BMC Genomics

RESEARCH ARTICLE   **Open Access**

Integrated RNA-seq and sRNA-seq analysis identifies novel nitrate-responsive genes in *Arabidopsis thaliana* roots

Elena A Vidal[1†], Tomás C Moyano[1†], Gabriel Krouk[2,3†], Manpreet S Katari[3†], Milos Tanurdzic[4,5†], W Richard McCombie[4], Gloria M Coruzzi[3*] and Rodrigo A Gutiérrez[1*]

Abstract

Background: Nitrate and other nitrogen metabolites can act as signals that regulate global gene expression in plants. Adaptive changes in plant morphology and physiology triggered by changes in nitrate availability are partly explained by these changes in gene expression. Despite several genome-wide efforts to identify nitrate-regulated genes, no comprehensive study of the Arabidopsis root transcriptome under contrasting nitrate conditions has been carried out.

Results: In this work, we employed the Illumina high throughput sequencing technology to perform an integrated analysis of the poly-A + enriched and the small RNA fractions of the *Arabidopsis thaliana* root transcriptome in response to nitrate treatments. Our sequencing strategy identified new nitrate-regulated genes including 40 genes not represented in the ATH1 Affymetrix GeneChip, a novel nitrate-responsive antisense transcript and a new nitrate responsive miRNA/TARGET module consisting of a novel microRNA, miR5640 and its target, AtPPC3.

Conclusions: Sequencing of small RNAs and mRNAs uncovered new genes, and enabled us to develop new hypotheses for nitrate regulation and coordination of carbon and nitrogen metabolism.

Keywords: Arabidopsis, Nitrate, RNA-seq, Roots, MicroRNA, Transcriptomics

Background

Nitrogen (N) is an essential macronutrient and a key factor controlling plant growth and development. Nitrate is the main form of N available in agricultural soils [1-3]. Nitrate is taken up by the cell by specific nitrate transporters and is reduced to nitrite in the cytoplasm by nitrate reductase. Nitrite is reduced to ammonium in the plastid by nitrite reductase and is incorporated into amino acids by the glutamate synthase/glutamine synthetase cycle (GS/GOGAT cycle). Nitrate metabolism is tightly coordinated with carbon metabolism, since carbon skeletons in the form of 2-oxoglutarate are required for ammonium assimilation [1,4]. One of the most striking examples of plant plasticity in response to changing environmental conditions is root

system architecture modulation by changes in nitrate availability (for reviews see [5-7]). In order to identify molecular mechanisms underlying these changes, transcriptomics analyses of the nitrate response of Arabidopsis have been performed, most of them utilizing the Affymetrix ATH1 GeneChip. Analyses with the ATH1 chip showed that nitrate is able to regulate more than 2,000 genes in roots, some of them responding as fast as 3–6 minutes after nitrate exposure [8] and including genes involved in nitrate transport, reduction and assimilation, hormone signaling pathways, transcription factors, kinases and phosphatases, among others [8-12]. However, a detailed view of the transcriptomics changes triggered by nitrate has been limited by the representation of genes in the ATH1 microarray. ATH1 contains probe sets representing approximately 21,000 genes allowing for the detection of only 71% of the genes annotated in the Arabidopsis genome v.10. Moreover, these probes do not include important regulatory elements of the genome such as small (sRNAs).

High-throughput sequencing technologies allow for quantitative determination of RNA levels and RNA

* Correspondence: gloria.coruzzi@nyu.edu; rgutierrez@bio.puc.cl
†Equal contributors
³Center for Genomics and Systems Biology, Department of Biology, New York University, New York, NY 10003, USA
¹FONDAP Center for Genome Regulation. Millennium Nucleus Center for Plant Functional Genomics, Departamento de Genética Molecular y Microbiología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago 8331010, Chile
Full list of author information is available at the end of the article

sequencing (RNA-seq) is becoming the technology of choice to investigate the transcriptome. RNA-seq offers several advantages over hybridization-based techniques like microarrays [13-18]. RNA-seq is not limited to detection of transcripts that correspond to annotated genes, thus it allows for identification of new genes. RNA sequencing can also be utilized to analyze the sRNA component of the transcriptome when libraries are prepared from low-molecular weight RNA fractions [19-24]. microRNAs (miRNAs), short interfering RNAs (siRNAs) and other types of sRNAs have been shown to play important roles in a broad range of biological processes, such as plant development and response to biotic and abiotic stresses [25-29], including plant responses to various nutrients [30-37].

In plants, the sRNA transcriptome is primarily composed of 23–24 nt siRNAs and 21–22 nt miRNAs [36,38,39]. Since miRNA precursors have distinctive secondary structures, many bioinformatics programs have been developed to predict new miRNAs based on sequencing of a sRNA in a library and inspection of the genome sequence containing this sequence for putative miRNA precursors [40-42]. Combination of deep sequencing approaches and bioinformatics predictions have identified 19,724 miRNAs related sequences across different phyla out of which 266 correspond to Arabidopsis miRNAs in miRBase v.17 [43].

miRNA regulation of nitrate-responsive genes has been shown to be a key mechanism of plant responses coordinating nitrate availability and root developmental responses. miR167 is down-regulated by nitrate treatments in pericycle cells and this leads to an induction of its target, the auxin response factor ARF8 [44]. Regulation of ARF8 by miR167 causes a change in the ratio of initiating and emerging lateral roots in response to nitrate [44]. Another nitrate regulatory module, consisting of miR393 and the AFB3 auxin receptor has been shown to control root system architecture in response to external and internal nitrate availability [37]. Microarray analysis suggests that other miRNAs can be involved in root responses to nitrate, since several miRNA targets are regulated by nitrate [45].

In this paper, we used Illumina sequencing technology to characterize the poly-A + and sRNA component of nitrate- and control-treated Arabidopsis roots to identify new nitrate-responsive genes. Using bioinformatics analysis of our libraries and miRNA prediction algorithms we were able to find new root expressed genes including new mRNAs and miRNAs. We discovered a new miRNA/target module that might act as an integrator of N and carbon metabolism in Arabidopsis roots.
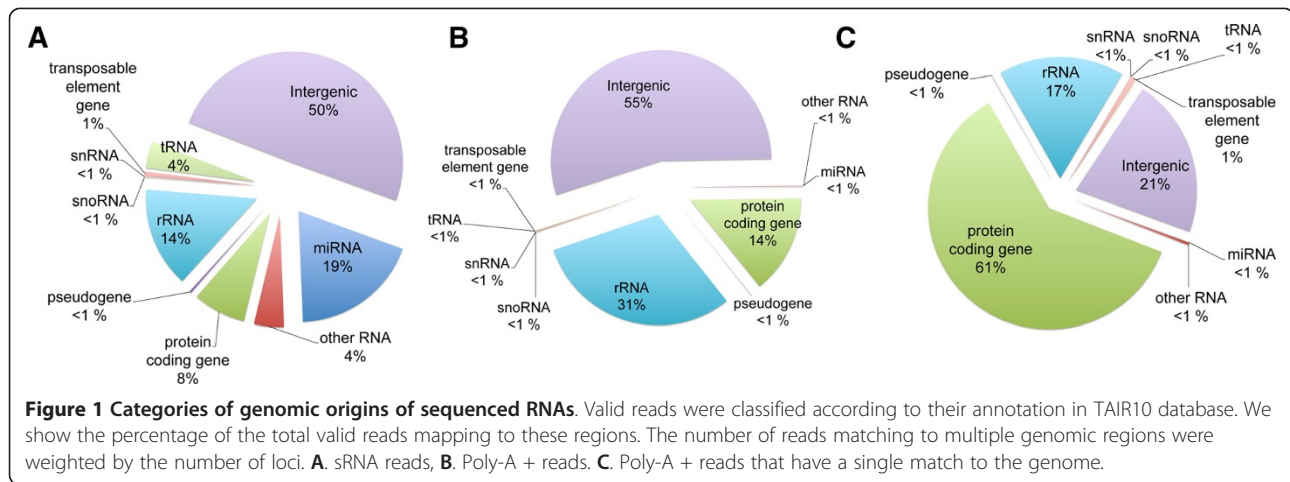
Results

Deep sequence analysis of the root transcriptome

In order to determine poly-A + and sRNA expression of Arabidopsis roots and their changes in response to nitrate, we grew plants in hydroponic nitrate-free medium with 0.5 mM ammonium succinate as the only N-source for two weeks and treated them with 5 mM KNO_3, or 5 mM KCl as control, for 2 hours. These experimental conditions have been previously shown to elicit robust gene expression responses to nitrate [10,44,45]. Total RNA from two independent sets of plants (biological replicates) was extracted from roots, and poly-A + enriched and sRNA fractions were used to construct libraries for Illumina sequencing (see Methods for details). The sequencing yielded ~5 to 8 million 35 bp long (sRNA libraries) or 50 bp long (poly-A + libraries) raw reads per sample library. After quality control filtering and trimming adaptor sequences (see Methods), the reads were mapped to the *Arabidopsis thaliana* genome using the Arabidopsis genome annotation available at The Arabidopsis Information Resource (TAIR) v.10 (www.arabidopsis.org). Approximately two thirds of the total Illumina reads perfectly matched the genome and were used for further analysis (Additional file 1).

