Open access • Journal Article • DOI:10.1042/BJ20091234

The Arabidopsis thaliana NAC transcription factor family: structure–function relationships and determinants of ANAC019 stress signalling — Source link

Michael Krogh Jensen, Trine Kjaersgaard, Michael Martin Nielsen, Pernille Galberg ...+3 more authors

Institutions: University of Copenhagen

Published on: 01 Mar 2010 - Biochemical Journal (Portland Press)

Topics: DNA-binding domain

The Arabidopsis thaliana NAC transcription factor family: structure-function relationships and determinants of ANAC019 stress signaling

Michael K Jensen, Trine Kjaersgaard, Michael M Nielsen, Pernille Galberg, Klaus Petersen, Charlotte O'Shea, Karen Skriver

The *Arabidopsis thaliana* NAC Transcription Factor Family: Structure-Function Relationships and Determinants of ANAC019 Stress Signaling

Michael K. JENSEN, Trine KJAERSGAARD, Michael M. NIELSEN, Pernille GALBERG, Klaus PETERSEN, Charlotte O'SHEA, and Karen SKRIVER*

Department of Biology, University of Copenhagen, Ole Maaloesvej 5, DK-2200 Copenhagen N, Denmark
*To whom correspondence should be addressed: Karen Skriver, KSkriver@bio.ku.dk, Phone: +45 35321712

SYNOPSIS

Transcription factors (TFs) are modular proteins minimally containing a DNA-binding domain (DBD) and a transcription regulatory domain (TRD). NAC proteins comprise one of the largest plant TF-families. They are key regulators of stress perception and developmental programs, and most share an N-terminal NAC domain. Based on analyses of gene expression data and phylogeny of *Arabidopsis thaliana* NAC TFs we systematically decipher structural and functional specificities of the conserved NAC domains and divergent C-termini. Nine of ten NAC domains analyzed bind a previously identified NAC DNA-target sequence with a CGT[GA] core, although with different affinities. Likewise, all but one of the NAC proteins analyzed is dependent on the C-terminal region for transactivational activity. *In silico* analyses show that NAC TRDs contain group-specific sequence motifs and are characterized by a high degree of intrinsic disorder. Furthermore, ANAC019 was identified as a new positive regulator of abscisic acid (ABA)-signaling, conferring ABA-hypersensitivity when ectopically expressed in plants. Interestingly, ectopic expression of ANAC019 DBD or TRD alone also resulted in ABA-hypersensitivity. Expression of stress-responsive marker genes (*COR47*, *RD29b*, *ERD11*) was also induced by full-length and truncated ANAC019. Domain-swapping was used to analyze the specificity of this function. Replacement of the NAC domain of ANAC019 with analogous regions of other NAC TFs resulted in chimeric proteins, which also have the ability to regulate ABA-signaling positively. In contrast, replacing the ANAC019 TRD with other TRDs abolished ANAC019-mediated ABA-hypersensitivity. In conclusion, our results demonstrate that biochemical and functional specificity is associated with both DBDs and TRDs in NAC TFs.

Key words: NAC transcription factor, DNA-binding, Intrinsic disorder, Domain-swap, Abscisic acid (ABA), *Arabidopsis*

Short title: Structural Determinants of *Arabidopsis* ANAC019 Function

ABBREVIATIONS

ABA	Abscisic acid
ABRC	Arabidopsis biological resource center
AGI	Arabidopsis genome initiative
CaMV	Cauliflower mosaic virus
CASTing	Cyclic amplification and selection of targets
Col-0	Columbia ecotype 0
CUC	CUP-SHAPED COTYLEDON (CUC)
DBD	DNA-binding domain
EMBL	European molecular biology laboratory
EMSA	Electrophoretic Mobility Shift Assays
GST	Glutathione-S-Transferase
ID	Intrinsic disorder
NAC	NAM/ATAF1,2/CUC2
MEME	Multiple EM for motif elicitation
MAST	Multiple alignment and search tools
MORF	Molecular Recognition Features
NACBS	NAC-binding site
palNACBS	Palindromic NACBS
TRD	Transcription regulatory domain
TF	Transcription factor
Rap1p	Repressor activator protein1p
TAIR	The Arabidopsis Information Resource

INTRODUCTION

Gene specific transcription factors (TFs) are DNA-binding regulatory proteins which activate or repress the basal transcription apparatus at target gene promoters. TFs are grouped into families based on sequence similarities, most often in the DNA-binding domain (DBD)[1]. TFs from the same family often bind DNA in a sequence-specific manner thereby only targeting promoters with a given consensus sequence [1]. Apart from the DBD, TFs are characterized by having a transcription regulatory domain (TRD). In contrast to the sequence similarities and well-defined structures of DBDs in members of a specific TF family, TRDs have traditionally been classified according to their amino acid profile as acidic, glutamine-, proline-, or serine/threonine-rich. Moreover, TRDs often have a high degree of low-complexity sequences and propensity for flexible protein segments that fail to self-fold into an ordered three-dimensional structure, commonly referred to as intrinsic disorder (ID) [2].

Ongoing efforts to understand the structural and functional modularity of TFs have traditionally been based on examination of the activity of truncated versions of the proteins. Early in the dissection of TFs, it was found that the DBD could be separated from the TRD with no loss of function of either module, as illustrated by the yeast GAL4 TF; a paradigm for eukaryotic transcriptional regulation studies [3]. Additionally, Johnston *et al*. [3] showed that expression of a cytoplasmicly located version of GAL4, without the N-terminal DBD, in a wild-type background enabled induction of the expression of reporter genes to the same extent as expression of full-length GAL4. This implied that GAL4 consists of separate domains conferring nuclear localization, protein-protein interactions, DNA-binding, and transcriptional activation [3]. This seemingly independent modularity has proved general to many TFs [4;5], and has led to the question as to how functional specificity is substantiated.

Though a wealth of studies has aimed to characterize TF functionality, a few of them should be highlighted as important attempts to shed light on functional specificity from analysis of independent TF modules. In yeast, this question has been addressed by deciphering regions within the transcriptional activator Rap1p (repressor activator protein1p), showing that the C-terminal TRD has only a minor effect on cell growth modulated by Rap1p [6]. Furthermore, domain swap experiments with combinations of elements from DBDs of Rap1p homologues from different yeasts revealed that major changes can be made to the amino acid sequence of this region without affecting Rap1p function in telomere binding [6]. In another study, using a series of domain-swap chimeres in which different parts of *Drosophila melanogastor* T-domain TFs, encoded by the genes *optomotor-blind* (*omb*) and *optomotor-blind related-1* (*org-1*), were mutually exchanged, Porsch and co-workers investigated the relevance of individual domains for proper eye development [7]. Their findings suggested that both transcriptional activation and repression properties as well as DNA-binding specificity can contribute to the functional characteristics of T-domain factors [7], highlighting the complexity of parsing the functional specificity of seemingly independent TF domains.

In plants, the NAC (NAM/ATAF1,2/CUC2) family constitutes a prominent group of TFs [8]. The genomes of *Arabidopsis thaliana* (thale cress), tobacco and rice all contain more than 100 genes encoding NAC TFs, making it one of the largest gene families in plants [9;10]. Genes encoding NAC TFs were originally identified from forward genetic screens as key regulators of developmental processes [8]. More recently, NAC TFs have also been shown to be involved in regulation of stress responses in both model plants and agronomically important crops [11;12].

We have previously identified the NAC protein ANAC019 as an interactor of the RING (Really Interesting New Gene)-H2-type E3 ubiquitin-protein ligase, RHA2a, and solved the crystal structure of the ANAC019 NAC domain [13;14]. This revealed a novel protein fold consisting of a twisted β-sheet packing against an α-helix on both sides. Moreover, using ANAC019 as a reference we have elucidated residues important for both NAC homo- and heterodimerization and DNA-binding to the core consensus element CGT[GA] [15]. Studies of *ANAC019* gene expression have implicated a role in different types of stress responses [14;16] and over-expression of *ANAC019* caused increased drought tolerance [16]. Moreover, it has been shown that ANAC019 regulates defense responses upon fungal attack through a signal pathway related to the plant hormone jasmonic acid [17]. Finally, Bu *et al.* [18], more recently verified our result on the ANAC019-RHA2a interaction and coined RHA2a as a new positive regulator of abscisic acid (ABA) phytohormone signaling [18].

With respect to parsing functional determinants of NAC DBDs and TRDs, Taoka *et al.* [19], have shown that for members of the *CUP-SHAPED COTYLEDON* (*CUC*) sub-family of NAC TFs the DBD confers functional specificity in terms of inducing adventitious shoot formation on calli [19;20]. They showed that TRDs of three different NAC TF origins could be fused to the NAC domain of CUC2 and still induce shoot formations. Contrastingly, substituting CUC NAC domains with a NAC domain from a member of the ATAF subfamily of NAC TFs did not induce shoot formation. However, the NAC domain of the CUC subfamily member, CUC1, did not retain shoot formation ability when fused to the potent TRD of the herpes simplex virus VP16 protein. This suggested that certain structural features of the TRD which are difficult to predict from the amino acid sequence were necessary for the function of CUC [19].

In this study, we have used a genome-wide approach taking advantage of the functional diversity of the NAC proteins to dissect structure-function aspects of NAC TF modularity. Our analyses showed that phylogenetic relationships based on NAC domain sequences generally are in accordance with C-terminal sequence motifs of intrinsically disordered TRDs and global NAC gene expression clusters. Additionally, our study includes examination of truncated versions of ten phylogenetically distinct NAC proteins to characterize the structural specificity of NAC DBDs and TRDs, and a chimeric-strategy based on our identification of ANAC019 as a novel ABA regulator, to investigate structure-function relationship of ANAC019 DBD and TRD. The findings suggested multi-specificity of NAC TFs in terms of DNA-binding sites of target genes and provide evidence for positive ABA-regulatory functionality to be associated with both the ANAC019 TRD and DBD.

EXPERIMENTAL

Oligonucleotides and restriction enzymes
Oligonucleotides and restriction enzymes used in this paper are listed in Supplementary Table S1. DNA sequence integrity of all expression constructs was confirmed by sequencing.

Accession numbers
Sequence data from this article can be found in TAIR9 (The Arabidopsis Information Resource) and European Molecular Biology Laboratory (EMBL) data libraries under the following NAC nomenclature names and accession numbers: ANAC019, At1g52890; ANAC003, At1g02220; ANAC002 (ATAF1), At1g01720; VOZ2; At2g42400; ANAC008 (SOG1), At1g25580; ANAC105 (VND3), At5g66300; ANAC030 (VND7), At1g71930; ANAC092 (NAC2/ORE1), At5g39610; ANAC062 (NTL6), At3g49530; ANAC040 (NTL8), At2g27300.

