

**The Second Georgia Tech
International Conference on
Bioinformatics**

***In silico* Biology
Sequence & Structure
& Function**

Atlanta, Georgia, USA

November 11-14, 1999

An Automated Comparative Analysis of Seventeen Complete Microbial Genomes

Arvind K. Bansal

Department of Mathematics and Computer Science, Kent State University, Kent, Ohio, USA

As sequenced genomes become larger and sequencing becomes faster, there is a need to develop accurate automated genome comparison techniques and databases to facilitate derivation of genome functionality; identification of enzymes, putative operons, and metabolic pathways; and to derive phylogenetic classification of microbes. This paper [3] modifies and extends an automated pair-wise genome comparison technique [1, 2] used to identify orthologs and gene-groups to derive orthologous genes in a group of genomes, to identify genes with conserved functionality, and to identify genes specific to groups of genomes. Seventeen microbial genomes archived at <ftp://ncbi.nlm.nih.gov/genbank/genomes> have been compared using the extended technique to derive orthologs, orthologous gene-groups, duplications, gene-fusions, genes with conserved functionality, and genes specific to groups of genomes.

The comparison results [3] for *E. coli* and *B. subtilis* two of the microbes thoroughly explored in wet laboratories are consistent with the NCBI annotations. The results reveal that the genomes within the same family have a higher percentage of orthologs and orthologous gene-groups in terms of size of the smaller genomes in the genome-pairs. However, genome-pairs with large number of genes share a large number of orthologs and orthologous gene-groups. There are large numbers of gene-group duplications and duplications of single genes. Duplication of gene-groups is largely a function of genome size, and to a lesser extent is a function of genomes being in the same family. The duplication of single genes is random for some genomes. Fused genes are small in number. Around 85 genes have conserved function. The functions of many genes involved in transcription and translation are conserved. 21 genes corresponding to ribosomal proteins have no orthologs in archaea microbes. Archaea genomes share a relatively higher percentage of orthologs among themselves. There are number of genes which are specific to *E. coli* and various subsets of eight pathogens.

References

[1] Bansal, A. K., Bork, P., and Stuckey, P., "Automated Pair-wise Comparisons of Complete Microbial Genomes", *Mathematical Modeling and Scientific Computing*, 9, 1 - 23, (1998).

[2] Bansal A. K., and Bork, P., "Applying Logic Programming to Derive Novel Functional Information in Microbial Genomes," Lecture Notes in Computer Science, Springer Verlag, 1551, 274 - 289, (1999).

[3] Bansal, A. K., "An Automated Comparative Analysis of Seventeen Complete Microbial Genomes", Bioinformatics, in press.