Plant Physiology Preview. Published on June 4, 2015, as DOI:10.1104/pp.114.255877

Running head: **N-regulated network modules conserved across species.**

Correspondence should be addressed to:
Gloria M. Coruzzi
Center for Genomics and Systems Biology,
Department of Biology,
New York University
12 Waverly Place
New York, 10003 USA.
E-mail: gloria.coruzzi@nyu.edu
Tel 212-998-3963
Fax 212-995-4986

Paper submitted in the System Biology, Molecular Biology, and Gene Regulation research category of Plant Physiology.

Title:

Cross-species network analysis uncovers conserved nitrogen-regulated network modules in rice.

Authors: Mariana Obertello[1,2], Stuti Shrivastava[1], Manpreet Katari[1] and Gloria Coruzzi[1]

1. Center for Genomics and Systems Biology, Department of Biology, New York University, 12 Waverly Place, New York, 10003 USA.
2. Instituto de Ingeniería Genética y Biología Molecular (INGEBI-CONICET), Vuelta de Obligado 2490 Piso 2, C1428ADN Buenos Aires, Argentina.

One-sentence Summary

Integration of gene interaction data across a model dicot and a monocot identifies conserved and distinct regulatory network modules involved in nitrogen use, enabling translational discoveries from models to crops

The author responsible for distribution of Materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors is Gloria Coruzzi (gloria.coruzzi@nyu.edu).

Financial source

This work was supported by National Science Foundation (NSF) Arabidopsis 2010 Genome Grant MCB-0929338 to GC and a NSF Database Activities: DBI-0445666, "Conceptual Data Integration for the Virtual Plant" to GC and MK. MO was also supported by Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). SS was supported by NSF Integrative Organismal Systems award IOS-0922738, "Comparative Genomics of Seed Evolution".

Corresponding author

Gloria M. Coruzzi
Center for Genomics and Systems Biology,
Department of Biology,
New York University
12 Waverly Place
New York, 10003 USA.
E-mail: gloria.coruzzi@nyu.edu
Tel 212-998-3963
Fax 212-998-8210

Abstract

In this study, we used a cross-species network approach to uncover nitrogen-regulated network modules conserved across a model and a crop species. By translating gene "network knowledge" from the data-rich model Arabidopsis (*Arabidopsis thaliana*) to a crop (*Oryza sativa*), we identified evolutionarily conserved N-regulatory modules as targets for translational studies to improve N-use efficiency in transgenic plants. To uncover such conserved N-regulatory network modules, we first generated a N-regulatory network based solely on rice (*O. sativa*) transcriptome and gene interaction data. Next, we enhanced the "network knowledge" in the rice N-regulatory network using transcriptome and gene interaction data from Arabidopsis and new data from Arabidopsis and rice plants exposed to the same N-treatment conditions. This cross-species network analysis uncovered a set of N-regulated transcription factors (TFs) predicted to target the same genes and network modules in both species. Supernode analysis of the TFs and their targets in these conserved network modules uncovered genes directly related to nitrogen use (e.g. N-assimilation) and to other shared biological processes indirectly related to nitrogen. This cross-species network approach was validated with members of two TF families in the supernode network, bZIP-TGA and HRS1/HHO family, have recently been experimentally validated to mediate the N-response in Arabidopsis.

Introduction

The goal of this study is to translate "network knowledge" from Arabidopsis, a data-rich model species, to enhance the identification of nitrogen (N)-regulatory networks in rice, one of the most important crops in the world. With a significantly smaller genome size than other cereals (~430 Mb), the ability to perform genetic transformations (Hiei and Komari, 2008), and a finished genome sequence (Matsumoto T, 2005), rice is an excellent monocot model for genetic, molecular and genomic studies (Gale and Devos, 1998; Sasaki and Sederoff, 2003). In this study, we constructed N-regulatory gene networks in rice using "network knowledge" from Arabidopsis, a data-rich laboratory model for dicots. Thus, this cross-species network study exploits the best-characterized experimental models for dicot and monocot plants, respectively.

Nitrogen (N) is a rate-limiting element for plant growth. Rice plants absorb NH_4^+ at a higher rate than NO_3^- (Fried et al., 1965). Because NH_4^+ strongly inhibits NO_3^- uptake in agricultural soils where both NO_3^- and NH_4^+ are present (Kronzucker et al., 1999a), root NH_4^+ uptake may be favored as a result of the specific down-regulation of NO_3^- uptake systems (Kronzucker et al., 1999b). In rice, combinations of NO_3^- and NH_4^+ usually result in a greater vegetative growth than when either N form is supplied alone (Cramer and Lewis, 1993). Therefore, we designed our N-treatment experiments in this study to include both NO_3^- and NH_4^+.

In previous studies of the Arabidopsis N-response, we analyzed transcriptome data in the context of gene interactions to identify and validate N-regulated gene networks *in planta* (Gifford et al., 2008; Gutiérrez et al., 2008; Krouk et al., 2010). In this paper, we compare the N-regulated genes and gene networks between Arabidopsis and rice. This cross-species network analysis provides a unique opportunity to examine the conservation and divergence of N-regulated networks in the context of monocot and dicot transcriptomes. As rice and Arabidopsis are highly divergent phylogenetically, any evolutionarily conserved networks should be of special importance.

Establishing the architecture of gene regulatory networks requires gathering information on transcription factors (TFs), their targets in the genome, and their corresponding binding sites in gene promoter regions. Generation of N-responsive transcriptome data from rice and Arabidopsis enabled us to identify conserved N-regulatory gene network modules shared between dicots and

monocots. We analyzed the rice and Arabidopsis transcriptome (using Affymetrix GeneChips) in response to N-treatments in roots and shoots. The VirtualPlant software platform (Katari et al., 2010) which is operational for both Arabidopsis and rice, was used to perform much of the analysis including homology mapping analysis and significance of overlap in gene lists using the Genesect tool (www.virtualplant.org).

The N-regulated gene network includes expression data generated in this study and metabolic and protein-protein interactions from publicly available rice data (Rohila et al., 2006; Ding et al., 2009; Rohila et al., 2009; Gu et al., 2011; Dharmawardhana et al., 2013). Despite the fact that much of genomic and systemic rice data has been generated over the past years, a lot of information is still missing. For example, Arabidopsis has much more experimental data with regard to cis-binding sites and protein-protein interaction. To fill these gaps in rice "network knowledge", we integrated orthology-based Arabidopsis interaction data (Palaniswamy et al., 2006; Yilmaz et al., 2009; Gu et al., 2011; Ho et al., 2012) and searched for functional Arabidopsis cis-binding sites in rice, to identify N-regulatory network modules and biological processes ("network biomodules") conserved between dicots and monocots.

An important issue in this analysis is orthology. Monocots and dicots are quite distantly related with divergence estimation of 140-150 MYA (Chaw et al., 2004). A naïve and crude method for identifying putative orthologs, is to use Reverse Blast Hit thresholds – the putative orthologs must map to each other with a Blast e-value less than some cut-off. The identification of putative orthologs between monocots and dicots is confounded by the presence of paralogs (homologous genes originating from gene duplication events). There are several algorithms, such as OrthoMCL (Fischer et al., 2011), that are designed to help distinguish an ortholog from a paralog, by comparing sequences within species in addition to between species. However, even if these algorithms can detect true orthologs with greater specificity, there is always a possibility that different gene family members in each species take on the responsibility of responding to nutrients, like nitrogen. Here, we test and compare the performance of Reverse Blast Hit method and OrthoMCL in identifying genes and gene interactions whose *function* is conserved across species. From here on, the cross-species gene mapping based on BLASTP will be referred to as 'homologs', and the matches based on OrthoMCL will be called 'orthologs'.

Finally, this cross-species network study significantly contributes to two important areas: (i) studying N-regulated gene networks in rice, an important crop, and (ii) identifying conserved and distinct N-regulatory hubs controlling network "biomodules" which can be used to enhance translational discoveries between a model plant and crops. Our aim to identify N-regulated genes across a model dicot and a monocot crop, and to interpret it in a systems biology/network context, is essential to derive testable biological hypotheses. By applying network information, we can identify key regulators of these N-responsive gene networks and biomodules, which can be further manipulated to study N-use efficiency in transgenic plants. This approach has the potential to enhance translational discoveries from Arabidopsis to a crop (rice) with the goal of improving plant N-use efficiency, which will contribute to sustainable agricultural practices by diminishing the use of N fertilizers.

Results

Equilibrating nitrogen-treatment conditions for Arabidopsis and rice

The goal of this study is to identify conserved N-response networks in two species by comparison. Thus, in our study we made our N-treatments and growth conditions of rice and Arabidopsis as comparable as possible. We adapted our hydroponic system for Arabidopsis (Gifford et al., 2008) to grow and treat *O. sativa (*rice*)* seedlings, with only the plant roots submerged in liquid media. For plants with minimal seed reserves such as Arabidopsis, an external N-supply is required to allow plant growth and development. By contrast, rice can grow for longer periods using N-nutrients stored in their seeds. In order to equilibrate growth conditions of these two species, and to eliminate the seed-nutrient effect during N-treatment, the nutritive rice seed tissue was dissected away from the rice seedlings once the cotyledon and roots emerged, and only the germinated embryo was placed in the hydroponic system. For both species, the N-source during this initial growth phase contained 0.5 mM ammonium succinate, which was renewed every 2-3 days with fresh media to avoid NH_4^+ depletion due to different consuming rates between species. This growth on a low level of a N-source (ammonium), was a background in which to observe effects of transient treatments with nitrate (as in (Wang et al., 2000; Wang et al., 2004)) and/or high ammonium. As the N-regulation of gene expression is largely dependent on carbon (C) resource provision in Arabidopsis (Krouk et al., 2009), 0.5% (w/v) sucrose was included in the growth media as a constant nutrient to eliminate C-signaling effects during transient N-treatments. After 12 days, plants were N-starved for 24 h. Finally, at the start of their light cycle plants were N-treated for 2hr with a combination of NO_3^- (40 mM) and NH_4^+ (20 mM), the amount of N in MS media (Murashige and Skoog, 1962), referred here as 1xN (for more details see Materials and Methods). Shoot and root RNA samples were hybridized to the Arabidopsis ATH1 and Rice Genome Arrays from Affymetrix to evaluate changes in global gene expression (see Materials and Methods) in response to N-treatments. The normalized microarray data for each species has been deposited in the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE38102.

The effect of N-treatment on genome-wide expression in rice

Our first aim was to identify N-regulated genes and study their response in rice shoots and roots. Following RMA normalization, 2-way ANOVA analysis with FDR correction, and filtering of transcriptome data using 1.5 fold cut-off (described in Materials and Methods, Figure 1), we found a set of 451 genes in rice that were significantly regulated in rice by N-treatment (Table 1). In rice shoots, 103 genes were N-induced, and 39 genes were repressed in response to N-treatment. In rice roots, 234 genes were N-induced while 106 genes were repressed in N-treated samples, compared to control treatments. (Table 1; see Table S1 for a complete list of regulated genes and Figure S1 for organ specific gene response). Rice roots appear to have a much larger response in terms of number of genes, which has also been previously observed in Arabidopsis (Wang et al., 2003). Additionally, these results from the rice microarray data were confirmed by RT-*q*PCR for a number of selected genes (Figure S2).

