Pseucode for comphet.pl

The following pseudocode describes the algorithm we implemented to simulate compositionally heterogeneous datasets. For all clades (clade-A, clade-B, clade-C, clade-D; Figure 1A) we assign the amino acid frequencies estimated from the Chang et al. (2015) dataset. For clade-A and clade-C, we follow the algorithm below for the following amino acid group pairs: (A,L), (R,K), (N,M), (D,F), (C,P), (Q,S), (E,T), (G,W), (H,Y), (I,V).

```
inflation_parameter = choose 1 from (0.1,0.5,0.9)
For each amino acid pair frequencies (amino_a, amino_b):
      adjustment_value = min_frequency(amino_a, amino_b) * inflation_parameter
      min_frequency_adjusted = min_frequency(amino_a, amino_b) + adjustment_value
      max_frequency_adjusted = max_frequency(amino_a, amino_b) - adjustment_value
      return min_frequency_adjusted, max_frequency_adjusted
```