Bioinformatic analysis of the NAC family
BLASTP and PSI-BLAST searches were performed using the NCBI BLAST services (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Searches were performed against a non-redundant database but limited to include only hits within the *Arabidopsis* genome. All sequences were annotated according to the latest *Arabidopsis* genome annotation, TAIR9 (http://www.arabidoposis.org). Multiple alignments were generated using ClustalW or ClustalX [21] followed by BoxShade (http://www.ch.embnetorg/software/BOX_form.html) for producing graphical presentations of the alignments and manual adjustments. Phylogenetic analysis of NAC domain sequences was performed using MEGA4 software [22]. The input was a multiple alignment of NAC domain sequences generated in ClustalX. The reliability of the different groupings of the tree was measured by bootstrapping. MEME (Multiple EM for motif Elicitation) was used to search for statistically significant motifs in the NAC proteins using either all NAC sequences or group and sub-group specific sequences as data set. MAST (Motif Alignment & Search Tool) was used for further analysis of the NAC proteins. P-values were used to determine the overall match of the sequence to the input motif. In addition, the relevance of the motifs identified was evaluated by analysis of its occurrence in control groups, e.g. the NAC domain served as a control group for motifs identified in the C-terminal regions and *vice versa*. Amino acid profiling was performed using Composition Profiler (http://www.cprofiler.org/)[23] taking advantage of the standard amino acid data set, Swissprot51, provided by the program. PONDR VL3 (http://www.pondr.com/pondr-tut2.html)[24] was used for prediction of intrinsic structural disorder.

For global expression analysis, normalized At-TAX expression data for 107 NAC genes during selected developmental stages and in response to different environmental stresses was obtained from http://www.weigelworld.org/resources/microarray/at-tax [25]. Expression values were imported into the R software for processing and heatmap output

using Bioconductor microarray analysis packages; *affy*, *Heatplus* and *gplots* (http://www.bioconductor.org/)[26].

Plant materials, growth conditions and *Arabidopsis* transformation

Arabidopsis (*Arabidopsis thaliana*) ecotype Col-0 wild-type, mutant homozygotes *anac019* (SALK_096303) and transgenic seeds were grown in growth chambers at 22°C and light intensity 150 µmol/m^2/s from a combination of fluorescent and incandescent lamps. Transformation of *Arabidopsis* Col-0 plants was performed by a modified floral dip infiltration method [27] using *Agrobacterium tumefaciens* strain GV3101 (pMP90) and the binary vector pCAMBIA3300 (http://www.cambia.org/daisy/bios/home.html) including CaMV 35S promoter for constituent expression of transgenes and the *BAR* selectable marker conferring Basta (glufosinate ammonium, 50 mg/L) selection (http://www.cambia.org/daisy/bios/home.html). Transgenic plants were selected by Basta spraying and T3 seeds from transformants expressing transgenes were used for subsequent analyses. For control an empty pCAMBIA3300 vector was transformed into Col-0 wild-type plants.

ABA sensitivity assay

To determine the effect of ABA on germination and post-germination growth efficiency, surface-sterilized and stratified (3 days) seeds of *Arabidopsis* Col-0 wild-type, *anac019* mutant alleles (SALK_096303), and transgenic lines expressing full-length, truncated, or chimeric constructs of *ANAC019*, were grown on half-strength MS medium supplemented with different concentrations of ABA (0-5 µM) (mixed isomers, SIGMA). A seed was considered as germinated when the radicle penetrated the seed coat. The percentage of germinated seeds was scored at day 3 after stratification in three independent experiments (40 seeds per experiment/treatment). Post-germination growth efficiency was scored as primary root length at day 7 after stratification. Seeds from Col-0 wild-type, *anac019* mutants and transgenic lines used for phenotypic analysis were harvested and collected at the same time. Root elongation assays were performed as described in [28] and significance levels assessed by two-sided *t*-test.

Production of recombinant proteins and *in planta* expression

For production of recombinant NAC proteins, cDNA clones obtained from ABRC (http://www.biosci.ohio-state.edu/~plantbio/Facilities/abrc/abrchome.htm) were amplified by PCR to obtain the complete coding part or the region encoding the NAC domain of ANAC019, ATAF1, NAC2/ORE1, VOZ2, VND3, VND7, NTL8, SOG1, NTL6, and ANAC003 using the primer sets listed in Supplementary table 1. The PCR-products were inserted into pENTR/D-TOPO (Invitrogen) followed by recombination into pDEST-15 (Invitrogen) to obtain Glutathione-S-transferease (GST)-NAC recombinant proteins. The recombinant proteins were named to show the borders of the NAC domain fragment, e.g. GST-ANAC019(1-168), which consists of GST fused at its C-terminus to the 168 N-terminal residues of ANAC019. *Escherichia coli* strain BL21(DE3) was transformed with the expression constructs and used for expression of GST-NAC recombinant proteins. Transformed *E. coli* cultures were grown to a cell density (OD_{600}) of 0.6 in Luria Bertani medium before induction with 0.5 mM isopropyl-β-D-thiogalactopyranoside. After 3 h of incubation at 37 °C the cells were harvested by

centrifugation at 10000 *g* for 20 min. The cell pellet was resuspended in phosphate-buffered saline (pH 7.5)/1mM dithiothreitol and incubated on ice for 30 min before sonication using a Soniprep 150 instrument (MSE) at amplitude 7 μ for 30 s, pausing for 1 min, and then repeating the process three times. Following sonication, the lysate was cleared by centrifugation at 10000 *g* for 20 min. The GST-NAC fusion proteins were purified by affinity chromatography using a 2 ml Glutathione Sepharose 4B (GE Healthcare) column for 250 ml *E. coli* culture (at the time of harvest OD_{600} was approximately 3.0-4.0). The column was washed with 10 column vol. phosphate buffered saline (pH 7.5)/1 mM DTT. Protein was eluted in 1 ml fractions in a buffer of 50 mM Tris/HCl (pH 8.0)/10 mM reduced glutathione. If needed, the proteins were purified further by anion exchange chromatography. The protein solutions were dialyzed over night against a buffer of 20 mM Tris/HCl(pH 8.0)/1 mM DTT and loaded onto a one ml RESOURCE Q column and eluted using a gradient of NaCl.

Vector pCAMBIA3300 and the primers and restriction enzymes listed in Supplementary Table S1 were used for generation of chimeric constructs for *in planta* expression. Full-length *ANAC019* (residues 1-317) was amplified using a cDNA clone obtained from ABRC. Constructs for domain-swapping experiments including the ANAC019 NAC domain and divergent TRDs were constructed as follows: The DNA fragment corresponding to the ANAC019 NAC domain (residues 1-168 of ANAC019) was PCR amplified using a full-length cDNA clone from ABRC. For overlap extension with fragments encoding the ANAC019 NAC domain (1-168), the reverse primers included a 12-base extension in their 5′-termini. DNA fragments encoding the TRDs of NAC2/ORE1(175-285), NTL8(158-335) and SOG1(216-449) were PCR amplified using cDNA clones from ABRC. For overlap extension PCR, forward primers for DNA amplification of sequences encoding TRDs, included a 12-base extension in their 5'-termini. Upon successful generation of individual fragments, the two amplification products were mixed and amplified using the following PCR program: 4x (94 °C, 3 min / 53 °C, 30 s / 72 °C, 3 min). The PCR reaction was performed with 2.5 U of *Pfu* DNA polymerase (Fermentas)/100 μl reaction, 0.2 mM dNTP, and approximately 150 ng of individual DNA fragments. Immediately afterwards, 0.3 μM of external primers were added and the constructed insert was amplified using the following PCR program: 95 °C, 2 min / 30 x (94 °C, 1 min / 52 °C, 1 min / 72 °C, 2 min) / 72 °C, 15 min / 4 °C. Constructs for domain-swapping experiments including ANAC019 TRD and divergent NAC domains were constructed as mentioned above, except amplifying NAC domains instead of TRDs of NAC2/ORE1(1-174), NTL8(1-157), and SOG1(1-215), and TRD of ANAC019(169-317) instead of the ANAC019 NAC domain.

EMSA (Electrophoretic Mobility Shift Assay)

A double-stranded oligonucleotide containing an optimized palindromic version of the NAC-binding site (NACBS) selected by CASTing (Cyclic amplification and selection of targets) [15] was prepared by annealing complementary single stranded oligonucleotides of the following sequence: 5'-CAGTCTTGCGTGTTGGAACACGCAACAGTC-3'. The double stranded oligonucleotide was 3' end-labelled with digoxigenin (DIG) using a DIG Gel Shift Kit, 2nd Generation (Roche). DNA binding reactions contained 4 μl 5 × binding buffer consisting of 100 mM Hepes (pH 7.6)/5 mM EDTA/50 mM $(NH_4)_2SO_4$/5 mM

DTT/1% (w/v) Tween-20/150 mM KCl, 1 µg poly[d(I-C)], 31 fmol DIG-labelled DNA and the amount of GST-NAC domain protein indicated in a final volume of 20 µl. The binding reactions were incubated for 15 min at room temperature and separated in a 10% polyacrylamide gel (PAGE Gold Precast Gels, Cambrex). Electroblotting was performed using a Hybond-N nylon membrane (Amersham Pharmacia Biotech), and chemiluminescent detection was carried out according to the instructions of the manufacturer of the DIG Gel Shift Kit.

Transactivation analysis in yeast

NAC proteins were examined for the presence of an activation domain using a yeast-based reporter assay. Full-length or NAC domain truncated coding sequences of the ten NAC candidates were cloned into the DNA-binding domain vector pGBKT7 (Clontech). Primers and restriction enzymes used are listed in Supplementary Table 1. Subsequently, constructs encoding the GAL4 DNA-binding domain fused to full-length or NAC domain truncated versions of NAC protein coding sequences were transformed into yeast strain pJ694A harboring *HIS3* and *ADE2* reporter genes. The transformed yeast cultures were plated onto SD/-Trp and SD/-Trp-His-Ade plates for seven days at 30 ºC before inspection of transactivation properties of the reporter constructs. For easy assessment of transactivational activity, overnight cultures of transformants were diluted to OD_{600} = 0.6, and then tenfold serial dilutions were prepared. Finally, 5 µl of each dilution was spotted onto SD/-Trp and SD/-Trp-His-Ade plates, which were then incubated at 30 °C for another seven days. Constructs pVA3-1 (p53/DBD) and pTD1-1 (T-antigen/AD) were used as positive controls. Yeast cells transformed with pGBKT7 vector alone cannot activate the *GAL1* promoter, and thus served as a negative control.

RNA extraction, cDNA synthesis and quantitative RT-PCR analysis

For *Arabidopsis* transcript analysis, total RNA was isolated from three complete rosettes for each genotype, using the RNeasy mini kit (Qiagen). One microgram of purified DNase-treated (Promega) RNA was used for cDNA synthesis using Superscript III cDNA synthesis kit (Invitrogen) according to the manufacturer's instructions. Quantitative PCR reactions were performed in triplicate for each individual line, and quantification of threshold cycle (CT) values was achieved by calculating means of normalized expressions using Q-gene software [29]. *Actin2* was used as a reference to determine the relative expression of ABA-responsive *Cold-responsive 47* (*COR47*) (At1g20440), and *Responsive-to-dessication 29b* (*RD29b*) (At5g52300), and stress-responsive *Ferritin precursor* 1(*FER1*)(At5g01600) and *Early-responsive-to-dehydration 11*(*ERD11*)(At1g02930). Experiments were repeated three times with similar results. Primers for real-time PCR are listed in Supplementary Table S1.