Analysis of the size distribution of sequences in the sRNA libraries showed that 21 nt long RNA molecules were the most abundant followed by 24 nt long sequences (Additional file 2). The pattern of sRNA sizes reflects a typical population of sRNAs with abundant miRNAs and tasiRNA (21–22 nt) and siRNAs (23–24 nt) (Additional file 2). However, we did not find accumulation of tRNA fragments as described in roots of phosphate-starved plants [38] or nitrate-starved seedlings [33]. We did not observe any obvious effect of nitrate provision on RNA size distribution (Additional file 2), suggesting that nitrate treatments under our experimental conditions do not have a global effect on sRNA population structure. Next, valid sequences were classified according to the genomic regions they match. Most sRNA sequences matched intergenic regions (8,415,076 sequences, 50%), followed by miRNA (3,189,443 sequences, 19%) and rRNA genes (2,469,734 sequences, 14% of the total valid reads) (Figure 1A). We were able to detect 142 distinct mature Arabidopsis miRNA sequences, corresponding to 98 different miRNA families, according to the miRBase database v.17 (www.mirbase.org) (Additional file 3). The number of miRNA sequences identified represents 66.7% of the 212 miRNAs reported in miRBase v.17, indicating that a considerable proportion of known miRNAs are expressed in the root organ. This number greatly exceeds the previously reported number of miRNAs expressed in roots, that indicated expressed miRNAs are less than 40% of the annotated total miRNAs [36,46]. We were also able to identify sequences corresponding to trans-acting siRNAs (ta-siRNA), including ta-siRNAs arising from the TAS1, TAS2 and TAS3 genes (Additional file 3). It has recently been shown that a significant number of miRNAs have specific root developmental zone or root cell type

Figure 1 Categories of genomic origins of sequenced RNAs. Valid reads were classified according to their annotation in TAIR10 database. We show the percentage of the total valid reads mapping to these regions. The number of reads matching to multiple genomic regions were weighted by the number of loci. **A**. sRNA reads, **B**. Poly-A + reads. **C**. Poly-A + reads that have a single match to the genome.

expression profiles [47]. Most root miRNAs showed low expression levels under our experimental conditions (Additional file 3), suggesting developmental control or expression in specific cell-types of the Arabidopsis root.

For Illumina libraries made from poly-A + RNA, a considerable amount of sequences map to intergenic regions (9,542,618 sequences, 55% of the reads) (Figure 1B). Inspection of sequences matching intergenic regions showed that most of them arise from telomeric or centromeric regions. Transcription from intergenic zones has been reported in previous high-throughput sequencing and tiling array experiments [48-51]. When we considered sequences with a unique match to the genome, only 732,226 sequences (22%) mapped to intergenic regions (Figure 1C). A high proportion of these sequences is supported by Arabidopsis ESTs or cDNAs (710,814 sequences, 97%) obtained from TAIR.

As shown in Additional file 4, most of these sequences are located near the 5' or 3' of annotated genes. We found sequences matching intergenic regions from poly-A + enriched libraries matching the same strand as annotated genes (Additional file 4 A,C). Interestingly, we also found sequences near annotated genes in antisense orientation (Additional file 4 B,D). These could represent novel transcripts that could have a role in controlling the expression of corresponding genes.

Reads matching protein coding genes (2,094,509 sequences) represent ~60% of the unique reads in poly-A + libraries (Figure 1C). The number of expressed protein coding genes detected unambiguously (19,979 protein coding genes) represents 73% of the total annotated in the Arabidopsis genome. Similar to sRNAs, a considerable proportion of genes are expressed in a cell-specific manner [52,53], thus some of the low-expressed transcripts detected under our experimental conditions might be developmentally controlled and/or expressed in specific cell-types of the root.

To date, most transcriptomics studies on the root nitrate response have been performed using the Affymetrix

ATH1 GeneChip [8-11,44,45,54]. In order to determine how our sequencing data compares with data obtained with the Affymetrix ATH1 GeneChip, we used the same RNA samples for Illumina library preparation and ATH1 microarray hybridization. We used the *affy* package library from Bioconductor (www.bioconductor.org) to determine the number of present calls in the ATH1 microarrays as a measure of gene detection. We were able to find 13,964 probes with a present call, approximately 67% of the gene specific probes that are present in the ATH1 microarray (Additional file 5). The Illumina sequencing data detected 13,411 of these genes (96%, at least one read matching the gene) and 3,022 annotated elements that were called absent in the ATH1 array. We found that these 3,022 elements had low expression values when compared with the 13,411 Illumina-detected elements that had present calls in Affymetrix (Additional file 6A,B). Additionally, Illumina was able to detect 4,215 elements that had no probe on the ATH1 microarray (Additional file 5).

In order to determine how data on nitrate-responsive genes obtained with RNA-seq and Affymetrix ATH1 chips correlated, we calculated the correlation between the KNO_3/KCl ratio for RMA normalized Affymetrix gene expression and the KNO_3/KCl ratio obtained for normalized libraries at different average gene coverages (AGCs). We defined AGC as the number of reads matching a gene multiplied by read length and divided by gene length. We found correlation between KNO_3/KCl ratios increase hyperbolically as average gene coverage increases (Additional file 7). This indicates correlation between the two techniques depends on gene expression levels. We found excellent correlation ($r^2 \geq 0.9$) between RNA-seq and ATH1 arrays when gene coverage was 0.8 or higher (reads matching the gene represent 80% or more of the gene length) (Additional file 7). These results highlight the potential of the sequencing strategy to identify novel nitrate-responsive genes in Arabidopsis roots.

Deep sequencing reveals a new nitrate-responsive component of the arabidopsis root transcriptome

In order to identify known miRNAs that are N-regulated under our experimental N-treatment conditions, we used the DESeq package in R to analyze digital gene expression in the RNA-sequencing data [55]. Replicates were used independently for statistical analysis of gene expression. Surprisingly, we were not able to identify known miRNAs (reported in miRBase v17) regulated by nitrate in roots based on our RNA-sequencing data. In order to distinguish between a technical and biological explanation for this result, we calculated an RNA rarefaction curve considering increasing number of random sequences from our sRNA libraries and the number of different sequences that could be determined from each sample (Figure 2A). We found that even when considering the total number of sequences available in our experiments, almost 17 million reads, we were far from saturation. This analysis indicates that most molecules in our sRNA libraries were sequenced only a few times, making it difficult to obtain accurate quantitative results. Using the same RNA samples and quantitative real time PCR, we were able to corroborate induction of miR393 (Figure 2B), a miRNA previously identified as nitrate responsive [37]. This result indicates that a significantly higher depth of sequencing than the current standards [33,34,36,38] is required for quantitative comparison of the sRNA fraction of the Arabidopsis transcriptome. In contrast, when quantifying mRNAs a considerably lower number of sequences is required to reach saturation (Figure 2C).

Among the poly-A + sequences, we found 505 regulated genes, considering only sense sequences that have a unique match to known genes. From these genes, 392 were induced and 113 were repressed by the nitrate treatment (Additional file 8). Regulated genes had an overrepresentation of genes belonging to the "nitrate response", "nitrate transport", "nitrate metabolic process", "nitrate assimilation", "nitrogen cycle metabolic process" and "cellular nitrogen compound biosynthetic process" Gene Ontology annotation, indicating that RNA-seq was successful for identifying nitrate responsive genes. Among these nitrate-regulated genes, we found 40 protein coding genes without probes in the ATH1 GeneChip (Table 1). We selected eight genes and validated them using real time quantitative reverse transcription polymerase chain reaction (RT-qPCR) (Additional file 9). Among the new nitrate-responsive genes, we found transcription factors and components of signaling cascades such as a γ subunit of the heterotrimeric G protein, AGG2. These genes may represent novel targets in the nitrate regulatory pathways in plants.
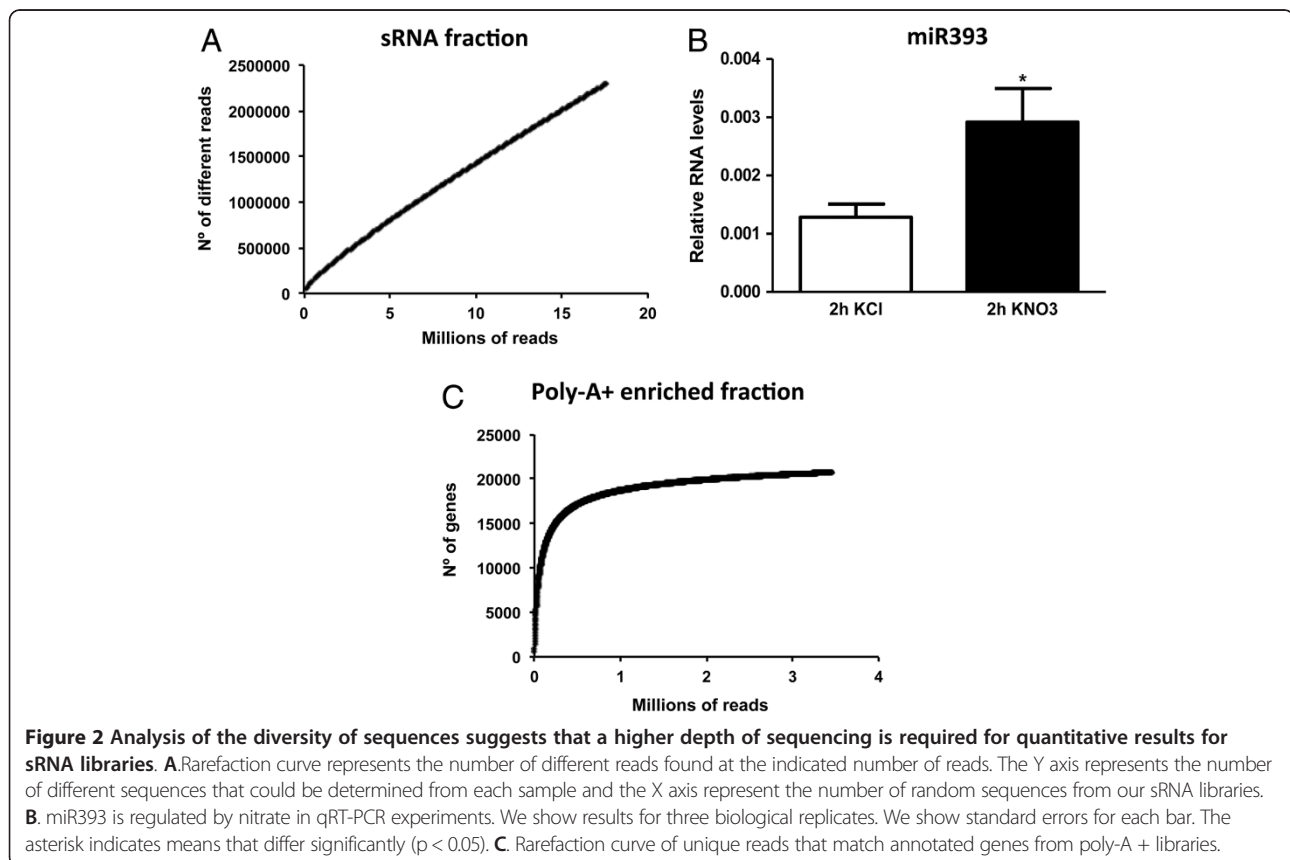

Figure 2 Analysis of the diversity of sequences suggests that a higher depth of sequencing is required for quantitative results for sRNA libraries. **A**.Rarefaction curve represents the number of different reads found at the indicated number of reads. The Y axis represents the number of different sequences that could be determined from each sample and the X axis represent the number of random sequences from our sRNA libraries. **B**. miR393 is regulated by nitrate in qRT-PCR experiments. We show results for three biological replicates. We show standard errors for each bar. The asterisk indicates means that differ significantly (p < 0.05). **C**. Rarefaction curve of unique reads that match annotated genes from poly-A + libraries.

