Research Article

Genomewide bioinformatic analysis negates any specific role for Dof, GATA and Ag/cTCA motifs in nitrate responsive gene expression in *Arabidopsis*

Ravi Ramesh Pathak, Suman K. Das, Devapriya Choudhury[1] and Nandula Raghuram

School of Biotechnology, GGS Indraprastha University, Kashmiri Gate, Delhi – 110403
[1]School of Biotechnology, Jawaharlal Nehru University, New Delhi – 110067

ABSTRACT

Nitrate response at the plant level is mediated by the transcriptional regulation of several hundreds of genes, but no common cis-acting nitrate-responsive elements (NREs) have been identified so far. Earlier, we bioinformatically ruled out the possibility that the previously published [$(a/t)_7$Ag/cTCA] motif could act as NRE on its own (Das *et al.*, 2007, Mol. Genet. Genomics, 278: 519-525). In the present study, we examined other motifs such as Dof and GATA binding elements in homologous as well as heterologous pairwise combinations in the *Arabidopsis* genome *in silico*. None of the above three motifs revealed any unique association with nitrate responsive genes or their subsets in any combination, either within their ORFs or 1 kb flanking sequences on either side. Additionally, twelve new, top-scoring candidate motifs that were generated using different online motif samplers were analyzed *in silico* using a subset of 21 'early' nitrate responsive genes, but did not reveal any specificity of occurence. These results underscore the need to continue the search for novel candidate NREs, as possible sites of intervention to understand/improve nitrate-responsive gene expression and nitrate use efficiency. **[Physiol. Mol. Biol. Plants 2009; 15(2) : 145-150]** *E-mail : raghuram98@hotmail.com*

Key words : *AG/CTCA, Arabidopsis, Dof, GATA, Gene expression, motif, Nitrate, NRE, response elements, transcription*

INTRODUCTION

The advent of high-throughput methods in the recent years began to reveal the true extent of the influence of nitrate in plant gene expression, metabolism, growth and development. The pervasiveness of nitrate response, which involves transcriptional regulation of several hundred genes, was recently revealed by microarray analyses in *Arabidopsis* (Wang *et al*., 2000, 2003, 2004; Scheible *et al*., 2004) and tomato (Wang *et al.,* 2001) and by subtractive hybridization analyses in rice (Wang *et al*., 2002).

Based on well known concepts in signal transduction and gene regulation, it is widely expected that common regulation of multiple genes by nitrate as a signal would necessitate the presence of common regulatory sequences or response elements (Raghuram *et al.*, 2006). Several attempts have been made to analyse the regulatory sequences flanking nitrate responsive genes in transgenic plants, but none of them identified the minimum consensus sequence specific to nitrate response (Dorbe *et al*., 1992; Vaucheret *et al*., 1992; Neininger *et al*., 1993; Rastogi *et al*., 1993; Vincentz *et al*., 1993; Lin *et al*., 1994). Using linker-scanning analysis, Hwang *et al*. (1997) proposed that *cis*-acting sequence elements comprising [$(a/t)_7$Ag/cTCA] motifs may mediate nitrate-dependent transcription of NR in *Arabidopsis thaliana*. However, by *in silico* examination of the occurrence of this motif in 625 nitrate-responsive genes in *Arabidopsis* and their 283 homologs in rice against their respective genomic backgrounds, Das *et al*, (2007) showed that the above motif is neither unique to nitrate responsive genes nor common to all of them.

While identification of specific nitrate-induced transcription factors in higher plants has not yet been successful, reports have suggested that some other transcription factors may also help in mediating nitrate response. Yanagisawa *et al*. (2004) raised transgenic *Arabidopsis* lines over expressing Dof1, a maize transcription factor that belongs to the Dof family of

Correspondence and reprint requests : Nandula Raghuram

plant-specific transcription factors known to activate the expression of several C-metabolizing genes associated with organic acid metabolism. The transgenic lines showed improved nitrogen content (by 30 %), higher levels of amino acids, better growth under low-nitrogen conditions and higher levels of mRNAs and enzyme activities for PEP carboxylase and pyruvate kinase, without any reduction of NR, GS and GOGAT transcripts. It is not clear whether Dof1 is inducible by nitrate, but the genes up-regulated by Dof1 over-expression clearly belonged to the list of known nitrate-responsive genes.

