



Proposal Status | MAIN ▶

Organization: New York University

Panel Summary #1

Proposal Number: 1025989**Panel Summary:**
Panel Summary

Intellectual Merit

Strengths:

The effort to make network inference applicable across plant species is important and timely. The integration with expression analysis of nitrogen response in rice and Medicago and the collaboration with Doug Cook was very well received. There was no doubt the proposed methods would be effective.

Weaknesses:

It remained unclear how the system would perform for less conserved processes. Is the Arabidopsis network sufficiently verified, and does it include the best possible function annotation and expression data? In rice, the expression data set used was found to be too limited.

Broader Impacts

Strengths:

There is excellent potential for tools from this project to be widely applied. The interdisciplinary training was rated as excellent. If possible, the R course should be put online for wider use.

Weaknesses:

There were concerns that the approach and network models would not work as well for questions other than nitrogen response and for the time series data presented in the proposal.

Summary

This was seen as a strong proposal from an excellent interdisciplinary team of researchers. Scientific approach and prospective outcomes are excellent. The proposal would have been even stronger if broader impact in other areas of plant research had been demonstrated. It is not obvious that the Arabidopsis network from this group is a gold standard or reference in the field.

This summary was read by the assigned panelists and they concurred that the summary accurately reflects the panel discussion.

Panel Recommendation: Highly Meritorious[◀ Back to Proposal Status Detail](#)Download [Adobe Acrobat Reader](#) for viewing PDF files

National Science Foundation
4201 Wilson Boulevard, Arlington, Virginia 22230, USA
Tel: 703-292-5111, FIRS: 800-877-8339 | TDD: 703-292-5090

[Privacy and Security](#)



Proposal Status | MAIN ▶

Organization: New York University

Review #1

Proposal Number: 1025989
Performing Organization: New York University
NSF Program: Plant Genome Research Project
Principal Investigator: Shasha, Dennis E
Proposal Title: TRMS:Cross Species Network Inference: From Models to Crops
Rating: Excellent

REVIEW:

What is the intellectual merit of the proposed activity?

This project proposes to leverage the VirtualPlant platform to create a pipeline of tools for cross-species network inference in plants. This is a highly relevant effort that will benefit many ongoing hypothesis-driven projects that lack the tools or capabilities to include network analysis.

Strengths:

This team is capable of applying network analysis to plant species.

Integration of a project to address nitrogen use efficiency in medicago and rice as proof of concept and for stimulative feedback into tool development. The PIs have excellent results in a previous analysis in arabidopsis that I consider cutting-edge. They discuss their data in several dimensions, and e.g. their comments that more timepoints may be more valuable than biological replicates have important impact on many experiments. Mostly, they present their science very well.

The analysis of medicago nitrogen response tackles interesting and novel questions, and relations to existing results are discussed.

Weaknesses:

MetaCyc is not mentioned, instead metabolic pathways for medicago will be assembled by this project.

What are the broader impacts of the proposed activity?

Strengths:

True interdisciplinary integration of biology and computer science with research and tools that can broadly impact plant genome research. Tool development efforts are well integrated in cyberinfrastructure, including iPlant, Galaxy.

Cross training of biologists and computer scientists is well established. There are very good training components for emerging scientists, but no outreach towards enthusing younger students for computational/biological science or specific efforts towards underrepresented.

Summary Statement

Strong proposal with an excellent team and important collaborations identified. The large effort in implementation is well justified as this will be a major resource and wide usability will depend on stability, power and ease of use. I think there will be a lot of "bang for the buck" including novel scientific insights.

[◀ Back to Proposal Status Detail](#)

Download [Adobe Acrobat Reader](#) for viewing PDF files

National Science Foundation
4201 Wilson Boulevard, Arlington, Virginia 22230, USA
Tel: 703-292-5111, FIRS: 800-877-8339 | TDD: 703-292-5090

[Privacy and Security](#)



Proposal Status | MAIN ▶

Organization: New York University

Review #2

Proposal Number: 1025989
Performing Organization: New York University
NSF Program: Plant Genome Research Project
Principal Investigator: Shasha, Dennis E
Proposal Title: TRMS:Cross Species Network Inference: From Models to Crops
Rating: Good

REVIEW:

What is the intellectual merit of the proposed activity?

