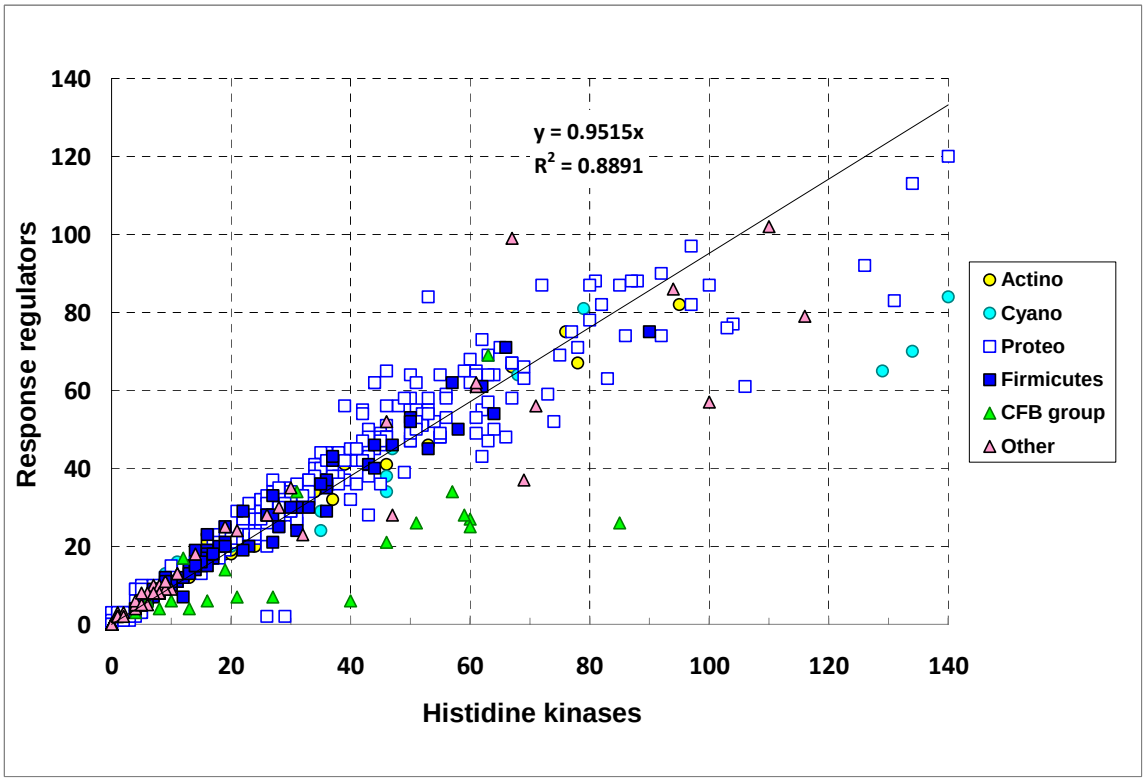
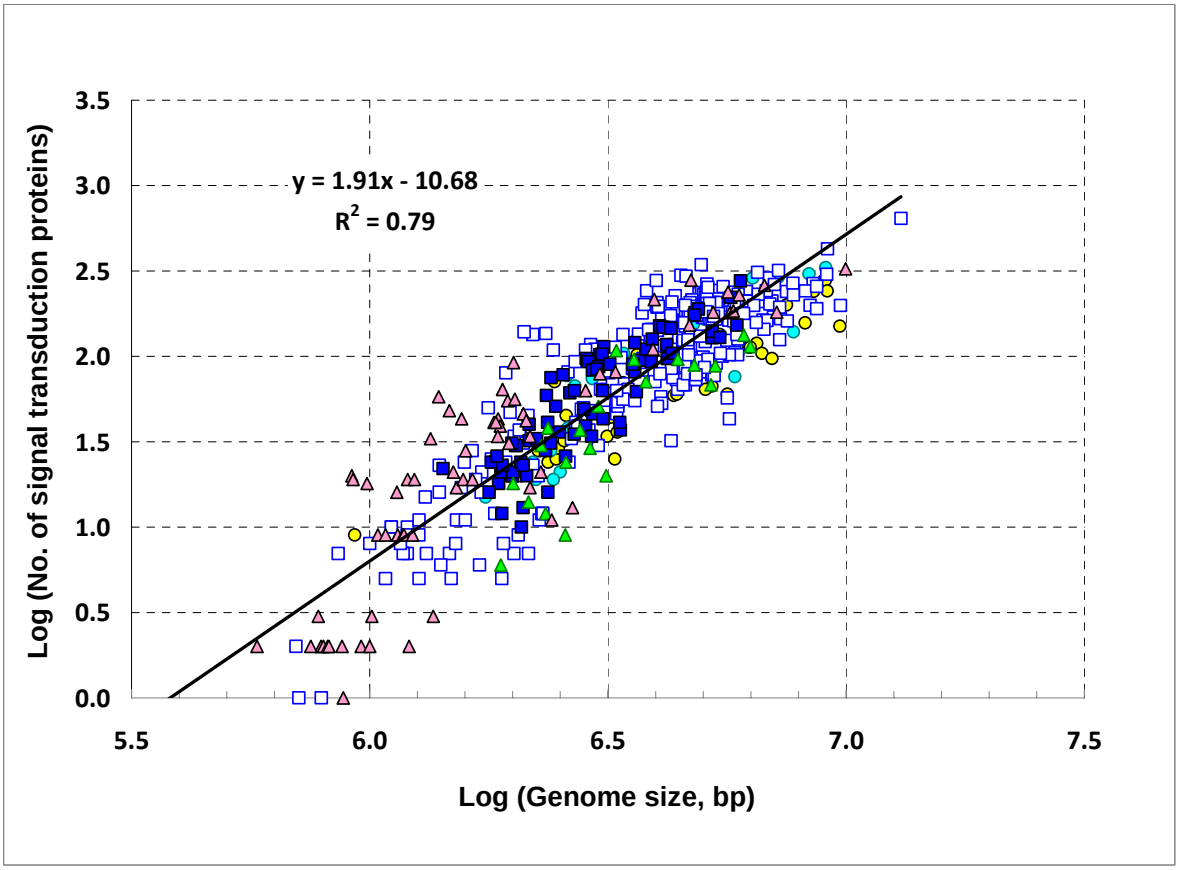


