THE PLANT CELL

The Plant Cell, 2025, **37**, koaf093

https://doi.org/10.1093/plcell/koaf093
Advance access publication 14 May 2025
Research Article

Model-to-crop conserved NUE Regulons enhance machine learning predictions of nitrogen use efficiency

Ji Huang,[1] Chia-Yi Cheng,[1,2] Matthew D. Brooks,[3] Tim L. Jeffers,[1] Nathan M. Doner,[1] Hung-Jui Shih,[1,†] Samantha Frangos,[1] Manpreet Singh Katari,[1] Gloria M. Coruzzi[1,*]

[1]Department of Biology, Center for Genomics and Systems Biology, New York University, New York, NY 10003, USA
[2]Department of Life Science, College of Life Science, National Taiwan University, Taipei 10663, Taiwan
[3]Global Change and Photosynthesis Research Unit, USDA-ARS, Urbana, IL 61801, USA
*Author for correspondence: gloria.coruzzi@nyu.edu (G.M.C.)
[†]Present address: Institute of Plant and Microbial Biology, Academia Sinica, Taipei 115201, Taiwan.
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 (https://academic.oup.com/plcell/pages/General-Instructions) is: Dr. Gloria M. Coruzzi.

Abstract

Systems biology aims to uncover gene regulatory networks (GRNs) for agricultural traits, but validating them in crops is challenging. We addressed this challenge by learning and validating model-to-crop transcription factor (TF) regulons governing nitrogen use efficiency (NUE). First, a fine-scale time-course nitrogen (N) response transcriptome analysis revealed a conserved temporal N response cascade in maize (*Zea mays*) and Arabidopsis (*Arabidopsis thaliana*). These data were used to infer time-based causal TF target edges in N-regulated GRNs. By validating 23 maize TFs in a cell-based TF-perturbation assay (Transient Assay Reporting Genome-wide Effects of Transcription factors), precision/recall analysis enabled us to prune high-confidence edges between ~200 TFs/700 maize target genes. We next learned gene-to-NUE trait scores using XGBoost machine learning models trained on conserved N-responsive genes across maize and Arabidopsis accessions. By integrating NUE gene scores within our N-GRN, we ranked maize TFs based on a cumulative NUE Regulon score. NUE Regulons for top-ranked TFs were validated using the cell-based TARGET assay in maize (e.g. ZmMYB34/R3→24 targets) and the Arabidopsis ZmMYB34/R3 ortholog (e.g. AtDIV1→23 targets). The genes in this NUE Regulon significantly enhanced the ability of XGBoost models to predict NUE traits in both maize and Arabidopsis. Thus, our pipeline for identifying TF regulons that combines GRN inference, machine learning, and orthologous network regulons offers a strategic framework for crop trait improvement.

Introduction

Improving agronomic traits in staple crops like maize (*Zea mays*) is essential to meet the demands of a growing global population (FAO 2023). Identifying key genes in crops that influence these traits such as nitrogen use efficiency (NUE) is crucial to both agriculture and the environment. In the age of genomics, gene regulatory networks (GRNs) and machine learning are 2 promising new approaches for this task. However, these approaches have been difficult to navigate and integrate in large and complex genomes of crop plants and even more challenging to functionally validate.

By contrast, GRNs inference and machine learning approaches can generate hypothesis to test in the well-studied model species Arabidopsis (*Arabidopsis thaliana*). This includes identifying genes of importance to a trait using machine learning and inferring their regulatory connections via network inference (e.g. TF → gene → trait).

Machine learning methods can successfully identify key genes important for traits using the increasing amount of omics data (Ma et al. 2014; Cheng et al. 2021). At the same time, GRNs provide a systematic approach to improving agronomic traits by elucidating transcription factor (TF)-to-gene regulation (Krouk et al. 2013; Springer et al. 2019). However, the challenge remains in transferring the GRN knowledge from model-to-crop and applying the findings in field settings (Ferrier et al. 2011; Liu et al. 2022). In this study, we integrate GRNs with machine learning and model-to-crop orthology mapping to enhance NUE, a key factor for maize yield (Lassaletta et al. 2014) and environmental sustainability (Zhang et al. 2015; Gao and Cabrera Serrenho 2023).

Previous GRN studies were conducted in the model species Arabidopsis, for which N-signaling and its links to genes involved in N-uptake and assimilation have been well studied. For example, novel regulators of a temporal N-signaling network were uncovered using network inference on time-course transcriptome N-response data and were functionally validated in Arabidopsis (Krouk et al. 2010; Varala et al. 2018; Brooks et al. 2019; Alvarez et al. 2020). Additionally, functional testing using a yeast-one-hybrid system was used to map TF binding to genes in an N-metabolism regulatory network (Gaudinier et al. 2018). Indeed, transferring such data-rich insights from Arabidopsis to less-studied crop species could be highly advantageous, as reviewed in Curci et al. (2022). One effective method for transferring knowledge from model-to-crop is through the conserved orthologs across species, as reviewed in Koonin (2005). Indeed, several recent studies have shown the conservation of ortholog functions under various stress conditions, including salt stress (Wu et al. 2021), cold stress (Zhang et al. 2017), oxidative stress (Hartmann et al. 2022), and leaf senescence (Liu et al. 2023).

Received March 19, 2025. Accepted April 7, 2025

In a previous study, we tested and validated the hypothesis that orthologous genes whose N-response was conserved across model-to-crop could enhance our ability to learn genes-of-importance to NUE (Cheng et al. 2021). By implementing a machine learning approach called XGBoost (Chen and Guestrin 2016), we discovered that using orthologs whose N-response is conserved between Arabidopsis and field-grown maize significantly enhanced our ability to predict NUE traits in left-out samples, compared with orthologs whose N-response was not conserved (Cheng et al. 2021). However, the regulatory networks involving these model-to-crop N-response genes-of-importance to NUE remain poorly understood. The goal of our current study is to infer and validate high-confidence nitrogen (N)-gene regulatory networks (N-GRNs) controlling NUE using fine-scale time-series data in maize, as we did for Arabidopsis (Varala et al. 2018; Brooks et al. 2019). The cross-species comparison would enable us to identify N-regulatory network modules conserved across a C3–C4 divide.

Indeed, a number of examples suggest a possible conserved function between Arabidopsis and maize orthologs, despite the fact that they span the C3/C4 divide (Cui 2021). In the case of NUE, Arabidopsis genes that are important for N-uptake/metabolism and regulation have orthologs in maize that are also critical for NUE. For example, maize lines that overexpress nitrate transporter/peptide transporter family 2 (ZmNPF6.6/ZmNRT1.1), the ortholog of the Arabidopsis nitrate transceptor, NITRATE TRANSPORTER 1.1 (Wang et al. 2020a), increase grain yield in maize under low nitrogen conditions (Cao et al. 2023). Maize genes from the cytosolic glutamine synthetase (GS1) family, Glutamine Synthetase 1.3 and 1.4 (ZmGLN1.3 and ZmGLN1.4), control grain yield in maize (Martin et al. 2006), and mutants in their Arabidopsis orthologs show reduced biomass and seed yield (Lothier et al. 2011; Guan et al. 2015). Lastly, a master N-response TF in Arabidopsis Nodule inception-like protein, NIN-like protein 7 (Castaings et al. 2009; Alvarez et al. 2020) has a maize ortholog NIN-like protein 5 (ZmNLP5, also known as ZmNLP13), which modulates N responses in maize under N-limited conditions (Ge et al. 2020). The finding that master TFs regulate the N-response cascade suggests that maize and Arabidopsis share an N-response GRN. However, despite these studies on individual genes, only a few transcriptome-wide studies in maize have explored systematic nitrogen signaling (Yang et al. 2011; Ge et al. 2020; Wang et al. 2020b; Buoso et al. 2021b; Cheng et al. 2021) but only focusing on a single time point and tissue type (Ge et al. 2020; Wang et al. 2020b; Buoso et al. 2021b).

To bridge these knowledge gaps, our study aims to: (i) infer and validate high-quality N-GRNs in maize, (ii) identify N-responsive gene orthologs in maize and Arabidopsis that control NUE traits using machine learning, and (iii) functionally validate conserved TFregulons (TF-to-target modules) (Aibar et al. 2017; Guillotin et al. 2023) whose expression can predict the NUE trait in each species. Toward this goal, in this model-to-crop study, we inferred time-based N-GRNs in maize (shoot and root) using RNA from a fine time-scale N-response experimental design that parallels a companion study in Arabidopsis (Varala et al. 2018; Brooks et al. 2019). Using a just-in-time (JIT) analysis of these data enabled us to identify a temporal cascade of N-responses conserved between model-to-crop. Importantly, we used this time-series data to infer a time-based N-response GRN in maize, which we validated and refined using precision/recall analysis. We did this by generating data for 23 maize TF whose genome-wide targets were validated using the cell-based Transient Assay Reporting Genome-wide Effects of Transcription factors (TARGET) assay (Bargmann et al. 2013; Brooks et al. 2019), modified for maize. Next, to identify the network modules of importance to NUE, we integrated our high-confidence N-GRNs with machine learning feature importance values into a NUE Regulon score. This enabled us to identify and validate an NUE Regulon conserved model-to-crop (AtDIV1 and ZmMYB34/R3) that can accurately predict NUE trait in left-out samples. Thus, our approach for identifying cross-species NUE Regulons offers promising candidates for NUE improvement via transgenics, genome editing, or molecular breeding.

Results

Identification of a temporal cascade of early-to-late nitrogen-responsive genes in maize

The quest to identify N-GRNs in maize began with microarray studies (Jiang et al. 2018) and later by RNA-seq analysis (Ge et al. 2020; Wang et al. 2020b; Buoso et al. 2021a). Recently, a major advance is the combined analysis of N-response RNA-seq and phenotype data in the machine learning of N-response genes-of-importance to NUE in maize (Cheng et al. 2021). Our current study aims to expand on these GRN studies by exploiting time, an unexplored dimension of the N-GRN response in maize. As causality moves forward in time, fine-scale time-series RNA-seq data are a valuable source to derive predictive networks. In our current study, we implement and validate a time-inferred causal GRN in maize. We then compare it to a parallel study in the model Arabidopsis (Varala et al. 2018) to reveal temporal N-regulatory networks conserved model-to-crop.

To uncover time-dependent effects of nitrogen (N)-treatment in maize, we conducted a fine-scale series of N-treatments over 10 time points (0 to 120 min) using a parallel design to the study we conducted in Arabidopsis (Varala et al. 2018). This N-treatment (20 mM KNO_3 + 20 mM NH_4NO_3) in the widely used Murashige and Skoog (MS) media was compared to a control (20 mM KCl), as done previously in Arabidopsis (Varala et al. 2018) (Supplementary Fig. S1). We measured N-related phenotypes including ^{15}N-uptake and collected samples for RNA-seq analysis from shoots and roots of maize B73 seedlings (see Materials and methods) (Supplementary Fig. S1). This replicated experimental design allowed us to follow the time-dependent N-responses in maize, as well as those conserved model-to-crop.

Our ^{15}N enrichment analysis showed a linear increase in total nitrogen uptake (nitrate and ammonium) in shoots and roots of maize B73 seedlings (Fig. 1, A and B and Supplementary Table S1). Moreover, principal component analysis (PCA) of gene expression data from shoots and roots (Supplementary Fig. S1, D and E) showed that PC1 is strongly associated with N-treatment, and PC2 correlates significantly with time (Supplementary Fig. S1, F and G). Thus, the effectiveness of our N-treatment experimental design is validated by both the PCAs of ^{15}N enrichment and RNA-seq data.

Leveraging our fine-scale time-course experimental design, we employed spline modeling (Storey et al. 2005) to identify genes regulated by N as a function of time (NxTime genes), as we previously did in Arabidopsis (Varala et al. 2018) (Supplementary Fig. S2A). By smoothing gene expression using a natural cubic spline across time samples, we could identify genes whose NxTime response curve is significantly different between N-treated vs. control KCl conditions (Fig. 1C). Spline analysis identified 2,732 NxTime response genes in maize shoots and 2,294 NxTime response genes in roots (Supplementary Tables S2 and S3). Among these gene sets, 453 maize genes, including 55 TFs, are NxTime responsive in both shoot and root (Supplementary Fig. S2B). A comparative analysis of the NxTime genes we uncovered in maize with published single-time-point N-treatment datasets captures significant overlaps with other N-treatment studies in maize (Supplementary Fig. S3, A and B). Importantly, 50% of the NxTime-responsive maize

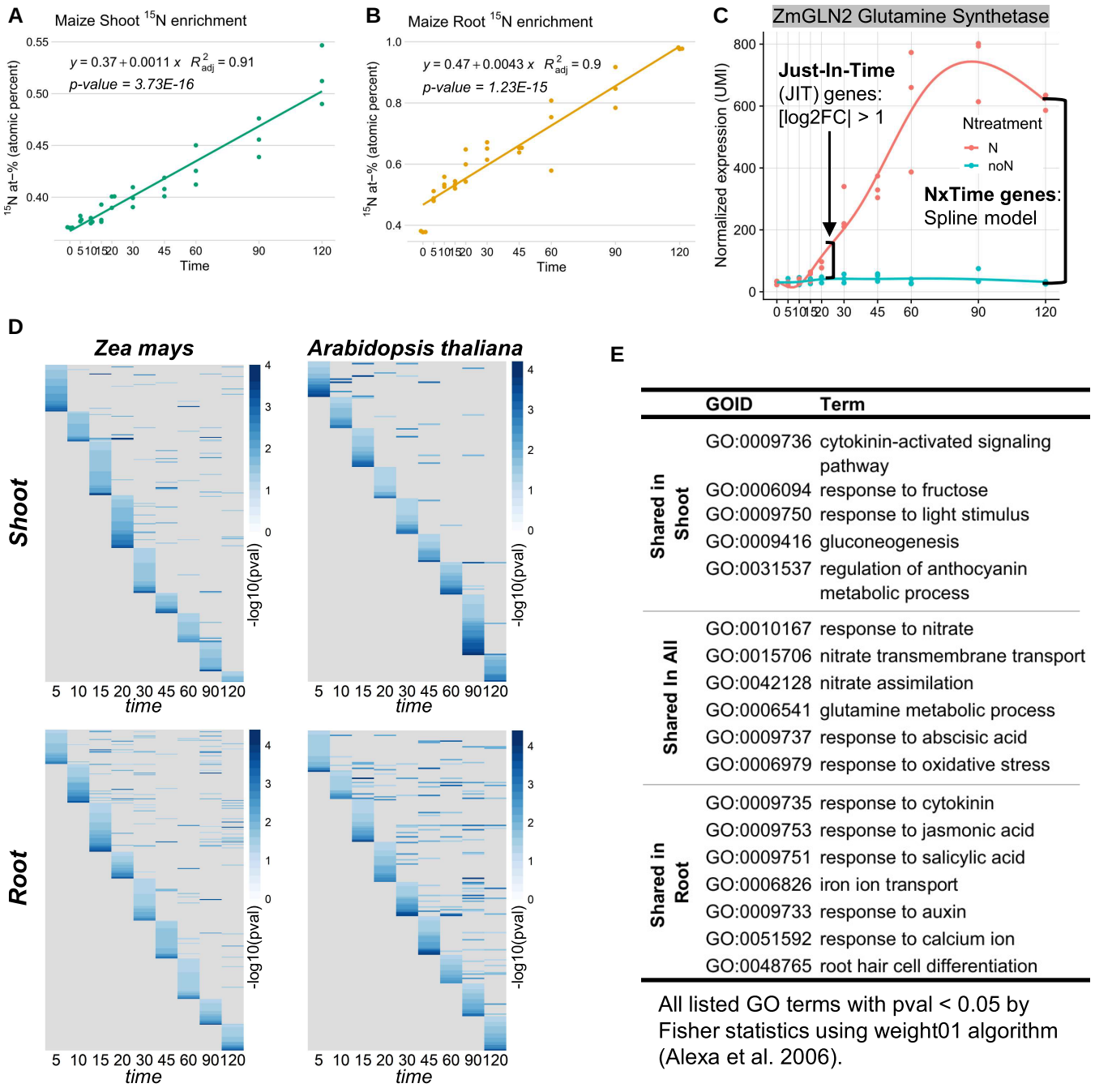


Figure 1. A temporal cascade of N-responsive "JIT" genes uncovers N-related GO process conserved between maize and Arabidopsis. N-treatment (+N/−N) of maize was used to monitor ^{15}N-uptake (**A** and **B**) and RNA-seq data collected across a fine-scale time course (5 min—2 h) (**C–E**) in shoots and roots (see Materials and methods). Maize **A)** shoot and **B)** root tissues exhibit a linear increase in total N-uptake, as detected by ^{15}N analysis. Maize genes regulated by N as a function of time (NxTime genes) were identified by Spline modeling of RNA-seq N-response curves, compared with controls (see Materials and methods). Each NxTime gene was assigned to a JIT bin, based on the *first* time point that N-response was > 2-fold (see Materials and methods). **C)** An example maize NxTime gene, ZmGLN2 (Zm00001d026501) encoding a cytosolic glutamine synthetase, was assigned to the 20 min JIT bin. **D)** All maize NxTime genes were chronologically arranged into JIT bins spanning early (5 min) to late (120 min) time points. For a model-to-crop comparison, this same NxTime and JIT analysis pipeline was performed on RNA-seq data from Arabidopsis exposed to an identical time-series N-treatment (Varala et al. 2018). JIT bins for each species show a cascade of GO term enrichment (*P*-values < 0.05) highlighted in blue. **E)** Key GO terms shared by JIT bins in maize and Arabidopsis are categorized by tissue specificity: shared in shoots, roots, or both tissues. These shared GO terms include nitrogen assimilation and processes in N-signaling, including hormone responses.

genes identified in our current study are novel N-responders, not previously detected in single-time-point studies of maize (Supplementary Fig. S3, C to F). Thus, our time-course RNA-seq analysis identified novel and dynamic N-responsive genes in shoots and roots of maize.

Temporal N-regulation of N-uptake and assimilation pathway genes is consistent with ^{15}N-uptake *in planta*

Our time-series N-treatment data of maize showed an increase of ^{15}N-uptake over time (Fig. 1, A and B). We thus asked whether genes

involved in N-uptake/assimilation respond temporally to the increase in internal N-dose as it accumulates over time. Indeed, within our NxTime-responsive gene lists, we identified 35 N-pathway genes in shoot and 29 in roots (see Materials and methods and Supplementary Fig. S4 and Table S4). Moreover, we found that 16/29 N-pathway genes NxTime-responsive expression pattern in roots are significantly correlated with our ^{15}N-uptake measurements (Supplementary Fig. S5). The Pearson's correlation coefficients for all 16 NxTime-responsive N-pathway genes are positive, suggesting that higher N-responsive expression of these genes leads to more nitrogen uptake. We also used the public maize gene atlas (Woodhouse et al. 2021) to explore the NxTime-responsive N-pathway gene expression patterns in the maize adult stages (Supplementary Fig. S6). In both shoot and root, the expression for N-pathway/NxTime-responsive genes extends beyond the seedling stage into the adult stage, with some genes peaking in expression during the reproductive stage (Supplementary Fig. S6). This underscores the potential of NxTime and N-pathway genes in plant improvement strategies related to N-uptake.

Just-In-Time (JIT) analysis uncovers a temporal N-response gene cascade conserved between maize and Arabidopsis

Our time-series N-response data give us the opportunity to explore the temporal dynamics of early-to-late N-response genes in maize as well as those conserved model-to-crop. To explore the conservation of NxTime gene responses, we first performed Fisher's exact test on the overlap of NxTime genes in maize (this study) and Arabidopsis (Varala et al. 2018) for both shoot and root tissues. These results indicate a significant conservation of the temporal nitrogen response across maize and Arabidopsis in both shoot (Fisher's exact test, *P*-value: 1.1E−08) and root (*P*-value: 3.2E−22) tissues. Next, to identify the timing of the N-responses, we applied the Just-In-Time (JIT) analysis to our new maize NxTime dataset, using the approach we previously developed in Arabidopsis (Varala et al. 2018). In our current study, the JIT approach used categorizes genes into bins based on the earliest time point at which they exhibit an absolute log2(Fold Change) > 1, when comparing N-treated versus control conditions over time (Fig. 1C).

We next tested whether gene sets within each JIT bin in maize were enriched in shared Gene Ontology (GO) terms and/or cis-regulatory motifs. GO enrichment analysis of the JIT genes uncovered a temporal N-response gene cascade enriched in nitrogen-related functions. In shoots, nitrate transport showed enrichment at early JIT time points (5, 10, and 15 min), while the glutamine metabolic process was enriched after 30 min, and photosynthesis terms emerged at later stages (120 min) (Supplementary Fig. S7A and Table S5). The trend of nitrate transport followed by glutamine/glutamate metabolic process is more obvious in maize roots, reflecting the sequential process of nitrogen uptake and assimilation in plants (Supplementary Fig. S7B and Table S5).

