Genome-wide identification and expression analysis of the NF-Y family of transcription factors in *Triticum aestivum*

Troy J. Stephenson[1,2], C. Lynne McIntyre[1], Christopher Collet[2] and Gang-Ping Xue[1]*

[1] CSIRO Plant Industry, 306 Carmody Road, St Lucia QLD 4067, Australia.

[2] School of Life Sciences, Queensland University of Technology, GPO Box 2434, Brisbane QLD 4001, Australia.

* Corresponding author:

Tel: 61 7 3214 2354

Fax: 61 7 3214 2920

E-mail: Gang-Ping.Xue@csiro.au

Abstract Nuclear Factor Y (NF-Y) is a trimeric complex that binds to the CCAAT box, a ubiquitous eukaryotic promoter element. The three subunits NF-YA, NF-YB and NF-YC are represented by single genes in yeast and mammals. However, in model plant species (*Arabidopsis* and rice) multiple genes encode each subunit providing the impetus for the investigation of the NF-Y transcription factor family in wheat. A total of 37 NF-Y and Dr1 genes (10 NF-YA, 11 NF-YB, 14 NF-YC and 2 Dr1) in *Triticum aestivum* were identified in the global DNA databases by computational analysis in this study. Each of the wheat NF-Y subunit families could be further divided into 4-5 clades based on their conserved core region sequences. Several conserved motifs outside of the NF-Y core regions were also identified by comparison of NF-Y members from wheat, rice and *Arabidopsis*. Quantitative RT-PCR analysis revealed that some of the wheat NF-Y genes were expressed ubiquitously, while others were expressed in an organ-specific manner. In particular, each TaNF-Y subunit family had members that were expressed predominantly in the endosperm. The expression of nine NF-Y and two Dr1 genes in wheat leaves appeared to be responsive to drought stress. Three of these genes were up-regulated under drought conditions, indicating that these members of the NF-Y and Dr1 families are potentially involved in plant drought adaptation. The combined expression and phylogenetic analyses revealed that members within the same phylogenetic clade generally shared a similar expression profile. Organ-specific expression and differential response to drought indicate a plant-specific biological role for various members of this transcription factor family.

Keywords CCAAT-binding factor · expression profile · Nuclear Factor Y · phylogeny · transcriptional regulation · *Triticum aestivum*

Abbreviations

AEL Apparent expression level

CBF CCAAT-binding factor

CYC1 Cytochrome c, isoform 1

EST Expressed sequence tag

HFM Histone-fold motif

LEC1 LEAFY COTYLEDON1

NC2 Negative Cofactor 2

NJ Neighbor-joining

NF-Y Nuclear Factor Y

TBP TATA-binding protein

Introduction

The CCAAT box is one of the most common elements in eukaryotic promoters. It is typically found between 60-100 bp upstream of the transcription start site and is functional in both direct and inverted orientations (Bucher 1990; Dorn et al. 1987b; Edwards et al. 1998; Mantovani 1998). CCAAT boxes are highly conserved within homologous genes across species in terms of position, orientation, and flanking nucleotides (Mantovani 1999). Genes under the control of promoters containing CCAAT boxes may be expressed in a ubiquitous manner or can be tissue- and/or stage-specific suggesting that patterns of gene expression are also determined by other *cis*- and *trans*-acting factors. Multiple *trans*-acting factors are associated with the CCAAT box, but only Nuclear Factor Y (NF-Y) has an absolute requirement for the

pentanucleotide (Dorn et al. 1987a). NF-Y from mammals preferentially binds CCAAT with specific flanking nucleotides: C[GA][GA]CCAAT[CG][AG]C[AC] (Mantovani 1999). NF-Y is comprised of three distinct subunits: NF-YA (CBF-B in vertebrates), NF-YB/CBF-A, and NF-YC/CBF-C (Romier et al. 2003). Each subunit is required for DNA-binding, subunit association and transcriptional regulation in both vertebrates and plants (Sinha et al. 1995). The NF-Y subunit proteins from vertebrates and plants also share high homology with the HAP2/3/5 complex from *Saccharomyces cerevisiae* (bakers yeast) (Hahn et al. 1988; McNabb et al. 1995; Pinkham et al. 1987). Each NF-Y subunit in animals and yeast is encoded by a single gene, but in plants each of the three subunits are encoded by multiple genes (Edwards et al. 1998; Gong et al. 2004; Riechmann et al. 2000). This indicates a more complex regulatory role for the various NF-Y proteins in plants than in other organisms.

Each NF-Y subunit has its own distinct, highly conserved core region (Mantovani 1999). The conserved core of the NF-YA subunit comprises two functionally distinct domains (Maity and de Crombrugghe 1992) whereby the N-terminal domain is required for subunit interaction with the NF-YB NF-YC heterodimer and the C-terminal domain is involved in DNA binding site recognition (Xing et al. 1994). Both the NF-YB and NF-YC conserved core regions also contain residues contacting DNA (Romier et al. 2003) and NF-YB has been found to contribute to DNA-binding specificity (Zemzoumi et al. 1999).

The NF-YB and NF-YC subunits contain the highly conserved histone-fold motif (HFM) and are structurally similar to core histone subunits H2B and H2A, respectively (Arents and Moudrianakis 1995; Baxevanis et al. 1995; Mantovani 1999). In contrast to the high levels of conservation of the NF-YA core regions across vertebrates, plants and yeast, the core sequences of the NF-YB and NF-YC subunits are not as conserved. Indeed, the sequences of members of the NF-YB and NF-YC subunit families of *Arabidopsis thaliana* are more divergent from each other than those of the NF-YA family members (Gusmaroli et al. 2001).

Within the NF-Y family, two groups of proteins, Dr1 (NC2β) and Drap1 (NC2α) are related to NF-YB and NF-YC respectively (Sinha et al. 1996).

The biological roles of the NF-Y family in plants have not been well studied. In *Arabidopsis,* many NF-Y subunit genes are expressed ubiquitously, although some are differentially expressed (Gusmaroli et al. 2001). For example, *AtNF-YC-4* RNA accumulates in seeds 7 days after germination, while *AtNF-YB-9* (*LEC1*) is only expressed in green siliquae (Gusmaroli et al. 2001). Over-expression of *LEC1* in transgenic *Arabidopsis* activates expression of embryo specific genes and results in the formation of embryo-like structures in leaves, indicating a role for *LEC1* in modulating embryo development (Lotan et al. 1998). *AtNF-YA-4* shows the most restricted expression patterns of the NF-YA members with RNA detected in senescent flowers, caulines, stems and 4 day old seedlings at varying levels of abundance (Gusmaroli et al. 2001). Three NF-YB subunits from *Oryza sativa* (*OsHAP3A-C*) are expressed ubiquitously (Miyoshi et al. 2003) and suppression of *OsHAP3* expression results in degeneration of the chloroplast and down regulation of photosynthesis genes indicating involvement of the gene products in chloroplast development (Miyoshi et al. 2003). In transgenic maize over-expression of a single ZmNF-YB subunit gene improves corn yield and increases chlorophyll content in the leaves of plants grown in the field under water-limited conditions (Heard et al. 2006).

Wheat boasts the largest total production of the cereal crops on the globe. Four of the five major wheat exporters (Argentina, Australia, Canada and U.S.A) grow wheat under conditions of moderate to severe drought, which is also the main abiotic constraint on wheat yield (Araus et al. 2002). Drought affects over 10% of arable land, reducing average yields of major crops by over 50% (Bray 2000). Response to abiotic stress requires plants to alter many biological processes such as cell cycle progression, metabolic rates and physiological balance. Transcription factors act as switches in the complex regulatory networks controlling these

biochemical and physiological processes (Shinozaki et al. 2003; Zhu 2002). There is no information about the NF-Y family of transcription factors in wheat and their potential role in drought response. This study aimed to use sequence information from the model plants *Arabidopsis* and rice to identify all unique *Triticum aestivum* NF-Y subunit members (TaNF-Y) from the global nucleotide sequence databases, to determine the nature and pattern of expansion of the TaNF-Y subunit gene families, and to identify the expression profile of the TaNF-Y genes.

Materials and methods

Database searches for TaNF-Y family members

Database searches were performed to collect all members of *T. aestivum* NF-Y family members using the conserved core of *O. sativa* and *A. thaliana* NF-Y subunit amino acid sequences. NF-Y subunit protein sequences were retrieved from the Rice Transcription Database (RiceTFDB) (version 2.0, http://ricetfdb.bio.uni-potsdam.de/v2.0/) and the *Arabidopsis* Transcription Database (ArabTFDB) (version 1.0, http://arabtfdb.bio.uni-potsdam.de/v1.0). The TBLASTN program was used with an E-value cut-off of $1.0e^{-08}$ to identify both assembled wheat EST (expressed sequence tag) contiguous sequences and EST singletons from TaGI (Release 10.0, http://compbio.dfci.harvard.edu/tgi/plant.html), PlantGDB (http://www.plantgdb.org/), and GenBank (http://www.ncbi.nih.gov/) available in October 2006. Many sequences initially collected were not unique, had incomplete domains or appeared to contain incorrect open-reading frames (ORFs). The latter two cases were excluded from further analysis. Pairwise sequence alignments and comparison of the EST contents of assembled sequences were used to find unique representatives of each gene. As some sequence differences in ESTs are due to

sequencing errors or DNA polymorphism in wheat cultivars, nucleotide sequences with 98% or greater identities over their length were considered as the same gene (many homoeologous genes in wheat are also likely to fall into this category) and a representative was chosen. For assembled sequences, EST contig contents were analysed. Assembled sequences containing ESTs identical to another were considered as the same sequence and one representative was used for further analysis. Assembled sequences were queried against Pfam (version 20.0, http://www.sanger.ac.uk/Software/Pfam/) and ProDom (Release 2005.1, http://protein.toulouse.inra.fr/prodom.html) to confirm their identity as NF-Y subunits. The HAP subunit sequences used in the phylogenetic analysis were retrieved from the *Saccharomyces* genome database (http://yeastgenome.org).

Consensus logos

A. thaliana, O. sativa and *T. aestivum* NF-Y subunit conserved core regions and identified motifs were used to create consensus logos using WebLogo (http://weblogo.berkeley.edu/) (Crooks et al. 2004). Default program parameters were used.

Alignments and phylogenetic analysis

All unique sequences were aligned initially using CLUSTALX version 1.83 (Thompson et al. 1997). Outputs were further refined by manual alignment of the N- and C-terminal sequences. Phylogenetic analysis was undertaken using the conserved core amino acid sequences of each subfamily as well as full-length amino acid sequences. TreePuzzle version 5.2 was used with exact Parameter estimation and Gamma distribution of rates among sites, to calculate the shape parameter α (Schmidt et al. 2002). The Coefficient of Variation (CV) parameter was calculated

by $CV = 1/\sqrt{\alpha}$ as described in the PHYLIP documentation (http://evolution.genetics.washington.edu/phylip/phylip.html). Distance analysis was performed using the PHYLIP program package version 3.65 (http://evolution.genetics.washington.edu/phylip.html) (Retief 2000). Sequences were first bootstrapped using the SEQBOOT program in order to obtain an estimate of the reliability for the analysis. The distance matrices were created for aligned and bootstrapped amino acid core regions using the PROTDIST program using Gamma distribution among sites and the JTT matrix for amino acid substitutions. The NEIGHBOR program was used to convert distance matrices into phylogenetic trees using a randomized input order. The *S. cerevisiae* HAP2, HAP3 and HAP5 sequences were used for the NF-YA, NF-YB and NF-YC out-groups respectively. For the bootstrapped data, the CONSENSE program was then used to create a consensus tree using the Majority rule extended consensus type.

Determination of conserved motifs

Identification of conserved motifs outside of the highly conserved NF-Y subunit core regions was accomplished with multiple sequence alignments and Multiple Em for Motif Elicitation (MEME) version 3.5.3 (http://meme.sdsc.edu) (Bailey and Elkan 1994). Options for MEME were adjusted to find motifs of lengths 4-12 amino acids. Input sequence data was modified to exclude the conserved core for each and was replaced with poly-X eliminating the possibility of misleading results (flanking regions to the conserved core are highly conserved among subunit families). Training sets for each subunit family were the non-redundant NF-Y subunit amino acid sequences from *A. thaliana, O. sativa,* and *T. aestivum*. Logos for identified motifs were created as per the method for conserved core consensus logos.

Plant materials

Spring wheat (*Triticum aestivum* L. cv. Babax) plants were grown in a controlled-environment growth room under both controlled and drought stressed growth conditions (Xue et al. 2006). Night and day conditions were 16-h light and 14/18˚C. Control plants were well watered. The drought treatment was achieved by water deprivation of 4-week-old plants grown in pots (~1.5L volume) containing a 3:3:1 mix of sand:soil:peat, until desired water contents were achieved. Relative leaf water contents of drought-stressed plants were determined as described by Xue and Loveridge (2004). Plant organs were harvested and immediately immersed in liquid nitrogen and stored at -80˚C.

Preparation of total RNA and cDNA synthesis

Total RNA was isolated from the leaf and root, pre-anthesis spike and stem of well-watered plants at the pre-anthesis stage, 18-20 day post-anthesis endosperm, 24-30 day post-anthesis embryo wheat organs and the leaf from drought stress and control plants using Plant RNA Reagent from Invitrogen following the manufacturer's directions. RNA was treated with the RNase-Free DNase (Xue et al. 2006) and purified using the RNeasy Plant Mini-kit column (Qiagen) following manufacturer's instructions. First strand complementary DNA (cDNA) was synthesised using an oligo (dT20) primer from purified total RNA using SuperScript III Reverse Transcriptase (Invitrogen) following the manufacturer's instructions.

Quantitative RT-PCR analysis

Transcript levels for TaNF-YA, TaNF-YB and TaNF-YC subunit families were quantified by real-time PCR with an ABI Prism 7900HT sequence detection system (Applied Biosystems)

using SYBR Green PCR Master Mix (Applied Biosystems) according to manufacturer's instructions. Gene-specific primer pairs used are listed in Supplementary Table 1. For this analysis both internal and external reference genes were used. Four house-keeping (HK) genes from *T. aestivum* were used as internal controls (Supplementary Table 1): *TaRPII36* is a RNA polymerase II 36 kDa subunit gene; *TaRP15* is a RNA polymerase 15 kDa subunit gene; *TaCCF* is a putative chromosomal condensation factor gene and *TaPGM2* is a phosphoglucomutase gene. A bovine transcript (C12B07) was used as an external control (Xue et al. 2006). Preliminary experiments were performed to establish the amplification efficiency for each of the primer pairs, to allow for a direct comparison of the expression levels of the NF-Y and Dr1 subunits. A dilution range of cDNA samples were subjected to real-time PCR to collect Ct values where Ct is related to the logarithm of the dilution factor and a slope of best-fit line is used to measure the reaction efficiency $E = 10^{(-1/slope)}$ (Rasmussen 2001). Relative quantification was calculated as $E_t^{cCPt-sCPt} \times E_r^{sC\Pr-cC\Pr}$ (Pfaffl 2001). Ct values were collected from three replicate PCR reactions for each biological sample and a mean was taken from three biological samples. The specificity of the PCR reactions was determined by melting curve analysis of the products.

