

## Poster H-84

### A functional genomics sequence analysis workbench



#### Authors:

CA Hefer (*University of Pretoria*)

F Joubert (*University of Pretoria*)

**Short Abstract:** A web-based biological sequence analysis workbench and management system integrated within a functional genomics platform is described. In addition to providing access to analysis packages, both locally and through web-services, relationships between functional genomic data types can be investigated, enabling users to identify patterns across different data types.

#### Long Abstract:

Keeping track of the multitude of sequence data being produced in modern laboratories is an ever increasing problem biologists are facing. In the post-genomics era scientists are confronted with a vast amount of biological data, especially sequence data, with which biological problems can be approached. Using the knowledge derived from functional genomic data types, sequence information serves as a basic link to derive relationships between genomic data types, and has the potential to reveal interesting biological interactions.

Analyzing biological sequences, albeit DNA or protein sequences, are daily tasks performed by biologists, but the myriad of analysis tools available, and the results produced by these tools are generally stored in a haphazard fashion on local computers. Laboratory data management and storage is generally not centralized, isolated from public data sources and not integrated with results from other functional genomic data types.

This poster describes the design and implementation of a web-based sequence data management system and analysis workbench in which complete control over the storage, analysis and annotation of the sequence data in the system is granted to the user. The system enables sequence capturing and collection from a typical small/medium sequence facility (or importing from local databases), quality assessment of the data, archiving, analysis and the annotation of the sequences. The system implements and extends the sequence ontology feature annotation (SOFA) data schema, which allows for the unambiguous annotation of sequence features.

Functionality of the system allows users to utilize analysis packages and tools, both locally and through web-services and to import customized sequence analysis pipelines and workflows to automate certain analysis procedures. A software agent will also be deployed on sequence data, exploring and searching for sequence features, similarities, homologies, annotations and relationships between different functional genomic data types and data sets.

The sequence analysis workbench forms part of a complex Functional genomics information management system (FunGIMS) currently in development at the University of Pretoria and aims to integratively manage various functional genomic data types.