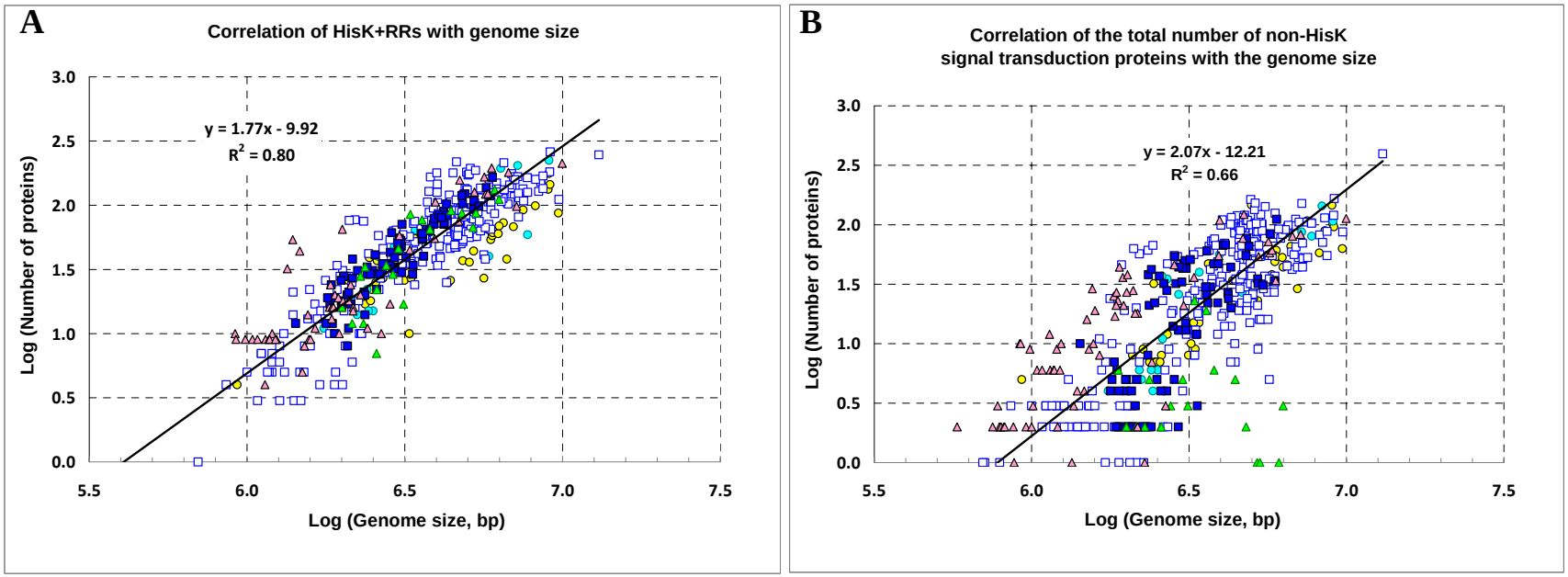
Supplementary Figure S1. Correlation between the numbers of histidine kinases and response regulators encoded in the same bacterial genomes.

