**Real data analysis**

Equipped with the equation for infer the number of HGT events between to genome based on their SI value, we analyzed a large set of real biological data, the EGGnog database. This database contains protein sequences of 1133 species, most of them bacteria. In addition, this database clusters all proteins into COGs (Clusters of Orthologous Groups). By this we could create for each organism a COG-file, i.e. a list of its genes which can serves as the input for SI method. This pre-processing stage is widely described in (9), and there we also described who SI induced 39 native clusters of closely related species. In the following table we present the average SI among each clique, and the average number (in % of the genome size average) of HGT events separating between each pair of species in each clique. We found that this parameter is normal distributed (Shapiro-Wilks test: ) with mean of 52.7%, median of 54.1% and SD of 23.78%. is simple word, we found that the number of HGT events separate between each pair of species inside the cliques is about 50% (±20). This is an interesting finding since SI values themselves are not normal distributed (Shapiro-Wilks test: ).

**Table 1:** Distribution of SI and the estimated number of HGT events among closely related species (cliques). In green- values fit the range of 1SD from the mean (i.e., ). In blue values higher more than 1SD from the mean (). In yellow values lower more than 1SD from the mean ().

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Clique number** | **Genus list** | **Avg SI** | **Clique size** | **Estimated number of HGT events (average)** | **Genome size (average)** | **Number of HGT events as % of genome size** |
| 1 | {'Borrelia'} | 0.554 | 8 | 320.0 | 1158.8 | 27.6 |
| 2 | {'unknow', 'Burkholderia', 'Ralstonia', 'Cupriavidus'} | 0.819 | 25 | 3657.1 | 6367.9 | 57.4 |
| 3 | {'Pelodictyon', 'Chlorobaculum', 'Chlorobium', 'Prosthecochloris'} | 0.851 | 10 | 1478.8 | 2271.3 | 65.1 |
| 4 | {'Shewanella'} | 0.764 | 19 | 2072.1 | 4262.0 | 48.6 |
| 5 | {'Streptococcus'} | 0.856 | 10 | 1331.7 | 2000.9 | 66.6 |
| 6 | {'Rickettsia'} | 0.654 | 13 | 434.3 | 1192.7 | 36.4 |
| 7 | {'Methanococcus'} | 0.632 | 6 | 584.6 | 1721.0 | 34.0 |
| 8 | {'Exiguobacterium', 'Oceanobacillus', 'Macrococcus', 'Bacillus', 'Geobacillus', 'Anoxybacillus', 'Staphylococcus', 'Listeria'} | 0.877 | 25 | 2283.6 | 3202.6 | 71.3 |
| 9 | {'Streptococcus', 'unknow'} | 0.713 | 14 | 863.8 | 2037.3 | 42.4 |
| 10 | {'unknow', 'Corynebacterium', 'Mycobacterium'} | 0.86 | 10 | 1568.4 | 2331.7 | 67.3 |
| 11 | {'Thermotoga'} | 0.501 | 6 | 439.3 | 1867.8 | 23.5 |
| 12 | {'Bartonella', 'Brucella'} | 0.669 | 15 | 961.1 | 2572.6 | 37.4 |
| 13 | {'Rhodopseudomonas', 'Nitrobacter', 'Bradyrhizobium', 'Oligotropha'} | 0.855 | 11 | 3222.2 | 4947.1 | 65.1 |
| 14 | {'Mycobacterium'} | 0.817 | 19 | 2761.3 | 4827.3 | 57.2 |
| 15 | {'unknow', 'Francisella'} | 0.608 | 9 | 518.0 | 1627.3 | 31.8 |
| 16 | {'Staphylococcus'} | 0.226 | 12 | 228.0 | 2647.8 | 8.6 |
| 17 | {'unknow', 'Shigella', 'Cronobacter', 'Serratia', 'Photorhabdus', 'Pectobacterium', 'Citrobacter', 'Klebsiella', 'Salmonella', 'Dickeya', 'Erwinia', 'Sodalis', 'Yersinia', 'Edwardsiella', 'Escherichia', 'Proteus', 'Enterobacter'} | 0.796 | 85 | 2405.4 | 4492.3 | 53.5 |
| 18 | {'Candidatus', 'Buchnera'} | 0.938 | 8 | 605.9 | 486.0 | 124.7 |
| 19 | {'Azotobacter', 'Pseudomonas'} | 0.826 | 18 | 3167.8 | 5382.3 | 58.9 |
| 20 | {'Clostridium'} | 0.749 | 13 | 1630.0 | 3496.0 | 46.6 |
| 21 | {'Photobacterium', 'Vibrio', 'Aliivibrio'} | 0.797 | 14 | 2369.4 | 4410.4 | 53.7 |
| 22 | {'Chlamydia', 'Chlamydophila', 'unknow'} | 0.444 | 14 | 194.2 | 966.0 | 20.1 |
| 23 | {'Lactobacillus', 'Pediococcus'} | 0.884 | 12 | 1574.2 | 2122.0 | 74.2 |
| 24 | {'Xanthomonas', 'Stenotrophomonas'} | 0.717 | 10 | 1864.1 | 4391.6 | 42.4 |
| 25 | {'Desulfotomaculum', 'Candidatus', 'Carboxydothermus', 'Pelotomaculum', 'Moorella', 'Ammonifex'} | 0.923 | 6 | 2172.2 | 2447.6 | 88.7 |
| 26 | {'Neisseria'} | 0.524 | 7 | 518.1 | 2065.3 | 25.1 |
| 27 | {'Agrobacterium', 'Rhizobium', 'Sinorhizobium', 'Ochrobactrum'} | 0.866 | 12 | 4149.9 | 6132.2 | 67.7 |
| 28 | {'Acidovorax', 'Variovorax', 'Delftia', 'Comamonas'} | 0.88 | 6 | 3517.6 | 4910.2 | 71.6 |
| 29 | {'Prochlorococcus', 'Synechococcus'} | 0.706 | 16 | 904.7 | 2179.1 | 41.5 |
| 30 | {'Bifidobacterium'} | 0.787 | 8 | 981.5 | 1857.1 | 52.8 |
| 31 | {'Bacillus', 'Geobacillus'} | 0.612 | 15 | 1687.6 | 5315.8 | 31.7 |
| 32 | {'Streptococcus'} | 0.61 | 19 | 604.3 | 1893.1 | 31.9 |
| 33 | {'Helicobacter'} | 0.528 | 8 | 390.5 | 1531.4 | 25.5 |
| 34 | {'Acinetobacter'} | 0.616 | 7 | 1121.1 | 3481.7 | 32.2 |
| 35 | {'Ehrlichia', 'Anaplasma'} | 0.679 | 8 | 390.0 | 992.9 | 39.3 |
| 36 | {'Geobacter'} | 0.892 | 6 | 2926.0 | 3871.7 | 75.6 |
| 37 | {'Campylobacter'} | 0.76 | 8 | 839.4 | 1721.1 | 48.8 |
| 38 | {'Methylobacterium'} | 0.676 | 6 | 2147.7 | 5681.5 | 37.8 |
| 39 | {'Sulfolobus'} | 0.453 | 6 | 559.6 | 2756.5 | 20.3 |