RESULTS

The *Arabidopsis* NAC family

To uncover the expansion of the *Arabidopsis* NAC TF family members from the latest genome annotation (TAIR9), the translated *Arabidopsis* genome was searched for NAC sequences using the NAC domain of ANAC019 [14] as search query and BLASTP as search program. This approach identified 92 non-redundant *Arabidopsis* proteins, and 18 additional NAC proteins were uncovered in an iterative PSI-BLAST search. A multiple ClustalX alignment of the 110 sequences revealed one sequence with only partial similarity to the NAC domain consensus, leaving 109 NAC sequences (Supplementary Table S2). To study the *Arabidopsis* NAC TF family in more detail, a phylogenetic tree was constructed in MEGA4 [22] using the neighbor-joining method, and bootstrap analysis to assess branch reliability (Figure 1). The conserved NAC domains were used to construct the phylogenetic tree. Since MEGA4 was unable to construct a tree based on alignments including At1g60240, this sequence was left out. The obtained phylogeny of 108 *Arabidopsis* NAC proteins is presented in Figure 1 and an alignment of selected NAC domains is presented in Figure 2A. Based on the phylogenetic analysis, the *Arabidopsis* NAC family was divided into ten major polytomies by collapsing branches with bootstraps <40%. Apart from the ten major groupings, several minor groupings were observed. To keep the *Arabidopsis* NAC nomenclature consistent, names from previous nomenclature systems [9;30] are maintained when possible in this updated nomenclature (see also Supplementary Table S2).

Global NAC gene expression analysis

Knowledge of NAC gene expression patterns has contributed significantly to our understanding of NAC gene involvement in plant stress tolerance and development [25]. To provide a global overview of the NAC transcriptome in relation to different developmental stages and environmental stresses we analyzed *Arabidopsis* gene expression data accessible online at the At-TAX homepage (http://www.weigelworld.org/resources/microarray/at-tax) [25]. As opposed to the limited number of NAC gene probe-sets available from the widely used ATH1 GeneChip, At-TAX expression data is available for almost all of the 108 NAC genes listed in Figure 1 (locus At1g64105 overlaps At1g64100 and has no unique probes), and therefore provided the basis for the expression analysis.

The At-TAX expression profiles showed that overall NAC gene expression is detected in all selected tissues and developmental stages (Figure 1). Likewise, NAC transcript levels are widely affected by a range of environmental stresses. Additionally, without posing any clustering of the expression data, minor groupings with similar expression patterns correlate with the phylogenetic analysis based upon NAC domain sequences. This is particularly evident for the tissue-specific expression patterns, *e.g.* the VND sub-group (II-1) of NAC genes, a fraction of which are predominantly expressed in both stems and roots, whereas others are expressed only in roots or stems. This data supports earlier

studies of VND members identifying them as important regulators in trans differentiation of various cells into vessel elements [31;32]. Another prominent sub-group is ATAF (IV-3) which shows the most abundant co-expression levels in old leaves and roots. This is also the case for the more distantly related subfamily VII-2 and for members of the transmembrane-achored sub-groups I-1 and I-2. Moreover, gene members of sub-group III-3 are highly salt-inducible, whereas those of subfamily I-1 and VII-2 mainly respond to elevated temperatures. Finally, the co-expression of the five NAC genes, encoding two NAC domains (VIII-1) in whole inflorescence and fruits deserves to be highlighted as a yet-to-be characterized sub-group with a distinct co-expression pattern correlating with phylogenetic distribution. Overall, among the displayed stress data sets, factors affecting the expression of most NAC genes are salt and extreme temperatures. With respect to the latter, most NAC genes induced by cold are repressed by heat, and *vice versa*. In conclusion, several phylogenetically related NAC genes are co-expressed in a developmental/tissue-specific, and to a lesser extent stress-specific manner.

Structural characterization and classification of the NAC proteins

To further analyze the structural diversity of *Arabidopsis* NAC proteins, sequence motifs were predicted using the programs MEME and MAST. This identified both shared and group-specific sequence motifs in the NAC domains (Figure 2A-B; Supplementary Table S2). The major motifs mapped to α-helices and β-strands of the NAC domain [13]. Whereas most of the NAC domains contained motifs *A*, *B*, *C*, *D*, *E*, and *F*, domains from group VI-X shared fewer motifs with the rest of the NAC domains, and some had sub-group specific motifs, such as motif *R* of sub-group IX-1 (Figure 2B).

The divergent C-terminal regions of NAC proteins, for which no tertiary structure information is available, were also examined for sequence motifs (Figure 2B; Supplementary Table S2). Sub-group II-1 NST/VND proteins shared motifs *WQ* and *LP*, of which *WQ* has been shown to be important for transactivational activity [33]. Three motifs; *LP (also named L;* [19]), *V* and *W*, dominate the C-termini of the CUC-like proteins, which includes NAC2/ORE1 [34], of sub-group II-3. Motif *W* was reported to be of importance for transcriptional activity [19;33]. Thus, most C-termini of group II proteins share motif *LP* and likely depend on a motif with a prominent tryptophan for activity (Figure 2B). Additional motifs, such as low-complexity histidine- and glutamine-rich *Q, U,* and *T* motifs were identified in some group II proteins. Group I includes most of the fourteen NAC proteins with a C-terminal transmembrane region [8;35]. For these proteins, the major part of the C-terminal motifs (e.g. *l*, *Y*, *b*, *K*, and *k*) is dominated by polar and negatively charged residues with conserved hydrophobic residues embedded in the polar matrix (e.g. isoleucine and leucine in motif *l* and tryptophan and phenylalanine in motif *K*). Sub-group III-3 contains the stress-related proteins, ANAC019, ANAC055, ANAC072, ATAF1, and ATAF2, which also have motifs found only in their sub-group (*o*, *g*, *h*, *n*, and *i*). These motifs, and motif *O* of sub-group V-II, are also dominated by a negatively charged matrix with a few conserved bulky and hydrophobic amino acid residues (Figure 2B; Supplementary Table S2).

Groups mentioned so far represent the most well studied NAC proteins, and comply with the classical modular outline presented in Figure 2C. Contrastingly, the proteins in sub-

group VIII-1 consist of two tandemly repeated NAC domains and have genes without annotated introns (Figure 2C). Sub-group IX-1 forms a distinct clade supported by high bootstrap values and contains the recently characterized *Suppressor of gamma response 1* (*SOG1*) [12]. The seven members of this sub-group deviate from the characteristic NAC structure by having N-terminal extensions of approximately fourty amino acid residues with a pair of conserved cysteine residues (motif *R;* Figure 2B-C). Some of these proteins also have typical, but specific C-terminal motifs (*d*, *e*). In the phylogenetic tree, sub-group IX-1 is close to the VOZ proteins (VIII-2;[36]), which were identified as NAC proteins in PSI-BLAST searches [15]. The VOZ proteins have an approximately 240 residues N-terminal region followed by a conserved DNA-binding domain, comprising a zinc-coordinating motif and a C-terminal basic region with structural similarity to the NAC domain (Figure 2A and 2C). In the N-terminal domain, the VOZ proteins contain motif *a* and *d* with chemical similarity to characteristic C-terminal NAC motifs. In conclusion, some sub-groups have specific motifs in the NAC domain, reflecting their divergence from the core NAC domain. Moreover, NAC C-termini contain several group-specific motifs which are enriched for polar and acidic residues.

Selection of representative NAC proteins

Most studies of NAC family members have focused on members from group I-III, which have typical NAC domains that bind DNA, and a C-terminal TRD (Figure 2C)[15;16;19]. To understand the modular structures of NAC TFs with respect to functionality, ten NAC proteins, representing functionally important clades spanning the phylogenetic diversity of the NAC family (Figure 1), were selected for characterization. Based on the extensive characterization of ANAC019 [13;15-17], we decided to use this protein as a reference in the comparative studies. The remaining nine NAC proteins included ATAF1, VND3, VND7, ANAC003, NAC2/ORE1, SOG1, VOZ2, NTL6, and NTL8 (Figures 1, 2A and 2B).

In vitro DNA-binding of structural variant NAC domains

We have previously selected a consensus NAC-binding site (NACBS) with the core sequence CGT[GA] using a re-iterative selection procedure on random oligonucleotides and ANAC019 and NAC2/ORE1 as baits [15]. A CGT[GA] or CGT core consensus sequence has been identified as DNA-binding sites of other NAC TFs from both *Arabidopsis* and wheat [15;16;37]. To examine the *in vitro* DNA-binding properties of selected NAC proteins, N-terminally GST-tagged NAC domains were produced heterologously. All proteins were purified by affinity chromatography and some were purified further by ion exchange chromatography. The quality of the recombinant proteins was examined by SDS-PAGE. The GST-NAC domain fusion proteins had the expected molecular mass values (between 47 and 54.1 kDa) and only minor contaminants (Figure 3A).

Based on the NACBS, a palindromic NACBS (palNACBS) was designed for EMSA-based analysis of NAC DNA-binding ([15]; Figure 3B). palNACBS was used to examine the ability of GST-NAC proteins to bind DNA (Figure 3C) in titration series using

increasing amounts (0-500 ng) of GST-NAC or GST as control. Initial experiments showed that GST-NAC proteins bound palNACBS with the same affinity as the corresponding NAC domain without a GST tag and that full-length NAC2/ORE1bound palNACBS with the same affinity as the NAC2/ORE1 NAC domain (data not shown). For each protein a minimum of three EMSAs were performed, each time with an independent protein batch and showing the same binding pattern. As expected, GST-ANAC019(1-168) and GST-NAC2/ORE1(1-165) bound palNACBS, and binding could be detected using 5 ng of the recombinant proteins. GST-ATAF1(1-165) also bound palNACBS, with an affinity, measured as the amount of protein needed to produced a shifted band, similar to that of ANAC019. This is in accordance with the close relationship of ANAC019 and ATAF1. In addition to the major band formed in the presence of GST-NAC and palNACBS, we sometimes observed a minor band, most often of slower migration (Figure 3C). This may represent non-native protein-protein interactions and has been observed in studies of other DNA-binding proteins [38]. A similar binding pattern was also seen for GST-VND7(1-163), whereas the closely related GST-VND3(1-165) bound with lower affinity. To our knowledge, the *in vitro* DNA-binding properties of proteins from the well-studied sub-group II-1 of VND proteins have not been analyzed previously. GST-VOZ2(205-420) also bound palNACBS, although with lower affinity than GST-ANAC019(1-168), which was also the case for both GST-NTL8(1-157) and GST-NTL6(1-168).

So far, ANAC003 and SOG1, or members of their clades, have not been characterized with respect to biochemical properties, and their sequences show a relatively low degree of similarity to the typical NAC domains. They both have substitutions in some of the positions corresponding to Lys-79, Arg-85, and Arg-88 in ANAC019 which are of importance to DNA-binding (Figure 2A)[15]. GST-SOG1(1-220) showed weak affinity for palNACBS. In contrast, attempts to show binding of GST-ANAC003(1-152) to palNACBS in EMSAs were unsuccessful. In addition, the recombinant protein did not show detectable binding to a 30 bp fragment of the cauliflower mosaic virus (CaMV) 35s promoter (positions -84 to -55), which is bound by several NAC proteins (data not shown) [39;40], and attempts to select binding sequences for GST-ANAC003(1-152) using CASTing were also unsuccessful. In conclusion, nine out of the ten NAC domains examined bound palNACBS, although with different affinities, generally reflecting their structural relations. No DNA-binding ability was detected for ANAC003.

