**Title**: NutriNet: A network inspired approach to improving Nutrient Use Efficiency (NUE) in crop plants

**Personnel**: Gloria Coruzzi (PI), New York University, Center for Genomics and Systems Biology

Dennis Shasha (co-PI) New York University, Courant Institute of Mathematical Sciences

Stephen Moose (co-PI) University of Illinois at Urbana-Champaign, Institute for Genomic Biology

Collaborators: Sandrine Ruffel (INRA, France), Gabriel Krouk (INRA, France), Richard McCombie (CSHL).

Senior Personnel: Kranthi Varala, Ying Li, Daniel Tranchina

**Intellectual Merit**: Recent advances in genome sequencing, functional genomics, and computational tools now enable a systems-level understanding of key physiological and developmental processes in model plant species such as Arabidopsis. However, translating this knowledge to agriculturally important phenotypes has been hindered by a gene-centric focus, and a limited capacity to empirically derive gene regulatory networks at a population scale in germplasm relevant to future crop improvement. The goal of our NutriNet project is to develop network-connected modules that are predictive of phenotypic variation and enhance the efficiency of genetic gain in crop species, using nutrient use efficiency (NUE) of cereals such as maize (Moose) and wheat (with our ERA-CAPS partners) as the target trait. The advantages of the network-oriented approach include: i) the ability to exploit detailed datasets for gene and protein interactions in Arabidopsis to inform analysis of data-poor crop species, ii) better definition of future experiments in model species that facilitate translational research, and iii) identification of robust network modules that can be applied in molecular breeding programs. ***Four aims are proposed to achieve these goals***:

**Aim 1. Exploiting Genetic Diversity in NUE (Coruzzi/Moose)**. New and existing transcriptome data for nitrogen-responsive genes and NUE [Gloria: you say nutrient use efficiency above, but now the genes are just for nitrogen] phenotypes in Arabidopsis accessions (Coruzzi, NYU) and Maize lines (Moose, Illinois) will be assembled across an initial training set of 5-10 genotypes representative of a broad phenotypic variation in NUE.

**Aim 2. Role of root-shoot N-signaling and root foraging on NUE (Coruzzi)**. A critical knowledge gap for NUE is the effect of heterogenous nutrient environments on NUE. Using a split-root experimental design, we will quantify the effect of heterogeneous environments on root nutrient foraging and 15N-assimilation among Arabidopsis, maize and wheat (EU funded) genotypes that vary for NUE. These data will guide transcriptomic (mRNA and small RNAs) analysis of regulatory networks that underlie local and systemic nitrogen signalling and NUE. [Gloria: I can see a reviewer writing. “Aim 2 has some nice preliminary results, but crop plants in practice will not be grown in a split root, so how will this help them?” One idea might be to say that if the soil is unevenly rich, then we know the genes that will help plants forage more efficiently, enabling them to use poor soils better or to use less fertilizer. You hint at this in the outcomes, but it needs to be more explicit.]

**Aim 3. Constructing NUE Networks (Shasha/Coruzzi).** Data from Aims 1 and 2 will be analysed in a network context, to identify hubs and modules associated with NUE. Network modules identified in Aim 3, will be validated experimentally and in field trials in Aim 4.

**3A**. **Crop-specific and Cross-species Networks**: We will use VirtualPlant Maize to enable the creation of crop-specific gene interaction networks. Users can also readily leverage additional interaction data (e.g. protein:protein, protein:DNA) available from Arabidopsis “network knowledge”.

**3B.** **The Nutri-Net Pipeline: Identifying NUE-predictive network modules**. We will develop the NutriNet pipeline to identify network modules predictive of NUE traits. Gene-to-NUE phenotype, gene-to-gene, and protein-to-protein interactions will be sequentially added into the analysis pipeline. The resulting NutriNet network modules will be tested for their ability to predict NUE outcomes on “left-out data using a “training-based” approach. NutriNet modules validated *in silico*, will be advanced to Aim 4 for lab and field tests.

**Aim 4. Nutri-Net Module Validation in Laboratory and Field Conditions**. **(Coruzzi/Moose)** The NutriNet modules developed from the analysis of aim 3 will be tested for prediction of improved NUE using species-appropriate methods. **Laboratory**: Gene components of NutriNet modules will be functionally tested for their role in NUE using mutant and transgenic Arabidopsis. **Field**: Initial field validation will be performed on diverse maize populations representing both historical and current elite hybrids, and wheat breeding germplasm (EU). Genotypes predicted to possess the optimal configurations of gene expression levels in the NutriNet network modules, will be evaluated in field trials with different regimes of nutrient availability. **Combined**: A comparative analysis of lab-to-field results will identify the “translation” of knowledge from Arabidopsis to crops.

**Broader Impacts**: Improving nutrient use efficiency in cereals is an important goal of plant biology, due its multiple positive impacts on sustainable agricultural production in the context of increasing global demand, climate change, and dwindling sources of inexpensive nutrient fertilizers. Evaluation of historical maize and wheat germplasm both suggest that strong selection for maximal grain yields achieved with excess levels of nutrient supply (e.g. N, P, K and S) may constrain the allelic and regulatory diversity available for future increases in NUE under low-input conditions. The NutriNet approach, aims to enable a more robust and efficient approach to identifying and selecting genotypes with enhanced NUE, by applying at a population scale the network knowledge gained from deep analysis of core genotypes. Iterative refinement of NutriNet will enhance both biological discovery and the translation of findings from model plant species to cereal crops. The NutriNet approach may be extended to other species and traits of agricultural importance. The project also offers a broad education and training platform for students and scientists that integrates computer science, systems biology, plant genomics, and applied genetics. In addition to leveraging the cross-disciplinary expertise with NUE for multiple species and nutrients, the planned collaboration with EU scientists improves trans-Atlantic cooperation in the research and training efforts that amplifies its potential global impact.

**Project Impact**:  Plant improvement for the future requires expertise in the areas of Functional Genomics, Crop Physiology, Molecular Breeding and Systems Biology.  Each of these disciplines are well-developed, but are rarely linked together in a team that is aimed specifically at the onset from moving fundamental discoveries to application in both C3 and C4 species.

**Novel Deliverable**:  This proposal describes a novel Network-inspired approach to Marker Assisted Selection (MAS) what will have specific application to the nutrient use efficiency problem as well as broad applications to plant species.

**Network Module Discovery**:  The new knowledge achieved in this project will consist of gene discovery, elucidation of regulatory circuits, identification of the molecular basis of nutrient physiology of better-adapted germplasm, and as practical deliverable, network-inspired MAS tools.

**Results from Prior NSF Support – PART I: Coruzzi, PI /Shasha, co-PI’s NSF DBI-O445666:** One outcome of the NutriNet proposal will be the creation and used of gene networks to enhance translational gene discovery from models-to-crops. The network approach is most closely related to a completed **NSF Grant DBI-0445666**, “Conceptual Data Integration for the Virtual Plant.” That project led to the creation of VirtualPlant ([www.virtualplant.org](http://www.virtualplant.org)) , a software platform to enable Systems Biology research in plants. This included the construction of the first Arabidopsis multinetwork - a molecular wiring diagram of the plant cell - which currently has 19,690 genes and 182,592 interactions [Katari et al 2010; Gutierrez et al 2007]. This Arabidopsis multi-network integrates plant gene/protein interaction data from a number of sources including biochemical pathways [Mueller et al 2003; Kanehisa et al 2004], protein:DNA interaction data [Davuluri et al 2003; Brady et al 2011], protein:protein interaction data [Cui et al 2008; DeFolter et al 2005; Popescu et al 2007], the Arabidopsis interactome data [Arabidopsis Interactome Consortium 2011], smallRNA data [Gustafson et al 2005; Lu et al 2005; Griffiths-Hones et al 2006] and literature-based interactions [Rzhetsky et al 2004]. In one application, a query against this Arabidopsis multi-network with 834 nitrogen-regulated genes resulted in a sub-network of 369 interacting genes [Gutierrez et al 2008] (Fig. 1C). At the top of the resulting list of network TF “hubs” - with 47 connections to targets in the N-regulatory network- were the transcription factors bZIP1, GLK1 and the central clock control gene CCA1, a Myb family transcription factor (TF) [Gutierrez et al 2008]. This discovery enabled us to derive and validate the novel hypothesis that nitrogen-regulation of CCA1 mRNA expression can phase-shift the circadian clock, which was experimentally validated [Gutierrez et al 2008]. Other examples of hypotheses derived and validated using the VirtualPlant multi-network are reported in [Gifford et al 2008; Vidal et al 2010; and Thum et al 2008]. We have recently extendedVirtualPlant to other data-rich species such as Rice, and to important crop species including maize (see [www.virtualplant.org](http://www.virtualplant.org)), to empower systems analysis in other species.

