Kullback Leibler Divergence in Complete Bacterial and Phage Genomes

Sajia Akhter, Ramy K. Aziz, Mona T. Kashef, Eslam S. Ibrahim, Barbara Bailey, Robert A. Edwards

Supplemental Material

T-test for testing the relationship between GC content and amino acid variations for bacteria and phage samples:

There are two samples: bacteria and phages.
The sample size of bacteria, m = 372
The linear model of bacterial sample: $y = 2x^2 - 2x + 0.5$ … (1)
The sample size of phages, n = 835
The linear model of phage sample: $y = 1.7x^2 – 1.7x + 0.44$ … (2)

The general form of these equations is: $y = \beta_0 + \beta_1 x + \beta_2 x^2$
We want to test whether the coefficients (β_1 and β_2) of these two equations are significantly different:

H_0: ß$_1^1$ = ß$_1^2$ and ß$_2^1$ = ß$_2^2$
H_A: ß$_1^1$ ≠ ß$_1^2$ and ß$_2^1$ ≠ ß$_2^2$
where ß$_1^1$ is the ß$_1$ coefficient of equation 1 (bacterial sample) and ß$_1^2$ is the ß$_1$ coefficient of equation 2 (phage sample).

T test for two independent unequal sample sizes:

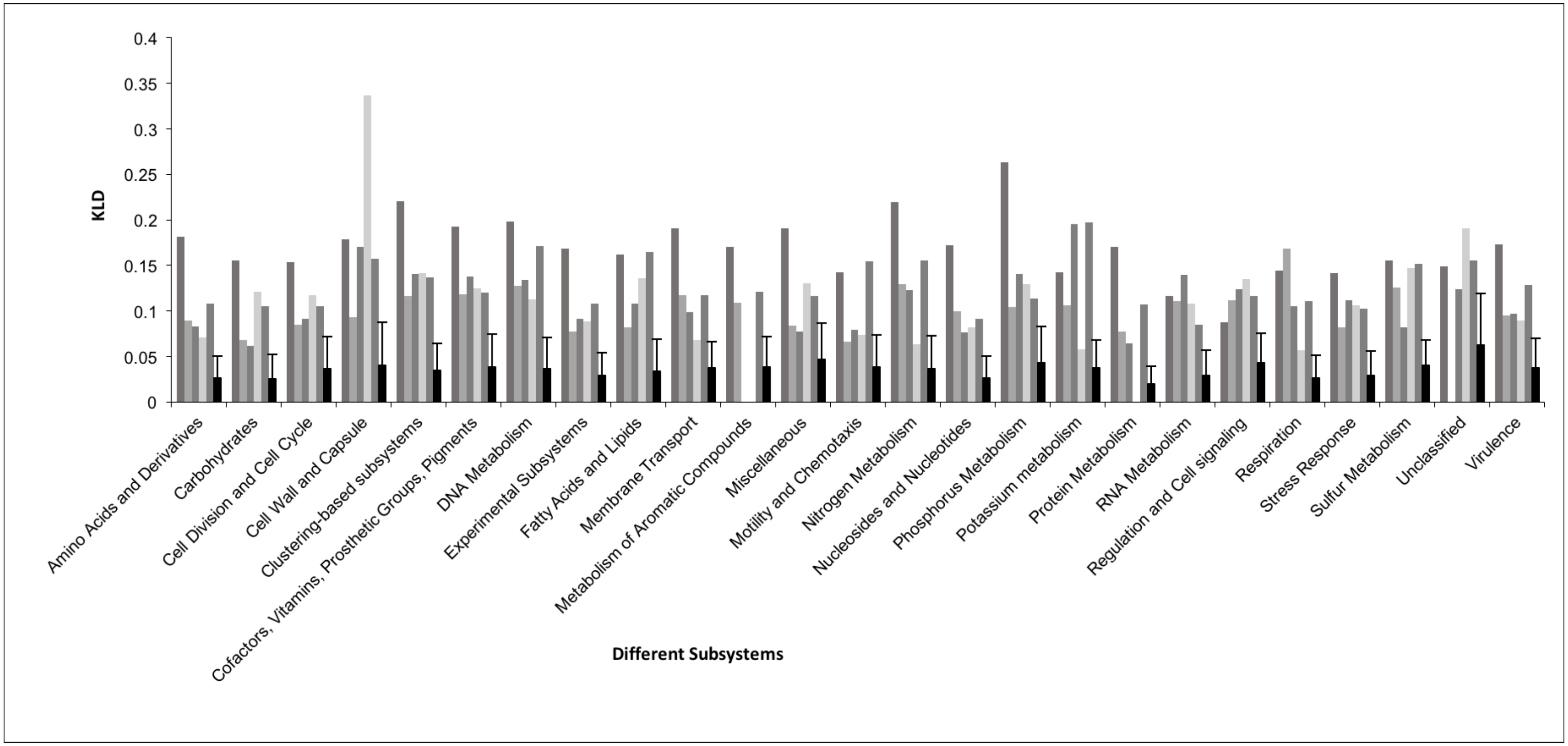
$$t = \frac{\bar{\beta}_p - \bar{\beta}_b}{\sqrt{SE(\beta_p)^2 + SE(\beta_b)^2}}$$, where SE is the standard error.