Next, we aimed to perform a comparative analysis of the temporal N-response cascade we uncovered in maize with the model Arabidopsis. To this end, we revisited the NxTime RNA-seq datasets we previously identified in Arabidopsis (Varala et al. 2018), using an identical JIT pipeline to our slightly modified scheme (see Materials and methods). Our model-to-crop comparison revealed a similarity in temporal N-response patterns and enriched GO term in JIT N-response genes shared between maize and Arabidopsis (*P*-value < 0.05), based on Fisher's exact test (see Materials and methods), in both shoot and root tissues (Fig. 1D and Supplementary Tables S5 and S6). Core processes in nitrogen metabolism and signaling were found to be common across the temporal N-response cascade in both species and tissues (Fig. 1E), although the exact timing of occurrence may not match. These include responses to nitrate (GO: 0010167), nitrate assimilation (GO: 0042128), and the glutamine metabolic process (GO: 0006541). Furthermore, shared GO terms in the N-response cascade in shoots are related to energy and metabolism, such as the response to light stimulus (GO: 0009750). In roots, shared terms in the NxTime response cascade are related to phytohormones, such as response to cytokinin (GO: 0009735) and auxin (GO: 0009733). In summary, the comparative JIT analysis revealed a conserved temporal cascade of N-responsive functions shared between the model species, Arabidopsis, and the crop, maize.

With regard to shared regulatory mechanisms, our analysis revealed that temporal N-response JIT genes in maize (this study) are enriched in specific cis-motif clusters analogous to those in Arabidopsis (Brooks et al. 2019) (Supplementary Fig. S8 and Tables S7 and S8). The temporal pattern for enriched cis-motifs was more obvious in the maize root JIT genes, likely due to the roots' direct exposure to nitrogen treatments. Notably, cis-motifs for the Barley B Recombinant/Basic PentaCysteine (BBR/BPC) TF family were prevalent across almost all JIT bins in shoots and roots in both species. The BBR/BPC motif was found in several key nitrogen pathway genes, such as ZmNiR1.2, ZmGLN1, ZmGLN1.5, and ZmGOGAT2. The BBR/BPC family is known for its role in cytokinin response regulation (Shanks et al. 2018). Cytokinin plays a crucial part in coordinating systematic nitrogen deficiency and demand signaling in Arabidopsis (Ruffel et al. 2011; Naulin et al. 2020). Moreover, BPCs are known to interact with the abscisic acid (ABA) signaling pathway (Mu et al. 2017) and the circadian clock (Lee et al. 2022), both of which are influenced by nitrogen availability (Gutierrez et al. 2008; Zhang et al. 2021; Contreras-López et al. 2022). Thus, the temporal N-response cascade across model-to-crop is linked to cytokinin, ABA, and circadian clock across model-to-crop.

The cis-motif clusters enriched in the temporal N-response JIT gene bins aggregate multiple TFs sharing similar position weight matrices (PWMs). To identify specific TF-binding motifs, we conducted cis-element enrichment using individual TF cis-motifs from the core JASPAR PWM database (Castro-Mondragon et al. 2022) for JIT genes. Similar to our results from cis-motif cluster analysis (described above), we observed a temporal enrichment of cis-motifs at the individual TF level in both maize and Arabidopsis JIT bins (Supplementary Fig. S9 and Tables S9 and S10). This analysis identified cis-motifs for several TFs previously validated to be pivotal to N-signaling in Arabidopsis within JIT gene bins, including AtCRF4 (Varala et al. 2018), AtTCP20 (Guan et al. 2014), or rice OsGRF4 (Li et al. 2018), and AtDIV1 (Cheng et al. 2021). In summary, our analyses emphasize a parallel temporal progression of cis-motifs enriched in the N-response cascade conserved between maize and Arabidopsis, suggesting a conserved mechanism across model-to-crop.

Validation of maize TFs that regulate genes in the temporal N-response cascade

Our next goal was to identify and validate candidate TFs that regulate genes in the temporal N-response cascade in maize, shown in Fig. 1. Drawing from our list of NxTime and JIT genes in maize, we selected 23 TFs spanning early-to-late JIT bins in both shoot and root tissues (Fig. 2). For validation, we selected TFs that span the N-temporal response cascade including first N-responder TFs (5 to 15 min), early-N responder TFs (20 to 45 min), or late N-responder TFs (60 to 120 min) (Fig. 2A). To uncover the direct

A

23 TFs→genome-wide targets validated by TARGET TF perturbation assays in cells

	First (JIT) (5, 10,15min)	Early (JIT) (20, 30, 45min)	Late (JIT) (60 ,90, 120min)	NxTime TF (spline)
Shoot	KN1 Orphan42 bHLH92 GLK44 EREB115	GLK13 MYB81 WRKY107 MYB121 MYBR3	HB54 WRKY26 MYB21 NAC61	MYB34 MYB38 MYB56 NAC66 bZIP17 NAC5 bHLH185 NLP3
Root	EREB115 GLK44 Orphan42 HB55 NAC5	MYBR3 MYB81	WRKY107 NAC61 NAC66	

B

TF	Overlap NxTime Direct Shoot	Specificity to Shoot NxTime
MYB34	744	17%
MYBR3	716	17%
NLP3	352	19%
MYB21	334	19%
G2	1151	16%
MYB81	496	17%
MYB56	347	17%
MYB38	721	15%
bHLH92	568	15%
WRKY36	709	15%
EREB115	431	16%
BZIP17	1181	15%
NACTF66	252	16%
BHLH185	435	15%
KN1	113	17%
Orphan42	593	15%
GLK44	339	15%
HB54	965	14%
NAC109	109	17%
MYB121	913	14%
NAC5	847	14%
WRKY107	854	13%
NAC61	1074	13%

-log10(padj$_{overlap}$) 1 11

C

TF	Overlap NxTime Direct Root	Specificity to Root NxTime
NLP3	385	21%
BZIP17	1180	15%
MYB38	750	16%
G2	1070	15%
HB54	987	15%
NACTF66	295	19%
MYB34	672	15%
GLK44	381	17%
MYB56	357	17%
MYBR3	645	15%
Orphan42	599	15%
bHLH92	552	15%
EREB115	431	16%
MYB81	443	15%
WRKY36	671	14%
MYB21	296	17%
KN1	132	20%
WRKY107	869	14%
NAC5	812	13%
BHLH185	414	14%
MYB121	847	13%
NAC61	1028	12%
NAC109	98	15%

-log10(padj$_{overlap}$) 1 32

D

	NxTime Shoot	NxTime Root
23 TF TARGET result collectively hit	2330 / 2732 (85.29%)	2063 / 2294 (89.93%)
	N-pathway Shoot	**N-pathway Root**
	105 / 118 (88.98%)	108 / 123 (87.8%)

Figure 2. Direct regulated targets of 23 maize TFs validated in the cell-based TARGET assay are enriched in NxTime genes identified in shoots or roots of maize plants. **A)** 23 NxTime TFs, spanning early to late JIT bins in shoots or roots were validated by adapting the medium-throughput cell-based TF–TARGET assay developed in Arabidopsis (Brooks et al. 2019). Validated TF–target genes of 23 TFs in maize cells show significant overlap with NxTime genes in **B)** shoots or **C)** roots of whole maize plants, with darker colors indicating a more significant overlap with whole tissue N-responses, as determined by Fisher's exact tests. Highlighted examples include: ZmNLP3 whose direct targets validated in maize cells are significantly enriched in NxTime genes identified in both shoots and roots of maize seedlings (black links); ZmMYB21 whose direct targets validated in maize cells are significantly enriched in shoot NxTime genes of maize seedlings (green links); ZmHB54 whose direct targets validated in maize cells are significantly associated with root NxTime genes in maize seedlings (brown links). **D)** Collectively, the 23 maize NxTime TFs validated via the TARGET assay in maize cells directly regulate ~85 to 90% of NxTime genes and N-pathway genes identified in whole maize seedlings.

genome-wide targets of these 23 maize TFs, we used the plant cell-based TARGET TF-perturbation system, a rapid medium through TF–target assay (Brooks et al. 2019), which overcomes the challenges of TF validation in maize transgenic plants. The TARGET assay can rapidly identify the direct regulated TF targets in plant cells isolated from whole tissue, which has been validated in Arabidopsis (Bargmann et al. 2013; Varala et al. 2018; Brooks et al. 2019) and recently in rice (Shanks et al. 2022).

In this present study of maize TFs, we adapted the TARGET TF-perturbation assay and TF expression vector to protoplasts derived from etiolated maize seedlings (see Materials and methods and Supplementary Fig. S10 for a depiction of the TARGET assay). Briefly, protoplasts were isolated from etiolated leaves of maize seedlings. The maize protoplasts (p1107) were transfected with a plasmid that overexpresses a TF glucocorticoid (TF-GR) fusion protein expressed from a ubiquitin (ZmaUBQ) (Bargmann et al. 2013; Brooks et al. 2019; Shanks et al. 2022). The plasmid also encodes a ZmaUBQ::RFP protein, which serves as a cell-sorting marker to isolate transfected cells using fluorescence-activated cell sorting (FACS). The GR-TF fusion protein design allows for the controlled entry of the maize TF-GR fusion protein into the nucleus using dexamethasone (DEX). Additionally, cycloheximide (CHX) + DEX treatment inhibits translation of mRNA for TFs. Therefore, only the target genes of the overexpressed TF are identified, when GR-TF samples are compared to the empty vector (EV) (GR-no TF) control (Brooks et al. 2019).

The maize cells expressing the TF-GR fusion (RFP-positive FACS-sorted cells) were collected for RNA-seq analysis and compared to cells transfected with an EV lacking the TF. We performed the TARGET assay on the 23 maize TFs which span early to late responders, and included a handful of NxTime TFs that did not pass a 2-fold change cutoff for the JIT analysis (Fig. 2A). On average, for the 23 maize TFs tested, each was validated to regulate ~4,000 direct target genes in the TARGET assay (Supplementary Table S11 and Data Set 1).

We next sought to validate the relevance of the direct regulated TF targets identified in the cell-based TARGET TF assay in maize cells to the N-response temporal cascade in maize in planta. To this end, we calculated the significance of the overlap between the direct regulated TF targets identified in the cell-based TARGET assay in maize cells with the NxTime genes identified in shoots and roots of maize seedlings (Fig. 2). Collectively, the direct regulated target genes of the 23 maize TFs validated in the cell-based TARGET assay encompass over 85.3% of NxTime genes in shoots and 89.9% in roots of maize plants (Fig. 2D). Using Fisher's exact test, we identified significant overlap of the targets identified for the 23 TFs in the cell-based assay, with the N-response genes identified in shoots or roots of maize plants, as -log10($padj_{overlap}$) (Fig. 2, B and C). This enabled us to determine the specificity of the TF to regulated nitrogen-responsive genes in each organ of maize. To this end, we calculated the percent overlap of the TF-regulated genes in the cell-based TARGET assay, with the NxTime genes within each tissue of maize seedlings, as done previously for Arabidopsis (Brooks et al. 2019). The overlap analysis identified ZmMYB34 (Zm00001d042830) and ZmMYBR3 (Zm00001d038270) as the top-ranked TFs controlling N-response in maize shoots (Fig. 2B). In support of this, in a many-to-one comparison, the maize MYB ortholog in Arabidopsis, AtDIV1 (AT5G58900), shares a role in nitrogen signaling (Varala et al. 2018). Moreover, supporting its role in nitrogen use vivo, an Arabidopsis *Atdiv1* mutant has a higher NUE, compared with wild type (Cheng et al. 2021). In maize roots, the highest-ranked TF for N-specificity *in planta* was ZmNLP3, whose ortholog in Arabidopsis, AtNLP8, is known as a master regulator of nitrate-dependent seed germination (Yan et al. 2016). We also found TFs that are highly ranked for target gene overlap in 1 maize tissue and low in another, for example, ZmMYB21 and ZmHB54 (Fig. 2, B and C). This suggests some TFs may have tissue-specific N-response (shoot versus root), but both can be captured by the cell-based TARGET assay.

We note that the TARGET assay was done in protoplasts isolated from maize leaves (this study) or roots in Arabidopsis (Brooks et al. 2019). We found that regardless of cell type used in the TARGET assay, both studies found that the direct TF targets identified in the TARGET cell-based assay significantly overlap with NxTime genes in both shoot and root tissues of whole plants (Fig. 2, B and C). This result suggests that the TARGET assay, when conducted in cells isolated from a single tissue type, can effectively identify core N-responsive genes and TF–target interactions relevant to either shoots or roots of whole plants.

To next categorize the 23 maize TFs validated in the TARGET assay as potential activators or repressors, we used 2 complementary approaches. First, we annotated the 23 maize TFs based on the presence of known activation domains in their Arabidopsis orthologs (Morffy et al. 2024). Second, for each of the 23 maize TFs, we analyzed its direct targets identified in our cell-based TARGET assays (this manuscript, Supplementary Data Set 1) and performed a binomial test to determine whether the TF was statistically more likely to activate or repress its targets (Supplementary Table S12). While there were specific examples of concurrence of these 2 approaches, overall, we did not observe a consistent correlation between the presence of activation domains (inferred from Arabidopsis orthologs) and the observed transcriptional changes of direct targets in maize (from maize TARGET assay) (see Supplementary Table S12). This suggests that the regulatory activity of these TFs as either activator or repressors of target genes may be context-dependent. For example, in our previous studies of 33 Arabidopsis TFs using the TARGET assay, we found that interacting TF partners and combinatorial cis-element effects likely contribute to the observed context-dependent regulation as activators or repressors, as explained/explored in detail in Brooks et al. (2019, 2021).

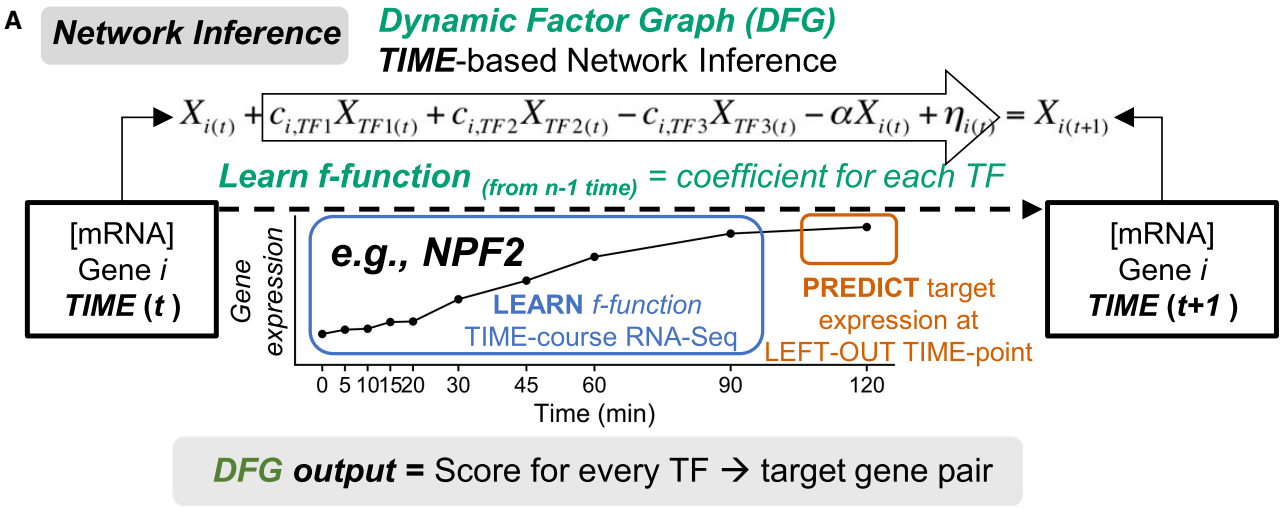
To make this TF–target validation set of 23 maize TFs a resource to the community, we integrated this dataset into the ConnecTF platform, as described in Brooks et al. (2021) (https://maize.connectf.org/). In the following section, we describe how we used this validated TF–target dataset of 23 maize TFs for GRN validation and refinement of our time-inferred nitrogen-signaling network.

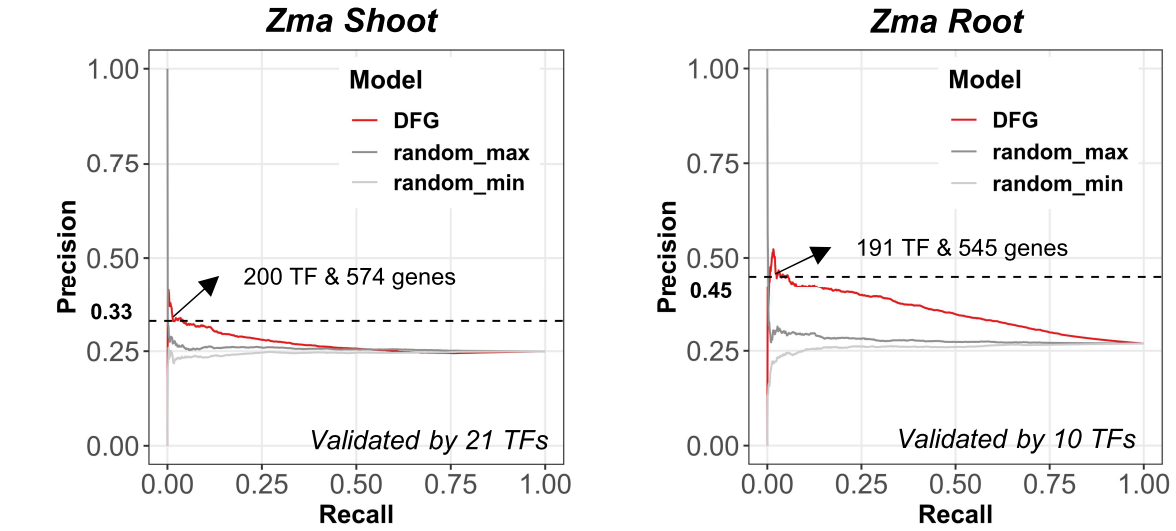
Time-based network inference and validation of TFs controlling the temporal nitrogen response cascade in maize

Inferring GRNs from time-course RNA-seq experiments offers the distinct advantage of identifying potential causal relationships between TFs and genes. Unlike analyses based on steady-state data, time-course experiments provide a dynamic perspective. Using time-series data to infer causality in TF-gene regulation enhances the ability to predict GRN responses at future time-points, a goal of systems biology (Krouk et al. 2010, 2013; Bar-Joseph et al. 2012; Marku and Pancaldi 2023).

Previously, we constructed and validated nitrogen-signaling GRNs (N-GRNs) in Arabidopsis shoots and roots using dynamic factor graphs (DFGs), a time-based machine learning GRN inference method (Krouk et al. 2010, 2013; Varala et al. 2018; Brooks et al. 2019). The DFG method uses fine-scale time-series RNA-seq data to predict the future expression of a target gene. DFG uses the target gene's current state and the expression of TFs at time (t) to estimate expression at the next time point ($t+1$) (Krouk et al. 2010, 2013) (Fig. 3).

In our current study, we applied this time-based DFG method to the maize temporal N-response RNA-seq data which identified 2,732 NxTime-responsive genes in shoots and 2,294 NxTime-responsive genes in roots (Supplementary Tables S2 and S3). Next, we used our validated TF–target data for 23 maize TFs from TARGET (Fig. 2 and Supplementary Data Set 1) to evaluate the performance of the DFG-inferred maize networks in predicting TF–target regulation using an area under the precision–recall (AUPR)





C

DFG	Zma Shoot	Zma Root
AUPR	0.268	0.351
Random-Max	0.256	0.278
Random-Min	0.245	0.260
Random-Mean	0.250	0.269
pvalue	< 0.001	< 0.001
Precision Cutoff	0.330	0.450
Weight Cutoff	1.267	1.124
Number of TFs	200 / 203	191 / 196
Number of Targets	774 / 2,732	736 / 2,294
Number of Edges	15,765 / 554,393	16,606 / 449,428

Figure 3. Generation of high-confidence time-based N-GRNs in maize using DFG and network pruning using TF–target validation data. **A)** The DFG method uses fine-scale time-series data to predict future target gene expression based on prior expression data of target genes and TFs (Krouk et al. 2010, 2013). Using left-out data to test and refine the models, DFG scores each TF's influence on target gene expression, where higher values indicate more confident TF–target edge predictions. **B)** PR analysis: maize shoot DFG network, PR performed using validated TF–target edges for 21 maize TFs in the TARGET cell-based TF assay (Fig. 2). PR analysis of the maize root DFG network performed using validated TF–target edges from 10 maize TFs using the TARGET assay (Fig. 2). **C)** AUPR analysis: DFG-inferred networks significantly outperform random networks (*P*-value < 0.001, permutation test; see Materials and methods). Precision cutoffs: precision cutoffs of 0.33 for shoot and 0.45 for root network were selected based on the flattening points of the PR curves, based on the validated TFs in shoots (21 TFs) and roots (10 TFs). These precision cutoffs determined from the validated TFs were then used to prune the predicted TF–targets for ALL the maize TFs in shoot (203 TFs) and root (196 TFs).

curve analysis. We also compared the performance of our DFG predictions of TF–target relationships, with networks inferred from 2 popular methods, GENIE3 (Huynh-Thu et al. 2010), dynGENIE3 (Huynh-Thu and Geurts 2018), or by correlation methods such as Pearson's correlation coefficient. We compared the performance of the above GRN methods using AUPR, which measures the accuracy of a TF–target prediction model by calculating the area under a curve that plots the precision (the ratio of true-positive predictions to the total number of positive predictions) against recall (the ratio of true-positive predictions to the total number of actual positives). AUPR analysis provides a single value that represents the model's overall performance. This analysis shows that our time-based DFG-inferred networks for both maize shoots (Supplementary Fig. S11A) and roots (Supplementary Fig. S11B) have the highest AUPRs, significantly outperforming networks derived from other network inference methods or random network models (P-value < 0.001, permutation test). We note that these AUPR values for the maize shoot and root GRNs are on par with those reported for high-confidence Arabidopsis GRNs (Varala et al. 2018; Brooks et al. 2019) (Supplementary Table S13). Therefore, the time-based DFG-inferred networks were selected for further downstream analysis in our model-to-crop N-GRN analysis.