The apparent expression level (AEL) of each gene relative to an internal reference gene, *TaRPII36*, was calculated using the following formula: $Er^{Ct} \div Et^{Ct} \times F$, where Ct is cycle threshold (PCR cycle number required for reaching the signal point used for detection across samples), Er is reference gene amplification efficiency (*TaRPII36*), Et is target gene amplification efficiency, and F is amplicon size factor (reference gene amplicon size ÷ target gene amplicon size). We tentatively used AEL values here to provide an approximate estimation of relative expression levels among various genes under the situation where the absolute quantification of mRNA levels for a large number of genes using a RNA (or cDNA) calibration curve is not possible.

Statistical analysis

The Students *t*-test was performed as a test of significance. P-values of ≤ 0.05 were considered to indicate statistically significant differences.

Results

Identification of NF-Y genes in *Triticum aestivum*

The conserved NF-Y domains from *Arabidopsis* and rice NF-Y subunit families were used to screen for *T. aestivum* NF-Y genes from public sequence databases resulting in the identification of 37 unique nucleotide sequences. Ten NF-YA, 11 NF-YB, 14 NF-YC and two Dr1 homologues were identified and assigned the following identifiers: TaNF-YA1-10, TaNF-YB1-11, TaNF-YC1-14 and TaDr1A-B (Table 1, Fig. 1). Some NF-Y-like EST sequences have incomplete domains or apparently incorrect ORFs and have not been included for further analysis. Searches of Pfam and ProDom domain databases using the assembled TaNF-Y proteins confirmed that the identified sequences were NF-Y homologues.

TaNF-YA subunits. Maity and de Crombrugghe (1992) report that the conserved core of the NF-YA subunit is less than 60 amino acids long and contains two functionally distinct domains of about 20 amino acids each separated by a non-conserved spacer region of approximately 10 amino acids. Each of the ten TaNF-YA subunits identified has support for containing a NF-YA/CBF-B/HAP2 domain matching domain family PF02045.5 in the Pfam database with E-

values less than $1.0e^{-34}$. The core sequence of the TaNF-YA subunit peptides were also found to contain >60% identity with the yeast HAP2 subunit (data not shown). The conserved core domain of TaNF-YA is 57 amino acids long. The subunit association domain (SAD) is 25 amino acids in length (Romier et al. 2003) and begins at the first residue of the core (Fig. 1A). A spacer region divides the SAD from the DNA-binding domain and is located at positions 26-36, spanning 11 amino acids (Fig. 1A). The DNA-binding region is 21 amino acids in length (Romier et al. 2003) and is located at the C-terminal region of the wheat NF-YA conserved core at positions 37-57 (Fig. 1A).

TaNF-YB subunits. Individual TaNF-YB peptide sequences were queried against the Pfam database and matches were found with Pfam family PF00808.12. In each case, strong support for containing a NF-YB/CBF-A/HAP3 domain was found, with E-values of $\sim 1.0e^{-30}$ in all cases except for TaNF-YB6 which had an E-value of $1.5e^{-15}$. Core regions of the TaNF-YB subunits share between 52% and 62% identity with yeast HAP3 (data not shown). The TaNF-YB subunit core region is 88 amino acids in length (Fig. 1B) which is consistent with an average length of 90 residues reported by Maity and de Crombrugghe 1992. Functional domains required for interaction between NF-YB, NF-YC and NF-YA homologues and subunit interaction with DNA have been identified by Kim et al. (1996). Based on the sequence similarity with the mammalian CBF-A homologue, amino acids 5-44 and 51-84 in the conserved core region of TaNF-YB may be involved in heterodimer formation, residues 34-41 could interact with the TaNF-YA subunit and amino acids 1-29 may well be required for DNA interaction in the heterotrimer complex (Fig. 1B).

TaNF-YC subunits. Queries using TaNF-YC sequences against the Pfam database returned matches with NF-YB/CBF-A/HAP3 family domains and E-values of $\sim 1.0\,e^{-10}$ indicated

regions of significant similarity. Identified TaNF-YC sequences were then queried against the ProDom database and matches were found in each instance with the PD003659 (NF-YC/CBF-C/HAP5). For each sequence, E-values were $<1.0e^{-22}$ except for TaNF-YC13 and TaNF-YC14, which had E-values of $4.0e^{-14}$ and $3.0e^{-14}$, respectively. TaNF-YC subunit members are more divergent from the yeast HAP5 subunit than the two other subunit family members, with identities in the range of 30% to 40% (data not shown). The TaNF-YC conserved core region is 74 amino acids long (Fig. 1C), which is shorter than the average length of 84 residues found in other organisms (Maity and de Crombrugghe 1992). Sequence similarity with the mammalian CBF complex suggests that amino acid residues 3-4 of the core region are likely to interact with DNA in the heterotrimer, residues 1-16 and 62-72 are likely to be involved in interaction with TaNF-YA in the heterodimer and amino acids spanning positions 17-61 of the conserved region are likely to be involved in dimerization with TaNF-YB (Fig. 1C).

Dr1 subunits. Dr1 is a β-subunit of NC2, NC2β, and is highly homologous to NF-YB at the amino acid level (Sinha et al. 1996). Two Dr1 homologues were identified through database searches using the NF-YB conserved core as a query sequence. Subsequent searches for Dr1 subunit members from the available wheat nucleotide sequence data using the wheat, rice and *Arabidopsis* Dr1 homologues did not result in the identification of additional unique Dr1 sequences. The high degree of similarity between TaDr1A, TaDr1B and TaNF-YB subunit members can be seen in Fig. 1D where a representative TaNF-YB subunit member (TaNF-YB3) has been aligned against the two TaDr1 sequences.

Conserved sequences in the NF-Y subunits

Conserved core consensus. In order to identify conserved amino acid residues in the core region of the NF-Y subunits among plant species sequence logos were created from multiple sequence alignments of the conserved core domains of *T. aestivum, A. thaliana* and *O. sativa* NF-Y subunit members (Fig. 2A-C). This analysis showed that the core sequence of the NF-YA family of proteins from each of these three plant species is the most highly conserved among the three subunits (Fig. 2A). The core sequence of the NF-YC proteins from these three plant species is the most divergent (Fig. 2C). There are only three absolutely conserved residues between *Arabidopsis*, rice and wheat within the NF-YC core region (Fig. 2C). There are 17 residues absolutely conserved in the core domain of the NF-YB subunit across the three species (Fig. 2B).

Conserved motifs outside of the NF-YA core region. Four motifs outside the conserved core region were identified in the NF-YA amino acid sequences of *Arabidopsis*, rice and wheat with the use of the MEME algorithm and multiple sequence alignments (Fig. 1A and Fig. 3). A non-polar/hydrophobic hexapeptide region found in all three species is located approximately 40 amino acids from the N-terminal end of the core region (Fig. 1A) and has absolute conservation for Gly and Tyr at the terminal residues (Fig. 3A). Figure 3B shows a five residue motif that is located around 5 amino acids from the N-terminus of the core. Only one of the five residues is not absolutely conserved where Val may be substituted for Met, both of which share non-polar and hydrophobic properties. Figure 3C shows a five residue motif located approximately 50 amino acids from the N-terminus of the core (Fig. 1A), two Tyr residues absolutely conserved and the physicochemically conservative substitution of Asp for Glu occurs in the first position. A motif, four residues in length has absolute conservation across the first three residues and is found approximately 30 amino acids from the N-terminus of the

core. Ile may be substituted with other non-polar hydrophobic residues at the last position (Fig. 1A and 3D).

Conserved motifs outside the NF-YB core region. Three motifs were identified outside of the conserved core sequence of the NF-YB peptide. A six residue motif adjacent the N-terminal side of the conserved core region is entirely conserved across the central four residues (Fig. 1B and 4A). Arg is usually present in the first position, but in a few cases is substituted by Lys, which has similar physicochemical properties. A motif of 10 amino acid residues was found in eight NF-YB subunit proteins and is located approximately 10 amino acids from the C-terminal residue of the core region (Fig. 1B and 4B). Positions 1, 3, 8 and 10 are absolutely conserved while small amino acid residues such as Ala, Ser or Gly occupy position 2. In position 4, Asp is substituted with Glu, which is also a negatively charged residue. In all but one case, position 6 is occupied by Ser and is substituted by Asn (both are small and polar). Position 7 is occupied in most cases by Val, but in some cases by Ala, which shows conservation for small residues. Position 9 usually contains Lys, but is substituted in one instance by Arg (both positively charged). A polar and hydrophobic motif, which is five amino acids in length, is located at the C-termini (Fig. 1B and 4C). Positions 1 and 3 are absolutely conserved, and all other residues have shared physicochemical properties within each site.

Conserved motifs outside of the NF-YC core region. Four motifs outside of the conserved core region were identified in the members of the NF-YC family from *Arabidopsis*, rice and wheat. One five residue motif is found in 19 NF-YC proteins and is located two amino acids from the N-terminal residue of the core region (Fig. 1C and 5A). All but one of the residues are 100% conserved. In 18 of the 19 proteins position 2 is occupied by Phe, but is occupied by another hydrophobic residue, Ile in TaNF-YC1. A Gln and hydrophobic residue rich motif of

11 amino acids is found in the NF-YC peptides approximately 20 amino acids from the N-terminal side of the core region (Fig. 1C and 5B) and may represent a motif within a transcriptional activation domain (Courey et al. 1989; Coustry et al. 1996; Gill et al. 1994). Flanking the C-terminus of the conserved core sequence is a seven residue motif that is conserved among 14 NF-YC family members (Fig. 1C and 5C). The fourth motif contains terminal Pro residues and Tyr at positions 2 and 4 (Fig. 1C and 5D). The remaining residues are hydrophobic, except in one case where Gln is found at position 3.

Phylogenetic analysis

To determine the nature and pattern of expansion of the TaNF-Y subunit genes, gene family trees were constructed using the Neighbor-joining (NJ) method for each subunit family. Outside the conserved core sequences there is considerable divergence making alignment difficult, thus for the initial analysis only the sequences of conserved core regions of each subunit were used. However, since many sequences were identical, in subsequent analysis the full-length protein sequences were used to obtain more detailed information about some of the relationships of the NF-Y subunits.

Phylogenetic trees of the TaNF-YA. Core region sequences of the TaNF-YA peptides are highly conserved, but can be divided into four clades assigned identifiers I-IV in wheat (Fig. 6A). Clade I contains TaNF-YA6 and TaNF-YA2, both of which have the most divergent core region sequences from the majority consensus with substitutions at positions 1, 2, 20, and 30 (Fig. 2A). The clustering of these two subunits as a separate clade is also supported in the NJ tree generated using the full-length protein sequences (Fig. 6B). Significant bootstrap values support the clustering of clade I. Clade III comprising TaNF-YA1 and TaNF-YA10 is present

in both trees and the clustering is well supported by very high bootstrap values. Similarly, clade IV comprising TaNF-YA4 and TaNF-YA9 is also present in both trees and their clustering is well supported by very high bootstrap values. Subunits clustering in clade IV have 100% amino acid identity between their conserved core regions, however the 65 amino acids in the N-terminal region of TaNF-YA9 are not found in TaNF-YA4 (Fig. 1A). Similarly, the conserved core sequences from TaNF-YA3, TaNF-YA7 and TaNF-YA8 have 100% amino acid identity; however, TaNF-YA3 does not have a 63 amino acid N-terminal sequence present in TaNF-YA7 and TaNF-YA8 (Fig. 1A). TaNF-YA5 resides on a separate branch in both trees and appears as an out-group to the other TaNF-YA subunits. This is expected, as this subunit is the most divergent of the TaNF-YA members in both the conserved core and full-length amino acid sequences. TaNF-YA5 has substitutions at positions 21, 27, 29, and 54 within the core region, which are unique and has a high level of divergence in both the N- and C-termini of the protein (Fig. 1A). To illustrate the sequence relationship of TaNF-YA members with those from model plant species, phylogenetic trees for NY-YA members from *T. aestivum*, *Arabidopsis* and rice were also constructed using conserved core and full-length amino acid sequences (Supplementary Fig. 1A-B).

Phylogenetic trees of the TaNF-YB. The NF-YB proteins have been divided into two classes in *Arabidopsis*, LEC1-like and non-LEC1-like and can be identified by 16 unique residues within the core region of the LEC1-like NF-YB subunits (Lee et al. 2003). The TaNF-YB subunit proteins cluster into four distinct clades in the NJ trees using either the core sequence or the full-length protein sequence (Fig. 6C-D). Clade IV contains the members TaNF-YB1 and TaNF-YB9 (Fig. 6C-D). Twelve of the LEC1 specific residues from *Arabidopsis* are conserved in the wheat peptides clustering in clade IV. The TaNF-YB subunit family has 12 non-LEC1-like NF-YB members that can be further divided into three clades. Clade III

contains the four members TaNF-YB2, TaNF-YB4, TaNF-YB10 and TaNF-YB11 and clustering is well supported with a high bootstrap value of 769 in the conserved core NJ tree. Clade II contains two members TaNF-YB5 and TaNF-YB6 that cluster with bootstrap values of 781 and 575 (Fig. 6C). Clade I has three members, TaNF-YB3, TaNF-YB7 and TaNF-YB8, and clustering is well supported in both trees with very high bootstrap values. Branching patterns on both trees indicate that clade I and II are more closely related to each other than they are to other clades in the protein subunit family. Phylogenetic trees for NY-YB members from *T. aestivum*, *Arabidopsis* and rice are shown in Supplementary Fig. 1C-D.

Phylogenetic trees of the TaNF-YC. The TaNF-YC subunit family is the most divergent of the three subunits. TaNF-YC subunit proteins cluster into five clades in the NJ trees (Fig. 6E-F). Clade II contains three members, TaNF-YC5, TaNF-YC11 and TaNF-YC12, and clustering is well supported with very high bootstrap values in both trees (Fig. 6E-F). All sequences in clade II share identical core regions (Fig. 1C). Clade I contains two members TaNF-YC3 and TaNF-YC9 that cluster together in both trees with strong support from bootstrap values (Fig. 6E-F). Like clade II, sequences in clade I have 100% identity in the core region (Fig. 1C). Clade IV contains two members, TaNF-YC6 and TaNF-YC8 clustering together in both trees and is well supported by a bootstrap value of 982 in the NJ tree based on the core region (Fig. 6E-F). TaNF-YC13 and TaNF-YC14 cluster together in clade V (Fig. 6E-F). TaNF-YC2 and TaNF-YC10 cluster together in clade III in both NJ trees and is well supported by high bootstrap values. Sequences in clade III contain identical core regions (Fig. 1C). There is support from high bootstrap values that clades IV and V are sister clades in both NJ trees (Fig. 6E-F). Phylogenetic trees for NY-YC members from *T. aestivum*, *Arabidopsis* and rice are shown in Supplementary Fig. 1E-F.