Shasha et al. propose inferring a network in rice based on an existing network they have developed for Arabidopsis, and validating the quality of their predictions based on gold standard or "ground truth" networks from the literature for rice. They propose to run some experiments on nitrogen use/signaling in rice to identify early stages of regulatory networks, ditto for Medicago (under naive and symbiont conditions), and finally to develop a cross-species network inference engine (CSNI) to leverage data from gold standard data sets, supplemented with user-supplied transcriptome or other data from a target species.

The proposal would strengthen existing collaborations between biologists and computer scientists at NYU.

Some questions/concerns re. the grant application:

VirtualPlant was a good start towards integrating data from multiple sources towards gene function prediction, and has been used in a handful of instances to identify novel roles for known regulatory proteins. It would be nice to see some effort made to weight the edges according to degree of support. For instance, in the Geisler-Lee interolog data, edges could be scored according to their CV value. Other gene function prediction approaches weight edges according to the predictive power of the given omic data set (e.g. GeneMANIA, Mostafavi et al., 2008). In addition, the predictive ability of a given network is only as good as the data that are used to generate it. Maybe this is part of the applicants other VirtualPlant NIH/NSF grant-funded efforts, but it would have been nice to see some emphasis on curation and addition of other data sets.

On page 5/6, what rice microarray data were used or are to be used? The data sets used can dramatically affect the coexpression scores (see Usadel et al., 2009, PC&E). I would argue or push for the development of multiple networks, based on different biological conditions. For while some interactions can be classified as "immutable", that is either a protein interacts or it does not with another protein, others change depending on what data one is using. I don't think it is an option to just use all possible data for coexpression either, as certain tissue-specific responses may be buried by larger responses in other tissues.

P7. Have the authors considered using Arabidopsis Reactome instead of KEGG?

There is no mention of GeneCAT (Mutwil et al., 2008) that also attempts to identify cross-species networks (in that case, for Arabidopsis and barley).

It may not be possible to identify the "best" network overall.

P8. Why not use InParanoid or OrthoMCL in the first place?

P10, in practice, $p < 0.01$ is easily met for Pearson CC.

P11. I like the proposed rice and Medicago nitrogen TF KO experiments

P11. mRNA-seq data may be more difficult to analyze than the authors suspect - many alternative splicing events (maize mRNA-seq data from Thomas Brutnell, pers. comm.)

P12. How well does rice-arabidopsis work? There are examples in the literature of expression of given genes being up-regulated in Arabidopsis in response to a given hormone, while the ortholog in rice is down-regulated.

What are the broader impacts of the proposed activity?

This is fine, but it is a bit surprising that a collaborative program between the Courant and the biology dept. at NYU has not yet been set up. The CSNI tool would be likely be used by the wider plant biology community.

Summary Statement

In general, I feel the authors' proposal has some intellectual merit (e.g. space-state modeling, CSNI, nitrogen profiling experiments in Medicago and rice) but I would have liked to have seen some more preliminary data. Benchmarking with other function prediction methods would help to convince that their approach is appropriate - again, I raise the concern of using all microarray data to build the coexpression part of the multinetwork.

[◀ Back to Proposal Status Detail](#)

Download [Adobe Acrobat Reader](#) for viewing PDF files



Proposal Status | MAIN ▶

Organization: New York University

Review #3

Proposal Number: 1025989
Performing Organization: New York University
NSF Program: Plant Genome Research Project
Principal Investigator: Shasha, Dennis E
Proposal Title: TRMS:Cross Species Network Inference: From Models to Crops
Rating: Very Good

REVIEW:

What is the intellectual merit of the proposed activity?

This team of investigators (except for co-PI Cook) has previously led efforts to develop the appropriate linkages between data generation and computational tools that enable systems biology approaches for plants. Understandably because of the extensive resources available, the tools and platforms have been developed for Arabidopsis. The NYU group (Gutierrez was in PI Coruzzi's lab) has led the way in applying systems biology tools to plant research, now accessible through the VirtualPlant software. With the emerging genome sequences and functional genomics datasets now available for other plant species, the time has now come to apply the gene network construction and analysis functions within VirtualPlant to crop plants. This proposal will develop a Cross Species Network Inference platform for rice and Medicago, using the primary N assimilation pathway for proof-of-principle.

