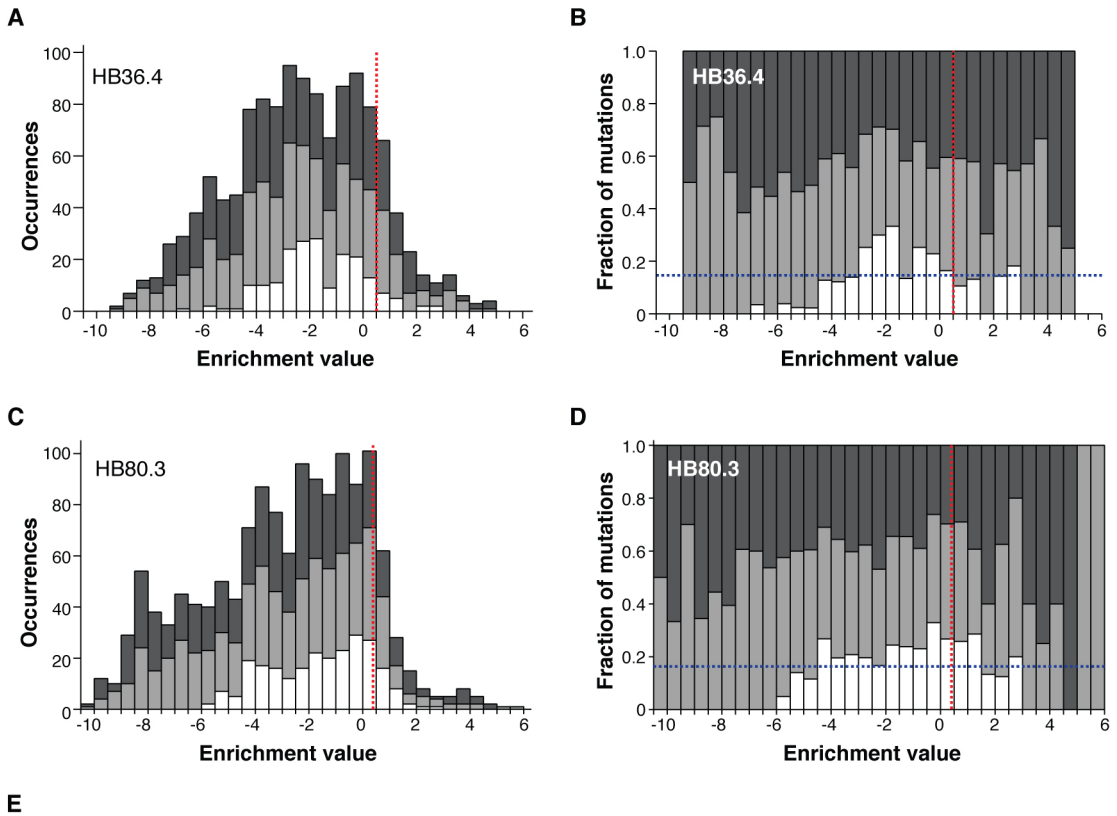
Department of Chemical and Biomolecular Engineering, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218 USA

**Supplementary Text**

**Strength of the *TEM-1* adaptive mutations for tazobactam and cefotaxime resistance.** A comparison with previously reported adaptive mutations for tazobactam resistance indicates that the selective conditions we utilized to identify tazobactam resistance alleles restricts against adaptive mutations with a small effect (15). For the cefotaxime resistance mutations, we compared our list of adaptive amino acid substitutions with Schenk et al's (41) extensive list of 48 cefotaxime resistance adaptive point mutations identified from an error prone PCR library of *TEM-1*. Twenty of our 30 amino acid substitutions are unique to our result, and many of these have either not been previously reported or are not previously known to confer resistance in isolation (42). Our selection readily identified all seven adaptive amino acid substitutions previously shown to confer a >3-fold improvement in resistance (41). We find 26 of the 28 possible codon substitutions that can give rise to these seven amino acids substitutions, suggesting our list contains ~93% of all of adaptive alleles with >3-fold improvement in resistance. We identify some substitutions previously shown to confer between 2.6 and 3-fold improvement, and do not identify any substitutions shown to confer <2.6-fold improvement. Substitutions with a ≤3-fold effect represent 85% of the adaptive point mutations (41). In summary, our selection identified a large set of adaptive alleles conferring about 2.5-fold or greater improvement in cefotaxime resistance, which we estimate to be the fittest 15% or more of all adaptive alleles with a single codon substitution.

