Structural, evolutionary and functional analysis of the NAC domain protein family in *Eucalyptus*

Steven G. Hussey[1], Mohammed N. Saïdi[2], Charles A. Hefer[3], Alexander A. Myburg[1], Jacqueline Grima-Pettenati[2]

[1]Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), Genomics Research Institute (GRI), University of Pretoria, Private Bag X20, Pretoria, 0028, South Africa

[2]Laboratoire de Recherche en Sciences Végétales (LRSV), Université Toulouse, UPS, CNRS, BP 42617, F-31326, Castanet-Tolosan, France

[3]Department of Botany, University of British Columbia, 3529-6270 University Blvd, Vancouver, B.C., V6T 1Z4, Canada

Corresponding author:
Prof. Alexander Myburg
Department of Genetics
University of Pretoria, Private Bag X20
Pretoria, 0028
South Africa
Email: zander.myburg@up.ac.za
Tel: +27-12-420-4945

Word count, main body:	6447
Word count, Introduction:	1141
Word count, Materials and Methods:	1013
Word count, Results:	2586
Word count, Discussion:	1575
Word count, Acknowledgements:	132
Number of tables:	1
Tables in colour:	Table 1
Number of figures:	6
Figures in colour:	Fig. 1, Fig. 2, Fig. 3, Fig. 4

Number of Supporting Files: 5

Summary

- NAC domain transcription factors regulate many developmental processes and stress responses in plants and vary widely in number and family structure. We analysed the characteristics and evolution of the NAC gene family of *Eucalyptus grandis*, a fast-growing forest tree in the rosid order Myrtales.
- NAC domain genes identified in the *E. grandis* genome were subjected to amino acid sequence, phylogenetic and motif analyses. Transcript abundance in developing tissues and abiotic stress conditions in *E. grandis* and *E. globulus* was quantified using RNA-seq and RT-qPCR.
- 189 *E. grandis* NAC (EgrNAC) proteins, arranged into 22 subfamilies, are extensively duplicated in subfamilies associated with stress response. Most *EgrNAC* genes form tandem duplicate arrays that frequently carry signatures of purifying selection. Sixteen amino acid motifs were identified in EgrNAC proteins, eight of which are enriched in, or unique to, *Eucalyptus*. New candidates for the regulation of normal and tension wood development and cold responses were identified.
- This first description of a Myrtales NAC domain family reveals a unique history of tandem duplication in stress-related subfamilies that has likely contributed to the adaptation of eucalypts to the challenging Australian environment. Several new candidates for the regulation of stress, wood formation and tree-specific development are reported.

Keywords: NAC, transcription factor, *Eucalyptus*, evolution, stress, wood formation

Introduction

Transcriptional regulators coordinate developmental processes and environmental responses in plants. A large family of NAC transcription factor proteins, defined by the conserved NAC (NAM/ATAF/CUC) DNA-binding domain in the N-terminal region, regulate diverse biological functions in terrestrial plants (Aida *et al.*, 1997; Olsen *et al.*, 2005). These proteins are chiefly involved in the response to biotic and abiotic stress (e.g. Tran *et al.*, 2004; Nakashima *et al.*, 2007; Jensen *et al.*, 2008; Meng *et al.*, 2009; Wu *et al.*, 2009), but also developmental processes (Guo & Gan, 2006; Yoo *et al.*, 2007; Willemsen *et al.*, 2008; Berger *et al.*, 2009; Morishita *et al.*, 2009), including secondary cell wall (SCW) biosynthesis during wood formation or xylogenesis (reviewed by Yamaguchi & Demura, 2010). It is thought that NAC proteins evolved over 400 million years ago and have to date only been identified in embryophytes (Zhu *et al.*, 2012). Most NAC subfamilies appeared before monocot-dicot divergence, with a few subfamilies restricted to tracheophytes, monocots, dicots or, rarely, specific plant families (Rushton *et al.*, 2008; Shen *et al.*, 2009; Zhu *et al.*, 2012).

The lignified SCWs that characterize vascular tissues of woody plants are rich in energy and biopolymers and therefore of significant agronomic importance (Plomion *et al.*, 2001; Mizrachi *et al.*, 2012). *Eucalyptus* is a woody plant genus encompassing some of the fastest growing plantation forest species. With over 20 million ha grown worldwide (Iglesias-Trabado & Wilstermann, 2008), it is a promising candidate for lignocellulosic biofuel production in addition to its extensive use in paper, pulp and raw cellulose products (Rockwood *et al.*, 2008; Carroll & Somerville, 2009). Lignified xylem fibers likely evolved through the sequential integration of independently evolved cellulosic cell wall thickenings, lignification (see Li & Chapple, 2010) and programmed cell death in a single cell type (Boyce *et al.*, 2003). NAC domain proteins feature prominently in the regulation of all these processes (Yamaguchi & Demura, 2010; Zhong *et al.*, 2010b; Ohtani *et al.*, 2011; Wang & Dixon, 2011). Knowledge of their transcriptional targets and biological functions provides a basis for developing approaches toward the improvement of wood and fiber properties. Considering the antiquity of xylogenesis, the apparent evolutionary conservation of most implicated NAC transcription factors (Zhong *et al.*, 2010a; Xu *et al.*, 2014) is not surprising. Yet, certain NAC subfamilies display distinct patterns

of evolution in particular plant lineages, associated with the unique evolutionary history of such lineages (Zhu *et al.*, 2012).

Almost all of the 700 known *Eucalyptus* species are endemic to Australia. Completing separation from Antarctica and drifting northwards around 50 Ma, the subcontinent's vegetation changed from tropical forest with high precipitation to a temperate, arid, grassland-dominated interior region during the Paleogene period (Kemp, 1981). *Eucalyptus* evolved by the Eocene (oldest fossils are ~52 Ma; Gandolfo *et al.*, 2011), diversifying in the cooler, drier conditions of the Paleogene and expanding throughout the continent (Beadle, 1981). While most eucalypts are adapted to xerophytic, fire-prone environments (Cary *et al.*, 2003), forest species such as *E. grandis* occupy wet, fertile regions of the eastern coast (Chippendale, 1988). The genetic basis for the successful adaptation of eucalypts to the widely variable and often harsh Australian environment and their rich diversity of phytochemical products, such as essential oils, remains poorly understood.

The evolutionary innovations allowing for the adaptation and unique properties of *Eucalyptus* could have involved diversification of transcription factors such as the NAC domain family. An understanding of how wood properties and environmental responses are transcriptionally controlled will help explain the adaptive potential of eucalypts, as well as their considerable capacity to produce woody biomass (Hinchee *et al.*, 2009). The structure, evolution, expression characteristics, and functions of the NAC domain family in *Eucalyptus* are currently unknown. We therefore analysed the gene and protein structure, phylogenetic relationships, and transcriptional dynamics of the *E. grandis* NAC domain family to elucidate the evolution and possible functions of NAC domain proteins in *Eucalyptus*. Within the core rosids (Eurosids), descriptions of NAC gene families have been reported for the Brassicales (Ooka *et al.*, 2003; Liu *et al.*, 2014; Ma *et al.*, 2014), Malvales (Shang *et al.*, 2013; Shah *et al.*, 2014), Sapindales (de Oliveira *et al.*, 2011), Fabales (Le *et al.*, 2011) and Malpighiales (Hu *et al.*, 2010). NAC family descriptions also exist for core Eudicot groups such as *Vitis* (Wang *et al.*, 2013) and the *Solanaceae* (Rushton *et al.*, 2008; Singh *et al.*, 2013; Kou *et al.*, 2014), as well as representative monocots (Nuruzzaman *et al.*, 2010; Christiansen *et al.*, 2011; Puranik *et al.*, 2013; Shiriga *et al.*, 2014). These family-wide surveys of NAC proteins have provided insights into their conservation, diverse evolutionary histories and possible functions. Comparative genomic

analysis of NAC proteins in other angiosperm lineages will facilitate a better understanding of NAC protein specialization and function. This study provides a thorough first description of *NAC* gene family structure in the Myrtales, an order basal to the core rosids (Myburg *et al.*, 2014), and contributes to our understanding of NAC domain evolution and function in the rosids.

Materials and Methods

Plant materials

E. globulus tissues were flash-frozen in liquid nitrogen. Juvenile and mature xylem samples (kindly provided by RAIZ, Portugal) were harvested from four- and ten-year-old trees (genotype VC9) respectively. Upright, tension and opposite xylem of two-year-old trees (genotypes GM52, BB3 and MB43, kindly provided by Altri Florestal, Portugal) were collected from main stems after three weeks of bending (45°). Fruit capsules and flower buds were harvested from genotype C33 (Altri Florestal, Portugal). For cold treatment experiments, one-year-old *E. globulus* genotype GM258 (Altri Florestal, Portugal) were subjected to cold (7°C) for 16 h in the dark. Control plants were maintained for 16 h in dark greenhouse conditions. Young and mature leaves, primary stems, secondary stems and roots were harvested. Each of three biological replicates consisted of bulked tissues from two trees.

RNA extraction and reverse transcription quantitative PCR (RT-qPCR)

Total RNA was extracted from frozen tissues as described elsewhere (Southerton *et al.*, 1998). cDNA synthesis, primer design and Fluidigm RT-qPCR analysis was conducted as described by Cassan-Wang *et al.* (2012). Primer set-specific PCR efficiencies and five control genes previously found to show stable expression across different tissues and various environments (Cassan-Wang *et al.*, 2012) were used for data normalization (Supporting Information Table S1). Expression data were subjected to QT clustering (Pearson correlation ≥ 0.5, minimum five genes per cluster) in TMEV (Saeed *et al.*, 2006).

Identification of NAC domain proteins

Genes in the *E. grandis* genome v.1.0 annotated with a Pfam NAM domain (PF02365; Punta *et al.*, 2012) were retrieved from Phytozome v7.0 (http://www.phytozome.net/eucalyptus.php). All but the longest splice variants were removed. Some genes encoding proteins lacking

initiation or termination codons were corrected with FGENESH (http://linux1.softberry.com/berry.phtml). Where possible, annotations were corroborated with RNA-seq data from Eucspresso (Mizrachi *et al.*, 2010) and EucGenIE (http://eucgenie.bi.up.ac.za; Hefer *et al.*, 2011) databases. Gene models that could not be corrected were discarded. Twenty-two gene models were located on smaller "satellite" scaffolds that have not yet been mapped to the eleven *E. grandis* chromosome scaffolds: those with ≥ 95% nucleotide identity to other NAC genes were considered allelic. The presence of a NAC domain in the proteins was evaluated with a hidden Markov model (HMM) constructed from the NAC domain alignment of representative proteins from diverse species (Olsen *et al.*, 2005), using HMMER 3.0 (Finn *et al.*, 2011).

