Conservation and divergence of regulatory architecture in nitrate-responsive plant gene circuits

Bian C[1,2,3], Demirer GS[1,2,4], Oz MT[1,5], Cai Y[5], Witham S[5], Mason GA[2], Chen R[5], Ahn J[5], Gaudinier A[2], Brady SM[2*], Patron NJ[4*]

[1]Equal contributions. [2]Department of Plant Biology and Genome Center, University of California, Davis, Davis CA 95616 USA. [3]Frontiers Science Center for Molecular Design Breeding, National Maize Improvement Center, State Key Laboratory of Maize Bio-breeding, Department of Plant Genetics and Breeding, China Agricultural University, Beijing, P. R. China. [4]Department of Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA. [5]Engineering Biology, Earlham Institute, Norwich Research Park, Norwich, UK, NR4 7UZ. * co-corresponding authors

ABSTRACT

Nitrogen is an essential element for all life processes in plants. As such, plant roots dynamically respond to nitrogen availability below-ground by executing a signaling and transcriptional cascade resulting in altered plant growth, optimized for nutrient uptake. The NIN-LIKE PROTEIN 7 (NLP7) transcription factor senses nitrogen and along with its closely related paralog NLP6, coordinates these transcriptional responses. Here, we dissect a sub-circuit of the Arabidopsis nitrogen transcriptional network comprising AUXIN RESPONSE FACTOR 18 (ARF18), ARF9, DEHYDRATION RESPONSE ELEMENT BINDING-PROTEIN 26 (DREB26), A NAC-DOMAIN CONTAINING PROTEIN 032 (ANAC032), NLP6 and NLP7 transcription factors and their regulation of NITRATE REDUCTASE 1 (NIR1). The *cis*-regulatory architecture of these factors are defined as well as their direct transcriptional regulation and demonstrate a network with multiple multi-node feedforward loops. Conservation and divergence of this network is similarly assessed in *Solanum lycopersicum*. The resulting network models are validated by measuring the transcriptional output of the network with the synthetic *NITRATE-REGULATED PROMOTER* in varying concentrations of available nitrogen revealing rewiring of nitrogen regulation across distinct plant lineages.

***Keywords*:** nitrogen, transcription factors, network, DNA binding, root, systems biology, synthetic biology, *Solanum lycopersicum*, tomato, *Arabidopsis thaliana*

BACKGROUND

Nitrogen (N) is essential for plant growth and basic metabolic processes. The supply of N through the energy-intensive Haber-Bosch process and fertilizer application has been critical to increasing crop yields but also has negative ecological and environmental consequences (1, 2). Knowledge of the diverse regulatory events that crop species employ to respond and adapt to N deficiency can be used to breed plants with increased nitrogen use efficiency and ultimately, to lessen reliance on externally applied N. In response to limiting N, plant roots initiate a complex series of molecular events (3–5) including rapid post-transcriptional, calcium- and phosphorylation-dependent signaling cascades that converge on transcriptional regulation of N transporters, assimilation enzymes, N signaling factors, carbon metabolism and hormone pathways (6–11). Below-ground, the outcome of these responses in the root includes altered lateral root elongation to forage for N and concomitant adjustment of N metabolism. Above-ground, changes are manifested in perturbed plant growth and a reduction in yield.

Plant responses to nitrate have best been studied in the model dicot *Arabidopsis thaliana* (Arabidopsis). Many TFs have been identified that coordinate transcriptional responses to nitrate (5, 11–17). Among these, AtNLP7 (14) plays a key role by modulating the expression of downstream target genes following Ca^{2+}-sensor protein kinase (CPK)-induced phosphorylation (8) and has recently been shown to directly bind nitrate (18). However, our understanding of the upstream regulation of this important TF and how its abundance might be controlled is relatively poor.

NLP7 acts within a complex transcriptional regulatory network (17) (15). An emergent and understudied property of this network is the prevalence of feedforward loop and combinatorial interactions upstream of *NLP7* and its closely related paralog *NLP6*. Interpreting the influence of these interactions on the regulation of nitrogen metabolism (as measured via expression) of *AtNIR1,* is non-trivial and depends on the sign of each regulatory interaction (activating or repressing) and the extent to which the gene is regulated by its upstream factors. Furthermore, to deploy the knowledge gained in Arabidopsis to the rational engineering of crop species, it is also necessary to obtain evidence of which aspects, if any, of the network are conserved. To date, while NLP7 has been shown to be conserved in some crop species (19, 20), we lack evidence for other TFs and the genetic basis of nitrate-responses. To date, few molecular players associated with N signaling and metabolism in tomato have been identified (21, 22). Although there is conservation of radial patterning between Arabidopsis and tomato, there are also key differences. Tomato has three cortex layers compared to the one in Arabidopsis, the outermost of which forms a barrier composed of lignin and suberin (23). A (24) predicted gene regulatory network for the exodermis demonstrated exodermal-enrichment of nitrogen metabolic genes with SlNLP7 as a central regulatory factor. These extensive differences between tomato and Arabidopsis further demonstrate the potential for changes in the underlying regulatory circuitry that controls responses to nitrate.

In an effort to better elucidate the network architecture upstream of AtNLP6/7, we systematically study DNA binding and regulation within an Arabidopsis gene regulatory circuit comprised of *AtARF9, AtARF18, AtANAC032, AtDREB26, AtNLP6, AtNLP7 (17)*, using *AtNIR1* as a final readout of transcriptional control of N metabolism. Regions of open chromatin associated with each gene are mined for DNA binding motifs and the ability of TFs to bind to these targets are subsequently determined using an *in vitro* methodology (25). *In vivo* binding and regulation are confirmed using a modification of the Transient Transformation System for Genome-Wide Transcription Factor Target Discovery (TARGET) assay (26) and a protoplast co-expression assay. In parallel, conservation of this network in tomato is tested using phylogenomics coupled with the *in vitro* and modified TARGET system. At the top and bottom of this circuit, ARF18/ARF9 and NLP6/7, respectively, act as activators in both Arabidopsis and tomato. However, in between these factors, gene regulatory interactions differ

between species. This network model was iteratively tested and validated using single, pairwise, and higher order genetic perturbations of all factors in a novel *in vivo* assay. Collectively, this systems level approach illustrates the conservation and divergence of transcriptional regulation of nitrogen metabolism in two dicot species. Few plant regulatory networks have been experimentally characterized to this extent. Our work provides a blueprint of how systems biology techniques can be used to inform synthetic biology approaches to crop engineering by predicting the outcomes of perturbations.

RESULTS

A putative transcriptional network regulating Arabidopsis response to nitrate.

In previous work, yeast-1-hybrid (Y1H) data was used to infer a putative transcriptional network that regulates the architecture of Arabidopsis root systems in response to N availability (17) (Figure 1). However, while Y1H is a powerful technique for identifying previously unknown interactions between TFs and target genes, false positives can result from transcription initiation by endogenous yeast TFs. Similarly, false negatives may result from improper folding, mislocalization, or because required post-translational modifications or protein interactions do not occur (27). In addition, though Y1H can identify candidate target genes, it is unable to identify the motifs responsible for binding and also cannot inform if interactions are likely to result in changes to the expression of target genes within the native species. This information is critical for understanding how regulatory networks control quantitative phenotypes, and for predicting the effects of perturbations. To address this, we first looked for evidence of TF-binding by identifying candidate binding sites in the regulatory regions of target genes.

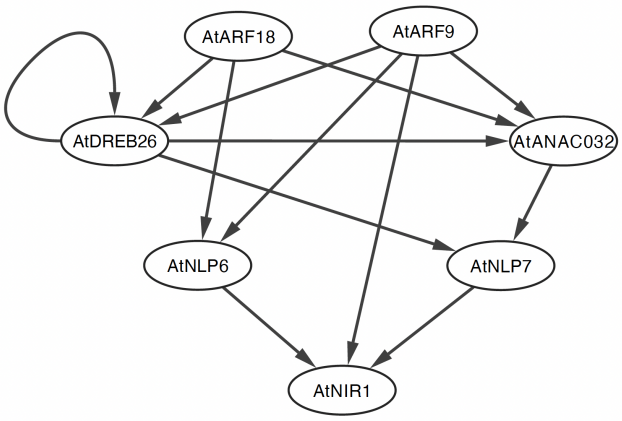

Figure 1. Nitrogen metabolism network sub-circuit in Arabidopsis that acts upstream of NLP6, NLP7 and NIR1 as determined by enhanced yeast one hybrid assays (17)**.**

In vitro characterization of candidate TF binding sites supports interaction within the regulatory circuit.

To identify candidate TF binding sites, we performed a systematic analysis of the upstream regulatory region of each gene in the regulatory circuit, with the aim of identifying candidate binding sites in regions of open chromatin. Regions for analysis were restricted to two kilobases upstream of the start of transcription, or until the next protein-coding gene. In addition, we analyzed one kilobase downstream of the start of transcription. To identify accessible chromatin, we mapped peaks from publicly available ATAC-seq data (28)(Figure 2a). We then used the FIMO package within the MEME software suite (29) to identify candidate binding sites for AtNLP7 and AtDREB26 using publicly available data to infer position weight matrices (PWMs) (30) (Supplementary Data 1). Equivalent DNA binding data for AtNLP6, AtARF9, AtARF18 or AtANAC032 was unavailable and PWMs from a closely related TF, for which the DNA-binding domain was confirmed (Supplementary Data 2), were used to identify candidate sites (see methods). This analysis identified candidate binding motifs that supported interactions within the putative Y-1-H regulatory circuit and identified the presence of further putative interactions (Figure 2a).

To validate the potential for TF-DNA-binding, the ability of each TF to bind to each candidate binding site was experimentally determined using a recently developed plate-base method for measuring the affinity of proteins for specific DNA sequences (25). To do this, we expressed recombinant proteins of AtARF18, AtANA032, AtNLP6, AtANLP7 and AtDREB26 with a genetically encoded N-terminal 9xHIS tag and a C-terminal HiBit tag. We were unable to purify sufficient soluble protein of AtARF9. We then tested the ability of purified protein to bind to 80 bp probes containing candidate binding sites, relative to random DNA probes. Example data supporting three edges within the network are shown in (Figure 2b). Data for all interactions is provided in Supplementary Data 3 and the sequences of all probes are provided in Supplementary Data 4.

From these data, we constructed an *in vitro* N network (IVNN) (Figure 2c) and identified several previously known interactions including between AtNLP6/ AtNLP7 and the Nitrogen Responsive Element (NRE) present in the promoter region of *AtNIR1 (31)*, as well as AtNLP7 with itself (32). These data also support four additional edges previously identified by Y-1-H: AtARF18 to *AtANAC032* and *AtDREB26*; AtANAC032 to *AtNLP7*; and AtDREB26 to *AtNIR1*. In addition, 12 further potential interactions were identified: AtNLP6 and AtNLP7 to *AtARF18* and *AtDREB26*; AtANAC032 to *AtDREB26;* AtDREB26, AtARF8 and AtANAC032 to *AtNIR1*; and AtDREB26 and AtANAC032 to *AtNLP6.* Some of these interactions could not have been identified by the earlier Y-1-H assay, as the motifs occur beyond the borders of the region cloned for Y-1-H assays. Finally, three instances of potential autoregulation were identified for AtANAC032, AtARF18, AtDREB26 and AtNLP6. Interestingly, AtNLP6/7, AtDREB26, and AtANAC032 were all observed to bind significantly to both AtNLP6 and AtNLP6, suggesting some conservation of network interactions. We also note that the formation of multimeric complexes involving these TFs may result in additional internal interactions. However, while these data indicate the potential for protein-DNA interactions, *in vitro* binding does not provide evidence that these interactions occur *in planta* or are regulatory in nature.

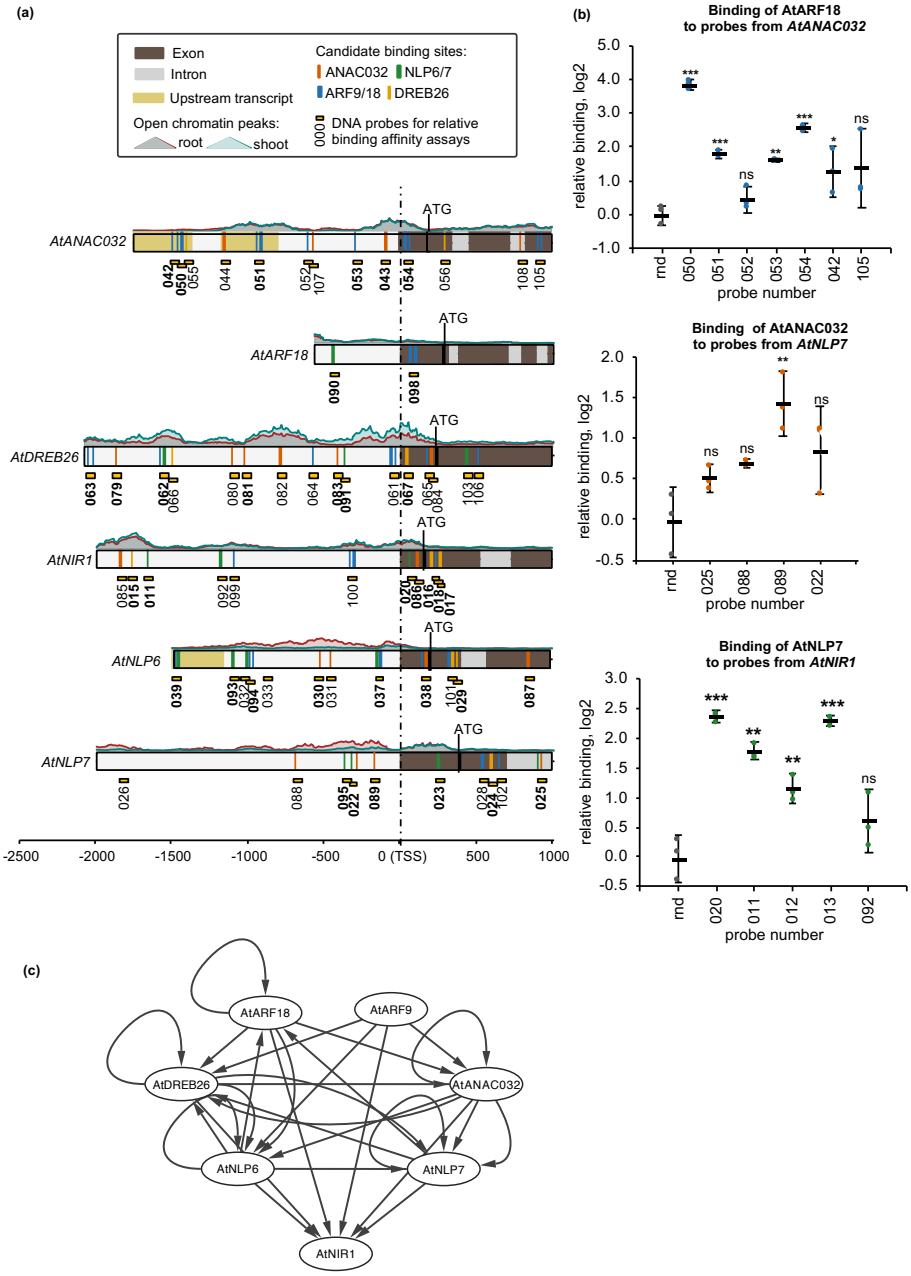

Figure 2. Transcription factor binding motifs and *in vitro* assays support the sub-network (a) Position weight matrices (PWMs) were used to identify candidate transcription factor binding sites in the 5' regions of *AtANAC032*, *AtANR1, AtARF18, AtDREB26, AtNIR1, AtNLP6* and *AtNLP7.* Binding of TFs to short DNA probes containing candidate sites (numbered boxes) was evaluated using an *in vitro* assay; bold text indicates that significant binding of at least one transcription factor was detected (Supplementary Data 3) **(b)** Representative *In vitro* binding assays showing interactions between AtARF18, AtANAC032 and AtNLP7 and probes from *AtANAC032, AtNLP7 and AtNIR1,* respectively**.** Error bars = 2 x standard error; n=3; *p*-values were calculated using an unpaired two-tailed Student's t-test of each sample to the random (rnd) control probe; *$p<0.05$, ** $p<0.01$, *** $p<0.001$; ns = not significant (c) an *in vitro* N network (IVNN) based on evidence of *in vitro* interactions between TFs and promoters.

