



To: NSF ABI Program

By signing below I acknowledge that I am listed as a collaborator on this ABI proposal, entitled "ABI Innovation: Cross Species Network Inference", with Dr. Dennis Shasha as the Principal Investigator. I agree to undertake the tasks assigned to me, as described in the proposal, and I commit to provide or make available the resources therein designated to me.

Signed: 

Print Name: Stephen A. Goff

Date: August 10<sup>th</sup>, 2010

Institution: BIOS Institute  
University of Arizona

## PROJECT SUMMARY “ABI Innovation: Cross Species Network Inference”

### 1. Senior personnel

PI: Dennis Shasha (NYU Courant Institute of Mathematical Sciences)

Co-PIs: Gloria Coruzzi & Manpreet Katari (NYU Biology, Center for Genomics & Systems Biology)

Senior Personnel: Arthur Goldberg (NYU Courant Institute of Mathematical Sciences)

Collaborators: Rodrigo Gutierrez (Catolica Universita de Chile)

2. Intellectual merit of the proposed activity We propose to develop a Cross Species Network Inference (CSNI) platform that will enable plant biologists to predict how an interacting network of genes/products in new genomes will react *as a system* in response to genetic modifications. To implement this ambitious goal, CSNI will integrate genome-scale data (e.g., RNA-Seq data) acquired in a Target Genome into *inferred* gene networks with the aid of experimentally validated data (e.g. metabolic, protein interaction, etc.) from Reference Genomes. The result will generate a set of testable hypotheses about gene networks in a target genome, as well as suggestions for future experiments, especially time series experiments. This project will leverage the facilities of the current VirtualPlant software platform ([www.virtualplant.org](http://www.virtualplant.org)) developed under an NSF Arabidopsis 2010 Grant (DBI-0445666) including Arabidopsis multinet network data, analysis, and manipulation tools {Katari, 2010 #1}. As output, we will provide a pipeline of tools for Cross Species Network Inference to the community via a website ([www.CrossSpecies.org](http://www.CrossSpecies.org)) and the NSF *iPlant* Project (see letter). In addition, our CSNI framework will build on the infrastructure of a generic bioinformatic analysis platform engine such as Taverna {Oinn, 2004 #2}, Kepler {Altintas, 2004 #3}, or Galaxy {Blankenberg, #4}. As a proof-of-principle, we will apply this Cross-Species Network Inference framework to predict metabolic and kinase interaction edges. We will then extend CSNI to other genomes, for which experimental data supporting network edges is not yet available. While CSNI is here described with respect to plants, the framework and basic algorithms extend to any under-analyzed species. This work will achieve one of the main goals of Systems Biology – predicting network states under untested conditions.

We divide the work into three aims:

Aim 1. Inference and validation of an interaction network in Rice as a proof-of-principle.

Using a known validated network in Arabidopsis (e.g. metabolic, protein:protein, and other validated interactions), homology between Arabidopsis and Rice genes, and transcriptome correlation data in Rice, we will create *inferred* interaction networks in Rice. To validate and refine the approach, these *inferred* Rice networks will be compared with experimentally validated networks from Rice. This can extend to other species.

Aim 2. Inference of Regulatory Networks Using Time Series As time series constitute a particularly informative method for inferring networks, we show the power of a model resulting from a closely spaced time series of an N-treatment experiment in Arabidopsis *even in the absence of other data*. We describe our “State-Space Analysis” machine learning method and validate the resulting network based on its predictive accuracy on out-of-sample data.

Aim 3. Develop A Bioinformatic Pipeline for Cross-Species Inference (CSNI). Provide a biologist-friendly CSNI software platform ([www.CrossSpecies.org](http://www.CrossSpecies.org)) that will infer networks in a target species, given experimentally validated networks in a reference species, homology information and experimental data in the target species.

3. Broader impacts of the proposed research This project is the result of a long-standing and highly successful collaboration between biologists at NYU and elsewhere, and computer scientists at NYU's Courant Institute of Mathematical Sciences. The systems biology tools and pipelines resulting from this project will empower biologists to use genomic data to predict a spectrum of gene networks in biology with broad applications to agriculture and the environment. In addition to scientific results, this collaboration extends to joint training of post-docs and graduate students in Systems Biology.