**PUBLICATIONS (most relevant) (PI Coruzzi and co-PI Shasha):**

Katari MS, …., Shasha D, Coruzzi G, Gutierrez R (2010) “VirtualPlant: A software platform to support Systems Biology research”. ***Plant Physiol***. Feb; 152:500-15

Nero D, Kelfer J, Katari MS, Tranchina D, Coruzzi G (2009) “*In silico* evaluation of predicted regulatory interactions in Arabidopsis thaliana”. ***BMC Bioinformatics***. Dec 21;10(1):435

Poultney C, Gutierrez R, Katari MS, Gifford M, Paley W, Coruzzi G and Shasha D (2007) “Sungear: Interactive visualization, exploration & analysis of genomic datasets”. ***Bioinf***, 23:259-61

Ferro A, Giugno R, Pigola G, Pulvirenti A, Skripin D, Bader G, Shasha D, “NetMatch: a Cytoscapeplugin for searching biological networks” ***Bioinf.***, 2007 23(7):910-912

**Applications of VirtualPlant: Hypothesis Generation and Testing**

Vidal EA, Araus V, Lu C, Parry G, Green PJ, Coruzzi GM, Gutiérrez RA (2010). Nitrate-responsive

miR393/AFB3 regulatory module controls root system architecture. ***PNAS.*** 107(9):4477-82

Krouk G, .. Shasha D, Coruzzi G & Gutierrez R (2009) “Systems approach uncovers restrictions for signal interactions regulating genome-wide responses .” ***PloS Comp Biol***. Mar;5(3):e1000326.

Gutierrez R, ..., Nero D, McClung R and Coruzzi G (2008) "Systems approach identifies a N-responsive gene network regulated by clock control gene CCA1" ***PNAS*** 105, 4939-4944. *(Factor 3).*

Gutierrez R, …Shasha D, Coruzzi G, Crawford N (2007) "Insights into the genomic nitrate response using genetics and the Sungear Software System" ***J Exp Bot*** doi: 10.1093/jxb/erm079

Gutierrez R, …, Shasha D, Coruzzi G (2007) "Qualitative network models & genome-wide expression data define C/N-responsive biomodules " ***Genome Bio***l 8: R7. *Faculty 1000 (Must Read: Fact 6)*

**Computational Publications**

Di Natale R, Ferro A, Giugno R, Mongiovi M, Pulvirenti A and Shasha D (2010) "SING: Subgraph

search In Non-homogeneous Graphs" ***BMC Bioinf***, 11:96doi:10.1186/1471-2105-11-96

Zhang X, D. Shasha, Y. Song and J. T. L. Wang (2010) “Fast Elastic Peak Detection for Mass Spec Data

Mining,” ***IEEE Transactions*** *on Knowledge & Data Engineering*. Issue 99. November 29, 2010,

**Plant Systems Biology: Reviews, Books and Outreach**

Ruffel S, Krouk G, Coruzzi G (2010). "A Systems View of Responses to Nutritional Cues in Arabidopsis: A Paradigm Shift for Predictive Network Modeling”. ***Plant Physiol***. 152;445-52

Coruzzi GM, Burga A, Katari MS, & Gutierrez RA (2009) “Plant Systems Biology”, ***Annual Plant Reviews***; Blackwell, , UK, 2009, Vol. 35. Pgs 3-31*.*

Gifford M, Gutierrez R, and Coruzzi G (2006) "Modeling the Virtual Plant ". Essay 12.2 Chapter 12.; In ***A Companion to Plant Physiology,*** , http://4e.plantphys.net/article.php?ch=12&id=352

**Education & Training**: Undergraduates (UG), master’s (MS) and PhD students have learned Systems Biology. **Undergraduates**: Steve Nowicki (NYU, CAS), Varuni Prabhakar (Barnard College), Rebecca Davidson (BS, Computer Science); **Masters Students (Computer Science)**: Ana F. Arroja , Ranjita Iyer, Jonathan Kelfer, Jesse Lingeman, Lee Parnell, Jarod Wang,; **PhD Students (NYU Courant)**: Chris Poultney, Aris Tsirigos, Saurabh Kumar; Damion Nero (NYU Biology). These students have gone on to PhD programs (Prabhakar and Parnell), post-docs (Poultney and Tsirigos) and to industry (Kelfer, Wang -Medidata Solutions; Damion Nero, Statistician Programmer, FOJP Service Corp). **High School Students**: Angela Fan (Stuyvesant HS) – Siemans Semi-Finalist 2011, Intel Finalist 2012; Jenny Kim (Chapin HS), Jack Stevenson (Stuyvesant HS), Kahmun Lo (Stuyvesant HS).

**Results from Prior NSF Support – PART II (Moose, co-PI, U. Illinois).** Co-PI Moose is PI of completed NSF-PGRP project (NSF No. XXXXXXXX) “Gene Discovery for Maize Responses to Nitrogen”. This six-year project was completed in 2011. Eleven project-supported publications have appeared to date. **[3 – 13]** , one additional manuscript is in review**[14]**, and four are in the final stages of preparation. Major objectives of the project were to integrate metabolite, gene expression and genetic mapping information with agronomic evaluations of nitrogen use efficiency in maize populations that are the focus of community genomics efforts. The RNA-profiling component of the project that used microarray technology was completed and data deposited at the Gene Expression Omnibus; however, we have since repeated many of these experiments with RNA-Seq and the combined results are the focus of the remaining four project publications. In addition, Moose is Co-PI on PGRP project (NSF No. XXXXX), “Exploring the Role of Noncoding RNAs in Heterosis”. This project initiated in 2009 and still active. Two publications have reported on the inheritance patterns of small RNAs in maize and Arabidopsis hybrids. **[16, 17]**. Two additional manuscripts have recently been submitted that describe the evolutionary dynamics of small RNAs in maize **[18]** and the impacts of 24-nt siRNAs on the inheritance of gene expression in maize**[19]**.

**Most Relevant Publications (co-PI Moose):**

Wang, D., Portis, A.R., Moose, S.P. and Long, S.P. (2008) Cool C4 photosynthesis – Pyruvate Pi

Dikinase expression and activity corresponds to the exceptional cold tolerance of carbon assimilation in *Miscanthus x giganteus*. *Plant Physiology* 148: 557-567.

Wang, D., Naidu, S.L., Portis, A.R., Moose, S.P. and Long, S.P. (2008) Can the cold tolerance of C4

photosynthesis in *Miscanthus x giganteus* relative to *Zea mays* be explained by differences in activities and thermal properties of Rubisco? *Journal of Experimental Botany* 59: 1779-1787.

Uribelarrea, M., Moose, S.P., and Below, F.E. (2007) Divergent selection for grain protein affects

nitrogen use efficiency in maize hybrids. *Field Crops Res*. 100: 82-90

Moose, S.P. and Mumm, R.H. (2008) Molecular plant breeding as the foundation for 21st Century Crop Improvement. *Plant Physiology* 147: 969-977.

Moose, S.P. and F. E. Below (2008) Biotechnology approaches to improving maize nitrogen use

efficiency, book chapter in *Molecular Genetic Approaches to Maize Improvement*, A.L. Kriz and B. A. Larkins (eds.) Springer-Verlag.

Lauter, N., Moscou, M.J., Habiger, J., and Moose, S.P. (2008) Quantitative genetic dissection of shoot architecture traits in maize: towards a functional genomics approach. *The Plant Genome* 1: 99-110.