Dof proteins are unique to plants. All Dof proteins analyzed so far, except for a protein of pumpkin, recognized an AAAG motif as the essential sequence element in DNA-binding assays *in vitro* (Yanagisawa 2004). Earlier, a systematic analysis with four maize Dof proteins and randomly synthesized DNA clarified the absolute requirement of an AAAG sequence and the relatively limited influence of its flanking sequences in the Dof–DNA interaction (Yanagisawa and Schmidt, 1999). Therefore, the conserved Dof domain appears to fundamentally provide all Dof proteins with a similar DNA-binding specificity. Transgenic *Arabidopsis* plants over expressing Dof1 showed better growth under low-nitrogen conditions, apart from other broader metabolic alterations, indicating a possible role for the AAAG motif in nitrate responsive gene expression. Such a possibility has not been tested for any nitrate responsive gene so far, let alone for all of them.

GATA motifs have been mostly implicated in light-dependent gene regulation in plants. I-boxes, originally defined as GATAA sequences, and other GATA-related motifs have been found in many light-regulated genes such as the ribulose bisphosphate carboxylase small subunit, chlorophyll A/B binding protein, and glyceraldydyde-3-phosphate dehydrogenase genes (Jeong and Shih, 2003). While several lines of evidence strongly suggest a role for GATA factors in light-mediated transcriptional regulation, this has not been conclusively demonstrated. GATA factors are implicated in nitrogen response in fungi, and *in vivo* footprinting experiments in spinach have demonstrated the existence of GATA motifs in the promoter of nitrite reductase gene that were differentially regulated by ammonium (Rastogi *et al*., 1997 and references cited therein). Many, if not all nitrate responsive genes are also known to be co-regulated by light, which also implicates GATA sequences in nitrate response. GATA-binding transcriptional factors normally recognize the consensus sequence WGATAR (W = T or A; R = G or A; Lowry and Atchley, 2000), although a considerable number of DNA motifs, including their non-conserved variants, could also be considered as potential targets of plant GATA factors (Reyes *et al.*, 2004).

Therefore, it would be of interest to evaluate the role of Dof and GATA elements in nitrate responsive gene expression. However, the fact that they are known to be involved in responses other than nitrate would make them relevant to nitrate response only if they have some unique interactive effects. While it would be tedious to test such interactions experimentally, *in silico* approaches permit such analyses in all possible combinations with all nitrate-responsive genes. In fact, even though Das *et al.* (2007) ruled out the possibility of [$(a/t)_7$Ag/cTCA] motif acting as NRE on its own, its interactive effects with other possible sequence motifs flanking nitrate responsive genes have not been tested. The availability of a curated list of 625 nitrate responsive genes and their various subsets in *Arabidopsis*, and an *in silico* approach developed in-house for evaluating candidate motifs (Das *et al.*, 2007) prompted the current investigation.

MATERIALS AND METHODS

Motif search among nitrate responsive and housekeeping genes

A set of 625 nitrate responsive genes and 25 housekeeping genes compiled by Das *et al*. (2007) were selected for the investigation. Their sequences, comprising of the ORFs plus 1 kb flanking regions on either side, were extracted from the *.ptt* file of the *Arabidopsis* genome downloaded from the ftp site of NCBI. These gene sequences were subjected to pattern search for various sequence elements under study, including the known elements *viz*, Dof, GATA and Ag/cTCA, as well as novel candidate motifs predicted using various motif sampling programs (see later). Pattern search was carried out using FUZZNUC program from the EMBOSS suite, for pairwise occurrence of the sequence elements under study in all possible combinations and orientations on both strands, as described previously (Das *et al*., 2007). A similar search was also conducted for the rest of the genes present in the Arabidopsis genome for background comparison.

