

Sequence analysis

In silico analysis of promoter regions from cold-induced genes in rice (*Oryza sativa* L.) and *Arabidopsis thaliana* reveals the importance of combinatorial controlAngelica Lindlöf^{1,*}, Marcus Bräutigam², Aakash Chawade³, Olof Olsson³
and Björn Olsson¹¹Systems Biology Research Centre, School of Life Sciences, University of Skövde, Box 408, 541 28 Skövde,²Department of Cell and Molecular biology, Medicinaregatan 9C, Box 462 and ³Department of Plant and Environmental Sciences, Box 461, Gothenburg University, SE405 30 Göteborg, Sweden

Received on October 27, 2008; revised on March 04, 2009; accepted on March 21, 2009

Advance Access publication March 25, 2009

Associate Editor: Limsoon Wong

ABSTRACT

Motivation: Cold acclimation involves a number of different cellular processes that together increase the freezing tolerance of an organism. The DREB1/CBFs are transcription factors (TFs) that are prominent in the regulation of cold responses in *Arabidopsis thaliana*, rice and many other crops. We investigated if the expression of DREB1/CBFs and co-expressed genes relies on combinatorial control by several TFs. Our results support this notion and indicate that methods for studying the regulation of complex cellular processes should include identification of combinations of motifs, in addition to searching for individual overrepresented binding sites.

Contact: angelica.lindlof@his.se

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

The productivity and growth of many plant species is severely limited by low temperatures (Mahajan and Tuteja, 2005). Cold stress, which includes chilling (0–12°C) and freezing (<0°C) temperatures, adversely affects crop yields by restraining sowing time, causing tissue damage and stunted growth. In order to overcome such limitations and increase the yields for agronomically important crops an improved tolerance to cold stress is desired.

Plants growing in temperate regions have evolved traits that make it possible for them to cope with freezing temperatures. In response to mild cold stress, ~4–6°C, a cascade of genetic reactions are triggered that greatly enhance the tolerance to later, more severe sub-zero temperatures. This ability to cope with low temperatures is known as cold acclimation (Guy, 1999). Moreover, this is a complex trait, involving numerous genes that interact in an intricate regulatory network. The activation and de-activation of genes participating in signaling pathways is achieved by several mechanisms, including the binding of transcription factors (TFs) to *cis*-elements in the promoter region of affected genes, which thereby become activated or suppressed.

Arabidopsis thaliana and rice (*Oryza sativa* L.) are two major model species for plants with apparent differences in overall physiology, adaptation to temperature variation and cold tolerance. *Arabidopsis thaliana* is a dicot with an ability to cold acclimate, whereas rice is a monocot that is susceptible to cold. There are three factors in *A.thaliana* that are prominent in the regulation of cold responses in this species: the AtCBF1-3 (*A.thaliana* C-repeat/dehydration responsive element-binding factor) (At4g25490, At4g25470 and At4g25480), which are all highly induced within the first hours of the response (Gilmour *et al.*, 1998). In rice, two of the identified orthologs, OsDREB1A-B (*O.sativa* dehydration-responsive element-binding factor) (Os09g35010 and Os09g35030), have been shown to be induced by cold stress, but also increase the tolerance to high-salt and drought when being overexpressed (Dubouzet *et al.*, 2003; Ito *et al.*, 2006).

Since it has been established that the expression of many genes rely on the combinatorial control of several TFs (Remenyi *et al.*, 2004; Singh, 1998), we wanted to investigate if this could be the case for the CBF/DREB factors. While conducting this study, we made observations that support the importance of combinatorial control of the response to cold stress. Consequently, we argue that identifying combinations of motifs is a necessary complement to searching for individual overrepresented binding sites when analyzing complex processes such as responses to cold stress.