Uribelarrea, M., Crafts-Bradner, S.J. and Below, F.E. (2009) Physiological N response of field-grown maize hybrids (Zea mays L.) with divergent yield potential and grain protein concentration. *Plant and Soil* 316: 151-160.

-------------------------------------------------------------------------------------------------------------------------------

**RESEARCH PLAN**

**Motivation**. The goal of the NutriNet project is to develop *network-connected modules* that are predictive of phenotypic variation in agricultural traits in the field, and enhance the efficiency of genetic gain in crop species. Recent advances in genome sequencing, functional genomics, and computational tools enable a systems-level understanding of key physiological and developmental processes in model plant species such as Arabidopsis. However, translating this knowledge to improve agriculturally important traits has been hindered by a gene-centric focus in crops, and a limited capacity to empirically derive gene regulatory networks at a population scale in germplasm relevant to future crop improvement. To bridge this knowledge-gap, the NutriNet project will use nutrient use efficiency (NUE) as the target trait, to explore a network-centric approach to crop improvement. The promise of this network-oriented approach includes: i) the ability to exploit detailed datasets for gene and protein interactions in Arabidopsis to inform analysis of data-poor crop species, ii) better definition of hypotheses for future experiments in model species that facilitate translational research, and iii) identification of robust network modules that can be applied in biomarker-assisted molecular breeding programs.

**Inspiration: Network-based classification of breast cancer metastasis**. The use of network modules to identify genes that underlie or predict traits such as NUE in agriculture - is inspired in part by the success of network-based diagnostics in other fields, including medicine. One proof-of-principle application by the lab of Trey Ideker, is the use of network-based classification of breast cancer metastasis [Chung et al 2007]. The Ideker study identified predictive breast cancer markers, not as individual genes, but as subnetworks extracted based on the intersection of transcriptome data and protein-interaction databases. Although genes with known breast cancer mutations are typically not detected through analysis of differential expression, they play a central role in the protein interaction network by connecting many differentially expressed genes. Thus, the study showed that subnetwork markers are more reliable predictors of outcome (e.g. metastasis vs. non-metastasis) than individual marker genes selected without network information [Chung et al 2007]. The goal of our NutriNet project, is to extend on these studies in medicine, to identify network modules that are diagnostic of – or involved in – NUE outcomes in agriculture.

**Prior studies of maize nitrogen biomarkers.**  While much is known about N-regulated transcriptome responses in Arabidopsis [Gutierrez 2008; Wang 2003; Kant et al 2011], and some work has been done in cereals including rice [Lian et al 2006] [Beatty et al 2009]. However, transcriptome studies of N-regulated genes in other crops are rare. The largest published study of N-responsive genes in maize identified biomarkers to detect nitrogen status [Yang et al 2011]. The goal of that study was to identify genes whose nitrogen response was *conserved across genotypes*, to use as reliable markers of N-status. By contrast, our NutriNet application proposes to identify genes whose nitrogen responsiveness *diverges across genotypes*, and to associate these differences with changes in NUE traits across the germplasm. Yang et al. note that “nitrogen-responsive gene expression does not necessarily indicate that a biomarker gene may help to detect and/or confer nitrogen use efficiency”. That is the goal of our NutriNet project.

**Exploiting Arabidopsis “Network knowledge” to identify genes involved in NUE in crops.** At the heart of the NutriNet project is using “network-knowledge” from Arabidopsis to inform network studies in crops. Network-based analysis in crops will enable scientists to interrogate biological function amongst lists of transcriptionally regulated genes in crops and derive testable hypotheses. To enable such network-based analysis in plants, the Coruzzi (NYU Biology) & Shasha (NYU Courant) labs collaborated on the creation of the Arabidopsis multi-network, which integrates all current knowledge about metabolic, DNA, RNA, and protein interactions in plants [Katari et al 2010], see Prior Studies. In one application, a query of the multinetwork with a list of 824 N-regulated Arabidopsis genes identified a connected sub-network of 346 genes. This network analysis identified transcription factor hubs to target genes in the N-assimilatory pathway involved in Gln synthesis and its conversion to Asn for N-transport to seed [Gutierrez et al, 2008]. The identification of CCA1 as a hub of this organic N-regulatory network uncovered new biology - nutrient regulation of the clock. Evidence that this network-oriented approach can be useful in agriculture, is supported by preliminary findings that N-regulatory networks in Arabidopsis [Gutierrez et al 2008] have aided the identification of genes associated with NUE in maize [Moose], as described below, and in a proof-of-principle example in Aim 3 of the Research Plan.

**Identifying NUE genes in Maize**. The Moose lab has been interested in identifying genes underlying NUE in maize using a combination of transcriptome and QTLs for NUE [REFS]. [Need a sentence or two here summarizing Moose maize work]. Remarkably, the network associated genes identified in the Arabidopsis organic-N regulatory network (e.g. bZIP1, CCA1, GLN1 and ASN1) [Gutierrez et al., 2008], were also found to control primary N-assimilation and Asn cycling in maize, which in turn contribute to NUE, as follows:

1. Levels of asparagine (Asn) in maize leaves sampled at night or during anthesis in ear-shoots are correlated with N-utilization. A positive relationship of Asn with grain-yield occurs at low-N, negative at high-N (positive with seed protein). These correlations of Asn with NUE traits are observed in all maize germplasm groups tested.
2. Promoter variation and mRNA levels for ASNase (asparaginase), which controls Asn degradation, shows that low ASNase expression reduces N-utilization in maize. This ASNase-trait association is observed in all maize germplasm groups.
3. bZIP1 mRNA levels is a trans-acting eQTL for AS, GS2, AspAT, NiR, which is associated with increases in N-utilization. The bZIP1-NUE trait association is observed in Illinois Protein selections and IBM RILs.
4. Promoter variation mediating response of the maize AS gene to diurnal rhythm – e.g. high level of AS gene activity during both night *and* day increases N-utilization. AS-trait association is observed in Illinois Long Term Protein selection and IBM RILs.

**Outcome**: The finding that N-regulatory gene networks discovered in Arabidopsis can help guide candidate gene analysis in a QTL analysis for NUE in maize, is a substantial finding which supports the notion that network-knowledge from Arabidopsis has the potential to enhance NUE studies in crops. This overlap is especially significant considering the experimental N-treatments in Arabidopsis (in the lab) and in the maize studies (in the field), were planned and executed completely independently in Arabidopsis by the Coruzzi lab (NYU) and in maize by the Moose Lab (U. Illinois), respectively. Our Research Plan highlights a collaborative effort between these two groups to conduct a set of coordinated experiments across a range of Arabidopsis and maize genotypes differing in NUE. The ensuing network modules will be used to develop testable hypothesis, enhance predictions of NUE based on gene expression data, and identify underlying genes. This network-module approach can be applied more broadly to enhance translational studies from models-to-crops for any agricultural trait of interest. In our application, this network-oriented approach will identify reliable markers as well as genes involved in NUE for use in molecular breeding programs, as detailed below.

-------------------------------------------------------------------------------------------------------------------------------

**Nutri-Net: A network-inspired solution to predicting NUE and to identify its underlying genes**. The approach and goal of NutriNet is distinct and novel compared to the previous maize N-responsive transcriptome studies [Yang et al 2010] which aimed to identify N-regulated genes conserved across maize genotypes in three important ways: i) ***Exploiting NUE Diversity***: Our goal is to exploit *differences* in gene expression across genotypes with diverse NUE profiles, to identify the genes that underlie or correlate with NUE, exploiting crop diversity. ii) ***Network Modules***: We will interrogate transcriptomic data from maize in the context of gene and protein interaction data using “network knowledge” from the data-rich model Arabidopsis, to identify network modules that predict or cause changes in NUE in crops. iii) ***Translational Networks***: Comparison of NUE networks across species and ecotypes, will enhance discovery of core vs specific nitrogen-regulated networks, as well as to create a fully-fleshed out cross-species network prototype to enhance translational studies from models-to-crops (Fig 1.1).