C-terminal regions of NAC proteins are intrinsically disordered

The C-terminal regions of NAC TFs were analyzed further using Composition Profiler (http://www.cprofiler.org/)[23], which analyses the amino acid composition of a protein. This showed an under-representation of hydrophobic, structure stabilizing amino acid residues and an over-representation of polar and negatively charged residues in the C-termini compared to the frequency in the reference dataset, Swissprot51 (Figure 4A). The only frequent basic residue in the C-termini is histidine as reflected by the histidine-rich motifs *Q* and *U* (Figure 2B). The N-terminal regions of the VOZ proteins have similar profiles (see Figure 2C for schematic comparison of canonical NAC modularity *vs*. VOZ

modularity). In contrast, the NAC domains have a higher frequency of basic and aromatic residues (Figure 4A).
The compositional profile of NAC TF C-termini is suggestive of ID [2]. For this reason, all NAC proteins were examined for disorder using the predictor PONDR VL3 ([24], http://www.pondr.com). This analysis confirmed our assumptions and showed that the C-termini have a large degree of ID, also compared to the NAC domains (Figure 4B). The analysis was supported by circular dichroism analysis of a C-terminal region of a barley NAC protein indicating that the region is largely unfolded (T. Kjærsgaard, unpubl. data). For some proteins, a sub-group specific conserved motif is located in a dip in the disorder prediction as exemplified by motif *l* of sub-group I-4 proteins, *i*, of some sub-group III-3 proteins, and *WQ* of the VND sub-group II-1 proteins. In other cases, the motif is located in similar positions in the border between ordered and disordered regions (e.g. motif *LP* of sub-group II-3, *K* of sub-group I-2, and *N* of sub-group VII-2 proteins) (Figure 4B; and data not shown). Such motifs may represent Molecular Recognition Features (MoRFs;[41]) with propensity for target-induced folding. In conclusion, *in silico* analysis of NAC C-termini showed that they are enriched in polar and acidic residue, and have a large degree of ID.

Divergent NAC transcription regulatory domains

In several cases NAC proteins have been reported to have a C-terminal transcription regulatory domain (TRD) [5;42]. This correlates with the high degree of ID in NAC C-termini (Figure 4B), characteristic of TRDs [2]. To provide a systematic characterization of NAC TRDs we examined the transactivational activity of full-length and N-terminally truncated versions of the candidate NAC proteins. For this purpose fragments encoding full-length and truncated versions of the ten NAC proteins were fused in-frame to the GAL4 DBD in the yeast expression vector pGBKT7. The GAL4-NAC-TRD fusions were subsequently expressed in a yeast strain containing *HIS3* and *ADE2* reporter genes.

Firstly, transactivation analysis of full-length and truncated versions of ANAC019 overlapped our previous findings, identifying the C-terminal of ANAC019 to harbor the transactivation domain (Figure 5A)[14]. However, and in accordance with Bu *et al.* [17], full-length ANAC019 did not show any transactivation potential in this assay, though Tran *et al*. [16] have shown ANAC019 to transactivate expression in a plant-based reporter assay [16]. Interestingly, removing the extreme N-terminal region (ANAC019(1-40) improved its transactivational potential (Figure 5A) a result reminiscent of NAC1 transactivational activity [5]. All together, this data showed that the ANAC019 C-terminal region has transactivational activity in yeast.

Secondly, full-length VND7, NTL8, ATAF1 and VOZ2 have all been shown to activate transcription in yeast or plant-based reporter assays [32;36;43;44]. Furthermore, it has been shown that VND3 activates transcription through hetero-dimerization with VND7 [32], and that the TRDs of NTL8 and ATAF1 are located in the C-terminal part of the proteins [43;44]. Our reporter gene assay corroborates the findings on the previously analyzed full-length and C-terminal constructs, apart from full-length NTL8 (Figure 5B). In this study this construct did not show transactivation, in contrast to observations by

Kim *et al*. [43](Figure 5B). However, apart from ANAC019 and NTL8, all full-length constructs were able to transactivate expression of the *HIS3* and *ADE2* reporter genes. Additionally, all analyzed N-terminal truncation constructs showed transactivational activity, apart from the non-canonical NAC transcription factor VOZ2 which was dependent on its N-terminal to confer transactivation (Figure 5B). This protein was shown to have its transactivating activity in the N-terminus (see Figure 2C). In conclusion, NAC domains are dispensable for conferring a transactivational potential, and most often the TRD reside in the C-terminal part of the protein.

ANAC019 regulates ABA signaling during seedling development

Expression of the *ANAC019* gene is induced by drought, salt and ABA [14;16](Figure 1). ABA plays a key role in seed germination and early seedling development, as well as during plant drought and salt stress perception [45]. To investigate the functional relation between ABA and ANAC019, the ABA-dose response of Col-0 wild-type plants was compared to that of transgenic plants over-expressing full-length ANAC019 (1-317), ANAC019(169-317), or ANAC019(1-168) and to *anac019* mutant plants using quantitative bioassays of germination (radicle emergence) and seedling development (Figure 6A-B). Ectopic expression of transgenes was verified using QRT-PCR (Supplementary Figure S1). In the absence of exogenous ABA all genotypes showed similar germination by day 3, except *anac019* mutant plants which showed a significant decrease in the ability to germinate ($P < 0.05$)(Figure 6A). In the presence of low ABA-concentrations, ANAC019(1-317) and ANAC019(1-168) showed lower germination efficiency than wild-type and *anac019* mutant plants. At 2 μM ABA, plants expressing either full-length or truncated versions of ANAC019 displayed ABA-hypersensitivity compared to wild-type and especially *anac019* mutant plants, and at 5 μM ABA only *anac019* mutant seeds germinated (Figure 6A). During early seedling development, primary root length was scored seven days after stratification in the absence or presence of different concentrations of ABA (Figure 6B). At low ABA-concentrations, plants expressing either ANAC019(1-317) or ANAC019 (1-168) were more sensitive than the wild-type and *anac019* mutant plants, as evidenced by shortened primary root elongation ($0.05 < P < 0.001$). At 2 μM ABA plants expressing ANAC019(1-317) and ANAC019(169-317), but not ANAC019 (1-168), were significantly hypersensitive to ABA compared to wild-type and *anac019* mutant plants ($P < 0.01$). In summary, ANAC019(1-317) and truncated versions of ANAC019 displayed ABA-hypersensitivity during germination and early seedling development, suggesting that ANAC019 is a positive regulator of ABA-signaling.

To further understand the molecular basis of ANAC019-mediated ABA-hypersensitivity, the expression profile of several ABA- and stress-responsive genes (*Responsive to dessication 29b (RD29b)*, *Cold regulated 47 (COR47)*, *Ferritin precursor 1 (FER1)* and *Early responsive to dessication 11 (ERD11)*)[16] were investigated by QRT-PCR (Figure 6C). Three out of four genes analyzed displayed increased transcript levels in plants expressing ANAC019(1-317), ANAC019(1-168) and ANAC019(169-317) compared to control and *anac019* mutant plants. The induction of *FER1* expression was dependent on the C-terminal TRD of ANAC019 (Figure 6C or 7). Taken together, the up-regulation of

ABA-responsive marker genes in plants expressing full-length or truncated versions of *ANAC019*, correlated largely with increased sensitivity of ectopically expressing *ANAC019* plants. The presence of NACBS in the 1kb promoter of all four genes (data not shown) highlights ANAC019 as a possible direct positive upstream regulator of ABA signaling genes.

Modular interdependency of ANAC019-mediated ABA signaling

Recently, Taoka *et al.* [19] showed that among close homologues of the CUC proteins of the NAC family, functional specificity resides in the N-terminal NAC domain [19]. To examine if ANAC019-mediated perturbation of ABA-sensitivity can be mapped to specific domains, six chimeric constructs were generated in which NAC domains of NAC2/ORE1, NTL8 and SOG1 were individually fused to the ANAC019 TRD, and the TRDs of NAC2/ORE1, NTL8 and SOG1 were fused to the ANAC019 NAC domain (Figure 7A). All constructs were transformed into Col-0 wild-type plants. Selection of these candidate fusion partners enabled analysis of both distant and close homologues of *ANAC019*, with or without overlapping expression profiles (see Figure 1). The effect of ectopic expression of these constructs was examined by scoring root length in the absence or presence of 2 μM ABA (Figure 7). In addition to confer reduced root length in plants expressing *ANAC019*, this ABA concentration has also been shown to compromise root length of other NAC members [28].

Interestingly, when substituting the ANAC019 NAC domain with the NAC domains of NAC2/ORE1 and NTL8 plants became ABA-hypersensitive, whereas introducing the SOG1 NAC domain had no effect on root elongation compared to control plants (Figure 7B). *Vice versa*, substituting the ANAC019 C-terminal TRD with NAC2/ORE1 and NTL8 TRDs had no effect on root elongation compared to control plants. However, substituting ANAC019 TRD with SOG1 TRD could have a lethal effect on plants or hinder T-DNA insertion events, as no transformants were recovered from several attempts to ectopically express the ANAC019-SOG1 chimere (Figure 7B). QRT-PCR analysis was used to analyze whether the observed phenotypes correlated with chimeric transcript dosage. All transcripts were expressed in the same range, though transcripts of ANAC019(1-168)-NTL8(158-335) and NAC2/ORE1(1-174)-ANAC019(169-317) showed the lowest abundances (Supplementary Figure S1). In conclusion, this data implies that the C-terminal region of ANAC019 encodes a major level of functional specificity alone and when fused to both distantly and closely related NAC domains of ABA-inducible origin.

DISCUSSION

In this manuscript the modular structure of plant-specific NAC transcription factors was parsed to improve the understanding of their structure-function relationship. Most NAC proteins have an N-terminal NAC DBD and a C-terminal TRD. Such modularity is a paradigm of TF architecture, and members of individual TF families often target similar core *cis*-elements [1]. In accordance with this, DNA-binding studies of NAC proteins from distant clades have shown that NAC TFs bind a palindromic consensus sequence containing the core CGT[GA] [15;16;36;37], supporting the existence of conserved TF family-wise DNA-target sites. However, CASTing using the barley NAC protein IDEF2 and the *Arabidopsis* calmodulin-binding NAC CBNAC (a group I protein) selected unrelated NACBSs with a GCTT core sequence [46;47]. In this study the structural divergence of the NAC proteins was used to examine their DNA-binding properties using a consensus sequence selected by two NAC proteins, ANAC019 and NAC2/ORE1, which belong to different NAC groups (Figure 1). Apart from the ANAC019 and NAC2/ORE1 NAC domains, the NAC domain derived from ATAF1 and VND7 bound palNACBS with comparable affinity, whereas the rest of the NAC domains bound this oligonucleotide probe with lower affinity or as in the case of ANAC003 with no detectable affinity. This difference in affinity cannot be explained by the quality of the recombinant proteins which were all soluble, homogenous and of a high degree of purity. Therefore, the most likely explanation is that NAC TFs from different groups can recognize different DNA-target sites.