Table 1 Illumina sequencing of poly-A + RNA enriched fraction identifies new nitrate responsive genes

AGI identifier	Description	$\log_2$ (KNO$_3$/KCl)
AT5G63160	BT1, BTB and TAZ domain protein 1	5.2
AT1G11655	Unknown protein	4.3
AT5G65030	Unknown protein	4.2
AT1G70260	nodulin MtN21 /EamA-like transporter family protein	4.0
AT2G33550	Homeodomain-like superfamily protein	3.7
AT1G68238	Unknown protein	3.7
AT4G34419	Unknown protein	3.6
AT1G02030	C2H2-like zinc finger protein	3.1
AT4G34800	SAUR-like auxin-responsive protein family	3.0
AT5G03330	Cysteine proteinases superfamily protein	2.9
AT1G60050	Nodulin MtN21 /EamA-like transporter family protein	2.9
AT2G45760	BAL, BAP2, BON association protein 2	2.8
AT1G70800	Calcium-dependent lipid-binding (CaLB domain) family protein	2.8
AT4G29905	Unknown protein	2.7
AT3G22942	AGG2, G-protein gamma subunit 2	2.2
AT1G23149	CPuORF29, conserved peptide upstream open reading frame 29	2.2
AT1G23150	Unknown protein	2.2
AT5G65980	Auxin efflux carrier family protein	2.2
AT3G14260	Protein of Unknown function (DUF567)	2.1
AT3G48180	Unknown protein	2.1
AT3G25717	DVL6, RTFL16, ROTUNDIFOLIA like 16	2.1
AT2G41440	Unknown protein	2.1
AT1G13245	DVL4, RTFL17, ROTUNDIFOLIA like 17	2.0
AT1G68825	DVL5, RTFL15, ROTUNDIFOLIA like 15	2.0
AT5G58320	Kinase interacting (KIP1-like) family protein	2.0
AT3G29034	Unknown protein	1.9
AT1G22882	Galactose-binding protein	1.6
AT4G04745	Unknown protein	1.6
AT4G09180	basic helix-loop-helix (bHLH) DNA-binding superfamily protein	1.3
AT1G45249	ABF2, abscisic acid responsive elements-binding factor 2	1.3
AT5G38200	Class I glutamine amidotransferase-like superfamily protein	1.2
AT2G18193	P-loop containing nucleoside triphosphate hydrolases superfamily protein	1.2
AT5G10200	ARM-repeat/Tetratricopeptide repeat (TPR)-like protein	1.2
AT5G52882	P-loop containing nucleoside triphosphate hydrolases superfamily protein	0.9
AT2G31141	Unknown protein	0.8
AT3G48340	Cysteine proteinases superfamily protein	−1.0
AT2G23790	Protein of Unknown function (DUF607)	−1.2
AT1G52120	Mannose-binding lectin superfamily protein	−1.5
AT4G39795	Protein of Unknown function (DUF581)	−1.9
AT3G06550	O-acetyltransferase family protein	−2.0

Reads were mapped to the Arabidopsis genome and regulated genes were determined using DESeq. We identified protein coding genes that were not represented on the Affymetrix ATH1 microarray. We show the $\log_2$(KNO$_3$/KCl) value for 2 biological replicates.

Prediction of new genes

In order to identify regions of the Arabidopsis genome that could encode new genes expressed under our experimental nitrate-treatment conditions, we searched for clusters of sequences that match the genome uniquely in regions without annotation (see Methods). These clusters could overlap annotated genes but in anti-sense orientation. Average exon length in the Arabidopsis genome (TAIRv10) is 298 nt, therefore we only considered clusters of 300 nt or more. We found 17 clusters with

these criteria (Additional file 10), 4 of which were located in the complementary strand of annotated genes and might represent natural antisense transcripts (NATs). Two clusters have been previously reported as cis-NATs in an analysis of Arabidopsis full-length cDNAs, At5g49440 and At3g19380 [56]. We found that one of the 17 clusters was induced by nitrate treatments. We labeled this cluster TCP23as as it is antisense to the TCP transcription factor TCP23 (At1g35560) (Figure 3A). We also found sRNAs matching the same region both in sense and antisense orientation (Figure 3B), however we did not find a correlation between their expression and TCP23as regulation by nitrate, suggesting these sequences most likely represent degradation products of TCP23 and

TCP23as. TCP23 was found not to be regulated by nitrate in our RNA-sequencing data (Additional file 5). In order to validate expression of this putative antisense transcript, we reverse-transcribed root RNA using strand-specific primers for TCP23 and for its antisense transcript and performed PCR with gene specific primers. As shown in Figure 3C, both TCP23 and TCP23as are expressed in roots, but only TCP23as is induced by the nitrate treatment. Our data suggest TCP23as could represent a novel nitrate-regulated transcript that might regulate TCP23 expression at the transcriptional or post-transcriptional level. Given the low expression levels of TCP23 in whole root sample, it is likely that its regulation by TCP23as occurs only in a subset of root cells.
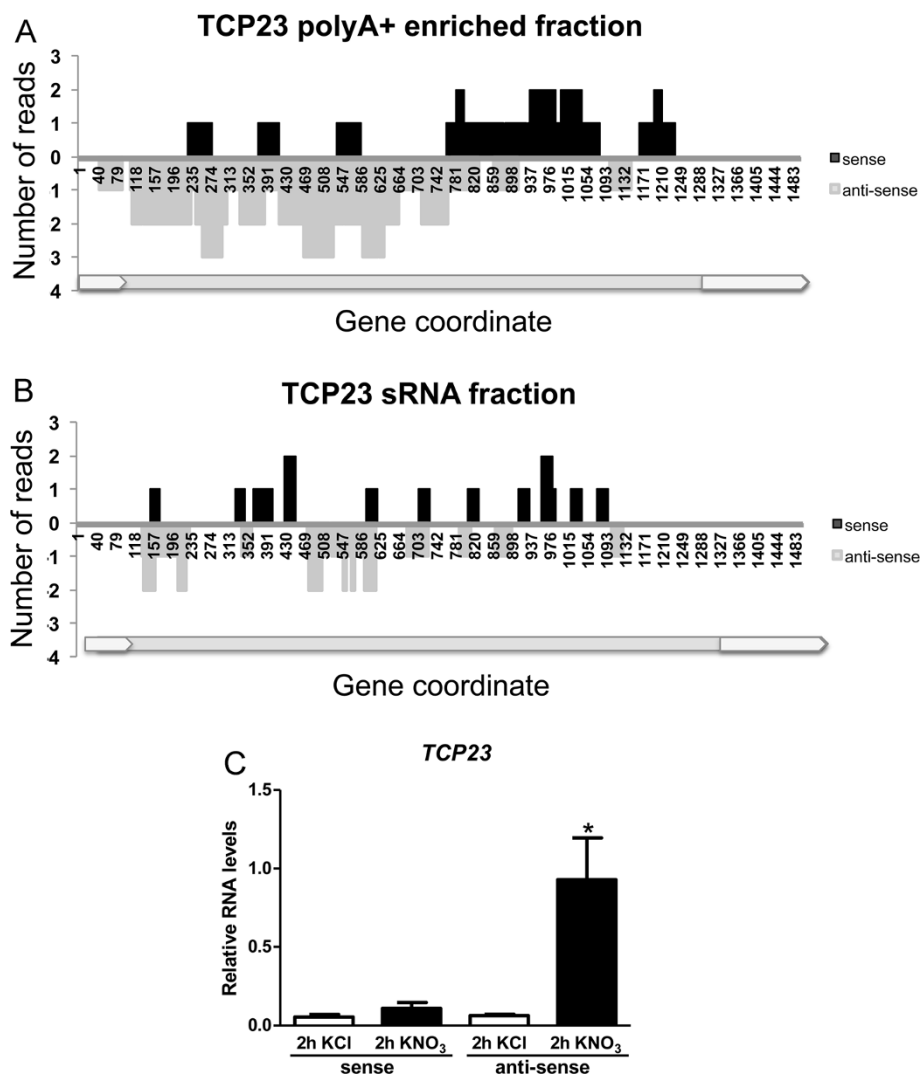

Figure 3 TCP23as is a novel nitrate-regulated gene that is anti-sense to TCP23. A. We represent RNA poly-A + transcripts mapping the TCP23 (AT1G35560) region. The black and grey bars represent reads sense or antisense to AT1G35560 respectively. We show the gene structure of TCP23 in gray. Gray represents 5'UTR and 3'UTR and dark gray represents the coding region. **B**. We represent sRNA transcripts mapping the TCP23 locus similar to panel A. **C**. cDNA was prepared using strand-specific primers for TCP23 and TCP23as. We quantified relative RNA levels of both transcripts using RT-qPCR. We show the results of three biological replicates and standard error.