2 METHODS

We have previously conducted a whole-genome expression profiling on chilling-susceptible rice, *O.sativa*, cv. *Indica v Jumla Marshi* (M.Bräutigam *et al.*, submitted for publication), where we identified 1450 genes as differentially expressed during cold stress. Rice seedlings were stressed at 4°C for 30 min, 2, 4, 8 and 24 h, and samples were thereafter hybridized to the Affymetrix GeneChip® Rice Genome Array. Additionally, data from a similar experiment for *A.thaliana* (D'Angelo *et al.*, 2005) was downloaded and analyzed in the same way. In this second dataset, 1753 genes were identified as differentially expressed. In more detail, probes were considered differentially expressed if their signal value was at least 3-fold changes different and had a Benjamini–Hochberg adjusted *P*-value of ≤0.05 in at least one time step. Genes corresponding to these probes

*To whom correspondence should be addressed.

were identified by matching each probe sequence to a *Japonica* and *A.thaliana* locus, respectively; through a BLASTn search against TIGR Rice (*cv. Japonica*) 5.0 pseudomolecules (TIGR pseudomolecules, ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/) and TAIR *A.thaliana* 7.0 cds sequences, (TAIR cds sequences, ftp://ftp.arabidopsis.org/home/tair/Sequences/blast_datasets/TAIR7_blastsets/) respectively. The full-length sequence of each probe was used, which was downloaded from Affymetrix (Affymetrix, www.affymetrix.com) web site. For genes having a significant match against several probes only the one with the lowest *E*-value was kept. The promoter regions, 1Kb upstream, of the identified cold-responsive genes in rice and *A.thaliana* were thereafter downloaded from TIGR (TIGR pseudomolecules, ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/) and TAIR (TAIR cds sequences, ftp://ftp.arabidopsis.org/home/tair/Sequences/blast_datasets/TAIR7_blastsets/), respectively.

In order to derive *cis*-elements that relate to TFs that are plausibly important in the regulation of the CBF/DREB factors, we looked at motifs present in genes that are co-expressed with the CBF/DREBs. One cluster per CBF/DREB factor was derived by identifying genes having highly similar expression profiles to one of the CBF/DREBs. This was done by calculating the Pearson correlation (PC) between each CBF/DREB and all other cold-responsive genes in each species. Only genes that had PC ≥ 0.95 with at least one CBF/DREB were included in the cluster.

In order to extract regulatory motifs that are overrepresented among genes in a cluster, we used a database of previously characterized motifs and searched for occurrences of these motifs in the promoter regions of the genes. The database was a merge of motifs from the PLACE (Higo et al., 1999) and plantCARE (Lescot et al., 2002) databases as well as motifs extracted from the literature. We choose to use consensus motifs, since this provided us with a high coverage of previously characterized plant motifs (in total 946 motifs). Thereafter, subsequences matching the motifs were searched for in both the forward and reverse strand of the promoter regions of the genes, by using an in-house developed Perl script, and each such match was considered as an occurrence of the motif. Only perfect matches to the motif were considered, i.e. the search did not include a scoring function. Additionally, we only regarded motifs with >4 IUPAC letters.

To identify significantly overrepresented individual motifs, we calculated how many genes in the cluster had $\geq T$ ($T=1, 2, \dots, 25$) occurrences of a motif *m* and the number that had <*T* occurrences of the same motif. The corresponding counts were calculated for all other genes in the genome. Fisher's exact one-sided test was then applied for detecting significantly overrepresented individual motifs in each cluster, using these counts. Overrepresented motif combinations were derived by calculating the number of genes having a combination of the overrepresented individual motifs. For example, for one of the clusters the motifs MACGYGB and VCGCGB had to be present at least eight and six times, respectively, in the promoter region of a gene in order to be considered as significantly overrepresented. Based on these thresholds, we then derived all significantly overrepresented combinations of two, three and four different motifs (each with the required number of occurrences), using Fisher's exact one-sided test ($P \leq 0.01$). In addition, we investigated significant motif combinations using a Benjamini–Hochberg false discovery rate (BH FDR) adjusted *P*-value threshold of ≤ 0.01 (Benjamini and Hochberg, 1995).

3 RESULTS

The clustering procedure resulted in one cluster for each of the CBF/DREB factors. The clusters are somewhat overlapping, since the CBFs have highly similar expression profiles. Moreover, the number of detected individual overrepresented motifs (IOMs) differed extensively, since the number of genes in each cluster differed, and ranged from 15 to 58 detected motifs.