The findings raise other intriguing themes associated with transcriptional regulators. From the phylogenetic analysis of the NAC domains major sub-groups correlate with conserved motifs in the C-terminal TRDs of NAC proteins (Figure 2B and Supplementary Table 2). Although divergent with respect to sequences, these motifs have common features including a dominance by polar residues with a few highly conserved hydrophobic residues, characteristic of TRDs [2]. The analyses also suggested, that C-terminal NAC regions have a high degree of ID. Based on these properties, the NAC proteins are well-suited as model proteins for systematic analysis of structural ID, a research topic of great current interest. Most NAC TRDs contain sequence motifs that may be recognized both by specific and general proteins of the transcriptional apparatus, and in cases such as motif *l* (Figure 4B), where they interrupt the large disordered TRD, they could serve as MoRFs for partner-induced protein folding [41]. A few of these motifs, e.g. motif *W* and *WQ*, have already been shown to be essential for TRD activity [19]. Further characterization of these motifs may depend on the identification of interaction partners, and screening using motif-based consensus sequences could represent a useful strategy for identifying such interactors.

The phylogenetic analysis also revealed tissue-specific NAC gene expression clusters correlating with phylogenetic clades (Figure 1). Such observations have been reported for members of other large gene families [48] and suggest that the NAC gene family has expanded through evolution by gene duplication creating paralogous genes with a high degree of sequence similarity and functional redundancy. However, the less prominent

correlation between phylogenetic clusters and stress expression patterns could indicate that functional redundancy of regulators from the same tissue, is diluted during fine-tuned environmental stress-perceptions needed for an adequate response of the given plant tissue. Future systems-oriented temporal studies of selected tissues in response to environmental stresses are needed for an improved understanding of overlapping NAC gene expression patterns versus potential functional redundancy.

Another interesting observation from the characterization of NAC TF modularity was the ability of ANAC019 DBD and TRD to mimic ABA-hypersensitivity associated with ectopic expression of full-length *ANAC019* in a wild-type background (Figure 6A-B). This suggested that powerful effects may result from engineering NAC TFs. The ability of the ANAC019 DBD to positively regulate ABA-signaling may be explained by its ability to dimerize with endogenous NAC domains and thereby enhance promoter binding [15] and subsequent activation of ABA-signaling genes. In contrast, the ability of a NAC TRD to function without its natural molecular context has not been reported previously. The functional specificity may be explained by the ability of isolated regions to functionally interact with other proteins of the transcriptional complex. That specific interactions are important for ANAC019 TRD function is supported by the inability of the TRD from NAC2/ORE1, NTL8, and SOG1 to functionally substitute ANAC019 TRD in the chimeric proteins (Figure 7B). Though knowledge of how the isolated ANAC019 TRD exerts its function *in vivo* remains elusive, it is intriguing to acknowledge research performed more than two decades ago on the GAL4 transcription factor [3]. Herein, an N-terminally truncated version of GAL4 disabled in DNA-binding, was able to induce the expression of a reporter gene and even localize to the nucleus when over-expressed in a GAL4 wild-type background [3]. All together, Johnston *et al*. [3] speculated that the truncated GAL4 protein was transported into the nucleus in association with the wild-type GAL4. This scenario is somewhat analogous to our findings on the expression perturbations of ABA and stress signalling reporter genes induced when full-length and truncated versions of ANAC019 are ectopically expressed (Figure 6C). Additionally, the hypothesis on co-transportation should be addressed by analysing the subcellular localization of ANAC019(169-317) and reporter gene expression in a 35S:*ANAC019(169-317)* expressing *anac019* mutant background.

Though truncated versions of ANAC019 excluding either the DBD or the TRD of ANAC019 retain ABA-hypersensitivity when ectopically expressed in a wild-type background, the potential of the ANAC019 TRD is of particular interest. In addition to its autonomous effect when ectopically expressed, ANAC019 TRD encodes a major level of functionality during ABA-signaling even when fused to NAC domains from NAC2/ORE1 and NTL8 of distantly related groups. The chimera, NAC2/ORE1(1-174)-ANAC019(169-317) and NTL8(1-157)-ANAC019(169-317), also mimic the ANAC019-mediated phenotype of ABA hypersensitivity. In contrast, substituting ANAC019(1-168) with SOG1(1-215) did not confer ABA-hypersensitivity when fused to ANAC019 TRD. Though seemingly contrasting to the results on NAC domain indispensability of CUC proteins to enhance adventitious shoot formation [19], we propose that the functionality of the NAC domain of ANAC019 during ABA-signaling is exchangeable by NAC2/ORE1 and NTL8 NAC domains due to the conservation of key residues. Hence,

several studies have demonstrated that NAC proteins form both homo- and heterodimers [5;13;39], and the structure of the NAC domain fold contains both dimerization and DNA-binding interfaces [13;15]. In the NAC domain salt-bridge forming Arg-19 and Glu-26 and conserved hydrophobic residues are important for dimerization [15]. Arg-19 of ANAC019 is conserved in 81 per cent of the NAC proteins, including NAC2/ORE1 and NTL8. In SOG1, and the additional sub-group IX-1 members, Arg-19 is replaced by a lysine, and the sequence of the dimerization region shows a relatively low degree of conservation in these proteins (Figures 2A and D). This supports the hypothesis that residues in this region play an important role for the exchangeability of the ANAC019 NAC domain with the NAC domain of NTL8 and NAC2/ORE1 during ABA signaling. This is further substantiated by our demonstration of *in vitro* ANAC019-NAC2/ORE1 hetero-dimerization [15]. Future studies using site-directed mutagenesis of SOG1, NTL8 and NAC2/ORE1 are needed to clarify this.

The DNA-binding motif of the NAC domain has not been finally identified and no atomic-level structure information is available for a NAC/DNA complex. However, studies of mutant ANAC019 NAC domains pinpoint the loop/β-strand region containing Lys-79, Arg-85, and Arg-88 of ANAC019 as the DNA-binding, or part of the DNA-binding, motif [15]; Figures 2A and D). Whereas Arg-88 is conserved in all NAC proteins, Arg-85 is substituted with a glutamine in NTL8 and ANAC003 and Lys-79 is substituted with an alanine in SOG1 and VOZ2. These substitutions may explain the decreased palNACBS affinity of the proteins. Since Lys-79, Arg-85, and Arg-88 are conserved in NTL6 and VND3, which also bind palNACBS with lower affinity, additional regions or residues, such as Arg-78 or Arg-76, must be of importance to DNA-binding. It is possible that the ability to dimerize with endogenous ANAC019 could overrule sub-optimal DNA-binding by for example the NTL8 NAC domain in enabling ANAC019-mediated ABA-sensing. However, in the case of the SOG1 NAC domain, which binds palNACBS with relatively low affinity, dimerization may not be sufficient to compensate for weak target DNA-binding. Although ANAC019 mutants deficient in dimerization cannot bind cognate DNA-binding sites *in vitro*, it is not known whether two NACBS are needed for *in vivo* DNA-binding by NAC TFs [15]. Hence, although the C-terminal TRD of ANAC019 is needed for conferring ABA hypersensitivity in NAC proteins with both a DBD and a TRD, residues in the NAC domain could modulate this activity through inter-domain communication for identification of interaction partners and target promoters from which ANAC019 TRD exerts its transactivation activity. All together, our results have laid out future directions for engineering of NAC TFs for an improved understanding and use of NAC networks and for investigating their role in plant stress perception.

ACKNOWLEDGEMENTS

Authors would like to thank Dr. Lars Ellgaard for critical revision of manuscript.

FUNDING

This project was supported by Danish Agency for Science Technology and Innovation (DFF | Technology and Production) to KS (no. 274-07-0173) and MKJ (no. 274-07-0385).

Reference List

1 Luscombe,N.M. and Thornton,J.M. (2002) Protein-DNA interactions: amino acid conservation and the effects of mutations on binding specificity. J.Mol.Biol., **320,** 991-1009.

2 Dyson,H.J. and Wright,P.E. (2005) Intrinsically unstructured proteins and their functions. Nat.Rev.Mol.Cell Biol., **6,** 197-208.

3 Johnston,S.A., Zavortink,M.J., Debouck,C. and Hopper,J.E. (1986) Functional domains of the yeast regulatory protein GAL4. Proc.Natl.Acad.Sci.U.S.A, **83,** 6553-6557.

4 Huang,J., Li,X., Yi,P., Hilf,R., Bambara,R.A. and Muyan,M. (2004) Targeting estrogen responsive elements (EREs): design of potent transactivators for ERE-containing genes. Mol.Cell Endocrinol., **218,** 65-78.

5 Xie,Q., Frugis,G., Colgan,D. and Chua,N.H. (2000) Arabidopsis NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. Genes Dev., **14,** 3024-3036.

6 Graham,I.R., Haw,R.A., Spink,K.G., Halden,K.A. and Chambers,A. (1999) In vivo analysis of functional regions within yeast Rap1p. Mol.Cell Biol., **19,** 7481-7490.

7 Porsch,M., Sauer,M., Schulze,S., Bahlo,A., Roth,M. and Pflugfelder,G.O. (2005) The relative role of the T-domain and flanking sequences for developmental control and transcriptional regulation in protein chimeras of Drosophila OMB and ORG-1. Mech.Dev., **122,** 81-96.

8 Olsen,A.N., Ernst,H.A., Leggio,L.L. and Skriver,K. (2005) NAC transcription factors: structurally distinct, functionally diverse. Trends Plant Sci., **10,** 79-87.

9 Ooka,H., Satoh,K., Doi,K., Nagata,T., Otomo,Y., Murakami,K., Matsubara,K., Osato,N., Kawai,J., Carninci,P., Hayashizaki,Y., Suzuki,K., Kojima,K., Takahara,Y., Yamamoto,K. and Kikuchi,S. (2003) Comprehensive analysis of NAC family genes in Oryza sativa and Arabidopsis thaliana. DNA Res, **10,** 239-247.

10 Rushton,P.J., Bokowiec,M.T., Han,S., Zhang,H., Brannock,J.F., Chen,X., Laudeman,T.W. and Timko,M.P. (2008) Tobacco transcription factors: novel insights into transcriptional regulation in the Solanaceae. Plant Physiol, **147,** 280-295.

11 Jensen,M.K., Rung,J.H., Gregersen,P.L., Gjetting,T., Fuglsang,A.T., Hansen,M., Joehnk,N., Lyngkjaer,M.F. and Collinge,D.B. (2007) The HvNAC6 transcription factor: a positive regulator of penetration resistance in barley and Arabidopsis. Plant Mol.Biol., **65,** 137-150.

12 Yoshiyama,K., Conklin,P.A., Huefner,N.D. and Britt,A.B. (2009) Suppressor of gamma response 1 (SOG1) encodes a putative transcription factor governing multiple responses to DNA damage. Proc.Natl.Acad.Sci.U.S.A.

13 Ernst,H.A., Olsen,A.N., Larsen,S. and Lo,L.L. (2004) Structure of the conserved domain of ANAC, a member of the NAC family of transcription factors. EMBO Rep., **5,** 297-303.

14 Greve,K., La Cour,T., Jensen,M.K., Poulsen,F.M. and Skriver,K. (2003) Interactions between plant RING-H2 and plant-specific NAC (NAM/ATAF1/2/CUC2) proteins: RING-H2 molecular specificity and cellular localization. Biochem.J, **371,** 97-108.

15 Olsen,A.N., Ernst,H.A., Leggio,L.L. and Skriver,K. (2005) DNA-binding specificity and molecular functions of NAC transcription factors. Plant Science, **169,** 785-797.