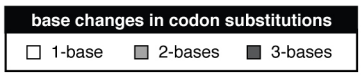
**References**

15. Firnberg, E. and Ostermeier, M. (2012) PFunkel: efficient, expansive, user-defined mutagenesis. *PLoS One*, **7**, e52031.
41. Schenk, M.F., Szendro, I.G., Krug, J. and de Visser, J.A. (2012) Quantifying the adaptive potential of an antibiotic resistance enzyme. *PLoS Genet*, **8**, e1002783.
42. Salverda, M.L., De Visser, J.A. and Barlow, M. (2010) Natural evolution of TEM-1 beta-lactamase: experimental reconstruction and clinical relevance. *FEMS Microbiol Rev*, **34**, 1015-1036.




**E**

| Median enrichment value | | | |
|---|---|---|---|
| | **Δ1-base** | **Δ2-bases** | **Δ3-bases** |
| **HB36.4** | -1.71 | -2.59 | -2.86 |
| **HB80.3** | -1.38 | -3.14 | -3.45 |




**Supplementary Fig. S1**. Distribution of fitness effects of codon substitutions in (**A**, **B**) *HB36.4* and (**D**, **E**) *HB80.3* determined using codon enrichment values for experimentally observed codon substitutions. The distribution is partitioned into codon changes with 1-, 2-, and 3-base changes. The red dashed vertical line indicates the enrichment value of the parental genes, and the blue dashed horizontal bar indicates the fraction of all possible mutations of the gene that are point mutations. The difference between Fig. 2 and this figure is that in Fig. 2, the amino acid enrichment values from Whitehead et al. (17) were used and all the synonymous codon substitutions that could encode an amino acid substitution (whether they were present in the library or not) were assigned the same enrichment value. In this figure, only experimentally observed codon substitutions with >100 counts in the unselected library were included, and the enrichment values were calculated on an individual codon basis. (**E**) Median enrichment values for types of codon substitutions.

**Supplementary Table S1. Replicate cefotaxime resistance of *TEM-1* alleles by plate assay**

| Colony | For positions 42-104-182-238 | | MIC[a](µg/ml) | | | |
|---|---|---|---|---|---|---|
| | Amino acids | Codons | Replicate 1 | Replicate 2 | Replicate 3 | Median |
| - | No TEM-1 gene | | 0.08 | 0.08 | 0.08 | 0.08 |
| *TEM-1* | A-E-M-G | gca-gag-atg-ggt | 0.08 | 0.08 | 0.08 | 0.08 |
| *GKTS* | G-K-T-S | gga-aag-acg-agt | 90.5 | 90.5 | 90.5 | 90.5 |
| | | | | | | |
| 43, 48 | G-K-M-A | ggg-aag-atg-gcg | 90.5 | 90.5 | 90.5 | 90.5 |
| 24 | G-K-M-S | ggg-aag-atg-tca | 45.3 | 45.3 | 64 | 45.3 |
| 2 | G-K-K-A | ggg-aag-aag-gct | 64 | 64 | 128 | 64 |
| 6 | G-K-T-A | gga-aag-acg-gct | 90.5 | 90.5 | 128 | 90.5 |
| 34 | | ggg-aag-acg-gcg | 181 | 181 | 181 | 181 |
| 9 | | ggg-aag-aca-gcc | 128 | 181 | 181 | 181 |
| 32 | G-K-T-S | ggg-aag-acg-tcg | 90.5 | 64 | 128 | 90.5 |
| 1 | G-K-A-A | ggg-aag-gcg-gct | 90.5 | 90.5 | 128 | 90.5 |
| 38 | G-K-A-S | ggg-aag-gcg-agc | 64 | 64 | 90.5 | 64 |
| 16 | | ggg-aag-gcc-agc | 90.5 | 90.5 | 90.5 | 90.5 |
| 14 | G-K-Q-A | ggg-aag-cag-gca | 90.5 | 90.5 | 128 | 90.5 |
| 5, 31 | | ggg-aag-cag-gcc | 64 | 64 | 64 | 64 |
| 7, 15 | | ggc-aag-caa-gca | 90.5 | 128 | 128 | 128 |
| 46 | G-K-S-A | ggg-aag-agc-gct | 181 | 128 | 90.5 | 128 |
| 33 | G-K-S-S | ggg-aaa-agt-agt | 90.5 | 64 | 128 | 90.5 |
| 3 | G-R-S-S | ggg-cgg-agc-tcg | 64 | 64 | 90.5 | 64 |
| 45, 49 | | ggt-aga-tct-tcg | 181 | 128 | 128 | 128 |

[a] Plate MIC assays performed in √2 increments of cefotaxime (Mueller-Hinton-agar, $10^4$ CFU/spot, 35°C for 20 hours).