The 451 N-regulated rice genes included genes involved in nitrate uptake and metabolism, sugar biosynthesis and ammonium assimilation among others (Table 2). Specifically, some of the genes in these groups are involved in producing reductants for nitrite uptake and also include enzymes of the pentose phosphate pathway, which generates the NADPH necessary for nitrogen assimilation (Table 2). We also observed N-induction of a gene that encodes the pentose-phosphate enzymes in both tissues: *G6PDH* (LOC_Os07g22350). Such genes involved in C-metabolism are related to the production of energy for nitrate or nitrite reduction. These types of genes have also been previously identified as N-responsive in Arabidopsis (Wang et al., 2003).

Finally, rice genes involved in ammonium assimilation were found to respond to N-treatments in our study (Table 2). NADH-GOGAT (LOC_Os01g48960) was N-induced in rice roots while *GLN* (LOC_Os04g56400) was found to be N-regulated (1.09 and 0.71 fold change respectively) in both roots and shoots (Table 2). The complete list of N-regulated genes in rice is shown in supplemental Table S1.

Genome-wide effects of nitrogen treatment in *Arabidopsis thaliana*

Arabidopsis seedlings were N-treated as described above for rice (for more details see Materials and Methods and Figure 1) and following RNA extraction, gene responses to N-treatments were analyzed using microarrays. Following normalization, 2-way ANOVA analysis, FDR correction and filtering for 1.5 fold change, 1,417 Arabidopsis genes were identified to be N-responsive compared to control treatment. In Arabidopsis shoots, 166 genes were N-induced and 184 genes were repressed in response to N-treatments. In Arabidopsis roots, 757 genes were N-induced and 424 genes were repressed (Table 1; for the complete list of regulated genes see Table S2). The N-regulated genes in Arabidopsis included genes involved in nitrate uptake and metabolism, genes in the Pentose Phosphate pathway and ammonium assimilation among others (Table 3).

As observed for rice, the majority of N-regulated genes in Arabidopsis are root-specific (also found previously (Wang et al., 2004)). For example, 75% of genes were uniquely N-regulated in Arabidopsis roots versus shoots, while only 16% of N-regulated genes were expressed exclusively in shoots (Figure S3). Several known Arabidopsis N-induced genes were also responsive to our treatments with ammonium nitrate, including: NIA1, NIA2, NIR, NRT2:1, NRT1:2, NRT3:1, ferredoxin 3, G6PD2, G6PD3, GLT1, ASN2 and GDH2 among others (Table 3, for a complete list see Table S2) (Wang et al., 2003; Krouk et al., 2010). Additionally, our microarray data was confirmed by RT-*q*PCR results in a number of selected Arabidopsis genes (Figure S4).

To determine whether the overlap between the rice and Arabidopsis N-responsive genes was significant, a permutation test was performed. 1,417 genes were selected randomly from Arabidopsis genes present on the Affy chip, and similarly 451 rice genes were selected randomly from genes present on the rice Affy chip. Using BLASTP homology, the overlap was measured in terms of rice and Arabidopsis genes. This was done 10,000 times and then the number of times the overlap was greater than or equal to the observed was counted. The overlap obtained from random sampling was never greater than or equal to the observed, making the *p*-value <0.0001. These results suggest that despite the difference in number of responsive genes, rice and Arabidopsis respond very similarly to the nitrogen treatments provided.

Network analysis identifies conserved genes involved in N-signaling in rice

It is known that the expression of many TFs is regulated by NO_3^-. However, to date, only a few of such NO_3^- regulated TFs have been shown to be involved in NO_3^- signaling in Arabidopsis (for review see (Castaings et al., 2011) and recent studies (Alvarez et al., 2014; Medici et al., 2015)).

Creation of a "Rice Arabidopsis N-regulatory Network" (RANN-Union).

To identify novel TFs that may play a global role in a N-regulatory network, we performed network analysis that exploited our microarray datasets from Arabidopsis and rice (Figure 2). We generated a network using the limited knowledge of known rice interactions and then, to enrich the existing network in rice, we introduced predicted interaction data based on homology to the large amount of Arabidopsis' "network knowledge". For this purpose, we started our network analysis by creating a "Rice Only N-response Network" (RONN) (Figure 2, Step 1). In Step 1, we used the rice experimental data generated in our study by looking at significant correlations among N-regulated rice genes (Pearson correlation coefficient with a *p*-value cut-off of 0.05), metabolic pathways from RiceCyc (Dharmawardhana et al., 2013), and experimentally determined protein-protein interactions in rice (Rohila et al., 2006; Ding et al., 2009; Rohila et al., 2009; Gu et al., 2011) for this network creation (for details see Materials and Methods). This "rice only" analysis resulted in a network of 451 N-regulated genes, with 36 TFs and 32,405 interactions among them (Figure 2, RONN).

Next, in Step 2 (Figure 2), predicted protein-protein interactions in rice and cis-binding site information from Arabidopsis were added to the RONN network. This generated a new predictive network: Rice Predicted N-regulatory Network (RPNN-predicted interactions) (for details see Materials and Methods). The RPNN-predicted interactions network included rice predicted regulatory interactions obtained from cis-binding site data in Arabidopsis, and transcription factor family information in rice from PlantTFDB (Jin et al., 2014) (for details see Materials and Methods). In the RPNN network, predicted regulatory edges are defined by the presence of a cis-binding site and a significant correlation between a transcription factor and target. In this analysis, 3,960 of the 32,225 correlation edges also contain cis-binding

information, thus re-categorizing them as regulatory edges. In the case where the target of one transcription factor (e.g. TF1) is another transcription factor (e.g. TF2), there is a possibility that TF1 is a target of TF2 (and vice versa), in which case one correlation edge between two TFs is converted to two regulatory edges. There are 168 such TF1-TF2 correlation edges, thus increasing the number of regulatory edges from 3,960 to 4,128 (Figure 2, RPNN). The RPNN-predicted interactions network had the same number of genes as the RONN network, however the addition of predicted protein-protein interactions along with regulatory data increases the total number of interactions to 32,839 in the RPNN-predicted interactions network (Figure 2).

Next, we were interested in further filtering the RPNN network to identify the N-regulatory genes and network modules whose regulation is conserved across two species, Arabidopsis and rice. To this end, in Step 3 (Figure 2), we introduced the Arabidopsis experimental data of N-responsive genes generated in our study into the RPNN-predicted interactions network. We approached this research question using two different orthology methods (BLASTP and OrthoMCL) to obtain two different Rice-Arabidopsis N-regulated Networks (RANN-BLAST and RANN-OrthoMCL, respectively). Both networks RANN-BLAST and RANN-OrthoMCL only contain rice genes where the rice gene and its putative ortholog in Arabidopsis is N-regulated in our experimental conditions. Additionally the correlation and regulatory edges between these conserved N-regulatory genes also had to be conserved (Figure 2; for details see Materials and Methods).

The RANN-BLAST network comprised 180 rice N-regulated genes, of which 23 are TFs. By contrast, the RANN-OrthoMCL network had only 48 rice N-regulated genes, of which 3 genes are TFs. It is not surprising that RANN-OrthoMCL network is smaller than RANN-BLAST, since OrthoMCL differentiates between orthologs and paralogs. It is important to note that out of 48 genes from RANN-OrthoMCL, only 2 additional genes were present uniquely in the RANN-OrthoMCL network and not in the RANN-BLAST network. These genes comprise a glycoprotein, LOC_Os10g41250 and a protein of unknown function, LOC_Os05g46340. As discussed below, we identified validated gene interactions using RANN-BLAST, which would have been missed had we only used RANN-OrthoMCL. Therefore, a union of the two conserved cross-species networks, RANN-BLAST and RANN-OrthoMCL, was performed to generate the

Rice-Arabidopsis N-regulatory Network (RANN-Union), which contains 182 rice N-regulated genes of which 23 genes are TFs (Figure 2, Step 4).

Of the 182 genes in the RANN-Union network (Figure 2, Step 4), some of the genes are known to be directly involved in N-assimilation; for example, nitrate transporters, nitrate and nitrite reductase, glutamine synthetase and glutamate synthase, among others (for the complete list of regulated genes see Table S3). The RANN-Union network also contains ferredoxin reductase genes (LOC_Os03g57120, LOC_Os05g37140 and LOC_Os01g64120) whose encoded proteins are indirectly involved in nitrite reduction by providing reducing power as shown in Arabidopsis (Wang et al., 2000). Additionally, LOC_Os03g57120 is orthologous to ATRFNR1 in Arabidopsis (At4g05390, based on BLASTP and OrthoMCL), which has also been shown previously to be involved in supplying reduced ferredoxin for nitrate assimilation (Hanke et al., 2005). In addition, two calcineurin B-like (CBL)-interacting protein kinases (CIPK) are present in the group of 182 N-regulated genes in the RANN-Union network. LOC_Os03g03510 has Arabidopsis CIPK23 as its ortholog (based on OrthoMCL and BLASTP), while, LOC_Os03g22050 is homolog to Arabidopsis CIPK23 only based on BLASTP (but not OrthoMCL). Interestingly, CIPK23 has been identified as NO_3^- inducible protein kinase (Castaings et al., 2011). Additionally, both rice CIPK loci (LOC_Os03g22050 and LOC_Os03g03510) are homologous to KIN11 and to MEKK1 (based on BLASTP but not OrthoMCL). KIN11, which is a Snf1-related kinase proposed to be part of an "energy-sensing" mechanism in Arabidopsis (Baena-González et al., 2007), and also found to be related to N-assimilation (Gutiérrez et al., 2008). Also, MEKK1 is involved in glutamate signaling in root tips of Arabidopsis (Forde, 2014). Moreover, LBD39 (LOC_Os03g41330) (Lateral Organ Boundary Domain), a transcription factor present in the RANN-Union, was found to be regulated at the transcriptional level by NO_3^- and involved in N-signaling in Arabidopsis (Rubin et al., 2009).

To study how TF connectivity changed throughout the network analysis, and to identify putative regulators that control the expression of conserved network modules, the transcription factors N-regulated in these networks were ranked based on their "hubbiness", the number of regulatory connections (Table 4). As mentioned previously, the number of connections found for TFs in the RPNN-predicted interactions (Step 2, Figure 2) decrease when the network was filtered with

Arabidopsis N-regulatory genes and their correlations (Step 3, Figure 2). The TF with the highest number of connections in the RANN-Union network is LOC_Os03g55590 (Table 4), a gene that belongs to the G2-like Transcription factor family, and sub-group HHO (for HRS1 Homolog). The HHO family has another member conserved in RANN-Union network, LOC_Os07g02800. A naïve assumption of the network analysis, is that the TF with the most connections has the most influential regulatory role. In previous studies, we have used the ranking of TF hubbiness to identify candidates for follow-up mutational studies in which they were validated (Gutierrez et al. 2008). To test the influence of orthology data, we investigated whether the rank of TFs based on hubbiness changed from the RPNN-predicted interactions network to the final network RANN-Union using the Wilcoxon test. A *p*-value of 1.423e-08 indicates that the connectivity rank presented in Table 4 has significantly changed through the network generation steps shown in Figure 2.