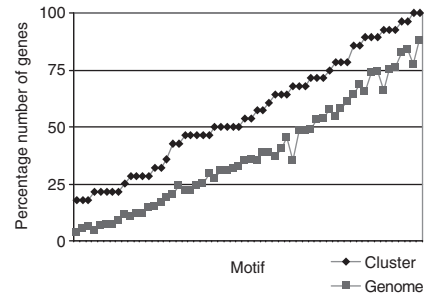


Fig. 1. Relative frequency of motif occurrence. Percentage number of genes in the *AtCBF1* cluster and the remaining genome, respectively, having at least *T* occurrence(s) of a motif. The y-axis indicates percentages and the x-axis represents the 58 significantly overrepresented ($P \leq 0.01$) motifs that were identified in this cluster.

When studying the presence of each IOM in more detail, it was revealed that motifs that are relatively common among genes in a cluster are also common among genes in the remaining genome, i.e. also in genes that are not cold-responsive. We calculated the relative occurrence fold-change (FC) of each motif as

$$FC(m) = \frac{x_1/(x_1 + x_2)}{y_1/(y_1 + y_2)} \quad (1)$$

where *m* is the motif in consideration, x_1 is the number of genes in a cluster having at least *T* occurrences of the motif, x_2 the number of genes in the cluster having <*T* occurrences of the motif, y_1 the number of genes in the remaining genome having at least *T* occurrences of the motif and y_2 the number of genes in the remaining genome having <*T* occurrences of the motif. We observed an overall strong correlation between the relative frequency in a cluster and the remaining genome (PC = 0.95–0.99 for all clusters), i.e. if a motif was commonly occurring among the genes in a cluster it was in most cases also common in the remaining genome, and vice versa (Fig. 1).

Regarding motif combinations, we derived all significantly overrepresented two-, three- and four-combinations and made observations supporting the importance of combinatorial control. In general, the *P*-value for the detected motif combinations decreased as the complexity of the motif combinations increased, i.e. the median *P*-value of the significantly overrepresented four-combinations was much lower than for the two-combinations (Fig. 2). This is explained by the fact that the number of genes having a significant combination decreased with increasing order of the combination to a greater extent among the genes in the remaining genome than in the cluster (compare middle boxplots with rightmost boxplots in Fig. 2). For example, using the cluster exemplified in Figure 2, the median number of genes having a combination decreased by 30% when comparing two- and four-combinations (from 10 to 7 genes), whereas in the remaining genome the decrease was 42% (from 3891 to 1647 genes). In many cases, a specific combination of four motifs frequently occurs among the genes in a specific cluster, but is totally absent among the genes in the remaining genome (w.r.t. *T*, in some cases a motif combination occurs in the remaining genome, but in those cases with a smaller value on *T* for one or several of the motifs included in the combination). These results indicate that a specific combination of TFs is important for the regulation.

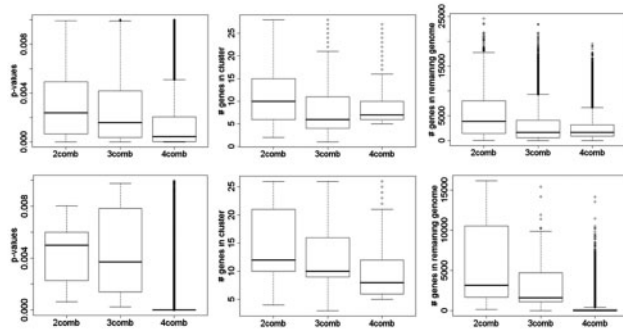


Fig. 2. Motif combinations. The results of significantly overrepresented motif combinations (SOMC) are shown for one cluster (*AtCBF2*). Lower row shows SOMC with FDR-adjusted $P \leq 0.01$ and upper row with standard $P \leq 0.01$. The left box-plots show the P -values, the middle box-plots show the number of genes in the cluster having a SOMC and the right box-plots show the number of genes in the remaining genome having a SOMC.