Expression profiles of the NF-Y gene family in wheat

Quantitative RT-PCR was used to examine the expression profiles of the wheat NF-Y and Dr1 genes. To estimate relative expression levels between genes, we analysed the apparent expression level (AEL) of each gene relative to an internal control gene, *TaRPII36,* using a combination of the following factors: the Ct value, PCR amplification efficiency and amplicon length (see Materials and methods). A gene with a higher level of expression generally has a lower Ct value when other conditions (such as PCR amplification efficiency and amplicon length) are the same (Pfaffl 2001), especially when the primers designed are predominantly located at or near the 3' region.

A moderate expression level was seen in the root for the TaNF-YA genes based on the AEL values (Supplementary Table 2). *TaNF-YA1* and *TaNF-YA6* were identified as being expressed at the highest level of all the TaNF-YA genes in the root (Supplementary Table 2). *TaNF-YA5* was found to be expressed the lowest of all the TaNF-YA genes in the root (Supplementary Table 2). Four genes (*TaNF-YA3, 4, 7* and 9) were predominantly expressed in the endosperm tissue (Fig. 7A). Two of these, *TaNF-YA3* and *TaNF-YA4* were expressed at their lowest levels in the leaf, while *TaNF-YA7* and *TaNF-YA9* were expressed at their lowest levels in the spike (Fig. 7A). As *TaNF-YA7* and *TaNF-YA8* share extremely high nucleotide sequence homology, real-time PCR primers for *TaNF-YA7* also amplify *TaNF-YA8.* Three genes (*TaNF-YA1, 2* and *10*) were expressed at their highest level in the root and at their lowest levels in the endosperm. *TaNF-YA6* was expressed at its highest level in the root and lowest in the leaf. *TaNF-YA5* was found at its highest expression level in the leaf.

The transcripts of TaNF-YB subunit genes were detected in all organs except for *TaNF-YB9*, which was not detected in the stem (Fig. 7B). Expression analysis of the *TaNF-YB5* gene was not carried out as no PCR products were amplified using three sets of primer pairs

attempted. *TaNF-YB5* is an EST singleton and thus may be a less reliable representation of the actual sequence. The transcript levels of the TaNF-YB genes in the root varied more than the TaNF-YA genes, with the highest transcript level being identified for the *TaNF-YB2* gene and the lowest for the *TaNF-YB9* gene (Supplementary Table 2). *TaNF-YB1* and *TaNF-YB9*, which are homologous to the *Arabidopsis LEC1* gene, were expressed predominantly in the endosperm. However, *TaNF-YB2* (which is not homologous to LEC1) was also highly expressed in the endosperm, compared to other plant organs. Of the TaNF-YB genes predominantly expressed in the endosperm, *TaNF-YB1* and *TaNF-YB2* were expressed at their lowest level in the leaf, while *TaNF-YB9* was expressed at the lowest level in the root. Five genes (*TaNF-YB3, 4, 6, 7* and *8*) were expressed at their highest levels in the leaf and their lowest in the embryo. *TaNF-YB10* was expressed at its highest level in the endosperm and leaf, while *TaNF-YB11* showed little variation between the organs.

Among the transcripts of TaNF-YC genes the highest expression level in the root was found with *TaNF-YC6* (Supplementary Table 2). *TaNF-YC10* was not detected in the leaf and is the single TaNF-YC gene not detectable in all organs. Analysis of *TaNF-YC4* expression was not undertaken for the same reasons encountered for *TaNF-YB5*. The expression level of *TaNF-YC1* was very low in all of the six organs (Supplementary Table 2, Fig. 7C). Two genes (*TaNF-YC2* and *TaNF-YC10*) were expressed predominantly in the endosperm and were either not expressed in the leaf (*TaNF-YC10*) or expressed at the lowest level in the leaf (*TaNF-YC2*). Three genes (*TaNF-YC5, 11* and *12*) were expressed at their highest levels in the leaf, spike and stem. Expression levels of *TaNF-YC13* and *TaNF-YC14* were at their highest in the embryo. Three genes (*TaNF-YC6, 7* and *8*) exhibited their highest expression levels in the leaf. *TaNF-YC3* and *TaNF-YC9* were found at their lowest expression levels in the root, with relatively little variation among the other five tissue types.

Expression analysis of the two TaDr1 genes revealed that both were expressed in all six wheat organs (Fig. 7D) with similar levels in the root (Supplementary Table 2). The highest expression level for both genes was in the endosperm, embryo and spike, while the lowest expression levels were in the leaf.

A preliminary expression analysis for drought responsiveness of all TaNF-Y genes was performed using one control leaf sample and one drought-stressed leaf sample (data not shown). Thirteen genes that showed >2-fold difference in levels of expression between control and drought-stressed samples were selected for further analysis. As shown in Fig. 8, eleven of the thirteen had statistically significant differences between expression levels when subjected to drought-stressed conditions. At least one member from each subunit family and both TaDr1 genes were found to be responsive to drought stress (i.e. one TaNF-YA, five TaNF-YB, three TaNF-YC and both TaDr1 genes). Eight of the eleven genes were down-regulated in drought-stressed wheat leaves (relative leaf water content ranging from 75% to 81%) and three genes were significantly up-regulated under conditions of drought. *TaNF-YA1* mRNA levels showed a significant reduction in drought-affected leaves to around a third of that seen in leaves of the non-stressed control (Fig. 8). A similar level of down-regulation in the drought-stressed leaves was seen for four TaNF-YB genes (*TaNF-YB3, 6, 7* and *8*) and two TaNF-YC genes (*TaNF-YC11* and *12*). *TaNF-YB2*, *TaDr1A* and *TaDr1B* were up-regulated genes under drought conditions by over 2-fold for each.

Correlation between the gene expression levels of TaNF-Y genes

Transcription factors that have correlated expression may be involved in transcriptional regulation of similar biological processes and strong correlation has been found to exist between some TaNF-Y genes. Three TaNF-YB genes (*TaNF-YB3, TaNF-YB7* and *TaNF-*

YB8), two TaNF-YA genes (*TaNF-YA3* and *TaNF-YA4*) and two TaNF-YC genes (*TaNF-YC11* and *TaNF-YC12*) had high correlation coefficients in expression within each of the three subunit families (Fig. 9A-C). Correlation of expression was also identified between members from different subunit families. One TaNF-YA subunit member (*TaNF-YA5*) and one TaNF-YC subunit member (*TaNF-YC8*) were found to be strongly correlated in expression (r = 0.94) (Fig. 9D). High correlation was also found between *TaNF-YA9* and *TaNF-YB10* (r = 0.96) (Fig. 9E) and between *TaNF-YB3* and *TaNF-YC12* (r = 0.98) (Fig. 9F).

Discussion

A large expansion in the members of the NF-Y family in plant genomes presents an interesting potential for differential gene regulation by NF-Y complexes. The findings of multiple NF-Y subunit genes in *Arabidopsis* and rice provided the impetus for the investigation of the biological role of this transcription factor family in wheat. In this study 10 NF-YA, 11 NF-YB, 14 NF-YC and 2 Dr1 genes were identified in *T. aestivum* from the sequence databases. The numbers (a total of 37) of TaNF-Y family members identified in this study are less than the apparent family sizes listed at the Plant Transcription Factor Database (Riano-Pachon et al. 2007) for *Arabidopsis* (42 NF-Y gene loci) and rice (45 NF-Y gene loci). It is likely this is the minimum number of genes present in *T. aestivum*, as a complete list of TaNF-Y genes will have to await the complete sequence of the wheat genome.

Given that wheat is an allohexaploid (genome AABBDD) (Feldman 2001), it would be expected that each of the three genomes would contain the same genes (homoeologues). To reduce the complexity of analysis in this study, we took sequences with 98% or greater identity to be a single gene. Therefore, it is likely that homoeologous EST sequences were assembled

into a single gene for analysis. However, this approach also has an additional potential error of combining very similar paralogous genes (Table 1), which arise through gene duplication events. Thus, the number of NF-Y genes present in the wheat genome could be higher than that reported here. Nevertheless, this study serves as a preliminary investigation into the genome-wide investigation of the NF-Y family in this important cereal crop species.

The analysis of the NF-Y transcription factor family in wheat has produced several novel findings. The eleven motifs identified here bioinformatically have not been reported elsewhere and this is the first time conservation has been identified outside of the core region in the plant NF-Y subunit families. The extremely high level of conservation of these motifs between three plant species in the otherwise highly divergent terminal regions provides support for the validity of these motifs. Furthermore, some of these motifs were common to subunits which share similar expression profiles and cluster together in the NJ trees. The identification of NF-Y subunit members which have identical conserved cores with highly similar expression profiles, such as those expressed predominantly in the endosperm is also an interesting finding. Here it has been identified that a number of NF-Y subunit members are drought responsive. Taken with the organ specific expression pattern of some subunit members, the expansion of the NF-Y TF family in plants may have resulted in subfunctionalisation of some of the subunit members.

The TaNF-YA subunit family separates into four clades based on the conserved core sequence region. Similarities within TaNF-YA subunit members clustering in clade II (TaNF-YA3, TaNF-YA7 and TaNF-YA8) seem to extend beyond their identical cores to include four short shared motifs (GVVAAY, RVPLP, DPYYG, and HPQI). The transcript levels of TaNF-YA members of clade II were higher in the endosperm than in other organs. TaNF-YA clade IV members (TaNF-YA4 and TaNF-YA9) were also expressed predominantly in the endosperm. Members of TaNF-YA clades II and IV contain two shared motifs (HPQI and

RVPLP) in addition to their shared expression profiles. Furthermore, one member from each of TaNF-YA clades I and IV (*TaNF-YA3* and *TaNF-YA4*) were identified to have strongly correlated expression profiles across all six wheat organs. It is possible that the genes in these two clades have a similar function and play a more active regulatory role in the endosperm than in other organs, but DPYYG and GVVAAY motifs are likely to be not required for this function due to the absence from members of clade IV. The TaNF-YA subunit members that cluster as clade III (TaNF-YA1 and TaNF-YA10) share 100% amino acid identity over the core sequence region and are quite similar over their terminal protein sequences. Expression of the *TaNF-YA1* and *TaNF-YA10* genes was the highest in the root; however, they showed different responses in the leaf to drought stress. *TaNF-YA1* was down-regulated in the drought-stressed leaves to a third of its control transcript levels, but from preliminary analysis *TaNF-YA10* expression was seemingly unaffected by drought stress.

TaNF-YB genes can also be divided into four phylogenetic groups. The TaNF-YB proteins clustering in clade IV (TaNF-YB1 and TaNF-YB9) are homologues of the *Arabidopsis LEC1* gene (At1g21970.1) (Supplementary Fig. 1C), that is involved in embryogenesis (Lotan et al. 1998). TaNF-YB1 and TaNF-YB9 are predominantly expressed in the endosperm, suggesting that they are not involved in embryogenesis as is LEC1. The REQDRF motif is present in both of these genes but is not unique to the LEC1-like NF-YB subunit members as seven other NF-YB genes also contain this conserved motif. In contrast to the members of clade IV, the members of clade I (TaNF-YB3, 7 and 8) exhibited a very low level of expression in the endosperm. These genes were expressed at a relatively higher level in leaf compared to the other organs examined. Furthermore, the members of TaNF-YB clade I were found to have strongly correlated expression profiles. These three TaNF-YB genes also shared a similar response to drought stress and were down-regulated to a third of their control levels in the leaf. TaNF-YB3, 7 and 8 share 100% amino acid identity across the conserved NF-YB core and are

unique over their N-and C-terminal regions, however each contains the two conserved motifs, REQDRF and QPQYH. The latter motif may be part of a transcriptional activation region; rich in Gln and hydrophobic residues (Courey et al. 1989; Coustry et al. 1996; Gill et al. 1994). The members of clade I may share a similar biological function based on the similarity in sequence and expression. TaNF-YB2 and TaNF-YB10 cluster in clade III and also contain the QPQYH motif. *TaNF-YB2* and *TaNF-YB10* were expressed at the highest level in the endosperm amongst the six organs examined. *TaNF-YB2* represents an interesting TaNF-Y gene in terms of positive drought stress responsiveness, while *TaNF-YB10* was not.

Five clades were identified in the TaNF-YC family. Members of TaNF-YC clade II (TaNF-YC5, 11 and 12) have identical conserved cores and slight variation at the C-termini. These members share one conserved motif (DFKNH) outside of the NF-YC core sequence. Expression patterns for these three TaNF-YC subunit genes are highly similar in that they are mainly expressed in leaf, spike and stem and two genes (*TaNF-YC11* and *TaNF-YC12*) had strongly correlated expression profiles across all of the organs analysed. Furthermore, *TaNF-YC5*, *TaNF-YC11* and *TaNF-YC12* were down-regulated in the drought stress leaves. Expression levels of members of clade IV were unchanged under conditions of drought stress. TaNF-YC13 and TaNF-YC14 of clade V shared a similar expression pattern across the six wheat organs but showed differential responses to drought stress. TaNF-YC clade III members (TaNF-YC2 and TaNF-YC10) may have the functions in the endosperm, as their expression levels were extremely high in the latter organ compared to other organs. In addition, TaNF-YC10 mRNA was not detectable in the leaf and was low in other tissues. TaNF-YC clade IV members (*TaNF-YC6* and *TaNF-YC8*) had a similar expression pattern; they were expressed at the highest in the leaf and were not responsive to drought.

The combined expression and phylogenetic analysis of the TaNF-Y genes in this study revealed that the relationships identified phylogenetically were reflected in their expression

profiles. In general, phylogenetic clades with proteins containing identical conserved core sequences were found to share a similar expression pattern. In contrast, the sequences of the N-terminal and C-terminal regions of these proteins did not appear to correlate with expression profiles. TaNF-YA clade IV, TaNF-YB clade IV and TaNF-YC clade III all contain members with identical core regions within each subunit family and all were highly expressed in the endosperm. It is interesting to find a clade in both TaNF-YA and TaNF-YC families that are also highly specific to endosperm and match with the expression pattern of TaNF-YB genes in clade IV that are homologous to *Arabidopsis LEC1*. This may indicate that these subunits potentially interact to form NF-Y trimer complexes in the endosperm. Furthermore, strong correlation in expression was found between some members from different subunit families in wheat. Whether the subunits members can form unique NF-Y complex in plants awaits further investigation.