Here, degree of freedom, df = m-2 + n -2 = 372 -3 + 835 -3 = 1201

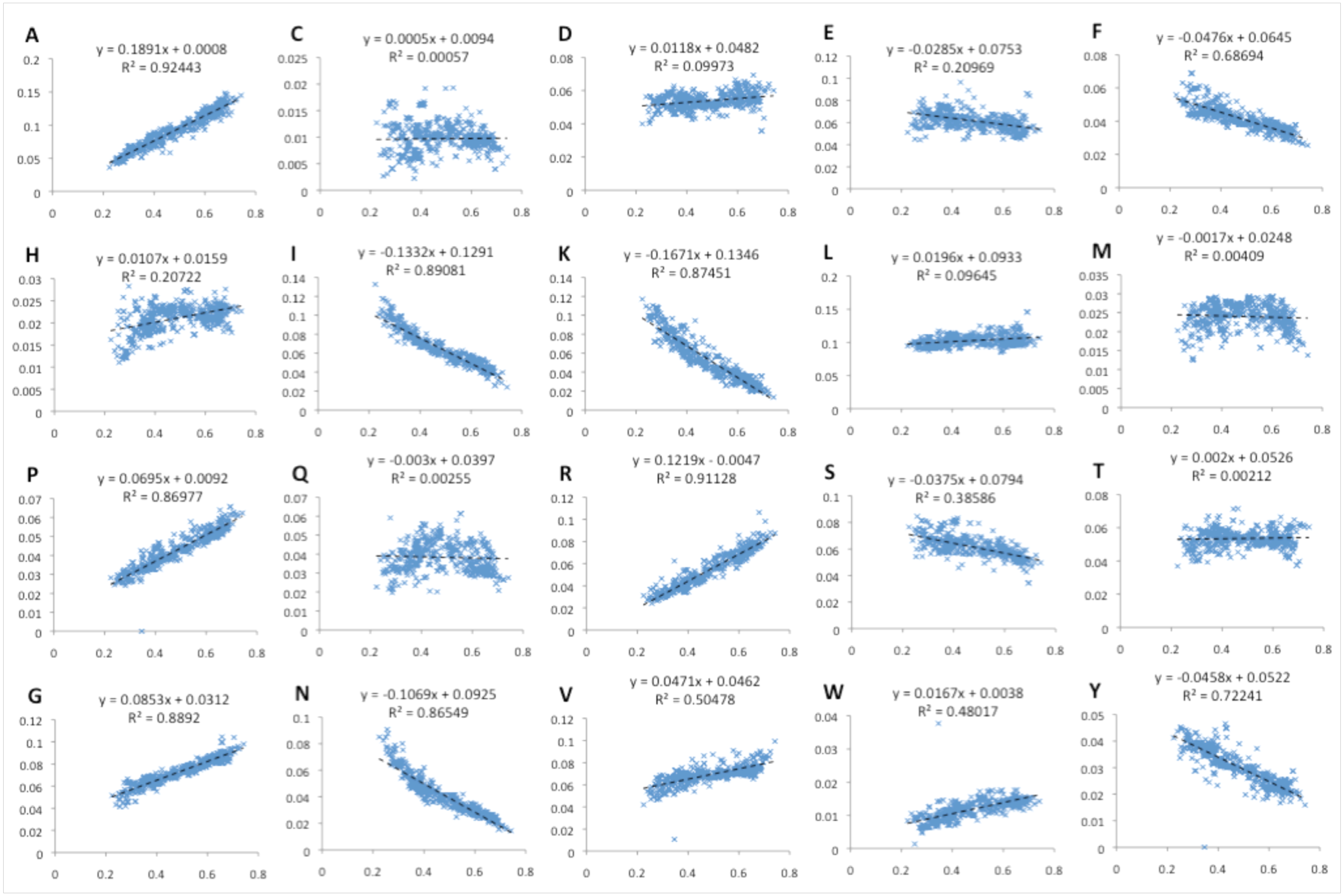
For coefficient ß$_1$, p value is 1.058481e-07, and for coefficient ß$_2$, p value is 1.631291e-06. So we can reject the null hypothesis.

This means Equation 1 and Equation 2 are significantly different.