Refining the network by protoplast expression assays to determine the regulatory consequences of TF binding.

To investigate if the TFs were able to modulate the expression of predicted targets, we tested interactions in two transient expression assays. We first used a variation of the Transient Transformation System for Genome-Wide Transcription Factor Target Discovery (TARGET) assay, in which glucocorticoid receptor (GR)-tagged TFs are expressed in protoplasts (26). This assay quantifies the expression of endogenous target genes following the treatment of cells with dexamethasone (DEX), which stimulates the translocation of TF into the nucleus. The addition of cycloheximide (CHX) in the presence of DEX prevents transcriptional cascades and tests if regulation is direct. In previous studies, transcriptional changes were quantified using RNA-seq, however, in this variation, which we term transient transformation system for specific transcription factor targets (TARSET), changes to the expression of specific target genes were quantified using qPCR. Secondly, we further investigated predicted direct interactions using an exogenous ratiometric transient luciferase (LUC) assay in which each target promoter was fused to a nanoluciferase reporter (LucN) and coexpressed in protoplasts with either a constitutively expressed TF or constitutively expressed YFP to maintain an equal transcriptional load. For experimental details, see Supplementary Data 5.

In a previous study, the application of CHX was found to have a substantial effect on the expression of *AtDREB26*, depressing expression several-fold (33). Therefore, regulatory responses of *AtDREB26* were only assessed in LUC assays, which showed activation by AtNLP7 and AtNLP6 and repression by AtARF18 (Supplementary Data 6). We also tested if the expression of other target genes was affected by CHX finding that *AtANAC032* is also upregulated by CHX. In spite of this, AtARF18 was found to repress *AtANAC032* in the TARSET assay, a change that was also observed in LUC assays (Figure 3a and b). Both assays also indicated that AtANAC032 directly represses *AtNLP7* and, as previously shown, AtNLP7 directly activates *AtNIR1* (Figure 3c-f). However, while it would therefore be expected that AtARF18 would indirectly activate expression of *AtNLP7* via repression of its direct negative regulator, AtNAC032, TARSET indicates that AtARF18 indirectly represses expression of *AtNLP7*, suggesting the involvement of additional as yet unidentified factors (Supplementary Data 6). Consistent with its role as a net negative regulator of AtNLP7, ARF18 was found to indirectly repress expression of *AtNIR1.* However, another unexpected result was the net activation of *AtNIR1* by AtNAC032, which was found to be both a direct repressor of *AtNIR1* and *AtNLP7*, again suggesting the presence of unidentified connections.

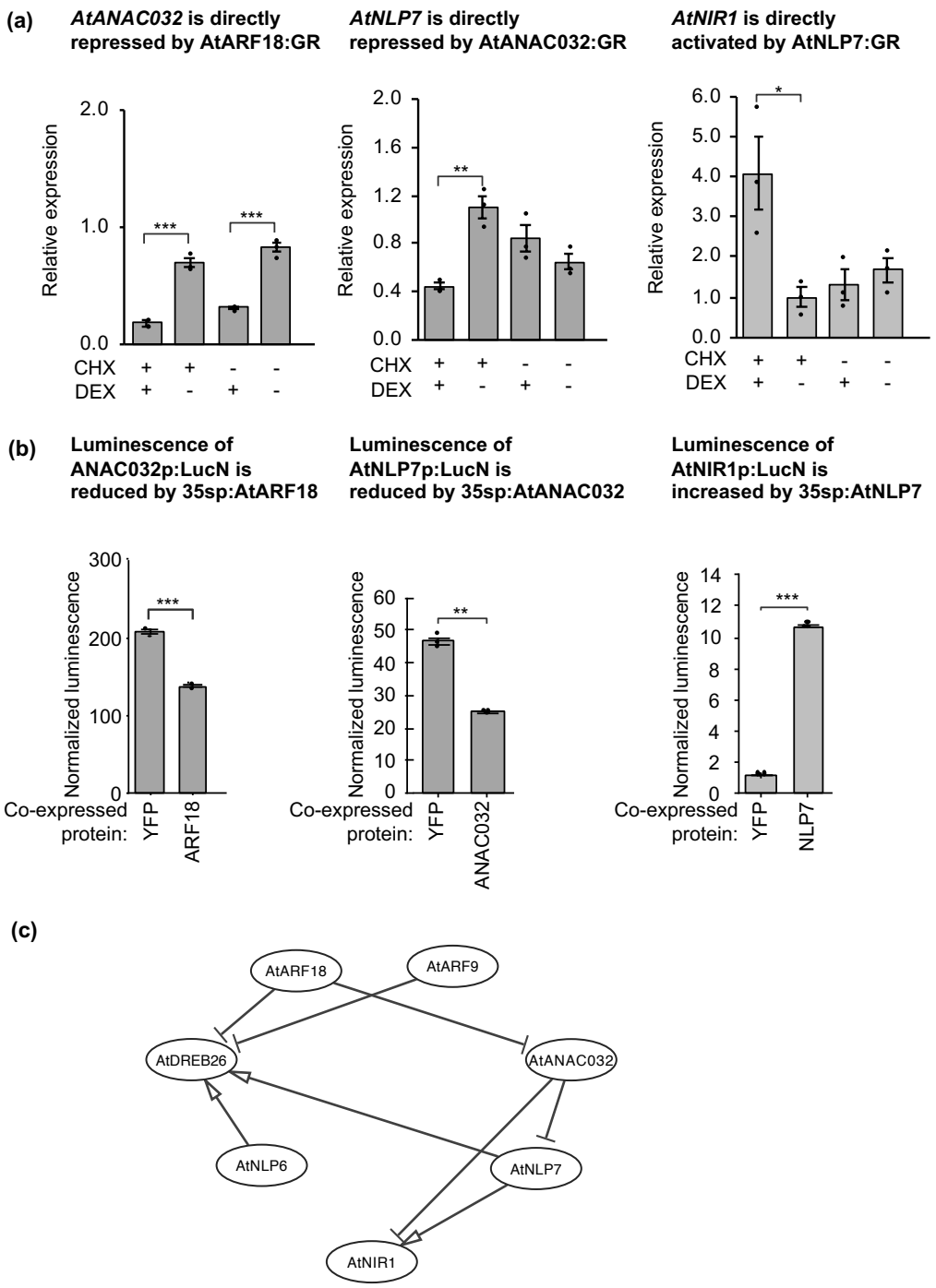

Figure 3. Examples of regulatory consequences of transcription factor expression in protoplasts. (**a**) Changes in endogenous gene expression of target genes in TARSET assays. Values represent the mean and standard error of three biological replicates of which each is the mean of two technical replicates. (**b**) Changes in normalized luminescence of nanoluciferase (LucN) in response to transcription factor overexpression. P-values were calculated using an unpaired two-tailed Student's t-test *P<0.05, ** P<0.01, *** P<0.001 **(c)** Schematic of sub-network illustrating the regulatory consequences (TARSET and/or co-expression data) of transcription factors that directly bind their target genes.

AtNLP6 and AtNLP7 were found to directly activate *AtDREB26* (Supplementary Data 6). Reciprocally, AtDREB26 indirectly repressed *AtNIR1* and *AtNLP7* (Supplementary Data 6), however the node via which this repression might be executed remains unidentified. Despite the identification of binding sites for AtDREB26 in the regulatory regions of *AtNAC032*, *AtNIR1* and *AtNLP7* (Supplementary Data 3), we did not identify any instances of direct regulation by AtDREB26.This may indicate that AtDREB26 requires an interacting partner for function, which may need to be co-expressed; previous CryY2H assays suggest that AtDREB26 interacts with at least three other TFs (34). However, we did detect strong indirect repression of both *AtNLP7* and *AtNIR1*. Finally, while *in vitro* binding to their own promoters was detected for AtNLP6 and AtNLP7(Supplementary Data 3) direct autoregulation was not detected; though both were observed to indirectly self-repress (Supplementary Data 6). The resulting regulatory network contains coherent three node regulation of *AtNIR1* by AtANAC032 via AtNLP7 and incoherent four node regulation of AtDREB26 via AtARF18, AtANAC032 and AtNLP7.

A conserved role for homologues of AtNLP7 in nitrate responsiveness has been demonstrated in several species including rice (20), and maize (19). Further, AtNLP7 is known to play a key role in rapid responses to nitrate by modulate the expression of downstream target genes following Ca2+-sensor protein kinase (CPK)-induced phosphorylation, and is also able to directly bind nitrate (8, 18). Thus, an understanding of the regulatory network controlling expression levels of this protein may provide insights into how protein levels of this important transcription factor are regulated.

We therefore investigated if this regulatory network is conserved in other plant lineages. We selected tomato as one of the most important horticultural crops, with a preference for nitrate as an inorganic nitrogen source (35). While tomato homologs of AtNLP7 have been identified (21), the molecular mechanisms of nitrate uptake and assimilation are poorly understood in this species. Further, recent data has demonstrated that some aspects of root development and gene regulation are not conserved between Arabidopsis and tomato (23).

Root system architecture and transcriptional regulatory responses to nitrate are conserved in tomato.

We first confirmed that there is conservation of root-system and transcriptional regulatory responses to nitrate. To do this, we germinated and grew seeds for ten days on media supplemented with 0 to 10 mM KNO_3. We observed that lateral root length and total root size increases from 0 mM to 1 mM with a decrease from 1-10 mM KNO3. However, this is non-linear, with growth decreasing at higher concentrations (Figure 4a-e). This is consistent with the scavenging behavior observed in Arabidopsis (36).

Samples were taken from roots grown in media supplemented with 0, 1 or 10 mM KNO_3 and RNA was extracted. RNAseq analysis revealed that canonical N regulatory genes increased in expression dependent on nitrate availability (Figure 4f). Nitrate-dependent increases in expression of *SlNIR1* and *SlNIR2* were confirmed by qPCR (Figure 4g). Analysis of GO category enrichment indicates a likely conservation of downstream biological processes including nitrogen and nitrate assimilation, glutamine and glutamate biosynthesis (Figure 4h and Supplementary Data 7).

