The productivity and growth of many plant species is severely impaired during response to mild cold stress, making it possible for them to cope with freezing temperatures. Important crops, like rice, usually have limited tolerance to 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 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 have 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.Bru¨ntigam 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
were identified by matching each probe sequence to a *Arabidopsis* and *A.thaliana* locus, respectively, through a BLASTn search against TIGR Rice (cv. *Arabidopsis*) 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, 1 Kb 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 ≥ T 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 MACOG8B and VGCGB were 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 ≤ 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 ACRBF1 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 ≤ 0.01) motifs that were identified in this cluster.](image)

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

\[
\text{FC}(m) = \frac{x_1/(x_1 + x_2)}{y_1/(y_1 + y_2)}
\]

where m is the motif in consideration, x1 is the number of genes in a cluster having at least T occurrences of the motif, x2 the number of genes in the cluster having < T occurrences of the motif, y1 the number of genes in the remaining genome having at least T occurrences of the motif and y2 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.
Combinatorial control of cold-induced genes

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 Arabidopsis, 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).

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 | $T$ |
|---------|------|-------|-----|
| AtCBF1  | GT-1 binding | GRWAAW | 7   |
|         | TATA-box | TATAA | 2   |
|         | TATA-box | TATTAA | 5   |
|         | ABRE-related | MACGYGB | 1   |
|         | POLASIG3 | AATATAT | 1   |
|         | ROOTMOTIFTAPOX1 | ATATT | 3   |
| AtCBF2  | GT-1 binding | GRWAAW | 6   |
|         | HSF (heat shock factor) | NTTCNNNGAAN | 4   |
|         | TATA-box | TTTTA | 5   |
|         | TATA-box | TATATA | 1   |
|         | W-box | TTGAC | 1   |
|         | ABRE-related | MACGYGB | 1   |
| AtCBF3  | WRKY binding | YTGA CY | 1   |
|         | CAAT-box | ATTTG | 2   |
|         | TATA-box | ATATAT | 1   |
|         | ABRE-related | MACGYGB | 1   |
|         | TATA-box | TATATT | 2   |
|         | W-box | TTGAC | 1   |
|         | TATA-box | TATAA | 3   |
| OsDREB1A | LTR | CCGAC | 1   |
|         | bZIP binding | ACACNN | 3   |
|         | Root-hair-specific | KCACCW | 2   |
|         | Light responsive | CCRCCC | 1   |
|         | Sugar starvation | CGACG | 1   |
|         | MAD8 binding | CWG | 1   |
| OsDREB1B | Polles-related | AGAA A | 3   |
|         | TATA-box | TATAA | 1   |
|         | SEF5 binding | AACCCA | 1   |
|         | Sugar starvation | CGACG | 2   |
|         | E2 binding | WTTSSCSR | 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, $P < 0.01$).

For the OsDREB1A and OsDREB1B clusters, the MSNM four-combinations were based on seven and eight motifs, respectively.

sequences for the A. thaliana genes is 68.6 ± 3.9%, whereas in the rice promoters it is 54.4 ± 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 LTR 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, 2006; Benedict et al., 2006) consens by democracy. Using meta-analyses of microarray data to model the cold acclimation signaling pathway in Arabidopsis.

et al
Agarwal, M. (2006) The OsDREB1s have previously been shown to not only enhance tolerance to cold, but also drought and high-salt, indicating that they participate 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, 2006; Benedict et al., 2006). This complicates the elucidation of signaling pathways directly coupled to an increased cold tolerance.

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 ICFR3 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 ACFRBs (M.Braunit 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 participate 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, 2006; Benedict et al., 2006). This complicates the elucidation of signaling pathways directly coupled to an increased cold tolerance.

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

Conflict of Interest: none declared.

REFERENCES

Agarwal, M. et al. (2006) A RR33 type MYB transcription factor is involved in the cold regulation of cbf genes and in acquired freezing tolerance. J. Biol. Chem., 281, 37563-37565.

Baker, S.S. et al. (1994) The 5’-region of Arabidopsis thaliana cbf15a has cis-acting elements that confer cold-, drought- and ABA-regulated gene expression. Plant Mol. Biol., 29, 701-713.

Bo, M.A. and Tavazoei, 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 Physiol., 141, 1219-1232.