Phylogenetic analysis

For the EgrNAC protein tree, 189 curated EgrNAC protein sequences were aligned using MUSCLE (http://www.ebi.ac.uk/Tools/msa/muscle/; parameters pre-set). The alignment was trimmed with Gblocks (Castresana, 2000) using parameters: minimum sequences per conserved position, $n/2 + 1$; minimum sequences per flank position, $n/2 + 1$; maximum number of contiguous nonconserved positions, 10; minimum block length, 2; allowed gap positions, all. After visual inspection, poorly aligning sequences (EgrNAC91, EgrNAC187) were removed. For the five-species NAC gene tree, NAC protein sequences from *Arabidopsis thaliana*, *Oryza sativa*, *Populus trichocarpa*, *Vitis vinifera* and the 189 EgrNAC sequences (678 total) were similarly aligned with MUSCLE and trimmed with Gblocks. Two poorly aligning sequences (EgrNAC29, EgrNAC91) were removed. The alignments were submitted to PhyML 3.0 (Guindon *et al.*, 2010), initiated with a BIONJ tree using estimated Gamma distribution, proportion of invariable sites fixed at 0.0, four substitution rate categories, an LG substitution model with empirical equilibrium frequencies, and Shimodaira-Hasegawa-like aLRT branch support testing (Anisimova & Gascuel, 2006). Trees were visualized in MEGA 5.05 (Tamura *et al.*, 2011). The trees were rooted at the midpoint due to the lack of a known outgroup (Shen *et al.*, 2009; Zhu *et al.*, 2012).

Identification of conserved protein motifs

Sequences of the 189 aligned EgrNAC proteins were analysed using MEME v.4.7.0 (Bailey *et al.*, 2006) with parameters: distribution of motifs, zero or one per sequence; maximum number

of motifs, 25; minimum number of sites, two; maximum number of sites, 189; minimum motif width, six; maximum motif width, 50. Overrepresented motifs were annotated using the PfamA and PfamB databases (http://pfam.sanger.ac.uk), and schematically represented using DomainDraw (Fink & Hamilton, 2007). HMMs were constructed from the MEME alignment of each motif using hmmbuild in HMMER 3.0 (Finn *et al.*, 2011). NAC protein sequences of *Populus trichocarpa*, *Glycine max*, *Carica papaya*, *Arabidopsis thaliana*, *Vitis vinifera*, *Oryza sativa* subsp. *japonica*, *Brachypodium distachyon* and *Zea mays* retrieved from the Plant Transcription Factor Database v.2.0 (Zhang *et al.*, 2011) were searched with the HMMs using HMMER 3.0. The TMHMM server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) was used for transmembrane helix (TMH) prediction, using a probability threshold of 1.0.

Gene structural analysis

Genomic sequences of the *E. grandis* v.1.0 annotation were downloaded from Phytozome v7.0 (http://www.phytozome.net/eucalyptus.php), untranslated regions were removed and where applicable genomic sequences re-annotated corresponding to curated gene models. Coding sequences were aligned to genomic sequences and schematics generated using GSDS (http://gsds.cbi.pku.edu.cn) (Guo *et al.*, 2007).

Chromosomal localization and test for selection neutrality

MapChart 2.2 (Voorrips, 2002) was used for chromosomal linkage visualization. Coding sequences of genes in individual tandem duplicate blocks were aligned using MUSCLE in MEGA 5.05 (Tamura *et al.*, 2011). A codon-based Z-test was performed for each block in MEGA 5.05 using Pamilo-Bianchi-Li (Li, 1993; Pamilo & Bianchi, 1993) substitution model, bootstrap variance estimation method (1000 replicates), and pairwise deletion. Blocks were assessed for neutrality (H_A: $dN \neq dS$), positive selection (H_A: $dN/dS > 1.0$) and purifying selection (H_A: $dN/dS < 1.0$). Only results that remained significant for most of the substitution models in MEGA were considered, and all blocks had more than ten synonymous or nonsynonymous substitutions as advised by Zhang *et al.* (1997).

***E. grandis* transcriptome analysis**

EgrNAC coding sequences were aligned to the *E. grandis* (v.1.1) genome sequence using Exonerate (Slater & Birney, 2005), and the genome locations calculated. RNA-seq data of six

tissues for three field-grown *E. grandis* individuals, and root samples prepared from young seedlings, were obtained from EucGenIE (http://eucgenie.bi.up.ac.za; Hefer *et al.*, 2011). The absolute transcript abundance values (FPKM) were obtained for the 189 NAC domain sequences with TopHat (Trapnell *et al.*, 2009) and Cufflinks (Trapnell *et al.*, 2010). The expression values were clustered using the QT clustering tool (Pearson correlation ≥ 0.75, minimum five genes per cluster) in the Multiple Array Viewer (Saeed *et al.*, 2006).

Results

Identification of NAC domain genes in *E. grandis*

Gene models in the v.1.0 annotation of the *E. grandis* genome containing a NAC domain were identified using a superfamily search (see Methods), yielding 254 candidates (Supporting Information Table S2). Thirty-nine alternative splice variants were removed except for two splice variants that displayed only partial gene sequence overlap. Apart from these, only primary transcripts were considered. Sixteen genes on scaffolds other than those comprising the eleven main linkage groups (chromosomes) were considered putative alleles of gene models on the chromosome scaffolds since they showed ≥95% nucleotide similarity. One pair of adjacent gene models was collapsed into a single gene model (Supporting Information Table S2). Ten genes were removed following manual curation and 43 were corrected (see Supporting Information Note S1 for details). The remaining proteins were inspected for a significant match to the NAC domain (E-value < 0.001) using a hidden Markov model (HMM) of the NAC domain, rejecting one gene model (Eucgr.D00593.1). This process yielded 189 nonredundant candidates. They encoded proteins of 82 to 799 residues, a range similar to NAC proteins from other plants (Supporting Information Table S3). We sorted the gene symbols alphanumerically, i.e. in their order of appearance on chromosomes A through K and their sequential appearance in scaffolds not linked to the chromosomes, and renamed them *EgrNAC1* through *EgrNAC189* (Supporting Information Table S4).

Phylogeny of the EgrNAC proteins in relation to angiosperm NAC proteins

The evolution of the EgrNAC proteins was evaluated through maximum likelihood analysis incorporating well-described NAC domain sequences in the dicots *Arabidopsis* (http://www.arabidopsis.org/), *Vitis* (Shen *et al.*, 2009), and *Populus* (Hu *et al.*, 2010), and the

monocot *Oryza* (Shen *et al.*, 2009). The non-*Eucalyptus* NAC proteins analysed are listed in Supporting Information Table S3. To improve the reliability of the phylogeny, only conserved positions in the alignment, represented by at least 50% of the sequences, were considered. Preliminary classification of the phylogeny according to the 21 subfamilies identified by an extensive analysis of the NAC protein family from nine lineages (Zhu *et al.*, 2012) revealed good agreement with the topology in our study. We therefore used the cited study to annotate subfamilies (i.e. roman numerals and referring to subfamily names identified by Ooka *et al.* (2003) where applicable), but some differences should be noted (see Supporting Information Note S2). This culminated in 22 subfamilies with acceptable bootstrap support (>70) and distinguishable topologies (Fig. 1a). The biological functions of the respective members of the subfamilies, where known, were investigated in the literature. Generally, members of the same subfamily appeared to share similar biological processes (Supporting Information Table S5), as observed elsewhere (Shen *et al.*, 2009).

The phylogeny representation in Fig. 1b was linearized with respect to evolutionary distance. NAC proteins from different species were generally interspersed; however, EgrNAC proteins were overrepresented in subfamilies IVa, IVc, Vb and VII. Subfamily VII had previously been found to contain *Populus* and *Carica* sequences, but not those of *Vitis* or herbaceous plants (Zhu *et al.*, 2012). Consistent with this, in our phylogeny *Populus* but no *Arabidopsis*, *Oryza* or *Vitis* sequences constituted this apparently ancient subfamily, while we additionally identified several *Eucalyptus* NACs in subfamily VII (Fig. 1b). These sequences displayed greater intraspecific than interspecific homology, as previously observed in *Populus* and *Carica* (Zhu *et al.*, 2012). This indicates independent gain of subfamily VII NAC sequences in *Eucalyptus*, *Populus* and *Carica* as opposed to their loss in *Arabidopsis*, *Oryza* and *Vitis*, and suggests that parallel evolution in these genera may have facilitated the retention of duplicated genes in subfamily VII. Similarly, most of the *Eucalyptus* NACs overrepresented in subfamilies IVa, IVc and Vb appear to be lineage-specific paralogs (Fig. 1b).

Because of the importance of *Eucalyptus* as a wood fiber crop, we explored whether the *E. grandis* genome contains homologs of *Arabidopsis* NACs involved in SCW biosynthesis (reviewed by Yamaguchi & Demura, 2010; Hussey *et al.*, 2013) using the phylogeny (Fig. 1b, detailed dendrogram available in Supporting Information Fig. S1). We found single putative

