**Supplementary Table 4: Correlation of genome size with genome G+C% in bacterial groups**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Phylum  (number of bacteria) | Number of bacteria | Pearson *r* (Genome G+C%, Genome Size) | Max Genome Size / Min Genome Size in the group | Max Genome G+C% / Min Genome G+C% in the group |
| δ-proteobacteria (20) | 10 (Low GC) | 0.789 | 2.988 | 1.748 |
| 10 (High GC) | 0.515 | 3.560 | 1.281 |
| Firmicutes (76) | 36 (Low GC) | -0.030 | 3.168 | 1.324 |
| 36 (High GC) | 0.175 | 3.217 | 1.849 |
| Actinobacteria (41) | 20 (Low GC) | 0.591 | 10.463 | 1.446 |
| 21 (High GC) | 0.293 | 3.529 | 1.102 |
| γ-proteobacteria (83) | 41 (Low GC) | 0.533 | 6.264 | 1.413 |
| 42 (High GC) | 0.033 | 2.862 | 1.522 |
| β-proteobacteria (42) | 21 (Low GC) | 0.835 | 6.236 | 1.437 |
| 21 (High GC) | -0.080 | 2.982 | 1.065 |
| α-proteobacteria (72) | 36 (Low GC) | 0.827 | 6.605 | 2.147 |
| 36 (High GC) | 0.147 | 3.116 | 1.190 |

Bacteria belonging to one phylum are divided into two groups having low genome (G+C)% and high genome (G+C)%. In each group, Pearson *r* (Genome G+C%, Genome Size) was calculated. Among all the phyla, no consistent pattern of Pearson *r* values between low and high (G+C)% groups were observed even though genome size and genome (G+C)% were comparable.