In rice, there would appear to be no more than three Dr1 (NC2β) subunits and one Drap1 (NC2α) subunit (Song et al. 2002), while *Arabidopsis* has at least one NF-YC subunit gene that exhibits significant sequence similarity to the human Drap1 subunit (Kusnetsov et al. 1999). Dr1 and Drap1 form a NC2 complex which is a global transcriptional repressor (Kim et al. 1997). TaNF-YC6 is a homologue to the rice OsDrap1 protein (Os11g34200.1) (Supplementary Figure 1E). Therefore, it appears that wheat has a minimum of two Dr1 homologues and one Drap1 homologue. Both TaDr1 genes were up-regulated during drought stress. NF-YB and NF-YC can interact with TATA binding-protein (TBP) in the absence of NF-YA indicating that some NF-YB and NF-YC subunits function independently on the formation of a trimer with NF-YA (Bellorini et al. 1997). Furthermore, Dr1 and Drap1 are highly homologous to NF-YB and NF-YC, which presents the question of whether the Dr1/Drap1 complex inhibits transcription by acting as an antagonist to the NF-Y subunits, preventing subunit association and subsequence binding to TBP.

In contrast to vertebrates and fungi, which have a single gene for each NF-Y subunit, plants have multiple genes for each subunit. The most intriguing questions are why plants have multiple members of each subunit, how they form trimer complexes among the members of the three subunit families and whether individual NF-Y trimer complexes share the same DNA-binding specificity. Plants require dynamic developmental programs that are able to adjust differentiation, growth and metabolism in response to the continuous changes in the environment. The presence of multiple genes for all three NF-Y subunits in wheat with different primary structure and expression patterns indicates a high level of complexity of regulation for this gene family in plants. Thus, the evolution of large NF-Y gene families may support the development of flexible regulatory mechanisms. Organ-specific expression of TaNF-Y genes and their differential response to drought stress suggest that individual members of the wheat NF-Y genes may have specific physiological roles, including their involvement in regulating gene expression in wheat adaptation to drought stress. Importantly this study identified a single wheat NF-YB gene (*TaNF-YB2*) that was significantly up-regulated in response to drought. Given that transgenic maize plants that over-express a NF-YB subunit tolerate drought stress better than wild type (Heard et al. 2006), it would be interesting to investigate whether over-expression of *TaNF-YB2* could produce a drought-tolerant wheat.

Acknowledgements

The authors are grateful to Jason Kam and Lindsay Shaw for providing RNA samples. The laboratory expense for this study was supported by a grant from the Australian Grains Research & Development Corporation.

Supplementary data

Supplementary Table 1. *Triticum aestivum* NF-Y subunit gene-specific and reference gene primers.

Supplementary Table 2. Ct and AEL values of TaNF-Y genes analysed by real-time PCR.

Supplementary Fig. 1. Phylogenetic trees of the NF-Y subunit families in *Arabidopsis*, rice and wheat.

References

Araus JL, Slafer GA, Reynolds MP, Royo C (2002) Plant breeding and drought in C3 cereals: what should we breed for? Ann Bot (Lond) 89 Spec No: 925-940

Arents G, Moudrianakis EN (1995) The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. Proc Natl Acad Sci U S A 92: 11170-11174

Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol 2: 28-36

Baxevanis AD, Arents G, Moudrianakis EN, Landsman D (1995) A variety of DNA-binding and multimeric proteins contain the histone fold motif. Nucleic Acids Res 23: 2685-2691

Bellorini M, Lee DK, Dantonel JC, Zemzoumi K, Roeder RG, Tora L, Mantovani R (1997) CCAAT binding NF-Y-TBP interactions: NF-YB and NF-YC require short domains adjacent to their histone fold motifs for association with TBP basic residues. Nucleic Acids Res 25: 2174-2181

Bray EA, Bailey-Serres, J. and Weretilnyk, E (2000) Responses to abiotic stresses. Biochemistry and molecular biology of plants, pp. 1158-1203. American Society of Plant Physiologists.

Bucher P (1990) Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. Journal of Molecular Biology 212: 563-578

Courey AJ, Holtzman DA, Jackson SP, Tjian R (1989) Synergistic activation by the glutamine-rich domains of human transcription factor Sp1. Cell 59: 827-836

Coustry F, Maity SN, Sinha S, de Crombrugghe B (1996) The transcriptional activity of the CCAAT-binding factor CBF is mediated by two distinct activation domains, one in the CBF-B subunit and the other in the CBF-C subunit. J Biol Chem 271: 14485-14491

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res 14: 1188-1190

Dorn A, Bollekens J, Staub A, Benoist C, Mathis D (1987a) A multiplicity of CCAAT box-binding proteins. Cell 50: 863-872

Dorn A, Durand B, Marfing C, Le Meur M, Benoist C, Mathis D (1987b) Conserved major histocompatibility complex class II boxes--X and Y--are transcriptional control elements and specifically bind nuclear proteins. Proc Natl Acad Sci U S A 84: 6249-6253

Edwards D, Murray JA, Smith AG (1998) Multiple genes encoding the conserved CCAAT-box transcription factor complex are expressed in Arabidopsis. Plant Physiol 117: 1015-1022

Feldman M (2001) The origin of cultivated wheat. In: Bonjean AP, Angus WJ (eds) The World Wheat Book. Lavoisier Publishing, Paris.

Gill G, Pascal E, Tseng ZH, Tjian R (1994) A glutamine-rich hydrophobic patch in transcription factor Sp1 contacts the dTAFII110 component of the Drosophila TFIID complex and mediates transcriptional activation. Proc Natl Acad Sci U S A 91: 192-196

Gong W, Shen YP, Ma LG, Pan Y, Du YL, Wang DH, Yang JY, Hu LD, Liu XF, Dong CX, Ma L, Chen YH, Yang XY, Gao Y, Zhu D, Tan X, Mu JY, Zhang DB, Liu YL, Dinesh-

Kumar SP, Li Y, Wang XP, Gu HY, Qu LJ, Bai SN, Lu YT, Li JY, Zhao JD, Zuo J, Huang H, Deng XW, Zhu YX (2004) Genome-wide ORFeome cloning and analysis of Arabidopsis transcription factor genes. Plant Physiol 135: 773-782

Gusmaroli G, Tonelli C, Mantovani R (2001) Regulation of the CCAAT-Binding NF-Y subunits in Arabidopsis thaliana. Gene 264: 173-185

Hahn S, Pinkham J, Wei R, Miller R, Guarente L (1988) The HAP3 regulatory locus of Saccharomyces cerevisiae encodes divergent overlapping transcripts. Mol Cell Biol 8: 655-663

Heard JE, Nelson D, Adams TR, Purcell J (2006) Transgenic approaches to improving drought stress tolerance in maize Proc. 8th Intl. cong. Plant Mol.Biol., Sym, pp. 41.

Kim IS, Sinha S, de Crombrugghe B, Maity SN (1996) Determination of functional domains in the C subunit of the CCAAT-binding factor (CBF) necessary for formation of a CBF-DNA complex: CBF-B interacts simultaneously with both the CBF-A and CBF-C subunits to form a heterotrimeric CBF molecule. Mol Cell Biol 16: 4003-4013

Kim S, Na JG, Hampsey M, Reinberg D (1997) The Dr1/DRAP1 heterodimer is a global repressor of transcription in vivo. Proc Natl Acad Sci U S A 94: 820-825

Kusnetsov V, Landsberger M, Meurer J, Oelmuller R (1999) The assembly of the CAAT-box binding complex at a photosynthesis gene promoter is regulated by light, cytokinin, and the stage of the plastids. J Biol Chem 274: 36009-36014

Lee H, Fischer RL, Goldberg RB, Harada JJ (2003) Arabidopsis LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor. Proc Natl Acad Sci U S A 100: 2152-2156

Lotan T, Ohto M, Yee KM, West MA, Lo R, Kwong RW, Yamagishi K, Fischer RL, Goldberg RB, Harada JJ (1998) Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. Cell 93: 1195-1205

Maity SN, de Crombrugghe B (1992) Biochemical analysis of the B subunit of the heteromeric CCAAT-binding factor. A DNA-binding domain and a subunit interaction domain are specified by two separate segments. J Biol Chem 267: 8286-8292

Mantovani R (1998) A survey of 178 NF-Y binding CCAAT boxes. Nucleic Acids Res 26: 1135-1143

Mantovani R (1999) The molecular biology of the CCAAT-binding factor NF-Y. Gene 239: 15-27

McNabb DS, Xing Y, Guarente L (1995) Cloning of yeast HAP5: a novel subunit of a heterotrimeric complex required for CCAAT binding. Genes Dev 9: 47-58

Miyoshi K, Ito Y, Serizawa A, Kurata N (2003) OsHAP3 genes regulate chloroplast biogenesis in rice. Plant J 36: 532-540

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29: e45

Pinkham JL, Olesen JT, Guarente LP (1987) Sequence and nuclear localization of the Saccharomyces cerevisiae HAP2 protein, a transcriptional activator. Mol Cell Biol 7: 578-585

Rasmussen R (2001) Quantification on the LightCycler instrument. In: Meuer S, Wittwer C, Nakagawara K (eds) Rapid Cycle Real-time PCR, Methods and Applications, pp. 21-34. Springer Press, Heidelberg.

Retief JD (2000) Phylogenetic analysis using PHYLIP. Methods Mol Biol 132: 243-258

Riano-Pachon DM, Ruzicic S, Dreyer I, Mueller-Roeber B (2007) PlnTFDB: An integrative plant transcription factor database. BMC Bioinformatics 8: 42

Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman

BK, Yu G (2000) Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science 290: 2105-2110

Romier C, Cocchiarella F, Mantovani R, Moras D (2003) The NF-YB/NF-YC structure gives insight into DNA binding and transcription regulation by CCAAT factor NF-Y. J Biol Chem 278: 1336-1345

Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18: 502-504

Shinozaki K, Yamaguchi-Shinozaki K, Seki M (2003) Regulatory network of gene expression in the drought and cold stress responses. Curr Opin Plant Biol 6: 410-417

Sinha S, Kim IS, Sohn KY, de Crombrugghe B, Maity SN (1996) Three classes of mutations in the A subunit of the CCAAT-binding factor CBF delineate functional domains involved in the three-step assembly of the CBF-DNA complex. Mol Cell Biol 16: 328-337

Sinha S, Maity SN, Lu J, de Crombrugghe B (1995) Recombinant rat CBF-C, the third subunit of CBF/NFY, allows formation of a protein-DNA complex with CBF-A and CBF-B and with yeast HAP2 and HAP3. Proc Natl Acad Sci U S A 92: 1624-1628

Song W, Solimeo H, Rupert RA, Yadav NS, Zhu Q (2002) Functional dissection of a Rice Dr1/DrAp1 transcriptional repression complex. Plant Cell 14: 181-195

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25: 4876-4882

Xing Y, Zhang S, Olesen JT, Rich A, Guarente L (1994) Subunit interaction in the CCAAT-binding heteromeric complex is mediated by a very short alpha-helix in HAP2. Proc Natl Acad Sci U S A 91: 3009-3013

Xue GP, Bower NI, McIntyre CL, Riding GA, Kazan K, Shorter R (2006) TaNAC69 from the NAC superfamily of transcription factors is up-regulated by abiotic stresses in wheat and recognises two consensus DNA-binding sequences. Functional Plant Biology 33: 43-57

Xue GP, Loveridge CW (2004) HvDRF1 is involved in abscisic acid-mediated gene regulation in barley and produces two forms of AP2 transcriptional activators, interacting preferably with a CT-rich element. Plant J 37: 326-339

Zemzoumi K, Frontini M, Bellorini M, Mantovani R (1999) NF-Y histone fold alpha1 helices help impart CCAAT specificity. J Mol Biol 286: 327-337

Zhu JK (2002) Salt and drought stress signal transduction in plants. Annu Rev Plant Biol 53: 247-273

Table 1 *Triticum aestivum* NF-Y proteins identified in the sequence databases. Either the GenBank accession or the Tentative Consensus (TC) identifier from TaGI is provided. Sequence identifiers in parentheses are redundant and are represented in this study by the non-parenthesised sequence.

Name	GenBank/TC	Name	GenBank/TC
NF-YA			
TaNF-YA1	BT008936	*TaNF-YB3*	BT009265
TaNF-YA2	BT009063		(TC238716)
TaNF-YA3	BT009512	*TaNF-YB4*	TC247628
	(TC270348)		(BT009393)
TaNF-YA4	BT009594	*TaNF-YB5*	CK203103
TaNF-YA5	TC240846	*TaNF-YB6*	CV776390
	(BT009624)	*TaNF-YB7*	TC238715
	(CK207902)	*TaNF-YB8*	TC238717
TaNF-YA6	DR739322	*TaNF-YB9*	TC240894
*TaNF-YA7**	TC238244	*TaNF-YB10*	TC248171
	(AY456087)	*TaNF-YB11*	TC269816
	(AY568307)	**NF-YC**	
	(AY568304)	*TaNF-YC1*	AL829454
	(AY568299)	*TaNF-YC2*	BT008988
	(AY568298)		(TC266361)
	(AY568297)	*TaNF-YC3*	BT009224
	(AY568295)	*TaNF-YC4*	DN829033
	(AY568303)**	*TaNF-YC5*	DR738968
*TaNF-YA8**	TC253181	*TaNF-YC6*	TC233433
	(AY568306)	*TaNF-YC7*	TC237647
	(AY568305)	*TaNF-YC8*	TC241235
	(AY568302)	*TaNF-YC9*	TC255016
	(AY568301)	*TaNF-YC10*	TC266360
	(AY569300)	*TaNF-YC11*	TC268430
TaNF-YA9	TC253407	*TaNF-YC12*	TC268432
TaNF-YA10	TC256735	*TaNF-YC13*	BJ308764
	(BT009542)	*TaNF-YC14*	TC270995
NF-YB		**Dr1**	
TaNF-YB1	BT009029	*TaDr1A*	TC236077
	(AY058921)		(AF464903)
TaNF-YB2	BT009078	*TaDr1B*	TC236076
	(TC247302)		(BT009234)

* The assembled *TaNF-YA7* and *TaNF-YA8* genes may include paralogues in addition to homoeologues.
** Alternative spliced form of the assembled *TaNF-YA7*.

Figure legends

Fig. 1 *Triticum aestivum* TaNF-Y family protein sequence alignment. All sequences identified from the current sequence databases are designated TaNF-YA 1-10 (A); TaNF-YB1-11 (B); TaNF-YC1-14 (C); TaDr1A-B (D). Sequence alignments were created using the CLUSTAL X (v1.83) program. Dashes indicate gaps in the sequences. Asterisks indicate positions that have a single, fully conserved residue. Semicolons indicate strongly conserved residues. Periods indicate weakly conserved residues. Amino acids are coloured to reveal the consensus levels of physicochemically related amino acids. Red boxes indicate conserved core regions. Black boxes indicate the locations of possible motifs found using MEME 3.5.3. Within the conserved core regions: black lines under the alignments indicate regions involved in contacting DNA, green lines under the alignments indicate regions involved in heterodimerization and blue lines under the alignments indicate NF-YA interaction regions.