Prediction of novel miRNA genes

Numerous approaches have been utilized to predict and discover miRNAs [57,58]. However, few experiments have been performed under contrasting N nutrient conditions [33,34,36]. To generate a list of putative new miRNAs that may be expressed under our experimental conditions, we used the miRNA gene prediction tool available in the University of East Anglia (UEA) sRNA toolkit, miRCat (http://srna-tools.cmp.uea.ac.uk) [59]. We chose this prediction tool because it is optimized for the identification of plant miRNA hairpins, and it has been trained and tested with published *Arabidopsis thaliana* high-throughput sRNA sequence data. We used as input for miRCat the filtered sRNA sequences obtained from our 4 sRNA libraries. miRCat was able to predict 123 mature miRNA sequences corresponding to 87% of the known miRNAs identified in our samples, indicating that the prediction algorithms implemented in miRCat are highly efficient in identifying plant miRNAs. The miRCat program was able to predict 51 new miRNA sequences when compared with miRBase v17 (Table 2). From these 51 new miRNA sequences our studies uncovered, 12 were recently cross-validated by other groups, suggesting the veracity of our results [47,60-63]. 21 putative new miRNA sequences were found in intergenic regions, likely representing new transcriptional units (Table 2). 10 miRNA sequences were found inside introns and 2 were found in the 5'UTR of protein coding genes suggesting they are transcribed along with the gene they overlap and 1 miRNA was found in a pseudogene (Table 2). 17 new miRNA sequences were located inside the region coding for the stem loop of known miRNAs (Table 2). Sequences that map onto miRNA precursors and that do not correspond to the mature miRNA or miRNA* sequences have been previously reported in Arabidopsis and are potentially functional miRNAs that are generated by the miRNA pathway [64,65]. All these new miRNAs have low expression levels, most of them being sequenced less than 50 times in our libraries, which probably explains why they have not been reported previously.

A novel nitrate-responsive miRNA/target regulatory module (AtPPC3/miR5640)

In order to further characterize the role of the novel miRNAs in the root nitrate response, we predicted target genes for new miRNAs sequenced in our libraries using the target prediction tool Target finder from the UEA sRNA toolkit, (Additional file 11). The program is based on a set of rules determined specifically for plant miRNA/TARGET interactions [66,67]. We looked in the target list for genes that could be related to N metabolism or to root growth regulation and that were either induced or repressed by nitrate based on our Illumina

results. One of the predicted targets was the transcript for PHOSPHOENOL PYRUVATE CARBOXYLASE 3 (AtPPC3, At3g14940) (Additional file 11), an enzyme that catalyzes CO_2 incorporation with phosphoenol pyruvate to form oxaloacetate [68]. *AtPPC3* is induced in roots after nitrate treatment based on our sequencing data (Additional file 8). The miRNA predicted to target *AtPPC3* has recently been reported as miR5640 [47]. miR5640 has been shown to be expressed in Arabidopsis primary root in the apical half of the meristematic zone (early meristematic zone), the elongation zone, and the maturation zone, according to sequencing data, but no additional validation on its expression or additional characterization of its function or target prediction has been performed [47]. In order to validate miR5640 as a *bona fide* miRNA, we confirmed its expression and expression of its precursor in roots using RT-qPCR. In addition, miR5640 precursor accumulated in the DCL1 (*dcl1-9*) mutant plants (Figure 4A), indicating that miR5640 precursor is processed by DCL1 as most miRNA precursors [69]. In order to experimentally confirm that AtPPC3 is a miR5640 target and to map the miR5640 cleavage site, we performed a modified RLM-RACE procedure [70]. We were able to detect and clone an amplification product corresponding to the expected size of a miR5640-cleaved AtPPC3 fragment. It has been described that cleavage of the target transcripts occurs near the middle of the base-pairing interaction [71,72]. As shown in Figure 4B, 30 out of 32 clones sequenced had a cleavage site inside the miRNA complementary sequence, between the 8th and 9th complementary bases from the miRNA 5' end. This result suggests that *AtPPC3* is a target of miR5640 and further corroborates miR5640 as a *bona fide* miRNA. Based on our sequencing data, we did not find differential expression of miR5640 2 hours after nitrate treatment, although *AtPPC3* is induced by this treatment. In order to determine if miR5640/AtPPC3 could represent a nitrate-responsive miRNA/TARGET module, we analyzed the nitrate response of the miR5640/AtPPC3 pair on a time course using RT-qPCR. As shown in Figure 4C, *AtPPC3* peak of induction by nitrate correlates with miR5640 repression by nitrate. The reduction of *AtPPC3* levels over time also correlates with the de-repression of miR5640, suggesting that *AtPPC3* levels are post-transcriptionally regulated by this miRNA in response to nitrate. Thus, miR5640/AtPPC3 represents a nitrate-responsive module that could be important for modulating carbon/N balance for nitrate assimilation in Arabidopsis roots.

Discussion

High throughput sequencing approaches have become powerful tools to identify the transcriptome of Arabidopsis and other systems. Besides the ability to

Table 2 Illumina sequencing identifies novel miRNAs

miRNA	Chr	miRNA Start	miRNA End	Mature Sequence	miRNA*	miRNA previoulsy reported	miRNA located in
miR5640	1(−)	1653540	1653560	AUGAGAGAAGGAAUUAGAUUC	YES	ath-miR5640 [47]	AT1G05570.1 intron
ath-MIR472-5p	1(−)	4182266	4182286	AUGGUCGAAGUAGGCAAAAUC	NO	Novel	ath-MIR472 stem loop
ath-MIR8166	1(−)	4525316	4525337	AGAGAGUGUAGAAAGUUUCUCA	NO	Novel	Intergenic region AT1G13240-AT1G13245
miR5654-3p	1(+)	11786350	11786371	GAAGAUGCUUUGGGAUUUAUUU	NO	miR5654-3p [47,63]	AT1G32583.1, 5'UTR
ath-MIR829-5p	1(−)	11834153	11834173	ACUUUGAAGCUUUGAUUUGAA	YES	Novel	ath-MIR829 stem loop
miR5014a	1(+)	24554009	24554029	UGUUGUACAAAUUUAAGUGUA	YES	ath-miR5014a [47,60]	AT1G65960.1 intron
ath-MIR840-3p	1(−)	771385	771405	UUGUUUAGGUCCCUUAGUUUC	YES	Novel	ath-MIR840 stem loop
ath-MIR398a-3p	2(+)	1040948	1040968	AAGGAGUGGCAUGUGAACACA	YES	Novel	ath-MIR398a stem loop
ath-MIR8180	2(+)	2063980	2063998	UGCGGUGCGGGAGAAGUGC	NO	Novel	Intergenic region AT2G05580-AT2G05590
ath-MIR8175	2(+)	3740938	3740957	GAUCCCCGGCAACGGCGCCA	NO	Novel	Intergenic region AT2G09880-AT2G09890
ath-MIR396a-3p	2(−)	4142331	4142351	GUUCAAUAAAGCUGUGGGAAG	YES	Novel	ath-MIR396a stem loop
ath-MIR8168	2(+)	5080690	5080710	AGGUGCUGAGUGUGCUAGUGC	NO	Novel	Intergenic region AT2G12490-AT2G12500
ath-MIR5632-5p	2(−)	8392588	8392608	UUGAUUCUCUUAUCCAACUGU	YES	Novel	ath-MIR5632 stem loop
ath-MIR8167a	2(+)	8894985	8895006	AGAUGUGGAGAUCGUGGGGAUG	NO	Novel	Intergenic region AT2G20620-AT2G20625
miR5995b	2(−)	10026977	10026997	AAAGAUGCAGAUCAUAUGUCC	YES	ath-miR5995b [63])	Intergenic region AT2G23540-AT2G23550
ath-MIR831-5p	2(+)	10247259	10247280	AGAAGCGUACAAGGAGAUGAGG	NO	Novel	ath-MIR831 stem loop
miR5637	2(−)	12270195	12270216	UAGAGGAAAAUAUAGAGUUGGG	NO	ath-miR5637 [47]	Intergenic region AT2G28620-AT2G28625
ath-MIR8170.1	2(+)	14100020	14100040	AUAGCAAAUCGAUAAGCAAUG	YES	Novel	AT2G33255.1 intron
ath-MIR8170.2	2(+)	14100079	14100099	UUGCUUAAAGAUUUUCUAUGU	YES	Novel	AT2G33255.1 intron
ath-MIR160a-3p	2(+)	16340342	16340362	GCGUAUGAGGAGCCAUGCAUA	YES	Novel	ath-MIR160a stem loop
ath-MIR8171	2(+)	16890466	16890486	AUAGGUGGGCCAGUGGUAGGA	NO	Novel	AT2G40440.1 intron
ath-MIR166a-5p	2(+)	19176128	19176148	GGACUGUUGUCUGGCUCGAGG	YES	Novel	ath-MIR166 stem loop
ath-MIR408-5p	2(+)	19319866	19319886	ACAGGGAACAAGCAGAGCAUG	YES	Novel	ath-MIR408 stem loop
miR5650	2(+)	19686959	19686979	UUGUUUUGGAUCUUAGAUACA	YES	ath-miR5650 [47]	AT2G48140.1 intron
miR173-5p	3(+)	8236161	8236182	UUCGCUUGCAGAGAGAAAUCAC	YES	ath-miR173-5p [62,63]	ath-miR173-5p stem loop
ath-MIR8169	3(+)	8836359	8836379	AUAGACAGAGUCACUCACAGA	NO	Novel	Intergenic region AT3G24340-AT3G24350
ath-MIR8183	3(−)	11747799	11747819	UUUAGUUGACGGAAUUGUGGC	NO	Novel	AT3G30110.1, pseudogene
ath-MIR8165	3(−)	16538510	16538530	AAUGGAGGCAAGUGUGAAGGA	NO	Novel	Intergenic region AT3G45170-AT3G45180
ath-MIR8174	3(−)	16589431	16589451	AUGUGUAUAGGGAAGCUAAUC	NO	Novel	Intergenic region AT5G38460-AT5G38470
miR5651	3(+)	17178489	17178509	UUGUGCGGUUCAAAUAGUAAC	YES	ath-miR5651 [47]	Intergenic region AT3G46616-AT3G46620
ath-MIR8167b	3(−)	8894985	8895006	AGAUGUGGAGAUCGUGGGGAUG	NO	Novel	Intergenic region AT3G47410-AT3G47420
ath-MIR8167c	3(−)	17469946	17469967	AGAUGUGGAGAUCGUGGGGAUG	NO	Novel	Intergenic region AT3G50700-AT3G50710
miR5633	3(+)	19544786	19544807	AUGAUCAUCAGAAAACAGUGAU	NO	ath-miR5633 [47])	Intergenic region AT3G52730-AT3G52740
ath-MIR393b-3p	3(+)	20691778	20691798	AUCAUGCGAUCUCUUUGGAUU	YES	Novel	ath-MIR393 stem loop

