BIOINFORMATICS

Mentor: Boumediene Belkhouche
Electrical Engineering and Computer Science Department
Tulane University
301 Stanley Thomas Hall
New Orleans, LA 70118
(504) 862-3372
E-mail: bb@eecs.tulane.edu

Overview:

The major objectives of this project are: (1) development of a local database; (2) analysis of data contained in the database; (3) and visualization.

1. Development and Maintenance of Databases

The simplest tasks used in bioinformatics concern the creation and maintenance of databases of biological information. Nucleic acid sequences (and the protein sequences derived from them) comprise the majority of such databases. While the storage and organization of millions of nucleotides is far from trivial, designing a database and developing an interface whereby researchers can both access existing information and submit new entries is only the beginning. Further, there is a need to induce maximum amount of intelligence in the controlling software so that the storage and maintenance tasks can be performed autonomously with minimal human intervention.

2. Analysis

The power of a database comes not from the collection of information, but in its analysis. A sequence of DNA does not necessarily constitute a gene. It may constitute only a fragment of a gene or alternatively, it may contain several genes. This task involves the analysis of sequence information and it involves the following sub-tasks:

- Finding the genes in the DNA sequences of various organisms
- Developing methods to predict the structure and/or function of newly discovered proteins and structural RNA sequences.
- Clustering protein sequences into families of related sequences and the development of protein models.
- Aligning similar proteins and generating phylogenetic trees to examine evolutionary relationships.

3. Visualization

Visualization is used to enhance analysis, pattern discovery, and data mining. Visual tasks include statistical displays, as well as graphical modeling of biological entities (e.g., genes, proteins, organs).

Prerequisites

Computer Science: data structures (CPSC 118 at Tulane)