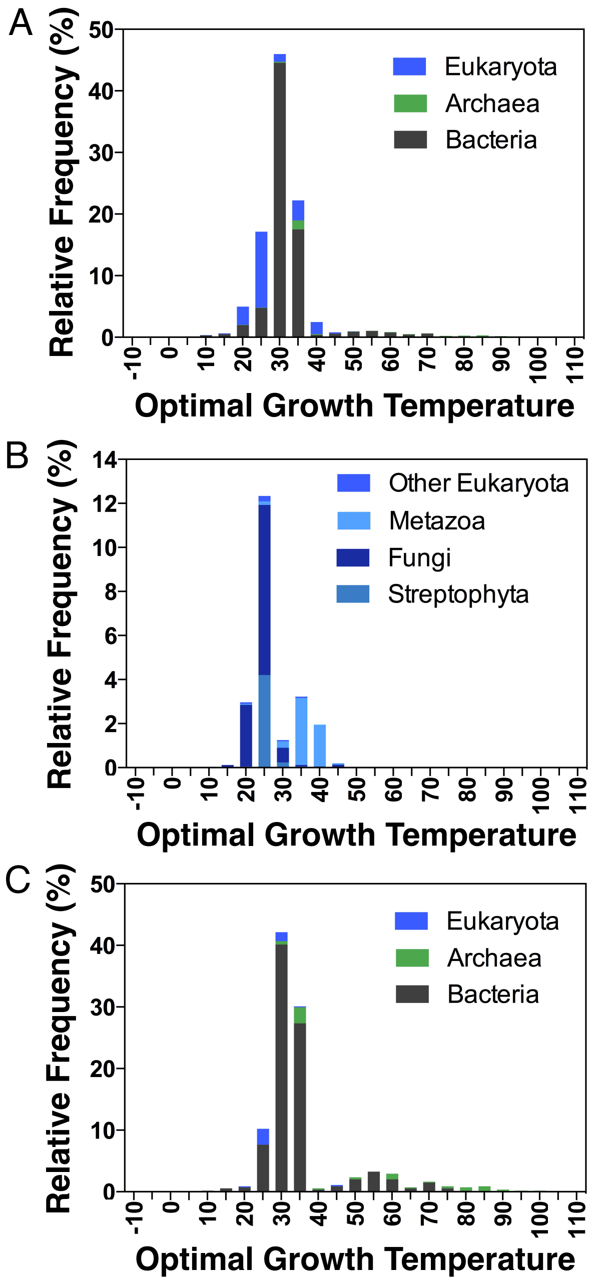
Biophysical Journal

Supporting Material

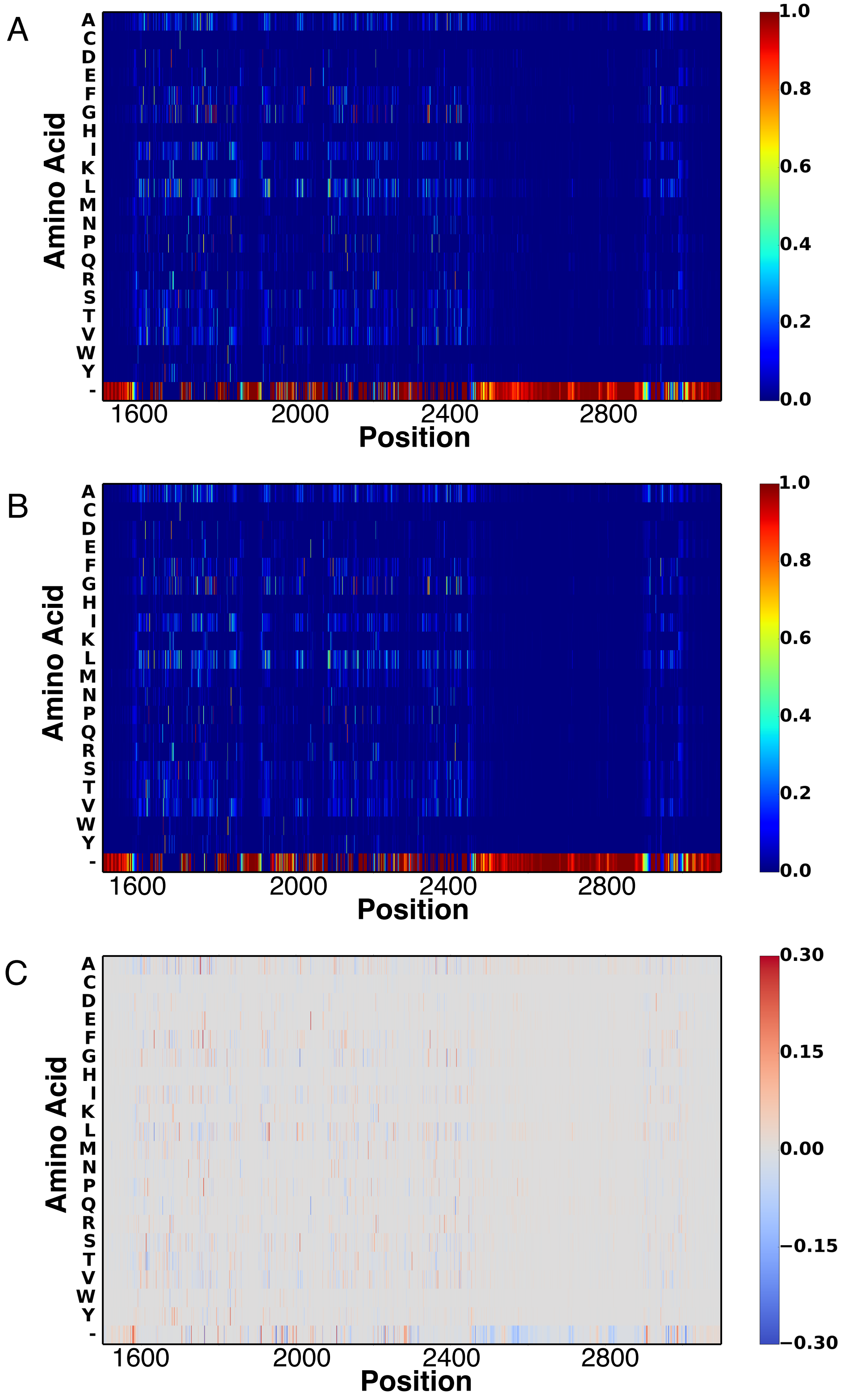
Rapid Bioinformatic Identification of Thermostabilizing Mutations

David B. Sauer,[1,*] Nathan K. Karpowich,[1] JinMei Song,[1] and Da-Neng Wang[1,*]

