

Poster H-62

Chisel - a Framework for Identification and Characterization of Taxonomic and Phenotypic Versions of Enzymes



Authors:

Alexis Rodriguez (*Argonne National Laboratory*)

Yi Zhang (*Argonne National Laboratory*)

Natalia Maltsev (*Argonne National Laboratory*)

Elizabeth Marland (*Argonne National Laboratory*)

Short Abstract: Chisel is a bioinformatics environment that supports both automated and expert-driven identification, characterization and comparative analysis of taxonomic and evolutionary variations of enzymes. Application of Chisel for the needs of biomedical research, bioremediation, high-throughput analysis of genomes and metagenomes will be discussed.

Long Abstract:

All organisms share a common ancestor, however, in the process of adaptation to the environment they develop different variations of molecular function, metabolic organization and phenotypic features. Availability of large volumes of sequence and enzymatic data for taxonomically and phenotypically diverse organisms allows for systematic exploration of adaptive mechanisms that led to diversification of enzymes (e.g. variations in their enzymatic properties, subunit composition, cofactor preferences and other).

The Bioinformatics group at the Mathematics and Computer Science division of Argonne National Laboratory has developed Chisel available at <http://compbio.mcs.anl.gov/CHISEL>. This computational framework and its algorithmic approach allows for identification, characterization and comparative analysis of evolutionary variations of enzymes. Analysis of enzymatic sequences in Chisel provides basis for reasoning about evolutionary history of particular enzymatic function. It also allows mapping particular implementations of the same function onto metabolic networks, and approaching the questions such as “what variants of the same enzymatic functions have preferential use in certain phylogenetic neighborhoods or in particular ecological niche?”

The identification of taxonomic and phenotypic variation of enzymes also facilitates the development of a number of biomedical applications such as, bacteriological diagnostics, identification of non-culturable organisms, interpretation of meta-genomes, and correction of annotation errors.

Chisel was used for automated annotation of microbial genomes in the framework of the PUMA2/Gnare system (<http://compbio.mcs.anl.gov/gnare>); for identification of the taxonomy – specific enzymes and metabolic pathways in the analysis of Hanford site metagenome (<http://compbio.mcs.anl.gov/PNNL1>) and for identification of potential antibacterial drug targets for the NIH Great Lakes RCE.

The current release of Chisel contains 8771 taxonomy-specific clusters and subsequent libraries of models that represent 1334 enzymatic functions. These models are used by Chisel classification tools and for the development of biomedical applications. Every cluster in Chisel is annotated with information regarding protein features (e.g. active sites, conserved domains and motifs), connection of enzymatic functions to the corresponding metabolic pathways and structural information. Chisel also supports comparative analysis of sequences in the framework of taxonomic and phenotypic information.

The Chisel framework is composed of:

- a) Enzymatic Knowledge base (EKB) – an integrated database containing sequence data and annotations, metabolic and enzymatic information, structural, taxonomic and phenotypic information from over 20 public databases and literature. EKB also includes results of analysis of enzymatic sequences using a variety of bioinformatics tools (e.g. Blast, Blocks, TMHMM). EKB provides a basis for the rules-based hierarchical clustering of the enzymatic sequences and their annotation.
- b) Rules-based clustering algorithm for identification and characterization of taxonomic and phenotypic variations of enzymatic functions. Clustering is based on the following criteria: composition of conserved functional domains and motifs, sequence length, variety of sequence features and phylogenetic analysis of enzymatic sequences.
- c) Library of models for taxonomic and phenotypic variations of enzymes represented by HMM profiles and PSSMs to be used by the tools for classification and characterization of proteins and prediction of gene functions; and a library of taxonomy-specific PCR primers and oligonucleotides to be used by experimentalists.
- d) Flexible user interface that allows for interactive and automated analysis of protein families and annotation.

The presentation will provide more detailed information regarding Chisel algorithm, framework, capabilities and offer some examples of its use for analysis of scientific data.