

METABOLIC NETWORK ANALYSIS OF APICOMPLEXAN PARASITES

A PROBABILISTIC METHOD FOR IMPROVED ENZYME PREDICTION

Stacy Hung (Ph.D. Candidate)

Program in Molecular Structure and Function, Hospital for Sick Children, Toronto

Department of Molecular Genetics, University of Toronto

December 7, 2008

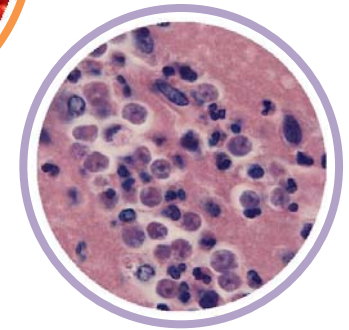
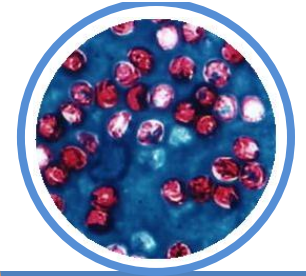


Motivation

- Apicomplexa
 - ▣ Unicellular eukaryotic parasites
 - ▣ Food-borne, water-borne illnesses
 - *Cryptosporidium spp.*, *Toxoplasma gondii*
 - ▣ Malaria
 - 500 million cases and 1 million+ deaths annually
 - *Plasmodium spp.*

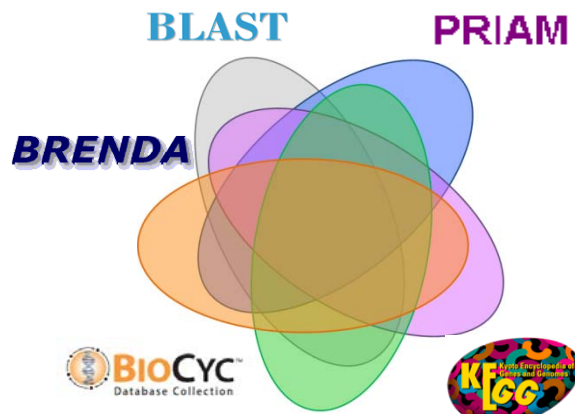
- Lack of preventative treatments and increase in drug resistant strains

- Identify potential drug targets *in silico*



Metabolic network analysis

Enzyme datasets

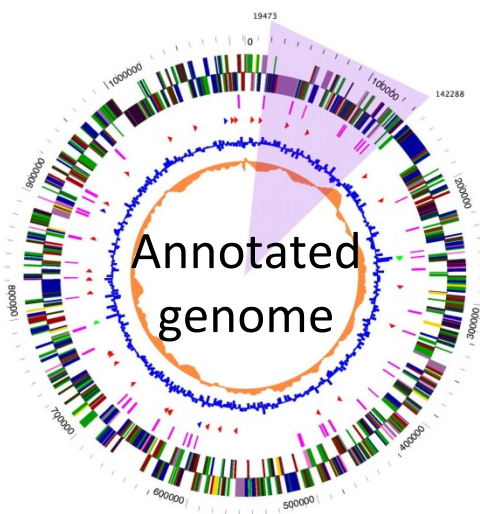


Experiments

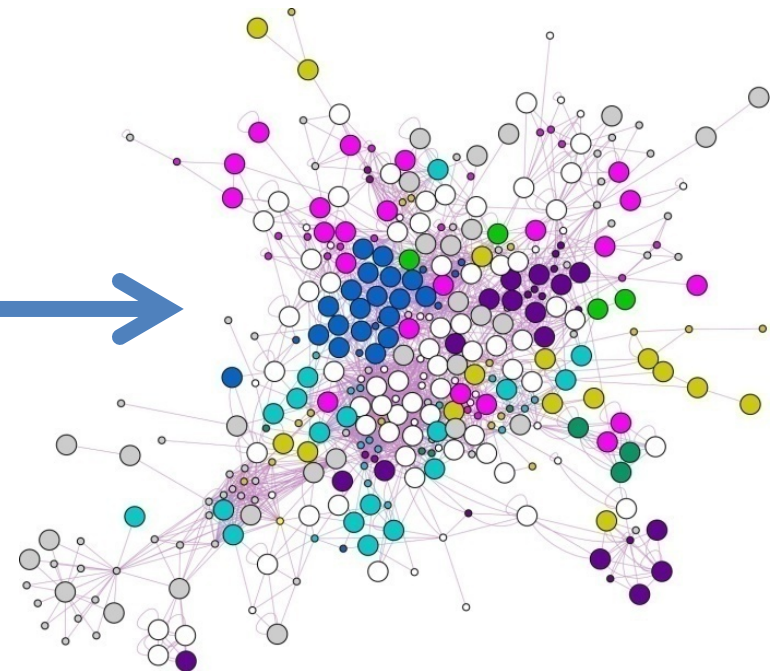


PubMed

Literature



Metabolic network
(*Plasmodium falciparum*)