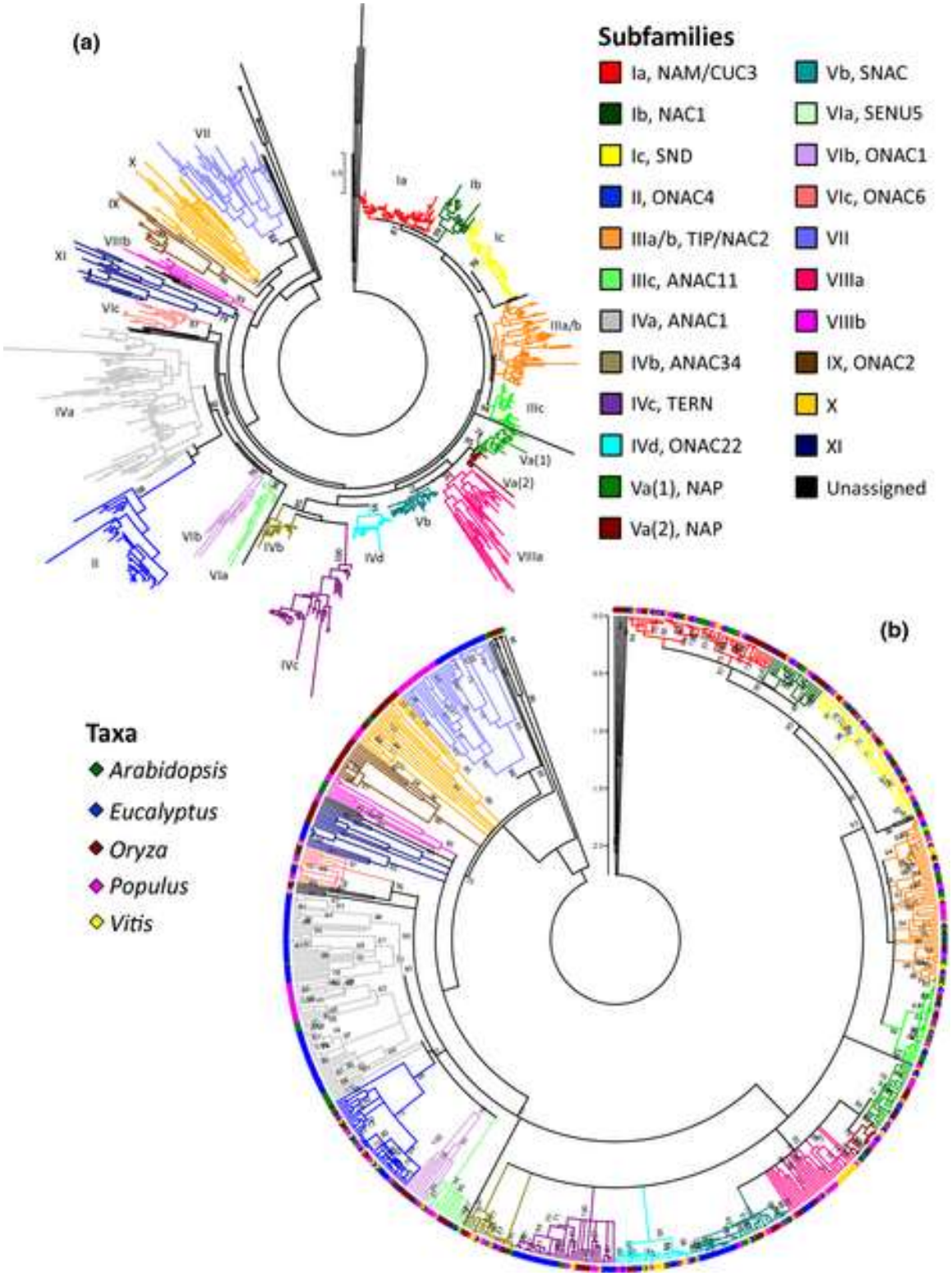


Fig. 1. Maximum likelihood phylogeny of the *Eucalyptus* NAC family in relation to angiosperm lineages *Arabidopsis*, *Populus*, *Oryza* and *Vitis.* Trees were rooted at the midpoint. (a) Circular phylogram showing subfamilies described by Zhu *et al.* (2012); aLRT branch support values for each subfamily are indicated. (b) Linearized circular representation, normalized with respect to evolutionary distance. aLRT values >70 are indicated and the organism of origin of each respective taxon is indicated with a diamond symbol. A detailed dendrogram is available in Supporting Information Fig. S1.

Eucalyptus orthologs of *Arabidopsis* fiber-associated SND1, SND2 and NST1, vessel-associated VNI2, VND6 and VND7, and multiple co-orthologs of SND3 and XND1 (Supporting Information Table S6). Interestingly, we did not identify a *Eucalyptus* ortholog of NST2, which is associated with endothecium SCWs in *Arabidopsis* anthers (Mitsuda *et al.*, 2005). Overall, this suggests that NAC-mediated transcriptional regulation of SCW biosynthesis in *Eucalyptus* is relatively well conserved with, but not identical to, *Arabidopsis*.

Phylogenetic relationships and expression patterns of *EgrNAC* genes

A gene tree of 187 EgrNAC proteins was constructed using the maximum likelihood approach after removing two poorly aligning proteins (Fig. 2a). This phylogeny was used to assess the evolutionary conservation of gene expression patterns, amino acid motifs and gene structure of *EgrNAC* genes.

Tissue-specific transcript abundance is suggestive of a gene's biological function. To generate hypotheses of the functions of unknown *EgrNAC* genes, we examined their expression patterns in shoot tips, young leaves, mature leaves, flowers, roots, phloem and developing (secondary) xylem. RNA-seq data for three field-grown *E. grandis* individuals were obtained from the *Eucalyptus* Genome Integrative Explorer (EucGenIE; http://eucgenie.bi.up.ac.za/), and reads were re-mapped to the *EgrNAC* coding sequences to accommodate corrected gene models (data provided in Supporting Information Table S7). Transcripts of closely related genes showed broadly similar abundance profiles (Fig. 2b). No expression was detected for 19 (~10%) of the *EgrNAC* genes in the sampled tissues. Conversely, 93 genes (~50%) were expressed in all tissues.

To identify transcripts with similar expression patterns, the expression data of the 189 *EgrNAC* genes were hierarchically clustered using a quality threshold algorithm, yielding 13 clusters (Supporting Information Fig. S2). Cluster 4 contained genes preferentially expressed in various stages of leaf development, enriched for paralogs belonging to stress response-associated subfamily Vb (Supporting Information Table S5). Transcripts preferentially expressed in tissues containing vascular cells (roots, phloem, developing xylem) were located in Clusters 6, 10 and 11, including (co-)orthologs of SND1, NST1, SND2 and XND1 (Supporting Information Fig. S2, Table S6). Based on their preferential expression in vascular tissues, *EgrNAC24*, *EgrNAC32*,

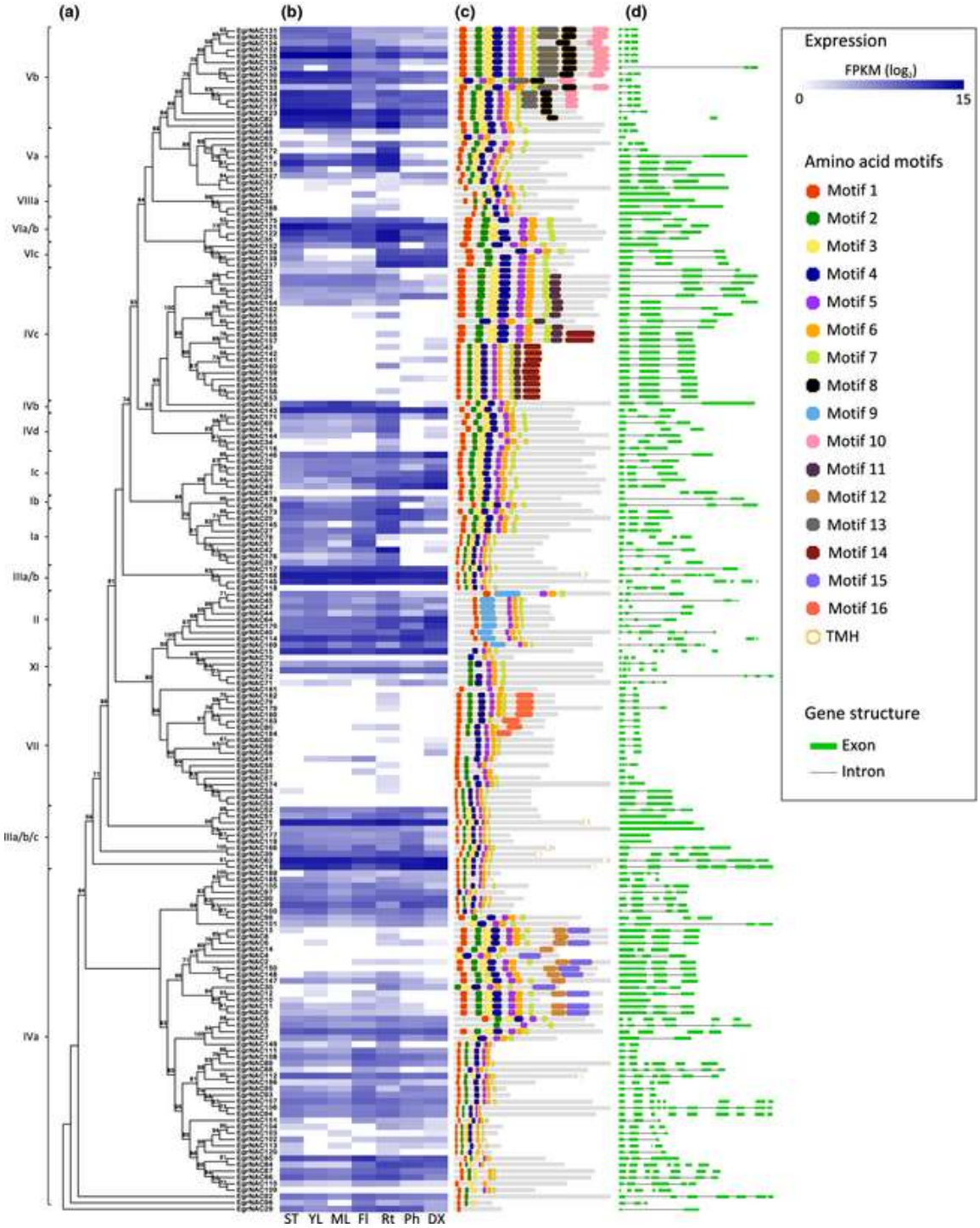


Fig. 2. Predicted EgrNAC protein phylogeny, transcript abundance profiles, conserved amino acid motifs and gene structure. From left to right: (a) unrooted maximum likelihood hylogeny of EgrNAC proteins, showing subfamily classification and aLRT values greater than 50. (b) RNA-seq transcript abundance of *EgrNAC* genes in shoot tip (ST), young leaf (YL), mature leaf (ML), flowers (Fl), root (Rt), phloem (Ph) and developing xylem (DX) of three field-grown *E.grandis* trees. Values are expressed as the $\log_2$ value of average fragments per ilobase of exon per million fragments mapped (FPKM) per tissue. (c) Composition and distribution of overrepresented amino acid motifs (see Supporting Information Table S8). Grey bars indicate relative protein lengths. (d) Position of exons and introns in the *EgrNAC* gene models.

EgrNAC58, *EgrNAC59*, *EgrNAC90*, *EgrNAC141* and *EgrNAC157* are novel candidates for the regulation of xylogenesis-related processes since they have no *Arabidopsis* orthologs associated with this process (Supporting Information Fig. S1, S2). Remarkably, of the 21 *EgrNAC* genes that were expressed in only one tissue, 14 were restricted to roots (Cluster 3). These included *EgrNAC81*, a homolog of *BRN1*/*BRN2* known to be specifically expressed in root tips (Bennett *et al.*, 2010). One gene was expressed only in mature leaves, three were restricted to flowers and another three to developing xylem (Supporting Information Fig. S2, unassigned cluster).