**Aim 1. Exploiting Genetic Diversity in NUE (Coruzzi/Moose)**.

**Rationale**: This aim exploits the comparative analysis of genetic diversity – across maize lines and Arabidopsis ecotypes- to generate a dataset that will allow us to identify gene network modules associated with NUE. An initial set of genotypes, known to cover a broad phenotypic variation in NUE, will be scored for changes in gene expression *and* NUE traits under distinct N-treatments. This genomic and phenotype data will be used to drive a multivariate network analysis, that will identify network modules that are predictive of, or underlying NUE in Aim 3. We expect to find network modules that are species and/or genotype specific, as well as ones that are broadly conserved.

**EXPERIMENTAL APPROACH:** We will collect transcriptome data and NUE phenotypes in a set of selected set of Arabidopsis accessions (Coruzzi, NYU) and Maize lines (Moose, Illinois) that differ in NUE traits. See details for the selection of lines from each species below. This genomic and phenotypic data will be assembled across an initial set of 10 genotypes, representative of a broad phenotypic variation in NUE. The resulting data will be used to fuel the gene network analysis in Aim 3A, and the gene network-to-NUE phenotype analysis in Aim 3B.

**EXPLORING NUE DIVERSITY IN MAIZE:**

**Selection of NUE lines and N-treatments:** Maize has a rich history of selective breeding to optimize numerous agronomic traits. The Nested Association Mapping (NAM) population was developed using 25 parent lines that capture the genetic diversity incorporated in current elite Maize lines, hence serves as an excellent basis for our studies. In our previous studies we measured NUE in 50 inbred lines (replicated trials, 3 years) and showed that large variation in NUE exists across the NAM parent lines (Fig 1.2). We used total biomass here, because unlike nearly all other studies of maize NUE that only consider grain yield, our measure of N utilization is less confounded by the variation in harvest index that is well-documented to exist between maize inbreds and hybrids, and between photoperiod-sensitive versus insensitive maize genotypes. Inbreds have very different phenotypic distributions from hybrids, but the relative ranking of inbreds or their hybrids is similar, which indicates that although hybrid vigor impacts the phenotypic value for N utilization, genetic correlations remain high.  Therefore, understanding gene-phenotype associations for N utilization in inbreds can be relevant to hybrids when total biomass is used to normalize the harvest index effect. As part of this study, we propose to sample the transcriptome of 20 maize lines from multiple tissues. Based on the phenotypic data from our previous studies, we select 10 initial lines that span the NUE diversity shown in Fig 1.2B. Fig 1.2B shows, in hybrids of B73 and each of 8 other lines chosen for sampling, the accumulation per plant in four categories influenced by Nitrogen supply. IHP1, MS71 and Oh43 show only modest improvements in yield when supplied with excess N while Mo17, OHG84, Oh7B and ILP1 show dramatic increase in yield at high N conditions. Mo18W represents tropical maize lines that generate higher biomass overall and shows marked improvement under high N. B73 and W22 were chosen as reference lines as the former is the reference genome for Maize and the latter is the background for the maize mutant population to be used in subsequent sampling. The analysis of transcriptome data and phenotypic data generated from this aim and its analysis in Aim 3 will guide our choice of 10 lines for subsequent sampling. Based on the findings of the cross-species gene network analysis in Aim 3 using this initial dataset (“learning set”) we will choose other genotypes that best test perturbations of the network, eg. mutants, transgenics, near-isogenic lines for natural variants for the second round of sampling.

**N-treatments and sampling:** All lines will be grown in controlled field conditions at two levels of Nitrogen, one N deficient (XX kg/ha) and the other excess N (XX kg/ha). Multiple plants for each line will be grown in a random block design and sampled at 8 days after pollination. The leaf, ear and developing seed tissues will be sampled separately to allow addressing the question of N remobilization. An efficient sequencing strategy similar to the one described below will be followed to allow sampling of 20 lines and 3 tissues with replications in a cost-effective manner.

**NUE traits:** For each of the proposed lines, except W22, we have measured the Seed yield, stover biomass, and Nitrogen in plant and seed separately in replicated trials over 3 years. A total of 50 inbred lines including the NAM and PVP parents were phenotyped for these traits. W22 and any mutants included in the second round of sampling will be phenotyped during the proposed study.

**EXPLOITING ARABIDOPSIS NUE DIVERSITY:**

**Selection of Arabidopsis Lines:** Previous measurements of Arabidopsis accessions grown on low (2mM) vs. high (10mM) nitrate uncover a wide variation of NUE traits in 19 accessions [Ikram et al 2012, Chardon et al 2012]. NUE traits measured included seed dry weight (DW), harvest index (HI), nitrogen harvest index (NHI), and N%SEEDS [Chardon 2012]. These accessions were then classified into four ideotypes as models for specific crop types based on their NUE traits (See Fig. 1.3 as adapted from Fig 4 in Chardon et al 2012). Specifically, accessions in *ideotype 1* could serve as models for rice and wheat, that require high grain mass, large HI and NHI, and high protein content in grain. *Ideotype 2* accessions could serve as a model for maize, rapeseed, and barley, for which low N in seeds is required because it is antagonistic to starch or oil accumulation. *Ideotype 3* accessions could serve as models for silage crops, for which require high vegetative biomass with a high content of proteins. Finally, accessions in *ideotype 4* could serve as a model for biofuel crops with high total biomass at harvest and low-N needs. Transcriptome studies were not performed in this initial analysis [Chardon 2012].

**N-treatments and data collection**: Briefly, Arabidopsis seedlings from the 19 accessions will be grown (in duplicate) under conditions reported in [Chardon et al 2012] under low vs. high nitrogen supply, using 15N as tracer (see below). The first replicate will be harvested to collect RNA from roots and leaves separately at the fully-grown adult stage, but prior to bolting. The second replicate will be harvested at the 5 silique stage, and RNA will be collected separately from leaves, roots, and influorescence. We will collect total RNA, and extract mRNA and smRNA simultaneously. The “balanced block” design and bar coding will make it feasible and economical to collect mRNA data from all 19 lines. However it will be cost prohibitive to sequence both mRNA and smRNA on all 19 line. We will therefore first examine variation in levels of specific mirRNAs identified to be N-responsive in Aim 2, and perform RNA-seq on a subset of lines. We will also collect NUE phenotype data including dry weight, total N content, and 15N-uptake (see below), to ensure comparability to the previous studies [Chardon et al 2012].

**Quantifying NUE traits.** Using 15N as a tracer, and performing 14N measurements, the rate and amount of NO3 incorporation into total N can be measured as a function of dry weight (*e.g.* biomass). 15N measurements will be performed as described [[Lejay et al 1999](#_ENREF_65)]. Briefly, 15N accumulation in plants is measured by supplementing media with 1% 15N excess for long labeling times. The total N-content and atom % of 15N abundance will be determined by Continuous-Flow Mass Spec, as described [Clarkson 1996], using a Euro-EA *Eurovector* elemental analyzer coupled with an *IsoPrime* Mass Spectrometer (GV instruments, Crewe, UK). This work will be performed collaboratively with Drs. Krouk and Ruffel (INRA, Montpellier). Previously unpublished studies show that this method is able to quantify the range of total N in total biomass (leaves + roots) across 45 Arabidopsis accessions (Fig 1.4). We will use the 15N and total-N and dry weight data from leaves, roots and seeds to calculate the NUE parameters and relate them to the traits previously measured by Chardon et al. [Chardon 2012]. In our study, we will repeat measurements of DWSEEDS, DWVEG, and HI (Harvest Index), in addition to the N-uptake measurements with 15N and N abundance measurements at the vegetative and reproductive stages. The overlapping traits DWSEEDS, DWVEG, and HI between the two studies will ensure the comparability of experiments and ability to cross-utilize trait measurements.

**TRANSCRIPTOMIC APPROACH: Maize AND Arabidopsis**

**RNA-Seq Analysis: A “balanced block” design for RNA-Seq enables efficiency-of-scale**. We aim to perform reliable quantitative comparison of gene expression across tissues (i.e. leaves, roots, influorescence) covering a wide range of NUE phenotypes across natural variants (e.g. 10 maize, and 19 Arabidopsis lines). To satisfy these demands in an efficient and economical way, we will use a “balanced block” RNA-Seq design with multiple levels of replication as described in [Auer and Doerge, 2010]. This design, which we have successfully implemented in the Coruzzi lab, involves preparing RNA-Seq libraries from 3 replicates of 3 tissue-types simultaneously, and bar-coding each library with a different index. These barcoded libraries are then pooled in equimolar ratios and sequenced on all 8 lanes of a single Illumina flow cell (see **Fig. 1.5**). This design accounts for biological variation between individuals and lane-to-lane variation during sequencing.