Sequence diversity across EC categories

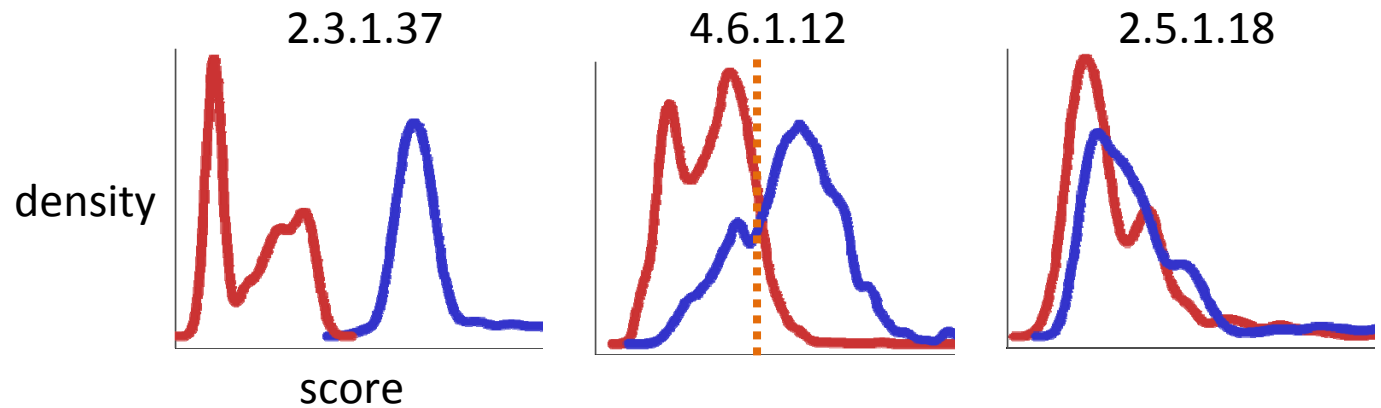
Globally align 117,000 enzymatic proteins



Categorize hits for 2,400 EC categories



Plot scores for hits in
(1) same EC and
(2) different EC



Probabilistic density function for each EC:

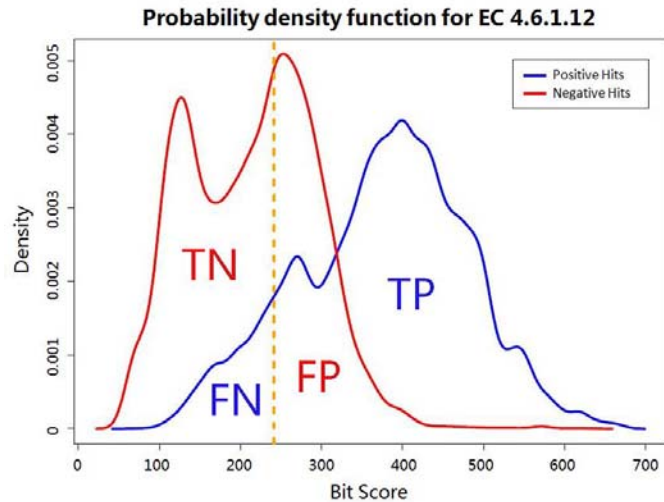
A probabilistic method for classifying unannotated enzymes

Current enzyme sources are missing information on

1. Sequence diversity within a functional class
2. How reactions were assigned
3. Degree of confidence for each assignment

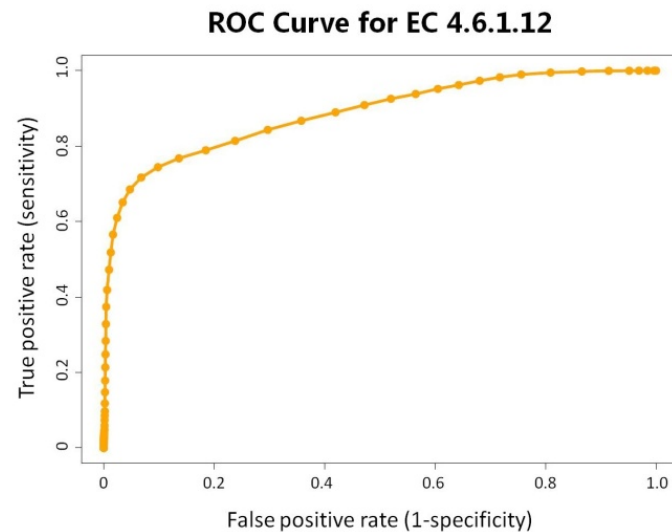
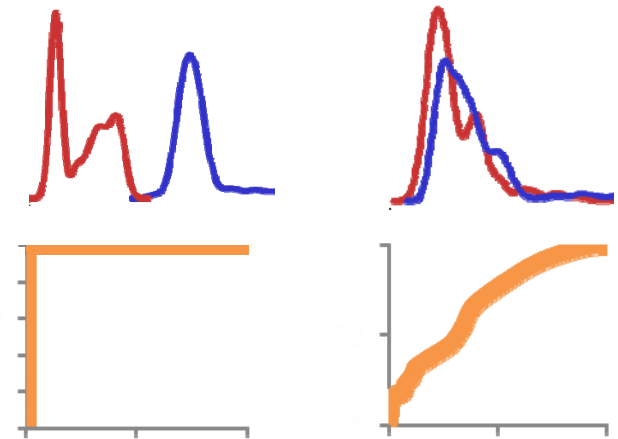
Insight into large-scale sequence evolution

Performance assessment



$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

$$\text{Specificity} = \frac{TN}{TN+FP}$$



- 5-fold cross-validation:
 - ▣ EC numbers with 30+ proteins
 - ▣ Eliminate multifunctional enzymes
- Nearly 100% accuracy
- *P. falciparum* gold standards

Probabilistic enzyme prediction

Calculate the probability that an unknown protein belongs to a particular enzyme category

$P(\text{unknown enzyme belongs to EC with score } x)$

$$= \frac{P(\text{EC}) * P(x | \text{EC})}{P(\text{EC}) * P(x | \text{EC}) + P(\text{EC}') * P(x | \text{EC}')}$$



$P(\text{unknown enzyme belongs to EC})$

$$= -2 \sum_{i=1}^n \ln(p_i)$$

Rank	EC	Confidence score
1	2.7.4.3	0.95
2	2.7.4.9	0.05
3	2.7.4.8	0