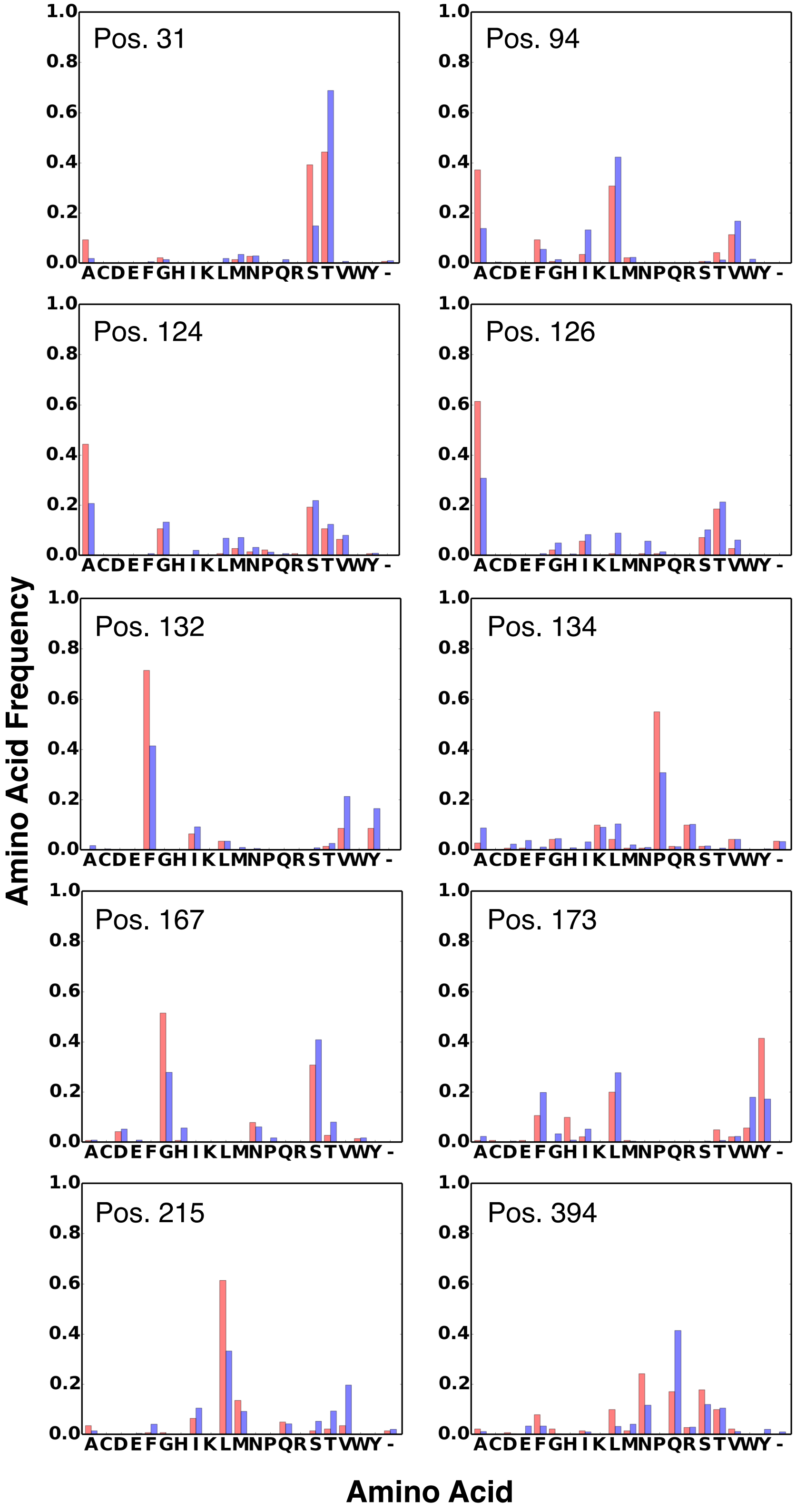
[1]Department of Cell Biology, The Helen L. and Martin S. Kimmel Center for Biology and Medicine, Skirball Institute of Biomolecular Medicine, New York University School of Medicine, New York, New York



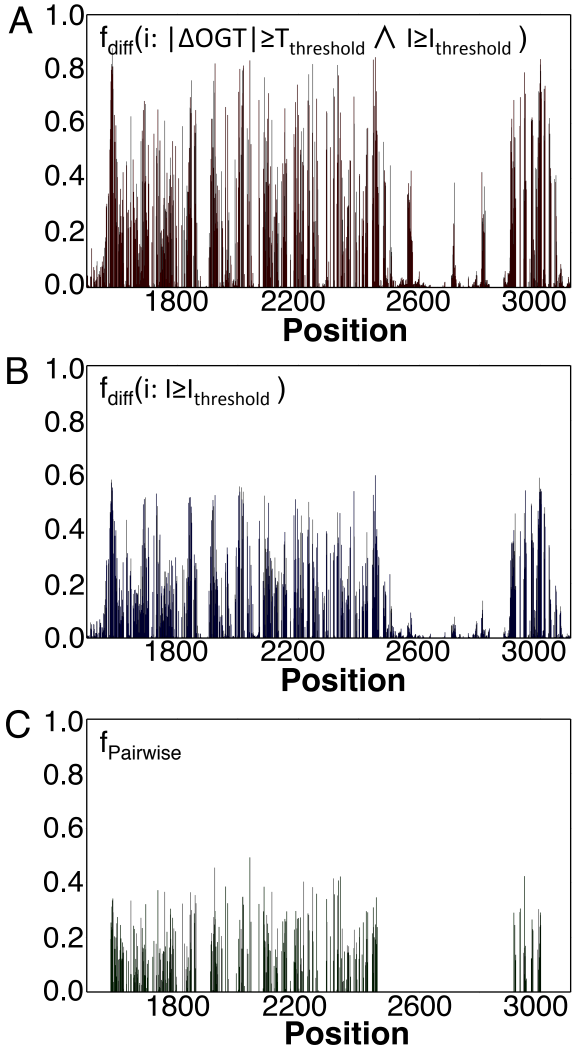
Supplementary Figure 1. OGT distributions within distinct taxa. A. Histogram of OGT by domain for all species data collected. B. Sub-histogram of A, OGT by kingdom for eukaryotes for all species data collected. C. Histogram of OGT assignments by kingdom for the BsTetL analysis.