Fig. 2 *Arabidopsis*, rice, and wheat NF-Y subunit conserved core consensus sequence logos. NF-YA subunit family (A); NF-YB subunit family (B); NF-YC subunit family (C), respectively. Sequence logos were created using WebLogo (http://weblogo.berkeley.edu/) (Crooks et al. 2004). The NF-YA logo was created with 16 rice, 14 *Arabidopsis* and 10 wheat sequences. The NF-YB logo was created with 13 rice, 10 *Arabidopsis* and 11 wheat sequence. The NF-YC logo was created with 19 rice, 18 *Arabidopsis* and 14 wheat sequences. The *Arabidopsis* and rice sequence are the non-redundant sequence members collected from the RiceTFDB and the ArabTFDB. Below the logos is a text representation of the majority consensus created from the three species. Overall height in each stack indicates the sequence conservation at that point. Height of each residue letter indicates relative frequency of the corresponding residue at that position. Amino acids are coloured according to their chemical properties: polar amino acids (G,S,T,Y,C,Q,N) are green, basic (K,R,H) blue, acidic (D,E) red and hydrophobic (A,V,L,I,P,W,F,M) amino acids are black.

Fig. 3 Motifs outside of the conserved core domains in NF-YA. GVVAAY (A); RVPLP(B); DPYYG (C); HPQI (D). Motifs identified with the use of MEME 3.5.3 (http://meme.sdsc.edu). Alignments produced with CLUSTALX (v1.83). Alignments are of *Arabidopsis* (starts with At), rice (starts with #) and wheat (starts with Ta) NF-YA subunits containing motifs depicted in the logos above. Four motifs were found to be common to the three plant species in the NF-YA subunit family. Below each is a text representation of the consensus sequence.

Fig. 4 Motifs outside of the conserved core domains in NF-YB. REQDRF (A); KSGDGSVKKD (B); QPQYH (C). Motifs identified with the use of MEME 3.5.3 (http://meme.sdsc.edu). Alignments produced with CLUSTALX (v1.83). Alignments are of *Arabidopsis*, rice and wheat NF-YB subunits containing motifs depicted in the logos above. Three motifs were found to be shared between the three plant species in the NF-YB subunit family. Below each is a text representation of the consensus sequence.

Fig. 5 Motifs outside of the conserved core domains in NF-YC. DFKNH (A); QQQQQQLQxFW (B); VPRDEAK (C); PYYYPP (D). Motifs identified with the use of MEME 3.5.3 (http://meme.sdsc.edu). Alignments produced with CLUSTALX (v1.83). Alignments are of *Arabidopsis*, rice and wheat NF-YC subunits containing motifs depicted in the logos above. Four motifs were found to be common to the three plant species in the NF-YC subunit family. Below each is a text representation of the consensus sequence.

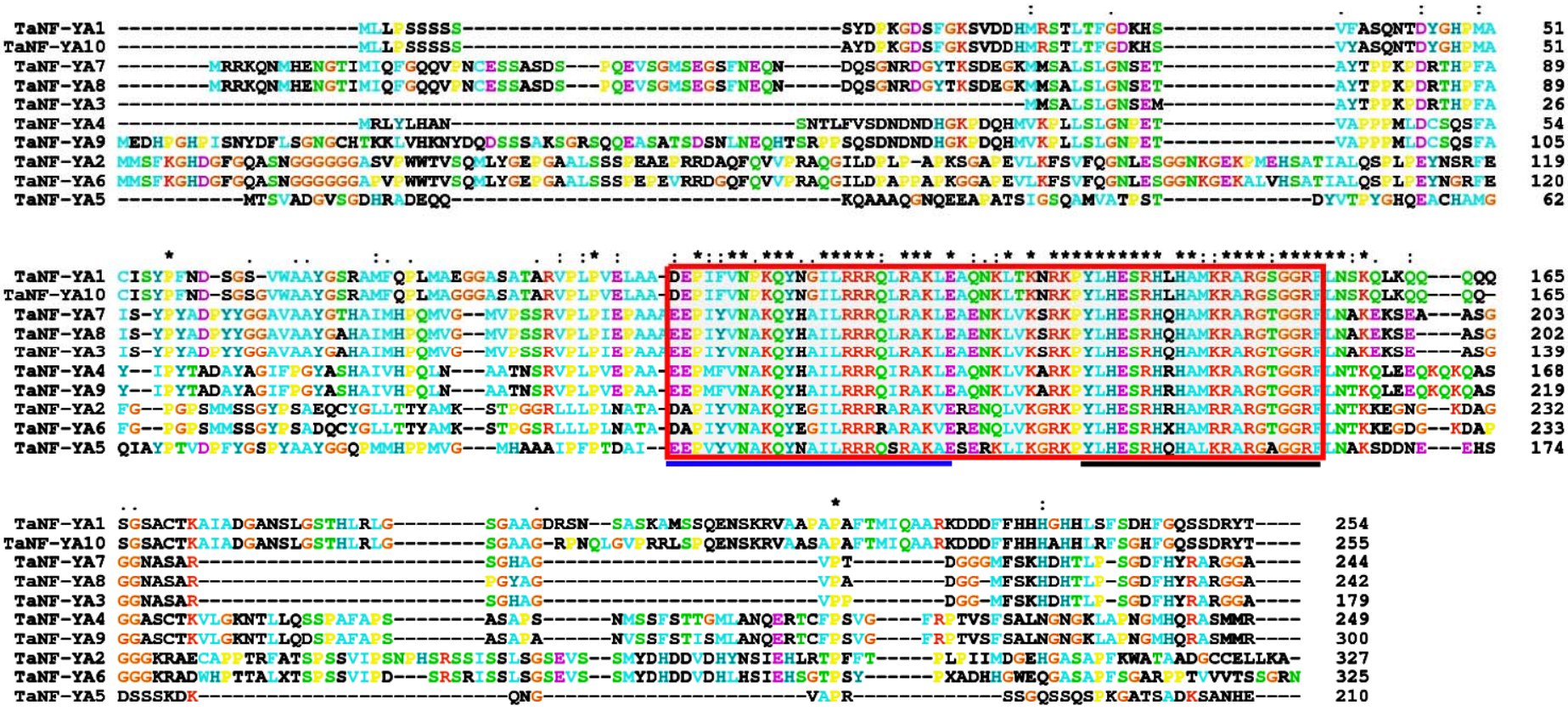
Fig. 6 Phylogenetic trees of TaNF-Y subunit families. TaNF-YA conserved core tree (A); Full-length TaNF-YA gene family tree (B); TaNF-YB conserved core tree (C); TaNF-YB full-length tree (D); TaNF-YC conserved core tree (E); TaNF-YC full-length tree (F). Each tree created with the PHYLIP program package (Retief 2000) using the Neighbor-joining method. Bootstrap values from 1000 replicates have been used to assess the robustness of the trees. Bootstrap values are shown in red. Each tree has been rooted using the *Saccharomyces cerevisiae* HAP homologues. Identifiers I-V have been used to indicate sequences which cluster with support from high bootstrap values.

Fig. 7 Expression profiles of NF-Y genes in wheat. The organ specificity of TaNF-YA genes (A); the organ specificity of TaNF-YB genes (B); the organ specificity of TaNF-YC genes (C); the organ specificity of TaDr1 genes. Expression level is expressed relative to the root on a logarithmic scale. Error bars indicate SD of the mean of three biological samples. Each sample was analysed with triplicate real-time PCR assays. Real-time primers for *TaNF-YA7* also amplify *TaNF-YA8*.

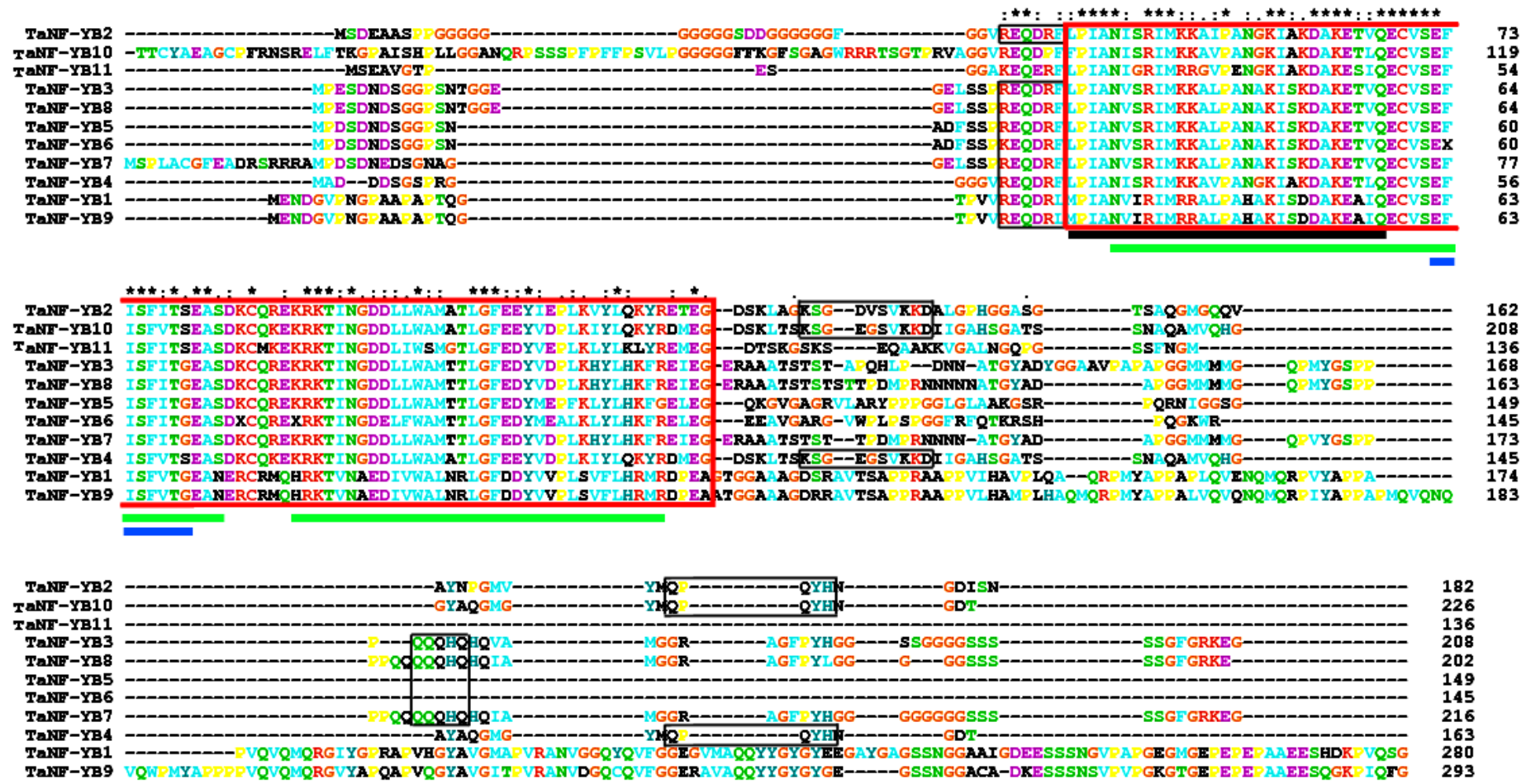
Fig. 8 Changes in the mRNA levels of wheat NF-Y genes in the drought-stressed leaves. Transcript level is expressed relative to the levels in non-stressed control leaves. Insert graph shows there was no significant difference in expression levels for the control genes (TaRPII36 and TaRP15; the relative expression levels were normalised with another internal control gene TaSnRK1). The relative leaf water contents of the drought-stressed plants ranged from 75%-81%. Error bars indicate SD of the mean of three biological samples. Each sample was analysed with triplicate PCR assays. Double asterisks indicate statistically significant differences with $P \leq 0.01$ and triple asterisks indicate statistically significant differences with $P \leq 0.001$ using Students *t*-test.

Fig. 9 Correlation between gene expression of TaNF-Y genes across six wheat organs. Correlated gene expression between *TaNF-YB3*, *TaNF-YB7* and *TaNF-YB8* (A); *TaNF-YA3* and *TaNF-YA4* (B); *TaNF-YC11* and *TaNF-YC12* (C); *TaNF-YA5* and *TaNF-YC8* (D); *TaNF-YA9* and *TaNF-YB10* (E); *TaNF-YB3* and *TaNF-YC12* (F). Relative expression levels are used for each gene. These data have been fitted using linear regression analysis.