**AIM 1 OUTCOME:** Experiments proposed in this aim will produce two parallel and related data sets that are critical to uncovering the molecular basis of variation in NUE. These studies will sample the breadth of variation in NUE among diverse germplasm as measured through phenotypic traits and capture the underlying transcriptome for the same diversity in N response. Analyses of the transcriptomic measurements in Aim 3 will help us resolve both, the common network modules governing global nitrogen-response across multiple lines and highly divergent species, as well as the variation within these modules among the different lines. Overlaying the NUE trait information on top of the transcriptome data, allows us to see the effect of the differences in expression of individual and network modules of genes on the overall Nitrogen Use Efficiency of the plant. Sampling the plants at the vegetative and reproductive stages, allows insights into the sub-processes of nitrogen uptake and remobilization that together compose the overall nitrogen utilization by a crop species.

**Aim 2. Role of root-shoot N-signaling and root foraging on NUE (Coruzzi)**. **Rationale.** The mechanisms underlying a plant’s ability to forage for and acquire nitrogen and other nutrients in a heterogeneous soil environment, is a critical aspect of NUE that is currently under-explored. We propose to fill this knowledge-gap using a split-root system which is uniquely suited to studying the impact of a heterogeneous nutrient environment on NUE. This system, highlighting a root-shoot-root signalling mechanism [Ruffel et al., 2011], will also enable us to gain the first insights into the role that *shoots* play in feedback mechanisms controlling root nutrient foraging. Split-root experiments will be conducted in selected lines of Arabidopsis and maize, and NUE traits will be measured (using 15N tracer) and transcriptome responses monitored in both shoots *and* roots. The comparative and integrated analysis of this data in Aim 3, will identify potential NUE network modules involved in root nutrient foraging and root-shoot communication for validation and field-testing in Aim 4.

**Approach. Split-roots: Measuring NUE in the face of a heterogeneous nitrogen environment.** The split-root system provides a unique perspective on the impact of heterogeneous soil N-conditions on plant NUE. In the ***split-root system,*** roots of a *single plant* are exposed to two distinct nitrogen environments; N-replete (Sp.KNO3) vs. N-deplete (Sp.KCl). As each ½ of the roots of the split (Sp.) and control (C.) plants are exposed to the same *local* nutrient environment (e.g. KNO3 or KCl), any differences between Sp. and C. plants are due to the distal or systemic signals on the other root ½ - e.g. **N-demand** (Sp.KCl) or **N-supply** (Sp.KNO3) [Ruffel et al. 2011; Liu et al 2010]. Of particular relevance to this project, is that split-root studies conducted in Arabidopsis [Ruffel 2011] and maize [Liu et al 2010] both uncovered a stimulation of lateral root (LR) outgrowth on the N-replete side, when the other root 1/2 encounters a N-deplete environment (Fig 2.1). Using a genomic approach in Arabidopsis, we discovered that this “N-demand” signal operates through a root-shoot-root signal relay, involves nitrate-signaling, and is cytokinin-dependent [Ruffel et al 2011]. While the maize split root studies revealed parallels to the Arabidopsis findings –e.g. distal N-deprivation signal stimulates lateral root outgrowth in roots exposed to N-rich environments - the molecular basis was not explored [Liu et al 2010]. Another unexplored aspect in both Arabidopsis and maize split-root studies - is the role of *shoots* in mediating the nutrient foraging responses of the roots. Indeed, shoots provide the energy- and C-supply that drive nitrate reduction and assimilation into organic form. By comparing results of split-root experiments across shoots and roots of Arabidopsis and maize, we should uncover novel molecular insights into the shoot-root interplay in C3 and C4 species- both commonalities and differences.

**Selection of NUE varying lines for split-root studies.**  We willperform a comparative split-root study on maize and Arabidopsis lines whose NUE phenotypes varying in response to nitrogen treatments. Specifically, we will select lines that show: 1. High NUE response to N-treatment, 2. Intermediate NUE response to N-treatment, and 3. No difference in NUE response to distinct N-treatments, as shown in Fig. 2.2 A&B. A comparison of lines with these three NUE states can help identify network modules responsible for the N-responsiveness of NUE, by selecting gene expression patterns that are coherent with the NUE states (see a schematic example in Fig 2.2C).

**Experimental Plan**: Split-root analysis will be performed with seedlings of the selected maize and Arabidopsis lines, and we will collect data on NUE phenotypes (15N, total-N, biomass) as well as RNA-seq data profiling both mRNA and sRNA from shoot and root tissue – including each ½ of the split roots (as shown in Fig 2. 3). For Arabidopsis, we will follow the protocols in [Ruffel et al. 2011], but this time using 15N as tracer to measure N-uptake and NUE traits, as described in preliminary studies below. For maize, we will follow the protocols for split-root analysis of maize seedlings, used in [Liu et al 2010], for which we also have preliminary results as below.

**Preliminary Studies: Split-root studies in maize:** In a pilot study, maize line B73 (the accession with full genome sequence available) was grown in hydroponic conditions for 3 weeks before the seedlings reached 20cm high, when the roots were separated into two groups in a split-root setup: one root 1/2 was exposed to 5mM KNO3while the other root 1/2 was exposed to 5mM KCl as control (Fig 2.4A). The root morphology of each root ½ was analyzed 12 days after the split-root treatments. We observed the previously reported enhanced LR growth on the sp.KNO3 side (Fig 2.4B) [Liu et al, 2010], a response which is conserved between Arabidopsis and maize (Fig 2.1). A new observation in our studies, was an increase in secondary lateral root growth on the sp.KNO3 side in maize (Fig 2.4B&C), which is not observed in Arabidopsis in a comparable experimental set-up [Ruffel et al 2011]. This partially conserved and partially distinct phenotypic responses to the heterogeneous N-environments of maize and Arabidopsis roots, suggest that by comparing the transcriptome response and N-use traits of these two species, we will likely uncover commonly employed mechanisms for plants to forage for nutrient-rich patches in the soil, as well as identifying additional species (e.g. crop)-specific mechanisms.

**Using 15N to measure NUE parameters in split-root plants.** Using 15N as a tracer, and methods described in Aim 1, in a previous unpublished studies conducted with Drs Ruffel and Krouk, we show that this method is able to quantify the difference in 15N incorporation in the Sp.KNO3 vs. Sp.KCl ½ of the split-roots, and to quantify the range of total N across a range of Arabidopsis accessions (Fig 1.4). We will use this method to derive NUE traits for Arabidopsis and maize seedlings, using some of the parameters for NUE in vegetative tissues described by [Chardon et al 2012.]

**Transcriptome responses to heterogeneous N-environments: Role of shoots and small RNAs.** In a pilot study, Illumina deep-sequencing technology was used to profile mRNAs and sRNAs in Arabidopsis (Col-0), 4h after starting the split-root N-treatment*.* In addition to identifying root genes previously described using microarrays [[Ruffel et al 2011](#_ENREF_8)], **this new RNA-seq analysis of mRNA** identified 73 genes in Arabidopsis that are specifically induced *in the shoots* by heterogeneous nitrogen states of the split-*roots*. These 73 genes are enriched in functional GO-categories like energy metabolism, photosynthesis and C-metabolism. This result enables us to derive a hypothesis that carbon export from *shoots* provides energy and C-skeletons to drive lateral root outgrowth and nitrate reduction/assimilation in the Sp.KNO3 roots. This hypothesis is supported by previous physiological studies of C-flux in split-root plants in *Medicago* ([[Jeudy et al 2010](#_ENREF_51)] and S. Ruffel, unpublished data). Comparing shoot responses to split-root nitrogen treatments in Arabidopsis and maize, will provide mechanistic insights into how and whether genes involved in C-flux from shoot-to-root are similar or distinct across a C3 and C4 species. Indeed, we found that 22% of the 195 genes responsive to the split-root conditions in Arabidopsis ( [Ruffel et al 2011] and the pilot study) are also present in a maize N-response network described in Aim 3A (probability of the observed overlap = <0.038). **Our pilot study on small RNAs** also uncovered a set of known miRNAs that are specifically regulated by a supply/depletion of nitrogen in Arabidopsis. A comparison between these Arabidopsis N-regulated miRNAs and a published maize study [Zhao et al, 2012] reveals that despite the distinct experimental conditions and C3 vs. C4 species, some N-regulated miRNAs are shared by maize and Arabidopsis which may play a role in mediating N-responsive root development (Fig 2.5). In addition we will look for miRNAs and siRNAs that differentially accumulate in the split-root plants compared to the control (*e.g.* Sp.KNO3 vs. C.KNO3), because they possibly mediate systemic N-signaling.