**Supplementary Table S2. Cefotaxime resistance of *TEM-1* alleles by liquid assay**

| Colony | For positions 42-104-182-238 | | MIC[a](µg/ml) | | | |
|---|---|---|---|---|---|---|
| | Amino acids | Codons | Replicate 1 | Replicate 2 | Replicate 3 | Median |
| - | No TEM-1 gene | | 0.08 | 0.08 | 0.08 | 0.08 |
| *TEM-1* | A-E-M-G | gca-gag-atg-ggt | 0.08 | 0.08 | 0.08 | 0.08 |
| *GKTS* | G-K-T-S | gga-aag-acg-agt | 2048 | 4096 | 2048 | 2048 |
| | | | | | | |
| 43, 48 | G-K-M-A | ggg-aag-atg-gcg | 2048 | 2048 | 2048 | 2048 |
| 24 | G-K-M-S | ggg-aag-atg-tca | 2048 | 2048 | 2048 | 2048 |
| 2 | G-K-K-A | ggg-aag-aag-gct | 2048 | 2048 | 2048 | 2048 |
| 6 | G-K-T-A | gga-aag-acg-gct | 2048 | 2048 | 2048 | 2048 |
| 34 | | ggg-aag-acg-gcg | 2048 | 2048 | 2048 | 2048 |
| 9 | | ggg-aag-aca-gcc | 2048 | 2048 | 2048 | 2048 |
| 32 | G-K-T-S | ggg-aag-acg-tcg | 4096 | 4096 | 4096 | 4096 |
| 1 | G-K-A-A | ggg-aag-gcg-gct | 2048 | 2048 | 2048 | 2048 |
| 38 | G-K-A-S | ggg-aag-gcg-agc | 4096 | 2048 | 2048 | 2048 |
| 16 | | ggg-aag-gcc-agc | 2048 | 2048 | 2048 | 2048 |
| 14 | G-K-Q-A | ggg-aag-cag-gca | 2048 | 2048 | 2048 | 2048 |
| 5, 31 | | ggg-aag-cag-gcc | 2048 | 2048 | 2048 | 2048 |
| 7, 15 | | ggc-aag-caa-gca | 2048 | 2048 | 2048 | 2048 |
| 46 | G-K-S-A | ggg-aag-agc-gct | 2048 | 2048 | 2048 | 2048 |
| 33 | G-K-S-S | ggg-aaa-agt-agt | 2048 | 2048 | 2048 | 2048 |
| 3 | G-R-S-S | ggg-cgg-agc-tcg | 2048 | 2048 | 2048 | 2048 |
| 45, 49 | | ggt-aga-tct-tcg | 2048 | 2048 | 2048 | 2048 |

[a] Liquid MIC assays performed in 2-fold cefotaxime increments (Mueller-Hinton broth, $5x10^5$ CFU/culture, 35°C for 20 hours)

**Supplementary Table S3. Replicate cefotaxime resistance of *TEM-1* alleles for Figure 1.**

| Allele | Codon Substitutions | Cefotaxime MIC[a] (µg/ml) | | | |
|---|---|---|---|---|---|
| | | Replicate 1 | Replicate 2 | Replicate 3 | Median |
| AEMG | 0 | 0.06 | 0.08 | 0.08 | 0.08 |
| GEMG | 1 | 0.08 | 0.06 | 0.08 | 0.08 |
| AKMG | 1 | 0.08 | 0.11 | 0.16 | 0.11 |
| AEQG | 1 | 0.08 | 0.08 | 0.08 | 0.08 |
| AEMA | 1 | 0.32 | 0.32 | 0.32 | 0.32 |
| GKMG | 2 | 0.32 | 0.32 | 0.32 | 0.32 |
| GEQG | 2 | 0.06 | 0.08 | 0.08 | 0.08 |
| GEMA | 2 | 1.81 | 1.81 | 1.81 | 1.81 |
| AKQG | 2 | 0.11 | 0.23 | 0.23 | 0.23 |
| AKMA | 2 | 10.24 | 10.24 | 10.24 | 10.24 |
| AEQA | 2 | 0.32 | 0.45 | 0.45 | 0.45 |
| GKQG | 3 | 0.32 | 0.32 | 0.32 | 0.32 |
| GKMA | 3 | 115.85 | 115.85 | 115.85 | 115.85 |
| GEQA | 3 | 1.81 | 2.56 | 2.56 | 2.56 |
| AKQA | 3 | 10.24 | 10.24 | 10.24 | 10.24 |
| GKQA | 4 | 115.85 | 115.85 | 115.85 | 115.85 |

[a]MIC assays performed in √2 increments of cefotaxime (Mueller-Hinton-agar, $10^4$ CFU/spot, 35°C for 20 hours).

