Final Technical Report
Foundation for a syntactic pattern recognition system for genomic DNA sequences

The goals of this project were to establish database access to several databases of interest, customizing bioJava for database applications, improving performance of bioJava applets, development of additional applets, and maintenance of GenLang for web applications.

Database access to several databases of interest including ASN.1 Entrez and AceDB was established as proposed by adapting drivers developed for Kleisli to Java-based versions. These drivers utilized the Java RMI. These drivers are used with a Java implementation of Kleisli called K2. K2 drivers to a relational database were used in a web interface to query DoTS, a prototype database of transcribed sequences.

The components for building bioJava applets underwent a major design and implementation change to achieve the goals of database customization, performance improvements, and new applications. The bioJava components were designed as Java beans or "biowidgets" that could be reutilized to build a variety of applications. The applications built include MapView, SeqView, and BlastView. MapView provides a graphical depiction of features on biological sequences such as the location of exons or repeats. SeqView provides biological sequences in an interactive window that can share objects with MapView. The result is that sequence features highlighted in MapView are highlighted on the sequence in SeqView. BLASTView incorporates both MapView and SeqView to graphically display results of a BLAST sequence similarity search. In addition, BLASTView also incorporates an alignment biowidget for display of multiple sequences aligned by different methods. Each of these applications have been designed to interact directly with data sources such as a relational database. A description of the biowidget architecture and sample application was published in Fischer et al Bioinformatics 1999.

Maintenance of GenLang for web applications was accomplished using the EpoDB web site. EpoDB is an erythropoiesis database (www.cbil.upenn.edu/EpoDB) containing highly curated sequence entries for genes expressed during vertebrate erythropoiesis. The site allows retrieval of sequences from a specified region such as the promoter. GenLang is used to search for instances of user-specified sequence motifs. The motifs can be in either orientation and may have one variation from the sequence. Furthermore, two motifs can be specified and searched for using GenLang.

In summary, substantial progress was made on the four goals specified in the revised project description.

DISCLAIMER

DISCLAIMER

**Portions of this document may be illegible
in electronic image products. Images are
produced from the best available original
document.**