



Development of NASA-Specific Bioinformatics Environment

Andrew Pohorille, NASA-ARC

Biomolecular Systems Research Program

The proposed work has two main objectives:

- To design, built, deploy and test bioinformatics environment needed for solving NASA-specific biology problems. This environment will consist of tools, databases and knowledge management system suited for NASA goals in space genetics, fundamental biology and astrobiology. It will provide a unified, integrated platform for NASA scientists and their collaborators to share data, information and analyses and, by doing so, greatly improve the capabilities for interdisciplinary, collaborative research.
- To enhance the capabilities of the environment by adapting to the needs of bioinformatics several sophisticated tools, originally developed at NASA to support other research areas, and by integrating bioinformatics capabilities with massively parallel computational environment at NASA-Ames.

Description

- a. We will develop or extend NASA-specific algorithms, such as bayesian classification, and pattern recognition to problems of expression analysis.
- b. We will compile a database of genomic, structural and expression data specific to the needs of NASA researchers, and develop an integrated knowledge management platform, which will allow scientists to collaborate and access these data. Such tools currently do not exist in the public domain, and are in an early developmental stage in industry.
- c. We will create a novel, interactive, collaborative interface between bioinformatics tools and 3D parallel graphics.

We also expect a combination of experimental results and the newly developed and applied tools of bioinformatics will lead to significant advances in the understanding of how microgravity affects physiology and improved understanding of fundamental processes required for life.

Innovative Claims/NASA Significance

The proposed research will have applications to virtually every NASA biology program that involves genomics and proteomics studies. Examples of such studies are:

1. Flight and ground based experiments on the response of organisms to conditions in space,
2. Reconstruction of evolutionary history of organisms from genomic data,
3. Examination of ecological coexistence of microorganisms,
4. Comparison between organisms living in ambient and extreme conditions in order to determine limits of life.

Plans

Deliverables at the end of the first year will consist of a working bioinformatics infrastructure, accessible to users through a web-interface. Programs to perform homology search, multiple alignment, motif search and analysis, Hidden Markov Models, sequence editing and utility tools, and database access will be provided. Software and data will be implemented on an 8-16 node Linux cluster. A retrieval and storage mechanism will be in place to update automatically data from public servers. The initial version of the interface with the visualizations software will be deployed.

Deliverables at the end of the second year will consist of automated processes to perform analysis on user data and will be interfaced to a relational database system. The clustering of protein databases and subsequent HMM construction, and protein annotation will be completed and the clusters/models will be validated. Information will be incorporated in to a relational database and will be available online. The models will be available to help annotate unknown sequences from organisms of interest to NASA, and as a tool for researchers. In addition, the evaluation of current methods for expression analysis will be performed, and the applicability of NASA-specific tools to this problem will be evaluated.

Deliverables at the end of the project will consist of an integrated informatics platform accessible through a web interface, which will house all data from sequence analysis, microarray, and structural data. The platform will be fully integrated with Ames' high performance computing facilities and collaborative, interactive 3D visualization environment. It will allow users to input data into the database, and to modify and share data between users. We will also develop new techniques and databases aimed at understanding the effects of microgravity on living organisms and identifying genes in a large number of novel organisms (for eco-genomics).