Search for composite motifs in 21 'early' nitrate responsive genes

As none of the individual sequence elements tested was uniquely over-represented in all the nitrate responsive

genes, the possibility of composite motifs consisting of heterologous pairs of the above three sequence elements was also tested, using the above method. However, this search was limited to the 21 'early' nitrate responsive genes as per the rationale described in Das *et al*. (2007).

Search for novel motifs

A set of 12 motif-finding algorithms were short-listed out of 50 programs available online (Table 1), on the basis of their consistency in performance and whether they were published and cited. They were used for the prediction of novel over-represented motifs in the upstream sequences of all 625 nitrate responsive genes for validation as possible candidates for NRE. The AT:GC ratio of the *Arabidopsis* genome as well as in the upstream sequences of the 25 house keeping genes used in this study were considered as background noise for this purpose. As these algorithms were designed to predict novel motifs as stand-alone sequence elements, 25 top scoring motifs were further tested by checking for their pairwise occurrence in 21 'early' nitrate responsive genes and housekeeping controls, using the FUZZNUC method described above. All the 625 possible pairwise combinations of these novel motifs, including homologous and heterologous composites, were tested.

RESULTS AND DISCUSSION

Nitrate is not only a substrate for nitrate reductase, the first enzyme of nitrate assimilation in plants, but also a signal for the regulation of over a thousand genes (Wang *et al*., 2000, 2003, 2004; Raghuram *et al*., 2006). Majority of these genes and their interactions are yet to be annotated and organized into distinct processes and pathways. The identification of nitrate response elements, if any, that could mediate nitrate-responsive gene expression, offers an excellent starting point to delineate the nitrate signaling mechanism. Earlier efforts that identified the motif [$(a/t)_7$Ag/cTCA] as a NRE were based on promoter deletion experiments with the upstream sequences of one or two genes in one or two plants (Hwang *et al*., 1997; Warning and Hachtel, 2000). However, bioinformatic searches using this element in the sequences flanking 625 nitrate-responsive genes of *Arabidopsis* and 283 of their rice homologs has revealed that on its own, the above motif is neither unique to nitrate responsive genes nor common to all of them. The only other sequence elements that were implicated in nitrate response were the Dof-binding element (AAAG) (Yanagisawa *et al*., 2004) and GATA

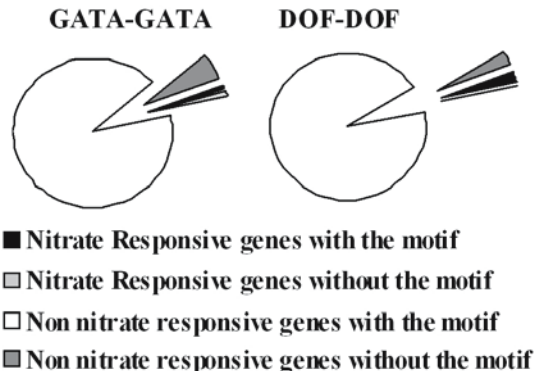

Fig. 1. Genome wide pair wise occurrence of candidate

element (WGATAR) (Rastogi *et al*., 1997), although they were already known to be involved in other responses. The present study sought to test them bioinformatically using the approach of Das *et al.* (2007) on 625 genes of *Arabidopsis*. A pair-wise search for the presence of Dof or GATA sequence elements in their promoter regions revealed their occurrence in ~ 90 % and 50 % of the 625 nitrate-responsive genes examined (Figure 1). However, these elements were also found widely in the rest of the *Arabidopsis* genome, confirming that on their own, these elements are neither specific to nitrate response, nor common to all the nitrate responsive genes.

Since nitrate responsive genes are not a homogeneous group and include smaller subsets of genes that express only in some tissues (eg. root/shoot) or at some times (early/late response) or may belong to secondary responses (Das *et al*., 2007), we needed to factor these differences into our analysis. Moreover, the poor specificity of occurrence of stand-alone motifs could not rule out the possibility that the above 3 motifs may act in heterologous combinations, as composite motifs are known in plants (Kato *et al*., 2004). Therefore, their composite occurrence was tested in various combinations on 21 'early' nitrate responsive genes and 25 housekeeping controls.