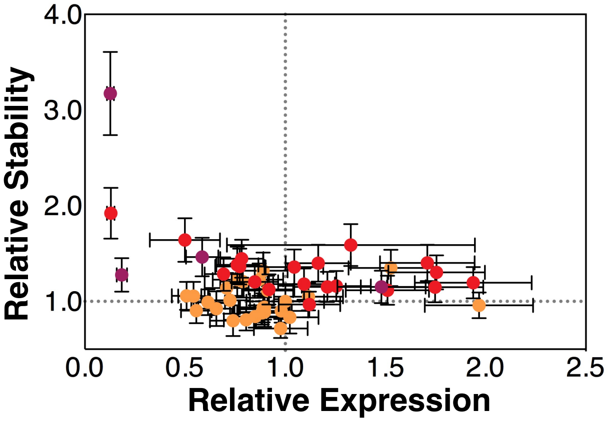
Supplementary Figure 2. Heatmaps of amino acid frequencies from the Global method. The heat maps of positional amino acid frequency in the multiple sequence alignment for thermophiles (A) and mesophiles (B). C. The difference in amino acid frequency by position between thermophiles and mesophiles. Positive (red) values indicate over-represented amino acids in thermophiles, negative (blue) values indicate amino acids under-represented (or over-represented in mesophiles).



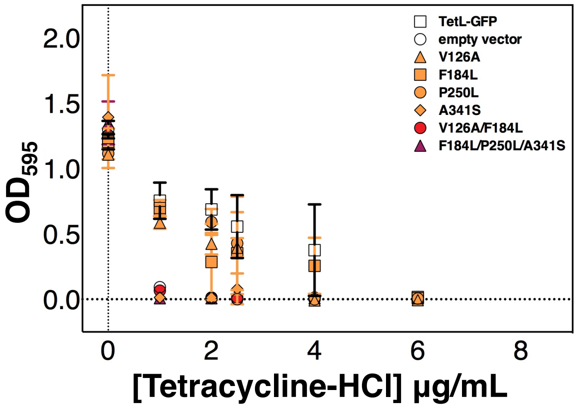
Supplementary Figure 3. Amino acid frequencies for the Global method at various positions. Histogram of amino acid frequency are shown for BsTetL positions 31, 94, 124, 126, 132, 134, 167, 173, 215, 394. Amino acid frequency is plotted in red or blue for thermophiles or mesophiles, respectively.



