New York University
A private university in the public service

Dennis Shasha
Courant Institute of Mathematical Sciences
Department of Computer Science
New York University
251 Mercer Street
New York, N.Y. 10012, USA
Telephone: (212) 998-3086
Electronic mail: shasha@cs.nyu.edu
Web: http://cs.nyu.edu/cs/faculty/shasha/index.html

July 13, 2007

Dr. Gloria Coruzzi
Dept of Biology
New York University

Dear Gloria,

I am most pleased to continue our fruitful collaboration, and to apply the systems biology tools we have developed and will develop to model the regulation of asparagine metabolism and nitrogen efficiency in your DOE-funded project.

Whereas many collaborations between biologists and computer scientists involve data mining and machine learning, the mining usually occurs after all the experiments have been completed. By contrast, our collaboration has used an "activist" data mining paradigm called adaptive combinatorial design that involves computation starting with the design of experiments. As we have realized it, adaptive combinatorial design starts when the combinatorial design algorithm suggests a few well-separated experiments over a variety of values of input variables (technically, covering arrays of strength two). The results from the experiment are then used to determine input variables on which to focus, which in turn suggests new experiments that again use combinatorial design but this time with constraints. After a few iterations, our approach permits the isolation of input variables that cause an outcome of interest as well as the identification of genes that might mediate those input variables.

In parallel, my group has collaborated with your post-doctoral fellows Rodrigo Gutierrez and Manpreet Katari to build the VirtualPlant system, incorporating a multinetwork of interconnections among genes (e.g., metabolic connections, protein-protein interactions, transcriptional regulation) in order to determine connections among genes identified by a series of experiments. Included in Virtual Plant are a set of tools that permit bench biologists to visualize the functionality (e.g. in the GO ontology) of identified gene groups across experiments as well as cell-specific interconnections with very minimal (under five minutes) training.

Finally, we have worked together on machine learning tools to help infer the functionality of genes of unknown function based on their connections to genes of known function.

As has happened in the past, we will find many other ways – some unanticipated – in which computing can contribute to experimental design, analysis, and model building.

I look forward to continuing our work together.

Best Regards,

Dennis Shasha