Conserved motifs in *Eucalyptus* NAC domain proteins and conservation of gene structure

Overrepresented amino acid motifs tend to represent functional regions that are evolutionarily conserved across or within specific lineages. We subjected the 189 EgrNAC sequences to motif overrepresention analysis using MEME (Bailey *et al.*, 2006). Sixteen significantly overrepresented motifs (E-value < 10^{-161}) of 11-50 residues were identified, present in 7-182 of the sequences (Supporting Information Table S8). Motif composition and arrangement (Fig. 2c) were in good agreement with the gene tree (Fig. 2a).

Using HMMs describing subdomains A to E of the NAC domain (Ooka *et al.*, 2003), we assigned Motif 1 to subdomain A, Motif 2 to subdomain B, Motif 3 and Motif 4 to subdomain C, Motif 5 and Motif 6 to subdomain D, and Motif 7 to subdomain E. As expected, these motifs occurred in the N-terminal half of EgrNAC sequences (Fig. 2c). Because they were also found in the majority (> 65%) of EgrNAC sequences (Supporting Information Table S8), they were classified as "general motifs". The remaining "specific motifs" (Motif 8 through Motif 16) were restricted to 7-20 EgrNAC proteins. Aside from Motif 9, none of the specific motifs had any hits (E-value < 0.01) to Pfam-A or Pfam-B databases (http://pfam.sanger.ac.uk), and occurred in the diverse C-terminal region, outside the NAC domain (Fig. 2c). Similarly, Motif 9 aside, none of the specific motifs matched those previously identified by Ooka *et al.* (2003), Fang *et al*. (2008), Jensen *et al.* (2010) or Hu *et al.* (2010).

The distribution of the motifs in other plant genomes was assessed using HMMs designed from the *Eucalyptus* alignments of each motif. Matching motifs (E-value < 0.01) were identified in the NAC proteins from *Populus*, *Glycine*, *Arabidopsis*, *Carica*, *Vitis*, *Oryza*, *Brachypodium* and *Zea*, as well as *Eucalyptus* as a positive control. The general motifs corresponding to

subdomains A through E of the NAC domain were detected at high frequencies in NAC proteins from all eight genomes as expected, while specific motifs appeared enriched in *Eucalyptus* compared to other genomes, with the exception of Motif 9 (Table 1). The latter, which was found in a minority of NAC proteins in all nine genera, is homologous (E-value = 8.4×10^{-5}) to the NAM domain (PF02365) and appears to replace Motifs 2, 3 and 4 (Fig. 2c). Motif 10 was present in all dicots, while 13 and 14 were only present in some dicots (Table 1). Motifs 12, 15 and 16 were exclusively found in EgrNAC proteins and may thus represent motifs unique to *Eucalyptus* (Table 1). It is unlikely that *Eucalyptus*-specific motifs were an artefact of HMMs built on *E. grandis* alignments, since no bias was observed in the cumulative frequency of general motifs identified in *E. grandis* compared to other genomes (Supporting Information Fig. S3).

Some NAC proteins are tethered to the endoplasmic reticulum or plasma membrane via transmembrane helices (TMHs) (Chen *et al.*, 2008; Seo *et al.*, 2008). We identified seven putative membrane-tethered EgrNAC proteins, all with single C-terminal TMHs comprising at least twenty residues (Fig. 2c). All EgrNAC proteins with predicted TMHs had putative *Arabidopsis* orthologs known to contain TMHs (Kim *et al.*, 2010) (Supporting Information Table S9). Interestingly, multiple membrane-tethered *Arabidopsis* co-orthologs were found for each TMH-containing EgrNAC protein (Supporting Information Table S9). All except one TMH-containing EgrNAC protein occurred in subfamily IIIa/b, which prominently features NACs involved in stress response and development (Supporting Information Table S5).

We next analysed the conservation of intron and exon arrangements in the *EgrNAC* genes (Fig. 2d). An average of 3.3 exons was observed, similar to most *Arabidopsis* NAC genes (Duval *et al.*, 2002), ranging from one to eleven. The numbers of exons were similar between closely related genes. Intron phase was also well conserved (Supporting Information Fig. S4), as reported previously for *Populus* NAC domain genes (Hu *et al.*, 2010).

Physical distribution of *Eucalyptus* NAC genes

A large proportion (34%) of genes in the *E. grandis* genome have expanded through tandem duplication (Myburg *et al.*, 2014). We studied the distribution of the 185 *E. grandis* NAC domain genes located on the eleven main chromosome scaffolds in the v.1.0 annotation (Fig. 3).

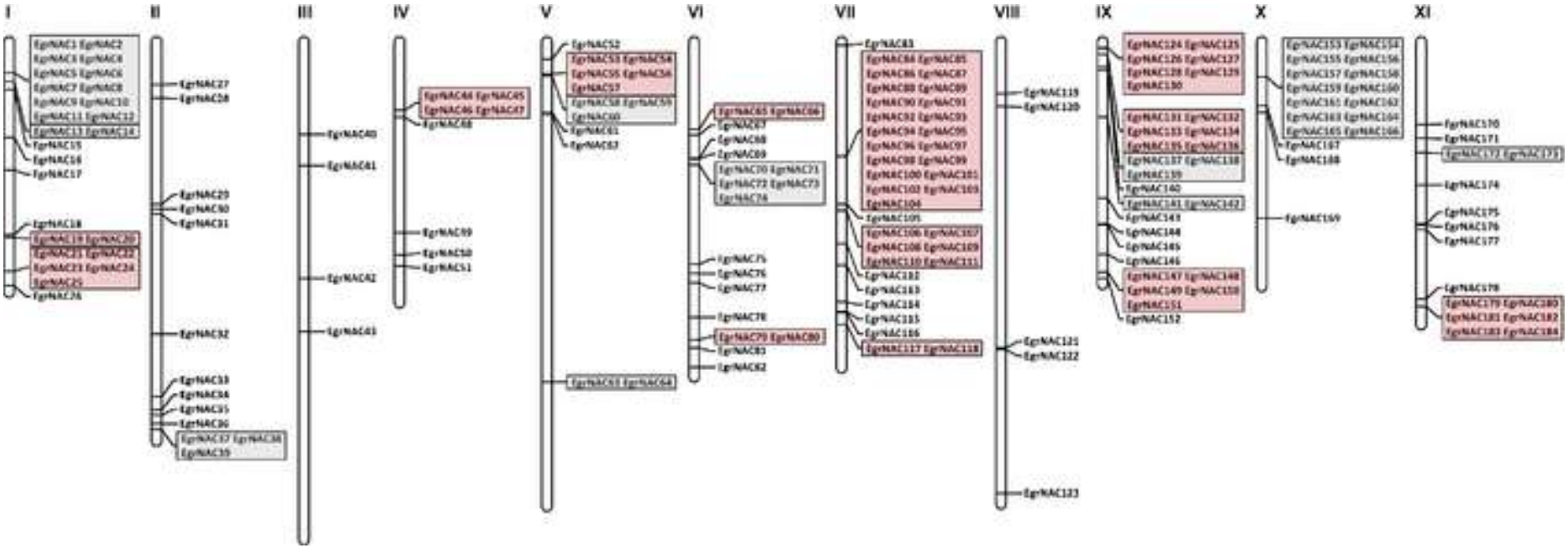


Fig. 3. Chromosomal locations of *EgrNAC* genes. Tandem duplicates are represented by shaded blocks; red shading indicates blocks with dN/dS < 1.0 ($P < 0.05$; codon-based Z-test). *P*-values for individual pairs of tandem duplicate pairs are available in Supporting Information Table S11.

We defined tandem duplicates according to Hanada *et al.* (2008) as pairs of NAC gene models within 100 kb of each other, having ten or fewer non-homologous genes in-between. Based on this definition, 121 (~64%) of the NAC domain genes were distributed amongst 23 blocks of tandem duplicate arrays of 2-21 members (Fig. 3). In most cases, the members of each tandem array belonged to a single subfamily (Supporting Information Table S10).

The fate of tandem duplicates include nonfunctionalization, neofunctionalization, subfunctionalization and redundancy (Rastogi & Liberles, 2005). To assess if members of tandem duplicate arrays are under natural selection in *E. grandis*, we used a codon-based Z-test (Tamura *et al.*, 2011) based on the ratio of nonsynonymous to synonymous substitution rates between all pairs of sequences in each tandem array. Out of the 23 tandem arrays, a test for overall purifying selection between pairs of genes in a given array was significant for 13 arrays ($P < 0.05$; Fig. 3). *P*-values for individual gene pairs in each of these arrays, which were not all significant alone, are shown in Supporting Information Table S11. No evidence of positive selection acting on any of the arrays overall or between any pairs of sequences in a given array was detected. These results suggest that most tandem arrays are under purifying selection.

Next, we analysed the expression patterns of *EgrNAC* paralogs in the seven EucGenIE tissues (http://eucgenie.bi.up.ac.za/) and compared them to public expression data of close homologs in *Arabidopsis*, *Oryza*, and *Populus*. Here, we define paralogs as proteins that are more closely related to each other than to homologs from the other genomes (inferred from Supporting Information Fig. S1). *EgrNAC* paralogs were composed mostly of tandem duplicate arrays. Most groups of *EgrNAC* paralogs featured a "dominant" transcript at a marked level of expression, accompanied by paralogs with reduced overall expression of similar, slightly diverged or undetected transcript abundance (Supporting Information Fig. S5). Expression data for *Arabidopsis*, *Oryza* and *Populus* homologs from Genevestigator (Hruz *et al.*, 2008) and Poplar Expression Angler (Toufighi *et al.*, 2005; Wilkins *et al.*, 2009) are shown for similar tissues, where available, underneath each *EgrNAC* paralog group (Supporting Information Fig. S5).

Paralog groups a, b, f, i, n and o contained one or two "dominant" *EgrNAC* transcripts expressed similarly to at least one non-*Eucalyptus* homolog (Supporting Information Fig. S5). However, we observed conflicting expression patterns between *EgrNAC* genes and non-

Eucalyptus homologs in groups d, h and m, suggesting functional divergence. Also, certain transcripts in group b (*EgrNAC24*), e (*EgrNAC43*, *EgrNAC141*, *EgrNAC154*, and *EgrNAC157*) and i (*EgrNAC50*, *EgrNAC59*, and several root-specific transcripts) showed expression patterns differing from *Eucalyptus* paralogs and non-*Eucalyptus* homologs (Supporting Information Fig. S5), three of which are wood development candidates (described above).