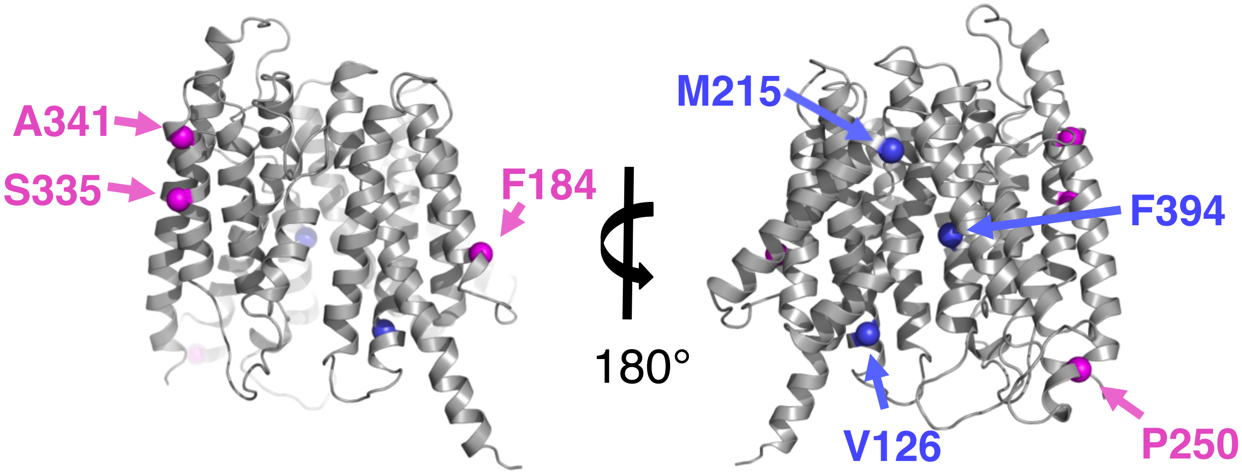
Supplementary Figure 4. Positional frequency of amino acid change among high similar homologs. A. The positional frequency of amino acid change between all close homologs with IΔOGTI > 10°C. B. Positional frequency of amino acid change, irrespective of OGT differences. C. Positional frequency of amino acid change after subtracting the frequency of change without OGT difference, positions of greater than 50% gap are hidden.



Supplementary Figure 5. Stability versus expression of BsTetL mutants. Stability and over-expression level for all mutants tested, normalized to wild type BsTetL. Single, double, and triple mutants are plotted as orange, red, and maroon circles, respectively. Error bars indicate standard deviation of duplicate experiments.



Supplementary Figure 6. Tetracycline resistance of BsTetL mutants at 41°C. Growth curve of BsTetL and mutants in the presence of tetracycline. Mutants are colored orange, red and maroon for single, double, and triple mutants, respectively. Wild type and empty vector control are indicated by open squares and circles, respectively.



Supplementary Figure 7. Location of BsTetL thermostabilizing mutants within the Major Facilitator Superfamily fold. Equivalent positions to the thermostabilizing mutations are noted on a BsTetL homology model. Mutants predicted by the global and pairwise methods are colored blue and magenta, respectively.

Equation S1.

$$
\begin{aligned}
&f_{Pairwise}(i) \\
&= \frac{\sum_{b=a+1}^{n}\sum_{a}^{n} H\left(AA_{a,i}, AA_{b,i}\right) \times \left(\left(|\Delta OGT(a,b)| > T_{Threshold}\right) \wedge \left(I(a,b) > I_{Threshold}\right)\right)}{\sum_{b=a+1}^{n}\sum_{a}^{n}\left(\left(|\Delta OGT(a,b)| > T_{Threshold}\right) \wedge \left(I(a,b) > I_{Threshold}\right)\right)} \\
&- \frac{\sum_{b=a+1}^{n}\sum_{a}^{n} H\left(AA_{a,i}, AA_{b,i}\right) \times \left(I(a,b) > I_{Threshold}\right)}{\sum_{b=a+1}^{n}\sum_{a}^{n}\left(I(a,b) > I_{Threshold}\right)}
\end{aligned}
$$

$$H(x,y) = \begin{cases} 0 \ if \ x = y \\ 1 \ if \ x \neq y \end{cases}$$

$$\Delta OGT(x,y) = OGT_x - OGT_y$$

$n \equiv number\ of\ sequences\ in\ MSA$

$AA_{x,i} \equiv the\ amino\ acid\ of\ sequence\ x\ at\ position\ i$

$OGT_x \equiv the\ OGT\ of\ sequence\ x$

$I(x,y) \equiv sequence\ identity\ between\ sequences\ x\ and\ y$