Creation of "Arabidopsis-Rice N-regulatory Network" (ARNN-Union).

Considering that there is more information available in Arabidopsis than in rice, we performed a similar network analysis as in Figure 2, but now using Arabidopsis N-regulated data as the starting point (Figure S7). We filtered the Arabidopsis network with rice experimental data generated in our study using BLASTP and OrthoMCL (see Supplemental Figure S7). The resulting Arabidopsis-Rice N-regulatory Network (ARNN-Union) has 276 genes. By definition, the identities of the genes from the Arabidopsis-Rice N-regulatory Network (ARNN-Union, 276 genes) (Figure S7) are equal to the Rice-Arabidopsis N-regulatory Network (RANN-Union, 182 genes) (Figure 2). The number of genes is different however, because in most of the cases rice genes have more than one N-regulated ortholog in Arabidopsis. Following this rationale, the ARNN-Union contains 76 TFs (Figure S7), while the RANN-Union contains only 23 TFs (Figure 2) (For a list, see Supplemental Table S4). We also studied how TF connectivity changed throughout the steps of the network analysis in Figure S7, by ranking TF's based on the number of regulatory connections (Table S4). In the top 5 highly ranked TFs of the ARNN-Union network (Table S4), we found 3 members of the HRS1/HHO family, and TGA1, which were each validated to be involved in the nitrogen response in Arabidopsis (Alvarez et al., 2010; Medici et al., 2015), in addition to WRKY28, a novel finding of our study. We also investigated

if the rank of TFs based on connectivity changed from the AONN network, to the final network ARNN-Union, again using a Wilcoxon test. A *p*-value of 1.391e-10 denotes that the connectivity rank of TFs (e.g. numbers of connections) in Supplemental Table S4 has changed significantly through the network generation process used in Figure S7.

Supernode analysis of Rice-Arabidopsis N-regulatory Network (RANN-Union).

The supernode analysis groups genes with the same biological processes, functional terms and annotations into a single node whose size is proportional to the number of genes in the supernode. To gain an understanding of how the conserved genes were connected to each other when categorized with plant metabolic network pathways information, a supernode network analysis was performed using transcription factor families (PlantTFDB, (Jin et al., 2014)) and OryzaCyc pathways associations (OryzaCyc v1.0 (Dharmawardhana et al., 2013) for the 182 genes in the RANN-Union network (Figure 2). The resulting supernode network of the RANN-Union network identified several well-represented transcription factor families highly connected to major metabolic pathways (Figure 3). The supernode network analysis also revealed that the transcription factor families with the highest number of members in this network are bZIP and WRKY.

The RANN-Union top transcription factor hubs include four members of the bZIP TF family in rice (LOC_Os05g37170, LOC_Os01g64020, LOC_Os06g41100 and LOC_Os01g64000). Homologs of these family members have been validated to be involved in N-responses in Arabidopsis (Gutiérrez et al., 2008; Hanson et al., 2008; Jonassen et al., 2009; Obertello et al., 2010; Para et al., 2014) (Figure 3 and Table 4). Three members of the bZIP TF family belong to the subfamily TGA, which has been recently indicated to be involved in nitrogen regulation (see below, Alvarez et al., 2010). The supernode network analysis also shows that the TF families: bZIP, bHLH, WRKY and G2-like (HHO) are involved in the N-regulation of genes related to "Nitrogen compound metabolism", which contains genes involved in the N assimilation pathway.

OsWRKY23, the second in the rank of most connected TFs in the RANN-Union network (Table 4), is homologous to Arabidopsis WRKY75 (At5g13080) based on BLASTP only, which has been shown to be related to phosphate acquisition (Devaiah et al., 2007). Also, OS-WRKY23 is

orthologous to Arabidopsis WRKY28 (At4g18170) based on BLASTP and OrthoMCL, which has been shown to be involved in activation of salicylic acid (SA) biosynthesis (van Verk et al., 2011).

Two predicted transcription factor families conserved in the Rice/Arabidopsis N-regulatory Network (RANN-Union) are biologically validated.

Among the list of 23 TFs present in the RANN-Union network, we found two TF families whose role in N-signaling has been experimentally validated. We first investigated the HHO/HRS1 family. This TF family has two N-regulated members in rice and four homologs in Arabidopsis (Supplemental Figure 5). To gain insights into the HHO/HRS family and their conserved N-regulation, we performed a phylogenetic analysis and found that the common N-responsive members of the HHO family from rice and Arabidopsis fall in the same clade (Supp Figure 5). The phylogenetic tree was built by ClustalW alignment and maximum likelihood method. This group of HHO family members present in the same clade is also orthologous to each other using either OrthoMCL or BLASTP (Supplemental Figure S5). This result is an *in-silico* validation of our cross-species network approach. Also, it has been recently validated that two members of this TF family, HRS1 and HHO1, have an important role in integrating nitrate signaling in the Arabidopsis root (Medici et al., 2015).

Based on supernode analysis, the bZIP family has 20 connections to biological processes making it the third most highly connected TF family in the RANN-Union network. The three N-regulated rice TGA family members (LOC_Os01g64020, LOC_Os05g37170 and LOC_Os06g41100) are putative homologs to the four N-regulated Arabidopsis TGA family members: At1g22070 (TGA3), At1g77920 (TGA7), At5g10030 (TGA4) and At5g65210 (TGA1) (Supplemental Figure S6). Based on our supernode network analysis, discussed above, these TFs have connections with “Biosynthesis” and “Degradation/ Utilization and Assimilation” metabolic pathway processes (Figure 3). We performed a phylogenetic tree analysis using all TGA family members in Arabidopsis and rice identified by BLASTP. The phylogenetic tree (Figure S6) shows that the rice and Arabidopsis N-regulated members of the TGA family are

paralogs, as confirmed by OrthoMCL. As shown in Supplemental Figure S6, all N-regulated TGA family members in each species were identified by homology based on BLASTP. However, it is important to point out that two of the members of the TGA transcription factor family identified in our RANN-BLAST network (TGA1 and TGA4) were recently validated as important regulatory components of the nitrate response in Arabidopsis (Alvarez et al., 2014). We also observed a significant overlap (*p*-value 0.008) between the validated targets identified *in-planta* in *tga1/4* double mutants, available data from Alvarez el al. 2014, and the predicted targets from our RANN-Union network analysis (analysis done using Genesect tool on VirtualPlant) (www.virtualplant.org). These TGA1/TGA4 targets identified in our analysis and validated *in planta* include two proteins that have been shown to be involved in N-signaling. These TGA1 targets include HRS1, a TF involved in N-signaling as mentioned earlier (Medici et al., 2015) and CIPK3, one of the several kinases identified to have a role in nitrogen signaling (Hu et al., 2009). The last gene present in this intersect set of validated HRS1 targets in the RANN network is a proteasome subunit, a potential gene hypothesis to be involved in nitrogen regulation (RPT5B), a potential new hypothesis for N-signaling via the proteasome that our analysis has uncovered. Thus, the conservation of function across rice and Arabidopsis implicated the role of TGA family in the N-response. It is noteworthy that this prediction, which is also supported by recent experimental data (Alvarez et al., 2014), would have been missed if we relied only on orthology based on OrthoMCL. Importantly, our cross species network analysis has also opened new hypotheses for testing about N-regulatory mechanisms in plants.

Discussion

This study provides a novel analysis of N-regulated gene networks conserved across two highly divergent species: *O. sativa* (a monocot) and Arabidopsis (a dicot). Despite their large phylogenetic distance, our analysis revealed a set of N-regulated genes, TFs and network modules conserved in rice and Arabidopsis, exposed to the same N-treatment conditions. Our analysis shows a statistically significant overlap, indicating that rice and Arabidopsis respond very similarly to the N-treatments. The list of genes regulated by nitrogen treatments in rice includes many of the known nitrate/ammonium regulated genes previously identified in Arabidopsis, including, genes known to respond to nitrate (NR, NiR, Fd, FNR, G6PDH). These results are not surprising in hindsight, given that the former are important to reduce the plant's risk of nitrite toxicity. Selected genes from the N-responsive lists were corroborated by RT-*q*PCR analysis. One of the important aspects of this genomic analysis is that the N-treatment performed on rice and Arabidopsis were comparable, so that the gene responses could be directly compared. Genome profiling revealed that 1.32% of the rice genome is regulated in response to N-treatment, while 6.76% of the Arabidopsis genome responds to N-treatment, and in both cases, roots were more sensitive to N than shoots. The result of the permutation test, which was performed to determine whether the overlap between the rice and Arabidopsis N-responsive genes was significant, suggests that despite the difference in number of N-responsive genes, rice and Arabidopsis respond very similarly to nitrogen treatment.

The rice genome size is more than three times that of Arabidopsis, and is estimated to have significantly more genes (Yu et al., 2005). According to that estimate, we would have expected more N-regulated genes in rice; however, the difference in total number of N-regulated genes between species might be mainly due to the fact that the N treatment used here affects these two plants differently. In support of that notion, it has long been known that rice can form natural associations with endophytic diazotrophs, which are responsible for supplying the plants with fixed N, increasing plant height, root length and dry-matter production. In rice and maize, associative nitrogen fixation can supply 20–25% of total N requirements (Santi et al., 2013). The experiments performed here were done on a sterile environment, so the difference in number of

N-regulated genes might be due to the fact that N-response pathway in rice needs the bacterial association to be completely active.

The N-signaling network has gained new levels of complexity during very recent years and is as yet far from being completely understood (Vidal et al., 2010; Castaings et al., 2011; Bargmann et al., 2013; Medici et al., 2015). In addition, it is an open question how well gene networks derived from model dicots, such as Arabidopsis, might faithfully reconstruct pathways in a monocot, such as rice.

Our hypothesis was that the conserved network nodes (genes) and edges (interactions) among species would provide an initial framework to understand the complex functional genomic and genetic knowledge of N-regulatory networks. To address this, we generated a gene expression network based on co-expression and homologs based on BLASTP and orthologs based on OrthoMCL to reveal conserved co-expression relationships between rice and Arabidopsis. Our results suggest that using BLASTP homology produced a more complete core N-regulatory network between rice and Arabidopsis compared to OrthoMCL alone. When we use OrthoMCL to distinguish between orthologs and paralogs, we lose promising candidates from the network. For example, if we only used OrthoMCL to obtain orthology information, we would have missed the TGA family members and their interaction to regulate N-responsive biological processes. From the phylogenetic analysis, it is clear that the TGA family members evolved in their function so much that different members of the family have taken on the responsibility to be N-responsive in each species. Since it is well accepted that different members of the TF family bind to the same binding site, this hypothesis is quite reasonable. As described in the results section, our predicted TGA1 and TGA4 target genes from the RANN-Union network overlap significantly with published and biologically validated *in planta* data in Arabidopsis (Alvarez et al., 2014).