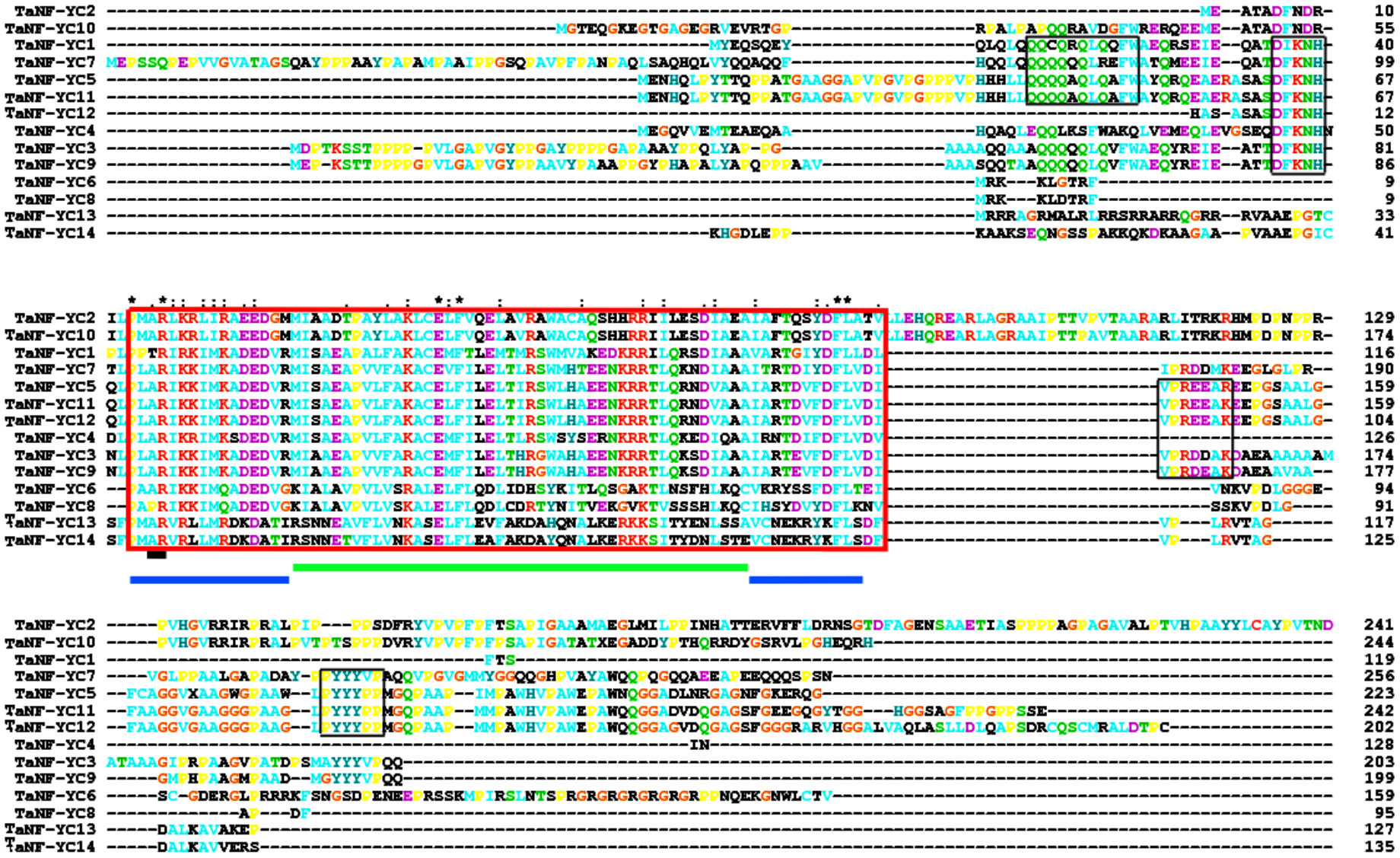
(A)


TaNF-YA1
TaNF-YA10
TaNF-YA7
TaNF-YA8
TaNF-YA3
TaNF-YA4
TaNF-YA9
TaNF-YA2
TaNF-YA6
TaNF-YA5

(B)


TaNF-YB2
TaNF-YB10
TaNF-YB11
TaNF-YB3
TaNF-YB8
TaNF-YB5
TaNF-YB6
TaNF-YB7
TaNF-YB4
TaNF-YB1
TaNF-YB9

(C)


TaNF-YC2
TaNF-YC10
TaNF-YC1
TaNF-YC7
TaNF-YC5
TaNF-YC11
TaNF-YC12
TaNF-YC4
TaNF-YC3
TaNF-YC9
TaNF-YC6
TaNF-YC8
TaNF-YC13
TaNF-YC14

(D)

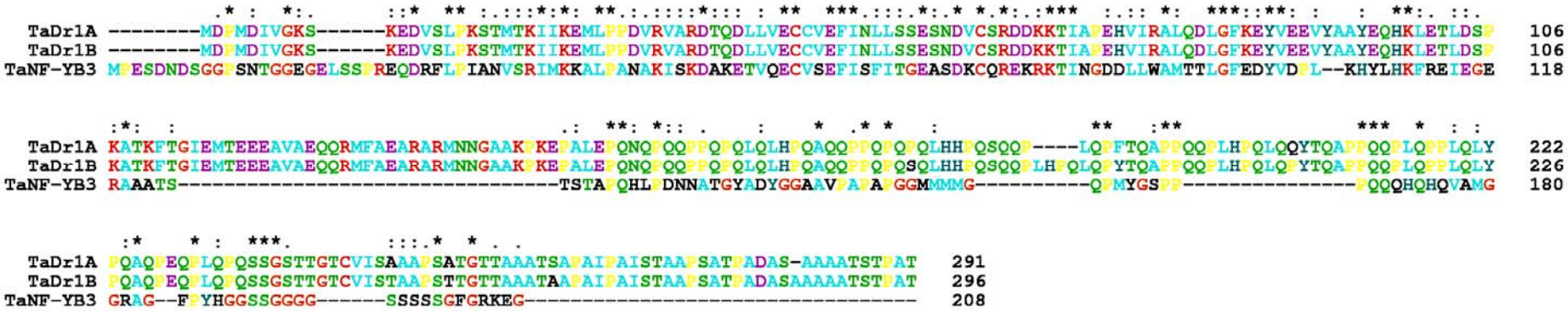


Fig. 1 *Triticum aestivum* TaNF-Y family protein sequence alignment. All sequences identified from the current sequence databases are designated TaNF-YA 1-10 (A); TaNF-YB1-11 (B); TaNF-YC1-14 (C); TaDr1A-B (D). Sequence alignments were created using the CLUSTAL X (v1.83) program. Dashes indicate gaps in the sequences. Asterisks indicate positions that have a single, fully conserved residue. Semicolons indicate strongly conserved residues. Periods indicate weakly conserved residues. Amino acids are coloured to reveal the consensus levels of physicochemically related amino acids. Red boxes indicate conserved core regions. Black boxes indicate the locations of possible motifs found using MEME 3.5.3. Within the conserved core regions: black lines under the alignments indicate regions involved in contacting DNA, green lines under the alignments indicate regions involved in heterodimerization and blue lines under the alignments indicate NF-YA interaction regions.

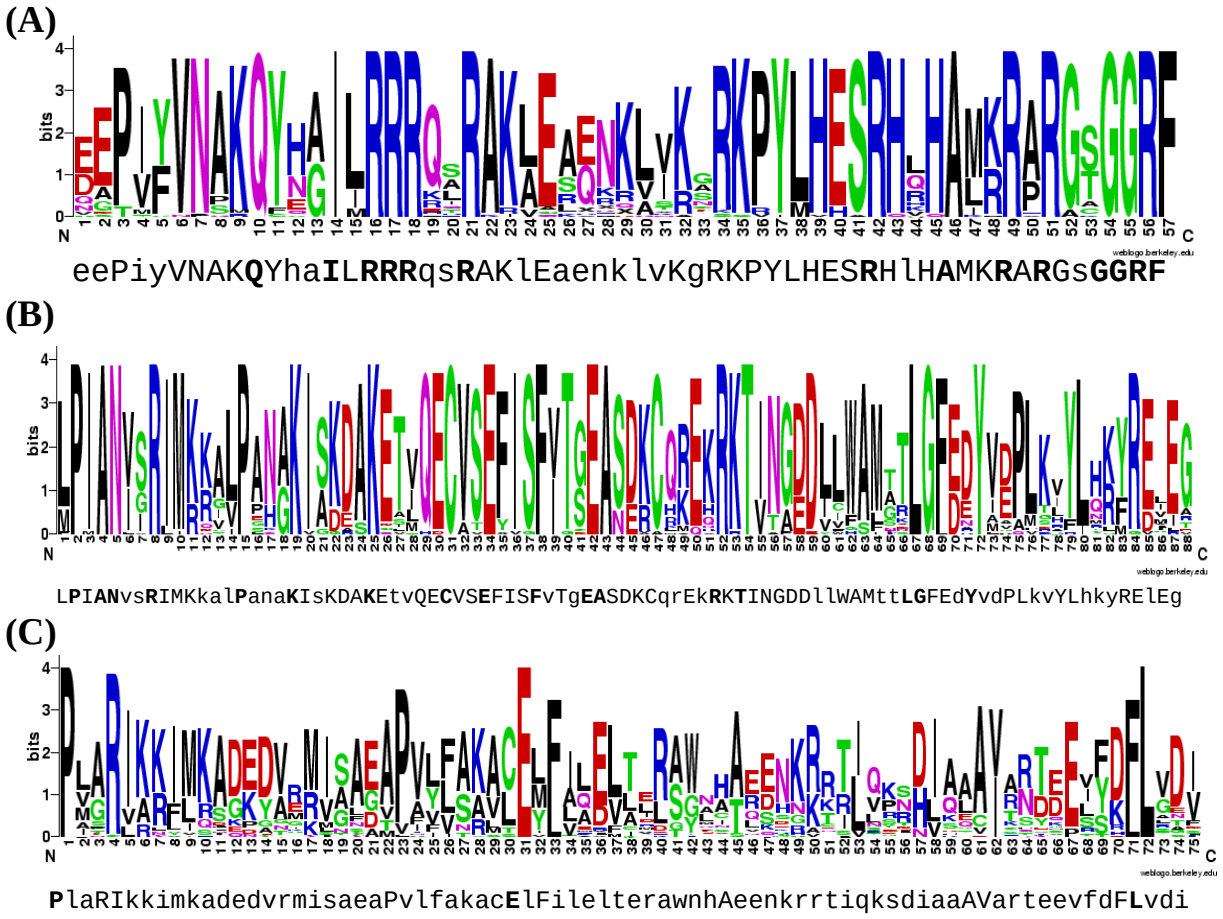


Fig. 2 *Arabidopsis*, rice, and wheat NF-Y subunit conserved core consensus sequence logos. NF-YA subunit family (A); NF-YB subunit family (B); NF-YC subunit family (C), respectively. Sequence logos were created using WebLogo (http://weblogo.berkeley.edu/) (Crooks et al. 2004). The NF-YA logo was created with 16 rice, 14 *Arabidopsis* and 10 wheat sequences. The NF-YB logo was created with 13 rice, 10 *Arabidopsis* and 11 wheat sequence. The NF-YC logo was created with 19 rice, 18 *Arabidopsis* and 14 wheat sequences. The *Arabidopsis* and rice sequence are the non-redundant sequence members collected from the RiceTFDB and the ArabTFDB. Below the logos is a text representation of the majority consensus created from the three species. Overall height in each stack indicates the sequence conservation at that point. Height of each residue letter indicates relative frequency of the corresponding residue at that position. Amino acids are coloured according to their chemical properties: polar amino acids (G,S,T,Y,C,Q,N) are green, basic (K,R,H) blue, acidic (D,E) red and hydrophobic (A,V,L,I,P,W,F,M) amino acids are black.

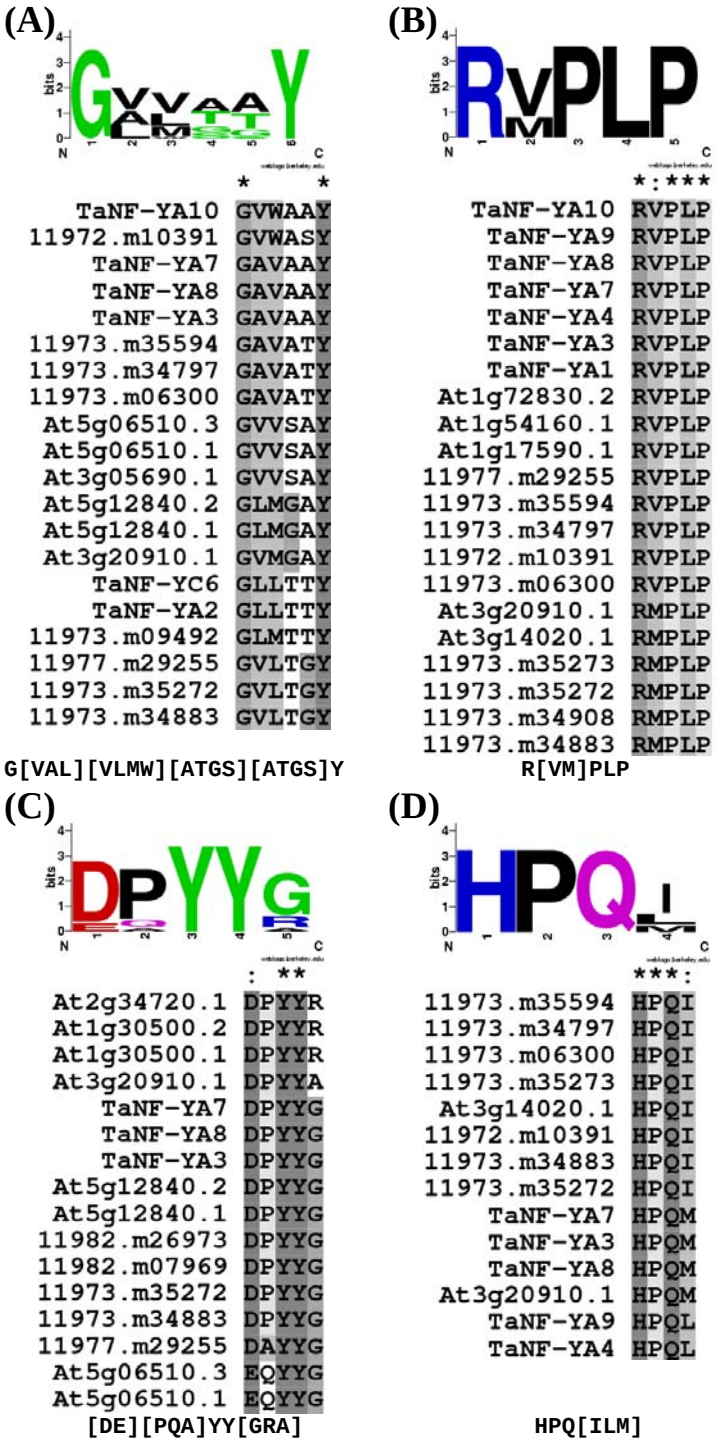


Fig. 3 Motifs outside of the conserved core domains in NF-YA. GVVAAY (A); RVPLP(B); DPYYG (C); HPQI (D). Motifs identified with the use of MEME 3.5.3 (http://meme.sdsc.edu). Alignments produced with CLUSTALX (v1.83). Alignments are of *Arabidopsis* (starts with At), rice (starts with #) and wheat (starts with Ta) NF-YA subunits containing motifs depicted in the logos above. Four motifs were found to be common to the three plant species in the NF-YA subunit family. Below each is a text representation of the consensus sequence.