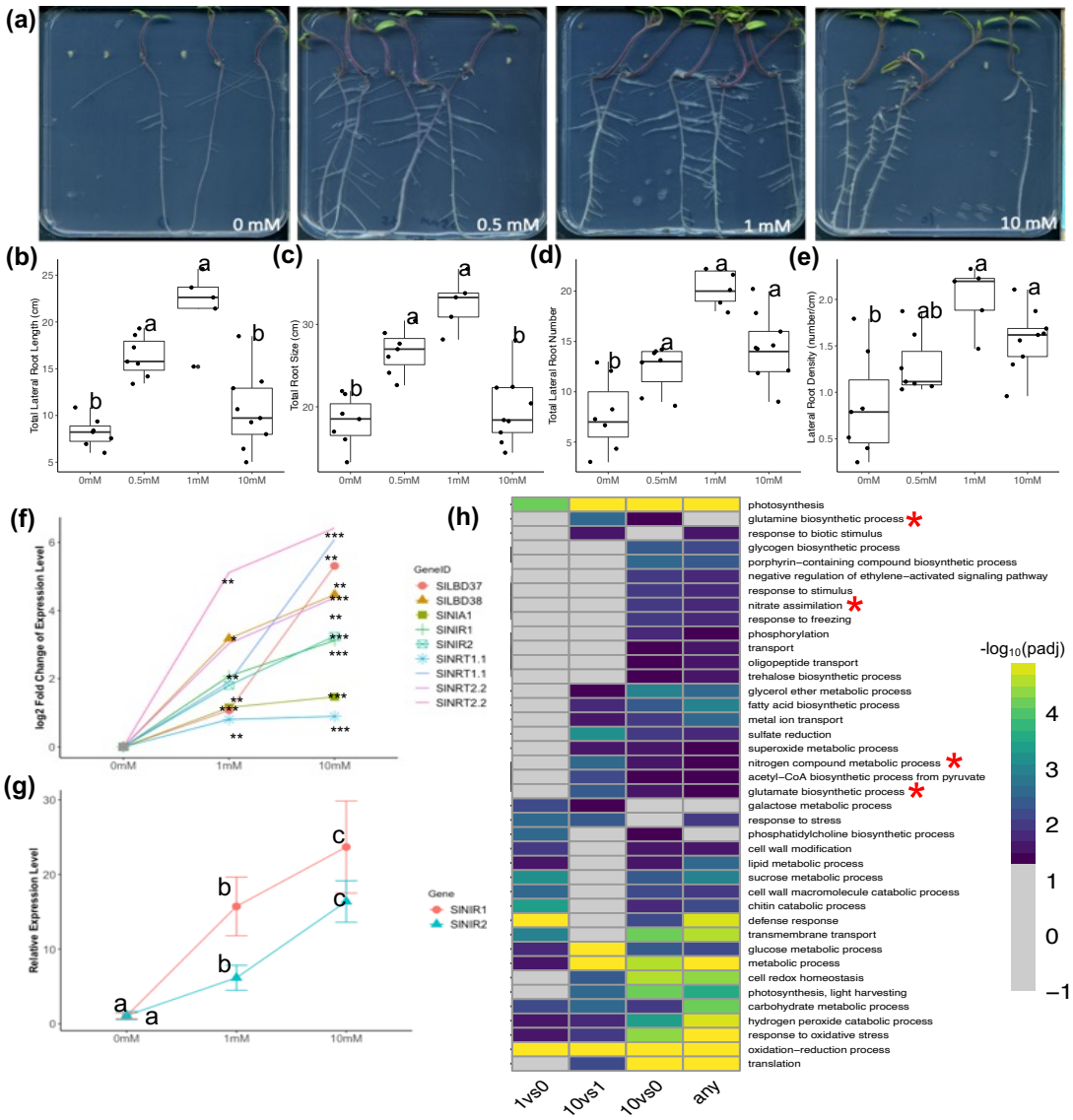

Figure 4. Root growth and gene expression response of tomato (M82) seedlings to potassium nitrate KNO_3. (a) Exemplar images of root system architecture (RSA) in increasing nitrate. Plants were imaged 10 days after germination on MS media (no nitrogen) supplemented with 0, 0.5, 1, and 10 mM KNO_3. **(b-e)** Quantification of root system architectures including total lateral root length (b), total root size (sum of main and lateral root lengths, (c), total lateral root number (d) and lateral root density (ratio of lateral root number to main root length, (e) from M82 seedlings 10 DAG grown on 0, 0.5, 1, and 10 mM KNO_3 (n=7). Boxplots show data from the 25th to the 75th percentile centered to the data median. Dots represent individual values. Statistical analysis was performed using one-way ANOVA with Tukey post-hoc test (adjusted p-value <0.05). Different small case letters indicate significance of differences between N concentrations. **(f)** Log_2 fold-changes of nitrogen-responsive genes. Asterisks indicate differentially expressed between N conditions by limma (false discovery rate <0.05). *** p-value < 0.001, ** p-value < 0.01,* p-value < 0.05. **(g)** qRT-PCR verifies that the expressions of tomato nitrogen-responsive genes SlNIR1 and SlNIR2 are induced by higher nitrogen levels in WT hairy roots. Statistical analysis was performed using one-way ANOVA with Tukey post-hoc test (adjusted p-value<0.05). Different letters indicate the significance of differences between tested N concentrations for each gene. **(h)** GO enrichment analysis of DEGs between nitrate conditions in M82 RNA-seq experiment. Some canonical nitrogen-mediated metabolisms are marked with asterisks.

What is the regulatory architecture of the tomato nitrogen response?

Given the conservation in N-mediated changes in root system architecture and transcription, we hypothesized that the underlying regulatory network is likely conserved between Arabidopsis and tomato. Using the elaborated network from Arabidopsis (Figure 3C) as a base, we investigated if the tomato genome encoded likely orthologs of *AtARF18, AtARF9, AtDREB26, AtANAC032, AtNLP6, AtNLP7* and *AtNIR1.* As expected, given the history of tomato genome triplication (37), duplicates or triplicates of many of these Arabidopsis genes were identified. Putative orthologs were first identified using phylogenetic trees (23) (Supplementary Data 8), and this gene set was refined by identifying the homolog(s) with expression in the tomato root (23). A clear likely ortholog was identified for *AtARF18* (*Solyc01g096070*), two potential *AtARF9* paralogs (*Solyc08g008380* and *Solyc08g082630*), a single *AtDREB26* ortholog (*Solyc11g012980*), two genes corresponding to either *AtNLP6* or *AtNLP7* (*Solyc08g008410* and *Solyc08g082750*) and paralogs of *AtNIR1* whose expression was confirmed to be N concentration-dependent (*Solyc10g050890* and *Solyc01g108630,* Figure 4f-g, Supplementary Figure 8). *Solyc08g008380* was chosen for further analysis given its expression in the tomato root, and was called *SlARF9B,* consistent with naming in (38). No clear *AtANAC032* ortholog was identifiable given the phylogeny. The first NLP gene, *Solyc08g008410* was found to be mis-annotated and to contain two NLP duplicates (Supplementary Data 9), the latter of which is extremely lowly expressed, and which has no support in transcript abundance. The genes were named *SlNLP7A* (*Solyc08g008410*) and *SlNLP7B* (*Solyc08g082750*) for further analyses, as it was impossible to differentiate which could be the correct ortholog for *AtNLP6* or *AtNLP7*.

Given the conservation of the wild type tomato root nitrogen response relative to Arabidopsis, we hypothesized that mutations of the genes in the top (*ARF18)* and the bottom (*SlNLP7)* of the circuit should also have similarly perturbed responses to changes in N concentration, and in the same direction (17). Root system architecture traits (primary root length, number of lateral roots, lateral root length, total lateral root length, total root length, lateral root density and ratio of lateral root length to total root length) were measured in limiting (1 mM KNO_3 for Arabidopsis, 0 and 1 mM KNO_3 for tomato) and sufficient (10 mM KNO_3) for mutant alleles of these genes in Arabidopsis and tomato. This included two previously described mutant alleles of Arabidopsis *ARF18*, *arf18-2* and *arf18-3 (17)*, and a new CRISPR-Cas9 generated allele ($arf18_c$) (Supplementary Data 10 and 11). In tomato, guide RNAs were designed to target *SlARF18* and *SlARF9b* to reduce redundancy (Supplementary Data 10 and 11); and to *SlNLP7b*. In both the Arabidopsis and tomato lines containing at least a mutation of *ARF18* (*arf18-3* and *slarf18arf9b*), a smaller root system was observed relative to wild type in at least one concentration of KNO_3 for all traits except primary root length (Figure 5 and Supplementary Data 12). In the Arabidopsis *nlp7-1* mutant, the root system was larger compared to wild type for four of the seven traits measured in at least one concentration of KNO_3 consistent with previous reports (Figure 5 and Supplementary Data 12) (17). In contrast, the tomato *nlp7b-2* mutant allele had a smaller root system compared for six of the seven traits quantified (Figure 5 and Supplementary Data 12). Although not perfectly concordant, this suggests a general genetic conservation of this circuit between Arabidopsis and tomato. Given these similarities and differences, we next looked at the degree of conservation of the network at molecular resolution.

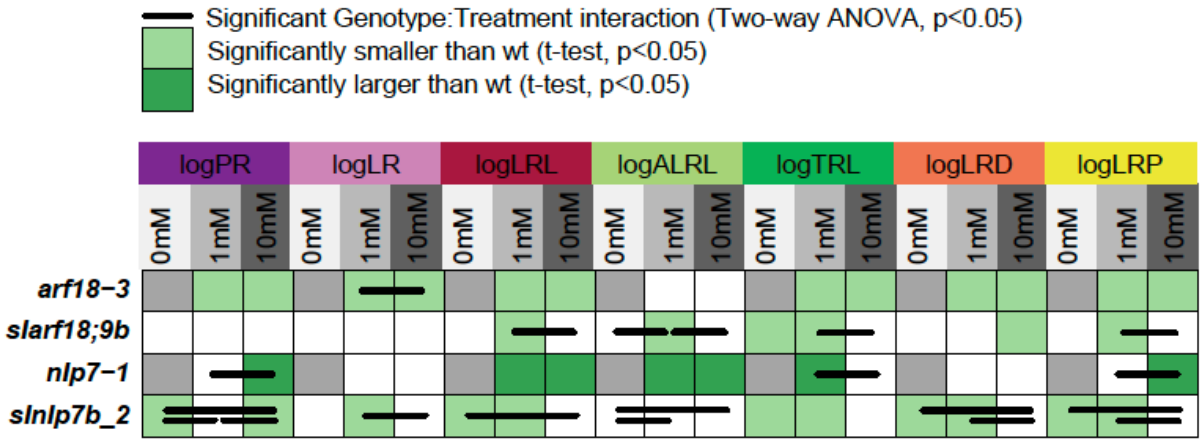

Figure 5. Tomato and Arabidopsis root system architecture traits in orthologous mutant backgrounds. Traits measured include the natural logarithm transformed primary root length (logPR), number of lateral roots (logLR), lateral root length (logLRL), total lateral root length (logALRL), total root length (logTRL), lateral root density (logLRD) and the ratio of lateral root length to total root length (logLRP). Traits were measured in 0, 1 and 10 mM KNO_3 for tomato, and 1 and 10 mM KNO_3 for Arabidopsis. Light green represents the trait as significantly smaller relative to wild type, while dark green represents the trait as significantly larger relative to wild type (t-test, p<.05). A black bar indicates a significant genotype x treatment interaction as measured by a two-way ANOVA (p<.05).

To investigate conservation of the network, we first investigated the upstream regions of *SlNIR1, SlNIR2, SLNLP7A, SLNLP7B, SLDREB26* to determine if there were candidate binding sites for ARF18/9; DREB26 and NLP6/7 TFs by using FIMO with publicly available PWM data (29, 30) (Figure 6a). There is extensive conservation of DNA binding domains between species (30, 39), and we tested if this was the case for the DNA-binding domains of the Arabidopsis and tomato transcriptions factors within the network. Indeed, extensive conservation between Arabidopsis and tomato was observed for the ARF, DREB and NLP factors (Supplementary Data 13). As a proof-of-principle, we tested if Arabidopsis NLP7 and ARF18 transcription factors were able to bind *in vitro* to candidate sites identified in the *SlNLP7A, SlNLP7B, SlNIR1* and *SlNIR2* promoters (Figure 6b). Interactions between AtARF18 and *SlDREB26* as well as *SlNIR1* were conserved, as were interactions between NLP7 and *SlNIR1/SlNIR2*. In contrast, AtARF18 was able to bind to the promoters of SlNLP7A and NLP7B, and the corresponding AtARF18-AtNLP6; and AtARF18-NLP7, were not observed (Supplementary Data 3).

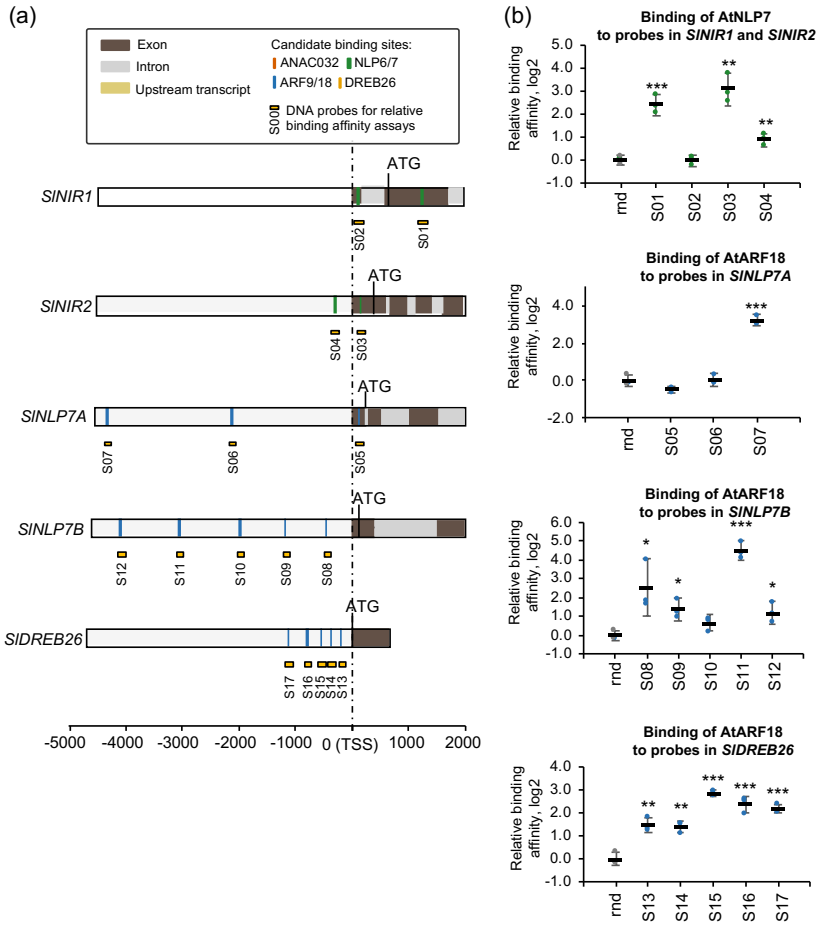

Figure 6. Transcription factor binding motifs and *in vitro* assays support conservation of the regulatory sub-network in tomato (a) Position weight matrices (PWMs) were used to identify candidate transcription factor binding sites in the 5' regions (up to 5 kilobases 2 kilobase downstream of the start of transcription) of *SlNIR1, SlNIR2, SlNLP7A, SlNLP7B and SlDREB26.* **(b)** *In vitro* relative binding assays showing binding of AtARF18 and AtNLP7 for candidate binding sites in *SlNIR1, SlNIR2, SlNLP7A, SlNLP7B and SlDREB26,* respectively. Error bars = 2 x standard error; n=3; P-values were calculated using an unpaired two-tailed Student's t-test of each sample to the random (rnd) control probe; *p<0.05, ** p<0.01, *** p<0.001; ns = not significant.

Regulatory interactions are partially conserved between Arabidopsis and tomato.

The *in vitro* TF-DNA binding assays, plus the lack of identification of a clear tomato ANAC032 ortholog, suggested distinct differences in the wiring of these nitrogen metabolic regulatory networks both in terms of binding and regulation. To investigate these regulatory interactions, we generated glucocorticoid fusions of the SlARF9B, SlARF18, SlDREB26, SlNLP7A and SlNLP7B transcription factors. In order to eliminate dosage-dependent effects, we introduced these constructs into their respective mutant backgrounds, as generated by CRISPR-Cas9 gene editing. We utilized transformation by *Rhizobium rhizogenes*, which generates transgenic roots in weeks. These roots, hereby called “hairy roots" have similar growth and anatomy to primary roots (40) and transcriptional reporters result in the same expression patterns, when comparing *A. tumefaciens*-transformed plant roots to *R. rhizogenes*-transformed hairy roots (23). Edits were confirmed by amplicon-based

sequencing (Supplementary Data 11). As with the assays in Arabidopsis (Figure 3, Supplementary Data 6), transcription factor activity was induced by application with dexamethasone, and direct targets were confirmed by the addition of cycloheximide.

Although the Arabidopsis ARF18 was able to bind to the *SlNLP7A* and *SlNLP7B* promoters, this was not reflected in the TARSET assay. However, an indirect repressive interaction was observed (Supplementary Data 14). SlARF18 and SlARF9B directly bound to the promoter of *SlDREB26* and repressed its expression (Figure 7a-b). In turn, SlDREB26 was found to directly repress the expression of *SlNLP7B* (Figure 7c). SlNLP7A and SlNLP7B both bind to and activate the expression of *SlNIR1* (Figure 7d-e).