The numbers of encoded sensor histidine kinases and response regulators are nearly identical in the members of the bacterial phyla *Actinobacteria* (yellow circles), *Cyanobacteria* (cyan circles), *Proteobacteria* (empty squares), *Firmicutes* (blue triangles) and members of other phyla (pink triangles), except for the members of the CFB group (phyla *Bacteroidetes* and *Chlorobi*, green triangles). In the latter group, the correlation still holds (R^2 ~0.57) but only half of histidine kinases have cognate response regulators. In *Bacteroidetes*, many sensor kinases are fused to receiver domains (forming 'hybrid' kinases) and, sometimes, additionally contain C-terminal AraC-type DNA-binding domains (forming truly 'one-component' signal transduction systems, see Ulrich *et al*., 2005). The two squares at the bottom relate to the two strains of *Orientia tsutsugamushi*, strain Boryong (26 HisKs, 3RRs) and strain Ikeda (29 HisKs, 3 RRs).



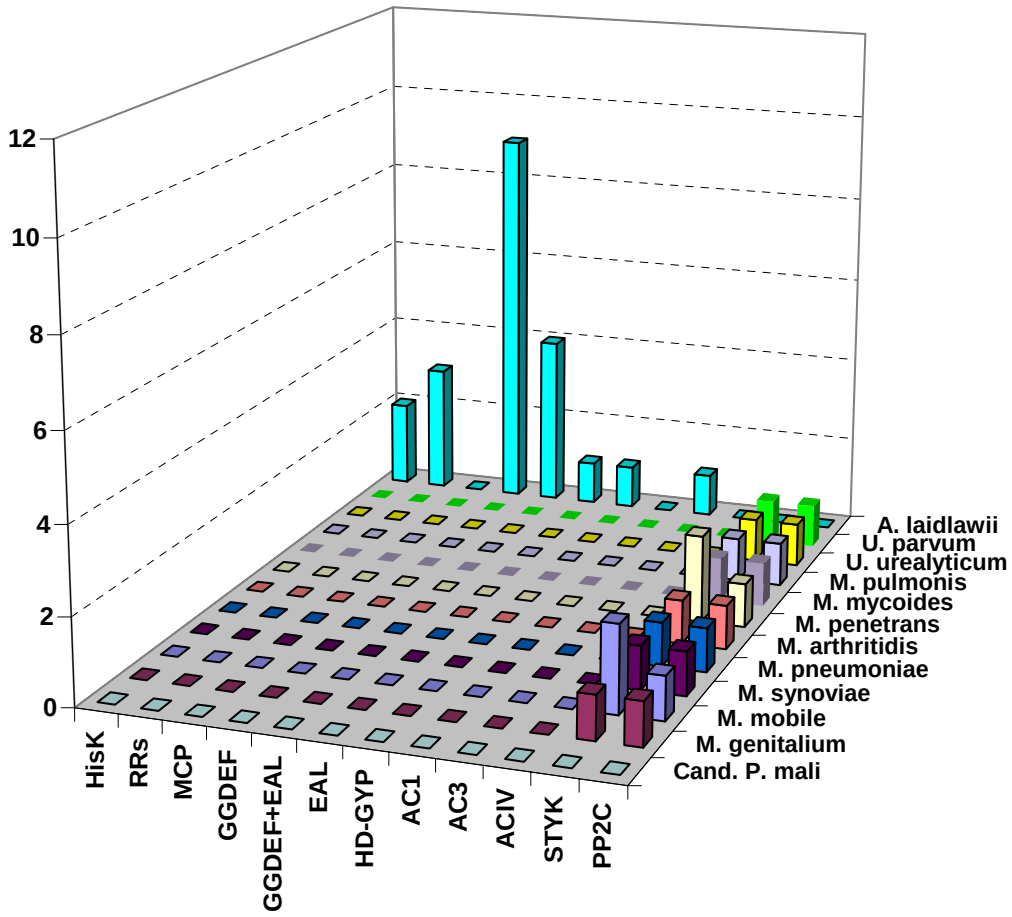
Supplementary Figure S2. Correlation between the organism's genome size and the number of encoded signal transduction proteins.