**AIM 2 OUTCOME**: This aim will generate data on N-use trait measurements and transcriptomes in a unique split-root experimental setup that highlights a plants’ ability to forage for nutrients in the soil. The split-root system also provides a window into shoot-root communication mechanisms, and thus provides windows into two important aspects contributing to NUE. Since the data will be generated with three maize varieties and three Arabidopsis accessions displaying diversity in NUE, the analysis will shed light on the mechanisms underlying the NUE differences within and between species. This data will feed into the comparative and integrated network analysis in Aim 3, which will identify potential NUE network modules involved in root foraging (including miRNA-target pairs) and root-shoot communication for validation and field-testing in Aim 4.

**AIM 3. Constructing NUE Networks (Shasha/Coruzzi).**

**Rationale**: Here, we will integrate transcriptome and NUE data from Aims 1 and 2, to identify network hubs and modules regulated by nitrogen (Aim 3A), and to “learn” the networks associated with NUE (Aim 3B). Components of these network modules validated *in silico* using “left-out” data in Aim 3B, will be experimentally validated using mutants/transgenics in Arabidopsis, and tested as predictors of NUE in maize in field tests, as described in Aim 4.

**Approach**: We will use two complementary approaches to integrate and analyze transcriptome and NUE data generated in Aims 1 and 2. In Aim 3A, we will generate *crop-specific* and *cross-species* N-regulated gene networks using functions available in VirtualPlant (www.virtualplant.org). The interoperability of our newly created VirtualPlant maize with the existing VirtualPlant Arabidopsis [Katari et al 2010], will make it possible to interpret the maize data in the context of Arabidopsis “network knowledge”. This will provide annotation and protein:protein interaction data which is available only in this data-rich model. In a complementary approach in Aim 3B, we will build multivariate networks that link network modules to NUE phenotypes based on gene expression data and also using protein:protein interaction data. These modules will be tested for their ability to predict NUE based solely on gene expression data and validated using left-out data, as described in Aim 3B. The network modules identified in either Aim 3A and/or Aim 3B, will also provide a focus for hypothesis-driven testing of candidate modules *in planta* in Aim 4.

**Aim 3A**. **Building crop-network using Arabidopsis “network knowledge”**. We will use VirtualPlant maize (and construct VirtualPlant wheat for ERA-CAPS), to enable the creation of crop-specific gene interaction networks. With these tools, users can construct gene networks on-the-fly based on crop-specific genomic data (e.g. transcritpome, metabolome, etc.). They can also enhance the crop networks by readily leveraging Arabidopsis “network knowledge” which will provide annotation, interaction data (e.g. protein:protein, protein:DNA, small RNA:RNA, etc.) available through the multi-network analysis feature in VirtualPlant. Note that the quantitative metabolite data generated by our ERA-CAPs colleagues, can be readily incorporated into metabolite nodes in the Arabidopsis multinetwork [Katari et al 2010], enabling us to interpret transcriptome and metabolome data in an integrated network.

**ERA-CAPS and Networks:** If the companion ERA-CAPS grant is funded, we will, in addition to existing maize and rice versions of VirtualPlant (www.virtualplant.org), create a VirtualPlant wheat database. We have reached a fair degree of automation in inducting new species into VirtualPlant. At present, the wheat genome is a draft genome that is not fully annotated; therefore the current LCG assembly of hexaploid wheat [Brenchley et al. 2012] will be used to construct the initial version of VP wheat. It is important to note that RNA-Seq data from wheat can be remapped to newer gene models as they get updated, without a need to repeat the experiments. The ERA-CAPS collaboration, will also provide a greater dataset input, more species and, additional to nitrogen (N), also treatments on sulphur (S) and phosphate (P). Cross-species network analyses with these additional datasets will allow comparison and identification of robust nutrient network modules (e.g. to N, P or S) and their interrelation.

***Rationale:*** To demonstrate how network analysis and tools can be used to translate knowledge from one species to another, especially to leverage Arabidopsis data for translational studies in crops, we performed a proof-of-principle study using a publicly available microarray N-treatment dataset that discovered gene expression biomarkers for the nitrogen status of maize *in planta* (Yang et al. 2011).  To illustrate the cross-species network analysis, we present a concise step-by-step walkthrough of this N-treatment maize dataset and its comparison to Arabidopsis multi-networks (see Fig. 3.1).

**A Case study of maize networks derived from cross-species comparisons to Arabidopsis.**

**Step 1. The maize data:** We created a database for *Zea mays* in VirtualPlant maize ([www.virtualplant.org](http://www.virtualplant.org)) using two publicly available sources Maize Sequence [http://maizesequence.org] and Phytozome v8.0 [http://www.phytozome.net]. The microarray dataset used in this case study contains a total of 90 microarray samples from nitrogen-treated maize plants from the Yang et al 2010 study [Yang et al. 2010].

**Step 2. Building a maize correlation network.** Using the current functions in VirtualPlant ([www.virtualplant.org](http://www.virtualplant.org)) [Katari et al 2010, Plant Physiol.] recently updated for maize, the N-treatment (is N here nitrogen or nutrient?) microarray data from maize [Yang et al 2010] was normalized using the RMA method, followed by a 2-way ANOVA. This analysis identified 5,000+ genes that are differentially regulated in response to N-treatment and time of day (FDR cutoff <0.05). Next, VirtualPlant maize was used to generate pair-wise gene correlation using the Pearson method, and all gene pair correlation values with a p-value <0.05 were used to create a maize gene expression correlation network, consisted of ~5,000 genes (Fig. 3.1). This maize correlation network is too large to enable focused hypothesis generation, and >50% of the maize genes in the network are unannotated, making hypothesis generation impossible. To aid the interpretation of the maize correlation network, we next analyzed the maize correlation network in the context of i) the knowledge of gene interactions in the Arabidopsis multiNetwork [Gutierrez et al 2007, Katari et al 2010], and ii) correlation networks based on Arabidopsis N-responsive genes (Fig 3.1), as described below.

**Step 3. Creating a maize translational network using Arabidopsis “network knowledge”.** We translated the ~5,000 N-responsive genes from maize to 3,756 Arabidopsis homologs using the “one-to-many” [I think this is many-to-one if we’re using the best hit] homology mapping function in VirtualPlant, which uses the maize “best-hit” to Arabidopsis data provided by Phytozome ([www.phytozome.net](http://www.phytozome.net)). Mapping “network knowledge” from Arabidopsis (i.e. protein:protein interaction, metabolic edges, and text mining) resulted in a network of ~2,000 connected maize genes and 10,800 edges. It is challenging to derive focused hypotheses for testing from such a large network. Therefore, we further refined our analysis by selecting genes that are N-regulated in both maize and Arabidopsis. To this end, we created an Arabidopsis nitrogen response gene set (1,254 genes) which is a union of the 866 genes responsive to nitrate/ammonium in seedlings [Gutierrez et al 2008] with the 462 genes responsive to nitrate in roots [Wang et al 2004]. The 1,254 Arabidopsis “nitrogen genes” and the 3,756 maize “nitrogen genes” were then intersected to produce a common gene list of 327 genes (a highly significant overlap (p <0.001)) according to the “GeneSect” function in VirtualPlant. This list of 327 N-regulated genes conserved in maize and Arabidopsis includes 15 transcription factors. Next, the Arabidopsis multinetwork was queried with these 327 N-regulated genes, shared between Arabidopsis and maize, by selecting all KEGG metabolic, protein interaction, and regulatory edges – determined based on correlation (>0.7 or ,-0.7) (within maize) and based on over represented cis-element binding site (using AGRIS), as described in [Gutierrez et al 2008], and [Nero et al 2009, BMC Bioinformatics]. This analysis resulted in a network of 286 genes that are N-regulated in both maize and Arabidopsis, and connected in a gene network based on metabolic, protein-DNA, protein:protein or other interactions. Importantly, the Arabidopsis “network knowledge” not only helped to focus on maize genes for translational studies, but also was crucial to network annotation, as it annotated 50% of the genes in the maize network, which are annotated as “hypothetical” or “unknown” genes in the maize genome.