The data shown in Fig. 2 reveal that none of the above three motifs were present in all the 21 early nitrate responsive genes in either homologous or heterologous combinations. The role of homologous pairs of Ag/cTCA elements was already ruled out by Das *et al.* (2007), and our data in Fig. 1 has already shown that the homologous GATA-GATA pair and the Dof-Dof pair is not specific or unique to nitrate response, as they also occur widely elsewhere in the genome. Therefore, even though the Dof-Dof pair occurred in 18 out of the 21 genes examined, it is difficult to attach much

Table 1. List of short listed motif samplers

Motif Sampler	Search Method	Availability	Reference
Gibbs Sampler	Gibbs	http://bayesweb.wadsworth.org/gibbs/gibbs.html	Thompson *et al*. (2003)
MEME	EM	http://meme.sdsc.edu/meme4_1/intro.html	Bailey and Elkan (1994)
AlignACE	Gibbs	http://atlas.med.harvard.edu/cgi-bin/alignace.pl	Hughes *et al.* (2000)
Ann-Spec	Gibbs	Unix-Source code	Workman and Stormo (2000)
YMF	String Enum	http://abstract.cs.washington.edu/~saurabh/YMFWebRSH/YMFInput.pl	Sinha and Tompa (2000)
ELPH	Gibbs	Unix-Sourcecode	http://www.cbcb.umd.edu/software/ELPH
Co-Bind	Gibbs	Unix-Sourcecode	Guha and Stormo (2001)
Weeder	Enum,PD	Unix-Source code	Pavesi *et al.* (2004)
Mitra	Prefix/Graph	http://www.ccls.columbia.edu/compbio/mitra/	Eleazar and Pevzner (2002)
DWE	Enum,PD	Unix-Source code	Sumazin *et al*. (2005)
A-GLAM		Unix-Source Code	Tharakaraman *et al.* (2005)
Wordspy	Enum	http://cic.cs.wustl.edu/wordspy/	Wang *et al.* (2005)

significance to it. Even the heterologous Dof -GATA, Dof- Ag/cTCA, GATA- Ag/cTCA pairs did not show substantial overrepresentation in the 21 'early' nitrate responsive genes as compared to their housekeeping controls (Fig. 2). Thus, none of the known motifs that were previously implicated in nitrate response could qualify to be NREs based on the criterion of specific, pairwise occurence, even in a small subset of the 625 nitrate responsive genes in *Arabidopsis*. Recent reports concerning some transcription factors implicated in nitrate response such as ANR1 (Gan *et al*., 2005) and NLP7 (Krap *et al.* 2009) have not yet revealed any sequence elements through which they may mediate nitrate-responsive gene expression, and were therefore not factored into our analysis. Otherwise, our results conclusively demonstrate the need for a fresh search for novel candidate motifs common to all nitrate responsive genes that may qualify to be NREs.

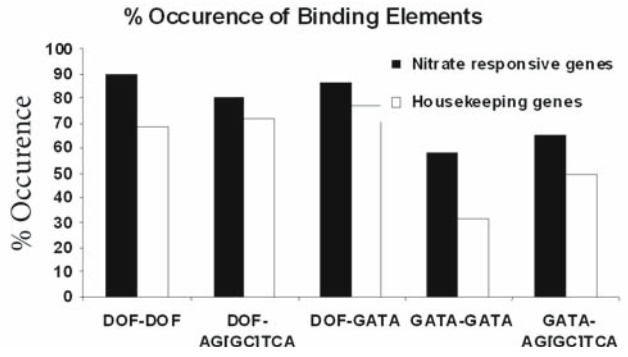

Fig. 2. Occurrence of combinatorial motifs in early nitrate responsive genes

The search for new candidate motifs was carried out in the upstream region of the 21 early nitrate responsive genes, using 12 motif sampling algorithms available online (Table 1), based on their novelty, robustness and flexibility to scan different genomes. The presence of repeat sequences was eliminated from the set to reduce false positives. As each of these motif samplers reported different sequence motifs and each sampler reported multiple motifs, only the top scoring sequence motifs shown in Fig. 3 have been chosen for further analysis. All these motif samplers reported single copies of sequence motifs, whereas our approach defined a response element as having at least two copies of the motif. Accordingly, the top scoring motifs reported in Fig. 3 were reanalyzed for their pairwise occurrence (homologous as well as heterologous) in the 21 early nitrate responsive genes and the 25 housekeeping genes using the FUZZNUC method. However, none of these combinations were found to be over-represented or specific with regard to nitrate responsive genes (data available on request).