The total numbers of encoded signal transduction proteins were calculated for the members of the bacterial phyla *Actinobacteria* (yellow circles), *Cyanobacteria* (cyan circles), *Proteobacteria* (empty squares), *Firmicutes* (blue triangles), members of the CFB group (phyla *Bacteroidetes* and *Chlorobi*, green triangles) and members of other phyla (pink triangles). For the purposes of this plot, the organisms encoding no signal transduction proteins were ignored (undefined log0).



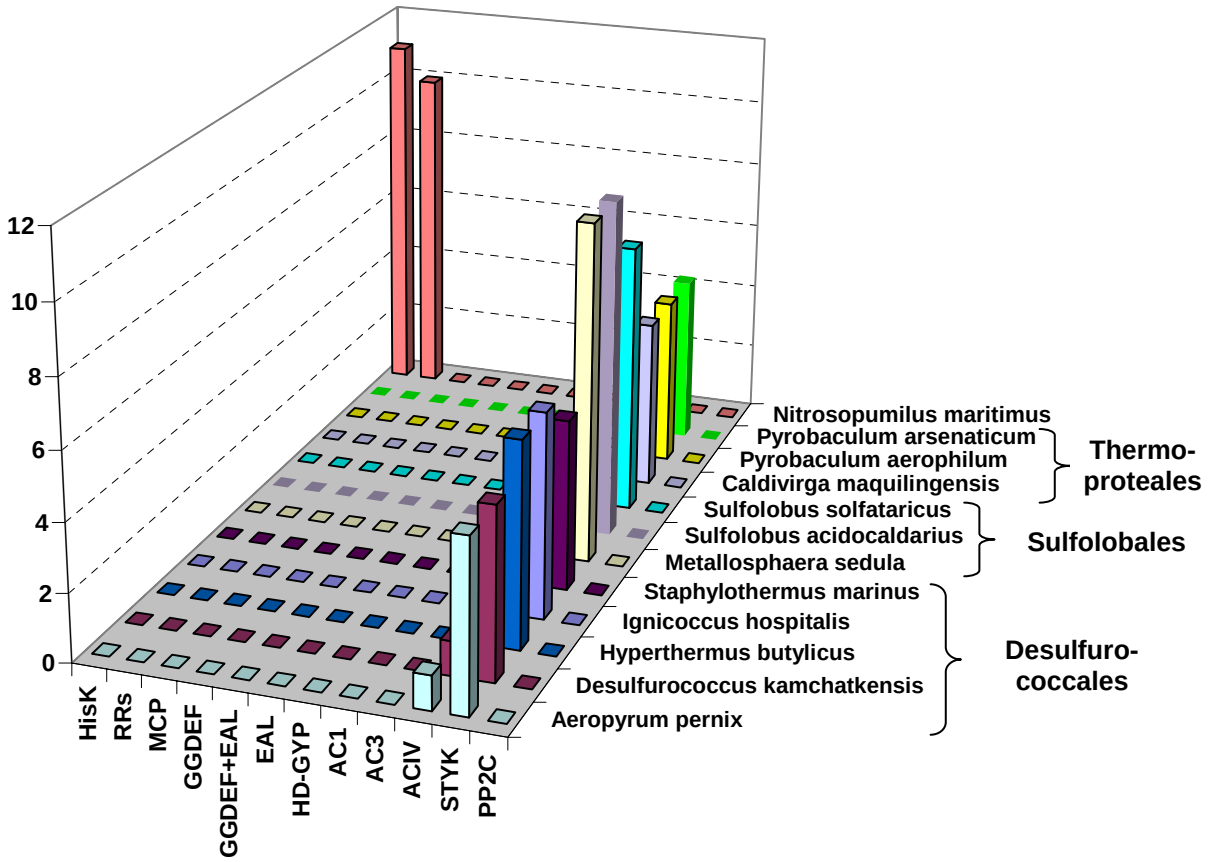
Supplementary Figure S3. Correlation between the organism's genome size and the number of (A) encoded histidine kinases and response regulators and (B) other signal transduction proteins.