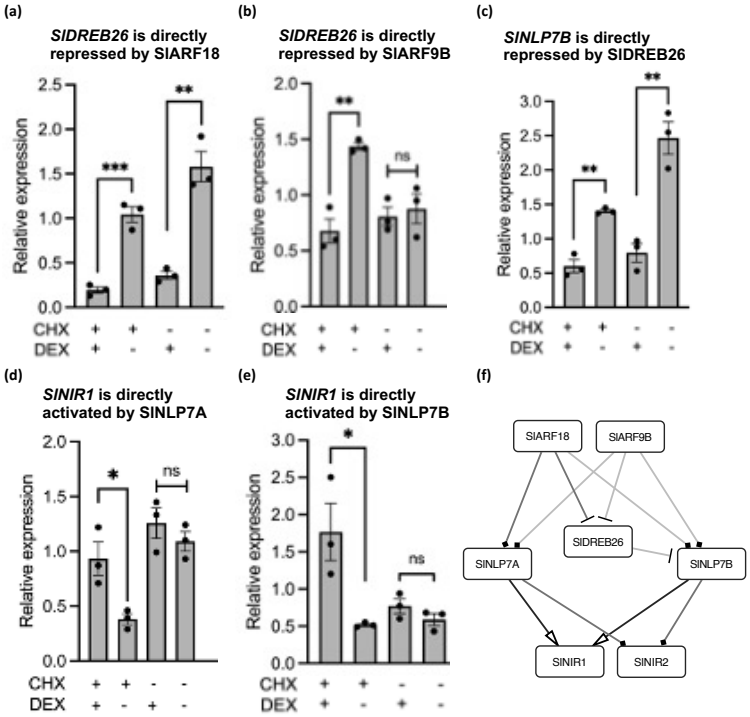

Figure 7. Examples of regulatory consequences of transcription factor expression in tomato root protoplasts via TARSET assays. Changes in endogenous gene expression of target genes in TARSET assays. SlDREB26 expression levels changed by TFs (**a**) SLARF18 and (**b**) SLARF9B, (**c**) SLNLP7B expression levels by SLDREB26, SlNIR1 levels by (**d**) SlNLP7A and (**e**) SlNLP7B. Values represent the mean and standard error of three biological replicates of which each is the mean of two technical replicates. P-values were calculated using an unpaired two-tailed Student's t-test, *P<0.05, ** P<0.01, *** P<0.001, and **** P<0.0001. N = 3. (**f**) Schematic of sub-network illustrating the regulatory consequences of transcription factors that directly bind their target genes in tomato.

In several cases, these represent conserved regulatory interactions between Arabidopsis and tomato. ARF18 and ARF9 both repress the transcription of *DREB26*, and NLP7A/B/NLP6/7 activate the transcription of *NIR1*. Major differences in these networks include the tomato DREB26 repression of *NLP7B* which is not observed in Arabidopsis, the Arabidopsis NLP6 and NLP7 activation of *DREB26* which is not observed in tomato, and the Arabidopsis NAC032 interactions which were not able to be tested in tomato due to the lack of a clear ortholog. No feedforward loops (three- or four-node) were identified in tomato, although they were present in Arabidopsis.

The network models predict the phenotypes of perturbation.

Collectively, these data in both Arabidopsis and tomato provide a predictive model of genetic regulation of plant nitrogen metabolism, and demonstrate differences in their wiring and components. In summary, DREB26 regulates *NLP7* in tomato and not Arabidopsis. ANAC032 is a component of the Arabidopsis network and is likely absent in tomato. NLP6 and NLP7 activate expression of *DREB26* and result in interconnected feedforward loops in Arabidopsis (Figure 3c), but these regulatory interactions are absent in tomato (Figure 7).

How are these differences and similarities in the mapped regulatory networks different between Arabidopsis and tomato? Major differences between the two include DREB26 as a key intermediary between ARF18/9 and NLP7A/7B in tomato, as opposed to a target of NLP7 as opposed to a target of both ARF18/9 and NLP6/7 in Arabidopsis. We tested these predictions using a single, synthetic output of the nitrogen transcriptional regulatory network. The synthetic *NITRATE-REGULATED PROMOTER (NRP) (41)* is robustly activated as concentrations of available nitrate increase, and has been used as a molecular tool to elucidate N regulation in both Arabidopsis and crop plants (11, 19, 41). We confirmed that the *NRP* acts as a robust N response reporter in tomato by generating hairy (*R. rhizogenes*-transformed) roots with the *NRP*::nls-GFP, the *AtNIR1*::nls-GFP reporter (Supplementary Data 15) and compared it to the N-inducibility of *SlNIR1* (Supplementary Data 15. These were further confirmed in *A. tumefaciens*-transformed stable tomato lines (Supplementary Data 15). Therefore, the *NRP* was used as a proxy of the output of the tomato and Arabidopsis transcriptional nitrate response.

We tested our models (Figures 3, 7), using a synthetic biology approach called PAROT: Protoplast Assay Reporting Overall Effects of TFs. Here, transfection of an NRP-driven luciferase reporter (Figure 8a) is introduced into either Arabidopsis or tomato protoplasts derived from wild type (Col-0 or M82, in Arabidopsis or tomato, respectively), Arabidopsis mutants or tomato hairy root mutant roots (Figure 8e,f, Supplementary Data 10,11) (relative to Col-0 primary roots or empty vector-transformed M82) and tested in limiting and sufficient N (0, 1 and 10 mM KNO_3). Equivalent to observations in intact plants, the NRP reporter is activated in wild type protoplasts with increasing concentrations of N (Figure 8b). We first tested the genetic regulation of the NLP6/7 paralogs (Arabidopsis - *nlp6, nlp7-1)* and tomato (*nlp7a/7b*) (Figure 8c,e, Supplementary Data 16c). In both Arabidopsis and tomato, a single *nlp6* or *7* mutant alone shows reduced sensitivity in its NRP response, with the double in tomato showing near complete insensitivity (Figure 8c,f, Supplementary Data 16a,d).This is consistent with AtNLP7 as a major regulator of *AtNIR1* expression in Arabidopsis (32, 42), and tomato (Figure 5).

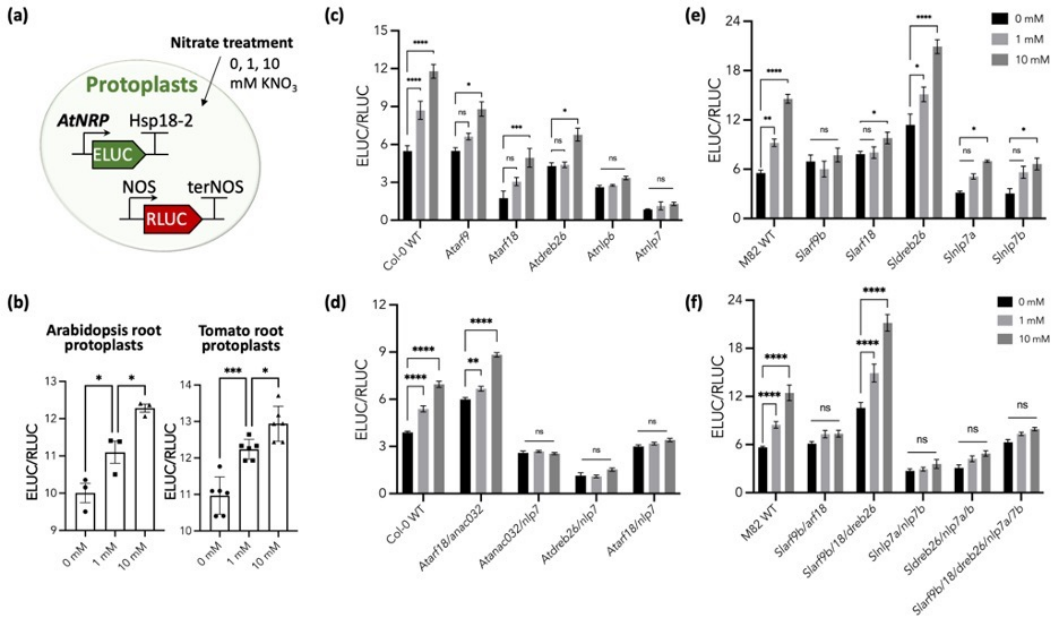

Figure 8. NRP:reporter (PAROT) assay in Arabidopsis and tomato mutants. (a) Schematic of the PAROT assay, in which protoplasts are transformed with a dual vector setup, where nitrate-responsive AtNRP drives the expression of Emerald luciferase (ELUC) and constitutive NOS promoter expresses the Red luciferase (RLUC). (b) Changes in normalized luminescence of Eluc in response to nitrate concentration in extracted Arabidopsis and tomato wild-type root protoplasts. P-values were calculated using an unpaired two-tailed Student's t-test *P<0.05, ** P<0.01, *** P<0.001, and N = 3 for Arabidopsis and N = 6 for tomato independent protoplast transformations. Normalized ELUC response of (c) Arabidopsis Col-0 WT, *Atarf9b*, *Atarf18-2, Atdreb26$_c$, Atnlp6*, and *Atnlp7-1* (d) Arabidopsis Col-0 WT, *arf18-2/anac032, anac032/nlp7-1, dreb26$_c$/nlp7$_c$, and arf18-2/nlp7-1* (e) tomato M82 WT, *Slarf9b #8, Slarf18 #2, Sldreb26 #10, Slnlp7a #1, and Slnlp7b #13* (f) tomato M82 WT, *Slarf9b/18 #4, Slarf9b/18/dreb26 #9, Slnlp7a/7b #21, Sldreb26/nlp7a/7b #6, and Slarf9b/18/dreb26/nlp7a/7b #2* at 0, 1, and 10 mM KNO_3. Error bars are standard error; N=4; P-values were calculated using two-way ANOVA with Tukey's multiple comparisons test; *p<0.05, ** p<0.01, *** p<0.001; **** p < 0.0001, ns = not significant.

Moving from the bottom of this regulatory circuit to the most upstream regulatory gene, we next tested the outcome of single or pairwise interactions in Arabidopsis. A CRISPR allele of *DREB26* showed reduced responsiveness of the NRP, but not complete insensitivity (Figure 8a, Supplementary Figure 16a). The double *dreb26* and *nlp7* mutant was insensitive to different concentrations of KNO_3, a near identical NRP phenotype to that of *nlp7-1* which is consistent with *NLP7* acting downstream of DREB26 (Figure 8b, Supplementary Figure 16a-b). The double *anac032* and *nlp7-1* mutant also demonstrated a near identical phenotype (KNO_3 insensitivity) to that of *nlp7-1* although the insensitivity of *dreb26/nlp7-1* was more extreme than that of *anac032/nlp7-1* (p value = 0.0065, Supplementary Data 16b). These are consistent with a synergistic interaction of DREB26 in repression of *NLP7* and the output of this network. *ARF18* and *ARF9* mutant alleles showed reduced NRP responsiveness although not as extreme as that of the *nlp6* or *7* mutants (Figure 8). When in the presence of the *anac032* mutant allele, the *arf18-2/anac032* allele no longer demonstrates reduced sensitivity to N, but instead a slightly enhanced responsiveness (Supplementary Data 16). Finally, an *arf18-2/nlp7-1* double mutant shows insensitivity; consistent with their roles as both primary upstream and downstream regulators of this circuit.

Similarly, in tomato, the *nlp7a* and *7b* single mutants show reduced responsiveness, while the double *nlp7a7b* mutant shows insensitivity to changes in N concentration (Figure 8). A mutant allele of *dreb26* shows elevated levels of the NRP reporter compared to wild type, and thus enhanced sensitivity (Figure 8). The *dreb26/nlp7a/7b* mutants show reduced sensitivity to nitrogen, indicating epistasis of the *nlp7a/7b* phenotype (Figure 8). Similar to Arabidopsis, the *arf9b* and *arf18* single mutants show reduced sensitivity to N, with insensitivity in the double mutant (Figure 8). The *arf9/18/dreb26* mutant shows a similar phenotype to that of *dreb26*, while the *arf9/18/dreb26/nlp7a/7b* mutant is insensitive to changes in N concentration, albeit with an increased magnitude of NRP activation (Figure 8). These data collectively represent differences in the intermediate regulation between ARF18/9B and NLP7(A/B or 6/7).

DISCUSSION

Changes in nitrogen availability are perceived and transduced via a series of signaling events culminating in rapid changes in transcription followed by mediation of plant growth for adaptive purposes (7, 11). Here we present data supporting interactions within a module of transcription factors that regulate responses to nitrate in Arabidopsis and tomato. At the nexus of this module is the NLP7 transcription factor which has recently been shown to act as a direct sensor of N itself (18) along with its paralog NLP6 (43). The ultimate goal in the study of plant nitrogen responses is to sufficiently elucidate this regulation to bypass the excessive application of nitrogen-based fertilizer via improvement of plant nitrogen use efficiency. Our approach and findings complement the many elegant network-focused studies that have elucidated several factors (9, 10, 32, 44–49) by focusing on the largely understudied transcriptional regulation of NLP6 and NLP7. We have further extended our knowledge of how this regulation is conserved in the crop species, tomato.

One means by which nitrogen use efficiency can be improved is by the targeted engineering of regulatory circuits. Although NLP7 directly senses nitrate and its activity is post-translationally controlled, an alternative and complementary mode is by control of *NLP7* abundance and thus, its availability to directly regulate nitrogen metabolic enzyme transcription. In Arabidopsis, ARF18 and ARF9 genetically regulate *ANAC032* and *NLP6/7 (17)*. Enhanced yeast one hybrid assays identified DREB26 as a gene with high betweeness centrality via connection of upstream factors (including ARF18/9) with that of NLP6/7, nitrate and nitrite reductase, glutamine synthetase and glutamine synthase (17). We sought to clarify regulatory relationships between these six genes and NIR1 as a representative nitrogen metabolic enzyme; via identification of the relevant *cis*-regulatory elements. Ideally, these will provide the logic by which these factors dictate coordinate transcriptional regulation of nitrogen metabolism. Putative transcription factor binding sites were intersected with previously defined regions of chromatin accessibility (28) to generate a set of candidate *cis*-regulatory elements (Figure 2a). *In vitro* binding assays (25) further restricted these to elements which were able to directly recruit the factors under study (Figure 2b). Collectively these indicated a complex sub-circuit comprising 29 interactions via 46 *cis*-elements, albeit without the contribution of ARF9 (Figure 2b-c). Binding affinity for ARF9 was not able to be tested as the protein was unable to be purified. Given the similarity in its DNA binding domain relative to ARF18, we anticipate that it would have had similar affinity as ARF18.