Expression characteristics of *E. globulus* NAC domain genes

Besides having superior wood properties, *E. globulus* is more frost-tolerant than *E. grandis*, but still suffers from frost damage (Hasey & Connor, 1990; Skolmen & Ledig, 1990; Tibbits *et al.*, 2006). To better understand NAC gene functions in eucalypts, we profiled the expression patterns of *E. globulus* orthologs in various tissues and in response to cold and tension stress. We used Fluidigm RT-qPCR to analyse expression patterns of 33 *EgrNAC* orthologs in *E. globulus* ("*EglNAC*"). Transcript profiles across nine *E. globulus* tissues were hierarchically clustered, revealing three prominent expression clusters: xylem-enriched, xylem-deficient, and unassigned (Fig. 4). These are robust clusters, since tissues were sampled from trees of different ages, genotypes and sites (see Methods). *EglNAC* transcripts were also quantified in cold-treated trees relative to control in primary stems, secondary stems, young leaves and roots. Three genes in primary stems, two in secondary stems, five in young leaves and five in roots showed a significant transcriptional response to cold treatment (Fig. 5).

Candidate *EglNAC* genes involved in tension wood formation were identified by comparing selected *EglNAC* transcripts in upright stem xylem to those in xylem from tension and opposite wood in an *E. globulus* bending trial. Two genes were significantly upregulated (*EglNAC31* and *XND1* homolog *EglNAC152*) and two downregulated (*SND3* homolog *EglNAC44* and *XND1* homolog *EglNAC139*) in tension wood relative to upright control (Fig. 6). In opposite wood, *EglNAC139* and *EglNAC141* were downregulated and upregulated relative to upright control, respectively (Fig. 6).

We assessed the evolutionary conservation of NAC gene expression between *Eucalyptus* species by comparing the correlation of expression of *EglNAC* and *EgrNAC* orthologs. We compared the xylem/leaf expression ratio calculated from Fluidigm and RNA-seq data (see Methods) to account for developmental variation and environmental effects. These data

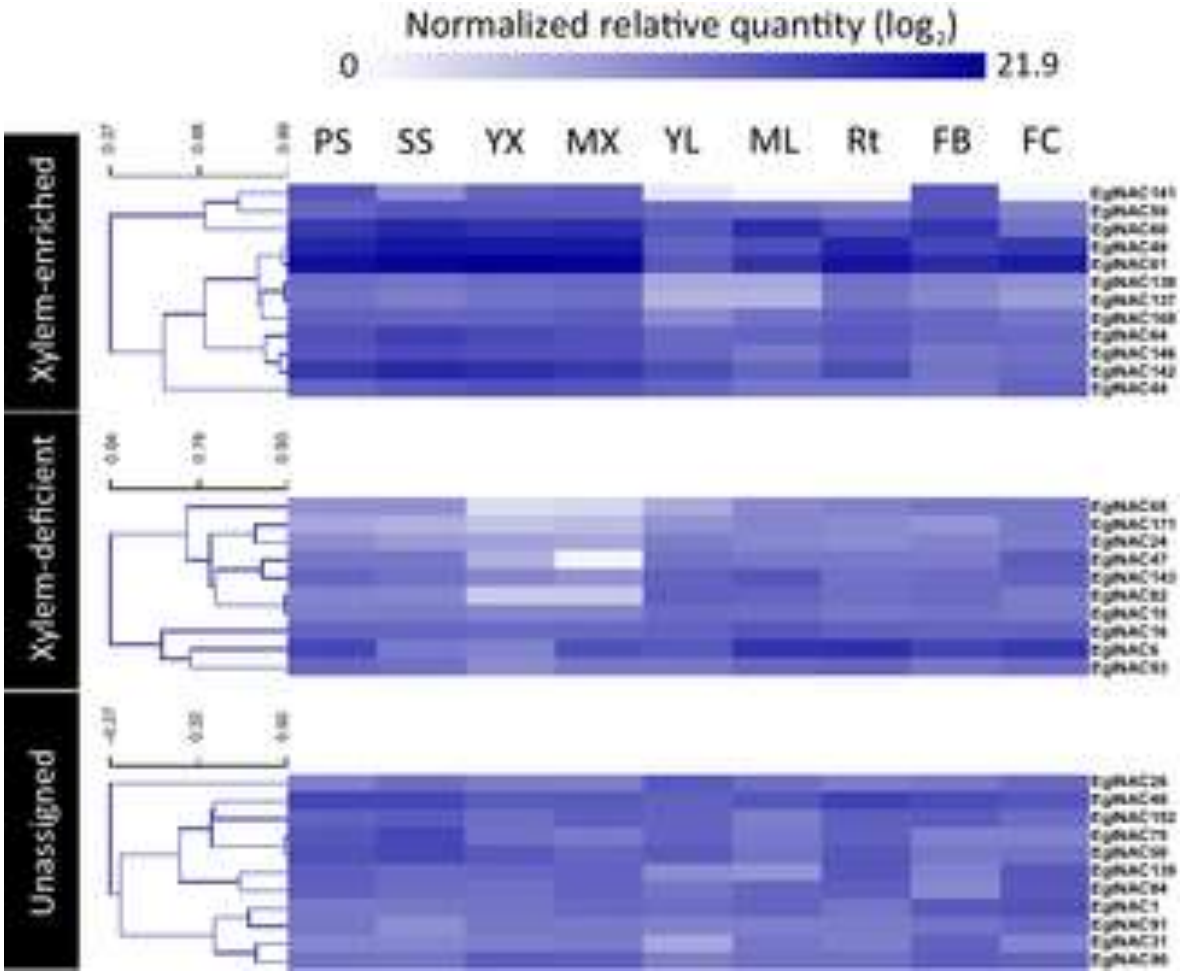


Fig. 4. Tissue and organ expression data for 33 *EgrNAC* orthologs from *E. globulus*, herein denoted *EglNAC*. Fluidigm RT-qPCR data were hierarchically clustered using a quality threshold (QT) algorithm. PS, primary stem; SS, secondary stem; YX, young xylem; MX, mature xylem; YL, young leaf; ML, mature leaf; Rt, root; FB, flower bud; FC, flower capsule. Branch distances indicate Pearson's correlation.

correlated significantly (Supporting Information Fig. S6; $r = 0.69$, $P < 0.0001$), suggesting a high overall conservation in expression of *EgrNAC* and *EglNAC* orthologs. Amongst the exceptions, *EgrNAC142* expression was not detected using RNA-seq in *E. grandis*, while that of its ortholog *EglNAC142* was detected by RT-qPCR analysis in *E. globulus* secondary tissues (Fig. 4).

Discussion

As representative of the Myrtales, an order basal to the core rosids (Myburg *et al.*, 2014), the *E. grandis* genome adds resolution to understanding gene family structure and function evolution in the rosids. We identified 189 nonredundant NAC domain proteins in the *E. grandis* genome, one of the largest NAC domain families known (Jin *et al.*, 2014). We modelled our subfamily annotation on that proposed by Zhu *et al.* (2012), which is based on nine diverse lineages and good bootstrap support for each subfamily. We included novel clustering such as the merging of subfamilies IIIa and IIIb, the subdivision of subfamily VIII and the annotation of a new subfamily, XI, resulting in 22 well-supported subfamilies in our study. The number of subfamilies proposed for the NAC domain family in different plants has been highly variable (Ooka *et al.*, 2003; Mitsuda *et al.*, 2005; Fang *et al.*, 2008; Rushton *et al.*, 2008; Pinheiro *et al.*, 2009; Hu *et al.*, 2010; Nuruzzaman *et al.*, 2010; Zhu *et al.*, 2012), and additional modifications will likely be proposed in the future as more plant genomes are analysed.

However defined, most NAC subfamilies are represented in angiosperm genomes (Shen *et al.*, 2009; Zhu *et al.*, 2012), with only one example of a subfamily unique to a lineage, the Solanaceae (Rushton *et al.*, 2008; Singh *et al.*, 2013). We found no evidence of a *Eucalyptus*-specific subfamily, although expansion of EgrNAC proteins was apparent in subfamilies IVa, IVc, Vb and VII due to five, three, two and four arrays of tandem duplication, respectively (Fig. 1b, Supporting Information Table S10). Subfamily VII was hypothesized to represent a tree-specific expansion involved in the regulation of wood formation (Zhu *et al.*, 2012). Although no functional data is yet available, the apparent expansion of EgrNAC and PNAC sequences in this subfamily supports this hypothesis, with at least two *EgrNAC* members (*EgrNAC58*, *EgrNAC59*) specifically expressed in developing xylem (Fig. 2a) and one upregulated in tension wood (discussed below). Our analysis shows that this tree-specific expansion occurred after *Vitis*-*Eucalyptus* divergence but before the Myrtales-Eurosid split. Although most subfamily VII genes

were expressed only in roots (Fig. 2b), these organs contain significant amounts of lignified vascular tissue. Similarly the expression patterns of most subfamily IVc genes duplicated in *Eucalyptus* were biased toward vascular tissues (roots, phloem, developing xylem) (Fig. 2b). Subfamilies firmly associated with transcriptional regulation of SCW formation also contained small-scale expansions, resulting in five SND3 co-orthologs in subfamily II, and four co-orthologs of XND1 in subfamily VIc (Supporting Information Table S6), some *E. globulus* orthologs of which we implicate in tension wood formation (*EglNAC44*, *EglNAC139*, *EglNAC152*; Fig. 6). In general, however, subfamilies with members known to regulate wood formation (Ic, II, VIc) did not exhibit notable expansion in *Eucalyptus*. Since the expanded subfamilies IVa and Vb contain members involved in stress response (Supporting Information Table S5), we hypothesize that the large blocks of tandem duplications contained within them reflect adaptations to environmental stress. A predominant stress response function for retained duplicates has been observed previously in model representatives across the three domains of life (Kondrashov *et al.*, 2002).

