

Database Structure and Visualization Software for Microbial Physiology Data

Nicholas Turner, Haquiu Liu, Jeremy Frank, Rob Edwards

INTRODUCTION

Viruses in the environment play a large role in community nutrient cycles through complex interactions with their hosts, by causing host death or modifying host metabolism. These interactions, though, are relatively poorly understood because on average, only around 10-30% of their proteins have similarity to any other known protein. To address this, we are functionally characterizing these proteins through how they affect host physiology using two different tools: metabolomics and phenotype microarrays, each of which address a different aspect of metabolism. Metabolomics give a picture of how the expressed protein affects the cell's overall metabolism while phenotype microarrays provide insight into how the protein affects the cell's ability to utilize specific compounds.

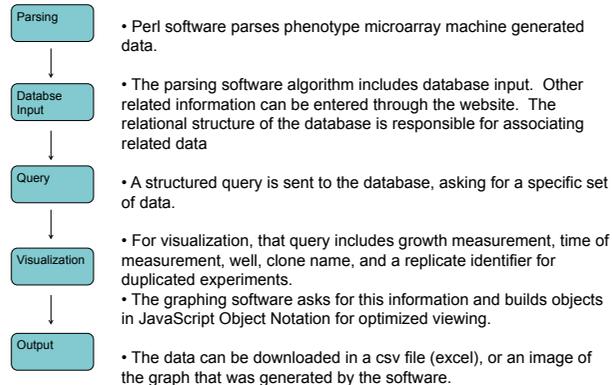
These datasets are information rich and by nature, extremely large. In order to parse, order, compare and visualize the data, computational methods in the form of parsing software, database input/output and visualization software were used. The microarray data was output from the plate reader and uploaded to the web-based application that first parses the data and assigns relevant information to each piece such as time of measurement. The data is then funneled into a database rich with extraneous information.

Most importantly, the database maps the important relationships between the data and associates relevant information that can be used in various methods of visualization. This pipeline forms the basis upon which subsequent metabolic modeling software can be used to predict protein function.

PROBLEM

In order to compare and visualize growth data from a large number of *Escherichia coli* clones quickly, a system specific to the data was needed. The growth data associated with each clone 96 wells by about 48 time points (or 4608 points per replicate). When it is multiplied by hundreds of clones, with each clone experiment being replicated numerous times, the problem set quickly becomes quite large. The meta-data associated with each piece of growth data must also be accounted for, such as growth medium composition and concentration; growth supplement composition and concentration; and bacteria ID, expressed protein, and expression vector information.

METHODS



STORAGE AND VISUALIZATION OF DATA

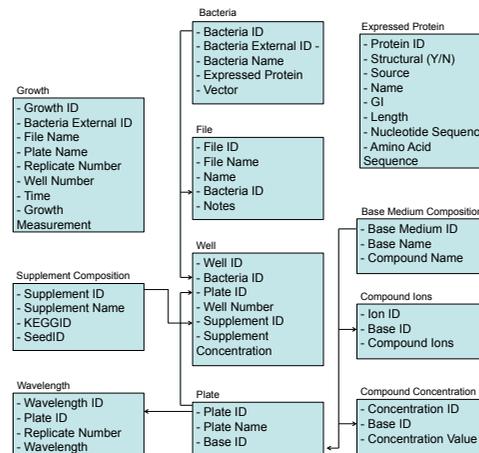
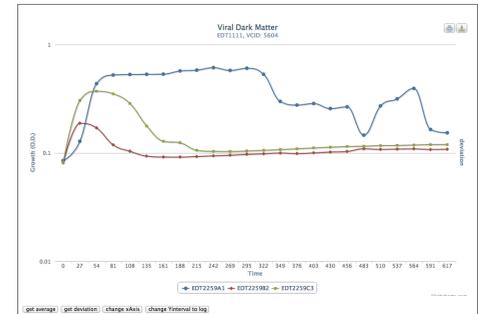
Storage:

A MySQL database stores the data in 11 tables: bacteria, compound concentration, compound ions, file, growth, base medium composition, supplement medium composition, plate, expressed protein, wavelength, well.

Visualization:

The data can be viewed online from any computer or mobile device. It includes functions for average, deviation, and viewing the axis in log base 2.

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
B	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12
C	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
D	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
E	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12
F	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
G	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12
H	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12



Representation of database with references

DIRECTION

The next phase of the project will include the metabolomics data in the database. We will also give the visualization software more power, including the ability to cross reference external databases. Metabolic modeling software is also on the horizon to predict the protein function.