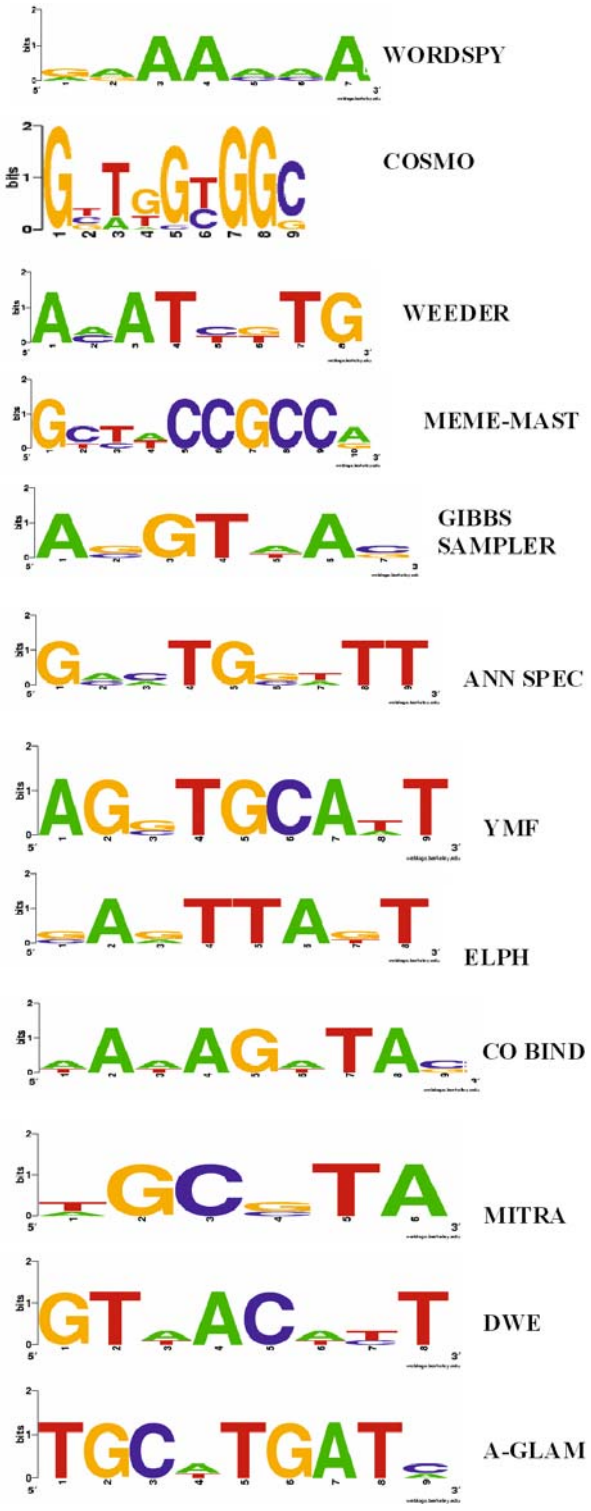

Fig. 3. Candidate motifs generated by various motif samplers, from upstream regions of 21 early nitrate sequenosite genes. None of these was found to qualify as NRE in subsequent analysis (data not shown)

In conclusion, our comprehensive bioinformatic analysis negates the possibility that any of the previously reported motifs in Arabidopsis, as well as novel motifs predicted by most of the motif samplers, qualify as nitrate responsive elements. We are currently extending this approach to analyse the flanking sequences of nitrate responsive genes identified recently in rice by microarray analyses in our laboratory (unpublished data). Such studies help to avoid wasteful experiments that rely on poorly characterized motifs, as well as to narrow down the number of possibilities that can be tested experimentally.