Over half of the tandem duplicate arrays showed significant overall purifying selection, suggesting that at least some of the retained duplicates are still functional and may provide adaptive advantages. Functional buffering, protein dosage benefits or subfunctionalization of paralogs could serve as the basis for this retention. Interestingly, paralogs with detected expression frequently exhibited dissimilar expression profiles (Supporting Information Fig. S5). This has previously been shown to be more common of small-scale duplicates than whole-genome duplications (Casneuf *et al.*, 2006), and is therefore an expected result due to the large number of tandem duplicates amongst these paralogs. Diverged expression may be explained by a transcriptional version of the duplication-degeneration-complementation (DDC) model of subfunctionalization, whereby duplicates accumulate deleterious mutations in regulatory regions that result in the complementary expression of functionally redundant copies in different tissues, and the subsequent retention of these complementary duplicates through purifying selection (Force *et al.*, 1999). For example, most *Arabidopsis* tandem duplicates have undergone rapid expression divergence in accordance with the DDC model (Haberer *et al.*, 2004). Protein sequence evolution of duplicated genes, rather than divergence in gene expression, is a more prominent mechanism towards morphological diversification (Hanada *et al.*, 2009). No evidence for positive selection (i.e. neofunctionalization of paralogs) was found amongst the tandem

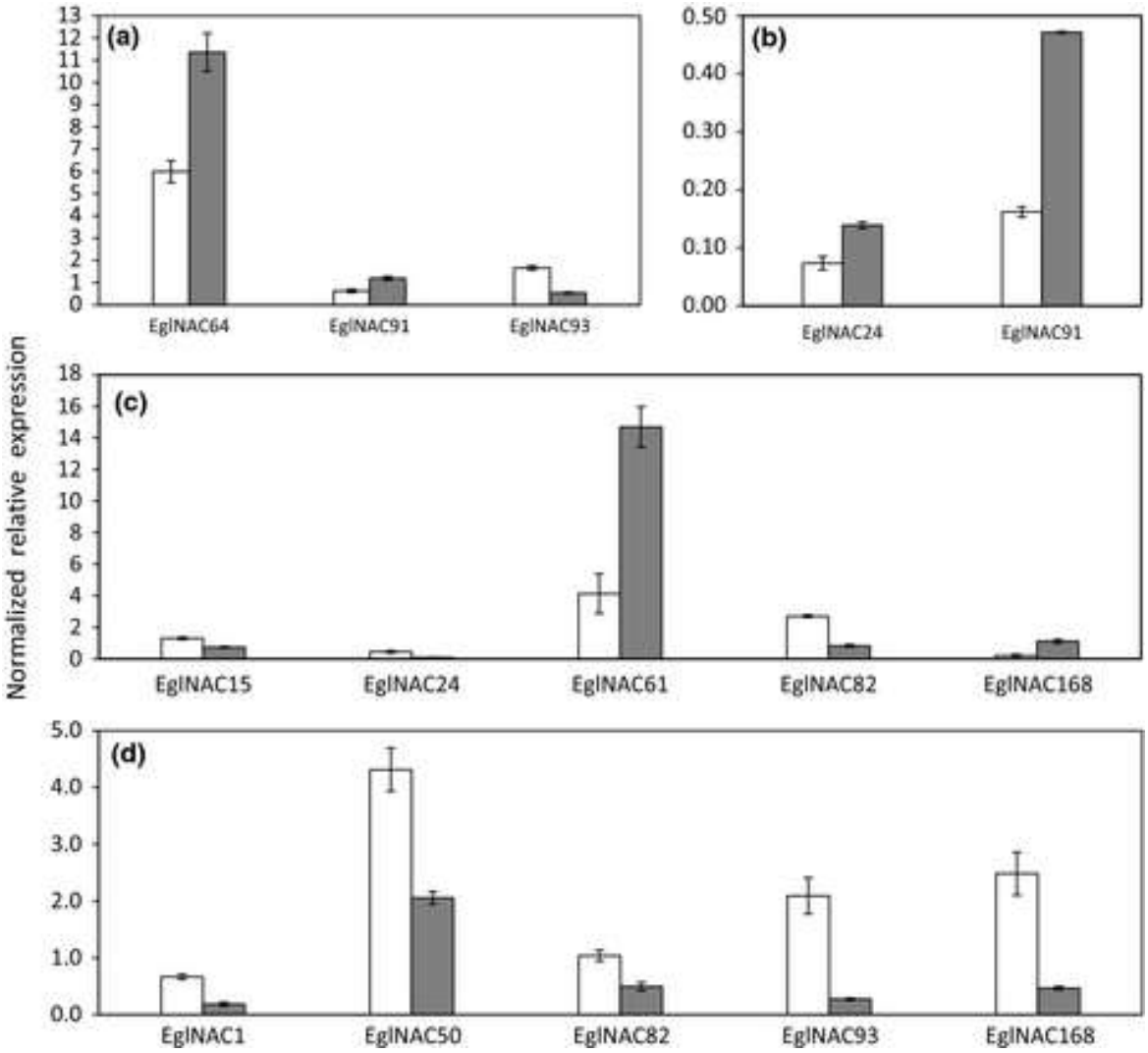


Fig. 5. *Eucalyptus globulus* NAC *EglNAC* genes showing a positive or negative transcriptional response towards cold treatment in primary stems (a), secondary stems (b), young leaves (c) and roots (d) in *E.globulus*. Error bars indicate the standard deviation of three biological replicates. A Bonferroniadjusted *P*-value of 0.05 was applied to a two-tailed Student's *t* test for each of 33 *EglNAC* genes. Only genes exhibiting significant responses are shown.

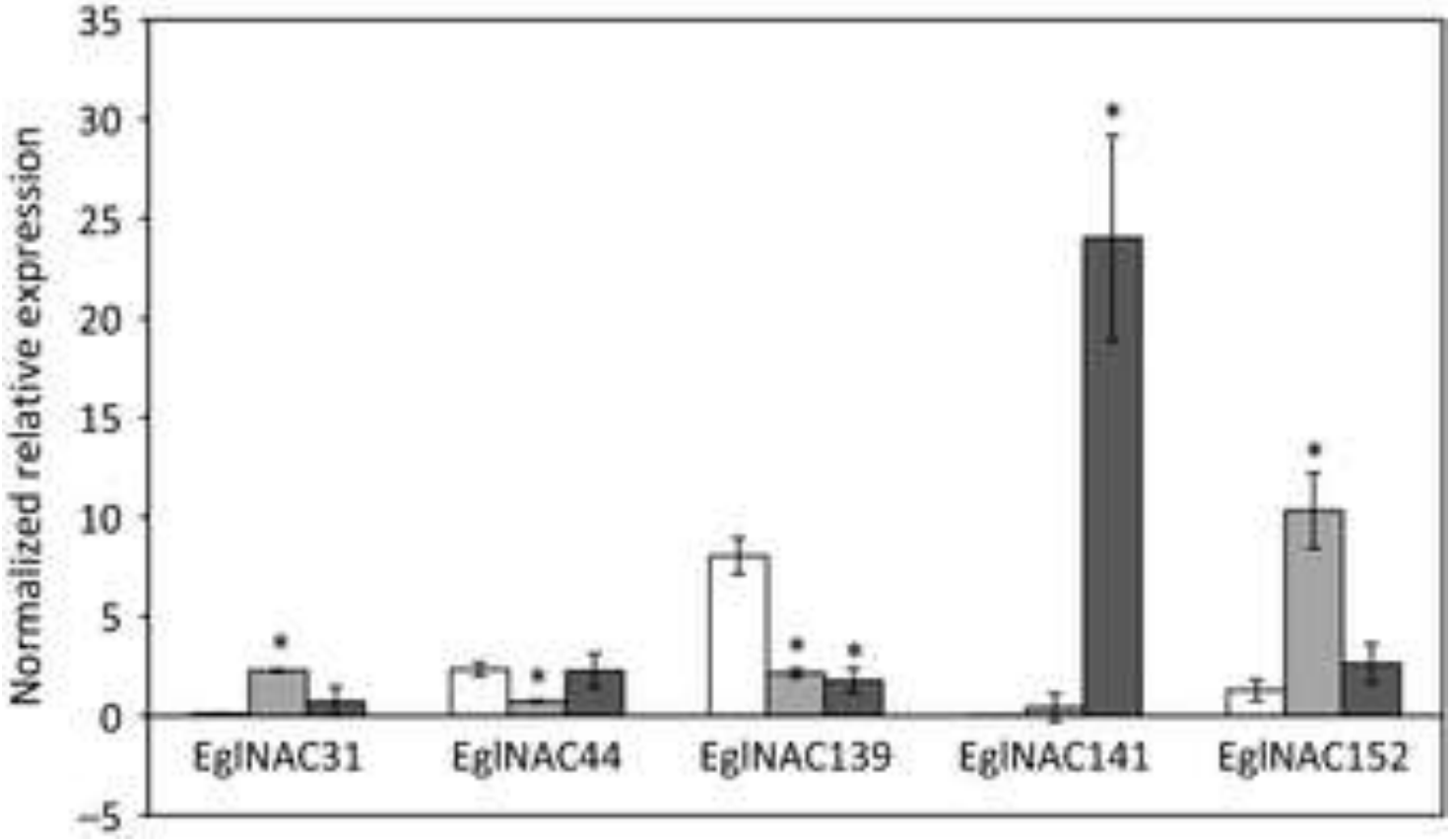


Fig. 6. Transcriptional profile of *EglNAC* genes in xylem tissue from tension or opposite wood relative to upright control in *E. globulus*. Error bars indicate standard deviation across three genotypes, using the average value of three ramets per genotype. *Significant difference relative to upright control according to two-tailed Student's *t* test, using a Bonferroni-corrected *P*-value (*P** = 0.05/33).

duplicates, but positive selection is rare, episodic and difficult to detect (Raes & van de Peer, 2003). Furthermore, subfunctionalization is considered a temporary state that ultimately facilitates the acquisition of novel functions (Rastogi & Liberles, 2005). Functional analysis of paralogous *EgrNAC* genes will help to unravel whether functional diversification has occurred.

The absence of detectable expression for 19 genes, all of which were found in tandem duplicate arrays, is suggestive of nonfunctionalization. However, most of these genes were predicted to encode full-length proteins containing the amino acid motifs present in related, expressed paralogs (Fig. 2c), suggesting that these genes may be functional. Five of these tandem duplicates tested statistically significant for purifying selection on their coding regions (Supporting Information Table S11), further suggesting a functional role. It is quite likely that these genes are expressed in tissues, cell types or conditions not sampled in this study. For example, removing just one tissue dataset, root, increases the proportion of non-expressed genes from ~10% to ~20%. Alternatively, the penetrance of paralog expression may differ between populations, genotypes or species. For example, *EgrNAC142* might be classified as a pseudogene from the *E. grandis* RNA-seq dataset, but a possible xylogenesis regulator from the *E. globulus* expression data (Fig. 2b, Fig. 4).