Table 2 Illumina sequencing identifies novel miRNAs *(Continued)*

ath-MIR8182	3(+)	22678166	22678187	UUGUGUUGCGUUUCUGUUGAUU	NO	Novel	AT3G61270.1, 5'UTR
ath-MIR166b-5p	3(+)	22922212	22922232	GGACUGUUGUCUGGCUCGAGG	YES	Novel	ath-MIR166 stem loop
ath-MIR8172	4(−)	7102572	7102592	AUGGAUCAUCUAGAUGGAGAU	YES	Novel	Intergenic region AT4G11800-AT4G11810
ath-MIR8179	4(−)	7161930	7161950	UGACUGCAUUAACUUGAUCGU	NO	Novel	AT4G1192.1 intron
ath-MIR8176	4(+)	11795199	11795219	GGCCGGUGGUCGCGAGAGGGA	NO	Novel	Intergenic region AT4G22320-AT4G22330
ath-MIR8178	4(+)	18087285	18087305	UAACAGAGUAAUUGUACAGUG	NO	Novel	AT4G38760.1 intron
ath-MIR8184	5(−)	3311974	3311994	UUUGGUCUGAUUACGAAUGUA	NO	Novel	Intergenic region AT5G10504-AT5G10510
miR5629	5(+)	3802933	3802954	UUAGGGUAGUUAACGGAAGUUA	NO	ath-miR5629 [47]	Intergenic region AT5G11790-AT5G11800
ath-MIR865.2	5(+)	5169992	5170011	UCUGGGAUGAAUUUGGAUCU	NO	Novel	ath-MIR865 stem loop
miR1888	5(+)	7168879	7168899	UAAGUUAAGAUUUGUGAAGAA	NO	ath-miR1888 [61,62]	AT5G21100.1 intron
ath-MIR8173	5(−)	7478572	7478592	AUGUGCUGAUUCGAGGUGGGA	NO	Novel	Intergenic region AT5G22510-AT5G22520
ath-MIR8177	5(−)	9362634	9362655	GUGUGAUGAUGUGUCAUUUAUA	NO	Novel	Intergenic region AT5G26617-AT5G26620
miR5638b	5(+)	14100017	14100037	ACAGUGGUCAUCUGGUGGGCU	NO	ath-miR5638b [47]	Intergenic region AT5G35945-AT5G35950
ath-MIR160c-3p	5(−)	19009095	19009115	CGUACAAGGAGUCAAGCAUGA	YES	Novel	ath-MIR160c stem loop
ath-MIR870-5p.1	5(−)	21395592	21395612	UUAGAAUGUGAUGCAAAACUU	NO	Novel	ath-MIR870 stem loop
ath-MIR870-5p.2	5(−)	21395604	21395624	AAGAACAUCAAAUUAGAAUGU	NO	Novel	ath-MIR870 stem loop
ath-MIR8181	5(−)	21641289	21641308	UGGGGGUGGGGGGGUGACAG	NO	Novel	AT5G5333.1, intron

Sequences from the Illumina libraries were queried for sequences representing putative new miRNAs with the miRCat program from the UEA sRNA toolkit. We show the genomic location and sequences of these miRNAs. Presence of reads corresponding to miRNA* are indicated (Yes or No). miRNAs that are not included in miRBase 17 but that have been cross-validated by other groups are indicated.
The name of mature sequences derived from the same arm of known hairpin precursor were named with the suffix .1 or .2. If the sequences derive from the opposite arm from the previously annotated sequence, were named with the suffix -5p, -3p.
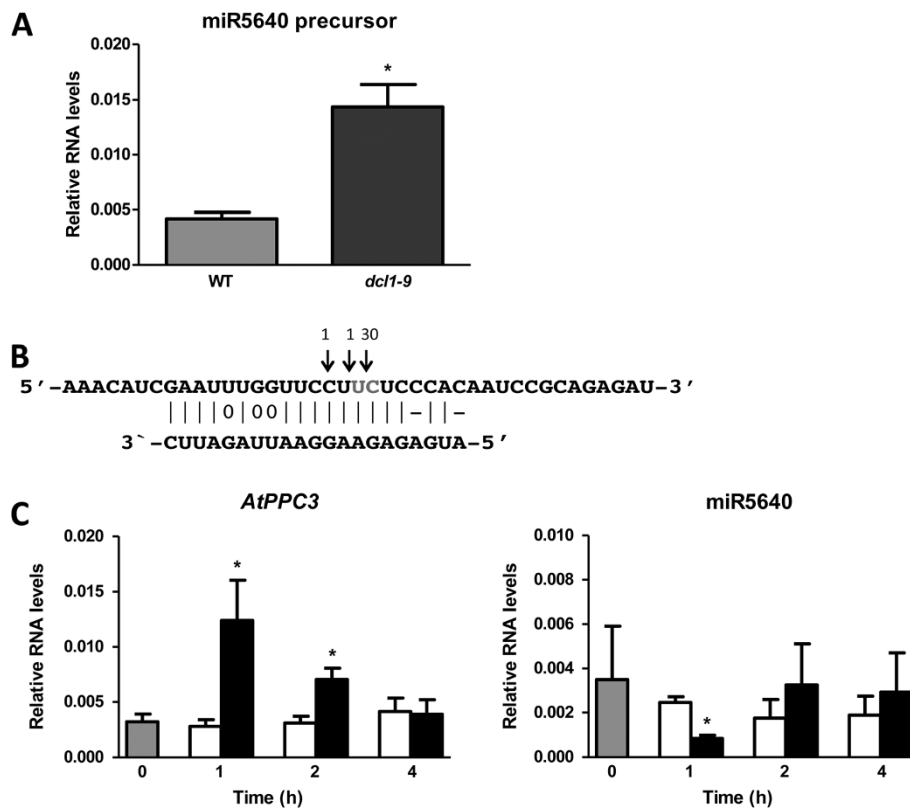
Figure 4 Illumina sequencing identifies a novel miRNA/TARGET module consisting of miR5640 and its target, *AtPPC3*. A. We analyzed the RNA levels of miR5640 predicted precursor using RT-qPCR in WT plants and in the *dcl1-9* mutant. **B**. We used RLM-RACE to validate *AtPPC3* as target of miR5640. The arrows show the numbers of colonies found with the cleavage product. **C**. We determined the RNA levels of *AtPPC3* and of mature miR5640 after 1, 2 and 4 hours of nitrate (black bars) or KCl (white bars) treatments. We show the results of three biological replicates and standard error.

profile novel genes expressed at low levels which could not be identified by traditional cloning and sequencing approaches, the high depth of sequencing obtained by these techniques allows for the absolute quantification of genes, and the comparison of gene expression under different experimental conditions [38,73,74]. Our high throughput sequencing results provided a detailed view of poly-A + RNAs and sRNAs expressed in Arabidopsis roots. We found that roots express a considerable portion of known protein coding genes and miRNA genes. However, most of these genes are expressed at low levels. These transcripts might represent cell specific transcripts whose expression is diluted when considering the whole root. Transcriptomics analysis of specific root cell types has shown that gene expression has an important cell-specific component that gives rise to functional diversification of cells [52,53].

Even though the sequencing depth used to characterize the sRNA component did not allow for accurate quantitative estimates, we were able to discover novel miRNAs that have eluded previous efforts. Our bioinformatics analysis predicted 51 putative miRNAs expressed in roots under the experimental conditions. Most of these sequences were

poorly expressed with less than 1 transcript per million transcripts. A recent publication that analyzes miRNA expressed in specific developmental zones and cell types of the root shows that 9 of these new miRNAs have cell or developmental zone specific expression [47] which can explain their low expression in the whole root samples. We were able to validate one of the predicted miRNAs, miR5640, as a putative miRNA expressed in roots. This miRNA is located inside intron 23 of the *CALLOSE SYNTHASE 1* gene (*CALS1*, AT1G05570). Intronic miRNAs represent the majority of the miRNAs of animal systems but there are only a few examples in Arabidopsis [75,76]. Characterized intronic Arabidopsis miRNAs include miR162a and miR838 which are involved in the regulation of *DCL1* [24,77,78]. However, analyzing our sequencing results, we found that the CALS1 transcript was not regulated by nitrate, thus miR5640 could have an independent nitrate-responsive promoter or pri-miR5640 processing to generate the mature miRNA could be a nitrate-regulated process.

We found miR5640 targeted the transcript that codes for *AtPPC3*, one of the four phosphoenolpyruvate carboxylase enzymes in Arabidopsis [79]. AtPPCs are important enzymes of carbon metabolism that catalyze the

β-carboxylation of phosphoenolpyruvate to yield oxalo-acetate. In C3 plants and algae, it has been shown that ATPPCs are important for the production of carbon skeletons for nitrogen assimilation [68,80,81]. Although there has been an extensive biochemical characterization of the AtPPCs enzymes in Arabidopsis, there are no reports of their function in N metabolism. AtPPC3 is a root specific AtPPC [82] and we found that it was nitrate-induced in our experiments, which is in agreement with the positive effect on nitrate assimilation predicted for this AtPPC. We also found evidence indicating that nitrate induction of *AtPPC3* might depend on a miR5640-mediated post-transcriptional regulation of *AtPPC3* levels in response to nitrate. Although we found AtPPC3 cleavage products that might be generated by miR5640 action over this transcript, we need further experiments to validate *AtPPC3* as a miR5640 target (i.e. to analyze *AtPPC3* levels in a miR5640 overexpressor plant), and to validate the role of this miRNA/*TARGET* module in nitrate assimilation in roots.