The proposal has four aims. The first is quite feasible, to infer and validate the interaction network in rice via methods developed for Arabidopsis, using sequence homology between rice and Arabidopsis genes and available expression profiling data from rice. One innovation here is to include protein-DNA and miRNA-mRNA interactions to enrich a "multinetwork". The second goal is create a state-space regulatory network for rice, by first obtaining time-series RNA expression profiles from rice shoots and roots following nitrate treatment, and then applying the network construction and "State-Space" analysis approaches. The validation here is computational, and in my view circular, and I would prefer to have seen validation of network predictions with some wet-lab experiments. However, reporting the network would enable others to do so. The next aim is to then develop a similar cross-species network inference for primary N assimilation in Medicago, with the added wrinkle of constructing networks under naive (no symbiotic Rhizobia) or N-fixing (with Rhizobia) conditions. Co-PI Cook will assist with the Medicago experiments, and it is likely this effort will yield some interesting hypotheses to test with respect to transcription factors that might be important regulators of the primary N-assimilation pathway and its feedback interactions with excessive N level and Rhizobia.

The major questions here are can this approach be extended to pathways beyond something like primary N assimilation that is highly conserved from microbes through plants, and will the Cross Species Network/VirtualPlant platforms be truly useful to biologists beyond this group that is focused on Arabidopsis N metabolism?

What are the broader impacts of the proposed activity?

Good cross-training in systems biology bridging the generation of functional genomics datasets through mathematical modelers, computational biologists and software developers. The R workshop seems valuable, and it would be nice to consider leveraging the computational expertise in the project to deliver these kind of activities on-line. It was good to see the postdoc as educators concept be included in the project description, instead of only in the supplementary documents.

Summary Statement

This is a leading group in development of platforms for systems biology network predictions in Arabidopsis, who now wish to explore the utility of their effort to crop plant species. It is evident the proposed activities are feasible, the question is whether they can be extended to a broader set of plant metabolism and plant species.

[◀ Back to Proposal Status Detail](#)

Download [Adobe Acrobat Reader](#) for viewing PDF files

National Science Foundation
4201 Wilson Boulevard, Arlington, Virginia 22230, USA
Tel: 703-292-5111, FIRS: 800-877-8339 | TDD: 703-292-5090

[Privacy and Security](#)



Proposal Status | MAIN ▶

Organization: New York University

Review #4

Proposal Number: 1025989
Performing Organization: New York University
NSF Program: Plant Genome Research Project
Principal Investigator: Shasha, Dennis E
Proposal Title: TRMS:Cross Species Network Inference: From Models to Crops
Rating: Excellent

REVIEW:

What is the intellectual merit of the proposed activity?

This proposal builds on an already established and productive collaboration between biologists and computer scientists at NYU. There is no doubt that this group is qualified to do the proposed research. The group plans to leverage data and programs established to look at nitrogen assimilation in the model plant Arabidopsis and use that information to infer networks in crop species. First, as proof of principle, they will predict a network for rice, a crop in which there is ample experimental data allowing for testing of their predicted network. Not only will this work allow the group to test their model building, but it will also generate testable hypotheses for further work in rice. Once they have done this, they will move into the nitrogen-fixing crop Medicago. During this, they will also be generating software tools and a pipeline for this type of work in a variety of crops and on a variety of questions, not just nitrogen assimilation. This work will produce many novel insights on nitrogen use in plants and will also develop a tool that will advance research in many areas of plant biology.

What are the broader impacts of the proposed activity?

The broader impacts of this proposal are strong. A resource will be created for the entire scientific community, i.e. the Cross Species Network Inference Pipeline, which will be freely available on the web. The close collaboration between biologists and computer scientists in the project, and the training of personnel in this environment, will create well-trained scientists ready to work in systems biology. A well-articulated plan for cross training of biologists and computer scientists is in place. A weakness of the broader impacts of the proposal is that it is not addressed how to recruit underrepresented minorities into the research, although female personnel are well represented.

Summary Statement

This is an exceptionally strong proposal that truly integrates computer scientists and biologists to forward a systems biology approach in plants. The training and broader impacts are also good.