In systems biology, networks with high precision offer a more focused selection of high-confidence true TF–gene interactions, thereby facilitating more targeted experimental validation. Thus, to further refine the accuracy of our time-based N-GRNs, we employed a precision–recall (PR) curve-based "pruning" approach—which prunes the network to include only high-confidence TF–target edges (Marbach et al. 2012; Chen and Mar 2018). For the AUPR analysis, we used TF–target interactions validated in the maize cell-based TARGET assay for the 23 TFs of maize whose N-responsive span the temporal N-response RNA-seq dataset (Fig. 2). This analysis enabled us to set precision cutoffs at 0.33 for the maize shoot GRN and 0.45 for the maize root GRN (Fig. 3B). These thresholds to identify high-confidence edges in all the TFs in the N-GRN were chosen based on the point at which the PR curve plateaued on the recall x axis. The TF–target network edges that passed these cutoff thresholds were deemed high-confidence TF–targets in the nitrogen-signaling GRNs (N-GRNs) (Supplementary Tables S14 and S15). This GRN pruning for high-confidence TF–targets retained nearly all maize NxTime-responsive TFs (~200/203), while only ~30% of target N-response genes (~700/2,500) were retained (Fig. 3C). Several N-related GO terms are still enriched in the pruned genes, including nitrate assimilation (GO: 0042128), glutamine family amino acid biosynthetic process (GO:0009084), and glutamate metabolic process (GO: 0006536), as well as energy-related GO terms, such as translation (GO: 0006412), regulation of proton transport (GO: 0010155), and electron transport chain (GO: 0022900) (Supplementary Table S16).

We next sought to ensure that our GRN pruning step was not overly influenced by the choice of TF validation datasets for the 23 maize TFs tested. To this end, we allocated 70% of the 23 maize TFs whose regulated targets were validated in TARGET assay to the pruning set, and the remaining 30% TFs for testing, with the division process randomized and repeated 4 times. Despite slightly different AUPR scores depending on the subset of TFs selected, both pruning and testing phases returned PR curves significantly superior to those of random networks (Supplementary Figs. S12 and S13), indicating that our GRN pruning method was not biased by a few TFs among the 23 maize TFs whose targets we validated in the TARGET assay.

Next, to elucidate a high-confidence temporal maize N-GRN, we built ensemble N-GRNs by integrating TF–targets identified from: (i) our pruned DFG time-based nitrogen response network, (ii) validated TF–target regulation data from the 23 TF TARGET assays, and (iii) TF–target binding data from the published ChIP-seq dataset for 104 maize TFs (22 TFs in the shoot and 18 TFs in the root NxTime genes) (Tu et al. 2020). In our temporal N-response network in maize, the temporal action of 92 TFs in shoot and that of 111 TFs in root are based on their >2-fold N-induction in JIT time bins spanning from early (5 min) to late (120 min) N-responders (Supplementary Tables S2 and S3). As a GRN visual, we highlighted the subnetwork of TFs regulating the nitrogen-uptake/assimilation pathway genes in maize shoots (Fig. 4 and Supplementary Data Set 2) or roots (Supplementary Fig. S14 and Data Set 3). For TF–target edges with multiple types of experimentally validated support (DFG, TARGET, and ChIP), see Supplementary Table S17. This ensemble approach reveals the temporal nature of the GRN controlling nitrogen uptake/assimilation in maize.

"Network Walking" uncovers a new role for ZmKN1 in the maize nitrogen response

Our temporal response N-network uncovered a set of novel 1,047 early N-response genes including 70 TFs whose relevance to the N-response was not previously known (see Supplementary Fig. S3). Surprisingly, one such TF is the well-studied TF Knotted1 (ZmKN1) known to be involved in maize leaf development (Greene et al. 1994; Kerstetter et al. 1997) which we identified as a "first responder" in the temporal nitrogen-signaling cascade. This finding was based on our analysis of our temporal N-response GRN and time-based JIT responses of its genes. ZmKN1 expression was significantly altered within 5 min of N-treatment (Fig. 2A). Previous studies have shown that ZmKN1 is essential for shoot apical meristem function, orchestrating key developmental processes in maize leaves, tassels, and ears (Greene et al. 1994; Bolduc et al. 2012). A new role for ZmKN1 in the temporal N-response in maize is supported by our discovery of a significant overlap between maize shoot NxTime genes uncovered in our study and validated ZmKN1 target genes *in planta* identified as misregulated in zm*kn1* mutants (Bolduc et al. 2012) (P-value <2.2E-16, Fisher's exact test) (Supplementary Fig. S15 and Table S18). These findings support a potential new role for ZmKN1 in mediating the nitrogen response in maize shoots *in planta*.

To validate the hypothesis that ZmKN1 is an early N-responder that mediates the N-GRN response in maize, we sought to decipher the network path that ZmKN1 might mediate to regulate these N-response target genes *in planta*. To this end, we first identified the direct regulated ZmKN1 targets in maize cells by conducting a TARGET TF-perturbation experiment (Fig. 2A and Supplementary Data Set 1). Interestingly, despite ZmKN1's early response to nitrogen, its direct targets identified in the TARGET assay rank 15th out of 23 TFs significance of overlap with NxTime genes in the shoot (Fig. 2B). This result suggests that ZmKN1's regulatory effect on NxTime genes might be mediated by secondary transcription factors (TF_{2s}) regulated directly by ZmKN1 (Fig. 5A).

To test this hypothesis, we used "Network Walking" approach described in Brooks et al. (2019, 2021) to trace the connections from ZmKN1 to direct regulated TF_2 genes, which would link ZmKN1 to its indirect targets *in planta* (Fig. 5A). For the Network Walking approach, we used the ensemble N-GRN in shoots including edges from pruned high-confidence DFG, TF–targets validated in the cell-based TARGET assay, and cell-based ChIP-seq data. The resulting "Network Walking" enabled us to connect ZmKN1 indirectly to 63.1% of its *in planta* N-responsive targets (309/490), via 10 N-responsive TF_{2s} that are direct regulated targets of ZmKN1 (Fig. 5B and Supplementary Data Set 4). The targets in this

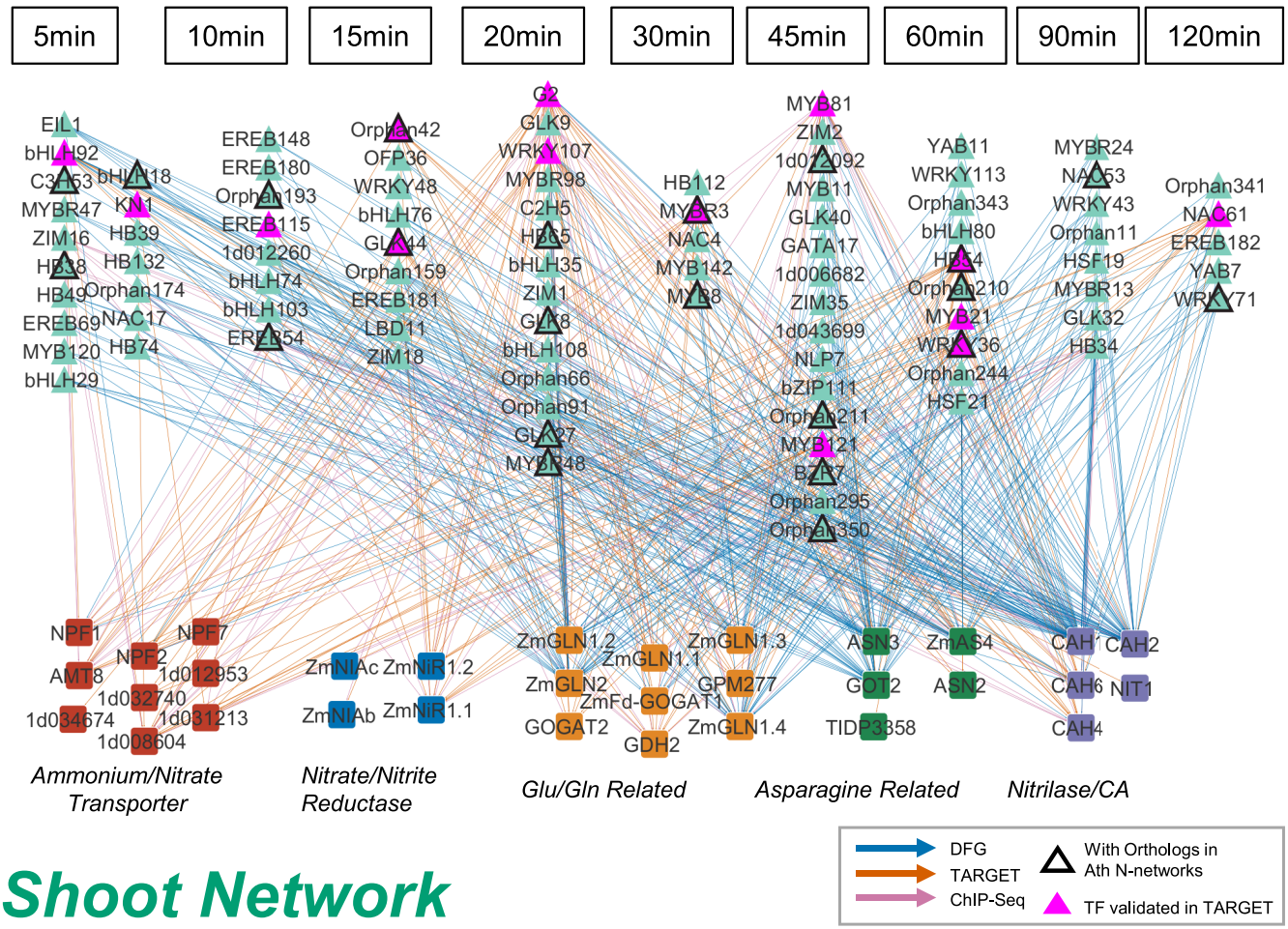


Figure 4. A temporal N-responsive "ensemble" network of the N-uptake/assimilation pathway in maize shoots. In this temporally resolved N-response network, edges connect NxTime TFs (organized by their JIT bins) with genes in the nitrogen-uptake/assimilation pathway. The "ensemble" network consists of 92 TFs. "This "ensemble" network exclusively includes high-confidence TF–target edges from the DFG network (blue), validated TF–target edges of 14/23 maize TF validated by the TARGET TF assay (vermilion), and TF–target edges confirmed by ChIP-seq for 10 TFs (reddish purple) (Tu et al. 2020). Maize TFs with orthologs in the Arabidopsis yeast-one-hybrid TF–DNA binding dataset (Gaudinier et al. 2018) or the Arabidopsis shoot NxTime gene list (Varala et al. 2018) are labeled with a black border line. For edges with multiple support (DFG, TARGET, ChIP), see Supplementary Table S17.

ZmKN1 influenced network of N-response genes share GO terms with targets identified using *kn1* mutants *in planta* (Bolduc et al. 2012) (Supplementary Table S19), such as "cell population proliferation" (Fig. 5B). Interestingly, GO term analysis reveals a novel involvement of ZmKN1 in nitrogen metabolism (Fig. 5B), extending its known role as a master regulator in shoot meristems (Smith et al. 1992; Kerstetter et al. 1997).

Finally, to compare the ZmKN1 binding data in maize cells and *in planta*, we performed additional TF–target validation by performing ChIP-seq on ZmKN1-tagged with a small biotin ligase recognition peptide (BLRP) overexpressed in maize protoplasts (Lau and Bergmann 2015; Matsuda et al. 2017) (see Materials and methods). Our new ChIP-seq data showed a significant overlap between the ZmKN1 target genes bound in leaf protoplast (this study) and in mature ear tissue (Bolduc et al. 2012) (Supplementary Fig. S16, A and B). Similar cis-binding motifs for ZmKN1 are detected both in maize leaf protoplasts (our study) and in ear tissue (Bolduc et al. 2012) (Supplementary Fig. S16, C and D). Overall, these results suggest that protoplast-based overexpression of ZmKN1 accurately reflects the gene binding and regulatory activities occurring in maize *in planta*.

Importantly, the 23 TF TARGET dataset of maize can also be used for validating various other networks. Utilizing AUPR analysis, we used the TF–target data for these 23 maize TFs to perform AUPR analysis of 45 GRNs from previously published research (Zhou et al. 2020). Our analysis identified 25 maize GRNs whose performance metrics surpassed random networks ($P < 0.05$, permutation test with 1,000 repeats), including 10 GRNs with *P*-values smaller than 0.001 (Supplementary Fig. S17). These 10 maize GRNs include various tissues and genotypes that are worthy of further study. Thus, the addition of our TF–target-validated TARGET for 23 maize TFs dataset enables AUPR analysis which is automated in the ConnecTF platform (Brooks et al. 2021) will facilitate the validation of externally submitted networks in the plant community.

Model-to-crop conserved N-response genes enhance machine learning of genes-of-importance to NUE

In the above sections, we focused on maize to infer and validate a temporal GRNs involved in N-responses in this major crop. In this section, we elaborate on how our model-to-crop analysis showed that the use of orthologous NxTime genes in machine learning significantly enhanced the predictive power of gene models for NUE, as in Cheng et al. (2021). We first set out to identify the set of orthologous NxTime genes shared between maize and Arabidopsis. In our

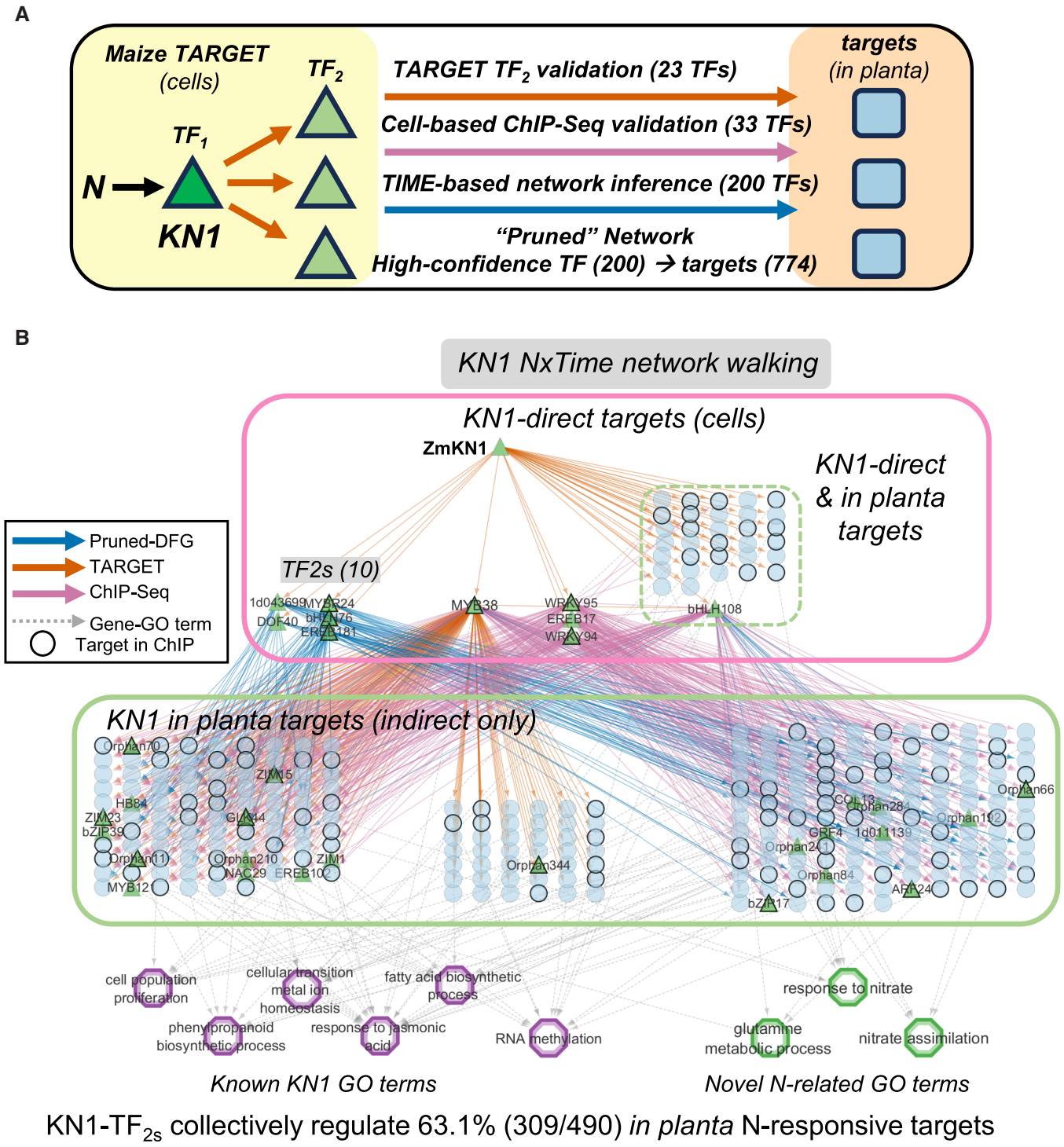


Figure 5. "Network Walking" uncovers a novel role for ZmKN1 in maize nitrogen signaling. **A)** A diagram depicting the Network Walking method that links ZmKN1's direct targets to indirect in planta targets. For detailed explanation of Network Walking method, see Brooks et al. (2019, 2021) and Huang et al. (2023). **B)** Hierarchical structure of the ZmKN1 regulatory network for N-responsive genes. Direct targets of ZmKN1 validated using the TARGET system are encompassed within a pink rectangle. The dashed green rectangle within the pink rectangle denotes ZmKN1 direct targets validated in TARGET and *in planta kn1* mutant (Bolduc et al. 2012). ZmKN1 directly regulates 10 TF_{2s}, which provide direct links to ZmKN1's indirect targets (encompassed in green rectangle). These 10 TF_{2s} which link ZmKN1 to its indirect targets are validated by: (i) the maize TARGET assays (vermilion edges) (Fig. 2), (ii) high-confidence edges in the "pruned" DFG network (blue edges) (Fig. 3), or (iii) ChIP-seq (reddish purple edges) (Tu et al. 2020). The indirect targets of ZmKN1, encompassed within in the green rectangle, consist of the overlap between in planta misregulated genes from the *kn1* mutant (Bolduc et al. 2012) and NxTime genes in shoots. Genes are linked to their GO annotations by gray edges. GO terms found for genes misregulated in *kn1* maize mutants in planta appear as purple hexagons, while novel nitrogen pathway-related GO terms discovered for ZmKN1 in this study are shown in green hexagons.

proof-of-principle example, we focused on the N-GRN subnetwork of TF predicted to regulate multiple genes in the N-uptake/assimilation pathway. We found that the shoot ensemble N-GRN for maize (Fig. 4) includes 92 TFs, 22 of which have orthologous TFs found in the Arabidopsis predicted networks (Gaudinier et al. 2018; Varala et al. 2018) (Fig. 4; Note, 22 validated maize TFs are denoted as

triangles with black boarder). With the discovery of orthologous TFs targeting genes in the N-pathway model-to-crop, we expanded our cross-species analysis to *all* the NxTime genes shared between maize and Arabidopsis. Through gene orthology, we found 353 shoot NxTime genes (34 TFs) and 406 root NxTime genes (41 TFs) in maize have orthologs in Arabidopsis that are also NxTime genes (Supplementary Tables S20 and S21), which we hereafter refer to as "conserved NxTime genes."

In our prior study, we found that evolutionary conserved model-to-crop genes could significantly enhance the accuracy of machine learning of gene models whose expression levels predict phenotypic outcomes in the field (Cheng et al. 2021). Therefore, in our current study, we next investigated whether using the conserved NxTime gene sets from our N-response time-course experiments in maize and Arabidopsis could enhance machine learning predictions of NUE phenotypes in the field. For a detailed description of our XGBoost analysis, see Materials and methods section: XGBoost machine learning prediction models of NUE trait from gene expression (Cheng et al. 2021). Our methods include feature selection, cross-validation, hyperparameter optimization, and application to an independent dataset using data from Li et al. (2023). These steps demonstrate that our modeling pipeline effectively mitigates overfitting and ensures the robustness of the XGBoost gene models learned. For our machine learning analysis, we used 331 out of 353 conserved maize NxTime genes identified in our study which were expressed in the Cheng et al. maize dataset (Cheng et al. 2021) along with the phenotypes (data downloaded from https://osf.io/avjph/).