An advantage of using high throughput sequencing is the ability to interrogate gene expression without the representation bias present in microarray experiments. We discovered 40 protein-coding genes that have not been reported to be nitrate-responsive in previous transcriptomics analysis of Arabidopsis roots. Among them, we found highly responsive genes such as *BT1* (At5g63160), a calmodulin-binding scaffold protein that acts redundantly with other BT proteins in female gametophyte development [83]. The closest homolog of BT1, BT2, has been reported to be responsive to multiple hormonal, stress and nutritional signals, including nitrate [84]. Interestingly, *BT1* is only expressed when nitrate is supplied, suggesting that it might have a nitrate-specific function in roots. The *AGG2* gene, one of the two genes encoding the gamma subunit of heterotrimeric G protein was also induced by nitrate. Heterotrimeric G protein in Arabidopsis has been involved in various developmental processes. In roots, it is involved in lateral root formation [85] and root apical meristem growth [86]. We have found that nitrate has an effect in primary and lateral root growth [37], thus nitrate regulation of *AGG2* might contribute to this response.

NATs are transcripts that fully or partially overlap with other transcripts. These pairs can mediate production of siRNAs to silence gene expression [87]. Additionally, NATs can modulate transcription, can affect mRNA stability and translation and can induce chromatin and DNA epigenetic changes [88]. Computational predictions have shown that the Arabidopsis genome potentially encodes sense-antisense transcript pairs representing approximately 7% of the protein coding genes [56]. We were able to identify 4 putative NATs of >300 bp in our sequencing data. One of these NATs was antisense to *TCP23* gene

and was induced by nitrate. *TCP* genes are transcription factors that promote growth and proliferation [89]. TCP23 is predicted to contain a chloroplast-targeting peptide, suggesting it might control transcription of chloroplast genes [90]. Although TCP23 has no described function, other class I TCP factors have been shown to be expressed in meristematic tissues and to control cell cycle genes such as *PCNA* and *CYCB1;1* [91,92]. Thus, *TCP23as* induction by nitrate might repress *TCP23* expression, controlling meristematic activity of the primary root. However, further studies are needed to analyze *TCP23as* role over *TCP23* expression on roots and on *TCP23* regulation by nitrate.

Conclusions

In summary, the sequencing of small RNAs and mRNAs uncovered new genes, and enabled us to develop new hypotheses for nitrate regulation and coordination of carbon and N metabolism. A highlight is the discovery of a novel microRNA, miR5640 and its target, *AtPPC3*. The data suggest that the nitrate-responsive miRNA/target module might be involved in controlling carbon flux to assimilate nitrate into amino acids. These findings suggest that microRNAs can have metabolic regulatory functions, as well as previously described developmental functions [37,44] in the nitrate response of Arabidopsis roots.

Methods

Growth and treatment conditions

Approximately 1,500 Arabidopsis seedlings were grown hydroponically on Phytatrays on MS-modified basal salt media without N (Phytotechnology Laboratories, M531) supplemented with 0.5 mM ammonium succinate and 3 mM sucrose under a photoperiod of 16 h of light and 8 h of darkness and a temperature of 22°C using a plant growth incubator (Percival Scientific, Inc.). After 2 weeks, plants were treated with 5 mM KNO_3 or 5 mM KCl as control for 2 hours.

Preparation of illumina libraries

Total RNA from from nitrate-treated or control roots was extracted using Trizol® (Invitrogen, cat. Number 15596–026). For poly-A + libraries, poly-A + RNA was enriched using the Poly(A)Purist™ MAG Kit (Ambion, cat, number AM1922M). Poly-A + RNA was decapped using tobacco acid pyrophosphatase and fragmented using RNA Fragmentation Reagents (Ambion, cat. Number AM8740). Low molecular weight RNA (<40 nt) was isolated from 100 µg of total RNA by PAGE on a FlashPAGE™ fractionator (Ambion, cat. Number AM13100). For construction of the libraries, cloning linker (AMP-5'p = 5'pCTG TAG GCA CCA TCA ATdideoxyC-3') was ligated to the 3' end of the RNA followed by purification of the ligation product on a 15% polyacrilamide/urea gel. The 3'-ligated

product was ligated to the 5' Solexa linker (5'-rArCrA rCrUrC rUrUrU rCrCrC rUrArC rArCrG rArCrG rCrUrC rUrUrC rCrGrA rUrC-3'). RNA with ligated adaptors was reverse transcribed into DNA using Illumina specific primer (5'- CAA GCA GAA GAC GGC ATA CGA TTG ATG GTG CCT ACA G-3') and cDNA was then PCR amplified using this primer and a specific primer (5'- AAT GAT ACG GCG ACC ACC GAA CAC TGT TTC CCT ACA CGA CG-3'). The libraries were gel purified using the QIAquick gel extraction kit (QIAGEN, cat. Number 28704). Libraries were sequenced on the Illumina 1G Genome analyzer.

Sequence analysis

Raw sequences from the Illumina 1G Genome analyzer in FASTQ format were analyzed with publicly available tools. Low quality reads were extracted with fastq quality filter by FASTX toolkit version 0.0.13 (http://hannonlab. cshl.edu/fastx_toolkit/). The Phred quality score was set to 20, a probability of incorrect base call of 1 in 100. 3' adaptor sequences were trimmed from the Illumina reads, and then were mapped to the Arabidopsis TAIR10 genome using Novoalign version 2.05.17 (www.novocraft. com). Perfect match sequences having passed the quality control, polynucleotide filter, and size filter (between 18 and 28 nt for sRNA libraries and ≥18 nt for poly-A + libraries) were selected for further analysis with custom made PERL scripts.

Determination of differentially expressed genes

To evaluate differential gene expression between KNO_3 and KCl treated samples, we used sequence counts corresponding to sRNAs or annotated elements as input for the DESeq package version 1.1.6 [55] available from Bioconductor (www.bioconductor.org). This tool uses a negative binomial distribution model to test for differential gene expression [55]. We found correlation values of 0.91 and 0.96 for controls and treatments respectively for sRNA-seq and of 0.99 for controls and treatments for RNA-seq data. Replicates were used independently for statistical analysis of gene expression. We adjusted for multiple testing using FDR correction [93] and filtered genes whose expression changed with corrected p-values ≥ 0.05.

New miRNA and target predictions

Quality filtered Illumina sequences were used as input for the MIRCAT tool [59], available at the University of East Anglia (UEA) sRNA toolkit (http://srna-tools.cmp. uea.ac.uk) using default parameters. To predict miRNA targets, we used the target prediction tool available from the UEA sRNA toolkit. The predicted targets, along with the putative cleavage site on these targets, were further validated using RNAhybrid version 2.1 [94].

Predicting novel transcribed regions

Novoalign alignments that did not overlap with annotated regions of the genome were pooled from all samples. Regions with continuous alignments in the same strand greater than 300 bp were identified as candidate novel transcribed regions.

Gene expression analysis using RT-qPCR

Gene expression analysis was carried out using the Brilliant® SYBR® Green QPCR Reagents on a Stratagene MX3000P qPCR system (Agilent) according to manufacturer's instructions. The RNA levels were normalized relative to the Clathrin adaptor complexes medium subunit family protein (At4g24550). Quantification of microRNA levels was carried out using the High-Specificity miRNA QRT-PCR Detection Kit from Stratagene on a Stratagene MX3000P qPCR system. The RNA levels were normalized relative to U6 snRNA (At3g14735). A list of RT-qPCR primers used in this work is provided in Additional file 12.

RLM-RACE

A modified procedure for RLM-RACE [70] was carried out using the GeneRacer™ kit. The GeneRacer RNA Oligo adapter was directly ligated to 250 ng of Poly-A + mRNA and the GeneRacer OligodT primer was used to synthesize first strand cDNA. This cDNA was subjected to a PCR amplification procedure with the GeneRacer 5′Primer and the GeneRacer 3′Primer to generate a pool of non-genespecific RACE products. Gene-specific 5′RACE reactions were performed with the GeneRacer 5′Nested Primer and a reverse gene-specific primer. The expected size of the PCR amplicons was checked on a 3% agarose gel. PCR products were cloned and sequenced to confirm predicted miRNA-mediated cleavage of the transcripts.

Availability of supporting data

The data sets supporting the results of this article are available in the NCBI GEO database [95] repository, under accession GSE44062.

Additional files

Additional file 1: Statistics of filtered sRNA and mRNA reads.

Additional file 2: Global profiling of Illumina-sequenced sRNA.

Additional file 3: Known miRNA and ta-siRNA expressed in roots.

Additional file 4: mRNA reads that unambigoulsly match intergenic regions are located near 5′ and 3′ of annotated genes.

Additional file 5: Expression analysis of annotated elements of TAIR v.10.

Additional file 6: Gene expression distribution of genes represented in the ATH1 microarray in Illumina poly-A + libraries.

Additional file 7: Correlation between nitrate-regulated genes obtained by RNA-seq and Affymetrix ATH1 microarrays depends on Average gene coverage.

Additional file 8: Nitrate regulated genes determined using Illumina deep sequencing.

Additional file 9: Deep sequencing of the Arabidopsis root poly-A + enriched fraction identifies new nitrate-regulated genes.

Additional file 10: Identification of putative new genes.

Additional file 11: Prediction of putative targets for novel miRNA sequences.

Additional file 12: List of primers used for qRT-PCR.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

EAV participated in the design of the study, performed molecular biology experiments and statistical analyses and wrote the paper. TCM performed the computational analysis of the data, performed the experiments and wrote the paper. GK participated in the design of the study, performed experiments, analyzed the data and helped to draft the manuscript. MK performed computational analysis of the data and helped to draft the manuscript. MT generated the libraries for sequencing and helped to draft the manuscript. WRM participated in sequencing of the libraries. GMC participated in the design and coordination of the study and helped to draft the manuscript. RAG participated in the design and coordination of the study and wrote the paper. All authors read and approved the final manuscript.