Dissecting the necessity of *cis*-regulatory elements or *cis*-regulatory modules has defined rules for cell type-specific expression in the bundle sheath (50) and for tomato fruit size and maize yield-related traits (51, 52). Testing this set of 47 *cis*-regulatory elements in these seven genes, however, is premature in the absence of evidence for direct transcriptional regulation. The modified TARGET assay (TARSET)

assay dramatically reduced the 29 possible interactions to eight via 15 putative *cis*-regulatory elements (Supplementary Data 3). This reduced set is highly conservative given that they were tested in mesophyll protoplasts bathed in only a single concentration of nitrogen and thus do not reflect nitrogen-inducibility nor cellular context. Transient interactions are also under-represented in this assay type but are prevalent within the nitrogen regulatory network at large (32). One critical target gene was not included in this refined sub-circuit - DREB26 due to autoactivation by CHX, as previously reported (33). Thus, four regulatory interactions remain untested *in planta* (ARF18, ARF9, NLP6 and NLP7 interactions with DREB26 as a target gene). Although not explicitly tested in the TARSET assay, *NLP6* and *NLP7* were not identified as regulatory targets of NLP7 in a previous study and thus there is no *in vivo* evidence for their autoregulatory interactions (32). The resulting sub-network architecture contains several interaction types - a feedforward loop and two four-node loops that can give rise to dynamic, non-linear behavior.

While Arabidopsis serves as an excellent organism to study genome-wide transcriptional regulation of nitrogen metabolism, only a limited number of studies have explored direct transcriptional targets in crop species - including rice, maize and wheat (53–59). This study complements those in another crop species - tomato. Instead of an unbiased genome-wide approach, the role of specific transcription factor genes in the ARF18/9 sub-circuit were examined. Tomato responses to nitrogen availability have been minimally studied (60, 61), likely due to its status as a specialty crop species that is most often grown in relatively smaller-scale agricultural systems with unlimited access to mineral nutrients via fertilizer application. Our analysis of root system architecture and the global transcriptional response to differing concentrations of available nitrogen indicate a remarkable similarity to the Arabidopsis response. However, the response of tomato to 0 mM KNO_3 is more similar to that of Arabidopsis at 1 mM KNO_3, perhaps due to differing amounts of N within maternal stores. At the molecular level, there were some differences in gene content with at least two tomato *NIR*, *NRT* and *ARF9* genes relative to their Arabidopsis counterparts. *ANAC032* was an exception: Arabidopsis *ANAC2, ANAC032, ANAC102* and *ANAC81* were found in an independent clade relative to three related tomato genes (*Solyc04g009440, Solyc06g060230* and *Solyc11g017470*), none of which could be reliably assigned to a respective Arabidopsis ortholog, even with cellular resolution data (Supplementary Data 8) (23). This can be interpreted in two distinct ways - either that there is no tomato *ANAC032* ortholog, or that one (or more) of the three tomato genes serve the function of the Arabidopsis gene. None of the tomato gene promoters tested contain candidate ANAC032 binding sites (Figure 6), which supports the former hypothesis. Analysis of tomato *nlp7b* and *arf18/9* root system architecture revealed increased complexity within the network. While the ARF mutants in each species are smaller relative to wild type for five traits, the Arabidopsis/tomato *nlp7* mutants have opposing growth phenotypes (Figure 5).

Dissection of the sub-circuit with both the *in vitro* and *in vivo* assays provided evidence of differences in the architecture of this module. Given the assumption that the Arabidopsis TFs serve as an appropriate proxy to examine binding of tomato *cis*-elements, interactions could exist between the seven tomato genes via twelve *cis*-elements (Figure 6). TARSET assays further confirmed five of these as direct and regulatory (Figure 7f). The remaining that could not be validated could represent a true lack of interaction, or transient interactions not able to be captured by this system (32). No feedforward or multinode loops were identified (Figure 7). Key changes in circuitry between the two species include (i) the repressive function of SlDREB26 on *SlNLP7B* while in Arabidopsis DREB26 does not regulate any target genes; (ii) in Arabidopsis NLP6 and NLP7 activate *DREB26* transcription with no such interaction in tomato; and (iii) the lack of ANAC032 to the tomato module. Morphologically, the root system architecture mutant phenotypes suggest the existence of a divergent developmental regulatory

network (Figure 5). Molecularly, the variation in sub-circuit wiring (Figure 3c, 7f) suggests differences in the final output of this network - changes in nitrogen metabolic gene expression. We chose to test this hypothesis and our model by systematic genetic perturbation. We drew on approaches used in Arabidopsis to monitor the transcriptional outcomes of auxin and cytokinin hormone signaling pathways via a synthetic reporter (62–64) - the nitrogen-responsive NRP (41). This reporter combined with the PAROT assay allowed rapid assessment of the network output - transcriptional regulation, specifically in differing concentrations of available nitrogen.

In Arabidopsis, NLP6 and NLP7 are activators of the nitrogen transcriptional response, supporting previous studies (7, 11). Both the *anac032/nlp7* and *dreb26/nlp7* mutants were insensitive to changes in KNO_3 with *dreb26/nlp7* displaying less expression of the NRP reporter (Figure 8d). The Arabidopsis network (Figure 3c) has no path connecting DREB26 to *NLP6* or *NLP7*. Therefore, there is likely a missing intermediate within the network. The net loss of *NRP* expression in *nac032/nlp7* supports the coherent feedforward loop between ANAC032, *NLP7* and *NIR1* (Figure 3c, 8d). The *arf18* and *arf9* mutants showed insensitivity to changes in KNO_3, similar to the *nlp6/7* mutants, although less extreme. ARF18 is connected to *NIR1* via *ANAC032* with two repressive interactions (Figure 3c). An additional set of three interactions connect ARF18 to *NIR1*, via *ANAC032* and *NLP7* (Figure 3c). The net outcome of both paths is coherent (net positive), which validates the current network model.

Like Arabidopsis, tomato *nlp7* mutants are insensitive to changing concentrations of KNO_3 supporting their roles as activators of the nitrogen transcriptional response. The NRP response is highly activated in the tomato *dreb26* mutant, consistent with the network model. Inconsistent with the model - the *dreb26nlp7* mutants are insensitive to changes in KNO_3 as measured by the NRP (Figure 8e-f). This could suggest a missing gene within the network that is downstream of DREB26 and which binds to and regulates *NLP7*, or a minor influence of DREB26 to the network overall relative to the major influence of NLP7. Also like Arabidopsis, tomato *arf18/9* mutants were insensitive to changing KNO_3 concentrations as measured by NRP activity (Figure 8f). The tomato *arf18/9/dreb26* mutant had an enhanced NRP response, while the *arf18/9/dreb26/nlp7a/nlp7b* mutant displayed an insensitive NRP response (Figure 8f). The latter quintuple mutant supports the network model, but not the former triple mutant. Again, there could be varying contributions of DREB26 to this network, or a missing gene.

Collectively, these data illustrate divergence and conservation of this regulatory circuit that determines transcriptional regulation of nitrogen metabolism. In Arabidopsis, these validated multi-node feedforward loops provide a framework to understand the dynamic ways in which the circuit can precisely regulate the nitrogen transcriptional response dependent on variable levels and times of availability of N within the system. In contrast, in tomato, although the quintuple mutant validates the overall network architecture, the triple mutants involving *dreb26* indicate missing pieces of information that need further testing - either through the identification of additional gene(s) or determining the circumstances by which DREB26 may have a minor input on network behavior.

This network was also validated in a protoplast-based assay, which provides an experimentally tractable approach to map regulatory interactions in the presence of multiple concentrations of nutrients (26, 65). However, the differences in RSA of the Arabidopsis and tomato *nlp* mutants do not correspond with the NRP reporter assays. The protoplast assays reflect a whole-organ response and root system architecture traits are determined by the activities of discrete populations of cells in either xylem pole pericycle cells, lateral root founder cells, and the root stem cell niche. Cell type-resolution assays can be used to elucidate these aspects of the network - either through cell type-specific nitrogen reporters, or through the enrichment of these specific cell populations.

In the future, engineering of these regulatory circuits in both species can be tackled by targeted dissection of the *cis*-regulatory modules mapped within the promoters of genes within the circuit. While parts of the network are conserved between Arabidopsis and tomato, the detection of rewiring indicates remodeling of regulation across plant lineages. It remains to be determined in each species exactly which *cis*-elements are responsible for transcriptional regulation of the target genes as well as how transcription factors act together to regulate target gene expression. Is this by distinct affinities of the transcription factor for its target or differing concentrations of the transcription factor itself? Is there competition between transcription factors in the cases of feedforward loops? Are there differences in transcription factor binding site usage dependent on the nutrient availability context? Indeed, many questions remain to be answered, but this work demonstrates how diverse assays can be used to provide blueprints for elucidation of network architecture and for future synthetic approaches to engineer improved nitrogen use efficiency in plant roots.

MATERIALS AND METHODS

Identification of candidate transcription factor binding sites.

Upstream sequence regions consisting of 2 kilobases (Kbs) upstream of the earliest annotated transcription start site (TSS) and 1 Kb downstream of *AtANAC032* (AT1G77450), *AtARF18* (AT3G61830), *AtARF9* (AT4G23980), *AtNLP6* (AT1G64530), *At NLP7* (AT4G24020), *AtDREB26* (AT1G21910) and *AtNIR1* (AT2G15620) were extracted from the Arabidopsis TAIR 10 (66) genome assembly. These regions were truncated if they overlapped with the transcript of a protein coding gene. Alternatively, they were extended if a longer region was included in previously reported yeast-1-hybrid data (17). ATAC-seq data from (28) were used to identify regions of open chromatin in Arabidopsis roots and shoot tissues. Individual .bed files for each replicate were concatenated and BEDTools (v2.29.2) merge (67) was used to combine overlapping peaks. Candidate binding sites for AtNLP7 (TGNCYYTT) and AtDREB26 (CCRCCGHC) were identified using position weight matrices (PWMs) from DAP-seq (30) using FIMO (v5.4.1) (68) with a p-value cut-off of 10^{-4} and a zero-order background model created using fasta-get-markov (29). Motifs with q-value <0.05 were considered high-confident binding sites. PWMs are provided in Supplementary Data 1. As no empirical binding motif data was available for AtNLP6, AtARF9, AtARF18 or AtANAC032, we identified closely related proteins from the same TF families for which data was available and compared sequence identity within the DNA binding domains (Supplementary Data 2). Finding these domains to be highly conserved, the PWM for AtNLP7 was also used to predict sites for AtNLP6 (TGNCYYTT), and the PWM for ATAF1 (KACGTR) was used to predict binding sites for ANAC032 (Supplementary Data 1). The PWM for AtARF2 was used to predict binding sites for AtARF9 and AtARF18. In addition, TGTCTC and TGTCGG motifs were also annotated as candidate sites as they have previously been shown to bind auxin response factors (ARFs) with higher affinity than other auxin response elements (69). Finally, pairs of TGTCNN motifs spaced less than 14 bp apart were also annotated as candidate ARF TFBSs, as paired motifs were previously shown to be enriched and preferentially bound in DAP-seq data (69, 70).

For tomato, upstream regions consisting of 5 Kbs upstream and 2 Kbs downstream of the annotated TSS of *SlDREB26* (*Solyc11g012980*), *SlNIR1* (*Solyc10g050890)*, *SlNIR2* (*Solyc01g108630), SlNLP7A* (*Solyc08g008410*) and *SlNLP7B* (*Solyc08g082750*) were extracted from *Solanum lycopersicum* ITAG5.0 (https://phytozome-next.jgi.doe.gov/). To confirm similarity, the DNA binding domains of tomato and Arabidopsis orthologs of ARF9, ARF18, DREB26, NLP6 and NLP7 were aligned (Supplementary data 13). Candidate binding sites in the upstream regions of the tomato genes defined

above were then analyzed using the same PWMs and parameters as defined above for Arabidopsis. A custom Jupyter Notebook (71) was used to create schematics of upstream regions illustrating gene features, candidate binding sites and open chromatin.

In vitro transcription factor binding assays.

Coding sequences of AtANAC032, AtARF18, AtARF9, AtNLP6, At NLP7, and AtDREB26 were amplified using primers to introduce the Gateway® recombinase sequences attB1 and attB2 and a C-terminal HiBIT tag. Amplicons were cloned into pDONR207 (attP1-ccdB-attP2) to create pENTR clones (attL1-TF:HiBiT-attL2) using Gateway®BP Clonase™II Enzyme Mix (Invitrogen). The resulting pENTR clones were combined with the Gateway®-ready pDEST vector, pH9GW(attR1-ccdB-attR2), a modified pET28a(+) vector (Novagen), using the Gateway®LR Clonase™II Enzyme Mix (Invitrogen). Details of all plasmids are available in Supplementary Data 17. Sequence-verified expression clones were transformed into *E. coli* BL21 and cultures were grown and protein purified as described in (25). Protein yield was quantified by measuring luminescence with the Nano-Glo® HiBiT Extracellular Detection System (N2420, Promega, Madison, WI) and HiBiT Control Protein (N3010, Promega). We were unable to obtain soluble protein of AtARF9. Details of all plasmids are available in Supplementary Data 17. Initially, binding of each TF to a double strand probe (DSPs) containing a known binding site was verified (25). Subsequently, the ability of each TF to bind to DSPs containing candidate binding sites from Arabidopsis and tomato promoters, relative to random DSPs, was quantified using the same method. Details of all probe sequences are available in Supplementary Data 4.

Transient transformation for specific transcription factor targets (TARSET).

