A new method for using co-conservation predict protein function

Anis Karimpour-Fard¹, Corrella Detweiler², Ryan T. Gill³, and Lawrence Hunter¹

¹University of Colorado School of Medicine
²MCD-Biology, University of Colorado, Boulder
³Department of Chemical and Biological Engineering, University of Colorado, Boulder

anis.karimpour-fard@uchsc.edu
http://compbio.uchsc.edu/Hunter
The problem …..

Close to 400 Microbial genomes are fully sequence and there is high percent of genes with unknown function.

The meaning of protein function

The function of protein A is its action on Substrate to form a Product

Biochemical view

The function of A is the context of its interactions with other proteins in the cell

Post genomic view
Pairs of genes that are present or absent together in genomes.

Conclusion: p2 and p7 are functionally linked. p3 and p6 are functionally linked.

Protein i: 110001111001001110001111
Protein j: 111000111110000110001111
19 matching bits out of 24
Network of Bacillus color coded based on gene function

Cell envelope and cellular process (red), Intermediary metabolism (green), Information pathway or central dogma (yellow), Uncategorized (gray), and others (blue),
The unique differences between an organism of choice and each organism are shown in red and conserved genes across species are shown in gray. Org (organism); org0 (organism of choice); P (protein).
Multiple Species Cluster Co-conservation Network

Bacillus
Salmonella
E. coli