To reduce dimensionality and avoid overfitting, we tested whether using the 331 conserved NxTime genes was a biologically principled way to enhance our XGBoost predictions of NUE. For our machine learning, we employed the XGBoost algorithm (Chen and Guestrin 2016) and used the match field expression data of these selected maize genes to predict NUE phenotypes determined in the field from (Cheng et al. 2021). We found that the 331 model-to-crop conserved NxTime genes significantly outperformed the same number of randomly selected NxTime genes (*P*-value: 1.25E−02) or expressed genes (*P*-value: 5.49E−03) in predicting field NUE phenotypes (Fig. 6). Each gene within this conserved NxTime set of 331 maize genes was assigned a feature importance score, termed NUEscore, to quantify its predictive value of the traits across the 16 maize genotypes (Supplementary Table S20). Further, we examined the correlation between the eigengenes of these 331 conserved NxTime maize genes and numerous phenotypic field traits (data downloaded from https://osf.io/avjph/) (Cheng et al. 2021). We found significant associations exclusively with NUE traits, rather than grain yield or biomass (Supplementary Fig. S18). These findings demonstrate the potential of the model-to-crop conserved NxTime genes identified in seedlings as reliable features for predicting NUE in mature maize grown under distinct N-field conditions.

NUE Regulon: integration of gene networks with machine learning scores for NUE trait predictions

To further enhance our model-to-crop GRN approach and identify conserved genes for NUE improvement, in the section below, we layout the NUE Regulon pipeline, which integrates: (i) high-confidence GRNs, (ii) gene orthology, and (iii) machine learning to learn network modules-of-importance to a trait. In this study, we focus on NUE traits, but the approach and pipeline can be applied to any species or trait of interest (see Graphical Abstract).

Step 1: NUE Regulons: integration of GRN and machine learning knowledge

This first step aims to integrate network knowledge from the GRNs with machine learning of genes of importance to NUE. The goal is to identify the TFs whose target genes are collectively ranked the highest for importance to NUE. To this end, we used the ensemble N-GRN for maize shoots (Supplementary Data Set 2) and incorporated the XGBoost NUE scores of the 331 NxTime genes conserved model-to-crop. We initiated this process by isolating a subnetwork within the ensemble N-GRN (Fig. 7A). The NUE-centric subnetwork includes the TF and their target genes (including TF_2 targets) that are NxTime genes conserved across model-to-crop (green nodes, respectively in Fig. 7A).

Step 2: NUE Regulon score: ranking network modules-of-importance to NUE

Next, we calculated the NUE Regulon score for each TF in the subnetwork. This NUE Regulon score represents a weighted importance to NUE (from XGBoost) of each TF and their target genes from the ensemble GRN (Fig. 7A). This NUE Regulon score is calculated from the NUE score for the TF times the aggregated NUE score of its NxTime target genes in the ensemble N-GRN (Fig. 7A). Ranking the NUE Regulon scores this way allowed us to identify TF hubs within NUE-centric subnetwork, distinguished not just by their network connectivity, but also by their high predictive value for targeting genes affecting NUE traits (Supplementary Table S22). In our prototype, ZmMYBR3 and ZmMYB34 emerged as the top 2 TFs relevant to maize NUE based on NUE Regulon score (Fig. 7B). This ranking is also supported by experimental data: (i) the direct targets of ZmMYBR3 and ZmMYB34 validated in the cell-based TARGET assay were the top-ranked TFs for specificity to regulating NxTime genes in maize shoots (Fig. 2B) and (ii) in our gene orthology analysis, each of these maize MYB TFs is orthologous to Arabidopsis AtDIV1, for which loss-of-function mutants have been shown to have enhanced NUE, specifically under low nitrogen conditions (Cheng et al. 2021).

Step 3: Cross-species NUE Regulons: TF–target modules conserved model-to-crop

Given the many-to-one orthologous relationship between maize ZmMYB34, ZmMYBR3 and Arabidopsis AtDIV1, we first tested the potential redundancy of these 2 maize MYBs functions. We did this experimentally by comparing the direct regulated target genes of ZmMYB34 and ZmMYBR3 derived from TARGET TF-perturbation assays performed in TF-transfected maize cells. This analysis revealed a significant overlap of the ZmMYB34 and ZmMYBR3 validated target genes (*P*-value < 2.2E-16, Fisher's exact test) (Supplementary Fig. S19), suggesting a functional redundancy between these 2 MYB TFs in maize. The intersection of target genes for ZmMYB34 and ZmMYBR3 ($n = 2{,}957$) called ZmMYB34/R3 was used as their validated targets for the ensuing comparison to targets of their orthologous Arabidopsis TF AtDIV1 targets validated in the TARGET assay in Arabidopsis leaf protoplasts (Supplementary Table S23). We found a significant overlap of orthogroups targets (*P*-value < 2.2E−16, Fisher's exact test) regulated by maize ZmMYB34/R3 and Arabidopsis ortholog AtDIV1 (Fig. 8A). Importantly, this finding indicates a conservation of orthologous TF–target regulatory function across 160 million years of evolutionary divergence between maize and Arabidopsis.

It is noteworthy that not every orthologous validated target shared between ZmMYB34/R3 and AtDIV1 is essential for the NUE trait. By overlapping the orthologous targets shared between ZmMYB34/R3 and AtDIV1 with the conserved NxTime genes whose

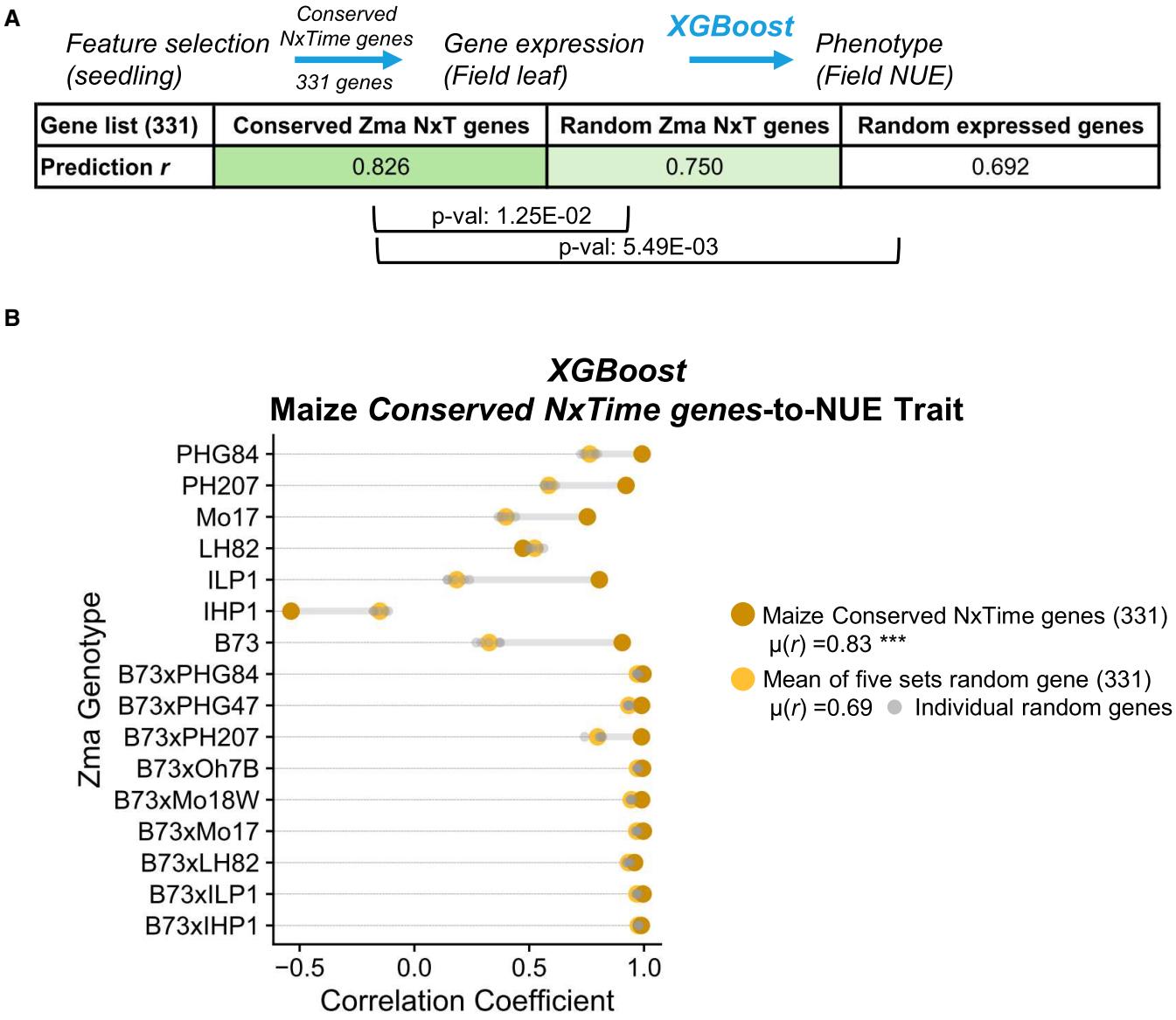


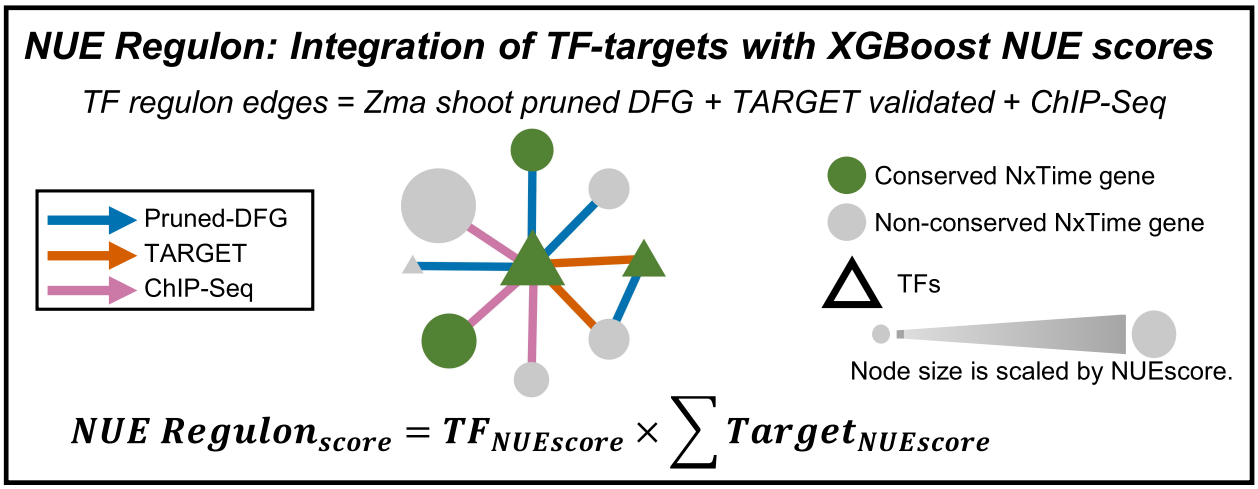
Figure 6. Model-to-crop conserved NxTime genes are significantly more effective in predicting NUE outcomes maize field samples. Out of 353 conserved NxTime genes, 331 genes are present in the training and testing dataset from Cheng et al. 2021. Using a selection of 331 conserved NxTime shoot genes, XGBoost was used to predict the maize NUE traits using RNA-seq and NUE data from 16 maize lines grown under −N/+N field conditions (Cheng et al. 2021). To evaluate the power of the XGBoost model to accurately predict the NUE trait based on the mRNA levels of the XGBoost gene models learned. To this end, Pearson's correlation (*r*) was calculated between XGBoost predicted NUE value vs. actual NUE value (in left-out samples). **A)** The *r* value was calculated for XGBoost models learned: (i) using the 331 out of 353 NxTime genes conserved model-to-crop, (ii) using an equivalent number of random maize NxTime genes, and (iii) the equivalent number of randomly expressed maize genes. *P*-values derived from a Wilcoxon signed-rank test show that the NxTime genes conserved model-to-crop are significantly more effective at predicting NUE in field-grown maize. **B)** Using genes from the 331 conserved NxTime genes in XGBoost to learn gene models predictive of NUE, significantly improves NUE predictions, compared with using an equivalent number of random genes in maize. ***: *P*-value < 0.001 (Wilcoxon's signed-rank test).

expression effectively predicted NUE outcomes in the field, we identified a specific subset of 23 Arabidopsis and 24 maize genes within a cross-species NUE Regulon (Fig. 8A). This conserved NUE Regulon includes key nitrogen pathway metabolic enzymes, such as nitrite reductase 1 (NIR1) and NADH-dependent glutamate synthase 1 (GLT1), as well as 3 TF_{2s} including validated TF regulators of the N-response in Arabidopsis such as HHO6 (Varala et al. 2018; Ueda et al. 2020) and CDF1 (Varala et al. 2018) (Fig. 8A).

Step 4: Evaluation of model-to-crop NUE Regulon in predicting NUE in field

Given the conservation of this NUE Regulon from model-to-crop (ZmMYB34/R3 and AtDIV1), we evaluated its potential to predict NUE traits in field-grown maize using the XGBoost pipeline. Impressively, the predictive performance of the cross-species NUE Regulon (ZmMYB34/R3 and AtDIV1) target genes significantly surpassed that of an equivalent number of randomly conserved NxTime genes in both maize (*P*-value: 1.5E−05—Wilcoxon's signed-rank test) (Fig. 8B) and Arabidopsis (*P*-value: 7.9E−04—Wilcoxon's signed-rank test) (Fig. 8C). We found that maize B73 hybrids generally have better predictions than inbreds, as was also observed in Cheng et al. (2021). We hypothesize that maize hybrids have adapted to the high nitrogen fertilizer condition used in breeding; thus, the gene response to nitrogen can better predict the NUE phenotype. From the XGBoost models, we retrieved the NUEscore for the conserved regulon genes and

A



B

Rank	Zma TF	TF score	Target score	# of genes	NUE Regulon score	Ath ortholog
1	ZmMYBR3	25.0	632.4	86	15,796	AtDIV1
2	ZmMYB34	21.1	655.0	113	13,851	AtDIV1
3	ZmHB54	22.8	406.7	127	9,255	AtHB5 / AtHB6
4	ZmMYB56	5.9	616.3	87	3,629	AtMYB48
5	ZmHB38	14.0	158.1	35	2,217	AtBEL1
6	ZmNLP3	2.0	686.6	73	1,397	AtNLP8
...				...		
34	See Supplemental Table S22 for full table					

TARGET TF-target validation (See Figure 7A)

* TFscore and Targetscore are based on the XGBoost derived feature importance scores.

Figure 7. Maize NUE Regulon integrates network connectivity with machine learning scores to prioritize TFs of importance to NUE. **A**) NUE Regulon's architecture integrates TF–target edges from pruned high-confidence DFG time-based networks (Fig. 3 and Supplementary Table S14), maize TF–targets validated in maize cells via TARGET (Fig. 2 and Supplementary Data Set 1), and ChIP-seq data for 2 maize TFs (Tu et al. 2020). Using these network edges, the NUE Regulon score for each TF was calculated as the product of the TF's own NUE score times the sum of all its targets NUE scores. **B**) TFs within NUE-centric subnetwork are ranked by their NUE Regulon scores (Supplementary Table S22). The Arabidopsis orthologs of these maize TFs are listed in the last column. The top-rank maize TFs, ZmMYBR3, ZmMYB34, and their ortholog AtDIV1 have been validated for TF–target using the maize cell-based TARGET system (Supplementary Data Set 1 and Table S23).

visualized it on an orthology mapping plot (Fig. 8A and Supplementary Tables S24 and S25). To further confirm the predictive power of the ZmMYB34/R3 regulon genes, we tested them on a different maize dataset which including expression and field phenotypes from 137 Chinese elite inbred lines (Li et al. 2023). It is known that maize ear development is significantly affected by nitrogen availability (Liao et al. 2012; Gheith et al. 2022). Using random forest models, we found the 24 genes in the ZmMYB34/R3 NUE Regulon show significantly better prediction performance for maize ear diameter and ear length, compared with an equal number of random conserved NxTime genes (Supplementary Fig. S20).

In conclusion, our cross-species NUE Regulon analysis pipeline identified a precise set of orthologous TF–target genes conserved model-to-crop whose expression levels significantly improved machine learning of NUE trait prediction in Arabidopsis and maize grown in the field. This finding reveals that bridging crucial model-to-crop regulatory links between model-to-crop can enhance our ability to identify genes-of-importance to NUE. Information gained from this NUE Regulon can be applied in molecular breeding or transgenesis to improve NUE in crops.

Discussion

In this study, we outline a combined computational and experimental strategy to identify genes-of-importance to an agronomic trait. In a systems biology approach, the NUE Regulon analysis pipeline integrates transcriptomics and traits by combining network inference, machine learning, and model-to-crop translation. Specifically, we show how a cross-species NUE Regulon significantly enhances machine learning of gene network regulons-of-importance to NUE in a model, Arabidopsis, and in field-grown maize plants. Impressively, our model-to-crop study indicates a conservation of TF–target regulatory module function across 160 million years of evolutionary divergence between maize and Arabidopsis. More broadly, this model-to-crop analysis approach can be applied to any species or trait-of-interest. Below, we discuss how transferring

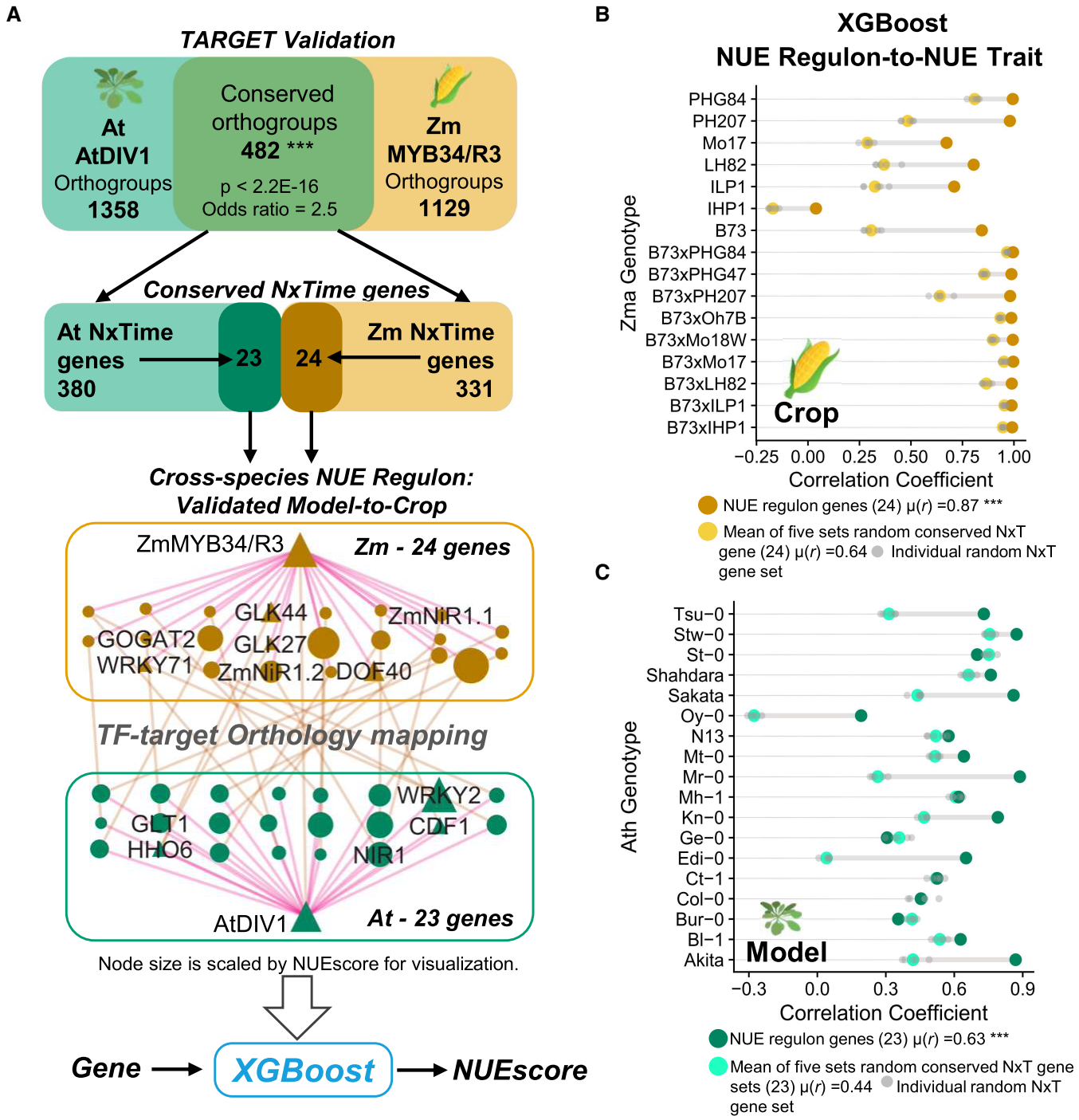


Figure 8. A model-to-crop conserved NUE Regulon significantly enhances predictions of NUE in both maize and Arabidopsis. **A)** Discovery of the conserved NUE Regulon shows that the orthologous TFs in maize (ZmMYB34/R3) and Arabidopsis (AtDIV1) and share a significant number of orthologous targets, validated by the plant cell-based TARGET assay (Supplementary Data Set 1 and Table S23). This TF network regulon includes 24 gene targets in maize and 23 orthologous gene targets in Arabidopsis, identified by overlapping genes in the conserved orthogroups with Arabidopsis or maize NxTime genes. Using genes from these cross-species NUE Regulon in XGBoost to learn gene models predictive of NUE significantly improves NUE predictions, compared with using an equivalent number of conserved NxTime genes in **B)** maize and **C)** Arabidopsis. *** P-value < 0.001 (Wilcoxon's signed-rank test).

integrated knowledge from data-rich experimental models-to-crops presents significant opportunities, as well as future challenges.