For Arabidopsis, constructs were generated using the Loop assembly toolkit (72) using one-step restriction-ligation protocols detailed in (73). Coding sequences of each TF were cloned (concurrently removing the stop codon as well as any internal instances of BsaI and SapI with point mutations to introduce synonymous codons with similar Arabidopsis codon usage) into pUAP4 (72) to create Level 0 standard parts comparable with the Phytobrick standard (74). These level 0 parts were assembled into the level 1 Loop acceptor (pCk2; Addgene #136696) in a one-step restriction-ligation reactions with a double CaMV35s- ΩTMV promoter/5′ UTR (pICH51288; Addgene #50269), a C-terminal GR tag (pEPOZ0CM0137; Addgene #197535) and the CaMV35s terminator (pICH41414). Details of plasmids are in Supplementary Data 17. Col-0 seeds were germinated and grown in potting medium (two-parts sieved compost to one-part sand) within controlled environment chambers with a 16 h photoperiod at 22 °C with 120–180 μmol/m^2/s light intensity. The photoperiod was reduced to 8 h two days before leaves were harvested for the preparation of mesophyll protoplasts. Mesophyll protoplasts were prepared from *A. thaliana* Col 0 as previously described (75). For each transcription factor, protoplasts were quantified and divided into twelve aliquots of 200 μL in transfection buffer (0.4 M mannitol, 15 mM $MgCl_2$, 4 mM MES, pH 5.6), each containing approximately 1 x 10^4/mL intact protoplasts. In addition, for each batch of protoplasts, three aliquots were transfected with a control plasmid expressing nuclear-localized YFP. Plasmid DNA was prepared using the Plasmid Plus Midi kit (Qiagen, Hilden, Germany) except that three additional wash steps were performed before DNA elution from the column. 10 μg of purified DNA in transfection buffer (2 g of PEG Mn 4000 (Sigma, 81240) in 2 mL of 500 mM mannitol and 0.5 mL of 1M $CaCl_2$) was mixed with each aliquot of protoplasts before protoplasts were washed and resuspended in 300 μL of washing buffer (154 mM NaCl, 125 mM, $CaCl_2$, 5 mM KCl, 2 mM MES; pH5.6). Following incubation at 24ºC with 70 μmol/m^2/s constant light, the transformation efficiency of each batch was confirmed to be >70% in the control aliquot by quantification of YFP fluorescence in the nuclei using an inverted fluorescence microscope (Zeiss Axio Observer Z1 or ThermoFisher Evos). Aliquots transfected with TF:GR plasmids were treated with 20 mm KNO3 and 20 mm NH4NO3 for 2

hours. Following this, either 10 μM dexamethasone (in 100% EtOH) for 3 hrs; 35 μM cycloheximide (in DMSO) for 20 mins; 10 μM dexamethasone and 35 μM cycloheximide applied sequentially; or no treatment, were added to three aliquots of protoplasts. When a treatment was not given, equivalent volumes of EtOH or DMSO were added. After 3 hours, cells were harvested and RNA was extracted using the Sigma Spectrum total RNA kit and cDNA was synthesized using M-MLV reverse transcriptase (Thermo Fisher). Quantitative reverse-transcription-PCR (qRT-PCR) was performed on a QuantStudio™ 6 Pro Real-Time PCR System (Applied Biosystems A43182) in 384-well plates. Amplifications were done in 10 μL reactions containing 1× SYBR® Green JumpStart™ Taq ReadyMix™ (Sigma S4438), 0.2 μm each primer and 6 ng cDNA template. No template and no reverse-transcriptase controls were also performed. Each amplification was performed in duplicate (technical repeat). A QuantStudio™ 6 Pro 384-well standard, relative quantification with melt program (comparative Ct with melt) was used with the following parameters: 94 °C for 2 minutes then 40 cycles of 94 °C for 15 seconds, 58 °C for 1 minute. A melt curve was performed at the end of the run to confirm the specificity of the amplification. Primer sequences are provided in Supplementary Data 18.

For tomato, TF coding sequences were either amplified from cDNA using Phusion polymerase (Primer sequences are provided in Supplementary Data 18) and cloned into pENTR™ /D-TOPO™ (ThermoFisher), or synthesized (Genewiz, South Plainfield, NJ). Sequence verified clones were used in LR cloning reactions with the Gateway-enabled pBEACON DESTINATION vector (26), which contains a C-terminal GR- tag. The resulting EXPRESSION clones were sequence verified. Details of plasmids are provided in Supplementary Data 17. Protoplast isolation from hairy root cultures (see below for methods of hairy root production) was performed as previously described (75) with the following changes: Two-weeks after sub-cloning, protoplasts were isolated from hairy roots tissues harvested from three petri dishes using 1.5% cellulase R10, 1.5% cellulase RS, 0.4% macerozyme and 0.13% pectolyase enzyme solution. Hairy roots were lightly chopped with a sharp razor blade using ~10 strokes and incubated in the enzyme solution for 4 hours at 28-30℃, in the dark, with shaking at 85 rpm. After incubation, the protoplast solution was filtered with 70 and 40 μm strainers, and centrifuged at 500 g 5 min at 4 ℃ once using 12 ml cold W5 solution in 14 ml round bottom culture tube, and second time using 2 ml W5 in a round bottom microcentrifuge tube. For downstream transfection, hairy root protoplasts are rususpended at a ~4-8 x 10^5 cells/ml concentration in MMG solution. Protoplast transfection, RNA purification and qRT-PCR was carried out as for Arabidopsis.

Luciferase co-expression assays.

Promoter regions of *AtARF18*, *AtANAC032*, *AtDREB26*, *AtNIR1*, *AtNLP6* and *AtNLP7* including ~1000 bp upstream of the TSS and the 5′ UTR, were amplified and cloned into pUAP4 (72) to create Level 0 standard parts comparable with the Phytobrick standard (74). During cloning, internal BsaI sites in *ANAC032*$_P$ (301 bp upstream of TSS) and *ARF18*$_P$ (258 bp upstream of TSS) promoters were removed by the introduction of a mutation. These mutations did not disrupt any candidate TFBSs. These Level 0 promoter parts were used in in one-step digestion-ligation reactions with the Level 1 Loop pCk1 (Addgene #136695) backbones together with parts containing LucN (pEPYC0CM0133, Addgene #154595), a C-terminal 3×FLAG® epitope tag (pICSL50007; Addgene #50308), and 3′ UTR and terminator sequences from *Agrobacterium tumefaciens* nopaline synthase (*AtuNOS*$_T$) (pICH41421, Addgene #50339). As a batch calibrator, an equivalent plasmid was constructed with *AtuNOS*$_P$ (pEPSW1KN0035) and experiment calibrator plasmids (pEPSW1KN0034 and pEPSW1KN0072) for ratiometric quantification were assembled from *CaMV35s*$_P$:tobacco mosaic virus (ΩTMV) (pICH51277, Addgene #50268) + firefly luciferase (LucF) (pEPAS0CM0008, Addgene #154594) + *CaMV35s*$_T$ (pICH41414), or *AtuNOS*$_P$ (pICH42211, Addgene #50255) + firefly luciferase (LucF) (pEPAS0CM0008,

Addgene #154594) + *AtuOCS*$_T$ (pICH41432, Addgene #50343). Finally, the coding sequence of each transcription factor and the YFP coding sequence were cloned into pUAP4 in one-step restriction-ligation reactions to create Level 0 parts (with stop codons) that were subsequently assembled into the level 1 Loop acceptor (pCk2; Addgene #136696) in one-step restriction-ligation reactions with a *CaMV35s*$_P$-ΩTMV (pICH51277; Addgene #50268) and *CaMV35s*$_T$ (pICH41414). Details of plasmids are in Supplementary Data 17.

Arabidopsis mesophyll protoplasts were prepared as described above. For each promoter, six protoplast aliquots were prepared. Three aliquots received three plasmids: *TF*$_P$-LucN-*AtuNOS*$_T$ + *CaMV35s*$_P$-ΩTMV-TF-*CaMV35s*$_T$ + *CaMV35s*$_P$-LucF-*CaMV35s*$_T$ at equimolar ratios totalling 1000 fmol. In the other three aliquots *CaMV35s*$_P$-ΩTMV-TF-CaMV35s$_T$ was substituted for *CaMV35s*$_P$-ΩTMV-YFP-*CaMV35s*$_T$. The entire experiment was repeated three times and, in every experiment, three protoplast aliquots were transfected with a batch calibrator (*CaMV35s*$_P$-LucN-*AtuNOS*$_T$ + *CaMV35s*$_P$-LucF-*CaMV35s*$_T$). Full experimental details are provided in Supplementary Data 5. Protoplasts were transfected as described above and, after 18 hours, collected by centrifugation. To each aliquot, 30 µL lysis buffer protease inhibitor mix (1x passive lysis buffer (Promega E1941) and 1x protease inhibitor cocktail (Sigma P9599)) was added. After 15 minutes on ice, debris was collected by centrifugation and 30 µL of lysate was added to each well of a 4titude 96-well white polystyrene microplate flat bottom plate (Azenta Life Sciences). The NanoGlo® Dual-Luciferase® Reporter Assay System (Promega N1610) was used to quantify NanoLuc and LucF luminescence in a CLARIOstar Plus plate reader with each well measured for ten seconds at 3600 gain with one second settling time.

Tomato transcriptome and gene expression analysis. Seeds were germinated on a nitrogen-free MS medium (Caisson labs, MSP07) supplemented with the appropriate concentration of KNO_3 at 22 °C in constant light. The plates were placed vertically in 22 °C growth chambers with 16-h light/8-h dark cycles and whole root samples were collected at 9 days post-germination. Total RNA was extracted using the Quick-RNA Plant Miniprep Kit (Zymo Research). First-strand cDNAs were synthesized using SuperScript IV Reverse Transcriptase (Thermo Fisher Scientific) with 500 ng of total RNA as a template. Quantitative RT-PCR reactions were prepared using a SYBR Green Master Mix kit (Thermo Fisher Scientific) and conducted in a CFX384 Touch Real-Time PCR Detection System (Bio-Rad). Three technical replicates were performed for each of the three biological replicates. Expression values were normalized to the reference gene *SlEXP (Solyc07g025390*) using the ΔCT method. Data are presented as mean ± SD. Primer sequences are provided in Supplementary Data 18. RNA-seq libraries were prepared using the QuantSeq 3' mRNA-Seq Library Prep Kit (Lexogen) according to the manufacturer's protocol. Libraries were submitted to the UC Davis DNA Technologies Core and sequenced with an Illumina single-end HiSeq 4000 SR100. Four biological replicates and three technical replicates were used for each sample. After sequencing, reads were pooled, trimmed, and filtered using Trim Galore! (v0.4.5) (Krueger, 2012) with the parameter -a GATCGGAAGAGCACA. Trimmed reads were pseudo-aligned to the ITAG4.1 transcriptome (cDNA) using Kallisto (v0.43.1) (Bray et al., 2016), with the parameters -b 100–single -l 200 -s 30. Raw RNA-seq read counts were filtered to remove genes with zero counts across all samples. Reads were converted to count per million (CPM) using the cpm() function in edgeR. Genes that were consistently lowly expressed across all samples were removed. Differential gene expression analysis was done using limma in R/Bioconductor, with empirical weights estimated using the voomWithQualityWeights function. Quantile normalization was used to account for different RNA inputs and library sizes. The linear model for each gene was specified as N treatment: log(counts per million) of an individual gene ~0+N_concentration. Differentially expressed genes were selected based on a false discovery rate <

0.05. Raw sequence data generated during this study have been deposited in the National Center for Biotechnology Information (NCBI) BioProject database under accession number PRJNA975125.

Clustering and gene ontology (GO) enrichment analysis.

Genes differentially expressed between N treatments were clustered in groups with similar expression patterns. The log2 CPM median across biological replicates were calculated for each gene and the expression of each gene was then scaled to the median expression across all samples. Hierarchical clustering was performed with the pheatmap v.1.0.12 R package with the Euclidean distance metric to quantify similarity. Genes assigned to each cluster are listed in Supplementary Data 7. Gene Ontology (GO) enrichment analysis was performed with the GOseq v.1.34.1 R package with correction for gene length bias (Young et al., 2010). The odds ratio for each ontology was calculated with the formula: (number of genes in GO category / number of all genes in input) / (number of genes in GO category / number of genes in all clusters). Enriched ontology terms were selected based on a p-value < 0.05 and an odds ratio > 1. GO categories enriched in each cluster are listed in Supplementary Data 7.

Phylogenetic tree construction.

Phylogenetic trees were generated as previously described (23). Briefly, representative proteomes were downloaded from Phytozome, Ensembl, or consortia sites. Next, BLASTp (76) was used to identify homologous sequences within each proteome. To refine this set of sequences, a multiple sequence alignment was generated with MAFFT v7 (77) and a draft tree was generated with FastTree (78). A monophyletic subtree containing the relevant sequences of interest was then selected. For the final trees, MAFFT v7 using L-INS-i strategy was used to generate a multiple sequence alignment. Next, trimal was used with the -gappyout option. To generate a phylogenetic tree using maximum likelihood, RAxML was used with the option -m PROTGAMMAAUTO and 100 bootstraps. Finally, bipartitions with bootstrap values less than 25% were collapsed using TreeCollapserCL4 (http://emmahodcroft.com/TreeCollapseCL.html). The resulting trees were rooted on sequences from the earliest-diverging species represented in the tree. All trees are provided in Supplementary Data 8.

Assembly of constructs for CRISPR/Cas9 mutagenesis.

All constructs were generated using the Loop assembly toolkit (72) using the methods described in (79). New level 0 DNA parts (e.g. promoters, terminators, CDSs) were either synthesized (Twist Bioscience, San Francisco, CA and Genewiz, South Plainfield, NJ) or amplified by PCR with overhangs containing SapI recognition sites and assembled into pUAP4 (Addgene #136079) to create parts compatible with the Phytobrick standard (74). Targets for mutation were selected using CRISPR-P 2.0 (80). Spacer sequences were synthesized (Integrated DNA Technologies, Coralville, IA or Twist Biosciences, San Francisco, CA) and integrated into single guide RNA (sgRNA) Level 1 expression constructs as described in (79). For Arabidopsis, each final construct contained the FASTred (81) visual non-destructive selectable marker cassette (*AtOLE*$_P$-OLE:RFP-*AtOLE*$_T$) in position 1; a Cas9 expression cassette (*AtYAO*$_P$-SpCas9-E9$_T$) or (*AtYAO*$_P$-SaCas9-*E9*$_T$) in a reverse position 2; a tandem arrays of two to four sgRNA cassettes; and a plant kanamycin resistance cassette (*AtuNOS*$_P$-nptII-*AtuOsc*$_T$). For tomato, each final construct contained the FASTgreen selection cassette (*AtOLE*$_P$-OLE:GFP-*AtOLE*$_T$) in position 1; the SpCas9 expression cassette (*AtRPS5A*$_P$-SpCas9-E9$_T$) in reverse position 2; two sgRNA expression cassettes in tandem in positions 3 and 4; and a plant kanamycin resistance cassette (*AtuNOS*$_P$-nptII-*AtuOsc*$_T$) or basta resistance cassette (*CaMV35s*$_P$-Bar-*CaMV35s*$_T$) in position 5.

Production of Arabidopsis lines with CRISPR/Cas9-induced mutations. Assembled CRISPR/Cas9 plasmids were transformed into *A. tumefaciens* GV3101 and liquid cultures were grown from single colonies in LB supplemented with 50 mg/mL rifampicin, 25 mg/mL gentamycin and 50 mg/mL kanamycin at 28 °C. Cells were collected by centrifugation and resuspended to OD_{600} 0.8 in 5% sucrose, 0.05% Silvet L-77 and sprayed onto the floral tissues of ten Col0 plants. Plants were sealed in black plastic bags for 24 hours. Seeds were collected from mature siliques and surface sterilized with 70% EtOH for 10 min followed by 3-5% sodium hypochlorite for 10 min. For selection of T_1 transgenic plants, sterilized seeds were germinated and grown on MS supplemented with 75 mg/mL kanamycin with 16 h light 22 °C. T_1 seedlings were transferred to soil and T_2 seed was collected. Seed batches were examined with a fluorescence microscope to assess segregation of transgene by expression from the FAST-red marker cassette. Non-fluorescent T_2 seeds from segregating seed batches were selected and germinated. DNA was extracted and PCR and sequencing was performed to confirm absence of the transgene and to identify mutations at the target loci. The genotypes of all lines analyzed in this study are provided in Supplementary Data 11. T_3 progeny of each line were grown to bulk seed and to confirm the stability of the genotype and T_4 seeds were used for subsequent analysis.