We further analyzed those significant four-combinations having $P \leq 0.01$ (BH FDR-adjusted), in order to investigate if the motifs occurring in these combinations had previously been coupled to cold stress. Applying the BH FDR procedure decreased the number of significant combinations, for example the number of four-combinations decreased with 55–88%. Additionally, to avoid the redundancy and the combinatorial complexity of higher order combinations, we considered only the most significant non-redundant, non-overlapping motif (MSNM) combinations from each cluster. Regarding *A.thaliana*, the MSNM four-combinations for the *AtCBF1* cluster were combinations of six different motifs, and for the *AtCBF2* and *AtCBF3* clusters they were combinations of six and seven motifs, respectively (Table 1 and Supplementary Table S1). For the *OsDREB1A* and *OsDREB1B* clusters, the MSNM four-combinations were based on seven and eight motifs, respectively.

In the list of motifs, the ABRE-related MACGYGB motif is distinguished, as it occurs in all but one cluster. This motif has previously been identified among genes responsive to cytosolic Ca^{2+} , which is a major secondary messenger for triggering cold acclimation signaling pathways (Kaplan *et al.*, 2006). Furthermore, the GT-1 binding site is represented in two of the *A.thaliana* clusters and one of the rice clusters. This motif has been found in many light-regulated genes in various plants, such as oat and rice. Additionally, GT-1 is activated by a calcium-dependent phosphorylation in response to light (Marechal *et al.*, 1999). Since cold stress peaks during the night it also coincides with light stress, which plausibly explains the presence of the GT-1 binding site.

In two of the *A.thaliana* clusters W-box and/or WRKY binding sites are represented, to which WRKY TFs bind (Eulgem *et al.*, 2000), but not in any of the rice clusters. Many of these TFs are rapidly and transiently induced by various abiotic stresses, including cold stress. However, WRKY motifs also occur frequently in the rice genome, but because of this ‘general’ occurrence, they are not detected as overrepresented individual motifs and thus not included in the identified combinations.

In the *A.thaliana* clusters, there are several different TATA-boxes present in the combinations, whereas this is not as common in rice. There are also other AT-rich motifs in the *A.thaliana* clusters, e.g. the POLASIG3 motif. Moreover, the AT-composition of the promoter

Table 1. Overrepresented four motif combinations

Cluster	Name	Motif	<i>T</i>
AtCBF1	GT-1 binding	GRWAAW	7
	TATA-box	TAATA	2
	TATA-box	TATAA	5
	ABRE-related	MACGYGB	1
	POLASIG3	AATAAT	1
	ROOTMOTIFTAPOX1	ATATT	3
AtCBF2	GT-1 binding	GRWAAW	6
	HSF (heat shock factor)	NTTCNNGAAN	4
	TATA-box	TTTTA	5
	TATA-box	TATATA	1
	W-box	TTGAC	1
	ABRE-related	MACGYGB	1
AtCBF3	WRKY binding	YTGACY	1
	CAAT-box	ATTGG	2
	TATA-box	ATATAT	1
	ABRE-related	MACGYGB	1
	TATA-box	TTATTT	2
	W-box	TTGAC	1
OsDREB1A	TATA-box	TAATA	3
	LTRE	CCGAC	1
	Unnamed	CGTGG	2
	bZIP binding	ACACNNG	3
	Root-hair-specific	KCACGW	2
	Light responsive	CCRCCC	1
OsDREB1B	Sugar starvation	CGACG	1
	MADS binding	CW ₈ G	1
	Pollen-related	AGAAA	3
	TATA-box	TATAA	1
	SEF3 binding	AACCCA	1
	Sugar starvation	CGACG	2
	E2 binding	WTTSSCSS	1
	REalpha	AACCAA	1
	GT-1 binding	GRWAAW	3
	ABRE-related	MACGYGB	1

Motifs included in the most significant non-redundant, non-overlapping motif (MSNM) four combinations found in each cluster. Cluster, the cluster; Name, a descriptive name of the motif; Motif, the sequence pattern of the regulatory motif; *T*, the lowest number of occurrences for which the motif is regarded as statistically overrepresented (BH FDR adjusted Fisher’s exact one-sided test, $P < 0.01$).

sequences for the *A.thaliana* genes is $68.6 \pm 3.9\%$, whereas in the rice promoters it is $54.4 \pm 6.8\%$. Whether this has any major impact on the regulation of the acclimation pathways remains to be elucidated.

