## **Supporting Information**

## Boyce et al. 10.1073/pnas.1405628111



(b) Default and Chained guide trees

Fig. S1. Total Column (TC) scores for increasing numbers of short-chain dehydrogenases/reductases sequences for Clustal Omega, Mafft (FFT-NS-2 algorithm), and Muscle (two iterations) with (A) default and balanced and (B) default and chained guide trees. Four sets of balanced and chained guide trees were created for each dataset, with different sequence ordering: Random, Traveling Salesman Problem (TSP) Minimum, Greedy Neighbor Minimum, and TSP Maximum (100 samples per dataset, except 25 samples for the largest Clustal Omega datasets).



Fig. 52. TC scores for increasing numbers of Cytochrome P450 sequences for Clustal Omega, Mafft (FFT-NS-2 algorithm), and Muscle (two iterations) with (A) default and balanced and (B) default and chained guide trees. Four sets of balanced and chained guide trees were created for each dataset, with different sequence ordering: Random, TSP Minimum, Greedy Neighbor Minimum, and TSP Maximum (100 samples per dataset, except 25 samples for the largest Clustal Omega datasets).



Fig. S3. TC scores for increasing numbers of zinc finger sequences for Clustal Omega, Mafft (FFT-NS-2 algorithm), and Muscle (two iterations) with (A) default and balanced and (B) default and chained guide trees. Four sets of balanced and chained guide trees were created for each dataset, with different sequence ordering: Random, TSP Minimum, Greedy Neighbor Minimum, and TSP Maximum (100 samples per dataset).



**Fig. 54.** TC scores for short-chain dehydrogenases/reductases sequences for Clustal Omega, Mafft (FFT-NS-2 algorithm), and Muscle (two iterations) with default and random chained guide trees. Five percent of the sequences in each dataset are Cytochrome P450 sequences (\*P < 0.05, \*\*P < 0.01, \*\*\*P < 0.001, 100 samples per dataset).



Fig. S5. Benchmark times for the alignment step for Clustal Omega, Mafft (FFT-NS-2 algo-rithm), and Muscle (two iterations) with default, balanced, and chained guide trees for shorter (zinc finger, zf-CCHH), medium length (short-chain dehydrogenases/reductases, sdr), and longer (Cytochrome P450, p450) sequences. Also included is the elapsed time for the default guide tree creation and the subsequence alignment. All times were determined using a single Intel Xeon 2.4GHz CPU with 16GB RAM.



Fig. S6. Average default tree depths for Clustal Omega, Mafft (FFT-NS-2 algorithm), Muscle (two iterations), and Kalign 2 using 1,024 sequences from each HomFam family. Families with less than 1,024 sequences were excluded (10 samples per family).

DNA C