Production of *Rhizobium rhizogenes*-induced tomato hairy roots with CRISPR/Cas9 mutations.

Tomato hairy root transformation was performed as previously described (40). Briefly, competent *R. rhizogenes* was transformed by electroporation with each CRISPR/Cas9 construct , plated on nutrient agar (Difco, BD 247940) with the appropriate antibiotics, and incubated for 2-3 days at 28-30 ℃. Single colonies were selected and used to inoculate liquid cultures that were grown overnight at 30 ℃ with shaking. The resulting cultures were used to transform cotyledons. For each construct, 40 fully expanded cotyledons were cut and dipped into liquid *R. rhizogenes* culture and co-cultivated in the dark on MS agar (MS basal media pH 5.8 with 1X vitamins, 3% w/v sucrose, 1% w/v agar) without antibiotic selection for three days. Cotyledons were then transferred to selective MS agar (MS basal media pH 5.8 with 1X vitamins, 3% w/v sucrose, 1% w/v agar, 200 mg/L cefotaxime, 50 mg/L kanamycin). At least 15 independent antibiotic-resistant roots were subcloned for each transformation for genotyping and subsequent analyses. Generally, roots were subcloned on selective MS agar for two rounds before maintenance on MS agar with Cefotaxime (MS basal media pH 5.8 with 1X vitamins, 1% w/v sucrose, 1% w/v agar, 200 mg/L cefotaxime).

For genotyping, samples of root tissues were collected aseptically and genomic DNA was purified. Target regions were amplified and adapters were ligated to produce barcoded pools of amplicon libraries. Libraries were normalized, pooled, and sequenced using Illumina chemistry (MGH CCIB DNA Core Facility, Massachusetts General Hospital, Boston). Sequencing reads were filtered, de-multiplexed, mapped to reference sequences using a customized script, and visualized in an Integrative Genomics Viewer genome browser (https://igv.org) to identify mutations. At least two independent lines, each with a different mutation, were selected from each transformation. The genotypes of all lines are provided in Supplementary Data 11.

Production of stable tomato lines with CRISPR/Cas9-induced mutations.

CRISPR/Cas9 constructs (as noted above) were introduced in *Agrobacterium tumefaciens* and transformed into *Solanum lycopersicum* cultivar M82 (LA3475) at the UC Davis Transformation Facility. For each construct, 12-20 first-generation (T0) seedlings were transferred to potting mix and cultivated in a growth chamber at 22 °C with 16-h light/8-h dark cycles for 14 days before transfer to the greenhouse. Genomic DNA was purified from leaf issues and target regions were amplified and

sequenced as described above. T1 seeds were collected from lines in which mutations at the target were identified and the resulting plants were genotyped to identify lines with homozygous or biallelic mutations at target sites in which the T-DNA had segregated away. The genotypes of all lines are provided in Supplementary Data 11. Plants were self-pollinated and progeny seeds were collected for further analysis.

Root system architecture (RSA) analysis.

Arabidopsis seeds were sieved to obtain seeds between 250 and 300 µm, sterilized with 50% (v/v) bleach and stored at 4°C for 48 hours before germination on sterile growth media supplemented with the appropriate concentration of KNO_3 as previously described (17). Arabidopsis mutants were grown in triplicate with Col-0 controls on the same plate. Nine seeds were placed on each plate. The plates were placed vertically in 22 °C growth chambers with 16-h light/8-h dark cycles for 9 days post germination before the plates were scanned and images processed using SmartRoot (82). Tomato seeds were germinated on a nitrogen-free MS medium (Caisson labs, MSP07) supplemented with the appropriate concentration of KNO_3. Tomato mutant seeds and their M82 control were grown in triplicate separately with 5 seeds placed on each plate. The plates were placed vertically in 22 °C growth chambers with 16-h light/8-h dark cycles for 8-10 d before the roots were scanned at 300 dpi with an Epson Perfection V700 scanner. Images were processed in Fiji ImageJ software. For both species, RSA traits including primary root length, total lateral root number, total lateral root length,average lateral root length, total root length (primary root length + total lateral root length), lateral root density (total number of lateral roots divided by primary root length), and the ratio of total lateral root length to total root length were quantified.

All statistical analyses were performed in R version i386 4.2.2. For RSA assays, two statistical tests were performed: unpaired t-tests between wild-type and mutant seedlings for each treatment combination as well as a two-way ANOVA to test for genotype x treatment interactions. Statistical tests were performed on the natural log of a given trait for each individual. After log transformation, outliers that were more than two standard deviations away from the mean within a genotype-treatment group were removed. To test for wild-type response to nitrogen treatment, a one-way ANOVA was performed using the car package's anova function (anova(lm(log.TraitValue ~ Treatment + Plate))). To test for significant genotype by treatment interactions, a two-way ANOVA was performed (anova(lm(log.TraitValue ~ Genotype*Treatment + Plate))). In the case of the tomato RSA analysis, the factor "Plate" was set to zero because it was highly correlated with the factor "Genotype". . Additional t-tests for each pairwise treatment combination were also performed. To test for genotype-specific responses to nitrogen treatment, the natural log of a trait was modeled using a two-way ANOVA (anova(log.TraitValue ~ Genotype*Treatment + Plate)). For the tomato data, the two-way ANOVA was run as anova(llog.TraitValue ~ Genotype*Treatment). Please see Supplementary Data 19 for all calculated p-values.

Assembly of nitrate-responsive reporter constructs.

A promoter fragment containing 1 kb upstream of the start codon of *AtNIR1*$_P$ was amplified from genomic DNA isolated from Arabidopsis Col-0 seedlings.In addition, 365 bps of a previously described synthetic arabidopsis nitrogen responsive promoter (*AtNRP*$_P$) (41) was synthesized (Twist Bioscience, San Francisco, CA). Both promoter fragments were cloned into pENTR™/5'-TOPO™ (ThermoFisher) and the resulting ENTRY clones were used in a LR Clonase II reaction with the binary DESTINATION vector pMR099, which contains the coding sequence of nuclear-targeted GFP (GFP_{nu}) and a cassette for constitutive expression of plasma membrane-targeted tagRFP (CaMV35s$_P$-TagRFP:LTI6b) (40).

The resulting expression clones encode *AtNIR*$_P$-GFP$_{nu}$ and *AtNRP*$_P$-GFP$_{nu}$. Similarly, *AtNRP*$_P$ was assembled with the ELuc coding sequence to produce a AtNRP$_P$-Eluc-Hsp18-2 expression vector.

Confocal imaging.

All confocal images were captured using a Zeiss Observer Z1 LSM700 (Zeiss) Confocal Laser Scanning Microscopy (water immersion, × 20 objective) with excitation at 488 nm and emission at 493–550 nm for GFP and excitation at 555 nm and emission at 560–800 nm for mRFP. The tomato seeds or hairy roots were germinated/transferred to MS medium with different KNO_3 concentrations for 10 days, and root tips of tomato seedlings or hairy roots were collected for imaging. Cell outlines were visualized using a constitutive mRFP expression cassette in the pMR099 vector. Confocal images were false-colored, and brightness/contrast was adjusted in Fiji ImageJ software.

Protoplast Assay Reporting Overall Effects of TFs (PAROT).

In this assay, vectors encoding a nitrate-responsive green luciferase (AtNRP$_P$-Eluc-Hsp18-2) and constitutively-expressed red luciferase (pNOS-Rluc-tNOS; pGREAT27, Addgene #170915) are delivered to root protoplasts. To obtain Arabidopsis root tissues, seeds were sterilized with 50% (v/v) bleach and vernalized for 3 days before plating on MS media (10g agar, 0.5 g MES, 10 g sucrose, 4.33 g MS salts, pH to 5.7 with KOH). Around 800 seeds were densely placed in two rows on 120 mm square plates with a Nitex mesh (Sefar) to facilitate recovery of roots from plates for protoplasting. The plates were placed vertically in growth chambers at 22°C with 16-h light/8-h dark cycles for 8-10 days before harvesting. Protoplasts were isolated as previously described (83) with modifications. Briefly, 20 mg pectolyase (Sigma-Aldrich P-5936) and 230 mg cellulase RS (Yakult) were dissolved in 15 mL of protoplast isolation buffer (600 mM mannitol, 20 mM MES, 20 mM KCl, 0.1% BSA, 10 mM $CaCl_2$, pH 5.7) and passed through a 0.45 µm filter. Root tissues were harvested from six plates, chopped several times and placed within a 70 µm strainer set within a 6-well petri dish containing 5 ml of the enzyme solution. After 3 hours at room temperature (or 25˚C) on an orbital shaker the solution was passed through a 40 µm cell strainer to remove any remaining tissues. Protoplasts were collected and washed in W5 buffer (2 mM MES, 154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl, pH 5.7) by centrifugation at 4°C 500g for 10 minutes with acceleration and deceleration set to zero. Tomato root protoplasts were prepared as described in the TARSET section.

Protoplasts of both species were counted with a hemocytometer and diluted in MMG buffer (4 mM MES, 0.4 M mannitol, 15 mM $MgCl_2$, pH 5.7) at a concentration of 2-8 x 10^5 cells/ml for transfection with dual luciferase vectors. Fresh 40% PEG transfection solution (4 g of PEG4000 (Fluka, cat. no. 81240) added into 3 mL of H_2O, 2.5 mL of 0.8 M mannitol, and 1 mL of 1 M $CaCl_2$) was prepared at least one hour before transfection to allow enough time to dissolve. In a round bottom microcentrifuge tube, 10 µg of each plasmid was mixed with 100 µL protoplasts and 110 µL PEG solution and incubated for 15 minutes. Protoplasts were again washed with W5 after transfection, and suspended in the WI buffer (4 mM MES, 0.5 M mannitol, 20 mM KCl, pH 5.7) with appropriate KNO_3 concentration (0, 1, or 10 mM) for 24 hours before quantification of luciferase in a plate reader, with each well measured for ten seconds at 3600 gain with one second settling time.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support of the United States Department of Agriculture (USDA), the National Science Foundation (NSF) of the United States of America, and the United Kingdom Research and Innovation's Biotechnology and Biological Sciences Research Council (UKRI BBSRC). This work was funded by a joint USDA/NSF/BBSRC Breakthrough Technologies Award Grant numbers: USDA:

2019-67013-29012, BBSRC: BB/S020853/1). Additional funding was provided by Strategic Programme Grant to the Earlham Institute (BB/CSP1720/1) work package 'Regulatory Interactions and Complex Phenotypes' (BBS/E/T/000PR9819). SW was supported by the BBSRC Norwich Research Park Doctoral Training Partnership (BB/M011216/1 Project No. 2116916).

CONTRIBUTIONS

NJP and SB conceived the study and were responsible for fundraising and supervision. SW, Y-MC, MTO, CB and GAM identified candidate binding sites; SW and Y-MC produced recombinant proteins and performed DNA binding assays; MTO, Y-MC, SW, GSD and CB cloned and produced constructs; MTO, SW and GSD performed transient protoplasts gene expression assays to detect regulatory interactions; CB produced and tested the N-responsive reporters; GSD, MTO and Y-MC conducted N-responsiveness assays in root protoplasts; CB generated and analyzed tomato transcriptome datasets; CB performed phylogenetic analysis of tomato genes; CB and GAM identified tomato homologs of Arabidopsis genes; MTO produced Arabidopsis CRISPR lines and conducted genotype analysis of all CRISPR and insertion lines; CB produced and genotyped tomato hairy root CRISPR lines and stable tomato CRISPR lines; MTO, Y-MC, GAM and RS conducted analyses of Arabidopsis plant root systems; CB conducted analyses of tomato root systems; all authors contributed to data analysis and the production of figures; NJP and SB drafted the manuscript text and all authors contributed to revisions.

CONFLICTS OF INTERESTS

The authors have no conflicts of interests to declare.

LIST OF SUPPLEMENTARY DATA

Supplementary Data 1. Position weight matrix logo plots used to identify candidate binding sites for transcription factors in the regulatory circuit.

Supplementary Data 2. Alignments of the DNA binding domains of Arabidopsis transcription factors.

Supplementary Data 3. *In vitro* transcription-factor DNA binding assays.

Supplementary Data 4. Sequences of DNA probes used in TRAM assays.

Supplementary Data 5. Schematic showing the methodology for protoplast co-expression assays.

Supplementary Data 6. Changes in endogenous gene expression of Arabidopsis target genes using TARSET assays.

Supplementary Data 7. Nitrate responsive genes in tomato roots.

Supplementary Data 8. Phylogenetic analysis of transcription factors in this study.

Supplementary Data 9. Domain analysis of tomato orthologues of AtNLP7.

Supplementary Data 10. Table of Arabidopsis and Tomato mutant lines .

Supplementary Data 11. Genotypes of CRISPR knock-out lines.

Supplementary Data 10. Alignments of the DNA binding domains of tomato and arabidopsis transcription factors.

Supplementary Data 11. Changes in endogenous gene expression levels of tomato target genes using TARSET assays.

Supplementary Data 12. Analysis of root system architecture of Arabidopsis and tomato mutant alleles of NLP7 and ARF18.

Supplementary Data 13. Alignments of the DNA binding domains of tomato and arabidopsis transcription factors.

Supplementary Data 14. Changes in endogenous gene expression levels of target genes (TARSET assays).

Supplementary Data 15. Validation of NRP:reporter assay.

Supplementary Data 16. NRP:reporter (PAROT) assay in Arabidopsis and tomato mutants.

Supplementary Data 17. Plasmids used in this study.

Supplementary Data 18. Primers used in this study.

Supplementary Data 19. Statistics for RSA analysis.