In this cross-species network approach, we used known rice annotation and experimental data to generate a “rice-only” expression network (RONN, Step 1, Figure 2), to which we added known Arabidopsis annotation data (Step 2, Figure 2), and subsequently filtered it with our Arabidopsis N-treatment experimental data generated in this study (Step 3, Figure 2). This analysis identified a core N-regulatory network conserved between rice and Arabidopsis (RANN-Union). This

cross-species network analysis enabled us to identify conserved N-regulated genes, network modules, TFs and biological process related to this essential nutrient. The list of potential N-responsive genes in rice is considerably reduced when we integrated our experimental data from Arabidopsis (Step 3, Figure 2). In addition, the supernode network analysis allowed us to visualize how N-responsive biological processes such as, "nitrogen compound metabolism" and "sugar biosynthesis", are related to each other and which transcription factor families are regulating them. The presence of metabolic pathways related to sugar metabolism and amino acid biosynthesis is important in this context since the production of reduced carbon is necessary to produce both the energy and carbon skeletons required for the incorporation of inorganic N into amino acids.

By starting with the experimental data from the model plant Arabidopsis, and subsequently filtering it with our rice experimental data generated in this study, we uncovered a subset of conserved TFs potentially involved in nitrogen regulation. However, compared to the N-regulated network information already known in Arabidopsis, we conclude that while we did not significantly improve our knowledge of Arabidopsis interactions by integrating rice data, we did identify a smaller evolutionarily conserved network. On the other hand, when we started with rice experimental data and then add predicted 'network knowledge' inferred from Arabidopsis, subsequently introducing Arabidopsis experimental data, we significantly improve our network connections and identified TF-target connections that have been experimentally validated in Arabidopsis. To summarize, using Arabidopsis "network knowledge" including gene interactions and experimental data highly refined our rice networks, enabled us to identify potential master TFs involved in the N-response, some of which have been biologically validated in Arabidopsis by independent experiments (e.g. members of the TGA and HHO transcription family members).

In plants, transcriptional regulation is mediated by a large number of transcription factors (TFs) controlling the expression of tens or hundreds of target genes in various signal transduction cascades. Interestingly, a recent transcriptome data analysis supports our predictions for the TFs controlling this core N-regulatory network uncovered in our analysis. Specifically, Canales et al. integrated publicly available root microarray data under contrasting nitrate conditions, and concluded that the most represented transcription factors families are AP2/ERF, MYB, bZIP and

bHLH (Canales et al., 2014). In our study, the TFs regulated by N-treatment were ordered by their network connectivity, under the premise that highly connected genes are more likely to be involved in biological processes. These transcription factor families are also present in our supernode analysis based on the Rice-Arabidopsis N-regulatory Network (RANN-Union). Additionally, our supernode analysis also revealed the G2-like (HHO) family in rice -based on orthology to Arabidopsis- as one of the most highly-connected TF families. In addition, there is recent experimental validation of several members of the HHO family being involved in the N-response in Arabidopsis (Medici et al., 2015). Another highly connected TF family obtained from the supernode analysis was the TGA family, three members of which were N-regulated and conserved in our RANN-BLAST network, but not in the RANN-OrthMCL network. With these results, we conclude that it is important to consider homologs based on BLASTP for retrieval of conserved network modules. We further validated the RANN-Union network by determining that our predicted targets of TGA1/4 significantly overlap (*p*-val 0.008) with validated targets identified *in planta* in *tga1/4* double mutants (Alvarez et al., 2014). Thus, this novel finding of transcription factors implicated in N-regulation of genes and network modules, conserved in both rice and Arabidopsis according to our predicted network, are strongly supported by the experimental study of *tga1* and *tga4* mutants (Alvarez et al., 2014).

Finally, our study addresses a major challenge of translational research, which is to transfer "network knowledge" from data-rich model species, such as Arabidopsis, to data poor crop species, such as rice. The results presented here describe the transfer of "network knowledge" from Arabidopsis to crops (e.g. Steps 2 and 3 of Figure 2), and how it can help develop effective and sustainable biotechnological solutions to enhance N acquisition by plants in natural or agricultural environments. Proper plant N nutrition in the environment will not only improve production but will also contribute to sustainable agricultural practices by diminishing the use of N fertilizers and thus reducing greenhouse gases, stratospheric ozone, acid rain, and nitrate pollution of surface and ground water.

Materials and Methods

Plant growth and treatment conditions

Rice seeds (*Oryza sativa* ssp. *japonica*) were kindly provided by Dale Bumpers of the National Rice Research Center (AR, USA). Seeds were surface-sterilized in 70% ethanol for 3 minutes followed by commercial H_2O_2 for 30 minutes with gently agitation, and washed with distilled water. Seeds were sown onto 1x Murashige and Skoog basal salts (custom-made; GIBCO) with 0.5 mM ammonium succinate and 3 mM sucrose, 0.8% BactoAgar at pH 5.5 for 3 days in dark conditions at 27°C. Following germination, embryos with developed root system and aerial tissue were dissected from the rest of the seed using a sterile blade and transferred to a hydroponic system (Phytatray II, Sigma Aldrich) containing basal MS salts (custom-made; GIBCO) with 0.5 mM ammonium succinate and 3 mM sucrose at pH 5.5. Fresh media was replaced every 3 days to maintain a steady nutritional state and optimal pH levels. After 12 days under long-day (16 h light: 8 h dark) growth conditions, at light intensity of 180 $\mu E.s^{-1}.m^{-2}$ and at 27°C, plants were transferred to fresh media containing only custom basal MS salts for 24 h prior treatment. On day 13, plants were transiently treated for 2 h at the start of their light cycle by adding nitrogen (N) at a final concentration of 20 mM KNO_3 and 20 mM NH_4NO_3 (referred here as 1xN). Control plants were treated with KCl at a final concentration of 20 mM. After treatment, roots and shoots were harvested separately using a blade, and immediately submerged into liquid nitrogen and stored at -80°C prior to RNA extraction.

Arabidopsis seeds were placed for 2 days in the dark at 4°C to synchronize germination. Seeds were surface-sterilized and then transferred to a hydroponic system (Phytatray I, Sigma Aldrich) containing the same media previously described for rice (pH 5.7). Growth conditions were the same as in rice, except that plants were under 50 $\mu E.s^{-1}.m^{-2}$ light intensity at 22°C. N-starvation and treatments were done as described above.

RNA isolation and RT-qPCR analysis

RNA was isolated from roots and shoots with the TRIzol reagent following manufacturer´s protocols (Invitrogen Life Technologies. Carlsbad, CA, USA). Standard manufacturer's

protocols were used to reverse-transcribe total RNA (1 to 2µg) to one-strand cDNA using Thermo™ script RT (Invitrogen). RT-PCR measurements were obtained for a set of selected genes using gene-specific primers (Table S5) and LightCycler FastStart DNA Master SYBR Green (Roche Diagnostics). Expression levels of tested genes were normalized to expression levels of the actin or clathrin gene as described in (Obertello et al., 2010).

Microarray experiments and analysis

- cDNA synthesis, array hybridization and normalization of the signal intensities were performed according to the instructions provided by Affymetrix. Affymetrix Arabidopsis ATH1 Genome Array and Rice Genome Array were used for respective species. The Affymetrix microarray expression data has been deposited in the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE38102.

- Gene expression values were transformed by taking the logarithm to the base 2 ($\log_2$) of the ratio of 1xN-treatment (experimental state) over KCl treatment (control state) to yield the magnitude of the deviations in up- and down-regulated genes symmetrically ($\log_2$ value of the ratio of 1-fold is 0). Data normalization was performed using the RMA (Robust Microarray Analysis) method in the Bioconductor package in R statistical environment.

- A two-way Analysis of Variance (ANOVA) was performed using a custom-made function in R to identify probes that were differentially expressed following N treatment. The *p*-values for the model were then corrected for multiple hypotheses testing using FDR correction at 5% (Benjamini and Hochberg, 1995). The probes passing the cut-off ($p \leq 0.05$) for the model and, N treatment or interaction of N treatment and tissue, were deemed significant. A Tukey's HSD post-hoc analysis was performed on significant probes to determine the tissue specificity of N-regulation at *p*-value cut-off ≤ 0.05 and |fold-change| ≥ 1.5-fold ($\log_2$ of 1.5 is 0.585). Probes mapping to more than one gene were disregarded. Finally, for the cases of multiple probe sets representing the same gene, the assumption was that the expression levels should be upregulated or down-regulated in all the probes representing the gene. Expression levels were combined for those that passed the criterion. A set of 451 N-regulated genes

differentially expressed in Rice and 1,417 N-regulated genes differentially expressed in Arabidopsis were obtained.

- For both species, Pearson correlation coefficient was calculated for probes that passed the 2-way ANOVA and FDR correction. Specifically, the Pearson correlation coefficient was computed between different pair of probe sets using the mean value of their expression data across the replicates using a custom script in R. Correlation was calculated separately for root genes and shoot genes in both species and the corresponding correlation edge was labeled accordingly.

Orthology Analysis

Sequence and annotation data for the *Oryza sativa* ssp. *japonica* genome was downloaded from the TIGR Rice Genome Annotation Database, version 6.1 (http://rice.plantbiology.msu.edu/). Similarly, data for the *Arabidopsis thaliana* genome was obtained from The Arabidopsis Information Resource (TAIR) website, version 10 (Lamesch et al., 2012). Homologous N-regulated genes between Rice and Arabidopsis were obtained using Reverse BLAST (Camacho et al., 2009) with an e-value $\leq$ 1e-20, thereby allowing for multiple orthologous gene hits. Orthology was determined using the data provided on the OrthoMCL website (Fischer et al., 2011).

Network analysis – data retreival

For the gene network analysis (Figure 2), rice network interaction data was obtained as follows:

- For Rice Only N-response Network (RONN) (Figure 2, Step 1), metabolic interactions were obtained from RiceCyc, Gramene Pathways (Dharmawardhana et al., 2013) and experimentally determined protein-protein interactions were obtained from the PRIN database (Gu et al., 2011) and Rice Kinase database (Rohila et al., 2006; Ding et al., 2009; Rohila et al., 2009).

- For Rice Predicted N-regulatory Network (RPNN-predicted interactions) (Figure 2, Step 2), computationally predicted protein-protein interactions were obtained from the PRIN database (Gu et al., 2011), and the Rice Journal database (Ho et al., 2012).

  Additionally for RPNN-predicted interactions (Figure 2, Step 2), regulatory interactions were predicted between a TF and its putative target. TF family membership in Rice was obtained from PlantTFDB (Jin et al., 2014) and cis-regulatory motifs were obtained from AGRIS (Palaniswamy et al., 2006). The upstream promoter sequences (1kb) in rice were retrieved from the RAP-DB (http://rapdb.dna.affrc.go.jp/). Cis-motifs in promoter regions were searched using the DNA pattern matching tool from the RSA tools – Plants server with default parameters (van Helden, 2003). HRS1-HHO family member targets were predicted similarly and cis-motifs for the TF family members were obtained from Medici et al. (Medici et al., 2015).