Model-to-crop translation of temporal regulatory networks controlling NUE

This study exploits time, a largely unexplored dimension of the N-GRN response, especially in maize. As causality moves forward in time, fine-scale time-series RNA-seq data are a valuable source to derive predictive networks, as shown in Arabidopsis (Krouk et al. 2010, 2013; Varala et al. 2018). Here, we implement and validate a time-inferred causal GRN in maize and compare it to a parallel study in the model Arabidopsis (Varala et al. 2018) to reveal N-regulatory temporal networks and network regulons conserved model-to-crop.

To do this, we first deciphered the temporal control of nitrogen-responsive gene regulatory networks (N-GRNs) in maize using

fine-scale time-course RNA-seq experiments and identified their temporal cascade of regulation using JIT analysis (see Materials and methods), as in Arabidopsis (Varala et al. 2018). This uncovered a temporally resolved nitrogen-signaling cascade of gene functions conserved between model-to-crop (Fig. 1).

Further leveraging our time-course N-responsive genes, we employed a time-based network inference method called DFG (Krouk et al. 2010, 2013; Varala et al. 2018; Brooks et al. 2019) to construct causal N-GRNs in maize shoots and roots. To experimentally validate these inferred networks, we used TARGET, a medium-throughput plant cell-based TF-perturbation assay (Fig. 2). The TARGET assay validated TF–target direct regulatory edges for 23 maize TFs whose targets encompass ~85 to 90% of the early-to-late temporally N-responsive genes *in planta* (Fig. 2D). These validated TF–target data were also used in a precision/recall AUPR analysis to prune the accuracy of predicted TF–target edges for ~200 TFs and 700 target genes in our N-GRNs (Fig. 3). We showed that this pruned N-GRN includes the validated and predicted TF regulators of 32 genes in the N-uptake/assimilation pathway in maize shoots (Fig. 4). Our integrating of these high-confidence TF–target edges identified in maize cells (via TARGET) with *in planta* data of a maize TF mutant in leaf development (Bolduc et al. 2012) led to the discovery of a novel role for the well-studied maize ZmKN1 TF in nitrogen signaling, using an approach called "Network Walking," described in Brooks et al. (2019, 2021) (Fig. 5).

Next, we aimed to identify how TFs network regulons in the pruned GRN might influence NUE outcomes. To this end, we identified TF NUE Regulons (Fig. 7), by integrating our temporal N-response maize GRNs with NUE phenotype field data from Cheng et al. (2021) using XGBoost machine learning analysis (Chen and Guestrin 2016), applied to maize as described. In a model-to-crop analysis, we found that conserved NxTime genes between maize and Arabidopsis can reliably predict NUE in plants. Moreover, by integrating network structures with NUE scores, we identified and validated a conserved maize–Arabidopsis NUE network regulon (ZmMYB34/R3—AtDIV1) of 24/23 genes whose expression levels significantly enhance the ability to predict NUE traits from RNA in left-out samples for both model-and-crop (Figs. 7 and 8).

Exploiting time-series data to enhance regulatory network inference and learn early N-signaling events in maize

Prior studies of nitrogen-signaling responses in maize have primarily focused on single time points (Ge et al. 2020, 5; Wang et al. 2020b; Buoso et al. 2021a; Cheng et al. 2021). Our analysis takes advantage of time-series data in maize to infer regulatory networks and perform a model-to-crop analysis. Unlike steady-state experiments that take a snapshot the transcriptome at a single point, time-course experiments find changes in gene expression throughout development or treatment, providing a deeper insight into gene regulatory dynamics (Storey et al. 2005; Aryee et al. 2009; Krouk et al. 2013). In addition to enhancing predictions of TF–target causality, measurements across multiple time points act as replicates, significantly boosting the statistical power for differential expressed gene analysis (Tai and Speed 2009; Sun et al. 2016).

In our present study, we used a spline model to identify NxTime genes (Fig. 1C). The spline model is known for its flexibility and efficiency in detecting nonlinear gene expression changes (Coffey et al. 2014; Spies et al. 2017). Since all time points are used in the differential expression (DE) testing, instead of treated versus control at individual time points, spline modeling handles noisy data and increases confidence in detecting DE genes (Luan and Li 2003; Krouk et al. 2010; Varala et al. 2018; Alvarez et al. 2021; Shanks et al. 2024). Recently, spline modeling has also been applied in detecting DE genes from temporal single-cell transcriptomic data in animals (Van den Berge et al. 2020; Fan et al. 2024). Our time-course experiment enabled us to pinpoint the first time-point that maize JIT genes that respond to nitrogen above a 2-fold threshold (Fig. 1C). Such JIT gene sets responding to the N-signal in maize were also observed in a parallel study in Arabidopsis (Varala et al. 2018). By comparison, we determined that temporal N-responsive conserved across model-to-crop shares biological processes (Figs. 1, D and E). In addition to identifying a large number of early- and dynamic N-response genes that were missed in steady-state studies of maize, we identified 353 NxTime genes in the shoot and 406 NxTime genes in the root conserved across model-to-crop.

These findings suggest a conserved plant mechanism for temporal transcriptional control of N-metabolic genes in plants in response to nutrient sensing, analogous to systems seen in *E. coli* and yeast (McAdams and Shapiro 2003; Zaslaver et al. 2004; Chechik et al. 2008). In other studies, in addition to the spline model, other promising methods for time-course transcriptome analysis include Granger causality (Heerah et al. 2021), hidden Markov model (Leng et al. 2015), and deep learning (Yuan and Bar-Joseph 2021).

GRN validation and refinement: precision/recall analysis and network pruning

One major bottleneck in transferring knowledge from model-to-crop is the validation of TF–target edges in the GRNs, especially in crops. One essential aspect of network inference studies is the incorporation of validated TF-to-target data which enables network refinement and "pruning" of high-confidence regulatory edges in a GRN.

Herein, we have enabled such GRN validations in a crop by adapting a medium-throughput cell-based TF-perturbation assay called TARGET to maize, from the protocols initially developed in Arabidopsis (Bargmann et al. 2013; Brooks et al. 2019). The TARGET assay enabled us to rapidly validate direct regulated TF–targets in isolated maize cells. Herein, we show that the genome-wide targets of 23 maize TFs validated in the maize cell-based TARGET system capture ~90% of the NxTime-responsive genes in maize shoots and roots (Fig. 2). Compared with ChIP-seq, and other cell-based TF-perturbation assays (Zhu et al. 2024), the TARGET assay identifies directly regulated TF–targets (Bargmann et al. 2013) (see Supplementary Fig. S10). This approach to identify direct regulated targets of a TF (e.g. TF-GR fusion + CHX) has also been used extensively *in planta* (Yamaguchi et al. 2015).

Importantly, the TARGET assay also provides an evidence-based method to validate and prune the GRNs. Instead of arbitrarily selecting the top 1% or 10% as the high-confidence TF–targets in GRNs, we used AUPR to guide where to prune the network (Fig. 3). AUPR is suggested to be the optimal performance measurement of a network, as it emphasizes true-positive predictions and ignores true negatives, which could overwhelm the GRN prediction (Aalto et al. 2020; Pratapa et al. 2020). In our study, we used the validated genome-wide targets of 23 maize TFs to prune our maize GRN for high-confidence edges connecting ~200 TFs to ~700 target genes (Fig. 3). Given that many agronomic traits are governed by such complex networks, having a high-quality GRN is crucial for identifying key genes that control traits, including NUE (Hirel et al. 2007; Gao et al. 2022). The recently proposed "omnigenic" model hypothesized that all genes expressed in a cell contribute to controlling a complex trait (Boyle et al. 2017; Mathieson 2021). We observed that the temporal nitrogen response GRN controlling nitrogen uptake/metabolism genes in

maize is highly interconnected and regulated by hundreds of TFs (Fig. 4 and Supplementary Fig. S14).

Model-to-crop conservation of GRN network regulons: NUE phenotype as a prototype working example

Previous studies in both plant (Obertello et al. 2015; Du et al. 2020) and animal research (Yu et al. 2012; Mueller et al. 2017; Möller et al. 2022) have highlighted the gene module conservation between model and nonmodel organism. However, the link between network module-to-phenotype remains largely unexplored (Langfelder and Horvath 2008). Our framework to rank the importance of regulatory network modules to a trait combines cross-species comparison and machine learning. Our rationale is if a conserved GRN network regulon is validated in diverse species, such as Arabidopsis and maize, which are separated by 160 million years (Kumar et al. 2017), then this regulon is highly likely to be important to a trait.

To test this hypothesis, we focused on a subnetwork regulon (NUE Regulon) for the NUE trait. To enhance our machine learning of genes-of-importance to NUE, we first reduced the dimensionality of genes used in our models to the 353 evolutionarily conserved maize–Arabidopsis NxTime gene sets (Fig. 6), as in Cheng et al. (2021). Using this evolutionarily conserved NxTime gene set, we showed the model-to-crop N-response genes selected at the seedling stage (Fig. 6A) can predict the NUE phenotype at adult stages in the field (Fig. 6B). For the NUE Regulon approach, we calculated the "NUE Regulon score" by integrating the machine learning-derived gene importance scores of the TF and its targets within the network context (Fig. 7). Moreover, we validated the top TFs regulation by TARGET in model (AtDIV1) and crop (ZmMYB34/R3) using the cell-based TARGET assay (Fig. 8). This validated that a model-to-crop conserved ZmMYB34/R3 regulon for NUE, which comprises only 24 genes, can accurately predict field NUE and ear phenotypes in field-grown maize (Fig. 8 and Supplementary Fig. S20). Using this precise gene set of TF regulon genes-of-importance to predict field outcomes in maize seedlings can facilitate improvements in NUE through molecular breeding programs, transgenics, or genome editing.

Challenges and limitations in model-to-crop translation

While our model-to-crop strategy effectively predicts the NUE trait, it also presents several challenges and limitations. First, our cross-species NUE Regulon relies on the orthology relationships, which are often difficult to infer. In our study, we applied OrthoFinder2, a stringent tree-based inference method for orthology mapping (Emms and Kelly 2019). Benchmark shows the OrthoFinder2 has high accuracy and good overall performance (Nevers et al. 2022). However, other large-scale ortholog databases may offer a more comprehensive list, such as InParanoiDB (Persson and Sonnhammer 2023), Phytozome (Goodstein et al. 2012), Orthologous Matrix (Altenhoff et al. 2024), and Ensembl Genomes (Yates et al. 2022). Meanwhile, the stringent tree-based orthology method may miss true orthologs that can be detected by more sensitive BLAST-based methods (Chen et al. 2007; Moreno-Hagelsieb and Latimer 2008; Obertello et al. 2015). Future studies could consider incorporating multiple orthology inference methods to ensure a more inclusive orthology identification.

Second, in this study, we used gene expression to limit the ortholog search, ensuring that maize–Arabidopsis orthologs had to be NxTime response genes in both model-to-crop. This approach greatly reduced the dimensionality for the machine learning by focusing on the conserved genes. However, it excluded the maize-specific and C4-specific genes that do not have orthologs in the Arabidopsis C3 plant. Also, genes with similar sequences (e.g. orthologs) may not have the identical functions. Genome duplication and gene family expansion after speciation may create orthologs with different functions (Patel et al. 2012; Das et al. 2016). For example, there may be TFs from the same family, but are not necessarily orthologs that regulate the same set of genes. To address this, future studies should incorporate maize-specific genes and investigate the functional divergence among orthologs, enhancing the power of model-to-crop knowledge transfer.

Overall, our work integrates GRN analysis with machine learning and gene orthology to identify regulons conserved model-to-crop species. Moreover, we validated the orthologous TFs regulons across species in plant cells. While we focused on the NUE trait, this strategy can be applied to any other crop or traits by utilizing information from any model species. Importantly, the selection for or modification of such NUE Regulons can be used to enhance NUE in crops, improving yield and mitigating N-fertilizer contribution to greenhouse gas emission.

Materials and methods

Plant material and nitrogen treatment

Z. mays (B73) seeds were obtained from US National Plant Germplasm System (https://npgsweb.ars-grin.gov/gringlobal/accessiondetail?accid=PI+550473). On d 1, seeds were soaked in sterilized water overnight. On the next day, seeds were sterilized in a glass beaker with 50% commercial bleach (CLOROX) for 20 min. Then, the seeds were rinsed 5 times with sterilized water, each time for 20 min. Six sterilized seeds were paper rolled on a double-layered, prerinsed 60 cm × 10 cm paper towel. Seeds were placed 2 cm below the paper edge. The paper rolls with seeds were covered with foil paper and stored in a plastic container filled with 200 mL water. Plants were checked daily.

All plants were grown in a PERCIVAL growth chamber at New York University [16-h light–8-h dark, 27 °C at light (150/μmol^2/s) 24 °C at dark cycle, 70% relative humidity]. Water in each container was replaced by nitrate-free MS hydroponic medium (Caisson Labs-MSP10-50LT) on d 6. On d 8, the germinated seedlings were taken out of the paper rolls and endosperms were removed with a razor blade by hand. After the removal of the endosperms, seedlings were grown in the fresh nitrate-free MS hydroponic medium. All plants were handled between 10 AM and 12 PM Eastern Standard Time, 2 h after the light cycle started. The seedlings were starved from nitrogen for 2 d. On the day of the sampling (d 10), 2 h after the start of the light period (10 AM), seedlings were transferred to the fresh hydroponic medium containing either the nitrogen in MS medium (20 mM KNO_3 + 20 mM NH_4NO_3) or 20 mM KCl and harvested at time intervals 0, 5, 10, 15, 20, 30, 45, 60, 90, and 120 min. Ten percent of nitrogen in the MS medium was labeled with the ^{15}N isotope (Cambridge Isotope Lab—NLM-390-1). Shoots and roots from 3 seedlings were combined as 1 biological replicate. Three biological replicates were collected per time point per treatment. They were immediately flash-frozen in liquid nitrogen and stored in an −80 °C freezer. The ground tissues were sent to UC Davis Stable Isotope Facility for ^{15}N measurement. The ^{15}N measurements in shoots and roots were plotted over time, and a simple linear regression model was applied using the "lm()" function in R. To calculate the correlation between gene expression and ^{15}N levels, Pearson's correlation coefficients were calculated using the "cor.test()" function in R. *P*-values were adjusted for multiple testing using the "p.adjust(method = 'BH')" function.

RNA extraction and library preparation for RNA sequencing

About 100 mg of finely ground tissues was used for RNA extraction using the Zymo Direct-zol RNA Miniprep Kits (R2052). The RNA quality was checked on a Nanodrop and Agilent TapeStation (5067-5576). The Lexogen QuantSeq 3′ mRNA-Seq Library Prep Kit (015.2 × 96) with UMI (081.96) was used for the mRNA library preparation. The libraries were pooled and sequenced on the Illumina NextSeq 500 platform at NYU-CGSB Genomics Core using the HighOutput 1 × 75 bp mode.

The UMI-incorporated RNA-seq libraries were analyzed following Lexogen's guidance (https://www.lexogen.com/quantseq-3mrna-sequencing/). The reads' UMI were extracted from raw fastq files using "extract" command from UMI-tools v1.1.1 (Smith et al. 2017). The fastq files were trimmed by fastp 0.21.0 (Chen et al. 2018). The clean fastq files were aligned to maize genome version 4 downloaded from Ensembl Plant (Kersey et al. 2018) using STAR 2.7.6a (Dobin et al. 2013). The identical UMI reads were deduplicated using "dedup" command from UMI-tools v1.1.1 (Smith et al. 2017). The read counts matrix was generated by featureCounts v2.0.1 (Liao et al. 2014) from the deduplicated bam files. All steps were organized using Snakemake 5.31.1 = workflow manager (Köster and Rahmann 2012) and Slurm job scheduler on the NYU High Performance Computing platform.

Identifying NxTime genes fitting a spline model

To identify the high-confidence nitrogen-responsive genes (referred to as NxTime genes), we fit the transcriptome data (+N vs −N) into a cubic spline model as previously described (Varala et al. 2018) with modifications (Fig. 1C). The data were modeled with a df equal to 5. Then, we used the generalized likelihood ratio test (lrt) in edge package (Storey et al. 2005) in R to test the NxTime genes whose N-response curve of mRNA levels differs significantly from −N controls (Fig. 1C). The bootstrap method with 1,000 iterations was applied in the lrt test. We set the cutoff for significant NxTime genes in shoot and root to include ~200 TFs. This cutoff prevents overfitting in the network inference step (avoiding too many regulators) and maintains consistency with the Arabidopsis time-course data. We identified 2,732 NxTime genes in shoot (q-value < 0.05) and 2,294 NxTime genes in root (q-value < 1E-5).

Identifying nitrogen pathway genes and cross annotation model-to-crop

We conducted a comprehensive search across 3 databases, including MaizeGDB (Andorf et al. 2015), UniProt (The UniProt Consortium 2019), and KEGG (Kanehisa et al. 2023), for key enzymes in the nitrogen metabolism pathways as listed in Supplementary Table S4. To avoid missing maize genes due to the lack of annotation, we performed reverse BLASTP search (Obertello et al. 2015) using Diamond (Buchfink et al. 2021) between well-annotated Arabidopsis genes and maize genes. Maize genes with e-value < 1E−20 in both queries (maize to Arabidopsis and Arabidopsis to maize) were annotated with Arabidopsis functions. The gene nomenclature adheres to MaizeGDB and TAIR conventions, except for those cited from published sources (Liu et al. 2022) (Supplementary Tables S26 and S27).

Just-In-Time (JIT) analysis, GO enrichment, and motif enrichment

For the JIT analysis (Fig. 1C), we normalized the expression matrix using "betweenLaneNormalization()" function in EDASeq (Risso et al. 2011). From the normalized expression matrix, we calculated the gene expression fold-change (FC) of + N/−N at each time point. Genes were binned to the JIT categories base on their first time showing |log2FC| > 1. The previously published Arabidopsis NxTime data (Varala et al. 2018) collected from identically treated Arabidopsis seedlings was reprocessed following the same criteria for consistency. For all the GO enrichment analyses, the topGO package (Alexa et al. 2006) with the Maize-GAMER annotation (Wimalanathan et al. 2018) was used for maize and the org.At.tair.db (Carlson et al. 2019) was used for Arabidopsis. The Fisher exact test using "weight01" algorithm (default in topGO) was applied to all the GO term enrichment tests.

Motif enrichment was performed by SEA from MEME suite (Bailey and Grant 2021) using the default setting with the order2 background model calculated from all promoter regions (–bfile). Two cis-motif sources were used. The first was the consensus cis-motif clusters from Brooks et al. (2019), containing 80 (sub)family-level TF motifs. The second was the JASPAR2022 Plant Core motifs (Castro-Mondragon et al. 2022) which includes 656 high-quality plant TF motifs. Motifs with QVALUE < 0.05 and ENR_RATIO > 3 from the SEA outputs were considered as enriched for the gene sets.

Maize TARGET experiments for medium-throughput validation of TF–targets in cells

All maize TARGET experiments listed in Fig. 2 were performed using the p1107 plasmid that contains a TF-GR fusion protein, and selectable marker gene for FACS was described previously (Shanks et al. 2022). The vector is available at the VIB-UGENT PSB plasmid repository with vector ID 25_55 and 25_56.

The maize transfection protocol was adapted from our Arabidopsis TARGET protocol (Bargmann et al. 2013; Brooks et al. 2019) and Dr. Jen Sheen's protocol (https://molbio.mgh.harvard.edu/sheenweb/protocols.html) with some modifications. The maize seeds were soaked overnight and planted in pots filled with Turface MVP. The pots were placed at the growth chamber, under 16-h light/8-h dark diurnal cycle, at temperatures 27 °C and 24 °C, respectively, and 70% humidity for 3 d. After seeds were germinated and about 5 cm above ground, the pots were moved to the dark room at temperature 27 °C and 70% humidity for etiolation for 7 d. On the day of the TARGET experiment, etiolated middle section (~4 cm) of the second leaves was cut into 1-mm pieces and soaked in the cell wall digestion solution (1.5% cellulase RS, 0.3% macerozyme R10 (Yakult Honsha), 0.6 M mannitol, 10 mM MES [pH 5.7], 1 mM $CaCl_2$, 5 mM β-mercaptoethanol, and 0.1% BSA) in a petri dish. The petri dish was vacuum-infiltrated for 30 min and shaken for 95 min at 50 rpm in the dark, followed by 10 min at 70 rpm to fully release cells. Maize protoplast cells were filtered through a 40-μm cell strainer (BD Falcon, USA) and spun down for 2 min at 280 g. The cells were washed with 10 mL buffer (10 mM MES, pH 5.7, 0.6 M mannitol, 10 mM KCl) 3 times. After the final wash, cells were resuspended in the buffer to 1.0 × 10^6 cells/mL and were ready for transfection.