16 Tran,L.S., Nakashima,K., Sakuma,Y., Simpson,S.D., Fujita,Y., Maruyama,K., Fujita,M., Seki,M., Shinozaki,K. and Yamaguchi-Shinozaki,K. (2004) Isolation and functional analysis of Arabidopsis stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress 1 promoter. Plant Cell, **16,** 2481-2498.

17 Bu,Q., Jiang,H., Li,C.B., Zhai,Q., Zhang,J., Wu,X., Sun,J., Xie,Q. and Li,C. (2008) Role of the Arabidopsis thaliana NAC transcription factors ANAC019 and ANAC055 in regulating jasmonic acid-signaled defense responses. Cell Res., **18,** 756-767.

18 Bu,Q., Li,H., Zhao,Q., Jiang,H., Zhai,Q., Zhang,J., Wu,X., Sun,J., Xie,Q., Wang,D. and Li,C. (2009) The Arabidopsis RING finger E3 ligase RHA2a is a novel positive regulator of abscisic acid signaling during seed germination and early seedling development. Plant Physiol, **150,** 463-481.

19 Taoka,K., Yanagimoto,Y., Daimon,Y., Hibara,K., Aida,M. and Tasaka,M. (2004) The NAC domain mediates functional specificity of CUP-SHAPED COTYLEDON proteins. Plant J., **40,** 462-473.

20 Daimon,Y., Takabe,K. and Tasaka,M. (2003) The CUP-SHAPED COTYLEDON genes promote adventitious shoot formation on calli. Plant Cell Physiol, **44,** 113-121.

21 Thompson,J.D., Gibson,T.J., Plewniak,F., Jeanmougin,F. and Higgins,D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res, **25,** 4876-4882.

22 Tamura,K., Dudley,J., Nei,M. and Kumar,S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol.Biol.Evol., **24,** 1596-1599.

23 Vacic,V., Uversky,V.N., Dunker,A.K. and Lonardi,S. (2007) Composition Profiler: a tool for discovery and visualization of amino acid composition differences. BMC.Bioinformatics., **8,** 211.

24 Peng,K., Vucetic,S., Radivojac,P., Brown,C.J., Dunker,A.K. and Obradovic,Z. (2005) Optimizing long intrinsic disorder predictors with protein evolutionary information. J.Bioinform.Comput.Biol., **3,** 35-60.

25 Zeller,G., Henz,S.R., Widmer,C.K., Sachsenberg,T., Ratsch,G., Weigel,D. and Laubinger,S. (2009) Stress-induced changes in the Arabidopsis thaliana transcriptome analyzed using whole-genome tiling arrays. Plant J..

26 Gentleman,R.C., Carey,V.J., Bates,D.M., Bolstad,B., Dettling,M., Dudoit,S., Ellis,B., Gautier,L., Ge,Y., Gentry,J., Hornik,K., Hothorn,T., Huber,W., Iacus,S., Irizarry,R., Leisch,F., Li,C., Maechler,M., Rossini,A.J., Sawitzki,G., Smith,C., Smyth,G., Tierney,L., Yang,J.Y. and Zhang,J. (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol., **5,** R80.

27 Clough,S.J. and Bent,A.F. (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J., **16,** 735-743.

28 Jensen,M.K., Hagedorn,P.H., de Torres-Zabala,M., Grant,M.R., Rung,J.H., Collinge,D.B. and Lyngkjaer,M.F. (2008) Transcriptional regulation by an NAC (NAM-ATAF1,2-CUC2) transcription factor attenuates ABA signalling for efficient basal defence towards Blumeria graminis f. sp. hordei in Arabidopsis. Plant J., **56,** 867-880.

29 Muller,P.Y., Janovjak,H., Miserez,A.R. and Dobbie,Z. (2002) Processing of gene expression data generated by quantitative real-time RT-PCR. Biotechniques, **32,** 1372-1379.

30 Mitsuda,N., Seki,M., Shinozaki,K. and Ohme-Takagi,M. (2005) The NAC transcription factors NST1 and NST2 of Arabidopsis regulate secondary wall thickenings and are required for anther dehiscence. Plant Cell, **17,** 2993-3006.

31 Kubo,M., Udagawa,M., Nishikubo,N., Horiguchi,G., Yamaguchi,M., Ito,J., Mimura,T., Fukuda,H. and Demura,T. (2005) Transcription switches for protoxylem and metaxylem vessel formation. Genes Dev., **19,** 1855-1860.

32 Yamaguchi,M., Kubo,M., Fukuda,H. and Demura,T. (2008) Vascular-related NAC-DOMAIN7 is involved in the differentiation of all types of xylem vessels in Arabidopsis roots and shoots. Plant J., **55,** 652-664.

33 Ko,J.H., Yang,S.H., Park,A.H., Lerouxel,O. and Han,K.H. (2007) ANAC012, a member of the plant-specific NAC transcription factor family, negatively regulates xylary fiber development in Arabidopsis thaliana. Plant J., **50,** 1035-1048.

34 Kim,J.H., Woo,H.R., Kim,J., Lim,P.O., Lee,I.C., Choi,S.H., Hwang,D. and Nam,H.G. (2009) Trifurcate feed-forward regulation of age-dependent cell death involving miR164 in Arabidopsis. Science, **323,** 1053-1057.

35 Kim,Y.S., Kim,S.G., Park,J.E., Park,H.Y., Lim,M.H., Chua,N.H. and Park,C.M. (2006) A membrane-bound NAC transcription factor regulates cell division in Arabidopsis. Plant Cell, **18,** 3132-3144.

36 Mitsuda,N., Hisabori,T., Takeyasu,K. and Sato,M.H. (2004) VOZ; isolation and characterization of novel vascular plant transcription factors with a one-zinc finger from Arabidopsis thaliana. Plant Cell Physiol, **45,** 845-854.

37 Xue,G.-P., Bower,N.I., McIntyre,M.C., Riding,G.A., Kazan,K. and Shorter,R. (2006) TaNAC69 from the NAC superfamily of transcription factors is up-egulated by abiotic stresses in wheat and recognises two consensus DNA-binding sequences. Func.Plant Biol., **33,** 43-57.

38 Bianchi,A., Stansel,R.M., Fairall,L., Griffith,J.D., Rhodes,D. and de,L.T. (1999) TRF1 binds a bipartite telomeric site with extreme spatial flexibility. EMBO J., **18,** 5735-5744.

39 Hegedus,D., Yu,M., Baldwin,D., Gruber,M., Sharpe,A., Parkin,I., Whitwill,S. and Lydiate,D. (2003) Molecular characterization of Brassica napus NAC domain transcriptional activators induced in response to biotic and abiotic stress. Plant Molecular Biology, **53,** 383-397.

40 Aida,M., Ishida,T. and Tasaka,M. (1999) Shoot apical meristem and cotyledon formation during Arabidopsis embryogenesis: interaction among the CUP-SHAPED COTYLEDON and SHOOT MERISTEMLESS genes. Development, **126,** 1563-1570.

41 Vacic,V., Oldfield,C.J., Mohan,A., Radivojac,P., Cortese,M.S., Uversky,V.N. and Dunker,A.K. (2007) Characterization of molecular recognition features, MoRFs, and their binding partners. J.Proteome.Res., **6,** 2351-2366.

42 Zheng,X., Chen,B., Lu,G. and Han,B. (2009) Overexpression of a NAC transcription factor enhances rice drought and salt tolerance. Biochem.Biophys.Res.Commun., **379,** 985-989.

43 Kim,S.Y., Kim,S.G., Kim,Y.S., Seo,P.J., Bae,M., Yoon,H.K. and Park,C.M. (2007) Exploring membrane-associated NAC transcription factors in *Arabidopsis*: implications for membrane biology in genome regulation. Nucleic Acids Res, **35,** 203-213.

44 Lu,P.L., Chen,N.Z., An,R., Su,Z., Qi,B.S., Ren,F., Chen,J. and Wang,X.C. (2007) A novel drought-inducible gene, ATAF1, encodes a NAC family protein that negatively regulates the expression of stress-responsive genes in Arabidopsis. Plant Mol.Biol., **63,** 289-305.

45 Finkelstein,R.R. and Gibson,S.I. (2002) ABA and sugar interactions regulating development: cross-talk or voices in a crowd? Curr.Opin.Plant Biol., **5,** 26-32.

46 Kim,H.S., Park,B.O., Yoo,J.H., Jung,M.S., Lee,S.M., Han,H.J., Kim,K.E., Kim,S.H., Lim,C.O., Yun,D.J., Lee,S.Y. and Chung,W.S. (2007) Identification of a calmodulin-binding NAC protein as a transcriptional repressor in Arabidopsis. J.Biol.Chem., **282,** 36292-36302.

47 Ogo,Y., Kobayashi,T., Nakanishi,I.R., Nakanishi,H., Kakei,Y., Takahashi,M., Toki,S., Mori,S. and Nishizawa,N.K. (2008) A novel NAC transcription factor, IDEF2, that recognizes

the iron deficiency-responsive element 2 regulates the genes involved in iron homeostasis in plants. J.Biol.Chem., **283,** 13407-13417.

48 Parenicova,L., de,F.S., Kieffer,M., Horner,D.S., Favalli,C., Busscher,J., Cook,H.E., Ingram,R.M., Kater,M.M., Davies,B., Angenent,G.C. and Colombo,L. (2003) Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world. Plant Cell, **15,** 1538-1551.

49 Bailey,T.L. and 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.

50 Bailey,T.L. and Gribskov,M. (1998) Combining evidence using p-values: application to sequence homology searches. Bioinformatics., **14,** 48-54.

FIGURE LEGENDS

Figure 1. Phylogenetic and expression analyses of the *Arabidopsis* NAC family
(**Left**) Phylogenetic presentation of the NAC TF family. The analysis is based on 108 NAC domain protein sequences and the tree was constructed in MEGA4 [22]. Individual NAC proteins are listed according to their TAIR9 (http://www.arabidopsis.org/portals/) AGI annotation, ANAC numbering as suggested by Ooka *et al*. [9] and, when possible, trivial name (for references see Supplementary table 2). (**Right**) Heatmap presentation of normalized At-TAX (http://www.weigelworld.org/resources/microarray/at-tax) expression data for 107 NAC genes during six different developmental stages and in response to six different stress treatments [25]. Averages of three replicate samples were used for all developmental stages and treatments analyzed. For the stress samples, expression profiles are presented relative to the appropriate mock treated samples. Color key at the bottom displays correlation between color and scaled log2 fold changes. Values below mean (developmental stages) or mock sample values (stress data set) are colored red and those above mean or mock sample values are colored white. Mean values and genes unchanged in their expression relative to mock control are colored orange. NAC TFs examined in this study are highlighted with an asterisk after the name.