The symbols indicate members of the bacterial phyla *Actinobacteria* (yellow circles), *Cyanobacteria* (cyan circles), *Proteobacteria* (empty squares), *Firmicutes* (blue triangles), CFB group (phyla *Bacteroidetes* and *Chlorobi*, green triangles), and members of other phyla (pink triangles). For the purposes of these plots, the organisms encoding no histidine kinases and response regulators (A) or any other signal transduction proteins (B) were ignored (undefined log0), leaving, respectively, 500 genomes in panel A and 541 genomes in panel B.



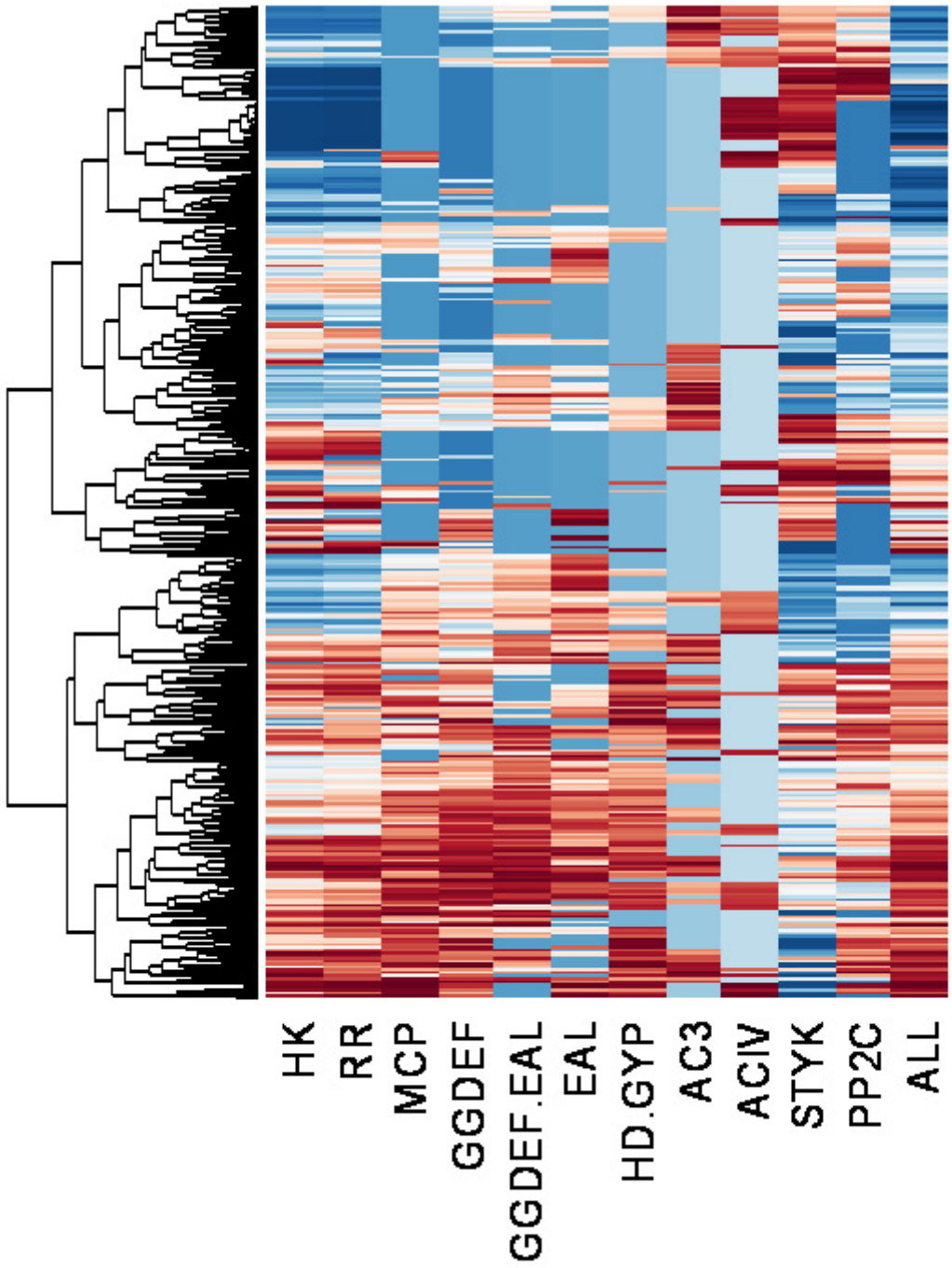
Supplementary Figure S4. Family profiles of signal transduction proteins encoded in the genomes of several representatives of *Tenericutes* (mollicutes).

