Eleventh International Conference on Mathematical and Computer Modelling and Scientific Computing

March 31 - April 3, 1997
Georgetown University Conference Center
Washington, DC, U.S.A.

\[ \rho \frac{\partial V_a}{\partial t} + \rho \left( V_R \frac{\partial V_R}{\partial R} + \frac{V_a}{R} \frac{\partial V_a}{\partial \alpha} + \frac{V_a V_R}{R} + \frac{V_0}{R \sin \alpha} \frac{\partial V_a}{\partial \theta} - \frac{V_0^3 \cot \alpha}{R} \right) \]
A FRAMEWORK FOR AUTOMATED COMPARISON OF COMPLETE
PROKARYOTIC GENOMES

Arvind K. Bansal (1), Peer Bork (2) and Peter J. Stuckey (3)
(1) Kent State University, Kent, Ohio, USA
(2) European Molecular Biology Laboratory (EMBL), Heidelberg, Germany
(3) University of Melbourne, Parkville, Australia

An automated pair-wise cross-species comparison of complete genome will provide a
cost-effective and time-efficient computational tool to derive the functions of genes, gene
conservation, mechanism of evolution, metabolic pathways and their variations within various
organisms. In this paper, we describe a framework for an automated comparison of complete
prokaryotic genomes to derive putative orthologues, paralogues, gene-groups, fused genes, and
gene-replications. The framework has three stages: blast comparison to filter gene-pairs with high
similarity, matching of filtered gene-pairs using Smith-Waterman's algorithm for pair-wise
alignment, and graph matching and fuzzy logic techniques to resolve ambiguities in identifying
orthologues and paralogues, gene-groups, fused-genes, and gene-replications. The application of
graph matching technique is the novel contribution in the paper. The graph matching stage
represents pair of complete genomes as bipartite graphs with each gene as a node, and develops
variations of stable marriage techniques in a bipartite graph to identify gene-groups, and resolve
ambiguities in gene-pair matching to identify orthologues and paralogues. Algorithms and a
general purpose software has been described. The results derived from automatic comparison
describing the presence of putative orthologues and gene-groups and fused-genes and
gene-replication have been discussed briefly.