For protoplast transfection, we mixed 150 μL protoplasts (1.5 × 10^5 cells), 40 μg plasmid, and 120 μL buffer in a 2-mm electroporation cuvette (BioRad #1652086). The BioRad Gene Pulser was used for the electroporation with 1 shot and the following condition: 150 V, 200 Ω, 125 μF. After electroporation, 700 μL cold buffer was added to the cuvette. Then, the protoplast mix was transferred to a new round-bottom 2-mL tube and stored at room temperature in the dark overnight. On the next day, protoplasts were treated with either +Nitrogen (20 mM KNO_3 and 20 mM NH_4NO_3) or −Nitrogen (20 mM KCl) for 2 h. The cells were treated with 30 μM CHX for 20 min (to block translation of secondary TF targets

genes) and then a 3-h 10 mM DEX treatment (to induce TF-GR nuclear import). After 3 h, TF vector and control EV-transfected protoplasts were FACS sorted for RFP signals into 150 μL TRI regent for RNA extraction (Zymo, R2061). The RNA quality was checked on the Agilent TapeStation (5067-5576). We used Lexogen QuantSeq 3′ mRNA-Seq Library Prep Kit (Lexogen, 015.2 × 96) for RNASeq library preparation.

Arabidopsis TARGET experiment to identify direct regulated TF–targets

The Arabidopsis TARGET experiments for TFs listed in Fig. 2 were performed using the pBOB11 plasmid, as described previously (Brooks et al. 2019). An EV plasmid (with no TF of interest) was also transfected to be used as a control for identifying differentially expressed genes (DEGs). The protocol for protoplasting and transfection leaf cell was adapted from the Sheen Lab (Yoo et al. 2007). In brief, rosette leaves from ∼4-wk-old soil-grown plants were harvested, sliced with a razor, and incubated in enzyme solution (1% cellulase R10, 0.2% macerozyme R10 [Yakult Honsha], 0.4 M mannitol, 20 mM MES pH 5.7, 20 mM KCl, 10 mM $CaCl_2$, and 0.1% BSA) for 90 min. Cells were strained through a 70-μM filter, collected by centrifugation at 300 g, and washed with W5 buffer (150 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl, 2 mM MES pH 5.7, 5 mM glucose). Cells were collected by centrifugation and transferred to MMg solution (400 mM mannitol, 10 mM $MgCl_2$, 4 mM MES pH 5.7) for transfection. Cells were diluted to a cell density of 6×10^5 cells/mL with MMg and mixed with 20 μL plasmid DNA (20 to 50 μg DNA) in a 2-mL microcentrifuge tube. 120 μL PEG solution (40% PEG4000, 200 mM mannitol, 100 mM $CaCl_2$) was added to the cells and mixed by inversion. After incubating for 2 to 5 min, W5 buffer was added, and the cells were washed twice in W5 buffer. Cells were then transferred to a 48-well plate and incubated overnight with shaking. To start the TARGET experiment, CHX was added to a final concentration of 35 μM for 20 min, before the addition of DEX to a final concentration of 10 μM for 3 h. After the incubation, cells were analyzed with a FACSAria II (BD Biosciences) and sorted directly into RNA Lysis Buffer from the Quick-RNA Miniprep Kit (Zymo Research), which was then used for RNA isolation following the manufacturer's instructions. Illumina library preparation was performed using the NEBNext Ultra II RNA Library Prep Kit for Illumina (NEB #E7770L) with NEBNext Multiplex Oligos for Illumina (NEB #E6609S) according to the manufacturer's instructions.

Maize ZmKN1 ChIP-seq experiment in isolated leaf cells

The maize protoplast ChIP protocol for ZmKN1 targets (Fig. 5B) was adapted from the previous uChIP-seq protocol (Lau and Bergmann 2015; Para et al. 2018) with modifications. In the p1107 vector, the BLRP tag was placed at the 5′ end of the inserted TF. The p1107-ZmKN1 plasmid was transfected into the etiolated maize leaf protoplasts as described above. After transfection, a final concentration of 50 μM biotin (Sigma B4639) was added to the buffer (10 mM MES, pH 5.7, 0.6 M mannitol, 10 mM KCl). Protoplasts were collected by centrifugation at 500 *g* for 2 min and resuspended in 500 μL buffer. About 60,000 transfected cells are sufficient for 1 ChIP reaction. For cross-linking, 13.7 μL of 36.5% formaldehyde was added and incubated at room temperature for 7 min. The reaction was quenched with 32 μL of 2 M glycine and incubated at room temperature for an additional 7 min. The cells were then centrifuged at 600 *g* for 2 min at 4 °C using a swing-bucket rotor. After removing the supernatant, the pellet was rinsed twice with 500 μL of ice-cold buffer, centrifuging at 600 g for 2 min at 4 °C between washes. The final pellet was snap-frozen in liquid nitrogen and stored at −80 °C.

Chromatin was isolated by adding 100 μL of prechilled lysis buffer (50 mM Tris–HCl, pH 8, 10 mM EDTA, 1% SDS, and cOmplete mini protease inhibitor) to the protoplasts, incubating on ice for 10 min, and resuspending by gentle pipetting. The lysate was transferred to a tube for sonication using a Bioruptor Pico Sonicator with 5 cycles of 30-s on, 30-s off in ice water. Subsequently, 500 μL of ChIP dilution buffer (10 mM Tris–HCl, pH 8, 150 mM NaCl, 1 mM EDTA, 1.1% Triton X-100, and cOmplete mini protease inhibitor-Roche 11836153001) was added to the lysate, and the mixture was centrifuged at 12,000 *g* for 10 min at 4 °C. The supernatant (chromatin) was transferred to a new prechilled tube. This was repeated by adding 400 μL of ChIP dilution buffer to the pellet, centrifuging again, and pooling the supernatants to dilute SDS from 1% to 0.1%.

During centrifugation, Dynabeads Protein A magnetic beads (Invitrogen 10001D) were washed by ensuring a homogeneous suspension and resuspending 20 μL of beads per sample in 1 mL of ChIP dilution buffer. The beads were placed on a magnetic stand for 30 s on ice, and the supernatant was discarded. This wash was repeated. The beads were then resuspended in 20**n* μL of ChIP dilution buffer (*n* = number of reactions). Prewashed Protein A (20 μL) was added to the chromatin mixture and incubated for 1 h at room temperature on a rotator. Beads were captured on a magnetic stand, and the supernatant was transferred to a new tube, setting aside 30 to 50 μL of chromatin as input for ChIP.

For immunoprecipitation, 50 μL of streptavidin beads (Invitrogen 65801D) was washed 3 times with 1 mL of ChIP dilution buffer and resuspended in 100**n* μL of ChIP dilution buffer. The beads were added to the precleared chromatin and incubated overnight with gentle agitation at 4 °C or for 1 h at room temperature. Beads were captured on a magnetic stand, the supernatant was discarded, and the beads were sequentially washed with 1 mL each of wash buffers 1 (2% SDS), 2 (20 mM Tris–HCl [pH 8], 500 mM NaCl, 2 mM EDTA, 0.1% deoxycholate, 0.1% (wt/vol) SDS, and 1% (vol/vol) Triton X-100), 3 (10 mM Tris–HCl [pH 8], 250 mM LiCl, 1 mM EDTA, 1% [vol/vol] NP-40%, and 0.5% [wt/vol] sodium deoxycholate), and 4 (10 mm Tris–HCl [pH 8], 1 mM EDTA), with each wash lasting 8 min at room temperature. Immune complexes were eluted by adding 95 μL of freshly prepared ChIP elution buffer (100 mM $NaHCO_3$, 1% SDS) and incubating at 65 °C for 40 min with a mixing frequency of 1,200 rpm. Tubes were briefly centrifuged (3,600 *g* for 10 s) and placed on a magnetic stand for 20 to 30 s, and 100 μL of eluates was transferred to new tubes.

For reverse cross-linking, 4 μL of 5 M NaCl was added to the 100 μL ChIPed DNA, and input samples were thawed, adjusted to a final volume of 100 μL with ChIP elution buffer, and supplemented with 4 μL of 5 M NaCl. Samples were incubated overnight at 65 °C. RNase A (1 μL, Thermo Scientific EN0531) was added to the eluates, mixed, and incubated at 60 °C for 30 min. Subsequently, 2.5 μL of protease solution (10-fold diluted proteinase K, Roche 3115887001) was added to each sample and incubated for 1 h at 63 °C. DNA was purified using the ChIP DNA Clean & Concentrator kit (Zymo D5205). DNA concentration was measured using Qubit. The cells were incubated in dark overnight. We used NEBNext Ultra II DNA Library Prep Kit (NEB E7645S) for sequencing libraries. The libraries were sequenced on the Illumina NextSeq 500 platform at NYU-CGSB Genomics Core using the HighOutput 1 × 75 bp mode.

Maize ZmKN1 protoplast ChIP-seq data analysis

The sequenced raw fastq reads were trimmed by fastp 0.21.0 (Chen et al. 2018). Clean reads were aligned to the maize genome version 4 downloaded from Ensembl Plant (Kersey et al. 2018) using bowtie

2.4.2 with default parameters (Langmead and Salzberg 2012). Aligned reads with MAPQ values smaller than 10 were discarded using samtools 1.11 (Li et al. 2009). The aligned bam files were converted int bigwig files using "bamCoverage" command from deepTools 3.5.0 (Ramírez et al. 2016). The peaks were called using macs 2.2.7.1 (Zhang et al. 2008) with non-ChIP DNA as control. The closet genes to the peaks were determined using bedtools 2.29.2 (Quinlan and Hall 2010). Top 600 peaks within +/1 kb of genes were kept for the motif analysis. The ±250 bp around the peak submits were used as input for MEME-ChIP (Bailey et al. 2015). The heatmap of the ChIP signal was plotted by "plotHeatmap" command from deepTools 3.5.0 (Ramírez et al. 2016). These data were used in Fig. 5 and Supplementary Fig. S16.

Maize and Arabidopsis TARGET data RNA-seq analysis

All RNA-seq raw reads for maize TARGET experiments were first processed to remove optical duplicates (Clumpify 37.50) and adapters (BBDuk 37.50). The trimmed reads were aligned to the maize genome version 4 using BBMap (37.50) (https://sourceforge.net/projects/bbmap/). The mapped reads were assigned to gene loci using featureCounts (Liao et al. 2014) with the latest annotation AGPv4.32. Genes passing the expression threshold (>1 CPM per sample) were normalized using upper-quantile normalization in EDASeq 2.18.0 (Risso et al. 2011) and replicate samples RUVSeq 1.18.0 (Risso et al. 2014). Differential gene expression analysis was conducted using edgeR (4.32) (McCarthy et al. 2012) to identify target genes for each TF, comparing EV- and TF-transfected samples with and without N treatment in parallel. The union of DEGs (false discovery rate [FDR] < 0.05) from minus- and plus-N conditions was considered the target genes for a given TF. For the Arabidopsis AtDIV1 TARGET assay, raw fastq files were trimmed by fastp 0.21.0 (Chen et al. 2018). The clean fastq files were aligned to TAIR10 genome using hisat2.2.1 with default parameters (Kim et al. 2015). The read counts matrix was generated by featureCounts v2.0.1 (Liao et al. 2014) from the aligned bam files. The DESeq2 (Love et al. 2014) was used to identify DEGs between AtDIV1 and EV with FDR <0.05. These data were used in Fig. 8.

Dynamic Factor Graph (DFG) network construction and network pruning

The DFG method was used for the network inference as described previously (Varala et al. 2018; Brooks et al. 2019) as depicted in Fig. 3. The DFG output is a table with all possible TF–target edges. The total TF list for maize is given in Supplementary Table S28. Every edge has a score indicating the confidence of that edge. In order to get the high-confidence N-signaling network, we applied network pruning based on AUPR curve (Varala et al. 2018; Brooks et al. 2019). We plotted the AUPR curve using maize TFs' TARGET results as gold standard. Then, we chose the cutoff of the DFG network where the precision score flats on the AUPR curve. Only edges with score higher than the cutoff were retained in the final high-confidence DFG networks. The P-value of the DFG network was computed from the permutation test. For this, we generated 1,000 random networks and calculated the AUPR values by permuting edge scores. The highest, lowest, and mean values of these 1,000 random networks were recorded and compared with the DFG network. The network AUPR values were calculated by the precrec package in R (Saito and Rehmsmeier 2017). We used the same input for networks that were constructed by GENIE3 (Huynh-Thu et al. 2010), dyGENIE3 (Huynh-Thu and Geurts 2018), and Pearson's correlation (cor(method = "p") function in R).

Gene orthology and N-response between maize and Arabidopsis orthologs

To infer the orthology between maize and Arabidopsis, we ran OrthoFinder2 (Emms and Kelly 2015) with the following command "orthofinder -a 8 -t 16 -M msa -A mafft." Protein sequences of the longest transcripts were downloaded from Phytozome 12 (Goodstein et al. 2012) for 8 species, including *A. thaliana*, *Brachypodium distachyon*, *Miscanthus sinensis*, *Oryza sativa*, *Panicum virgatum*, *Sorghum bicolor*, *Setaria italica*, and *Z. mays*. The protein sequences were used in OrthoFinder2, and the output ortholog table between maize and Arabidopsis was retrieved (Supplementary Data Set 5).

XGBoost machine learning prediction models of NUE trait from gene expression

To predict NUE phenotypes from gene expression levels, we used XGBoost (Chen and Guestrin 2016) the machine learning pipeline described previously (Cheng et al. 2021). The R version of XGBoost (https://cran.r-project.org/web/packages/xgboost) was used for the analysis. High-dimensional gene expression data pose challenges due to the risk of overfitting, especially when predicting phenotypes from gene expression matrices. To address this, we implemented multiple strategies. First, for feature selection, we used a conserved set of NxTime genes shared between maize and Arabidopsis, which represents a smaller and more biologically relevant subset compared to the complete NxTime gene set. As demonstrated in this study and in our previous work (Cheng et al. 2021), this conserved gene set improves model performance compared with an equivalent number of randomly selected NxTime genes. Second, we employed a leave-one-out cross-validation strategy for training and testing. In maize, each iteration used 15 genotypes for training and 1 for testing, repeated across 16 rounds so that each genotype served as the test set once. Third, during XGBoost model training, we applied 5-fold cross-validation to optimize hyperparameters (Cheng et al. 2021). Finally, we validated the model's robustness on an independent dataset (Li et al. 2023) by using the conserved ZmMYB34/R3 regulon as features, showing that this gene set achieved significantly better predictive performance than random gene sets (as shown in Supplementary Fig. S20).

Gene features were selected from either the conserved NxTime genes or conserved NUE Regulon genes. There are 16 genotypes in the maize field data with gene expression and NUE phenotype (Cheng et al. 2021). We trained the XGBoost model on the 15 genotypes and tested the model on the left-out genotype. A grid search was conducted to fine-tune the following hyperparameters, including "nrounds," "col- colsample_bytree," "eta," "subsample," and "max_depth." Model performance was evaluated by calculating Pearson's correlation coefficients between the predicted and actual values and reported the *r* from 100 iterations. The combined gene importance scores from all genotypes were named NUEscore. The same workflow was applied to the Arabidopsis dataset with 18 genotypes (from Cheng et al. 2021). The output of machine learning is given in Supplementary Table S20.

Network visualization and machine learning integration into NUE Regulons

To connect the ensemble N-GRN structure with the NUEscore, we computed the NUE Regulon score for each model-to-crop conserved N-responsive TF by aggregating the NUEscore of their target genes times the TF's own NUEscore (Fig. 7). The NUE Regulon score provides a weighted importance of TFs and their associated target genes, linking network connectivity with the gene importance. The ensemble N-GRN in shoot contains edges from pruned

DFG network, alongside validated TARGET and validated ChIP-seq regulations.

The AtDIV1 and ZmMYB34/R3 TARGET results were first mapped to their respective orthogroups based on the orthology from OrthoFinder2. The intersection of these orthogroups was tested using Fisher's exact test implemented in GeneOverlap R package (Shen 2023). All network visualizations were done in Cytoscape 3 (Shannon et al. 2003).

Analysis of external published datasets

For maize gene expression across developmental time in leaf and root, the fragments per kilobase of transcript per million mapped reads) values for NxTime genes were retrieved from qTeller (Woodhouse et al. 2021). The expression matrices were normalized using z-scores and plotted using ComplexHeatmap (Gu 2022). For "NGWAS" dataset in Supplementary Fig. S3, we retrieved genes mentioned in the published manuscript on GWAS meta-analysis for NUE (Xing et al. 2024), specifically Supplementary Tables S3, S9, S11, S13, S14, and S15. The maize v5 IDs were converted to v4 IDs according to MaizeGDB (Andorf et al. 2015).

For the Monsanto leaf dataset (Yang et al. 2011), the gene expression was measured by Affymetrix microarray. The probe sequences were obtained from NCBI Gene Expression Omnibus (GEO) (GPL14616). These sequences were aligned to the maize v4 cDNA sequence using the BLAST with the following parameters "blastn -task megablast -query probe.fa -db Zea_mays.AGPv4.cdna.all.fa -evalue 0.0001 -out monsanto_probe_raw_BLAST_result.txt -outfmt '7 qseqid sseqid evalue pident means' -best_hit_overhang 0.1 -max_target_seqs 1." Probed mapped to multiple genes were removed. To identify the N-responsive genes, we downloaded the raw array data from NCBI GEO (GSE32361) and used limma to identify probes that are differentially expressed between rescue nitrogen and low nitrogen conditions.

For the other published maize N-response RNA-seq datasets (Ge et al. 2020; Wang et al. 2020b; Buoso et al. 2021a; Li et al. 2023), we used the same pipeline as our maize NxTime dataset, except no UMI steps. We reprocess the raw fastq reads, downloaded from NCBI-SRA to minimize the differences between various processing methods as reported in the original publication. In general, raw fastq reads were downloaded from NCBI-SRA and trimmed by fastp 0.21.0 (Chen et al. 2018). The clean reads were aligned to the maize genome version 4 using STAR 2.7.6a (Dobin et al. 2013). The featureCounts v2.0.1 (Liao et al. 2014) was used to generate gene count matrices. The N-responsive genes were called using DESeq2 (Love et al. 2014) fitting model "~ Ntreatment" which compares nitrogen-treated with nontreated (or low concentration) samples. For the "Cheng_leaf" data, the N-responsive genes were retrieved from the Supplementary Data Set 2 of the publication (Cheng et al. 2021).

Accession numbers

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers listed in Supplementary Tables S2 and S3 and Supplementary Data Set 1.

Acknowledgments

The authors would like to thank Dr. Dennis Shasha for constructive discussions. The authors thank Drs. Jacapo Cirrone and Kranthi Varala for their discussion of computational analysis. The authors acknowledge members of the Coruzzi lab at NYU Biology and Shasha lab members from NYU Courant Institute for feedback and comment throughout the project. The authors thank Genomics Core at NYU-CGSB for cell sorting and sequencing and the NYU High Performance Computing (https://sites.google.com/nyu.edu/nyu-hpc/) for providing resources for the analysis in this manuscript.

Author contributions

J.H., C.-Y.C., and G.M.C. designed research; J.H., C.-Y.C., H.-J.S., N.M.D., and S.F. conducted experiments; J.H. and C.-Y.C. analyzed the data with input from M.D.B., T.L.J., M.S.K., and N.M.D.; and J.H. and G.M.C. wrote the manuscript with input from all authors.

Supplementary data

The following materials are available in the online version of this article.

Supplementary Figure S1. Gene expression PCA shows that PC1 is associated with nitrogen treatment, and PC2 is associated with time in the maize shoot and root NxTime dataset.

Supplementary Figure S2. Analysis pipeline for identifying NxTime genes in the maize data.

Supplementary Figure S3. The maize NxTime genes in shoot and root not only have significant overlap with published datasets, but also identify new N responsive genes.

Supplementary Figure S4. The functional nitrogen pathway genes in maize show significant activation by nitrogen treatment.

Supplementary Figure S5. Sixteen out of 26 maize root nitrogen pathway genes show a positive correlation with ^{15}N uptake.

Supplementary Figure S6. The identified maize seedling NxTime genes in this study also express in adult tissues.

Supplementary Figure S7. Enriched GO terms for maize shoot and root JIT genes show a temporal cascade from early to late time points.

Supplementary Figure S8. The maize JIT genes are enriched in unique motif clusters similar to those in Arabidopsis JIT genes.

Supplementary Figure S9. The individual motif enrichment for the maize and Arabidopsis JIT genes exhibits a sequential pattern over time, including several known nitrogen regulators.

Supplementary Figure S10. TARGET assay rapidly validate genome-wide TF–targets in cells.

Supplementary Figure S11. In the maize NxTime dataset, the DFG method performs significantly better than random networks and has the highest AUPR values compared with the other 3 tested methods.