Novel conserved and *Eucalyptus*-specific amino acid motifs were identified in the C-termini of EgrNAC proteins (Table 1). These motifs were restricted to particular clades of EgrNAC proteins (Fig. 2c) and, although their functions are yet to be investigated, may participate in protein-protein interactions or the transcriptional activation or repression activities of the associated members. Transmembrane helices (TMHs) are also relevant to transcriptional regulation because they facilitate rapid post-translational recruitment of transcription factors tethered to intracellular membranes (Seo *et al.*, 2008). In agreement with a predominant role of such proteins in stress response (Chen *et al.*, 2008), all EgrNAC proteins with predicted TMHs are homologous to *Arabidopsis* TMH-containing NACs (Supporting Information Table S9), and most belong to the stress and developmental process-associated subfamily IIIa/b (Supporting Information Table S5). Membrane tethering is therefore not a mechanism for regulatory novelty in *E. grandis*.

We showed that, between *E. grandis* and *E. globulus* tissues, orthologous NAC domain genes have correlated expression, suggesting that orthologous genes perform similar functions across *Eucalyptus* species. We implicated ten *EglNAC* genes in the transcriptional response to cold stress in leaves, stems and roots in *E. globulus*, none of which had close *Arabidopsis* homologs known to play a role in abiotic stress response. Five of these genes responded to cold treatment in more than one tissue, but two candidates (*EglNAC24* and *EglNAC168*) were differentially regulated in opposing directions in different tissues (Fig. 5). Interestingly, transcripts of *Arabidopsis* homologs of (secondary) cell wall-associated NAC genes *SND1* (*EgrNAC61*), *SND3* (*EglNAC64*) and *VND4*/*VND5* (*EglNAC50*) were also affected by cold treatment (Fig. 5). It is known that lignin content and composition, the accumulation of phenylpropanoid pathway derivatives and increased phenylalanine ammonia-lyase (PAL) activity is associated with cold stress (reviewed by Moura *et al.*, 2010). Possibly, these SCW-associated candidates prepare the tree for frost exposure through their direct or indirect regulation of the phenylpropanoid pathway. Other *E. globulus* homologs of *SND3* (*EglNAC44*) and *XND1* (*EglNAC139*, *EglNAC152*) showed transcriptional responses to tension and/or opposite wood formation (Fig. 6), as observed for *Populus* homologs of *SND3* and *XND1* (Grant *et al.*, 2010). These candidates regulate several secondary wall-associated genes in *Arabidopsis* and *Populus* (Zhao *et al.*, 2008; Zhong *et al.*, 2011). Transcript profiles of *EglNAC31* (of subfamily VII discussed above) and *EglNAC141* (subfamily IVc), which have no close *Arabidopsis* homologs, suggest they are novel candidates for regulating tension and opposite wood, respectively (Fig. 6), and strengthen the hypothesis that members of expanded subfamilies IVc and VII may play a role in regulating specialized aspects of wood formation in *Eucalyptus*.

The NAC gene family of *E. grandis* is one of the largest described to date. Pervasive tandem, and less frequent segmental, duplications have contributed significantly to *EgrNAC* expansion, a tenth of which appear to be transcriptionally silent in deeply sequenced *E. grandis* RNA-seq data. Although the functions of most *EgrNAC* genes remain to be investigated, limited duplication of homologs of known regulators of SCW biosynthesis suggests functional conservation, while subfamilies with paralog expansion appear to be associated with abiotic and biotic stress responses as well as vascular development, such as tension and opposite wood formation. It is thus postulated that duplication of *EgrNAC* genes has contributed to the favourable woody traits of *Eucalyptus* and its adaptation to the diverse and often harsh

Australian climate. Divergent expression and evidence of purifying selection acting on most groups of paralogs suggests a complex interplay of subfunctionalization, functional redundancy and nonfunctionalization in their evolution. Our analysis supports the existence of a tree-specific subfamily of NAC domain proteins, dating its appearance prior to Myrtales-Eurosid divergence. Additionally, several new candidates for vascular development, tension wood formation and cold response were found. Our study provides a first-level understanding of how one of the largest transcription factor families in plants may have contributed to the evolutionary success of the Myrtaceae and the accomplishment of *Eucalyptus* as a global fiber crop.

Acknowledgements

The authors thank J.P Paiva and Altri-Florestal (Portugal) for kindly providing samples from an *E. globulus* tension wood experiment, and H. San Clemente and B. Savelli for bioinformatic assistance. Thanks to N. Ladouce and the Plateforme Génomique Génopole Toulouse/Midi-Pyrénées (Genotoul) for technical assistance with Biomark Fluidigm dynamic arrays. We thank Albe van der Merwe (UP) for his assistance with the phylogenetic analysis. JGP and MNS acknowledge the financial support of the ANR (Project "Tree For Joules" ANR-2010-KBBE-007-01; Labex Tulip ANR-10-LABX-41), the CNRS, and the University Toulouse III (UPS). SH and AAM acknowledge funding from the Bioinformatics and Functional Genomics Programme of the National Research Foundation, South Africa (UID 71255). *E. grandis* sequence data were produced by the US Department of Energy Joint Genome Institute http://www.jgi.doe.gov/ in collaboration with the user community.

References

Aida M, Ishida T, Fukaki H, Fujisawa H, Tasaka M. 1997. Genes involved in organ separation in *Arabidopsis*: an analysis of the *cup-shaped cotyledon* mutant. *The Plant Cell* **19**(6): 841-857.

Anisimova M, Gascuel O. 2006. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Systematic Biology* **55**(4): 539-552.

Bailey TL, Williams N, Misleh C, Li WW. 2006. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Research* **34**: W369-W373.

Beadle NCW. 1981. *The Vegetation of Australia*. London: Cambridge University Press.

Bennett T, Toorn Avd, Sanchez-Perez GF, Campilho A, Willemsen V, Snel B, Scheres B. 2010. SOMBRERO, BEARSKIN1, and BEARSKIN2 regulate root cap maturation in *Arabidopsis*. *The Plant Cell* **22**: 640-654.

Berger Y, Harpaz-Saad S, Brand A, Melnik H, Sirding N, Alvarez JP, Zinder M, Samach A, Eshed Y, Ori N. 2009. The NAC-domain transcription factor GOBLET specifies leaflet boundaries in compound tomato leaves. *Development* **136**: 823-832.

Boyce CK, Cody GD, Fogel ML, Hazen RM, Alexander CMOD, Knoll AH. 2003. Chemical evidence for cell wall lignification and the evolution of tracheids in Early Devonian plants. *International Journal of Plant Sciences* **164**(5): 691-702.

Carroll A, Somerville C. 2009. Cellulosic biofuels. *Annual Review of Plant Biology* **60**: 165-182.

Cary G, Lindenmayer D, Dovers S. 2003. *Australia burning: fire ecology, policy and management issues*. Melbourne: CSIRO Publishing.

Casneuf T, De Bodt S, Raes J, Maere S, van de Peer Y. 2006. Nonrandom divergence of gene expression following gene and genome duplications in the flowering plant *Arabidopsis thaliana*. *Genome Biology* **7**: R13.

Cassan-Wang H, Soler M, Yu H, Camargo ELO, Carocha V, Ladouce N, Savelli B, Paiva JAP, Leplé J-C, Grima-Pettenati J. 2012. Reference genes for high-throughput quantitative reverse transcription–PCR analysis of gene expression in organs and tissues of *Eucalyptus* grown in various environmental conditions. *Plant and Cell Physiology* **53**(12): 2101-2116.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* **17**(4): 540-552.

Chen Y-N, Slabaugh E, Brandizzi F. 2008. Membrane-tethered transcription factors in *Arabidopsis thaliana*: novel regulators in stress response and development. *Current Opinion in Plant Biology* **11**: 695-701.

Chippendale GM 1988. *Eucalyptus, Angophora* (Myrtaceae). In: A. S. George ed. *Flora of Australia*. Melbourne: Brown Prior Anderson Pty Ltd, 198.

Christiansen MW, Holm PB, Gregersen PL. 2011. Characterization of barley (*Hordeum vulgare* L.) NAC transcription factors suggests conserved functions compared to both monocots and dicots. *BMC Research Notes* **4**: 302.

de Oliveira TM, Cidade LC, Gesteira AS, Filho MAC, Filho WSS, Costa MGC. 2011. Analysis of the NAC transcription factor gene family in citrus reveals a novel member involved in multiple abiotic stress responses. *Tree Genetics and Genomes* **7**: 1123-1134.

Duval M, Hsieh T-F, Kim SY, Thomas TL. 2002. Molecular characterization of *AtNAM*: a member of the *Arabidopsis* NAC domain superfamily. *Plant Molecular Biology* **50**: 237-248.

Fang Y, You J, Xie K, Xie W, Xiong L. 2008. Systematic sequence analysis and identification of tissue-specific or stress-responsive genes of NAC transcription factor family in rice. *Molecular Genetics and Genomics* **280**(6): 547-563.

Fink J, Hamilton N. 2007. DomainDraw: A macromolecular schematic drawing program. *In Silico Biology* **7**: 14.

Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research* **39**: W29-W37.

Force A, Lynch M, Pickett FB, Amores A, Yan Y-l, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**: 1531-1545.

Gandolfo MA, Hermsen EJ, Zamaloa MC, Nixon KC, González CC, Wilf P, Cúneo NR, Johnson KR. 2011. Oldest known *Eucalyptus* macrofossils are from South America. *PLoS ONE* **66**(6): e21084.

Grant EH, Fujino T, Beers EP, Brunner AM. 2010. Characterization of NAC domain transcription factors implicated in control of vascular cell differentiation in *Arabidopsis* and *Populus*. *Planta* **232**: 337-352.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* **59**(3): 307-321.

Guo A, Zhu Q, Chen X, Luo J. 2007. GSDS: a gene structure display server. *Yi Chuan* **29**(8): 1023-1026.

Guo Y, Gan S. 2006. AtNAP, a NAC family transcription factor, has an important role in leaf senescence. *The Plant Journal* **46**(4): 601-612.

Haberer G, Hindemitt T, Meyers BC, Mayer KFX. 2004. Transcriptional similarities, dissimilarities, and conservation of *cis*-elements in duplicated genes of *Arabidopsis*. *Plant Physiology* **136**: 3009-3022.

Hanada K, Kuromori T, Myouga F, Toyoda T, Shinozaki K. 2009. Increased expression and protein divergence in duplicate genes is associated with morphological diversification. *PLoS Genetics* **5**(12): e1000781.

Hanada K, Zou C, Lehti-Shiu MD, Shinozaki K, Shiu S-H. 2008. Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. *Plant Physiology* **148**: 993-1003.