ACKNOWLEDGEMENTS

This work was supported in part by research grants from DST (No. SP/SO/BB-27/2002) and CSIR (No. 60/ (0056)/02/EMR-II) awarded to N. Raghuram and fellowship provided to RRP. The authors wish to thank Sunila and Navjyoti Chakraborty for their assistance in proof reading of the manuscript.

REFERENCES

Bailey TL and Elkan C (1994) "Fitting a mixture model by expectation maximization to discover motifs in biopolymers", *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, AAAI Press, Menlo Park, California, pp. 28-36.

Das SK, Pathak RR, Choudhury D and Raghuram N (2007). Genomewide computational analysis of nitrate response elements in rice and Arabidopsis. *Mol. Genet. Genomics* 278: 19–525

Dorbe MF, Caboche M and Daniel-Vedele F (1992). The tomato *Nia* gene complements a *N. plumbaginifolia* nitrate reductase-deficient mutant and is properly regulated. *Plant Mol. Biol.* 18: 63–375

Eleazar E, Pevzner PA. (2002) "Finding Composite Regulatory Patterns in DNA Sequences.'' *Bioinformatics* July 18, Supplement 1:S354-63.

Gan Y, Filleur S, Rahman A, Gotensparre S and Forde BG (2005). Nutritional regulation of ANR1 and other root-expressed MADS-box genes in Arabidopsis thaliana. *Planta* 222: 730–742

Guha TD and Stormo GD (2001). Identifying target sites for cooperatively binding factors. *Bioinformatics.* 17, 608-621

Hughes JD, Estep PW, Tavazoie S, and Church GM (2000) Computational identification of *cis*-regulatory elements associated with groups of functionally related genes in Saccharomyces cerevisiae. *J Mol Biol*. 296 (5):1205-14

Hwang CF, Lin Y, D'souza T and Cheng CL (1997). Sequences necessary for nitrate-dependent transcription of *Arabidopsis* nitrate reductase genes. *Plant Physiol.* 113: 853–862

Jeong MJ and Shih MC (2003). Interaction of a GATA factor with cis-acting elements involved in light regulation of nuclear genes encoding chloroplast glyceraldehyde-3-phosphate dehydrogenase in Arabidopsis. *Biochem. Biophys. Res. Commun.* 300: 555–562

Kato M, Hata N, Banerjee N, Futcher B and Zhang MQ (2004). Identifying combinatorial regulation of transcription factors and binding motifs. *Genome Biol.* 5(8):R56.

Lin Y, Hwang CF, Brown JB and Cheng CL (1994). 5' proximal regions of *Arabidopsis* nitrate reductase genes direct nitrate-induced transcription in transgenic tobacco. *Plant Physiol.* 106: 477–484

Neininger A, Bichler J, Schneiderbauer A and Mohr H (1993). Response of a nitrite reductase 3.1-kilobase upstream regulatory sequence from spinach to nitrate and light in transgenic tobacco. *Planta* 189: 440–442

Pavesi G, Mereghetti P, Mauri G, Pesoli G (2004). Weeder Web: Discovery of transcription factor binding sites in a set of sequences from co-regulated genes. *Nucleic Acids Res.* 32: W 199–W203.

Raghuram N, Pathak RR and Sharma P (2006). Signalling and the molecular aspects of N-use efficiency in higher plants. In: RP Singh PK Jaiwal (eds), *Biotechnological approaches to improve nitrogen use efficiency in plants.* Studium Press LLC, Houston, pp 19–40

Rastogi R, Back E, Schneiderbauer A, Bowsher C, Moffatt B and Rothstein SJ (1993). A 330 bp region of the spinach nitrite reductase gene promoter directs nitrate-inducible tissue specific expression in transgenic tobacco. *Plant J.* 4: 317–326

Rastogi R, Bate N, Sivasankar S and Rothstein SJ (1997). Footprinting of the spinach nitrite reductase gene promoter reveals the preservation of nitrate regulatory elements between fungi and higher plants. *Plant Mol. Biol.* 34: 465-476.