Figure 2. Structural features of NAC proteins
(**A**) Alignment of the NAC domain sequence of the ten *Arabidopsis* NAC proteins analysed in this study. Residues surrounded by black are common to at least half of the sequences, whereas residues surrounded by gray are chemically similar in more than half of the sequences or similar to the dominating residue. An asterisk shows residues forming a salt bridge stabilising the dimerization interface (underlined) [13]. # indicates residues in a region containing several highly conserved residues of importance to DNA-binding [15]. The secondary structure elements, α-helix (α) and β-strands (β), of the ANAC019 NAC domain are indicated above the alignment. The element enclosed in parentheses was only present in one of the monomers of the NAC domain dimer [13]. The dominating sequence motifs identified by MEME/MAST analysis [49;50] are shown by the letters A-E. The consensus sequences of the motifs are shown in Supplementary Table 2. Proteins are labelled by their trivial name or ANAC numbering (for references see Supplementary table 2). (**B**) WebLogos of MEME/MAST sequence motifs for motif *I*, found in the N-terminal extension of group IX-1 proteins, and representative sequence motifs from NAC TRDs (for motif assignments to specific groups, sub-groups and proteins, see Supplementary Table 2). The height of a letter in the WebLogo indicates its relative frequency at the given position (x-axis) in the motif. (**C**) Schematic domain structures of NAC proteins; typical NACs, N-terminally extended sub-group IX-1 NAC proteins, VOZ with the zinc finger indicated (ZincF), and tandem NAC domain NACs. Black bars indicate NAC domain boundaries and black arrows indicate intron positions. The figure is not drawn to scale. (**D**) The structure of the ANAC019 NAC domain homo-dimer (Protein Data Base accession code 1UT7) is shown in ribbon format with different colours (green and magenta) for each monomer. The residues (Arg-19 and Glu-26) of the dimer interface (encircled) and Arg-88 of the likely DNA-binding motif (boxed) are

shown as sticks with nitrogen and oxygen atoms shown in blue and red, respectively. Residues Lys-79 to Arg-85 were not traced in the NAC019 crystal structure [13].

Figure 3. Analysis of NAC domain DNA binding using EMSAs
(**A**) SDS/PAGE analysis and Coomassie Blue staining of gel molecular-mass markers (molecular masses indicated in kDa at the left) (lane 1), approximately 5 µg of affinity-purified recombinant GST-ANAC019(1-168) (lane 2), GST-ATAF1(1-165) (lane 3), GST-NAC2/ORE1(1-176) (lane 4), GST-VOZ2(205-420) (lane5), GST-SOG1(1-220) (lane 6), GST-ANAC003(1-152) (lane 7), GST-NTL8(1-157) (lane 8), GST-VND7(1-163) (lane 9), GST-VND3(1-165) (lane 10), and GST-NTL6(1-168) (lane 11). (**B**) MEME WebLogos for the NAC-binding sequence selected for ANAC019 (top) and NAC2/ORE1 (middle). The selected sequences were used to design the palindromic NACBS, palNACBS (bottom), which was used for EMSA-based analysis of NAC DNA-binding [15]. (**C**) Representative EMSA using oligonucleotide palNACBS and 0 (lane 1), 5 ng (lane 2), 50 ng (lane 3), or 500 ng (lane 4) of the recombinant proteins indicated. All experiments were repeated at least three times with new preparations of the recombinant proteins, showing similar results with the major higher molecular mass band corresponding in size to the palNACBS-GST-ANAC019(1-168) complex.

Figure 4. Structural characteristics and predictions of NAC domains and adjacent regions
(**A**) Amino acid composition profiles of NAC C-terminal domains (not including C-termini of sub-group VIII-1 and VIII-2 proteins; left), VOZ N-terminal domains (middle), and NAC domains (right) compared to the reference data set, Swissprot51. (**B**) PONDR VL3 [24] analysis of NAC2/ORE1, representing group II NAC proteins, NTL8 representing sub-group I proteins, ANAC019 representing group III proteins, and SOG1 representing group IX proteins. These proteins were used for construction of chimeric proteins in this study (Figure 7). A threshold is applied with disorder assigned to values greater than or equal to 0.5 as indicated by the black bar. The position of the NAC domain is shown by a broken bar.

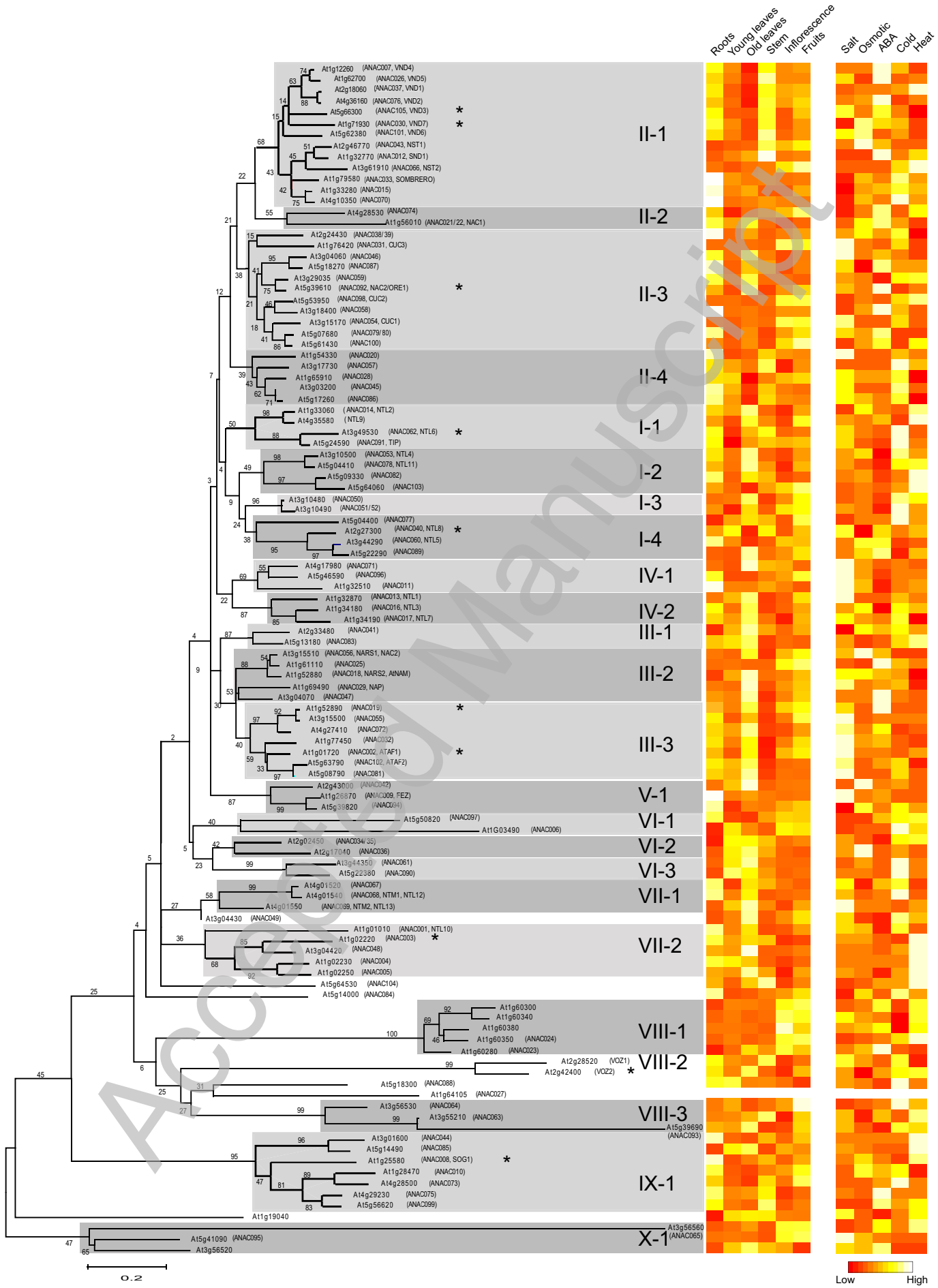
Figure 5. NAC transcription factors are transcriptional activators. (A) Fusion proteins of pBD-ANAC019(1-317), pBD-ANAC019(1-168), pBD-ANAC019(41-317), and pBD-ANAC019(169-317) were expressed in yeast and screened after seven days for their transactivation activity of the *HIS3* and *ADE2* reporter genes. **(B)** Full-length and truncated versions of ten NAC proteins were fused to GAL4-BD and analyzed for their transactivation activity of *HIS3* and *ADE2* seven days after yeast transformation. pVA3-1, pTD1-1 is a positive control, and empty pGBKT7 and pBD-ANAC019(1-168) are negative controls.

Figure 6. *ANAC019* is a positive regulator of ABA signaling. Transgenic plants over-expressing full-length ANAC019 (1-317), ANAC019(1-168) and ANAC019(169-317) are more sensitive to ABA than the Col-0 wild-type and *anac019* mutant. (**A**) Germination efficiency of Col-0, *anac019* mutant, 35S:ANAC019(1-317),

35S:ANAC019(1-168) or 35S:ANAC019(169-317) seeds grown on half-strength MS medium supplemented with different concentrations of ABA (0, 0.1, 0.5, 1, 2, or 5 μM) for 3 d after stratification. (**B**) Post-germination growth efficiency of Col-0, *anac019* mutant, 35S:ANAC019(1-317), 35S:ANAC019(1-168) or 35S:ANAC019(169-317) seedlings grown on half-strength MS medium supplemented with different concentrations of ABA for 7 d after stratification. (**C**) Expression profiles (+/- SEM) of ABA-signaling and stress-responsive genes in the Col-0, empty vector control, *anac019* mutant, 35S:ANAC019(1-317), 35S:ANAC019(1-168) or 35S:ANAC019(169-317) plants. The expressions of ABA-responsive (*COR47* and *RD29B)* and stress-responsive genes (*FER1* and *ERD11*) were analyzed by quantitative real-time PCR using *Actin2* as a reference.

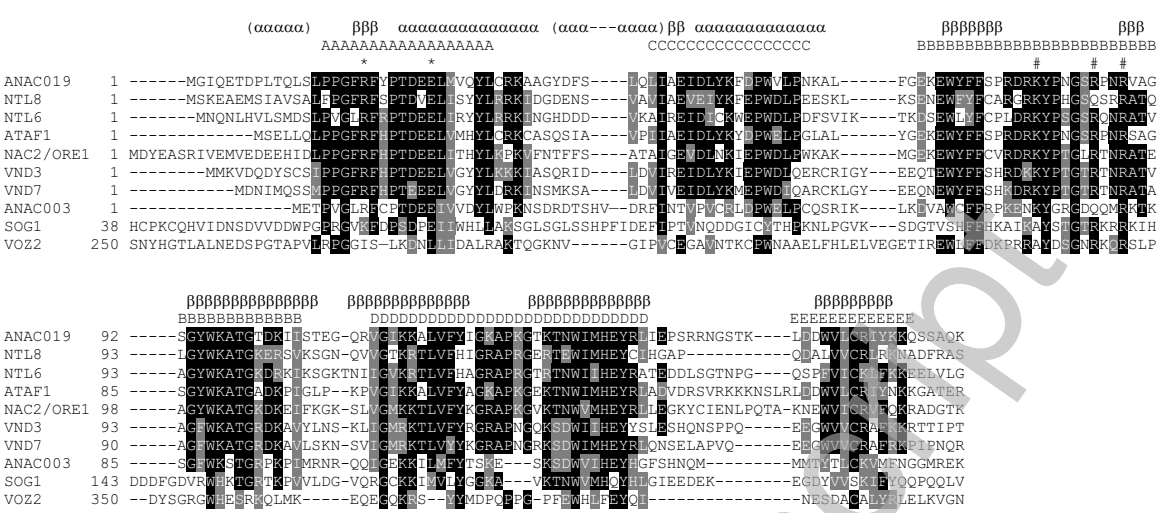
Figure 7. Analysis of ANAC019-mediated ABA sensitivity by domain-swap analysis. (**A**) Schematic representation of ANAC019-derived chimeric constructs expressed from the 35S promoter of pCAMBIA3300. (**B**) ABA-sensitivity of plants ectopically expressing *ANAC019* and domain-swapped NAC gene products relative to empty vector transformed plants was scored five days after end of stratification. For analysis of ANAC019 functional specificity TRDs (grey-shades) of NAC2/ORE1, SOG1 and NTL8 were fused to the ANAC019 NAC domain (black). *Vice versa*, NAC2/ORE1, NTL8 and SOG1 NAC domains (grey-shades) were fused to ANAC019 TRD (black). The assay was performed under the same condition as shown in Figure 6A-B. Asterisks indicate the significance of the difference from control plants as analyzed to two-sided student´s *t*-test. * = $P < 0.05$. Scale bar = 1 cm.