**Step 4. Validation of the maize translational N-regulatory network:** The cross-species network analysis in this case study enabled us to identify and help annotate a network of 286 N-regulated genes in maize, which form a set of bio-modules.

This list of 286 N-regulated interacting genes in the “maize translational network” (e.g. conserved between maize and Arabidopsis) includes 15 transcription factor hubs. Several of these TF hubs have been previously validated to regulate genes involved in N-assimilation in Arabidopsis [Gutierrez et al 2008] and have also been implicated in QTL studies of NUE in maize [NEED MOOSE REFS]. These TF hubs of this conserved N-regulatory network provide a focus for network module identification, hypothesis testing and validation. For example, an exploration of the “network neighborhood” around the transcription factor hub CCA1 uncovered a subnetwork module of 59 genes including targets involved in nitrate uptake and sulfur uptake (Fig. 3.2). Using functions in VirtualPlant, a BioMaps analysis of the Arabidopsis homologs of this bio-module (59 genes) reveals an overrepresentation in GO terms including photoperiodism (p-val <0.005) and nitrate transport (p-val <0.01). The conservation of this network module between maize and Arabidopsis, reinforces the discovery that nitrogen-regulation of CCA1 imparts nutrient regulation of N-assimilation and the circadian clock in Arabidopsis [Gutierrez et al 2008]. This case study now extends this discovery to maize, and also suggests that nitrogen regulation of CCA1 coordinates uptake of nitrogen, and sulfur which is highly supportive of our companion ERA CAPS application, which will add data on phosphate and sulfur nutrient responses to our network studies. Thus, this Arabidopsis-enabled maize network can now be used as targets for transgenic studies and as molecular markers for N-response in translational studies. [This subaim is awesome☺]

**Aim 3B.** **Identifying NUE-predictive network modules**. The goal of this subaim is to identify network modules that are predictive of NUE traits. The resulting network modules will be tested for their ability to predict NUE outcomes using a “training-based” approach. Specifically, we will create NutriNet network modules using gene expression and NUE phenotype data from several varieties (e.g. 5-10), and test their performance in accurately predicting known NUE phenotypes in “left out” varieties, based solely on gene expression data. NutriNet modules discovered using this approach will be advanced to Aim 4 for lab and field tests of the most promising predicted modifications.

**Approach**: This aim combines gene expression data with protein:protein and other interaction data from Arabidopsis, to test whether the resultant gene network modules allow better prediction of agricultural traits – e.g. NUE - as successfully as it did in the case-study for medicine (e.g. prediction of breast cancer metastasis) [Chuang et al 2007]. In the spirit of the Chuang et al. paper, we will make use of the protein-protein interaction network (and possibly other network data) to refine the gene set that could influence biomass for hypothesis generation. Procedurally, we will include both members of a protein-protein interaction gene pair in such networks, if their mean expression correlates significantly with biomass. To test whether protein-protein interaction pairs are helpful, we will compare the mean expression correlations of random pairs with those of protein-protein interaction pairs. Further, we will quanity how many new genes this method yields, compared to a method that looks only at single gene-to-trait correlations at the same p-value cutoff. [I suggest removing the following sentence. No reason to be argumentative] While we will follow the “spirit” of the breast-cancer network module study by the Ideker lab [Chuang et al 2007], we propose improvements to the statistical analysis in the study to increase specificity and sensitivity of the approach in two approaches described below. We will test the effectiveness of Approach 1 (see Fig 3.3) vs. Approach 2 for their ability to predict NUE using left-out data sets, to determine which approach gives the best results. [I suggest that we need either one approach or the other. I think approach 1 is fine and it reads better]

***Approach 1***: “**Paired scores” approach to identify network modules**. Note, that we describe here a method for pairs of genes, but the same approach can be extended to triplets, quartets etc.

**Step 1:** **Single Gene expression to NUE association:** The experiments from Aim 1 provide high confidence gene expression measurements across 19 Arabidopsis ecotypes, and at least 10 maize lines, under low and high nitrogen regimes, and at two developmental stages. The fold-change in gene expression between the low and high N [Again, N means Nitrogen here but “nutrient” in NUE] conditions across the ecotypes, will be correlated with the appropriate NUE trait. For example, at the vegetative stage, the appropriate NUE trait might be %N in shoots or total Biomass at high N. Thus, for each trait, each gene *i* has a correlation value (Ci) and an associated p-value (pi).

**Step 2: Calculating Paired scores:** A paired score is calculated from the p-value of the correlation of a gene pair instead of the correlation coefficient itself, since this allows the members of the pair to be inversely correlated with the trait. For every possible pair of gene i and gene j, PSij is calculated as a measure of additive correlation confidence of that pair to the NUE trait. Specifically, PSij = -log10(pi) + -log10(pj). The overall distribution of the paired scores PSij is analyzed to identify the underlying distribution model.

**Step 3: Protein-Protein interaction pairs:** We test the Paired Scores of gene pairs that are predicted to interact at a protein level by protein:protein interaction data in Arabidopsis against random pairs, of a matching sample size, from Step 2 to draw a significance cut-off. Protein-protein pairs whose paired score is higher than the significance cut-off are identified as components of network modules that affect the trait of interest.

**Step 4: Linking pairs to generate network modules:** Two significant pairs are linked if they share one of the genes. This process is repeated iteratively until all possible links are exhausted. The resulting modules should have sets of genes that have a) Protein interaction edge and b) Significant paired score with at least one other gene in the module.

**Step 5: Compare network modules across species:** The network modules identified independently in each species are overlapped to identify core conserved modules that affect the same trait across the species. Additionally, the direction of correlation for each gene in these modules is noted to find optimal configurations of gene expression that produce the desirable trait (Fig 3.4). These modules and configurations will be validated in Aim 4 under laboratory and field conditions.

**Approach 2: “Composite expression” scores approach.** This approach is similar in principle to the Chuang et al. 2007 method of correlating “composite scores” of pairs of genes to their trait of interest (e.g. breast cancer metastasis). The major improvement we suggest is to calculate the composite score in a way that would allow trait optimization by concurrent and antagonistic (or even more complex) changes in the gene expression level within a module. Such changes would be smoothed out when the composite expression value is calculated simply as an average expression of the module in Chuang et al 2007, and thus may miss opportunities to regulate our trait of interest, NUE. Our approach to the implementation is as follows:

**Step 1: Composite scores and correlation:** For every pair of genes, the composite score is calculated as the average expression of both genes and this score is correlated to the Trait measurement across lines. Note, that this approach can be trivially extended to triplets of larger groupings of genes. The output here would be the p-value for the correlation of the composite score to the trait. An alternative approach is to fit a linear regression model with the expression values of the genes as the factors, thus capturing the distinct effect of each gene, and the trait measurement as the response variable. The output would be the p-value of the best model for each set of genes and additionally the coefficients for each gene from the model.

**Step 2: Significant modules:** The distribution of p-values from Step 1, is used to identify a significance score cut-off. All pairs with a significant p-value are further grouped using shared genes to create network modules, as described in Approach 1. These modules will be prioritized for validation in Aim 4, based on net significance of the module, computed as the sum of –log10P values for each pair in the module.