- For the Rice-Arabidopsis N-regulatory Network using BLASTP (RANN-BLAST) and Rice-Arabidopsis N-regulatory Network using OrthoMCL (RANN-OrthoMCL) (Figure 2, Step 3), a correlation edge was considered as a 'conserved correlation edge' when the correlation between N-regulated gene pair in rice was supported by a significant correlation edge between its respective Arabidopsis N-regulated orthologous gene pair, with correct directionality (both correlation edges (in each species) were either both positive or both negative) and tissue-specificity (both correlation edges (in each species) were either both root correlation edge or both shoot correlation edge).

Network construction

In Step 1 (Figure 2), the 451 rice N-regulated genes were queried against the metabolic and experimentally determined protein-protein interaction databases, and all the significant correlation edges between them ($p \leq 0.05$) were used to generate RONN. Querying against the predicted protein-protein interactions databases in Step 2 (Figure 2) further enriched this network. Additionally, the predicted regulatory interactions, obtained using cis-motifs from Arabidopsis, were restricted to those TF:target gene pairs where the two were also significantly

correlated ($p \leq 0.05$). The resulting network, RPNN-predicted for Step 2 (Figure 2) had 451 rice genes with 36 TFs, and a total of 32,839 interactions between them.

The RPNN-predicted interactions network has reduced number of correlation-only edges compared to RONN because adding cis-motif information to the network resulted in some of the correlation-only edges to be reassigned as regulatory edges. This also increased the total number of regulatory (4,128) edges and correlation-only (28,265) edges in the network to 32,393 edges from 32,225 correlation-only edges (Figure 2). The 168 additional edges were a result of added directionality of regulation, accounting for cases where one TF (TF1) was targeting and was being targeted by another TF (TF2) in the network (Figure 2).

In Step 3 (Figure 2), Arabidopsis N-regulated experimental correlation data was introduced using BLASTP and OrthoMCL and individual networks were generated for each method following a similar workflow. Briefly, in both methods the rice experimental correlation data was filtered with Arabidopsis correlation data, inferred in rice using orthology, to yield conserved correlation edges. If the significant correlation edge between N-regulated gene pair in rice was also supported by a significant correlation edge between its respective Arabidopsis N-regulated orthologous gene pair, then it was considered a 'conserved correlation' edge. The resulting networks for Step 3 (Figure 2), RANN-BLAST and RANN-OrthoMCL comprised a total of 180 N-regulated rice genes with 2,212 total interactions, and 48 N-regulated rice genes with 383 total interactions, respectively.

Finally, the two networks RANN-BLAST and RANN-OrthoMCL were merged in Step 4 to yield the RANN-Union network, which had 182 N-regulated rice genes and 2,273 total interactions between them.

Network visualization and analysis

All network visualizations were created using Cytoscape (v2.8.3) software (Shannon et al., 2003). Custom-made script was used to analyze the total number of direct targets for a TF for each of the regulatory network. The summarized result for the analysis across all networks is presented in Table 4. The Wilcoxon signed-rank test was used in R to validate that the change in

number of direct targets for the TFs is significant across the network generation process (Hollander et al., 2014).

Supernode Network Analysis

The supernode analysis merges the individual nodes (genes) into a single node, its size proportional to the number of nodes merged, based on the classification system selected. The transcription factor families (Plant TFDB, Jin et al., 2014) and PlantCyc (OryzaCyc v1.0, PMN) pathways were the two major classification groupings used for our purposes, with level-3 subclass hierarchical classification (Figure 3). The individual gene pair interactions were merged appropriately for the supernodes and were similar interaction types as present in the gene network analysis.

Phylogenetic analysis

The sequences coding for G2-like (HHO) and TGA proteins were retrieved from the AGRIS (Arabidopsis Gene Regulatory Information Server; http://arabidopsis.med.ohio-state.edu/) database and from the Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/). The alignment of the full-length amino acid sequences was performed in ClustalW using standard settings. The phylogeny reconstruction was inferred by using the maximum likelihood method. The bootstrap values were obtained based on 500 replicates. Phylogenetic analysis was conducted in MEGA5 software (Tamura et al., 2011).

Figures

Figure 1. A schematic diagram of the experimental and data mining approach used in this study. Briefly, *O. sativa* (rice) and *A. thaliana* plants were grown for 12 days before a 2 hr treatment with 1xN vs. KCl control (see Methods). Genome-wide analysis using Affymetrix chips has been used in order to quantify mRNA levels. Modeling of microarray data, using ANOVA, homology/orthology and network analysis (detailed in Methods), were used to identify a core translational N-regulatory network shared between rice and Arabidopsis.

Table 1. Number of nitrogen regulated genes in *O. sativa* and *A. thaliana*. Percentage of regulated genes for each type of regulation is in parentheses.

Table 2. Selected rice genes regulated by nitrogen in shoots and roots (for more details see Materials and Methods). See Suppl. Table S1 for full list of genes. The fold change of nitrogen response genes were calculated as the ratio between N / KCl expression value. *p*-value cut-off ≤ 0.05 and fold-change $\geq$ 1.5-fold (shown on table is the $\log_2$ of values, fold-change cut-off $\log_2 1.5 = 0.585$). NC, no change.

Table 3. Selected Arabidopsis genes regulated by nitrogen in shoot and/or roots (for more details see Materials and Methods). The fold change of nitrogen response genes were calculated as the ratio between N / KCl expression value. *p*-value cut-off ≤ 0.05 and fold-change $\geq$ 1.5-fold (shown on table is the $\log_2$ of values, fold-change cut-off $\log_2 1.5 = 0.585$). NC, no change.

Figure 2. The workflow of the network analysis of N-regulated genes differentially expressed in rice resulting in "Rice-Arabidopsis N-regulatory Network (RANN-Union)". The input was 451 rice N-regulated genes. In each of the three steps, we introduced rice and Arabidopsis data in order to identify the RANN-Union network, which includes N-regulated genes and network modules conserved between rice and Arabidopsis (for more details see Materials and Methods).

Table 4. List of the transcription factors in the "Rice-Arabidopsis N-regulatory Network (RANN-Union)". For each step of the network construction (Figure 2), transcription factors were rank based on their number of connections in the network.

Figure 3. Supernode network analysis created from the 182 genes of "Rice-Arabidopsis N-regulatory Network" (RANN-Union). Individual nodes were clustered based on PlantCyc pathways and TF families classification to form supernodes. Genes which do not belong to either of the two classifications are not shown here. Triangles represent TFs families and squares represent PlantCyc pathways (Zhang et al., 2010). The size of the nodes is proportional to the number of genes within that particular category (from 1 to 5). Nodes are connected by TF:target (red dashed lines = predicted negative correlation; green dashed lines = predicted positive correlation) and predicted protein-protein interactions (blue dashed lines). All nodes are present in the "Rice-Arabidopsis N-regulatory Network" (RANN-BLAST) supernode network. Nodes

circled in thick grey lines are also present in the "Rice-Arabidopsis N-regulatory Network" (RANN-OrthoMCL) supernode network.

Supplemental Material

Table S1. Genes regulated by nitrogen in rice shoots and roots are sorted based on their regulation according to the ANOVA analysis (*p*val<0.05).

Figure S1. Rice N-regulated gene lists compared using the Sungear tool (Poultney et al., 2007) housed in Virtual Plant (www.virtualplant.org). The polygon shows the four lists of N-regulated genes at the vertices. The circles inside the polygon (vessels) represent the list of genes that are shared by the anchors (gene lists), as indicated by the arrows around the vessels with the number of shared genes in parenthesis. The area of each vessel is proportional the number of genes associated with that vessel.

Figure S2. Quantification of mRNA levels of *O. sativa* N-regulated genes. Transcript levels were determined by RT-qPCR and are shown as relative to expression of a housekeeping rice actin gene (LOC_Os10g36650). Values are the mean ±SE from three biological replicates. Asterisks indicate significant differences between control (N-) and treatment (N+) for each tissue according to ANOVA analysis ($p<0.05$).

Table S2. Genes regulated by nitrogen in shoots and roots of Arabidopsis are sorted based on their regulation according to the ANOVA analysis after FDR correction ($p<0.05$).

Figure S3. Arabidopsis N-regulated gene lists compared using the Sungear tool (Poultney et al., 2007) housed in Virtual Plant (www.virtualplant.org). The polygon shows the four lists of N-regulated genes at the vertices. The circles inside the polygon (vessels) represent the list of genes that are shared by the anchors (gene lists), as indicated by the arrows around the vessels with the number of shared genes in parenthesis. The area of each vessel is proportional the number of genes associated with that vessel.

Figure S4. Quantification of mRNA levels of *A. thaliana* N-regulated genes. Transcript levels were determined by RT-qPCR and are shown as relative to expression of a housekeeping Clathrin gene (At4g24550). Values are the mean ±SE from three biological replicates. Asterisks indicate significant differences between control (N-) and treatment (N+) for each tissue according to ANOVA analysis ($p<0.05$).

Figure S5. Arabidopsis and rice HRS1/HHO transcription factor family phylogenetic tree built by ClustalW alignment and maximum likelihood method. The bootstrap values displayed were calculated based on 500 replications (MEGA6). N-regulated genes are indicated under the shaded rectangles (red for rice genes and blue for Arabidopsis genes). Genes identified as homologs or orthologs based on BLASTP or OrthoMCL respectively, are indicated with a check mark.

Figure S6. Arabidopsis and rice TGA transcription factor family phylogenetic tree built by ClustalW alignment and maximum likelihood method. The bootstrap values displayed were calculated based on 500 replications (MEGA6). N-regulated genes are indicated by the shaded rectangles (red for rice genes and blue for Arabidopsis genes). Genes identified as homologs or orthologs based on BLASTP or OrthoMCL, respectively are indicated with a check mark.

Figure S7. The workflow of the analysis of N-regulated genes differentially expressed in rice resulting in "Arabidopsis-Rice N-regulatory Network (ARNN-Union)". The input was 1417 Arabidopsis N-regulated genes. In each of the three steps shown in Figure S7, we introduced rice and Arabidopsis data in order to identify the Arabidopsis core translational network, which includes N-regulated genes and network modules conserved between rice and Arabidopsis (for more details see Materials and Methods).

Table S3. List of the 182 genes in the "Rice-Arabidopsis N-regulatory Network" (RANN-Union).

Table S4. List of the transcription factors in the "Arabidopsis-Rice N-regulatory Network (ARNN-Union)" from Figure S7. For each step of the rice core translational network, transcription factors were rank based on their number of connections.

Table S5. Quantitative real-time PCR primers used in this study.

Acknowledgments

The authors thank Dan Tranchina for assistance in statistical analysis of microarray data in the early stages of this paper.

References

Alvarez J, Riveras E, Aceituno F, Tamayo K, Gutierrez R (2010) TGA1 and TGA4 transcription factors control nitrate responses in Arabidopsis thaliana root organs. 21st International Conference on Arabidopsis research.