Reyes JC, Muro-Pastor MI and Florencio FJ (2004). The GATA family of transcription factors in *Arabidopsis* and rice. *Plant Physiol.* 134:1718–1732

Scheible WR, Morcuende R, Czechowski T, Fritz C, Osuna D, Palacios-Rojas N, Schindelasch D, Thimm O, Udvardi MK and Stitt M (2004). 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.* 136: 2483–2499

Sinha S and Tompa M (2000). A Statistical Method for Finding Transcription Factor Binding Sites, *Eighth International Conference on Intelligent Systems for Molecular Biology*, San Diego, CA, 344-35

Sumazin P, Chen G, Hata N, Smith AD, Zhang T, and Zhang MQ (2005) DWE: discriminating word enumerator. *Bioinformatics*; 21:31–38.

Tharakaraman K, Mariño-Ramírez L, Sheetlin S, Landsman D, and Spouge JL (2005) Alignments anchored on genomic landmarks can aid in the identification of regulatory elements. *Bioinformatics*, 21, i440–i448.

Thompson W, Rouchka EC, and Lawrence CE (2003) Gibbs Recursive Sampler: finding transcription factor binding sites. *Nucleic Acids Res*. 31(13): 3580-3585.

Vaucheret H, Marion-Poll A, Meyer C, Faure JD, Marine E and Caboche M (1992). Interest in and limits to the utilisation of reporter genes for the analysis of transcriptional regulation of nitrate reductase. *Mol. Gen. Genet*. 235: 259–268

Vincentz M, Moureaux T, Leydecker MT, Vaucheret H and Caboche M (1993). Regulation of nitrate and nitrite reductase expression in *Nicotiana plumbaginifolia* leaves by nitrogen and carbon metabolites. *Plant J.* 3(2):315–324

Yanagisawa S, Akiyama A, Kisaka H, Uchimiya H and Miwa T (2004). Metabolic engineering with Dof1 transcription factor in plants: improved nitrogen assimilation and growth under low-nitrogen conditions. *Proc. Natl. Acad. Sci. USA*, 101: 7833–7838.

Yanagisawa S (2004). Dof domain proteins: plant-specific transcription factors associated with diverse phenomena unique to plants. *Plant Cell Physiol*. 45(4):386-391

Yanagisawa S and Schmidt RJ (1999). Diversity and similarity among recognition sequences of Dof transcription factors. *Plant J*. 17: 209–214

Wang, G., Yu, T., Zhang, W. (2005) WordSpy: identifying transcription factor binding motifs by building a dictionary and learning a grammar *Nucleic Acids Res*, . 33, W412–W416

Wang R, Guegler K, LaBrie ST and Crawford NM (2000). Genomic analysis of a nutrient response in *Arabidopsis* reveals diverse expression patterns and novel metabolic and potential regulatory genes induced by nitrate. *Plant Cell* 12:1491–1509

Wang YH, Garvin DF and Kochian LV (2001). Nitrate-induced genes in tomato roots: Array analysis reveals novel genes that may play a role in nitrogen nutrition. *Plant Physiol.* 127: 345–359

Wang X, Wu P, Xia M, Wu Z, Chen Q and Liu F (2002). Identification of genes enriched in rice roots of the local nitrate treatment and their expression patters in split-root treatment. *Gene* 297: 93–102

Wang R, Okamoto M, Xing X and Crawford NM (2003). Microarray analysis of the nitrate response in *Arabidopsis* roots and shoots reveals over 1,000 rapidly responding genes and new linkages to glucose, trehalose-6-phosphate, iron and sulfate metabolism. *Plant Physiol.* 132:556–567

Wang R, Tischner R, Gutiérrez RA, Hoffman M, Xing X, Chen M, Coruzzi G and Crawford NM (2004). Genomic analysis of the nitrate response using a nitrate reductase-null mutant of *Arabidopsis*. *Plant Physiol*. 136: 2512–2522

Workman C and Stormo GD (2000). ANN-Spec: A method for discovering transcription factor binding sites with improved specificity. Proc. Pacific Symposium on Biocomputing.

Zhang H and Forde BG (1998). An Arabidopsis MADS box gene that controls nutrient-induced changes in root architecture. *Science* 279: 407–409