**TESTING OF NETWORK MODULES IN SILICO***In silico* Testing with left-out varieties: Aim 3A and 3B both identify Network modules relevant to NUE traits. These network modules will be refined further by testing their ability to discriminate NUE outcomes in both species using “left out” data. Briefly, the network construction and module identification (Fig 3.3) will be performed with data from all but one of the lines. The resulting modules will be used to predict the NUE outcome of the left out line based on the direction of correlation of the module composite expression with NUE. For example, if the composite expression level of a given module is negatively correlated with the NUE trait and the left out line has a low composite expression for that module then we expect the line to have high NUE. Modules that consistently predict high NUE will be prioritized for validation in Aim 4.

**Testing configurations of network modules in Arabidopsis ideotypes:** The network modules we identity in Aim 3 are relevant to NUE. Here, we will test how the relative expression of genes in these modules correspond to the four model ideotype for crops. Expression patterns within each module can be compared across the Arabidopsis accessions from each ideotype, to identify configurations of the expression state of each module that are “signatures” for that ideotype. For example, for each network and/or network module associated with an NUE trait, we classify expression values of each gene across the 19 Arabidopsis lines representing the four ideotypes. Genes whose level of expression is significantly associated with Arabidopsis lines belonging to one ideotype will form the network signature for that ideotype, as depicted in Fig 3.4.

**Aim 3 Outcome:** The first outcome of Aim 3 is the identification of N-regulatory networks conserved between Aradiopsis and maize (Aim 3A). These conserved network modules provide a focus for translational studies, and as N-responsive biomarkers. In Aim 3B, the outcome is the identification of network modules of high relevance to NUE traits. The individual members of such modules, and in particular the highly connected hubs, provide targets for further studies in NUE. These modules can guide future breeding efforts to improve NUE by incorporating the higher performance alleles into elite hybrid lines. The second important outcome of Aim 3B, is to identify the absolute and relative expression levels of interconnected genes within each module that optimize NUE. Such optimal configurations will be lab and field tested in Aim 4 to validate strong associations with NUE.

**Aim 4. Nutri-Net module validation in laboratory and field conditions**. **(Coruzzi/Moose)** **Rationale**. The network modules identified in Aim 3 will be validated using species-appropriate methods. **Laboratory validation in Arabidopsis**: Gene components of NutriNet modules will be functionally tested for their role in NUE using mutant and transgenic Arabidopsis in both whole plant and split-root studies (Aim 4A). **Field validation in maize**: Initial field validation of network modules will be performed on diverse maize populations representing both historical and current elite hybrids (Aim 4C). Genotypes predicted to possess the optimal configurations of gene expression levels of genes in the network modules, will be evaluated in field trials with different regimes of nutrient availability. **Arabidopsis-to-Maize validations**: A comparative analysis of lab-to-field results will identify the “translation” of knowledge from Arabidopsis to crops. We will test the functional “translation” of Arabidopsis-to-maize networks using a cell-based assay for TF networks in Aim 4C. This high-risk, high-payoff sub-aim, could potentially provide “mode-of-action” data for potential commercialization of translational discoveries from models-to-crops.

**Prioritizing network modules and components for lab and field validation studies**. Aim 3 will identify a number of network modules and associated transcription factor hubs. Network modules shared by maize and Arabidopsis (Aim 3A) and ones associated with NUE traits (Aim 3B) will be prioritized for validation. Our preliminary studies in Aim 3A, already identified a conserved network of 287 genes, as we as 15 conserved TF hubs, some of which were previously validated to target genes in the N-assimilation pathway in Arabidopsis ( GLK1, CCA1 and bZIP) [Gutierrez et al 2008] and have been implicated in maize QTL NUE studies.

Aim 4A. Laboratory-based validation: Arabidopsis. In this subaim, components of NutriNet modules will be functionally tested for their role in NUE *in planta*, using mutant and transgenic Arabidopsis. We will test Arabidopsis T-DNA mutants and transgenics in these TFs for alterations in NUE (compared to wild-type Col-0) both in whole plant assays described in Aim 1, which allows us to monitor NUE traits in seeds. We will also test mutant/transgenic seedlings in the split-root assays described in Aim 2, to assess the role of the TF in controlling plant NUE response under heterogenous N-environment conditions, which is most reminiscent of plants grown in the field. The split root assay also provides a unique window into components affecting shoot-root interplay and its affect on NUE traits including root nutrient foraging.

Aim 4B. Field-based testing: Maize. Nitrogen-regulatory networks identified in Aim 3A and ones in Aim 3B that are associated with NUE traits will be assayed in field grown-lines. Initial field validation will be performed on diverse maize populations representing both diversity and current elite hybrids, and wheat breeding germplasm (EU ERA-CAPS). The elite lines of maize represent germplasm that has been selected for seed yield albeit under high N conditions. The diversity lines differ greatly in seed yield, and also in NUE, and provide a broad range of mechanisms governing NUE. We anticipate that the expression levels of important N-regulatory hubs identified in Aim 3A will vary appreciably across the maize diversity lines. Gene expression within network modules identified in Aim 3B as being predictive of NUE will be assayed across these maize lines. We will determine if the expression states of genes in conserved network modules are predictors of NUE within and across crop species. Genotypes predicted to possess the optimal configurations of gene expression levels in the NutriNet network modules, will be evaluated in field trials with different regimes of nutrient availability. This will be performed by Q-PCR or by transcriptome analysis of genes in network modules.

**Aim 4C. Maize-Arabidopsis cross validation of network hubs:** This aim will focus on the cross-validation of TF hubs conserved between network modules of maize and Arabidopsis, to identify mode-of-action and translational relevance. For this, we will use a rapid cell-based assay called *TARGET* (*T*ransient *A*ssay *R*eporting *G*enome-wide *E*ffects of *T*ranscription factors**)** that enables us to rapidly validate TF🡪targets genome-wide [Bargmann et al 2013]. The advantage of this cell-based assay, is that we can compare whether the genome wide targets of a TF from Arabidopsis and maize are functionally equivalent, as follows. Using Gateway™ technology, the *TARGET* system employs a vector with a RFP marker, for which any TF can be fused with a GR (the glucocorticoid receptor) tag, and successful transformants are isolated by Fluorescence-assisted cell-sorting. This 35S::TF-GR chimera allows one to i) overexpress the studied TF in the protoplasts using 35S, and to ii) control the TF entrance into the nucleus using a dexamethasone (DEX) treatment [Schena and Yamamoto 1988]. The TARGET system combined with RNA sequencing and/or ChIP-seq enables rapid and systematic assessment of TF function. This rapid system enables us to validate targets of TFs in Arabidopsis within 2 weeks. As part of this NSF Plant Genome grant, we will identify TF targets for Arabidopsis TFs. We will also work to adapt the *TARGET* system to maize protoplasts. This should be feasible, based on studies from the Sheen lab (MGH) in which both Arabidopsis and maize protoplasts are used in transient expression of signal transduction components (see Sheen, 2001). This will enable rapid cross-validation of our network predictions between Arabidopsis and maize. A big advantage of the TARGET system is that it can be adapted for a wide range of species – especially in crops where the generation of transgenic plant lines is either impossible or problematic and more time-consuming [Sheen, 2001], to enable rapid and systematic assessment of TF function in numerous plant species, e.g. important crop model species. Ideally, we would be able to perform cross-species validation studies, by expressing a maize TF in Arabidopsis and an Arabidopsis TF in maize. In these cases, a positive result (e.g shared targets) would be conclusive, while a negative result would be inconclusive (e.g. codon usage, or dependence on other factors). These cross-species validation studies performed in this rapid assay system could help prioritize TFs for translational studies *in planta*.

**Aim 4 Outcome**: Laboratory and field trials will confirm the NUE regulatory modules identified in Aim 3. Aim 4A confirms the major transcription factors that regulate genes primarily responsible for Nitrogen metabolism (Uptake and Remobilization) while Aim 4B validates that perturbations of the network modules leads to measurable changes in NUE of the plants under lab conditions. Aim 4C expands the validation to field conditions and in addition to testing the role of the modules themselves assays the absolute and relative expression levels of genes in the module and relates them to NUE under field conditions. This knowledge would allow breeders to create optimal combinations of alleles that might not exist in current germplasm.