Supplementary Figure S12. The maize shoot DFG network has a better PR curve than random networks, regardless of the validation TARGET TFs.

Supplementary Figure S13. The maize root DFG network has a better PR curve than random networks, regardless of the validation TARGET TFs.

Supplementary Figure S14. Time-resolved GRN of the nitrogen pathway in maize root.

Supplementary Figure S15. The ZmKN1 in planta misregulated genes have a significant overlap with the maize NxTime shoot genes.

Supplementary Figure S16. The protoplast ZmKN1 BLRP-ChIP-seq captures in planta binding and regulation.

Supplementary Figure S17. The maize TARGET dataset can be used to validate GRNs from other sources using AUPR analysis.

Supplementary Figure S18. The eigengenes of conserved NxTime genes (331 out of 353) are significantly correlated with NUE traits, but not with other agronomic traits.

Supplementary Figure S19. The maize ZmMYB34 and ZmMYBR3 share a significant number of TARGET-validated genes, suggesting redundant functions between these 2 TFs.

Supplementary Figure S20. The random forest model using 24 ZmMYB34/R3 regulon genes predicts significantly better than the same number of random conserved NxTime genes for ear diameter (ED) and ear length (EL).

Supplementary Table S1. 15N percentage in the maize shoot and root samples.

Supplementary Table S2. Maize shoot NxTime gene list.

Supplementary Table S3. Maize root NxTime gene list.

Supplementary Table S4. Maize N-pathway genes.

Supplementary Table S5. The JIT gene-enriched GO terms for maize shoot and root.

Supplementary Table S6. The JIT gene-enriched GO terms for Arabidopsis shoot and root.

Supplementary Table S7. The JIT gene-enriched motif clusters for maize shoot and root.

Supplementary Table S8. The JIT gene-enriched motif clusters for Arabidopsis shoot and root.

Supplementary Table S9. The JIT gene-enriched JASPAR motifs for maize shoot and root.

Supplementary Table S10. The JIT gene-enriched JASPAR motifs for Arabidopsis shoot and root.

Supplementary Table S11. The 23 maize TARGET results overlap with NxTime, N-pathway, and other N-datasets.

Supplementary Table S12. Predicted activator or repressor roles of 23 maize TFs with TARGET Data.

Supplementary Table S13. AUPR of maize and Arabidopsis GRNs. All networks were built using the DFG method with time-course RNA-seq data.

Supplementary Table S14. Pruned DFG network for maize shoot.

Supplementary Table S15. Pruned DFG network for maize root.

Supplementary Table S16. GO enrichment for genes in the shoot DFG pruned network.

Supplementary Table S17. Gene network edges supported by 2 or 3 methods.

Supplementary Table S18. The 490 overlap genes between ZmKN1 in planta RNA-seq (Bolduc et al. 2012) and maize NxTime shoot genes (this study).

Supplementary Table S19. The ZmKN1 Network Walking-enriched GO terms.

Supplementary Table S20. The maize conserved NxTime genes in shoot and their feature importance scores from XGBoost.

Supplementary Table S21. The maize conserved NxTime genes in root.

Supplementary Table S22. The NUE Regulon score for all maize TFs within the DFG network.

Supplementary Table S23. The Arabidopsis AtDIV1 leaf TARGET DE genes.

Supplementary Table S24. The AtDIV1-ZmMYB34/R3 conserved NUE Regulon genes in maize and their feature importance scores from XGBoost.

Supplementary Table S25. The AtDIV1-ZmMYB34/R3 conserved NUE Regulon genes in Arabidopsis and their feature importance scores from XGBoost.

Supplementary Table S26. All mentioned maize gene names with v4 IDs.

Supplementary Table S27. All mentioned Arabidopsis gene names with TAIR IDs.

Supplementary Table S28. The combined maize TF gene list from Grassius and PlantTFDB.

Supplementary Data Set 1. The 23 maize TFs DEGs from the TARGET assay.

Supplementary Data Set 2. The ensemble N-GRN for nitrogen-uptake/assimilation pathway genes in maize shoots in Cytoscape format.

Supplementary Data Set 3. The ensemble N-GRN for nitrogen-uptake/assimilation pathway genes in maize roots in Cytoscape format.

Supplementary Data Set 4. The ZmKN1 regulatory network for N-responsive genes in Cytoscape format.

Supplementary Data Set 5. The Arabidopsis–Maize orthology table used in this study.

Funding

This work was supported by the National Science Foundation Plant Genome Research Program (IOS-1339362) to G.M.C., National Institutes of Health Grant R01-GM121753 to G.M.C., and National Institutes of Health National Institute of General Medical Sciences Fellowship F32GM116347 to M.D.B.

Conflict of interest statement. No conflicts of interest to declare.

Data availability

The raw FASTQ data generated in this study have been deposited in the GEO database under accession GSE280353. The code used in this study is available on GitHub: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE280353.

References

Aalto A, Viitasaari L, Ilmonen P, Mombaerts L, Gonçalves J. Gene regulatory network inference from sparsely sampled noisy data. *Nat Commun*. 2020:11(1):3493. https://doi.org/10.1038/s41467-020-17217-1

Aibar S, González-Blas CB, Moerman T, Imrichova H, Hulselmans G, Rambow F, Marine J-C, Geurts P, Aerts J, van den Oord J. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods*. 2017:14(11):1083–1086. https://doi.org/10.1038/nmeth.4463

Alexa A, Rahnenführer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*. 2006:22(13):1600–1607. https://doi.org/10.1093/bioinformatics/btl140

Altenhoff AM, Warwick Vesztrocy A, Bernard C, Train C-M, Nicheperovich A, Prieto Baños S, Julca I, Moi D, Nevers Y, Majidian S, et al. OMA orthology in 2024: improved prokaryote coverage, ancestral and extant GO enrichment, a revamped synteny viewer and more in the OMA ecosystem. *Nucleic Acids Res*. 2024:52(D1):D513–D521. https://doi.org/10.1093/nar/gkad1020

Alvarez JM, Brooks MD, Swift J, Coruzzi GM. Time-based systems biology approaches to capture and model dynamic gene regulatory networks. *Annu Rev Plant Biol*. 2021:72(1):105–131. https://doi.org/10.1146/annurev-arplant-081320-090914

Alvarez JM, Schinke A-L, Brooks MD, Pasquino A, Leonelli L, Varala K, Safi A, Krouk G, Krapp A, Coruzzi GM. Transient genome-wide interactions of the master transcription factor NLP7 initiate a rapid nitrogen-response cascade. *Nat Commun*. 2020:11(1):1–13. https://doi.org/10.1038/s41467-020-14979-6

Andorf CM, Cannon EK, Portwood JL, Gardiner JM, Harper LC, Schaeffer ML, Braun BL, Campbell DA, Vinnakota AG, Sribalusu VV. MaizeGDB update: new tools, data and interface for the maize

model organism database. *Nucleic Acids Res.* 2015:44(D1): D1195–D1201. https://doi.org/10.1093/nar/gkv1007

Aryee MJ, Gutiérrez-Pabello JA, Kramnik I, Maiti T, Quackenbush J. An improved empirical Bayes approach to estimating differential gene expression in microarray time-course data: BETR (Bayesian estimation of temporal regulation). *BMC Bioinformatics*. 2009:10(1): 409. https://doi.org/10.1186/1471-2105-10-409

Bailey TL, Grant CE. SEA: Simple Enrichment Analysis of motifs. bioRxiv 457422. https://doi.org/10.1101/2021.08.23.457422, 24 August 2021, preprint: not peer reviewed.

Bailey TL, Johnson J, Grant CE, Noble WS. The MEME suite. *Nucleic Acids Res*. 2015:43(W1):W39–W49. https://doi.org/10.1093/nar/gkv416

Bargmann BOR, Marshall-Colon A, Efroni I, Ruffel S, Birnbaum KD, Coruzzi GM, Krouk G. TARGET: a transient transformation system for genome-wide transcription factor target discovery. *Mol Plant*. 2013:6(3):978–980. https://doi.org/10.1093/mp/sst010

Bar-Joseph Z, Gitter A, Simon I. Studying and modelling dynamic biological processes using time-series gene expression data. *Nat Rev Genet*. 2012:13(8):552–564. https://doi.org/10.1038/nrg3244

Bolduc N, Yilmaz A, Mejia-Guerra MK, Morohashi K, O'Connor D, Grotewold E, Hake S. Unraveling the KNOTTED1 regulatory network in maize meristems. *Genes Dev*. 2012:26(15):1685–1690. https://doi.org/10.1101/gad.193433.112

Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. *Cell*. 2017:169(7):1177–1186. https://doi.org/10.1016/j.cell.2017.05.038

Brooks MD, Cirrone J, Pasquino AV, Alvarez JM, Swift J, Mittal S, Juang C-L, Varala K, Gutiérrez RA, Krouk G, et al. Network walking charts transcriptional dynamics of nitrogen signaling by integrating validated and predicted genome-wide interactions. *Nat Commun*. 2019:10(1):1569. https://doi.org/10.1038/s41467-019-09522-1

Brooks MD, Juang CL, Katari MS, Alvarez JM, Pasquino A, Shih HJ, Huang J, Shanks C, Cirrone J, Coruzzi GM. ConnecTF: a platform to integrate transcription factor-gene interactions and validate regulatory networks. *Plant Physiol*. 2021:185(1):49–66. https://doi.org/10.1093/PLPHYS/KIAA012

Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods*. 2021:18(4): 366–368. https://doi.org/10.1038/s41592-021-01101-x

Buoso S, Tomasi N, Arkoun M, Maillard A, Jing L, Marroni F, Pluchon S, Pinton R, Zanin L. Transcriptomic and metabolomic profiles of Zea mays fed with urea and ammonium. *Physiol Plant*. 2021a:173(3):935–953. https://doi.org/10.1111/ppl.13493

Buoso S, Tomasi N, Said-Pullicino D, Arkoun M, Yvin J-C, Pinton R, Zanin L. Characterization of physiological and molecular responses of Zea mays seedlings to different urea-ammonium ratios. *Plant Physiol Biochem*. 2021b:162:613–623. https://doi.org/10.1016/j.plaphy.2021.03.037

Cao H, Liu Z, Guo J, Jia Z, Shi Y, Kang K, Peng W, Wang Z, Chen L, Neuhaeuser B, et al. ZmNRT1.1B (ZmNPF6.6) determines nitrogen use efficiency via regulation of nitrate transport and signalling in maize. *Plant Biotechnol J*. 2023:22(2):316–329. https://doi.org/10.1111/pbi.14185

Carlson M, Falcon S, Pages H, Li N. org.At.tair.db: Genome wide annotation for Arabidopsis, 2019:9. Preprint at https://doi.org/10.18129/B9.bioc.org.At.tair.db

Castaings L, Camargo A, Pocholle D, Gaudon V, Texier Y, Boutet-Mercey S, Taconnat L, Renou J-P, Daniel-Vedele F, Fernandez E, et al. The nodule inception-like protein 7 modulates nitrate sensing and metabolism in Arabidopsis. *Plant J*. 2009:57(3): 426–435. https://doi.org/10.1111/j.1365-313X.2008.03695.x

Castro-Mondragon JA, Riudavets-Puig R, Rauluseviciute I, Berhanu Lemma R, Turchi L, Blanc-Mathieu R, Lucas J, Boddie P, Khan A, Manosalva Pérez N, et al. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*. 2022:50(D1):D165–D173. https://doi.org/10.1093/nar/gkab1113

Chechik G, Oh E, Rando O, Weissman J, Regev A, Koller D. Activity motifs reveal principles of timing in transcriptional control of the yeast metabolic network. *Nat Biotechnol*. 2008:26(11): 1251–1259. https://doi.org/10.1038/nbt.1499

Chen F, Mackey AJ, Vermunt JK, Roos DS. Assessing performance of orthology detection strategies applied to eukaryotic genomes. *PLoS One*. 2007:2(4):e383. https://doi.org/10.1371/journal.pone.0000383

Chen S, Mar JC. Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data. *BMC Bioinform*. 2018:19(1):232. https://doi.org/10.1186/s12859-018-2217-z

Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018:34(17):i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Chen T, Guestrin C. 2016. XGBoost: a scalable tree boosting system. In. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, New York, NY, USA; Association for Computing Machinery. pp. 785–794. https://doi.org/10.1145/2939672.2939785

Cheng C-Y, Li Y, Varala K, Bubert J, Huang J, Kim GJ, Halim J, Arp J, Shih H-JS, Levinson G, et al. Evolutionarily informed machine learning enhances the power of predictive gene-to-phenotype relationships. *Nat Commun*. 2021:12(1):5627. https://doi.org/10.1038/s41467-021-25893-w

Coffey N, Hinde J, Holian E. Clustering longitudinal profiles using P-splines and mixed effects models applied to time-course gene expression data. *Comput Stat Data Anal*. 2014:71:14–29. https://doi.org/10.1016/j.csda.2013.04.001

Contreras-López O, Vidal EA, Riveras E, Alvarez JM, Moyano TC, Sparks EE, Medina J, Pasquino A, Benfey PN, Coruzzi GM. Spatiotemporal analysis identifies ABF2 and ABF3 as key hubs of endodermal response to nitrate. *Proc Natl Acad Sci U S A*. 2022:119(4): e2107879119. https://doi.org/10.1073/pnas.2107879119

Cui H. Challenges and approaches to crop improvement through C3-to-C4 engineering. *Front Plant Sci*. 2021:12:715391. https://doi.org/10.3389/fpls.2021.715391

Curci PL, Zhang J, Mähler N, Seyfferth C, Mannapperuma C, Diels T, Van Hautegem T, Jonsen D, Street N, Hvidsten TR, et al. Identification of growth regulators using cross-species network analysis in plants. *Plant Physiol*. 2022:190(4):2350–2365. https://doi.org/10.1093/plphys/kiac374

Das M, Haberer G, Panda A, Das Laha S, Ghosh TC, Schäffner AR. Expression pattern similarities support the prediction of orthologs retaining common functions after gene duplication events. *Plant Physiol*. 2016:171(4):2343–2357. https://doi.org/10.1104/pp.15.01207

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-Seq aligner. *Bioinformatics*. 2013:29(1):15–21. https://doi.org/10.1093/bioinformatics/bts635

Du H, Ning L, He B, Wang Y, Ge M, Xu J, Zhao H. Cross-species root transcriptional network analysis highlights conserved modules in response to nitrate between maize and Sorghum. *IJMS*. 2020:21(4):1445. https://doi.org/10.3390/ijms21041445

Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol*. 2015:16(1):157. https://doi.org/10.1186/s13059-015-0721-2

Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 2019:20(1):238. https://doi.org/10.1186/s13059-019-1832-y

Fan Y, Li L, Sun S. Powerful and accurate detection of temporal gene expression patterns from multi-sample multi-stage single-cell transcriptomics data with TDEseq. *Genome Biol.* 2024:25(1):96. https://doi.org/10.1186/s13059-024-03237-3

FAO. Agricultural production statistics 2000–2022. FAOSTAT Analytical Briefs (No. 79). 2023. https://doi.org/10.4060/cc9205en

Ferrier T, Matus JT, Jin J, Riechmann JL. *Arabidopsis* paves the way: genomic and network analyses in crops. *Curr Opin Biotechnol.* 2011:22(2):260–270. https://doi.org/10.1016/j.copbio.2010.11.010

Gao Y, Cabrera Serrenho A. Greenhouse gas emissions from nitrogen fertilizers could be reduced by up to one-fifth of current levels by 2050 with combined interventions. *Nat Food.* 2023:4(2):170–178. https://doi.org/10.1038/s43016-023-00698-w

Gao Y, Qi S, Wang Y. Nitrate signaling and use efficiency in crops. *Plant Comm.* 2022:3(5):100353. https://doi.org/10.1016/j.xplc.2022.100353

Gaudinier A, Rodriguez-Medina J, Zhang L, Olson A, Liseron-Monfils C, Bågman A, Foret J, Abbitt S, Tang M, Li B, et al. Transcriptional regulation of nitrogen-associated metabolism and growth. *Nature.* 2018:563(7730):259–264. https://doi.org/10.1038/s41586-018-0656-3

Ge M, Wang Y, Liu Y, Jiang L, He B, Ning L, Du H, Lv Y, Zhou L, Lin F, et al. The NIN-like protein 5 (ZmNLP5) transcription factor is involved in modulating the nitrogen response in maize. *Plant J.* 2020:102(2):353–368. https://doi.org/10.1111/tpj.14628

Gheith EMS, El-Badry OZ, Lamlom SF, Ali HM, Siddiqui MH, Ghareeb RY, El-Sheikh MH, Jebril J, Abdelsalam NR, Kandil EE. Maize (Zea mays L.) productivity and nitrogen use efficiency in response to nitrogen application levels and time. *Front Plant Sci.* 2022:13:941343. https://doi.org/10.3389/fpls.2022.941343

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 2012:40(D1):D1178–D1186. https://doi.org/10.1093/nar/gkr944

Greene B, Walko R, Hake S. Mutator insertions in an intron of the maize knotted1 gene result in dominant suppressible mutations. *Genetics.* 1994:138(4):1275–1285. https://doi.org/10.1093/genetics/138.4.1275

Gu Z. Complex heatmap visualization. *iMeta.* 2022:1(3):e43. https://doi.org/10.1002/imt2.43

Guan M, Møller IS, Schjoerring JK. Two cytosolic glutamine synthetase isoforms play specific roles for seed germination and seed yield structure in Arabidopsis. *J Exp Bot.* 2015:66(1):203–212. https://doi.org/10.1093/jxb/eru411

Guan P, Wang R, Nacry P, Breton G, Kay SA, Pruneda-Paz JL, Davani A, Crawford NM. Nitrate foraging by Arabidopsis roots is mediated by the transcription factor TCP20 through the systemic signaling pathway. *Proc Natl Acad Sci U S A.* 2014:111(42):15267–15272. https://doi.org/10.1073/pnas.1411375111

Guillotin B, Rahni R, Passalacqua M, Mohammed MA, Xu X, Raju SK, Ramírez CO, Jackson D, Groen SC, Gillis J, et al. A pan-grass transcriptome reveals patterns of cellular divergence in crops. *Nature.* 2023:617(7962):785–791. https://doi.org/10.1038/s41586-023-06053-0

Gutierrez RA, Stokes TL, Thum K, Xu X, Obertello M, Katari MS, Tanurdzic M, Dean A, Nero DC, McClung CR, et al. Systems approach identifies an organic nitrogen-responsive gene network that is regulated by the master clock control gene CCA1. *Proc Natl Acad Sci U S A.* 2008:105(12):4939–4944. https://doi.org/10.1073/pnas.0800211105

Hartmann A, Berkowitz O, Whelan J, Narsai R. Cross-species transcriptomic analyses reveals common and opposite responses in Arabidopsis, rice and barley following oxidative stress and hormone treatment. *BMC Plant Biol.* 2022:22(1):62. https://doi.org/10.1186/s12870-021-03406-7

Heerah S, Molinari R, Guerrier S, Marshall-Colon A. Granger-causal testing for irregularly sampled time series with application to nitrogen signalling in Arabidopsis. *Bioinformatics.* 2021:37(16):2450–2460. https://doi.org/10.1093/bioinformatics/btab126

Hirel B, Le Gouis J, Ney B, Gallais A. The challenge of improving nitrogen use efficiency in crop plants: towards a more central role for genetic variability and quantitative genetics within integrated approaches. *J Exp Bot.* 2007:58(9):2369–2387. https://doi.org/10.1093/jxb/erm097

Huang J, Katari MS, Juang C-L, Coruzzi GM, Brooks MD. Building high-confidence gene regulatory networks by integrating validated TF–target gene interactions using ConnecTF. In: Kaufmann K, Vandepoele K, editors. *Plant gene regulatory networks. Methods in molecular biology*. Vol. 2698. Humana; 2023. p. 195–220. https://doi.org/10.1007/978-1-0716-3354-0_13

Huynh-Thu VA, Geurts P. dynGENIE3: dynamical GENIE3 for the inference of gene networks from time series expression data. *Sci Rep.* 2018:8(1):3384. https://doi.org/10.1038/s41598-018-21715-0

Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. Inferring regulatory networks from expression data using tree-based methods. *PLoS One.* 2010:5(9):1–10. https://doi.org/10.1371/journal.pone.0012776

Jiang L, Ball G, Hodgman C, Coules A, Zhao H, Lu C. Analysis of gene regulatory networks of maize in response to nitrogen. *Genes (Basel).* 2018:9(3):151. https://doi.org/10.3390/genes9030151

Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* 2023:51(D1):D587–D592. https://doi.org/10.1093/nar/gkac963

Kersey PJ, Allen JE, Allot A, Barba M, Boddu S, Bolt BJ, Carvalho-Silva D, Christensen M, Davis P, Grabmueller C, et al. Ensembl genomes 2018: an integrated omics infrastructure for non-vertebrate species. *Nucleic Acids Res.* 2018:46(D1):D802–D808. https://doi.org/10.1093/nar/gkx1011

Kerstetter RA, Laudencia-Chingcuanco D, Smith LG, Hake S. Loss-of-function mutations in the maize homeobox gene, knotted1, are defective in shoot meristem maintenance. *Development.* 1997:124(16):3045–3054. https://doi.org/10.1242/dev.124.16.3045

Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 2015:12(4):357–360. https://doi.org/10.1038/nmeth.3317

Koonin EV. Orthologs, paralogs, and evolutionary genomics. *Annu Rev Genet.* 2005:39(1):309–338. https://doi.org/10.1146/annurev.genet.39.073003.114725

Köster J, Rahmann S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics.* 2012:28(19):2520–2522. https://doi.org/10.1093/bioinformatics/bts480

Krouk G, Lingeman J, Colon A, Coruzzi G, Shasha D. Gene regulatory networks in plants: learning causality from time and perturbation. *Genome Biol.* 2013:14(6):123. https://doi.org/10.1186/gb-2013-14-6-123

Krouk G, Mirowski P, LeCun Y, Shasha DE, Coruzzi GM. Predictive network modeling of the high-resolution dynamic plant transcriptome in response to nitrate. *Genome Biol.* 2010:11(12):R123. https://doi.org/10.1186/gb-2010-11-12-r123

Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol.* 2017:34(7):1812–1819. https://doi.org/10.1093/molbev/msx116

Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008:9(1):559. https://doi.org/10.1186/1471-2105-9-559

Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods*. 2012:9(4):357–359. https://doi.org/10.1038/nmeth.1923

Lassaletta L, Billen G, Grizzetti B, Anglade J, Garnier J. 50 year trends in nitrogen use efficiency of world cropping systems: the relationship between yield and nitrogen input to cropland. *Environ Res Lett*. 2014:9(10):105011. https://doi.org/10.1088/1748-9326/9/10/105011

Lau OS, Bergmann DC. MOBE-ChIP: a large-scale chromatin immunoprecipitation assay for cell type-specific studies. *Plant J*. 2015:84(2):443–450. https://doi.org/10.1111/tpj.13010

Lee Y-C, Tsai P-T, Huang X-X, Tsai H-L. Family members additively repress the ectopic expression of BASIC PENTACYSTEINE3 to prevent disorders in Arabidopsis circadian vegetative development. *Front Plant Sci*. 2022:13:919946. https://doi.org/10.3389/fpls.2022.919946

Leng N, Li Y, McIntosh BE, Nguyen BK, Duffin B, Tian S, Thomson JA, Dewey CN, Stewart R, Kendziorski C. EBSeq-HMM: a Bayesian approach for identifying gene-expression changes in ordered RNA-seq experiments. *Bioinformatics*. 2015:31(16):2614–2622. https://doi.org/10.1093/bioinformatics/btv193

Li C, Li Y, Song G, Yang D, Xia Z, Sun C, Zhao Y, Hou M, Zhang M, Qi Z, et al. Gene expression and expression quantitative trait loci analyses uncover natural variations underlying the improvement of important agronomic traits during modern maize breeding. *Plant J*. 2023:115(3):772–787. https://doi.org/10.1111/tpj.16260

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009:25(16):2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li S, Tian Y, Wu K, Ye Y, Yu J, Zhang J, Liu Q, Hu M, Li H, Tong Y, et al. Modulating plant growth–metabolism coordination for sustainable agriculture. *Nature*. 2018:560(7720):595–600. https://doi.org/10.1038/s41586-018-0415-5

Liao C, Peng Y, Ma W, Liu R, Li C, Li X. Proteomic analysis revealed nitrogen-mediated metabolic, developmental, and hormonal regulation of maize (Zea mays L.) ear growth. *J Exp Bot*. 2012:63(14):5275–5288. https://doi.org/10.1093/jxb/ers187

Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014:30(7):923–930. https://doi.org/10.1093/bioinformatics/btt656

Liu M, Guo C, Xie K, Chen K, Chen J, Wang Y, Wang X. A cross-species co-functional gene network underlying leaf senescence. *Hortic Res*. 2023:10(1):uhac251. https://doi.org/10.1093/hr/uhac251

Liu X, Hu B, Chu C. Nitrogen assimilation in plants: current status and future prospects. *J Genet Genomics*. 2022:49(5):394–404. https://doi.org/10.1016/j.jgg.2021.12.006

Lothier J, Gaufichon L, Sormani R, Lemaître T, Azzopardi M, Morin H, Chardon F, Reisdorf-Cren M, Avice J-C, Masclaux-Daubresse C. The cytosolic glutamine synthetase GLN1;2 plays a role in the control of plant growth and ammonium homeostasis in Arabidopsis rosettes when nitrate supply is not limiting. *J Exp Bot*. 2011:62(4):1375–1390. https://doi.org/10.1093/jxb/erq299

Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol*. 2014:15(12):1. https://doi.org/10.1186/s13059-014-0550-8

Luan Y, Li H. Clustering of time-course gene expression data using a mixed-effects model with B-splines. *Bioinformatics*. 2003:19(4):474–482. https://doi.org/10.1093/bioinformatics/btg014

Ma C, Zhang HH, Wang X. Machine learning for big data analytics in plants. *Trends Plant Sci*. 2014:19(12):798–808. https://doi.org/10.1016/j.tplants.2014.08.004

Marbach D, Costello JC, Küffner R, Vega NNM, Prill RJ, Camacho DM, Allison KR, Aderhold A, Allison KR, Bonneau R, et al. Wisdom of crowds for robust gene network inference. *Nat Methods*. 2012:9(8):796–804. https://doi.org/10.1038/nmeth.2016

Marku M, Pancaldi V. From time-series transcriptomics to gene regulatory networks: a review on inference methods. *PLoS Comput Biol*. 2023:19(8):e1011254. https://doi.org/10.1371/journal.pcbi.1011254

Martin A, Lee J, Kichey T, Gerentes D, Zivy M, Tatout C, Dubois F, Balliau T, Valot B, Davanture M, et al. Two cytosolic glutamine synthetase isoforms of maize are specifically involved in the control of grain production. *Plant Cell*. 2006:18(11):3252–3274. https://doi.org/10.1105/tpc.106.042689

Mathieson I. The omnigenic model and polygenic prediction of complex traits. *Am J Hum Genet*. 2021:108(9):1558–1563. https://doi.org/10.1016/j.ajhg.2021.07.003

Matsuda K, Mikami T, Oki S, Iida H, Andrabi M, Boss JM, Yamaguchi K, Shigenobu S, Kondoh H. ChIP-seq analysis of genomic binding regions of five major transcription factors highlights a central role for ZIC2 in the mouse epiblast stem cell gene regulatory network. *Development*. 2017:144(11):1948–1958. https://doi.org/10.1242/dev.143479

McAdams HH, Shapiro L. A bacterial cell-cycle regulatory network operating in time and space. *Science*. 2003:301(5641):1874–1877. https://doi.org/10.1126/science.1087694

McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res*. 2012:40(10):4288–4297. https://doi.org/10.1093/nar/gks042

Möller S, Saul N, Projahn E, Barrantes I, Gézsi A, Walter M, Antal P, Fuellen G. Gene co-expression analyses of health(span) across multiple species. *NAR Genom Bioinform*. 2022:4(4):lqac083. https://doi.org/10.1093/nargab/lqac083

Moreno-Hagelsieb G, Latimer K. Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics*. 2008:24(3):319–324. https://doi.org/10.1093/bioinformatics/btm585

Morffy N, Van den Broeck L, Miller C, Emenecker RJ, Bryant JA, Lee TM, Sageman-Furnas K, Wilkinson EG, Pathak S, Kotha SR, et al. Identification of plant transcriptional activation domains. *Nature*. 2024:632(8023):166–173. https://doi.org/10.1038/s41586-024-07707-3

Mu Y, Zou M, Sun X, He B, Xu X, Liu Y, Zhang L, Chi W. BASIC PENTACYSTEINE proteins repress ABSCISIC ACID INSENSITIVE4 expression via direct recruitment of the polycomb-repressive complex 2 in Arabidopsis root development. *Plant Cell Physiol*. 2017:58(3):607–621. https://doi.org/10.1093/pcp/pcx006

Mueller AJ, Canty-Laird EG, Clegg PD, Tew SR. Cross-species gene modules emerge from a systems biology approach to osteoarthritis. *npj Syst Biol Appl*. 2017:3(1):1–15. https://doi.org/10.1038/s41540-017-0014-3

Naulin PA, Armijo GI, Vega AS, Tamayo KP, Gras DE, de la Cruz J, Gutiérrez RA. Nitrate induction of primary root growth requires cytokinin signaling in Arabidopsis thaliana. *Plant Cell Physiol*. 2020:61(2):342–352. https://doi.org/10.1093/pcp/pcz199

Nevers Y, Jones TEM, Jyothi D, Yates B, Ferret M, Portell-Silva L, Codo L, Cosentino S, Marcet-Houben M, Vlasova A. The quest for orthologs orthology benchmark service in 2022. *Nucleic Acids Res*. 2022:50(W1):W623–W632. https://doi.org/10.1093/nar/gkac330

Obertello M, Shrivastava S, Katari MS, Coruzzi GM. Cross-species network analysis uncovers conserved nitrogen-regulated network modules in rice. *Plant Physiol*. 2015:168(4):1830–1843. https://doi.org/10.1104/pp.114.255877

Para A, Li Y, Coruzzi GM. μChIP-seq for genome-wide mapping of in vivo TF-DNA interactions in Arabidopsis root protoplasts. *Methods Mol Biol*. 2018:1761:249–261. https://doi.org/10.1007/978-1-4939-7747-5_19

Patel RV, Nahal HK, Breit R, Provart NJ. BAR expressolog identification: expression profile similarity ranking of homologous genes in plant species. *Plant J*. 2012:71(6):1038–1050. https://doi.org/10.1111/j.1365-313X.2012.05055.x

Persson E, Sonnhammer ELL. InParanoiDB 9: ortholog groups for protein domains and full-length proteins. *J Mol Biol*. 2023:435(14):168001. https://doi.org/10.1016/j.jmb.2023.168001

Pratapa A, Jalihal AP, Law JN, Bharadwaj A, Murali T. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat Methods*. 2020:17(2):147–154. https://doi.org/10.1038/s41592-019-0690-6

Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010:26(6):841–842. https://doi.org/10.1093/bioinformatics/btq033

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. 2016:44(W1):W160–W165. https://doi.org/10.1093/nar/gkw257

Risso D, Ngai J, Speed TP, Dudoit S. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat Biotechnol*. 2014:32(9):896–902. https://doi.org/10.1038/nbt.2931

Risso D, Schwartz K, Sherlock G, Dudoit S. GC-content normalization for RNA-seq data. *BMC Bioinform*. 2011:12(1):480. https://doi.org/10.1186/1471-2105-12-480

Ruffel S, Krouk G, Ristova D, Shasha D, Birnbaum KD, Coruzzi GM. Nitrogen economics of root foraging: transitive closure of the nitrate-cytokinin relay and distinct systemic signaling for N supply vs. demand. *Proc Natl Acad Sci U S A*. 2011:108(45):18524–-18529. https://doi.org/10.1073/pnas.1108684108

Saito T, Rehmsmeier M. Precrec: fast and accurate precision–recall and ROC curve calculations in R. *Bioinformatics*. 2017:33(1):145–147. https://doi.org/10.1093/bioinformatics/btw570

Shanks CM, Hecker A, Cheng C-Y, Brand L, Collani S, Schmid M, Schaller GE, Wanke D, Harter K, Kieber JJ. Role of BASIC PENTACYSTEINE transcription factors in a subset of cytokinin signaling responses. *Plant J*. 2018:95(3):458–473. https://doi.org/10.1111/tpj.13962

Shanks CM, Huang J, Cheng C-Y, Shih H-JS, Brooks MD, Alvarez JM, Araus V, Swift J, Henry A, Coruzzi GM. Validation of a high-confidence regulatory network for gene-to-NUE phenotype in field-grown rice. *Front Plant Sci*. 2022:13:1006044. https://doi.org/10.3389/fpls.2022.1006044

Shanks CM, Rothkegel K, Brooks MD, Cheng C-Y, Alvarez JM, Ruffel S, Krouk G, Gutiérrez RA, Coruzzi GM. Nitrogen sensing and regulatory networks: it's about time and space. *Plant Cell*. 2024:36(5):1482–1503. https://doi.org/10.1093/plcell/koae038

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003:13(11):2498–2504. https://doi.org/10.1101/gr.1239303

Shen L. GeneOverlap: an R package to test and visualize gene overlaps. *R Package*. 2023:3:1–17. https://doi.org/10.18129/B9.bioc.GeneOverlap

Smith LG, Greene B, Veit B, Hake S. A dominant mutation in the maize homeobox gene, knotted-1, causes its ectopic expression in leaf cells with altered fates. *Development*. 1992:116(1):21–30. https://doi.org/10.1242/dev.116.1.21

Smith T, Heger A, Sudbery I. UMI-tools: modeling sequencing errors in unique molecular identifiers to improve quantification accuracy. *Genome Res*. 2017:27(3):491–499. https://doi.org/10.1101/gr.209601.116

Spies D, Renz PF, Beyer TA, Ciaudo C. Comparative analysis of differential gene expression tools for RNA sequencing time course data. *Brief Bioinform*. 2017:20(1):1–11. https://doi.org/10.1093/bib/bbx115

Springer N, León N, Grotewold E. Challenges of translating gene regulatory information into agronomic improvements. *Trends Plant Sci*. 2019:24(12):1075–1082. https://doi.org/10.1016/j.tplants.2019.07.004

Storey JD, Xiao W, Leek JT, Tompkins RG, Davis RW. Significance analysis of time course microarray experiments. *Proc Natl Acad Sci U S A*. 2005:102(36):12837–12842. https://doi.org/10.1073/pnas.0504609102

Sun X, Dalpiaz D, Wu D, Liu J S, Zhong W, Ma P. Statistical inference for time course RNA-Seq data using a negative binomial mixed-effect model. *BMC Bioinformatics*. 2016:17(1):324. https://doi.org/10.1186/s12859-016-1180-9

Tai YC, Speed TP. On gene ranking using replicated microarray time course data. *Biometrics*. 2009:65(1):40–51. https://doi.org/10.1111/j.1541-0420.2008.01057.x

The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*. 2019:47(D1):D506–D515. https://doi.org/10.1093/nar/gky1049

Tu X, Mejía-Guerra MK, Franco JAV, Tzeng D, Chu P-Y, Shen W, Wei Y, Dai X, Li P, Buckler ES. Reconstructing the maize leaf regulatory network using ChIP-seq data of 104 transcription factors. *Nat Commun*. 2020:11(1):1–13. https://doi.org/10.1038/s41467-020-18832-8

Ueda Y, Nosaki S, Sakuraba Y, Miyakawa T, Kiba T, Tanokura M, Yanagisawa S. NIGT1 family proteins exhibit dual mode DNA recognition to regulate nutrient response-associated genes in Arabidopsis. *PLoS Genet*. 2020:16(11):e1009197. https://doi.org/10.1371/journal.pgen.1009197

Van den Berge K, Roux de Bézieux H, Street K, Saelens W, Cannoodt R, Saeys Y, Dudoit S, Clement L. Trajectory-based differential expression analysis for single-cell sequencing data. *Nat Commun*. 2020:11(1):1201. https://doi.org/10.1038/s41467-020-14766-3

Varala K, Marshall-Colón A, Cirrone J, Brooks MD, Pasquino AV, Léran S, Mittal S, Rock TM, Edwards MB, Kim GJ, et al. Temporal transcriptional logic of dynamic regulatory networks underlying nitrogen signaling and use in plants. *Proc Natl Acad Sci U S A*. 2018:115(25):6494–6499. https://doi.org/10.1073/pnas.1721487115

Wang W, Hu B, Li A, Chu C. NRT1.1 s in plants: functions beyond nitrate transport. *J Exp Bot*. 2020a:71(15):4373–4379. https://doi.org/10.1093/jxb/erz554

Wang Y, Xu J, Ge M, Ning L, Hu M, Zhao H. High-resolution profile of transcriptomes reveals a role of alternative splicing for modulating response to nitrogen in maize. *BMC Genomics*. 2020b:21(1):353. https://doi.org/10.1186/s12864-020-6769-8

Wimalanathan K, Friedberg I, Andorf CM, Lawrence-Dill CJ. Maize GO annotation-methods, evaluation, and review (maize-GAMER). *Plant Direct*. 2018:2(4):e00052. https://doi.org/10.1002/pld3.52

Woodhouse MR, Sen S, Schott D, Portwood JL II, Freeling M, Walley JW, Andorf CM, Schnable JC. Qteller: a tool for comparative multi-genomic gene expression analysis. *Bioinformatics*. 2021:38(1):236–242. https://doi.org/10.1093/bioinformatics/btab604

Wu T-Y, Goh H, Azodi CB, Krishnamoorthi S, Liu M-J, Urano D. Evolutionarily conserved hierarchical gene regulatory networks for plant salt stress response. *Nat Plants*. 2021:7(6):787–799. https://doi.org/10.1038/s41477-021-00929-7

Xing J, Zhang J, Wang Y, Wei X, Yin Z, Zhang Y, Pu A, Dong Z, Long Y, Wan X. Mining genic resources regulating nitrogen-use efficiency based on integrative biological analyses and their breeding applications in maize and other crops. *Plant J*. 2024:117(4):1148–1164. https://doi.org/10.1111/tpj.16550

Yamaguchi N, Winter CM, Wellmer F, Wagner D. Identification of direct targets of plant transcription factors using the GR fusion technique. *Methods Mol Biol*. 2015:1284:123–138. https://doi.org/10.1007/978-1-4939-2444-8

Yan D, Easwaran V, Chau V, Okamoto M, Ierullo M, Kimura M, Endo A, Yano R, Pasha A, Gong Y, et al. NIN-like protein 8 is a master regulator of nitrate-promoted seed germination in Arabidopsis. *Nat Commun*. 2016:7(1):13179. https://doi.org/10.1038/ncomms13179

Yang XS, Wu J, Ziegler TE, Yang X, Zayed A, Rajani MS, Zhou D, Basra AS, Schachtman DP, Peng M, et al. Gene expression biomarkers provide sensitive indicators of in planta nitrogen Status in maize. *Plant Physiol*. 2011:157(4):1841–1852. https://doi.org/10.1104/pp.111.187898

Yates AD, Allen J, Amode RM, Azov AG, Barba M, Becerra A, Bhai J, Campbell LI, Carbajo Martinez M, Chakiachvili M, et al. Ensembl genomes 2022: an expanding genome resource for non-vertebrates. *Nucleic Acids Res*. 2022:50(D1):D996–D1003. https://doi.org/10.1093/nar/gkab1007

Yoo S-D, Cho Y-H, Sheen J. Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat Protoc*. 2007:2(7):1565–1572. https://doi.org/10.1038/nprot.2007.199

Yu S, Zheng L, Li Y, Li C, Ma C, Li Y, Li X, Hao P. A cross-species analysis method to analyze animal models' similarity to human's disease state. *BMC Syst Biol*. 2012:6(S3):S18. https://doi.org/10.1186/1752-0509-6-S3-S18

Yuan Y, Bar-Joseph Z. Deep learning of gene relationships from single cell time-course expression data. *Brief Bioinform*. 2021:22(5):bbab142. https://doi.org/10.1093/bib/bbab142

Zaslaver A, Mayo AE, Rosenberg R, Bashkin P, Sberro H, Tsalyuk M, Surette MG, Alon U. Just-in-time transcription program in metabolic pathways. *Nat Genet*. 2004:36(5):486–491. https://doi.org/10.1038/ng1348

Zhang S, Zhang Y, Li K, Yan M, Zhang J, Yu M, Tang S, Wang L, Qu H, Luo L, et al. Nitrogen mediates flowering time and nitrogen use efficiency via floral regulators in rice. *Curr Biol*. 2021:31(4):671–-683.e5. https://doi.org/10.1016/j.cub.2020.10.095

Zhang X, Davidson EA, Mauzerall DL, Searchinger TD, Dumas P, Shen Y. Managing nitrogen for sustainable development. *Nature*. 2015:528(7580):51–59. https://doi.org/10.1038/nature15743

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W, et al. Model-based analysis of ChIP-seq (MACS). *Genome Biol*. 2008:9(9):R137. https://doi.org/10.1186/gb-2008-9-9-r137

Zhang Y, Ngu DW, Carvalho D, Liang Z, Qiu Y, Roston RL, Schnable JC. Differentially regulated orthologs in Sorghum and the subgenomes of maize. *Plant Cell*. 2017:29(8):1938–1951. https://doi.org/10.1105/tpc.17.00354

Zhou P, Li Z, Magnusson E, Cano FG, Crisp PA, Noshay JM, Grotewold E, Hirsch CN, Briggs SP, Springer NM. Meta gene regulatory networks in maize highlight functionally relevant regulatory interactions. *Plant Cell*. 2020:32(5):1377–1396. https://doi.org/10.1105/tpc.20.00080

Zhu J, Li S, Jiang H, Lv D, Ma S, Wang B, Lu X, Yang W, Chen R, Zhou X. Protoplast transient expression-based RNA-sequencing: a simple method to screen transcriptional regulation in plants. *Plant Physiol*. 2024:194(1):408–411. https://doi.org/10.1093/plphys/kiad495