REFERENCES

1. E. Sinha, A. M. Michalak, V. Balaji, Eutrophication will increase during the 21st century as a result of precipitation changes. *Science* **357**, 405–408 (2017).
2. S. Savci, Investigation of Effect of Chemical Fertilizers on Environment. *APCBEE Procedia* **1**, 287–292 (2012).
3. M. Stitt, Nitrate regulation of metabolism and growth. *Curr. Opin. Plant Biol.* **2**, 178–186 (1999).
4. H. Zhang, B. G. Forde, Regulation of Arabidopsis root development by nitrate availability. *J. Exp. Bot.* **51**, 51–59 (2000).
5. E. A. Vidal, T. C. Moyano, J. Canales, R. A. Gutiérrez, Nitrogen control of developmental phase transitions in Arabidopsis thaliana. *J. Exp. Bot.* **65**, 5611–5618 (2014).
6. R. Wang, M. Okamoto, X. Xing, N. M. Crawford, Microarray analysis of the nitrate response in Arabidopsis roots and shoots reveals over 1,000 rapidly responding genes and new linkages to glucose, trehalose-6-phosphate, iron, and sulfate metabolism. *Plant Physiol.* **132**, 556–567 (2003).
7. C. Marchive, *et al.*, Nuclear retention of the transcription factor NLP7 orchestrates the early response to nitrate in plants. *Nat. Commun.* **4**, 1713 (2013).
8. K.-H. Liu, *et al.*, Discovery of nitrate-CPK-NLP signalling in central nutrient-growth networks. *Nature* **545**, 311–316 (2017).
9. G. Krouk, P. Mirowski, Y. LeCun, D. E. Shasha, G. M. Coruzzi, Predictive network modeling of the high-resolution dynamic plant transcriptome in response to nitrate. *Genome Biol.* **11**, R123 (2010).
10. O. Contreras-López, *et al.*, Spatiotemporal analysis identifies ABF2 and ABF3 as key hubs of endodermal response to nitrate. *Proc. Natl. Acad. Sci. U. S. A.* **119** (2022).
11. P. Guan, *et al.*, Nitrate foraging by Arabidopsis roots is mediated by the transcription factor TCP20 through the systemic signaling pathway. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15267–15272 (2014).
12. H. Zhang, B. G. Forde, An Arabidopsis MADS box gene that controls nutrient-induced changes in root architecture. *Science* **279**, 407–409 (1998).
13. M. L. Gifford, A. Dean, R. A. Gutierrez, G. M. Coruzzi, K. D. Birnbaum, Cell-specific nitrogen responses mediate developmental plasticity. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 803–808 (2008).
14. L. Castaings, *et al.*, The nodule inception-like protein 7 modulates nitrate sensing and metabolism in Arabidopsis. *Plant J.* **57**, 426–435 (2009).
15. J. M. Alvarez, *et al.*, Systems approach identifies TGA1 and TGA4 transcription factors as important regulatory components of the nitrate response of Arabidopsis thaliana roots. *Plant J.* **80**, 1–13 (2014).
16. A. Medici, *et al.*, AtNIGT1/HRS1 integrates nitrate and phosphate signals at the Arabidopsis root tip. *Nat. Commun.* **6**, 6274 (2015).
17. A. Gaudinier, *et al.*, Transcriptional regulation of nitrogen-associated metabolism and growth. *Nature* **563**, 259–264 (2018).
18. K.-H. Liu, *et al.*, NIN-like protein 7 transcription factor is a plant nitrate sensor. *Science* **377**, 1419–1425 (2022).
19. H. Cao, *et al.*, Overexpression of the Maize ZmNLP6 and ZmNLP8 Can Complement the Arabidopsis Nitrate Regulatory Mutant nlp7 by Restoring Nitrate Signaling and Assimilation. *Front. Plant Sci.* **8**, 1703 (2017).
20. J. Wu, *et al.*, Rice NIN-LIKE PROTEIN 4 plays a pivotal role in nitrogen use efficiency. *Plant Biotechnol. J.* **19**, 448–461 (2021).
21. M. Liu, X. Zhi, Y. Wang, Y. Wang, Genome-wide survey and expression analysis of NIN-like Protein (NLP) genes reveals its potential roles in the response to nitrate signaling in tomato. *BMC Plant Biol.* **21**, 347 (2021).
22. J. Machado, *et al.*, Young Tomato Plants Respond Differently under Single or Combined Mild Nitrogen and Water Deficit: An Insight into Morphophysiological Responses and Primary Metabolism. *Plants* **12** (2023).
23. K. Kajala, *et al.*, Innovation, conservation, and repurposing of gene function in root cell type development. *Cell* (2021) https:/doi.org/10.1016/j.cell.2021.04.024.
24. M. Ron, *et al.*, Identification of novel loci regulating interspecific variation in root morphology and cellular development in tomato. *Plant Physiol.* **162**, 755–768 (2013).
25. Y.-M. Cai, S. Witham, N. J. Patron, Tuning plant promoters using a simple split luciferase method for quantifying

transcription factor binding affinity. *bioRxiv*, 2023.02.13.528283 (2023).

26. B. O. R. Bargmann, *et al.*, TARGET: a transient transformation system for genome-wide transcription factor target discovery. *Mol. Plant* **6**, 978–980 (2013).
27. J. S. Reece-Hoyes, *et al.*, Enhanced yeast one-hybrid assays for high-throughput gene-centered regulatory network mapping. *Nat. Methods* **8**, 1059–1064 (2011).
28. K. C. Potter, J. Wang, G. E. Schaller, J. J. Kieber, Cytokinin modulates context-dependent chromatin accessibility through the type-B response regulators. *Nat Plants* **4**, 1102–1111 (2018).
29. T. L. Bailey, J. Johnson, C. E. Grant, W. S. Noble, The MEME Suite. *Nucleic Acids Res.* **43**, W39–49 (2015).
30. R. C. O'Malley, *et al.*, Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell* **165**, 1280–1292 (2016).
31. M. Konishi, S. Yanagisawa, Identification of a nitrate-responsive *cis*-element in the Arabidopsis NIR1 promoter defines the presence of multiple *cis*-regulatory elements for nitrogen response. *Plant J.* **63**, 269–282 (2010).
32. J. M. Alvarez, *et al.*, Transient genome-wide interactions of the master transcription factor NLP7 initiate a rapid nitrogen-response cascade. *Nat. Commun.* **11**, 1157 (2020).
33. M. D. Brooks, *et al.*, Network Walking charts transcriptional dynamics of nitrogen signaling by integrating validated and predicted genome-wide interactions. *Nat. Commun.* **10**, 1569 (2019).
34. S. A. Wanamaker, *et al.*, CrY2H-seq: a massively multiplexed assay for deep-coverage interactome mapping. *Nat. Methods* **14**, 819–825 (2017).
35. M. Errebhi, G. E. Wilcox, Tomato growth and nutrient uptake pattern as influenced by nitrogen form ratio. *J. Plant Nutr.* **13**, 1031–1043 (1990).
36. T. Araya, *et al.*, CLE-CLAVATA1 peptide-receptor signaling module regulates the expansion of plant root systems in a nitrogen-dependent manner. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 2029–2034 (2014).
37. Tomato Genome Consortium, The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**, 635–641 (2012).
38. M. Zouine, *et al.*, Characterization of the tomato ARF gene family uncovers a multi-levels post-transcriptional regulation including alternative splicing. *PLoS One* **9**, e84203 (2014).
39. M. T. Weirauch, *et al.*, Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431–1443 (2014).
40. M. Ron, *et al.*, Hairy root transformation using Agrobacterium rhizogenes as a tool for exploring cell type-specific gene expression and function using tomato as a model. *Plant Physiol.* **166**, 455–469 (2014).
41. R. Wang, *et al.*, Multiple regulatory elements in the Arabidopsis NIA1 promoter act synergistically to form a nitrate enhancer. *Plant Physiol.* **154**, 423–432 (2010).
42. M. Konishi, T. Okitsu, S. Yanagisawa, Nitrate-responsive NIN-like protein transcription factors perform unique and redundant roles in Arabidopsis. *J. Exp. Bot.* **72**, 5735–5750 (2021).
43. Y.-H. Cheng, *et al.*, Interplay Between NIN-LIKE PROTEINs 6 and 7 in Nitrate Signaling. *Plant Physiol.* (2023) https:/doi.org/10.1093/plphys/kiad242.
44. C.-Y. Cheng, *et al.*, Evolutionarily informed machine learning enhances the power of predictive gene-to-phenotype relationships. *Nat. Commun.* **12**, 5627 (2021).
45. M. D. Brooks, *et al.*, ConnecTF: A platform to integrate transcription factor-gene interactions and validate regulatory networks. *Plant Physiol.* **185**, 49–66 (2021).
46. S. Heerah, M. Katari, R. Penjor, G. Coruzzi, A. Marshall-Colon, WRKY1 Mediates Transcriptional Regulation of Light and Nitrogen Signaling Pathways. *Plant Physiol.* **181**, 1371–1388 (2019).
47. K. Varala, *et al.*, Temporal transcriptional logic of dynamic regulatory networks underlying nitrogen signaling and use in plants. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 6494–6499 (2018).
48. J. Doidy, *et al.*, "Hit-and-Run" transcription: de novo transcription initiated by a transient bZIP1 "hit" persists after the "run." *BMC Genomics* **17**, 92 (2016).
49. A. Para, *et al.*, Hit-and-run transcriptional control by bZIP1 mediates rapid nutrient signaling in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 10371–10376 (2014).
50. P. J. Dickinson, *et al.*, A bipartite transcription factor module controlling expression in the bundle sheath of Arabidopsis thaliana. *Nat Plants* **6**, 1468–1479 (2020).

51. D. Rodríguez-Leal, Z. H. Lemmon, J. Man, M. E. Bartlett, Z. B. Lippman, Engineering Quantitative Trait Variation for Crop Improvement by Genome Editing. *Cell* **171**, 470–480.e8 (2017).

52. L. Liu, *et al.*, Enhancing grain-yield-related traits by CRISPR–Cas9 promoter editing of maize CLE genes. *Nature Plants* **7**, 287–294 (2021).

53. X. Sheng, H. Yu, J. Wang, Y. Shen, H. Gu, Establishment of a stable, effective and universal genetic transformation technique in the diverse species of Brassica oleracea. *Front. Plant Sci.* **13**, 1021669 (2022).

54. L. Cao, *et al.*, Genome-wide identification of NF-Y gene family in maize (Zea mays L.) and the positive role of ZmNF-YC12 in drought resistance and recovery ability. *Front. Plant Sci.* **14**, 1159955 (2023).

55. R. Wang, *et al.*, *Cis*-regulation of the amino acid transporter genes ZmAAP2 and ZmLHT1 by ZmPHR1 transcription factors in maize ear under phosphate limitation. *J. Exp. Bot.* **72**, 3846–3863 (2021).

56. M. Ge, *et al.*, The NIN-like protein 5 (ZmNLP5) transcription factor is involved in modulating the nitrogen response in maize. *Plant J.* **102**, 353–368 (2020).

57. X. He, *et al.*, The Nitrate-Inducible NAC Transcription Factor TaNAC2-5A Controls Nitrate Response and Increases Wheat Yield. *Plant Physiol.* **169**, 1991–2005 (2015).

58. C. Li, *et al.*, Genome-wide characterization of *cis*-acting DNA targets reveals the transcriptional regulatory framework of opaque2 in maize. *Plant Cell* **27**, 532–545 (2015).

59. M. Obertello, S. Shrivastava, M. S. Katari, G. M. Coruzzi, Cross-Species Network Analysis Uncovers Conserved Nitrogen-Regulated Network Modules in Rice. *Plant Physiol.* **168**, 1830–1843 (2015).

60. M. J. Asins, *et al.*, Genetic analysis of rootstock-mediated nitrogen (N) uptake and root-to-shoot signalling at contrasting N availabilities in tomato. *Plant Sci.* **263**, 94–106 (2017).

61. F. S. Chapin 3rd, C. H. Walter, D. T. Clarkson, Growth response of barley and tomato to nitrogen stress and its control by abscisic acid, water relations and photosynthesis. *Planta* **173**, 352–366 (1988).

62. T. Ulmasov, J. Murfett, G. Hagen, T. J. Guilfoyle, Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements. *Plant Cell* **9**, 1963–1971 (1997).

63. G. Brunoud, *et al.*, A novel sensor to map auxin response and distribution at high spatio-temporal resolution. *Nature* **482**, 103–106 (2012).

64. E. Zürcher, *et al.*, A robust and sensitive synthetic sensor to monitor the transcriptional output of the cytokinin signaling network in planta. *Plant Physiol.* **161**, 1066–1075 (2013).

65. A. Para, Y. Li, G. M. Coruzzi, µChIP-Seq for Genome-Wide Mapping of In Vivo TF-DNA Interactions in Arabidopsis Root Protoplasts. *Methods Mol. Biol.* **1761**, 249–261 (2018).

66. P. Lamesch, *et al.*, The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* **40**, D1202–10 (2012).

67. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

68. C. E. Grant, T. L. Bailey, W. S. Noble, FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).

69. A. Freire-Rios, *et al.*, Architecture of DNA elements mediating ARF transcription factor binding and auxin-responsive gene expression in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 24557–24566 (2020).

70. C. Cancé, R. Martin-Arevalillo, K. Boubekeur, R. Dumas, Auxin response factors are keys to the many auxin doors. *New Phytol.* **235**, 402–419 (2022).

71. B. E. Granger, F. Pérez, Jupyter: Thinking and Storytelling With Code and Data. *Comput. Sci. Eng.* **23**, 7–14 (2021).

72. B. Pollak, *et al.*, Loop assembly: a simple and open system for recursive fabrication of DNA circuits. *New Phytol.* **222**, 628–640 (2019).

73. Y.-M. Cai, *et al.*, Rational design of minimal synthetic promoters for plants. *Nucleic Acids Res.* **48**, 11845–11856 (2020).

74. N. J. Patron, *et al.*, Standards for plant synthetic biology: a common syntax for exchange of DNA parts. *New Phytol.* **208**, 13–19 (2015).

75. S.-D. Yoo, Y.-H. Cho, J. Sheen, Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat. Protoc.* **2**, 1565–1572 (2007).

76. T. Madden, The BLAST sequence analysis tool. *The NCBI handbook* (2003).

77. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

78. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).

79. Q. M. Dudley, O. Raitskin, N. J. Patron, Cas9-Mediated Targeted Mutagenesis in Plants. *Methods Mol. Biol.* **2379**, 1–26 (2022).

80. H. Liu, *et al.*, CRISPR-P 2.0: An Improved CRISPR-Cas9 Tool for Genome Editing in Plants. *Mol. Plant* **10**, 530–532 (2017).

81. T. L. Shimada, T. Shimada, I. Hara-Nishimura, A rapid and non-destructive screenable marker, FAST, for identifying transformed seeds of Arabidopsis thaliana. *Plant J.* **61**, 519–528 (2010).

82. G. Lobet, L. Pagès, X. Draye, A novel image-analysis toolbox enabling quantitative analysis of root system architecture. *Plant Physiol.* **157**, 29–39 (2011).

83. K. Birnbaum, *et al.*, Cell type-specific expression profiling in plants via cell sorting of protoplasts from fluorescent reporter lines. *Nat. Methods* **2**, 615–619 (2005).