Author details

[1]FONDAP Center for Genome Regulation. Millennium Nucleus Center for Plant Functional Genomics, Departamento de Genética Molecular y Microbiología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago 8331010, Chile. [2]Biochimie et Physiologie Moléculaire des Plantes, UMR 5004 CNRS/INRA/SupAgro-M/UM2, Institut de Biologie Intégrative des Plantes-Claude Grignon, Montpellier, France. [3]Center for Genomics and Systems Biology, Department of Biology, New York University, New York, NY 10003, USA. [4]Cold Spring Harbor Laboratory, New York, USA. [5]Present address: School of Biological Sciences, The University of Queensland, St Lucia, Australia.

References

1. Crawford NM: **Nitrate: nutrient and signal for plant growth.** *Plant Cell* 1995, **7**:859–868.
2. Redinbaugh MG, Campbell WH: **Higher plant responses to environmental nitrate.** *Physiol Plant* 1991, **82**:640–650.
3. Stitt M: **Nitrate regulation of metabolism and growth.** *Curr Opin Plant Biol* 1999, **2**:178–186.
4. Crawford NM, Glass ADM: **Molecular and physiological aspects of nitrate uptake in plants.** *Trends Plant Sci* 1998, **3**:389–395.
5. Vidal EA, Gutiérrez RA: **A systems view of nitrogen nutrient and metabolite responses in Arabidopsis.** *Curr Op Plant Biol* 2008, **11**:521–529.
6. Zhang H, Rong H, Pilbeam D: **Signalling mechanisms underlying the morphological responses of the root system to nitrogen in *Arabidopsis thaliana*.** *J Exp Bot* 2007, **58**:2329–2338.
7. Forde BG, Walch-Liu P: **Nitrate and glutamate as environmental cues for behavioural responses in plant roots.** *Plant Cell Environ* 2009, **32**:682–693.
8. Krouk G, Mirowski P, LeCun Y, Shasha D, Coruzzi G: **Predictive network modeling of the high-resolution dynamic plant transcriptome in response to nitrate.** *Genome Biol* 2010, **11**:R123.
9. Wang R, Okamoto M, Xing X, Crawford NM: **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* 2003, **132**:556–567.
10. Wang R, Tischner R, Gutierrez RA, Hoffman M, Xing X, Chen M, Coruzzi G, Crawford NM: **Genomic analysis of the nitrate response using a nitrate reductase-null mutant of Arabidopsis.** *Plant Physiol* 2004, **136**:2512–2522.
11. Wang R, Xing X, Crawford N: **Nitrite acts as transcriptome signal at micromolar concentrations in Arabidopsis roots.** *Plant Physiol* 2007, **145**:1735–1745.
12. Wang R, Xing X, Wang Y, Tran A, Crawford NM: **A genetic screen for nitrate regulatory mutants captures the nitrate transporter gene NRT1.1.** *Plant Physiol* 2009, **151**:472–478.
13. Blencowe BJ, Ahmad S, Lee LJ: **Current-generation high-throughput sequencing: deepening insights into mammalian transcriptomes.** *Genes Dev* 2009, **23**:1379–1386.
14. Malone J, Oliver B: **Microarrays, deep sequencing and the true measure of the transcriptome.** *BMC Biol* 2011, **9**:34.
15. Marguerat S, Bähler J: **RNA-seq: from technology to biology.** *Cell Mol Life Sci* 2010, **67**:569–579.
16. Morozova O, Hirst M, Marra MA: **Applications of new sequencing technologies for transcriptome analysis.** *Annu Rev Genomics Hum Genet* 2009, **10**:135–151.
17. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and quantifying mammalian transcriptomes by RNA-Seq.** *Nat Methods* 2008, **5**:621–628.
18. Wang Z: **RNA-Seq: a revolutionary tool for transcriptomics.** *Nat Rev Genet* 2009, **10**:57–63.
19. Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Law TF, Grant SR, Dangl JL, Carrington JC: **High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of miRNA genes.** *PLoS One* 2007, **2**:e219.
20. Henderson IR, Zhang X, Lu C, Johnson L, Meyers BC, Green PJ, Jacobsen SE: **Dissecting *Arabidopsis thaliana* DICER function in small RNA processing, gene silencing and DNA methylation patterning.** *Nat Genet* 2006, **38**:721–725.
21. Kasschau KD, Fahlgren N, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Carrington JC: **Genome-wide profiling and analysis of Arabidopsis siRNAs.** *PLoS Biol* 2007, **5**:e57.
22. Lu C, Kulkarni K, Souret FF, MuthuValliappan R, Tej SS, Poethig RS, Henderson IR, Jacobsen SE, Wang W, Green PJ, Meyers BC: **MicroRNAs and other small RNAs enriched in the Arabidopsis RNA-dependent RNA polymerase-2 mutant.** *Genome Res* 2006, **16**:1276–1288.
23. Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu A-L, Zhao Y, McDonald H, Zeng T, Hirst M, *et al*: **Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells.** *Genome Res* 2008, **18**:610–621.
24. Rajagopalan R, Vaucheret H, Trejo J, Bartel DP: **A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*.** *Genes Dev* 2006, **20**:3407–3425.
25. Huijser P, Schmid M: **The control of developmental phase transitions in plants.** *Development* 2011, **138**:4117–4129.
26. Khan G, Declerck M, Sorin C, Hartmann C, Crespi M, Lelandais-Brière C: **MicroRNAs as regulators of root development and architecture.** *Plant Mol Biol* 2011, **77**:47–58.
27. Martin RC, Liu P-P, Goloviznina NA, Nonogaki H: **microRNA, seeds, and Darwin?: diverse function of miRNA in seed biology and plant responses to stress.** *J Exp Bot* 2010, **61**:2229–2234.
28. Ruiz-Ferrer V, Voinnet O: **Roles of plant small RNAs in biotic stress responses.** *Annu Rev Plant Biol* 2009, **60**:485–510.
29. Sunkar R, Li Y-F, Jagadeeswaran G: **Functions of microRNAs in plant stress responses.** *Trends Plant Sci* 2012, **17**:196–203.
30. Chiou T-J, Aung K, Lin S-I, Wu C-C, Chiang S-F, Su C-I: **Regulation of phosphate homeostasis by microRNA in Arabidopsis.** *Plant Cell* 2006, **18**:412–421.
31. Fujii H, Chiou T-J, Lin S-I, Aung K, Zhu J-K: **A miRNA involved in phosphate-starvation response in Arabidopsis.** *Curr Biol* 2005, **15**:2038–2043.
32. Kawashima CG, Yoshimoto N, Maruyama-Nakashita A, Tsuchiya YN, Saito K, Takahashi H, Dalmay T: **Sulphur starvation induces the expression of**

microRNA-395 and one of its target genes but in different cell types. *Plant J* 2009, **57**:313–321.
33. Liang G, He H, Yu D: Identification of Nitrogen starvation-responsive microRNAs in *Arabidopsis thaliana*. *PLoS One* 2012, **7**:e48951.
34. Pant BD, Musialak-Lange M, Nuc P, May P, Buhtz A, Kehr J, Walther D, Scheible W-R: Identification of nutrient-responsive Arabidopsis and Rapeseed microRNAs by comprehensive real-time polymerase chain reaction profiling and small RNA sequencing. *Plant Physiol* 2009, **150**:1541–1555.
35. Zhao M, Ding H, Zhu J-K, Zhang F, Li W-X: Involvement of miR169 in the nitrogen-starvation responses in Arabidopsis. *New Phytol* 2011, **190**:906–915.
36. Zhao M, Tai H, Sun S, Zhang F, Xu Y, Li W-X: Cloning and characterization of Maize miRNAs involved in responses to nitrogen deficiency. *PLoS One* 2012, **7**:e29669.
37. Vidal EA, Araus V, Lu C, Parry G, Green PJ, Coruzzi GM, Gutierrez RA: Nitrate-responsive miR393/*AFB3* regulatory module controls root system architecture in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 2010, **107**:4477–4482.
38. Hsieh L-C, Lin S-I, Shih AC-C, Chen J-W, Lin W-Y, Tseng C-Y, Li W-H, Chiou T-J: Uncovering small RNA-mediated responses to phosphate-deficiency in Arabidopsis by deep sequencing. *Plant Physiol* 2009, **151**:2120–2132.
39. Pantaleo V, Szittya G, Moxon S, Miozzi L, Moulton V, Dalmay T, Burgyan J: Identification of grapevine microRNAs and their targets using high-throughput sequencing and degradome analysis. *Plant J* 2010, **62**:960–976.
40. Li F, Orban R, Baker B: SoMART: a web server for plant miRNA, tasiRNA and target gene analysis. *Plant J* 2012, **70**:891–901.
41. Adai A, Johnson C, Mlotshwa S, Archer-Evans S, Manocha V, Vance V, Sundaresan V: Computational prediction of miRNAs in *Arabidopsis thaliana*. *Genome Res* 2005, **15**:78–91.
42. Jones-Rhoades MW, Bartel DP: Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol Cell* 2004, **14**:787–799.
43. Kozomara A, Griffiths-Jones S: miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* 2011, **39**:D152–D157.
44. Gifford ML, Dean A, Gutierrez RA, Coruzzi GM, Birnbaum KD: Cell-specific nitrogen responses mediate developmental plasticity. *Proc Natl Acad Sci USA* 2008, **105**:803–808.
45. Gutierrez RA, Lejay L, Dean A, Chiaromonte F, Shasha DE, Coruzzi GM: Qualitative network models and genome-wide expression data define carbon/nitrogen-responsive molecular machines in Arabidopsis. *Genome Biol* 2007, **8**:R7.
46. Yang X, Zhang H, Li L: Global analysis of gene-level microRNA expression in Arabidopsis using deep sequencing data. *Genomics* 2011, **98**:40–46.
47. Breakfield NW, Corcoran DL, Petricka JJ, Shen J, Sae-Seaw J, Rubio-Somoza I, Weigel D, Ohler U, Benfey PN: High-resolution experimental and computational profiling of tissue-specific known and novel miRNAs in Arabidopsis. *Genome Res* 2012, **22**:163–176.
48. Bertone P, Stolc V, Royce TE, Rozowsky JS, Urban AE, Zhu X, Rinn JL, Tongprasit W, Samanta M, Weissman S, *et al*: Global identification of human transcribed sequences with genome tiling arrays. *Science* 2004, **306**:2242–2246.
49. Johnson JM, Edwards S, Shoemaker D, Schadt EE: Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments. *Trends Genet* 2005, **21**:93–102.
50. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermüller J, Hofacker IL, *et al*: RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* 2007, **316**:1484–1488.
51. Khaitovich P, Kelso J, Franz H, Visagie J, Giger T, Joerchel S, Petzold E, Green RE, Lachmann M, Pääbo S: Functionality of intergenic transcription: an evolutionary comparison. *PLoS Genet* 2006, **2**:e171.
52. Birnbaum K, Shasha DE, Wang JY, Jung JW, Lambert GM, Galbraith DW, Benfey PN: A gene expression map of the Arabidopsis root. *Science* 2003, **302**:1956–1960.
53. Brady SM, Orlando DA, Lee J-Y, Wang JY, Koch J, Dinneny JR, Mace D, Ohler U, Benfey PN: A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science* 2007, **318**:801–806.
54. Scheible WR, Morcuende R, Czechowski T, Fritz C, Osuna D, Palacios-Rojas N, Schindelasch D, Thimm O, Udvardi MK, Stitt M: Genome-wide reprogramming of primary and secondary metabolism, protein synthesis, cellular growth processes, and the regulatory infrastructure of Arabidopsis in response to nitrogen. *Plant Physiol* 2004, **136**:2483–2499.