**Supplementary Table S4 Median enrichment value of adaptive mutations as a function of the number of base changes in the codon.**

| Gene | Median enrichment value | | |
|---|---|---|---|
| | Δ1-base | Δ2-bases | Δ3-bases |
| *HB36.4* | 0.933 | 1.172 | 1.192 |
| *HB80.3* | 0.956 | 1.025 | 1.285 |

**Supplementary Table S5. Experimentally identified adaptive codon substitutions for cefotaxime resistance in *TEM-1***

| Ambler position | codon | | amino acid | | occurrences |
|---|---|---|---|---|---|
| | WT | mutated | WT | mutated | |
| 69 | atg | tgc | M | C | 1 |
| | | tgt | | | 1 |
| 104 | gag | aag | E | K | 12 |
| | | atg | | M | 2 |
| | | cgg | | R | 3 |
| 164 | cgt | gca | R | A | 1 |
| | | gcg | | | 1 |
| | | gct | | | 1 |
| | | gac | | D | 1 |
| | | gat | | | 3 |
| | | ggc | | G | 1 |
| | | ggg | | | 5 |
| | | ggt | | | 5 |
| | | cac | | H | 1 |
| | | cat | | | 17 |
| | | aac | | N | 7 |
| | | agc | | S | 1 |
| | | agt | | | 11 |
| | | tca | | | 2 |
| | | tcg | | | 14 |
| | | tct | | | 3 |
| 166 | gaa | ccc | E | P | 1 |
| | | ccg | | | 1 |
| 171 | gaa | gta | E | V | 2 |
| | | gtt | | | 1 |
| | | tat | | Y | 2 |
| 172 | gcc | cac | A | H | 3 |
| | | cat | | | 4 |
| | | cca | | P | 1 |
| | | ccc | | | 2 |
| | | cct | | | 2 |
| | | aca | | T | 1 |
| | | acc | | | 5 |
| | | tac | | Y | 4 |
| | | tat | | | 1 |
| 238 | ggt | gca | G | A | 1 |
| | | gcc | | | 2 |
| | | gcg | | | 4 |
| | | gct | | | 3 |
| | | gac | | D | 1 |
| | | gat | | | 16 |
| | | gaa | | E | 2 |
| | | gag | | | 3 |
| | | aac | | N | 1 |
| | | aat | | | 1 |
| | | agc | | S | 2 |
| | | agt | | | 45 |
| | | tca | | | 2 |
| | | tcc | | | 1 |

| | | | | | |
|---|---|---|---|---|---|
| | | tcg | | | 3 |
| | | tct | | | 2 |
| | | aca | | T | 2 |
| | | act | | | 1 |
| 240 | gag | gca | E | A | 1 |
| | | gcg | | | 3 |
| | | gct | | | 1 |
| | | gga | | G | 8 |
| | | ggc | | | 9 |
| | | ggg | | | 20 |
| | | ggt | | | 10 |
| | | cca | | P | 1 |
| | | ccg | | | 8 |
| | | cct | | | 2 |
| | | agc | | S | 5 |
| | | agt | | | 1 |
| | | tca | | | 2 |
| | | tcc | | | 1 |
| | | tcg | | | 1 |
| | | tct | | | 1 |
| | | aca | | T | 1 |
| | | act | | | 1 |
| 241 | cgt | cca | R | P | 15 |
| | | ccc | | | 13 |
| | | ccg | | | 25 |
| | | cct | | | 13 |
| 243 | tct | ggc | S | G | 3 |
| | | ggg | | | 1 |

**Supplementary Table S6. Enrichment of adaptive amino acid substitutions by the genetic code.**

| Gene | Adaptive advantage | Number of amino acid substitutions in protein | | Number of identified adaptive amino acids substitutions in protein | | Enrichment of adaptive amino acids[a] (%) | p-value[b] |
|---|---|---|---|---|---|---|---|
| | | All | Accessible with a 1-bp substitution | All | Accessible with a 1-bp substitution | | |
| *TEM-1* | cefotaxime resistance | 5434 | 1687 | 30 | 13 | 39.6 | 0.106 |
| *TEM-1* | tazobactam resistance | 5434 | 1687 | 19 | 8 | 35.6 | 0.210 |
| *HB36.4* | hemagglutinin binding | 932 | 309 | 127 | 55 | 30.6 | 0.0066 |
| *HB80.3* | hemagglutinin binding | 964 | 312 | 83 | 27 | 0.51 | 0.531 |
| Variable for calculations | | N | m | n | k | | |

[a] $\text{enrichment} = \left[\frac{\frac{k}{n} - \frac{m}{N}}{\frac{m}{n}}\right] \times 100\%$

[b] calculated from a hypergeometric distribution: $P(x = k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}$