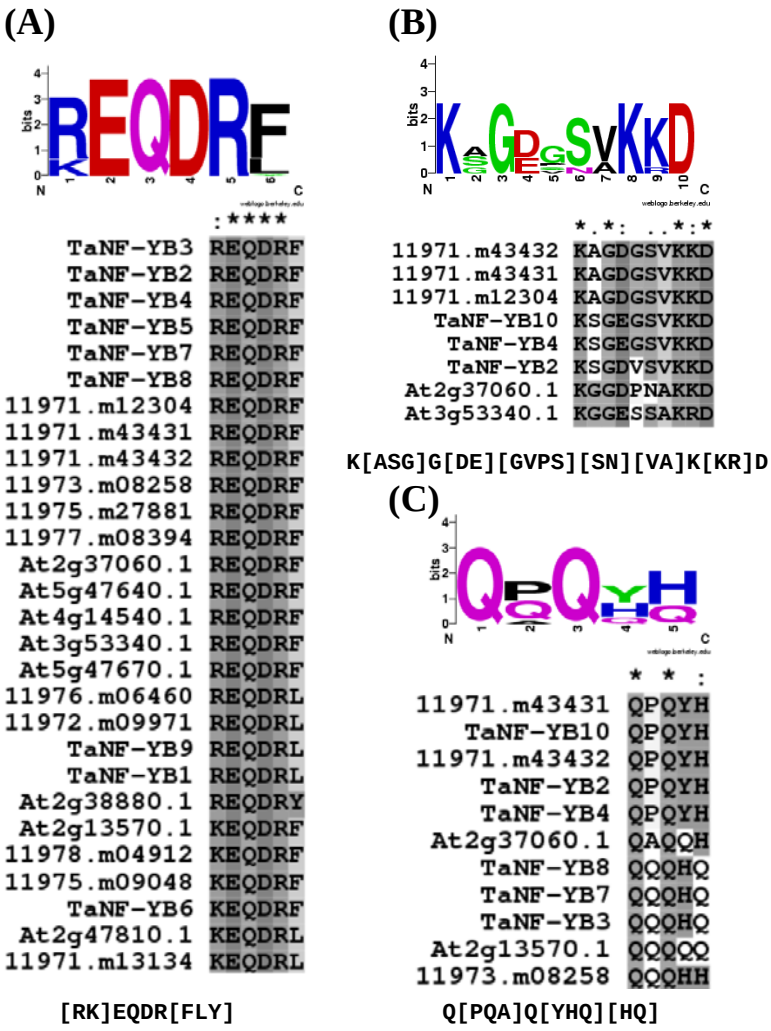


Fig. 4 Motifs outside of the conserved core domains in NF-YB. REQDRF (A); KSGDGSVKKD (B); QPQYH (C). Motifs identified with the use of MEME 3.5.3 (http://meme.sdsc.edu). Alignments produced with CLUSTALX (v1.83). Alignments are of *Arabidopsis*, rice and wheat NF-YB subunits containing motifs depicted in the logos above. Three motifs were found to be shared between the three plant species in the NF-YB subunit family. Below each is a text representation of the consensus sequence.

(A)

```
                *:***
      TaNF-YC4  DFKNH
      TaNF-YC3  DFKNH
      TaNF-YC5  DFKNH
      TaNF-YC7  DFKNH
11972.m06092    DFKNH
     TaNF-YC12  DFKNH
     TaNF-YC11  DFKNH
      TaNF-YC9  DFKNH
11976.m09092    DFKNH
11973.m06906    DFKNH
11978.m07861    DFKNH
11979.m06182    DFKNH
 At1g08970.1    DFKNH
 At1g54830.1    DFKNH
 At1g56170.1    DFKNH
 At3g48590.1    DFKNH
 At5g27910.1    DFKNH
 At5g63470.1    DFKNH
      TaNF-YC1  DIKNH
D[FI]KNH
```

(B)

```
                **    **:  **
11973.m06906    QQQQAQLQAFW
     TaNF-YC11  QQQQAQLQAFW
      TaNF-YC5  QQQQAQLQAFW
11978.m05189    QQQKEQLQAFW
11976.m09092    QQQQRQLQQFW
      TaNF-YC1  QQCQRQLQQFW
 At1g56170.1    QQQQQQLQMFW
 At5g63470.1    QQQQQQLQMFW
11972.m06092    QQQQQQLREFW
      TaNF-YC7  QQQQQQLREFW
 At3g48590.1    QQQQQQLQLFW
 At1g54830.1    QQLTQQLQSFW
 At1g08970.1    QQLAQQLQAFW
QQ[QCL][QKTA][QARE]QL[QR]xFW
```

(C)

```
               ***:  :  :
11972.m06092   VPRDEMK
 At3g48590.1   VPRDEIK
11978.m07861   VPRDEAK
     TaNF-YC9  VPRDEAK
11973.m06906   VPREEAK
    TaNF-YC12  VPREEAK
    TaNF-YC11  VPREEAK
     TaNF-YC6  VPREEAR
 At5g63470.1   VPREEIK
     TaNF-YC3  VPRDDAK
11979.m06182   VPRDDAK
 At1g54830.1   VPREDLR
 At1g08970.1   VPREDLR
11976.m09092   VPRDDLK
VPR[DE][ED][AILM][KR]
```

(D)

```
                ** * *
      TaNF-YC7  PYYYVP
11972.m06092    PYYYVP
      TaNF-YC6  PYYYPP
11973.m06906    PYYYPP
     TaNF-YC12  PYYYPP
     TaNF-YC11  PYYYPP
 At5g63470.1    PYYYPP
 At3g48590.1    PYYYPP
 At1g08970.1    PYGYLP
 At1g54830.1    PYGYLP
PY[YG]Y[PLV]P
```

Fig. 5 Motifs outside of the conserved core domains in NF-YC. DFKNH (A); QQQQQQLQxFW (B); VPRDEAK (C); PYYYPP (D). Motifs identified with the use of MEME 3.5.3 (http://meme.sdsc.edu). Alignments produced with CLUSTALX (v1.83). Alignments are of *Arabidopsis*, rice and wheat NF-YC subunits containing motifs depicted in the logos above. Four motifs were found to be common to the three plant species in the NF-YC subunit family. Below each is a text representation of the consensus sequence.

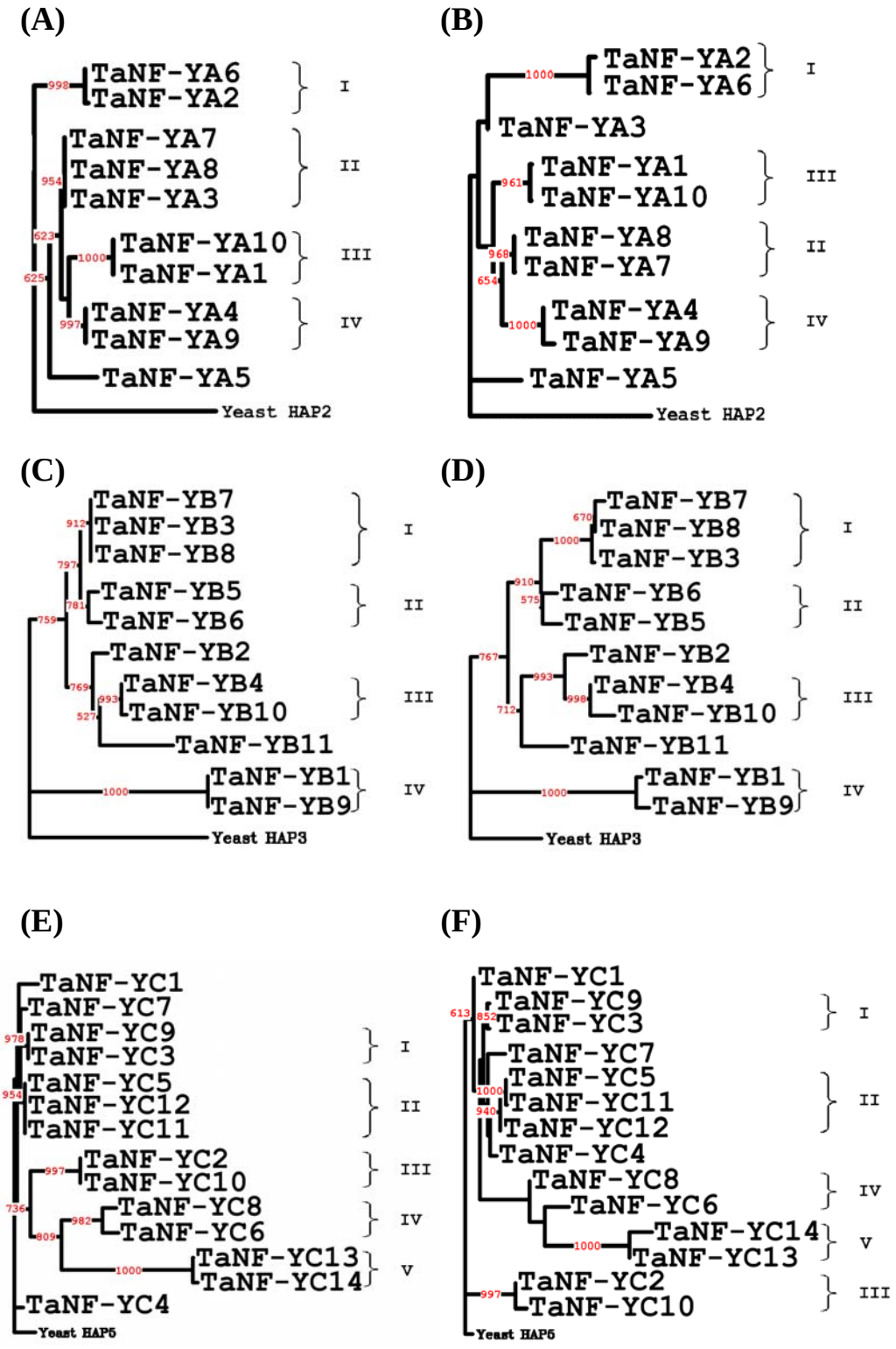
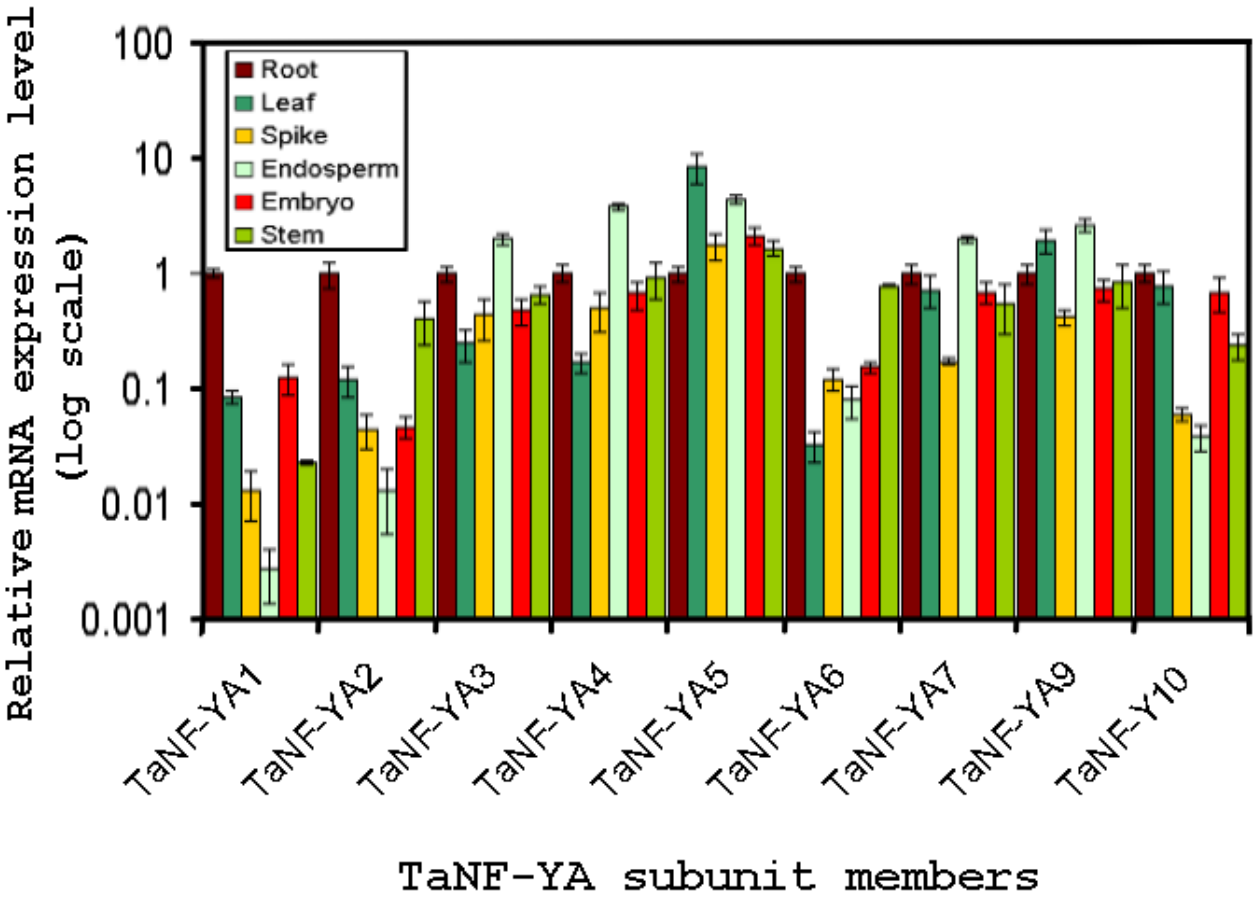
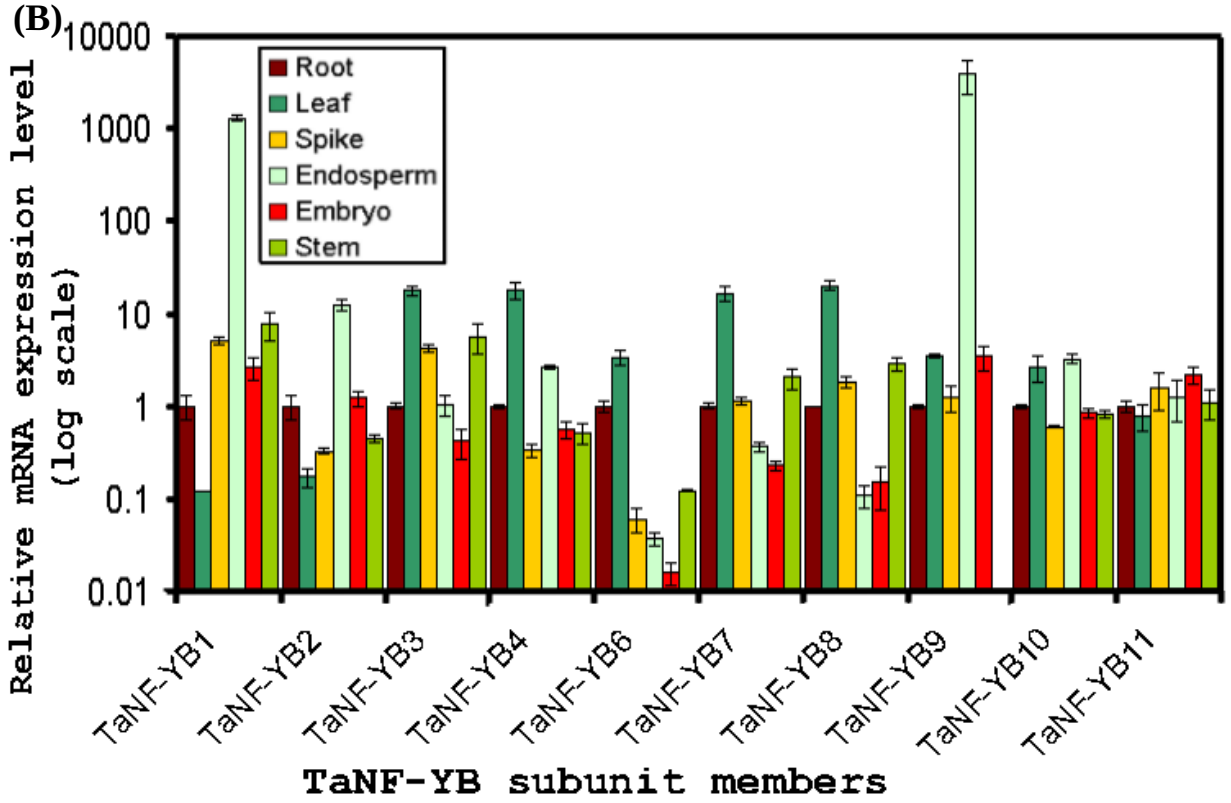


Fig. 6 Phylogenetic trees of TaNF-Y subunit families. TaNF-YA conserved core tree (A); Full-length TaNF-YA gene family tree (B); TaNF-YB conserved core tree (C); TaNF-YB full-length tree (D); TaNF-YC conserved core tree (E); TaNF-YC full-length tree (F). Each tree created with the PHYLIP program package (Retief 2000) using the Neighbor-joining method. Bootstrap values from 1000 replicates have been used to assess the robustness of the trees. Bootstrap values are shown in red. Each tree has been rooted using the *Saccharomyces cerevisiae* HAP homologues. Identifiers I-V have been used to indicate sequences which cluster with support from high bootstrap values.