[◀ Back to Proposal Status Detail](#)

Download [Adobe Acrobat Reader](#) for viewing PDF files

National Science Foundation
4201 Wilson Boulevard, Arlington, Virginia 22230, USA
Tel: 703-292-5111, FIRS: 800-877-8339 | TDD: 703-292-5090

[Privacy and Security](#)



Proposal Status | MAIN ▶

Organization: New York University

Review #5

Proposal Number: 1025989
Performing Organization: New York University
NSF Program: Plant Genome Research Project
Principal Investigator: Shasha, Dennis E
Proposal Title: TRMS:Cross Species Network Inference: From Models to Crops
Rating: Very Good

REVIEW:

What is the intellectual merit of the proposed activity?

> How important is the proposed activity to advancing knowledge and understanding within its own field or across different fields?

The proposed research is important to plant biology and bioinformatics. The proposed science is of high quality and internationally competitive. The application area is of the highest importance.

> How well qualified is the proposer (individual or team) to conduct the project? (If appropriate, the reviewer will comment on the quality of prior work.)

The team is a strong one. It involves collaboration between biologists at NYU Center for Genomics and Systems Biology, and computer scientists at NYU's Courant Institute of Mathematical Sciences. Both the wet and dry labs are well equipped. The team has a strong publication record, and I was particularly impressed by: Gutierrez R, Stokes T, Thum K, Xu X, Obertello M, Katari M, Tanurdzic M, Dean A, Nero D, McClung R and Coruzzi G (2008) 'Systems approach identifies an organic nitrogen-responsive gene network that is regulated by the master clock control gene CCA1' Proc. Natl Acad Sci USA 105, 4939-4944. I was surprised and confused that it was not mentioned in G. Coruzzi's CV!

> To what extent does the proposed activity suggest and explore creative, original, or potentially transformative concepts?

The proposed bioinformatic and systems biology methodologies are sensible and generally state-of-the art. Plants are of course particularly complex, with more genes than animals. I would have appreciated more comment on this.

I like the proposed work on 'state space' model learning. In my opinion too much focus is currently placed in systems biology of using ODE's even when there is insufficient data to justify such models. The proposed use of workflows is a positive.

One areas where where I think the proposed research is slightly behind the curve is that there is no suggestion of using semantic web technology ũ apart from implicitly with work flows. Much of bioinformatics is moving over to use RDF databases, and ontologies such as OBI to record experiments.

There was nothing in the application that surprised me and made me wish I had thought of it. No 'wow factor'.

> How well conceived and organized is the proposed activity? Is there sufficient access to resources?

The proposed researcher is well organized and planned. I think the timeline for deploying the services to the general community (4-5 years) is not very ambitious. I think they should be used much more quickly. It was not clear to me whether if I was a commercial company and didn't want to use a public service whether I could download the required tools and data.

What are the broader impacts of the proposed activity?

> How well does the activity advance discovery and understanding while promoting teaching, training, and learning?

The proposal would promote the training of more cross-disciplinary researchers in plant biology and computer science. Such researchers are essential for the future of plant biology.

> How well does the proposed activity broaden the participation of underrepresented groups (e.g., gender, ethnicity, disability, geography, etc.)?

It doesn't.

> To what extent will it enhance the infrastructure for research and education, such as facilities, instrumentation, networks, and partnerships?

The proposal does not propose to enhance infrastructure.

> Will the results be disseminated broadly to enhance scientific and technological understanding?

See above. I am concerned about the proposed slow pace of dissemination.

> What may be the benefits of the proposed activity to society?

The application area of 'food security' is of the highest importance. The proof-of-principle work concerning nitrogen use chosen for the cross species network inference studies is of clear scientific and agronomic interest.

Summary Statement

The proposed science is of high quality and internationally competitive. The application area is of the highest importance.

I do however have two concerns: The first is that the proposed work seems to overlap with two existing grants:

- . A Systems approach to regulatory networks controlling N-assimilation (NIH)
- . Nitrogen Networks in Plants (NSF)

It was not clear to me how these existing grants relate to the proposal, their titles seem very similar. If there is overlap then I would reduce my rating.