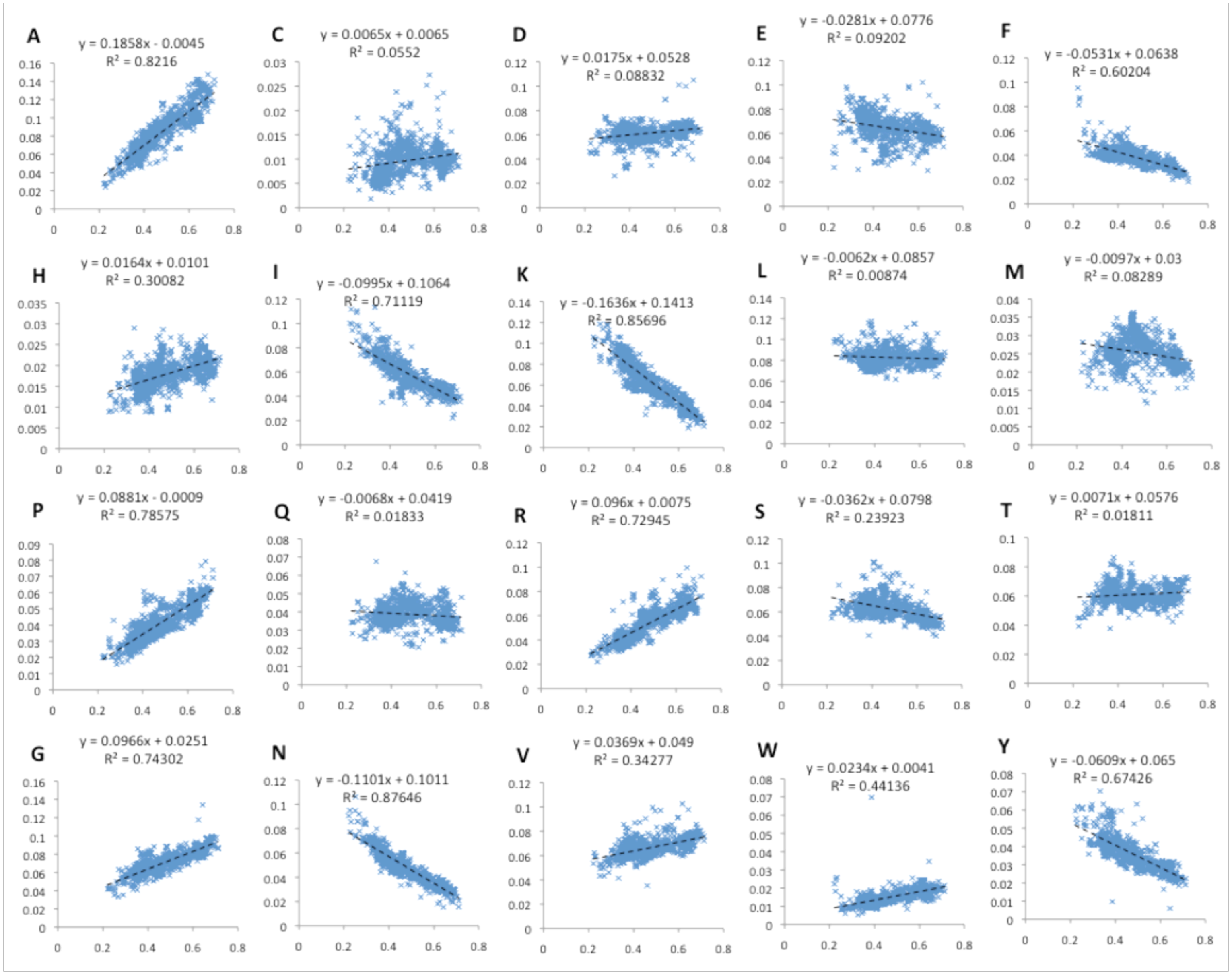
Supplemental Figure 1: Comparison of the divergence of amino acid composition and the phylogenetic group for the most divergent bacterial genomes. The first five bars represent *Wigglesworthia glossinidia*, *Borrelia garinii*, *Mycoplasma mycoides*, *Ureaplasma parvum* serovar and *Buchnera aphidicola*. The sixth bar is for the mean of amino acid utilization for each subsystem.



Supplemental Figure 2: Amino acid frequency of 446 bacterial genomes vs. GC%



Supplemental Figure 3: Amino acid frequency of 835 complete phage genomes vs. GC%


A
y = 0.1858x - 0.0045
R² = 0.8216
C
y = 0.0065x + 0.0065
R² = 0.0552
D
y = 0.0175x + 0.0528
R² = 0.08832
E
y = -0.0281x + 0.0776
R² = 0.09202
F
y = -0.0531x + 0.0638
R² = 0.60204
H
y = 0.0164x + 0.0101
R² = 0.30082
I
y = -0.0995x + 0.1064
R² = 0.71119
K
y = -0.1636x + 0.1413
R² = 0.85696
L
y = -0.0062x + 0.0857
R² = 0.00874
M
y = -0.0097x + 0.03
R² = 0.08289
P
y = 0.0881x - 0.0009
R² = 0.78575
Q
y = -0.0068x + 0.0419
R² = 0.01833
R
y = 0.096x + 0.0075
R² = 0.72945
S
y = -0.0362x + 0.0798
R² = 0.23923
T
y = 0.0071x + 0.0576
R² = 0.01811
G
y = 0.0966x + 0.0251
R² = 0.74302
N
y = -0.1101x + 0.1011
R² = 0.87646
V
y = 0.0369x + 0.049
R² = 0.34277
W
y = 0.0234x + 0.0041
R² = 0.44136
Y
y = -0.0609x + 0.065
R² = 0.67426