(A)


Relative mRNA expression level (log scale)
100
10
1
0.1
0.01
0.001
Root
Leaf
Spike
Endosperm
Embryo
Stem
TaNF-YA1
TaNF-YA2
TaNF-YA3
TaNF-YA4
TaNF-YA5
TaNF-YA6
TaNF-YA7
TaNF-YA9
TaNF-Y10
TaNF-YA subunit members

(B)


Relative mRNA expression level (log scale)
10000
1000
100
10
1
0.1
0.01
Root
Leaf
Spike
Endosperm
Embryo
Stem
TaNF-YB1
TaNF-YB2
TaNF-YB3
TaNF-YB4
TaNF-YB6
TaNF-YB7
TaNF-YB8
TaNF-YB9
TaNF-YB10
TaNF-YB11
TaNF-YB subunit members

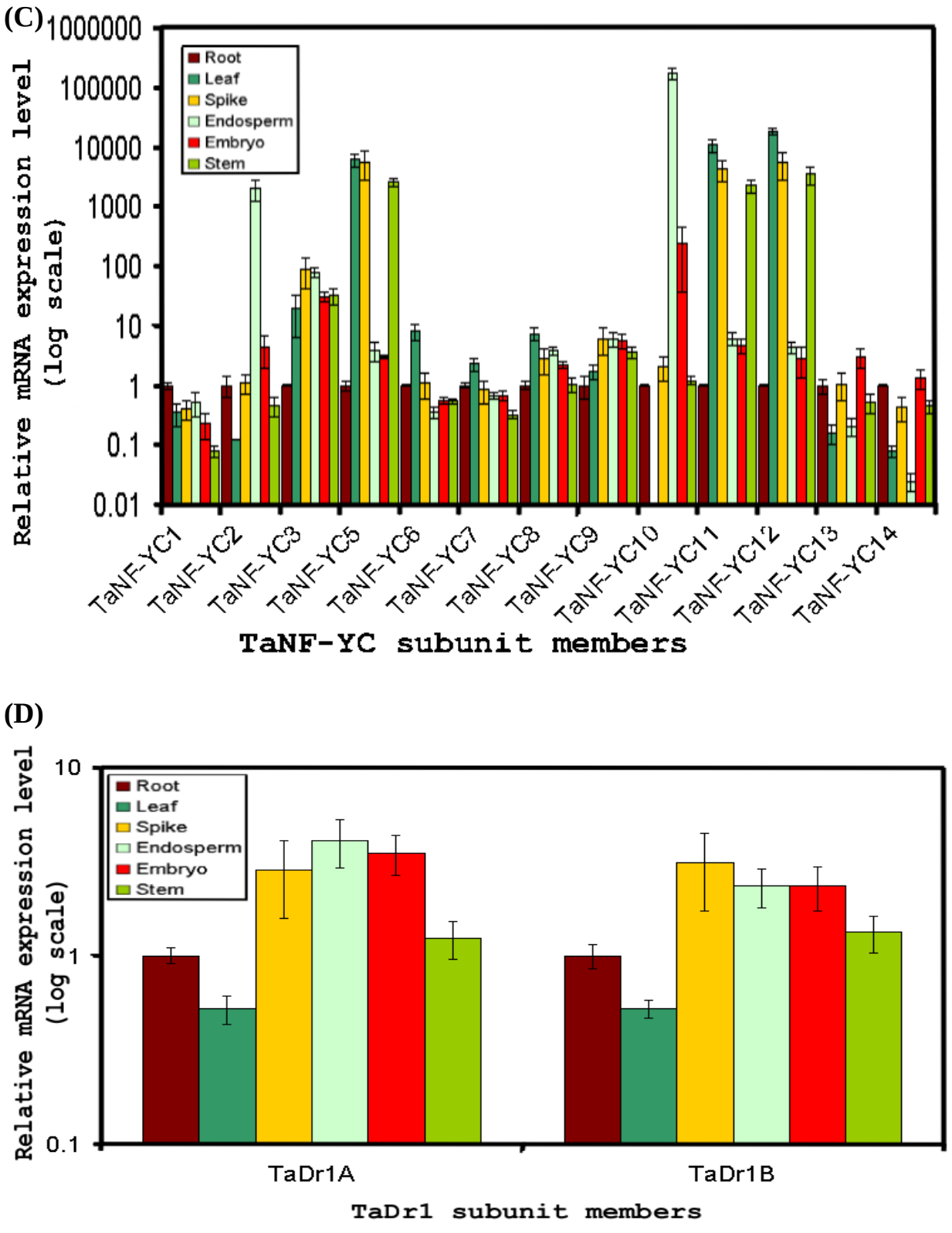


Fig. 7 Expression profiles of NF-Y genes in wheat. The organ specificity of TaNF-YA genes (A); the organ specificity of TaNF-YB genes (B); the organ specificity of TaNF-YC genes (C); the organ specificity of TaDr1 genes. Expression level is expressed relative to the root on a logarithmic scale. Error bars indicate SD of the mean of three biological samples. Each sample was analysed with triplicate real-time PCR assays. Real-time primers for *TaNF-YA7* also amplify *TaNF-YA8*.

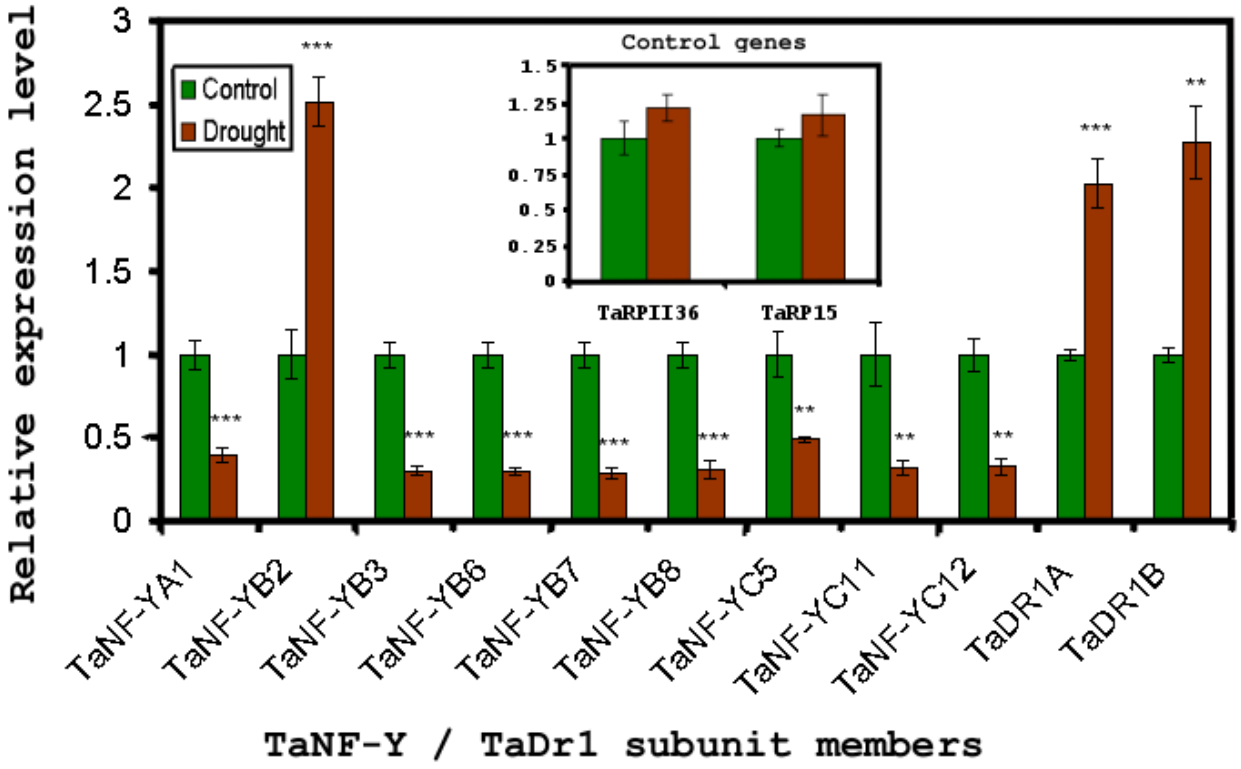


Fig. 8 Changes in the mRNA levels of wheat NF-Y genes in the drought-stressed leaves. Transcript level is expressed relative to the levels in non-stressed control leaves. Insert graph shows there was no significant difference in expression levels for the control genes (TaRPII36 and TaRP15; the relative expression levels were normalised with another internal control gene TaSnRK1). The relative leaf water contents of the drought-stressed plants ranged from 75%-81%. Error bars indicate SD of the mean of three biological samples. Each sample was analysed with triplicate PCR assays. Double asterisks indicate statistically significant differences with $P \leq 0.01$ and triple asterisks indicate statistically significant differences with $P \leq 0.001$ using Students *t*-test.

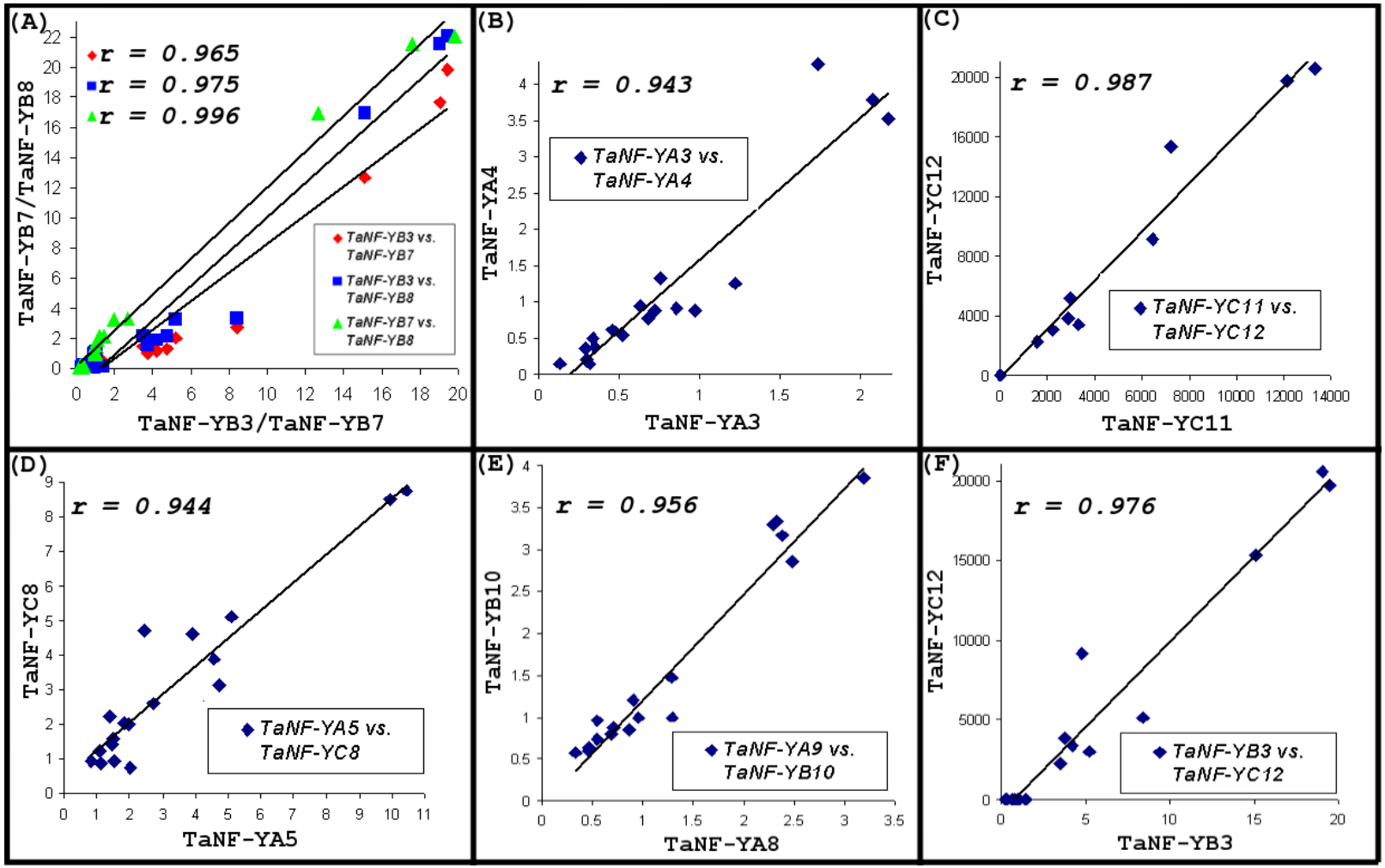


Fig. 9 Correlation between gene expression of TaNF-Y genes across six wheat organs. Correlated gene expression between *TaNF-YB3*, *TaNF-YB7* and *TaNF-YB8* (A); *TaNF-YA3* and *TaNF-YA4* (B); *TaNF-YC11* and *TaNF-YC12* (C); *TaNF-YA5* and *TaNF-YC8* (D); *TaNF-YA9* and *TaNF-YB10* (E); *TaNF-YB3* and *TaNF-YC12* (F). Relative expression levels are used for each gene. These data have been fitted using linear regression analysis.