Alvarez JM, Riveras E, Vidal E a, Gras DE, Contreras-López O, Tamayo KP, Aceituno F, Gómez I, Ruffel S, Lejay L, et al (2014) Systems approach identifies TGA1 and TGA4 transcription factors as important regulatory components of the nitrate response of Arabidopsis thaliana roots. The Plant journal **80**: 1–13

Baena-González E, Rolland F, Thevelein JM, Sheen J (2007) A central integrator of transcription networks in plant stress and energy signalling. Nature **448**: 938–42

Bargmann BOR, Marshall-Colon A, Efroni I, Ruffel S, Birnbaum KD, Coruzzi GM, Krouk G (2013) TARGET: A Transient Transformation System for Genome-Wide Transcription Factor Target Discovery. Molecular plant **6**: 978–80

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society **57**: 289–300

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. BMC bioinformatics **10**: 421

Canales J, Moyano TC, Villarroel E, Gutiérrez RA (2014) Systems analysis of transcriptome data provides new hypotheses about Arabidopsis root response to nitrate treatments. Frontiers in plant science **5**: 22

Castaings L, Marchive C, Meyer C, Krapp A (2011) Nitrogen signalling in Arabidopsis: how to obtain insights into a complex signalling network. Journal of experimental botany **62**: 1391–7

Chaw S, Chang C, Chen H, Li W (2004) Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. Journal of Molecular Evolution **58**: 424–41

Cramer M., Lewis OA. (1993) The influence of nitrate and ammonium nutrition on the growth of wheat (Triticum aestivum) and maize (Zea mays) plants. Annals of Botany **72**: 359–365

Devaiah BN, Karthikeyan AS, Raghothama KG (2007) WRKY75 transcription factor is a modulator of phosphate acquisition and root development in Arabidopsis. Plant physiology **143**: 1789–801

Dharmawardhana P, Ren L, Amarasinghe V, Monaco M, Thomason J, Ravenscroft D, McCouch S, Ware D, Jaiswal P (2013) A genome scale metabolic network for rice and accompanying analysis of tryptophan, auxin and serotonin biosynthesis regulation under biotic stress. Rice **6**: 15

Ding X, Richter T, Chen M, Fujii H, Seo YS, Xie M, Zheng X, Kanrar S, Stevenson RA, Dardick C, et al (2009) A rice kinase-protein interaction map. Plant physiology **149**: 1478–92

Fischer S, Brunk B, Chen F, Gao X, Harb O, Iodice J, Shanmugam D, Roos D, Stoeckert C (2011) Using OrthoMCL to Assign Proteins to OrthoMCL-DB Groups or to Cluster Proteomes Into New Ortholog Groups. Current Protocols in Bioinformatics 35:

Forde BG (2014) Glutamate signalling in roots. Journal of experimental botany **65**: 779–87

Fried M, Zsoldos F, Vose PB, Shatokhin IL (1965) Characterizing the NO3 and NH4 Uptake Process of Rice Roots by Use of 15N Labelled NH4NO3. Physiologia Plantarum **18**: 313–321

Gale MD, Devos KM (1998) Plant Comparative Genetics after 10 Years. Science **282**: 656–659

Gifford ML, Dean A, Gutierrez R a, Coruzzi GM, Birnbaum KD (2008) Cell-specific nitrogen responses mediate developmental plasticity. Proceedings of the National Academy of Sciences of the United States of America **105**: 803–8

Gu H, Zhu P, Jiao Y, Meng Y, Chen M (2011) PRIN: a predicted rice interactome network. BMC bioinformatics **12**: 161

Gutiérrez R a, Stokes TL, Thum K, Xu X, Obertello M, Katari MS, Tanurdzic M, Dean A, Nero DC, McClung CR, et al (2008) Systems approach identifies an organic nitrogen-responsive gene network that is regulated by the master clock control gene CCA1. Proceedings of the National Academy of Sciences of the United States of America **105**: 4939–44

Hanke GUYT, Okutani S, Satomi Y, Takao T, Suzuki A (2005) Multiple iso-proteins of FNR in Arabidopsis : evidence for different contributions to chloroplast function and nitrogen. 1146–1157

Hanson J, Hanssen M, Wiese A, Hendriks MMWB, Smeekens S (2008) The sucrose regulated transcription factor bZIP11 affects amino acid metabolism by regulating the expression of ASPARAGINE SYNTHETASE1 and PROLINE DEHYDROGENASE2. The Plant journal: for cell and molecular biology **53**: 935–49

Van Helden J (2003) Regulatory Sequence Analysis Tools. Nucleic Acids Research **31**: 3593–3596

Hiei Y, Komari T (2008) Agrobacterium-mediated transformation of rice using immature embryos or calli induced from mature seed. Nature protocols **3**: 824–34

Ho C, Wu Y, Shen H, Provart NJ, Geisler M (2012) A predicted protein interactome for rice. Rice **5**: 1–14

Hollander M, Wolfe D, E C (2014) Nonparametric Statistical Methods.

Hu H-C, Wang Y-Y, Tsay Y-F (2009) AtCIPK8, a CBL-interacting protein kinase, regulates the low-affinity phase of the primary nitrate response. The Plant journal : for cell and molecular biology **57**: 264–78

Jin J, Zhang H, Kong L, Gao G, Luo J (2014) PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. Nucleic Acids Research **42**: D1182–D1187

Jonassen EM, Sévin DC, Lillo C (2009) The bZIP transcription factors HY5 and HYH are positive regulators of the main nitrate reductase gene in Arabidopsis leaves, NIA2, but negative regulators of the nitrate uptake gene NRT1.1. Journal of plant physiology **166**: 2071–6

Katari MS, Nowicki SD, Aceituno FF, Nero D, Kelfer J, Thompson LP, Cabello JM, Davidson RS, Goldberg AP, Shasha DE, et al (2010) VirtualPlant: a software platform to support systems biology research. Plant physiology **152**: 500–15

Kronzucker H, Glass A, Yaeesh Siddiqi M (1999a) Inhibition of nitrate uptake by ammonium in barley. Analysis Of component fluxes. Plant physiology **120**: 283–92

Kronzucker H, Siddiqi M, Glass A, Kirk G (1999b) Nitrate-ammonium synergism in rice. A subcellular flux analysis. Plant physiology **119**: 1041–6

Krouk G, Mirowski P, LeCun Y, Shasha DE, Coruzzi GM (2010) Predictive network modeling of the high-resolution dynamic plant transcriptome in response to nitrate. Genome biology **11**: R123

Krouk G, Tranchina D, Lejay L, Cruikshank A a, Shasha D, Coruzzi GM, Gutiérrez R a (2009) A systems approach uncovers restrictions for signal interactions regulating genome-wide responses to nutritional cues in Arabidopsis. PLoS computational biology **5**: e1000326

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, et al (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic acids research **40**: D1202–10

Matsumoto T et al (2005) International Rice Genome Sequencing Project. The map-based sequence of the rice genome. Nature **436**: 793–800

Medici A, Marshall-Colon A, Ronzier E, Szponarski W, Wang R, Gojon A, Crawford NM, Ruffel S, Coruzzi GM, Krouk G (2015) AtNIGT1/HRS1 integrates nitrate and phosphate signals at the Arabidopsis root tip. Nature communications **6**: 6274

Murashige T, Skoog F (1962) A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. Physiologia plantarum **15**: 473–497

Obertello M, Krouk G, Katari MS, Runko SJ, Coruzzi GM (2010) Modeling the global effect of the basic-leucine zipper transcription factor 1 (bZIP1) on nitrogen and light regulation in Arabidopsis. BMC systems biology **4**: 111

Palaniswamy SK, James S, Sun H, Lamb RS, Davuluri R V, Grotewold E (2006) AGRIS and AtRegNet. A Platform to Link cis-Regulatory Elements and Transcription Factors into Regulatory Networks. Plant physiology **140**: 818–829

Para A, Li Y, Marshall-Colón A, Varala K, Francoeur NJ, Moran TM, Edwards MB, Hackley C, Bargmann BOR, Birnbaum KD, et al (2014) Hit-and-run transcriptional control by bZIP1 mediates rapid nutrient signaling in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America **111**: 10371–10376

Rohila J, Chen M, Chen S, Chen J, Cerny R, Dardick C, Canlas P, Fujii H, Gribskov M, Kanrar S, et al (2009) Protein-Protein Interactions of Tandem Affinity Purified Protein Kinases from Rice. PloS one 4:

Rohila JS, Chen M, Chen S, Chen J, Cerny R, Dardick C, Canlas P, Xu X, Gribskov M, Kanrar S, et al (2006) Protein-protein interactions of tandem affinity purification-tagged protein kinases in rice. The Plant journal : for cell and molecular biology **46**: 1–13

Rubin G, Tohge T, Matsuda F, Saito K, Scheible W-R (2009) Members of the LBD family of transcription factors repress anthocyanin synthesis and affect additional nitrogen responses in Arabidopsis. The Plant cell **21**: 3567–84

Santi C, Bogusz D, Franche C (2013) Biological nitrogen fixation in non-legume plants. Annals of botany **111**: 743–67

Sasaki T, Sederoff RR (2003) Genome studies and molecular genetics. The rice genome and comparative genomics of higher plants. Current Opinion in Plant Biology **6**: 97–100

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research **13**: 2498–504

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular biology and evolution **28**: 2731–9

Van Verk MC, Bol JF, Linthorst HJM (2011) WRKY transcription factors involved in activation of SA biosynthesis genes. BMC plant biology **11**: 89

Vidal E a, Tamayo KP, Gutierrez R a (2010) Gene networks for nitrogen sensing, signaling, and response in Arabidopsis thaliana. Wiley interdisciplinary reviews Systems biology and medicine **2**: 683–93

Wang R, Guegler K, LaBrie ST, Crawford NM (2000) Genomic analysis of a nutrient response in Arabidopsis reveals diverse expression patterns and novel metabolic and potential regulatory genes induced by nitrate. The Plant cell **12**: 1491–509

Wang R, Okamoto M, Xing X, Crawford NM (2003) Microarray Analysis of the Nitrate Response in Arabidopsis Roots and Shoots Reveals over 1,000 Rapidly Responding Genes and New Linkages to Glucose, Trehalose-6-Phosphate, Iron, and Sulfate Metabolism. Plant physiology **132**: 556–567

Wang R, Tischner R, Gutierrez RA, Hoffman M, Xing X, Chen M, Coruzzi G, Crawford NM (2004) Genomic Analysis of the Nitrate Response Using a Nitrate Reductase-Null Mutant of Arabidopsis. Plant physiology **136**: 2512–2522

Yilmaz A, Nishiyama MY, Fuentes BG, Souza GM, Janies D, Gray J, Grotewold E (2009) GRASSIUS: a platform for comparative regulatory genomics across the grasses. Plant physiology **149**: 171–80

Yu J, Wang J, Lin W, Li S, Li H, Zhou J, Ni P, Dong W, Hu S, Zeng C, et al (2005) The Genomes of Oryza sativa: a history of duplications. PLoS biology **3**: e38

Tables

	Roots		Shoots	
	Induced	**Repressed**	**Induced**	**Repressed**
Rice 451 genes	234 (51.8%)	106 (23.5%)	103 (22.8%)	39 (8.6%)
Arabidopsis 1417 genes	757 (53.4%)	424 (29.9%)	166 (11.7%)	184 (12.9%)

Table 1. Number of nitrogen regulated genes in *O. sativa* and *A. thaliana*. Percentage of regulated genes for each type of regulation is in parentheses.