The second concern is the costing of the proposal. It seemed to me that the project was a bit over-staffed relative to the amount of work planned. I was particularly unconvinced by the need for the 'senior personnel'.

[◀ Back to Proposal Status Detail](#)

Download [Adobe Acrobat Reader](#) for viewing PDF files

National Science Foundation
4201 Wilson Boulevard, Arlington, Virginia 22230, USA
Tel: 703-292-5111, FIRS: 800-877-8339 | TDD: 703-292-5090

[Privacy and Security](#)



Proposal Status | MAIN ▶

Organization: New York University

Review #6

Proposal Number: 1025989
Performing Organization: New York University
NSF Program: Plant Genome Research Project
Principal Investigator: Shasha, Dennis E
Proposal Title: TRMS:Cross Species Network Inference: From Models to Crops
Rating: Good

REVIEW:

What is the intellectual merit of the proposed activity?

Shasha et al. propose to develop, validate and deploy an analysis pipeline for cross-species comparative inference of gene function and interaction networks based on similarities in nucleotide sequence of the protein coding and regulatory regions and transcription patterns. Such a tool is sorely needed with the growing number of genome and transcriptome sequences coming available for emerging model and non-model plant species. The straightforward algorithms described in this proposal seem quite reasonable and likely to produce useful inferences for gene function non-model species. Given the complicated history of gene and genome duplication since the divergence of Arabidopsis and rice, however, the proposed experimental design seems naïve. Nonetheless, the PIs have an impressive record of success in designing tools for comparative inference and using them to elucidate gene function and the structure of interaction networks. Further, the results of pilot experiments describe in the proposal are surprisingly encouraging.

There is no question that the plant science community needs a framework for translational research to leverage knowledge from existing experimental models (most notably but not exclusively Arabidopsis) to emerging and non-model plant species. The PIs propose to launch such a framework (Aim 4) after developing and validating a prototype used to infer rice gene function through analysis of gene expression covariances and homology to Arabidopsis genes with known function (Aim 1). Initial analyses have been based on reciprocally best BLAST hits performed surprisingly well (Table II) given that there have been a minimum of three whole genome duplications in the lineage leading to Arabidopsis and two in the lineage leading to rice since they diverged at least 150 million years ago. The PIs do not claim they will use reciprocal best hits as their sole methods for making predictions based on cross-species comparisons, but the proposal could have gone further and acknowledging the potentially confounding effects of gene and genome duplications. Choosing a more closely related target (e.g. Brassica) or reference (e.g. maize) taxon would make homology-based inferences less problematic.

The complicated relationships between most Arabidopsis and rice genes may be especially confusing when attempting to predict regulatory networks in rice based on expression covariance and the structure of Arabidopsis interaction networks (Aim 2). The closer relationship between Arabidopsis and Medicago may make comparative analysis of genes and gene networks involved in nitrogen metabolism in (Aim 3) less challenging, but gene and genome duplications (two on lineage leading to Arabidopsis and one on the lineage leading to Medicago) may still obscure simple comparisons of gene expression and homology. Of course the absence of nitrogen fixation in Arabidopsis will also limit its utility as a model of nitrogen metabolism in legumes including Medicago.

What are the broader impacts of the proposed activity?

The PIs have an excellent record of productivity in terms of publications, training and the development of cyberinfrastructure for comparative analysis of fundamental processes in plant growth and development. If the objectives of this proposal were fully realized, positive impacts for crop improvement and human health could be expansive.

Summary Statement

In summary, the ultimate objective of the proposed research to develop a user-friendly platform for Cross-Species Network Inference is extremely important. Advancement of this objective could evolve synergistically with the work of the iPlant genotype-to-phenotype (iPG2P) and plant tree of life (iPToL) grand challenge initiatives. As such, the proposed development of a web-based Cross-Species Network Inference database and analysis tool would be a major contribution.

[◀ Back to Proposal Status Detail](#)

Download [Adobe Acrobat Reader](#) for viewing PDF files

National Science Foundation
4201 Wilson Boulevard, Arlington, Virginia 22230, USA
Tel: 703-292-5111, FIRS: 800-877-8339 | TDD: 703-292-5090

[Privacy and Security](#)