Several of the motifs in rice have been shown to be important in ABA and/or light stress signaling pathways, such as the bZIP, light responsive and REalpha motifs (Baker *et al.*, 1994; Degenhardt and Tobin, 1996; Finkelstein and Lynch, 2000). Moreover, the LTRE motif CCGAC present in one of the rice clusters is overlapping with the DRE motif RCCGAC, which is actually the target of the CBF/DREB factors (Stockinger *et al.*, 1997). This motif has also been coupled to drought and light signaling, amongst others, besides cold (Kim *et al.*, 2002).

4 DISCUSSION

Biologically, cold acclimation is a complex trait that involves the up- and down-regulation of numerous genes. In addition, a portion of

these genes also encode activities involved in protecting the cell from related stresses such as drought, light, salt and wounding (Fujita, et al., 2006). This complicates the elucidation of signaling pathways directly coupled to an increased cold tolerance.

As previously mentioned, activation and deactivation of genes participating in a specific biological process is achieved, amongst others, by the binding of TFs to *cis*-elements in the promoter region of affected genes. In addition, the coordination of multiple TFs in a combinatorial fashion makes it possible for the plant to control the expression of genes in response to a variety of environmental stimuli by using a limited number of TFs. Several studies address the importance of considering combinatorial control in order to understand the underlying regulatory network (Beer and Tavazoie, 2004; Chawade, et al., 2007; Hannehalli and Levy, 2002; Kato, et al., 2004; Pilpel, et al., 2001; Zhou and Wong, 2004). Our results further support this notion and we argue that the focus should therefore be on identifying motif combinations when studying complex processes such as cold acclimation.

For *A.thaliana*, the results indicate that a combination of an ABRE-related, GT-1, WRKY and AT-rich motif is important in the regulation of the cold response. Interesting is also that some of the previously identified motifs coupled to cold stress, such as the ICER3 and MYB15 motifs (Agarwal et al., 2006; Benedict et al., 2006), are present in the *A.thaliana* combinations but not in rice (data not shown). The derived combinations in rice indicate instead on a coupling to ABA, drought and/or light stress signaling pathways. In addition, in our microarray experiment the *OsDREB1s* are differentially expressed to a much lower level than the *AtCBFs* (M.Bräutigam et al., submitted for publication; data not shown), which further indicates an overall difference in response patterns. The *OsDREB1s* have previously been shown to not only enhance tolerance to cold, but also drought and high-salt, indicating that they are more universal TFs responding to various stresses (Dubouzet, et al., 2003). This is plausibly reflected in the identified motif combinations as well as the observed difference in the overall response patterns in the microarray experiments.

Funding: Swedish Research School in Genomics and Bioinformatics; Swedish Research Council (VR).

Conflict of Interest: none declared.