Biochemical Journal Immediate Publication. Published on 08 Dec 2009 as manuscript BJ20091234

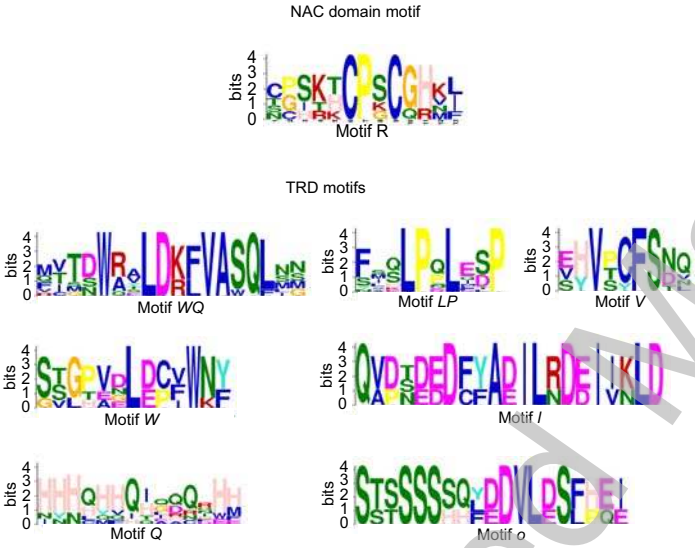


Biochemical Journal Immediate Publication. Published on 08 Dec 2009 as manuscript BJ20091234

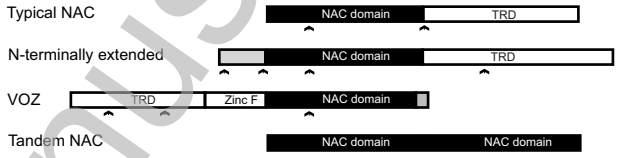
2A



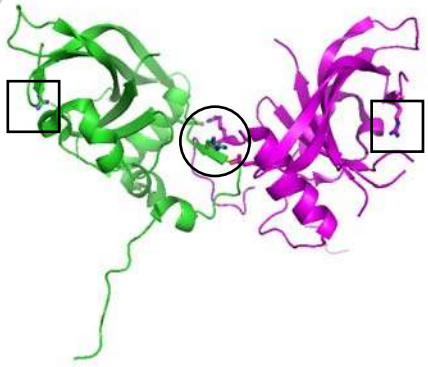
2B



2C

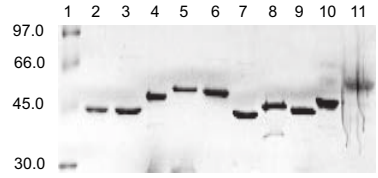


2D

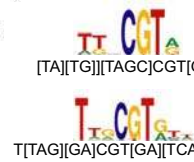


Biochemical Journal Immediate Publication. Published on 08 Dec 2009 as manuscript BJ20091234

3A

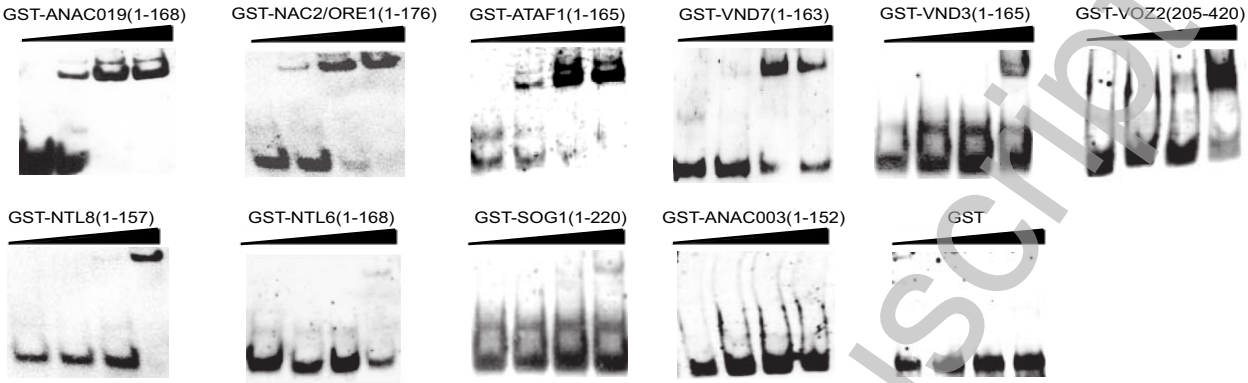


3B

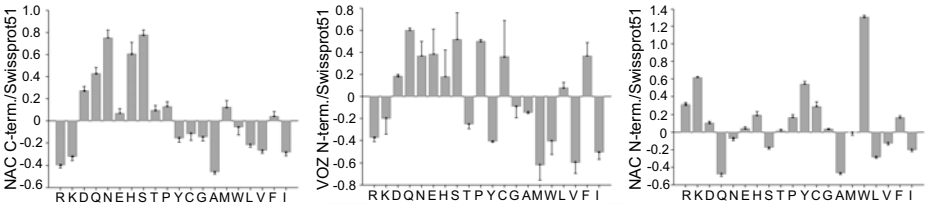


palNACBS
5'-CAGTCTTGCGTGTTGGAACACGCAACAGTC-3'
3'-GTCAGAACGCACAACCTTGTGCGTTGTCAG-3'

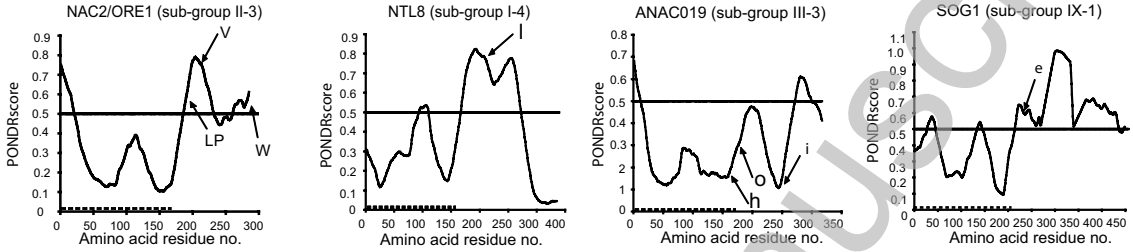
3C



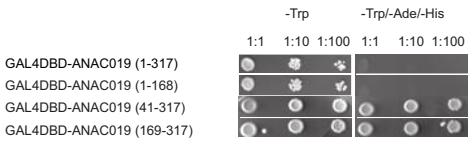
4A



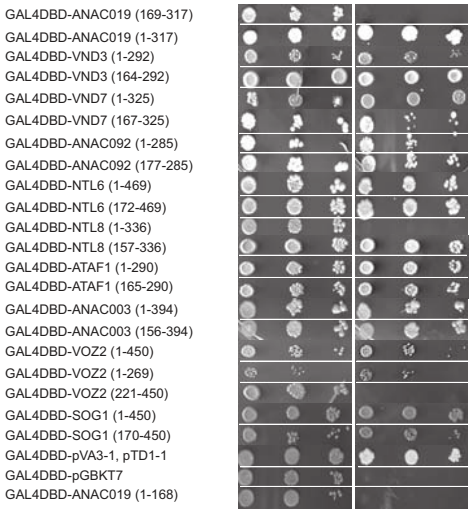
4B

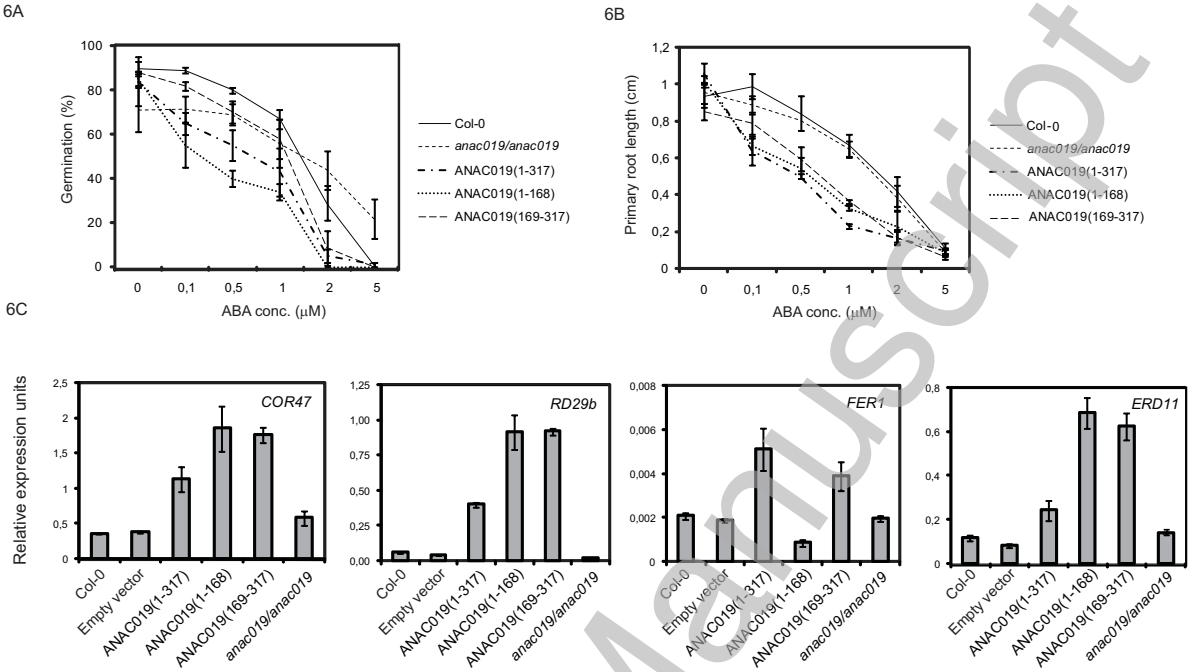


5A



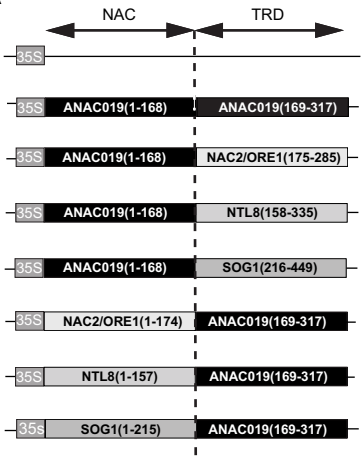
5B





Biochemical Journal Immediate Publication. Published on 08 Dec 2009 as manuscript BJ20091234

7A



7B