The organisms' full names in the public genome databases are as follows: *Candidatus* Phytoplasma mali, *Mycoplasma genitalium* G37, *Mycoplasma mobile* 163K, *Mycoplasma synoviae* 53, *Mycoplasma pneumoniae* M129, *Mycoplasma arthritidis* 158L3-1, *Mycoplasma penetrans* HF-2, *Mycoplasma mycoides* subsp. *mycoides* SC str. PG1, *Mycoplasma pulmonis* UAB CTIP, *Ureaplasma urealyticum* serovar 10 str. ATCC 33699, *Ureaplasma parvum* serovar 3 str. ATCC 700970, *Acholeplasma laidlawii* PG-8A.



Supplementary Figure S5. Family profiles of signal transduction proteins in *Crenarchaeota*.

The profiles of all organisms except for *Nitrosopumilus maritimus* include from 5 to 10 'eukaryotic-type' Ser/Thr protein kinases and a putative thermophilic adenylate cyclase (a member of CYTH superfamily whose activity in archaea has not been documented). *Nitrosopumilus maritimus* does not encode either Ser/Thr protein kinases or an adenylate cyclase but encodes 11 bacterial-type sensor histidine kinases and 10 response regulators, of which 8 are single-domain response regulators (stand alone receiver domains), see http://www.ncbi.nlm.nih.gov/Complete_Genomes/SignalCensus.html.



Supplementary Methods

to the paper by Michael Y. Galperin, Roger Higdon & Eugene Kolker "Interplay of heritage and habitat in the distribution of bacterial signal transduction systems"

1. Calculating the IQ scores

The number of sensor proteins and response regulators (N_{SP+RR}) was modeled as a function of the total number of proteins (N_T) in each organism. Both variables were transformed to the log scale to remove the skew from their distributions and because on the log scale their relationship was nearly linear (Figure 1). To account for organisms with no signal transduction proteins (undefined logarithm of zero) or with a total of one signal transduction proteins (Log1=0), they were removed from the from the IQ calculation and there IQ scores were assumed to be zero[1]. This relationship was fit using least squares regression resulting in the following regression line estimate:

$$\log(N_{SP+RR}) = 1.91 \cdot \log\ N_T\ + 11.3; R^2 = 0.76$$

IQ scores prior to normalization were measured by the residuals from this regression; however the slope was rounded to 2 for simplicity. This led to the non-normalized IQ scores:

$$IQ_{UN} = \log\left(\frac{N_{SP+RR}}{N_T^2}\right)$$

The scores were then normalized so that the minimum is 0 and the mean is 100, resulting in the final IQ formulation:

$$IQ = 50.3 \cdot \log\left(\frac{N_{SP+RR}}{N_T^2}\right) + 701.5,$$

which results in a fairly symmetric distribution centered about 100 (see the figure below). For simplicity, the normalizing constants were rounded, resulting in the following easy formula for calculating IQ scores in various bacteria and archaea:

$$\mathbf{IQ = 50*(\ln N_{ST} - 2\ln N_T) + 700}$$

[1] An alternative solution to the problem of zero counts, adding a small constant, for example, 0.5, to the number of proteins counted in every genome, did not result in a better fit.