55. Anders S, Huber W: Differential expression analysis for sequence count data. *Genome Biol* 2010, **11**:R106.
56. Wang X-J, Gaasterland T, Chua N-H: Genome-wide prediction and identification of cis-natural antisense transcripts in Arabidopsis thaliana. *Genome Biol* 2005, **6**:R30.
57. Ghosh Z, Chakrabarti J, Mallick B: miRNomics—The bioinformatics of microRNA genes. *Biochem Biophys Res Commun* 2007, **363**:6–11.
58. Meyers BC, Souret FF, Lu C, Green PJ: Sweating the small stuff: microRNA discovery in plants. *Curr Opin Biotechnol* 2006, **17**:139–146.
59. Moxon S, Schwach F, Dalmay T, MacLean D, Studholme DJ, Moulton V: A toolkit for analysing large-scale plant small RNA datasets. *Bioinformatics* 2008, **24**:2252–2253.
60. Borges F, Pereira PA, Slotkin RK, Martienssen RA, Becker JD: MicroRNA activity in the Arabidopsis male germline. *J Exp Bot* 2011, **62**:1611–1620.
61. German MA, Pillay M, Jeong D-H, Hetawal A, Luo S, Janardhanan P, Kannan V, Rymarquis LA, Nobuta K, German R, *et al*: Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nat Biotech* 2008, **26**:941–946.
62. Moldovan D, Spriggs A, Yang J, Pogson BJ, Dennis ES, Wilson IW: Hypoxia-responsive microRNAs and trans-acting small interfering RNAs in Arabidopsis. *J Exp Bot* 2010, **61**:165–177.
63. Zhang B, Jin Z, Xie D: Global analysis of non-coding small RNAs in Arabidopsis in response to jasmonate treatment by deep sequencing technology. *J Integr Plant Biol* 2012, **54**:73–86.
64. Sobkowiak L, Karlowski W, Jarmolowski A, Szweykowska-Kulinska Z: Non-canonical processing of Arabidopsis pri-miR319a/b/c generates additional microRNAs to target one RAP2.12 mRNA isoform. *Front Plant Sci* 2012, **3**:46.
65. Zhang W, Gao S, Zhou X, Xia J, Chellappan P, Zhou X, Zhang X, Jin H: Multiple distinct small RNAs originate from the same microRNA precursors. *Genome Biol* 2010, **11**:R81.
66. Allen E, Xie Z, Gustafson AM, Carrington JC: microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* 2005, **121**:207–221.
67. Schwab R, Palatnik JF, Riester M, Schommer C, Schmid M, Weigel D: Specific effects of microRNAs on the plant transcriptome. *Dev Cell* 2005, **8**:517–527.
68. Huppe HC, Turpin DH: Integration of carbon and nitrogen metabolism in plant and algal cells. *Annu Rev Plant Physiol Plant Mol Biol* 1994, **45**:577–607.
69. Sunkar R, Zhu J-K: Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis. *Plant Cell* 2004, **16**:2001–2019.
70. Llave C, Xie Z, Kasschau KD, Carrington JC: Cleavage of scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. *Science* 2002, **297**:2053–2056.
71. Brodersen P, Voinnet O: The diversity of RNA silencing pathways in plants. *Trends Genet* 2006, **22**:268–280.
72. Kidner CA, Martienssen RA: The developmental role of microRNA in plants. *Curr Op Plant Biol* 2005, **8**:38–44.
73. Lelandais-Briere C, Naya L, Sallet E, Calenge F, Frugier F, Hartmann C, Gouzy J, Crespi M: Genome-wide medicago truncatula small RNA analysis revealed novel microRNAs and isoforms differentially regulated in roots and nodules. *Plant Cell* 2009, **21**:2780–2796.
74. Zhang J, Xu Y, Huan Q, Chong K: Deep sequencing of Brachypodium small RNAs at the global genome level identifies microRNAs involved in cold stress response. *BMC Genomics* 2009, **10**:449.
75. Brown JWS, Marshall DF, Echeverria M: Intronic noncoding RNAs and splicing. *Trends Plant Sci* 2008, **13**:335–342.
76. Meng Y, Shao C: Large-scale identification of mirtrons in Arabidopsis and Rice. *PLoS One* 2012, **7**:e31163.
77. Xie Z, Kasschau KD, Carrington JC: Negative feedback regulation of Dicer-Like1 in Arabidopsis by microRNA-guided mRNA degradation. *Curr Biol* 2003, **13**:784–789.
78. Kim YK, Kim VN: Processing of intronic microRNAs. *EMBO J* 2007, **26**:775–783.
79. Sanchez R, Cejudo FJ: Identification and expression analysis of a gene encoding a bacterial-type phosphoenolpyruvate carboxylase from Arabidopsis and Rice. *Plant Phys* 2003, **132**:949–957.
80. Podestá FE, Plaxton WC: Regulation of cytosolic carbon metabolism in germinating *Ricinus communis* cotyledons. *Planta* 1994, **194**:381–387.
81. Scheible WR, Krapp A, Stitt M: Reciprocal diurnal changes of phosphoenolpyruvate carboxylase expression and cytosolic pyruvate kinase, citrate synthase and NADP-isocitrate dehydrogenase expression regulate organic acid metabolism during nitrate assimilation in tobacco leaves. *Plant Cell Environ* 2000, **23**:1155–1167.

82. Sánchez R, Flores A, Cejudo FJ: **Arabidopsis phosphoenolpyruvate carboxylase genes encode immunologically unrelated polypeptides and are differentially expressed in response to drought and salt stress.** *Planta* 2006, **223**:901–909.

83. Robert HS, Quint A, Brand D, Vivian-Smith A, Offringa R: **BTB and TAZ domain scaffold proteins perform a crucial function in Arabidopsis development.** *Plant J* 2009, **58**:109–121.

84. Mandadi KK, Misra A, Ren S, McKnight TD: **BT2, a BTB protein, mediates multiple responses to nutrients, stresses, and hormones in Arabidopsis.** *Plant Phys* 2009, **150**:1930–1939.

85. Ullah H, Chen J-G, Temple B, Boyes DC, Alonso JM, Davis KR, Ecker JR, Jones AM: **The β-subunit of the Arabidopsis G protein negatively regulates auxin-induced cell division and affects multiple developmental processes.** *Plant Cell* 2003, **15**:393–409.

86. Chen J-G, Gao Y, Jones AM: **Differential roles of Arabidopsis heterotrimeric G-protein subunits in modulating cell division in roots.** *Plant Physiol* 2006, **141**:887–897.

87. Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu J-K: **Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis.** *Cell* 2005, **123**:1279–1291.

88. Faghihi MA, Wahlestedt C: **Regulatory roles of natural antisense transcripts.** *Nat Rev Mol Cell Biol* 2009, **10**:637–643.

89. Martín-Trillo M, Cubas P: **TCP genes: a family snapshot ten years later.** *Trends Plant Sci* 2010, **15**:31–39.

90. Wagner R, Pfannschmidt T: **Eukaryotic transcription factors in plastids — Bioinformatic assessment and implications for the evolution of gene expression machineries in plants.** *Gene* 2006, **381**:62–70.

91. Kosugi S, Ohashi Y: **PCF1 and PCF2 specifically bind to cis elements in the rice proliferating cell nuclear antigen gene.** *Plant Cell* 1997, **9**:1607–1619.

92. Li C, Potuschak T, Colón-Carmona A, Gutiérrez RA, Doerner P: **Arabidopsis TCP20 links regulation of growth and cell division control pathways.** *Proc Natl Acad Sci USA* 2005, **102**:12978–12983.

93. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP: **Exploration, normalization, and summaries of high density oligonucleotide array probe level data.** *Biostatistics* 2003, **4**:249–264.

94. Rehmsmeier M, Steffen P, Höchsmann M, Giegerich R: **Fast and effective prediction of microRNA/target duplexes.** *RNA* 2004, **10**:1507–1517.

95. Edgar R, Domrachev M, Lash AE: **Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.** *Nucleic Acids Res* 2002, **30**:207–210.