

Comparative genomics: From genome sequences to genome biology

Michael Y. Galperin

NCBI, National Institutes of Health, USA

Genome biology aims at using the complete genome sequences to reconstruct all metabolic and signaling pathways that could operate in the target organisms and identify the likely regulatory hubs and potential drug targets. Such analysis requires comprehensive functional annotation of all proteins encoded in each sequenced genome. Standard sequence analysis typically fails to provide (confident) functional assignment for at least a third of the genes even in the relatively small prokaryotic genomes. As a result, comparative genomics has to deal with the constantly growing numbers of "hypothetical" proteins whose functions remain unknown. This talk will discuss using comparative genomics to improve our understanding of microbial metabolic and signaling pathways, including some recent examples of identification of "missing" enzymes and prediction of alternative enzyme variants. It will show that the number of truly enigmatic "conserved hypothetical" proteins is relatively small, particularly in the reduced genomes of pathogenic bacteria, which suggests that most of their cellular functions are already accounted for. In contrast, the number of uncharacterized genes in free-living organisms remains quite large and their functions remain obscure. Our current hypothesis is that many of these genes have "house-cleaning" function, which is almost as important as house-keeping, particularly for aerobic bacteria and for eukaryotic cells. We shall also briefly discuss how comparative genomics could be used for identification of priority targets for future research and the challenges in characterization of their functions.

Biography

Michael Y. Galperin completed his Ph.D. at the Lomonosov Moscow State University in Russia and received postdoctoral training at the University of Louisville, University of Connecticut, and the National Institutes of Health. Since 1999, he has been a Staff Scientist (now Lead Scientist) at the National Center Biotechnology Information (NCBI), a division of the National Library of Medicine at the U.S. National Institutes of Health. He has published two books and 170 papers on various aspects of microbial genomics. He currently serves as Executive Editor of the Nucleic Acids Research annual Database Issue and an Editor of Environmental Microbiology.

galperin@ncbi.nlm.nih.gov