Gene_ID	Gene description	Log2 Ratio	
		Root	Shoot
Nitrate uptake			
LOC_Os02g02170	high affinity nitrate transporter, putative, expressed	-0.93	NC
LOC_Os02g38230	component of high affinity nitrate transporter, putative, expressed	1.66	NC
Nitrate/nitrite assimilation			
LOC_Os02g53130	nitrate reductase, putative, expressed	2.15	3.32
LOC_Os01g25484	ferredoxin--nitrite reductase, chloroplast precursor, putative, expressed	2.37	2.67
Ferredixin Reduction			
LOC_Os01g64120	ferredoxin-6, chloroplast precursor, putative, expressed	3.28	4.22
LOC_Os05g37140	ferredoxin-6	1.33	0.89
LOC_Os03g57120	ferredoxin--NADP reductase, root isozyme, chloroplast precursor, putative, expressed	1.55	1.95
LOC_Os04g44650	ferredoxin-thioredoxin reductase	0.67	0.95
Pentose Phosphate Pathway			
LOC_Os08g08840	glucose-6-phosphate/phosphate translocator 2, chloroplast precursor, putative, expressed	0.84	NC
LOC_Os07g22350	glucose-6-phosphate 1-dehydrogenase 2, chloroplast precursor, putative, expressed	1.62	1.07
Ammonium assimilation			
LOC_Os01g48960	glutamate synthase, chloroplast precursor, putative, expressed	1.17	NC
LOC_Os04g56400	glutamine synthetase, chloroplast precursor, putative, expressed	1.09	0.71

Table 2. Selected rice genes regulated by nitrogen in shoots and roots (for more details see Materials and Methods). See Suppl. Table S1 for full list of genes. The fold change of nitrogen response genes were calculated as the ratio between N / KCl expression value. *p*-value cut-off ≤ 0.05 and fold-change ≥ 1.5-fold (shown on table is the $\log_2$ of values, fold-change cut-off $\log_2 1.5 = 0.585$). NC, no change.

Gene_ID	Gene description	Log2 Ratio	
		Roots	Shoots
Nitrate uptake			
At1g69850	nitrate transporter 1:2; calcium ion binding / transporter (NRT1:2)	-0.62	-1.04
At5g50200	Nitrate transmembrane transporters (NRT3.1)	1.79	2.04
At1g08090	high affinity nitrate transporter 2.1 (NRT2:1)	2.76	2.35
Nitrate/Nitrite assimilation			
At1g37130	nitrate reductase 2 (NIA2)	NC	2.65
At1g77760	nitrate reductase 1 (NIA1)	3.69	5.85
At2g15620	nitrite reductase; ferredoxin-nitrate reductase (NIR1)	3.33	6.33
Ferredixin Reduction			
At2g27510	ferredoxin 3; electron carrier (ATFD3)	1.52	3.16
At4g05390	ROOT FNR 1; oxidoreductase (ATRFNR1)	2.49	3.99
At1g30510	ROOT FNR 2; oxidoreductase (ATRFNR)	2.75	4.43
Pentose Phosphate Pathway			
At1g24280	Glucose-6-phosphate dehydrogenase 3 (G6PD3)	3.34	4.78
At5g13110	Glucose-6-phosphate dehydrogenase 2 (G6PD2)	1.93	2.99
Ammonium assimilation			
At5g35630	glutamine synthetase 2 (GS-GLN2)	1.14	NC
At5g16570	Glutamine synthetase 1;4 (GLN1;4)	-1.20	NC
At5g53460	NADH-dependent glutamate synthase 1 gene (GLT1)	1.38	2.40
Glutamate biosynthesis/degradation			
At1g51720	glutamate dehydrogenase, putative	1.14	NC
At5g07440	glutamate dehydrogenase 2 (GDH2)	1.76	NC

Table 3. Selected Arabidopsis genes regulated by nitrogen in shoot and/or roots (for more details see Materials and Methods). The fold change of nitrogen response genes were calculated as the ratio between N / KCl expression value. *p*-value cut-off ≤ 0.05 and fold-change ≥ 1.5-fold (shown on table is the $\log_2$ of values, fold-change cut-off $\log_2$ 1.5 =0.585). NC, no change.

Table 4

Rice Gene Locus	Rice Gene Description	Number of directed connections			
		Rice Predicted N-regulatory Network (RPNN-predicted interactions)	Rice-Arabidopsis N-regulatory Network (RANN-BLAST)	Rice-Arabidopsis N-regulatory Network (RANN-OrthoMCL)	Rice-Arabidopsis N-regulatory Network (RANN-Union)
LOC_Os03g55590	DNA binding protein, putative, expressed	180	45	17	46
LOC_Os01g53260	OsWRKY23 - Superfamily of rice TFs having WRKY and zinc finger domains, expressed	162	41		41
LOC_Os01g64000	ABA response element binding factor, putative, expressed	138	35		35
LOC_Os01g06640	DNA binding protein, putative, expressed	191	31		31
LOC_Os07g02800	myb-like DNA-binding domain, SHAQKYF class family protein, expressed	150	30	14	30
LOC_Os01g43650	OsWRKY11 - Superfamily of rice TFs having WRKY and zinc finger domains, expressed	166	26	10	29
LOC_Os11g47870	chitin-inducible gibberellin-responsive protein 2, putative, expressed	163	27		27
LOC_Os09g35030	sbCBF6, putative, expressed	229	16		16
LOC_Os09g25070	OsWRKY62 - Superfamily of rice TFs having WRKY and zinc finger domains, expressed	170	14		14
LOC_Os01g34060	DNA binding protein, putative, expressed	99	12		12
LOC_Os04g42950	DNA binding protein, putative, expressed	96	12		12
LOC_Os09g32260	ANAC079/ANAC080, putative, expressed	129	12		12
LOC_Os11g08210	NAC domain-containing protein 71, putative, expressed	99	11		11
LOC_Os02g15340	NAC domain-containing protein 76, putative, expressed	67	10		10
LOC_Os04g55970	DNA binding protein, putative, expressed	143	10		10
LOC_Os12g10660	salt tolerance-like protein, putative, expressed	50	10		10
LOC_Os03g04310	DNA binding protein, putative, expressed	84	7		7
LOC_Os10g42130	ANAC071, putative, expressed	78	6		6
LOC_Os01g64020	transcription factor HBP-1b, putative, expressed	219	5		5
LOC_Os08g42550	AP2 domain containing protein, expressed	61	5		5
LOC_Os06g41100	TGA10 transcription factor, putative, expressed	119	2		2
LOC_Os05g37170	transcription factor TGA6, putative, expressed	90	1		1
LOC_Os02g06910	auxin response factor 6, putative, expressed	163			
LOC_Os03g21710	WRKY DNA binding domain containing protein, expressed	79			
LOC_Os03g47730	knotted1-interacting protein, putative, expressed	61			
LOC_Os03g55220	helix-loop-helix DNA-binding, putative, expressed	175			
LOC_Os04g56990	transfactor, putative, expressed	293			
LOC_Os05g20930	transcriptional regulator RABBIT EARS, putative, expressed	195			
LOC_Os05g38140	bHLH transcription factor, putative, expressed	79			
LOC_Os06g07030	dehydration responsive element binding protein, putative, expressed	101			
LOC_Os09g26420	ethylene response factor, putative, expressed	69			
LOC_Os09g36160	SHI, putative, expressed	30			

Table 4. List of the transcription factors in the "Rice-Arabidopsis N-regulatory Network (RANN-Union)". For each step of the network construction (Figure 2), transcription factors were rank based on their number of connections in the network.

Figures

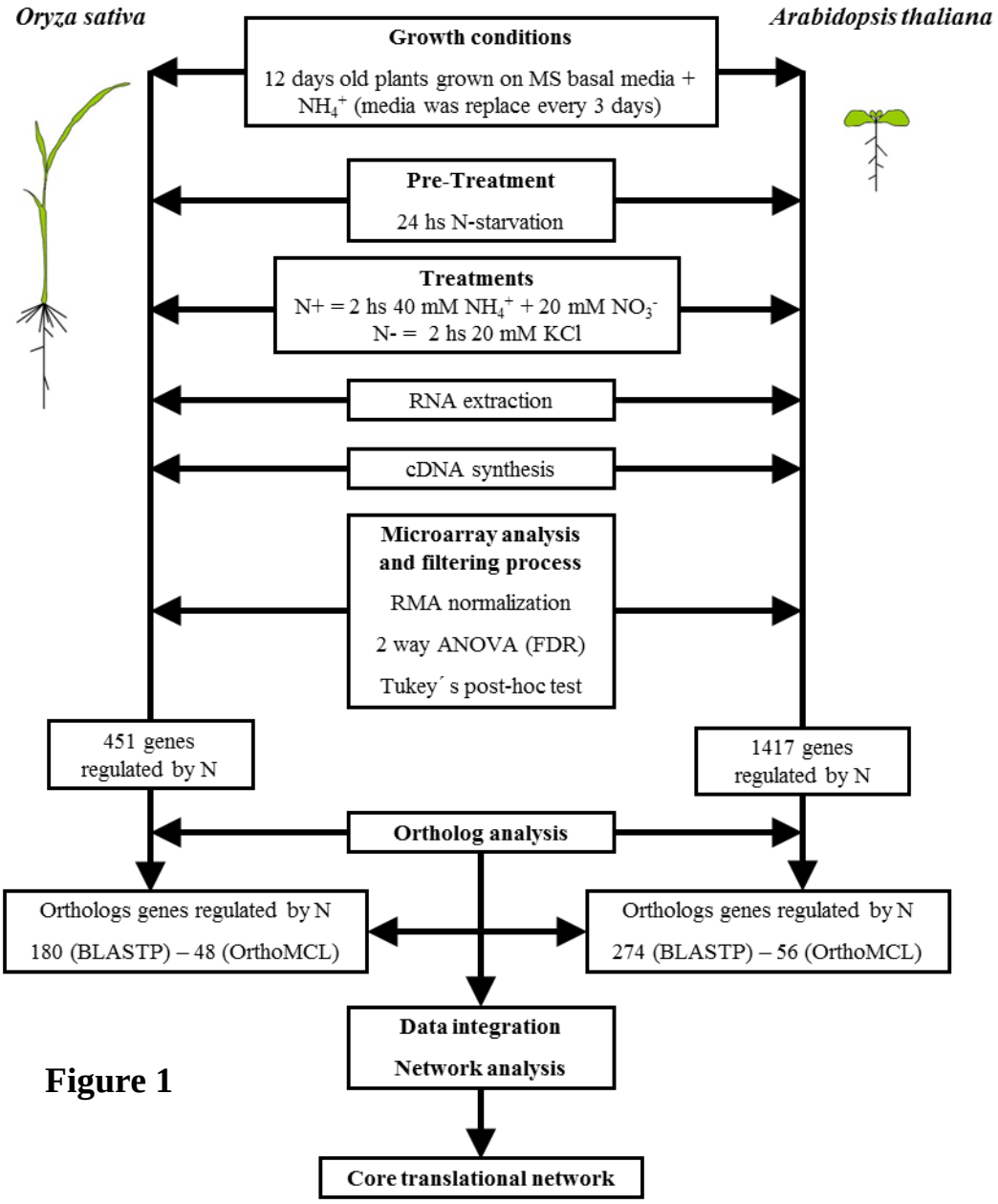
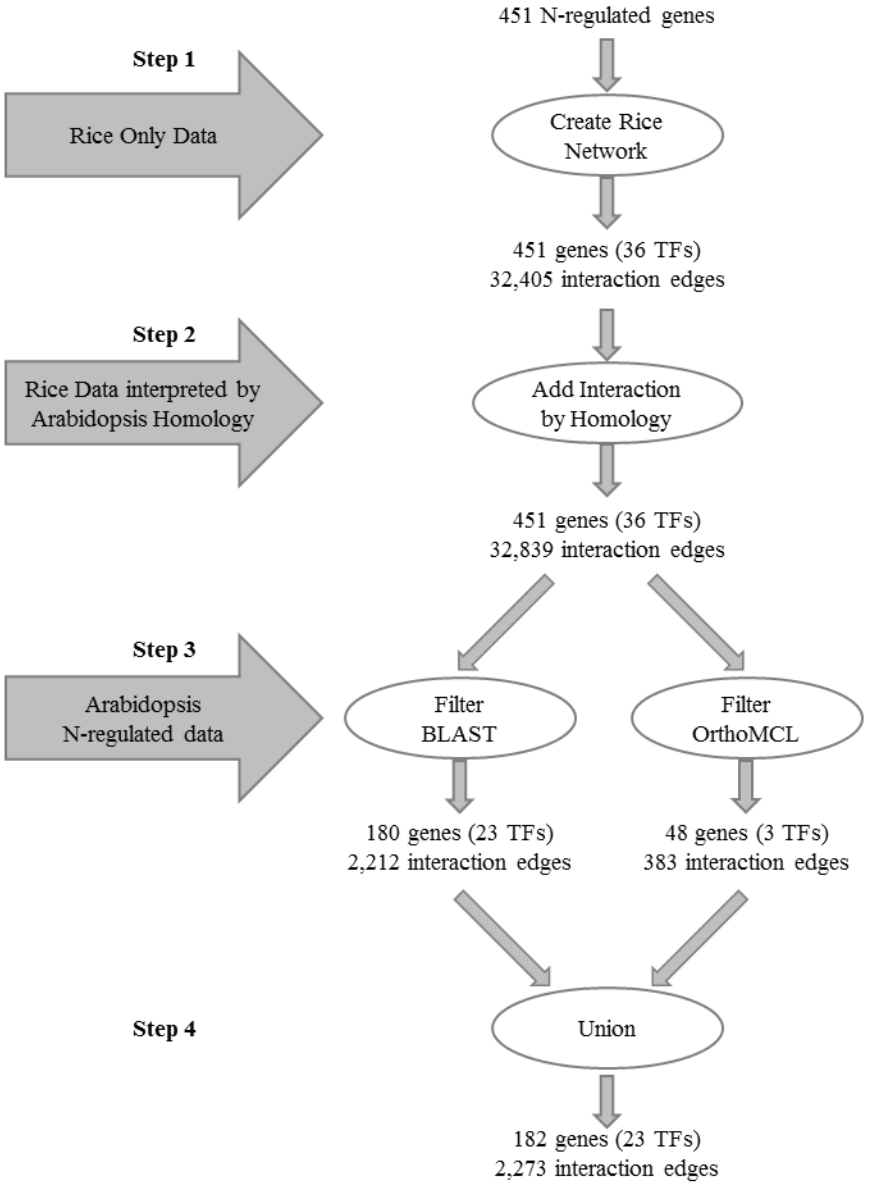