REFERENCES

- Agarwal,M. et al. (2006) A R2R3 type MYB transcription factor is involved in the cold regulation of *cbf* genes and in acquired freezing tolerance. *J. Biol. Chem.*, **281**, 37636–37645.
- Baker,S.S. et al. (1994) The 5'-region of *Arabidopsis thaliana* *cor15a* has *cis*-acting elements that confer cold-, drought- and ABA-regulated gene expression. *Plant Mol. Biol.*, **5**, 701–713.
- Beer,M.A. and Tavazoie,S. (2004) Predicting gene expression from sequence. *Cell*, **2**, 185–198.
- Benedict,C. et al. (2006) Consensus by democracy. Using meta-analyses of microarray and genomic data to model the cold acclimation signaling pathway in *Arabidopsis*. *Plant Phys.*, **141**, 1219–1232.
- Benjamini,Y. and Hochberg,Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B.*, **57**, 289–300.
- Chawade,A. et al. (2007) Putative cold acclimation pathways in *Arabidopsis thaliana* identified by a combined analysis of mRNA co-expression patterns, promoter motifs and transcription factors. *BMC Genomics*, **8**, 304.
- D'Angelo,C. et al. (2005) Cold stress time course. *AtGenExpress*. Available at <http://www.arabidopsis.org/info/expression/ATGenExpress.jsp> (last accessed date 1 April, 2009).
- Degenhardt,J. and Tobin,E.M. (1996) A DNA binding activity for one of two closely defined phytochrome regulatory elements in an *Lhcb* promoter is more abundant in etiolated than in green plants. *Plant Cell*, **1**, 31–41.
- Dubouzet,J.G. et al. (2003) *OsDREB* genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. *Plant J.*, **4**, 751–763.
- Eulgem,T. et al. (2000) The WRKY superfamily of plant transcription factors. *Trends Plant Sci.*, **5**, 199–206.
- Finkelstein,R.R. and Lynch,T.J. (2000) The *Arabidopsis* abscisic acid response gene *ABI5* encodes a basic leucine zipper transcription factor. *Plant Cell*, **4**, 599–609.
- Fujita,M. et al. (2006) Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. *Curr. Opin. Plant Biol.*, **9**, 436–442.
- Gilmour,S.J. et al. (1998) Low temperature regulation of the *Arabidopsis* CBF family of AP2 transcriptional activators as an early step in cold-induced *COR* gene expression. *Plant J.*, **16**, 433–442.
- Guy,C. (1999) Molecular responses of plants to cold shock and cold acclimation. *J. Mol. Microbiol. Biotechnol.*, **1**, 231–242.
- Hannehalli,S. and Levy,S. (2002) Predicting transcription factor synergism. *Nucleic Acids Res.*, **30**, 4278–4284.
- Higo,K. et al. (1999) PLACE: a database of plant *cis*-acting regulatory DNA elements. *Nucleic Acids Res.*, **26**, 358–359.
- Ito,Y. et al. (2006) Functional analysis of rice *DREB1/CBF*-type transcription factors involved in cold-responsive gene expression in transgenic rice. *Plant Cell Physiol.*, **47**, 141–153.
- Kaplan,B. et al. (2006) Rapid transcriptome changes induced by cytosolic Ca²⁺ transients reveal ABRE-related sequences as Ca²⁺-responsive *cis* elements in *Arabidopsis*. *Plant Cell*, **18**, 2733–2748.
- Kato,M. et al. (2004) Identifying combinatorial regulation of transcription factors and binding motifs. *Genome Biol.*, **5**, R56.
- Kim,H.J. et al. (2002) Light signalling mediated by phytochrome plays an important role in cold-induced gene expression through the C-repeat/dehydration responsive element (C/DRE) in *Arabidopsis thaliana*. *Plant J.*, **6**, 693–704.
- Lescot,M. et al. (2002) PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. *Nucleic Acids Res.*, **30**, 325–327.
- Mahajan,S. and Tuteja,N. (2005) Cold, salinity and drought stresses: an overview. *Arch. Biochem. Biophys.*, **444**, 139–158.
- Marechal,E. et al. (1999) Modulation of GT-1 DNA-binding activity by calcium-dependent phosphorylation. *Plant Mol. Biol.*, **40**, 373–386.
- Pilpel,Y. et al. (2001) Identifying regulatory networks by combinatorial analysis of promoter elements. *Nat. Genet.*, **29**, 153–159.
- Remenyi,A. et al. (2004) Combinatorial control of gene expression. *Nat. Struct. Mol. Biol.*, **11**, 812–815.
- Singh,K.B. (1998) Transcriptional regulation in plants: the importance of combinatorial control. *Plant Physiol.*, **118**, 1111–1120.
- Stockinger,E.J. et al. (1997) *Arabidopsis thaliana* *CBF1* encodes an AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE, a *cis*-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. *Proc. Natl Acad. Sci. USA.*, **29**, 1035–1040.
- Zhou,Q. and Wong,W.H. (2004) CisModule: de novo discovery of *cis*-regulatory modules by hierarchical mixture modeling. *Proc. Natl Acad. Sci.*, **33**, 12114–12119.