Figure 1

Figure 1. A schematic diagram of the experimental and data mining approach used in this study. Briefly, *O. sativa* (rice) and *A. thaliana* plants were grown for 12 days before a 2 hr treatment with 1xN vs. KCl control (see Methods). Genome-wide analysis using Affymetrix chips has been used in order to quantify mRNA levels. Modeling of microarray data, using ANOVA, homology/orthology and network analysis (detailed in Methods), were used to identify a core translational N-regulatory network shared between rice and Arabidopsis.



Network Edge Statistics

Rice Only N-response Network (RONN)

Metabolic	:	178 edges
Protein-protein	:	2 edges
Correlation Only	:	32,225 edges

Rice Predicted N-regulatory Network (RPNN-predicted interactions)

Metabolic	:	178 edges
Protein-protein	:	268 edges
Regulatory (Cis+Correlation)	:	4,128 edges
Correlation Only	:	28,265 edges

Rice-Arabidopsis N-regulatory Network (RANN-BLAST)

Metabolic	:	120 edges
Protein-protein	:	169 edges
Regulatory (Cis+Correlation)	:	368 edges
Conserved Correlation Only	:	1,555 edges

Rice-Arabidopsis N-regulatory Network (RANN-OrthoMCL)

Metabolic	:	46 edges
Protein-protein	:	21 edges
Regulatory (Cis+Correlation)	:	41 edges
Conserved Correlation Only	:	275 edges

Rice-Arabidopsis N-regulatory Network (RANN-Union)

Metabolic	:	120 edges
Protein-protein	:	169 edges
Regulatory (Cis+Correlation)	:	372 edges
Conserved Correlation Only	:	1,612 edges

Data Source

- Add RiceCyc metabolic interactions (Dharmawardhana et al., 2013).
- Add protein-protein experimentally determined interactions (Rohila et al., 2006, Rohila et al., 2009, Ding et al., 2009; Gu et al., 2011).
- Add rice experimental correlation data (this study).

- Add protein-protein predicted interactions (Interolog method) (Gu et al., 2011; Ho et al., 2012).
- Add predicted regulatory interactions using cis-motifs from Arabidopsis (Palanisway et al., 2006, Yilmaz et al., 2009).

- Filter Rice Predicted N-regulatory Network with Arabidopsis N-regulated genes and their correlation (this study). Arabidopsis-Rice homology was obtained using BLAST (Reverse BLAST multiple hits, e-value =< 1e-20) (Camacho et al., 2014).

- Filter Rice Predicted N-regulatory Network with Arabidopsis N-regulated genes and their correlation (this study). Arabidopsis-Rice homology was obtained using OrthoMCL (Stoeckert et al., 2011).

Figure 2. The workflow of the network analysis of N-regulated genes differentially expressed in rice resulting in "Rice-Arabidopsis N-regulatory Network (RANN-Union)". The input was 451 rice N-regulated genes. In each of the three steps, we introduced rice and Arabidopsis homology knowledge which includes N-

Figure 3

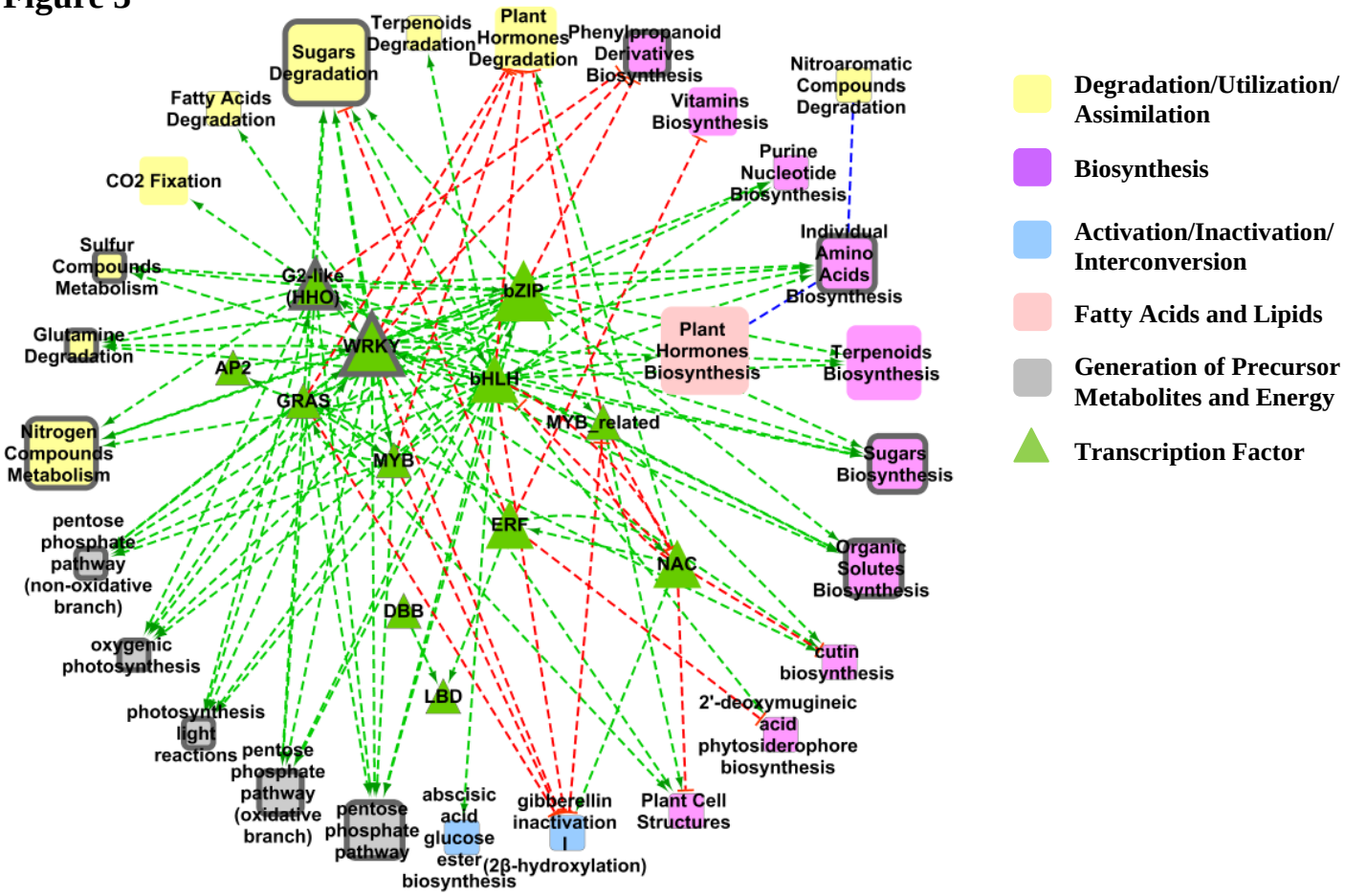


Figure 3. Supernode network analysis created from the 182 genes of "Rice-Arabidopsis N-regulatory Network" (RANN-Union). Individual nodes were clustered based on PlantCyc pathways and TF families classification to form supernodes. Genes which do not belong to either of the two classifications are not shown here. Triangles represent TFs families and squares represent PlantCyc pathways (Zhang et al., 2010). The size of the nodes is proportional to the number of genes within that particular category (from 1 to 5). Nodes are connected by TF:target (red dashed lines = predicted negative correlation; green dashed lines = predicted positive correlation) and predicted protein-protein interactions (blue dashed lines). All nodes are present in the "Rice-Arabidopsis N-regulatory Network" (RANN-BLAST) supernode network. Nodes circled in thick grey lines are also present in the "Rice-Arabidopsis N-regulatory Network" (RANN-OrthoMCL) supernode network.