Hasey JK, Connor JM. 1990. *Eucalyptus* shows unexpected cold tolerance. *California Agriculture* **44**(2): 25-27.

Hefer C, Mizrachi E, Joubert F, Myburg A. 2011. The *Eucalyptus* genome integrative explorer (EucGenIE): a resource for *Eucalyptus* genomics and transcriptomics. *BMC Proceedings* **5**(Suppl 7): O49.

Hinchee M, Rottmann W, Mullinax L, Zhang C, Chang S, Cunningham M, Pearson L, Nehra N. 2009. Short-rotation woody crops for bioenergy and biofuels applications. *In Vitro Cellular & Developmental Biology - Plant* **45**(6): 619-629.

Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L, Widmayer P, Gruissem W, Zimmermann P. 2008. Genevestigator V3: A reference expression database for the meta-analysis of transcriptomes. *Advances in Bioinformatics* **2008**: 420747.

Hu R, Qi G, Kong Y, Kong D, Gao Q, Zhou G. 2010. Comprehensive analysis of NAC domain transcription factor gene family in *Populus trichocarpa*. *BMC Plant Biology* **10**: 145.

Hussey SG, Mizrachi E, Creux NM, Myburg AA. 2013. Navigating the transcriptional roadmap regulating plant secondary cell wall deposition. *Frontiers in Plant Science* **4**: 325.

Iglesias-Trabado G, Wilstermann D 2008. *Eucalyptus universalis*. Global cultivated eucalypt forests map 2008 Version 1.0.1. In GIT Forestry Consulting's EUCALYPTOLOGICS: Information resources on *Eucalyptus* cultivation worldwide. Available online at http://www.git-forestry.com [accessed 26 September 2014].

Jensen M, Kjaersgaard T, Nielsen M, Galberg P, Petersen K, O'Shea C, Skriver K. 2010. The *Arabidopsis thaliana* NAC transcription factor family: structure-function relationships and determinants of ANAC019 stress signalling. *Biochemical Journal* **426**: 183-196.

Jensen MK, Hagedorn PH, de Torres-Zabala M, Grant MR, Rung JH, Collinge DB, Lyngkjaer MF. 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**(6): 867-880.

Jin J, He Zhang LK, Gao G, Luo J. 2014. PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Research* **42**(D1): D1182-D1187.

Kemp EM 1981. Tertiary palaeogeography and the evolution of the Australian climate. In: A. Keast ed. *Ecological Biogeography of Australia*. The Hague: Dr. W. Junk bv Publishers, 33-46.

Kim S-G, Lee S, Seo PJ, Kim S-K, Kim J-K, Park C-M. 2010. Genome-scale screening and molecular characterization of membrane-bound transcription factors in *Arabidopsis* and rice. *Genomics* **95**: 56-65.

Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV. 2002. Selection in the evolution of gene duplications. *Genome Biology* **3**(2): research0008.0001–0008.0009.

Kou X, Wang S, Wu M, Guo R, Xue Z, Meng N, Tao X, Chen M, Zhang Y. 2014. Molecular characterization and expression analysis of NAC family transcription factors in tomato. *Plant Molecular Biology Reporter* **32**(501-516).

Le DT, Nishiyama R, Watanabe Y, Mochida K, Yamaguchi-Shinozaki K, Shinozaki K, Tran L-SP. 2011. Genome-wide survey and expression analysis of the plant-specific NAC transcription factor family in soybean during development and dehydration stress. *DNA Research* **18**(4): 263-276.

Li W-H. 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *Journal of Molecular Evolution* **36**: 96-99.

Li X, Chapple C. 2010. Understanding lignification: challenges beyond monolignol biosynthesis. *Plant Physiology* **154**: 449-452.

Liu T, Song X, Duan W, Huang Z, Liu G, Li Y, Hou X. 2014. Genome-wide analysis and expression patterns of NAC transcription factor family under different developmental stages and abiotic stresses in Chinese cabbage. *Plant Molecular Biology Reporter* **32**(5): 1041-1056.

Ma J, Wang F, Li M-Y, Jiang Q, Tan G-F, Xiong A-S. 2014. Genome wide analysis of the NAC transcription factor family in Chinese cabbage to elucidate responses to temperature stress. *Scientia Horticulturae* **165**: 82-90.

Meng C, Cai C, Zhang T, Guo W. 2009. Characterization of six novel NAC genes and their responses to abiotic stresses in *Gossypium hirsutum* L. *Plant Science* **176**(3): 352-359.

Mitsuda N, Seki M, Shinozaki K, 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.

Mizrachi E, Hefer CA, Ranik M, Joubert F, Myburg AA. 2010. *De novo* assembled expressed gene catalog of a fast-growing *Eucalyptus* tree produced by Illumina mRNA-Seq. *BMC Genomics* **11**: 681.

Mizrachi E, Mansfield SD, Myburg AA. 2012. Cellulose factories: advancing bioenergy production from forest trees. *New Phytologist* **194**(1): 54-62.

Morishita T, Kojima Y, Maruta T, Nishizawa-Yokoi A, Yabuta Y, Shigeoka S. 2009. *Arabidopsis* NAC transcription factor, ANAC078, regulates flavonoid biosynthesis under high-light. *Plant and Cell Physiology* **50**(12): 2210-2222.

Moura JCMS, Bonine CAV, Viana JdOF, Dornelas MC, Mazzafera P. 2010. Abiotic and biotic stresses and changes in the lignin content and composition in plants. *Journal of Integrative Plant Biology* **52**(4): 360-376.

Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, Jenkins J, Lindquist E, Tice H, Bauer D, Goodstein DM, Dubchak I, Poliakov A, Mizrachi E, Kullan ARK, van Jaarsveld I, Hussey SG, Pinard D, Merwe Kvd, Singh P, Silva-Junior OB, Togawa RC, Pappas MR, Faria DA, Sansaloni CP, Petroli CD, Yang X, Ranjan P, Tschaplinski TJ, Ye C-Y, Li T, Sterck L, Vanneste K, Murat F, Soler M, Clemente HS, Saidi N, Cassan-Wang H, Dunand C, Hefer CA, Bornberg-Bauer E, Kersting AR, Vining K, Amarasinghe V, Ranik M, Naithani S, Elser J, Boyd AE, Liston A, Spatafora JW, Dharmwardhana P, Raja R, Sullivan C, Romanel E, Alves-Ferreira M, Külheim C, Foley W, Carocha V, Paiva J, Kudrna D, Brommonschenkel SH, Pasquali G, Byrne M, Rigault P, Tibbits J, Spokevicius A, Jones RC, Steane DA, Vaillancourt RE, Potts BM, Joubert F, Barry K, Jr. GJP, Strauss SH, Jaiswal P, Grima-Pettenati J, Salse J, Peer YVd, Rokhsar DS, Schmutz

J. 2014. The genome of *Eucalyptus grandis* - a global tree for fiber and energy. *Nature* **510**: 356-362.

Nakashima K, Tran L-SP, Nguyen DV, Fujita M, Maruyama K, Todaka D, Ito Y, Hayashi N, Shinozaki K, Yamaguchi-Shinozaki K. 2007. Functional analysis of a NAC-type transcription factor OsNAC6 involved in abiotic and biotic stress-responsive gene expression in rice. *The Plant Journal* **51**: 617-630.

Nuruzzaman M, Manimekalai R, Sharoni AM, Satoh K, Kondoh H, Ooka H, Kikuchi S. 2010. Genome-wide analysis of NAC transcription factor family in rice. *Gene* **465**: 30-44.

Ohtani M, Nishikubo N, Xu B, Yamaguchi M, Mitsuda N, Goue N, Shi F, Ohme-Takagi M, Demura T. 2011. A NAC domain protein family contributing to the regulation of wood formation in poplar. *The Plant Journal* **67**(3): 499-512.

Olsen AN, Ernst HA, Leggio LL, Skriver K. 2005. NAC transcription factors: structurally distinct, functionally diverse. *Trends in Plant Science* **10**(2): 79-87.

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, Kikuchi S. 2003. Comprehensive analysis of NAC family genes in *Oryza sativa* and *Arabidopsis thaliana*. *DNA Research* **10**(6): 239-247.

Pamilo P, Bianchi NO. 1993. Evolution of the *Zfx* and *Zfy*, genes: Rates and interdependence between the genes. *Molecular Biology and Evolution* **10**: 271-281.

Pinheiro GL, Marques CS, Costa MDBL, Reis PAB, Alves MS, Carvalho CM, Fietto LG, Fontes EPB. 2009. Complete inventory of soybean NAC transcription factors: Sequence conservation and expression analysis uncover their distinct roles in stress response. *Gene* **444**: 10-23.

Plomion C, Leprovost G, Stokes A. 2001. Wood formation in trees. *Plant Physiology* **127**: 1513-1523.

Punta M, Coggill PC, Eberhardt RY, Tate JMJ, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer ELL, Eddy SR, Bateman A, Finn RD. 2012. The Pfam protein families database. *Nucleic Acids Research* **40**(D1): D290-D301.

Puranik S, Sahu PP, Mandal SN, B. VS, Parida SK, Prasad M. 2013. Comprehensive genome-wide survey, genomic constitution and expression profiling of the NAC transcription factor family in foxtail millet (*Setaria italica* L.). *PLoS ONE* **8**(5): e64594.

Raes J, van de Peer Y. 2003. Gene duplication, the evolution of novel gene functions, and detecting functional divergence of duplicates in silica. *Applied Bioinformatics* **2**(2): 91-101.

Rastogi S, Liberles DA. 2005. Subfunctionalization of duplicated genes as a transition state to neofunctionalization. *BMC Evolutionary Biology* **5**: 28.

Rockwood DL, Rudie AW, Ralph SA, Zhu JY, Winandy JE. 2008. Energy product options for *Eucalyptus* species grown as short rotation woody crops. *International Journal of Molecular Sciences* **9**: 1361-1378.

Rushton PJ, Bokowiec MT, Han S, Zhang H, Brannock JF, Chen X, Laudeman TW, Timko MP. 2008. Tobacco transcription factors: novel insights into transcriptional regulation in the Solanaceae. *Plant Physiology* **147**: 280-295.

Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, Li J, Thiagarajan M, White JA, Quackenbush J. 2006. TM4 Microarray Software Suite. *Methods in Enzymology* **411**: 134-193.

Seo PJ, Kim S-G, Park C-M. 2008. Membrane-bound transcription factors in plants. *Trends in Plant Science* **13**(10): 550-556.

Shah ST, Pang C, Hussain A, Fan S, Song M, Zamir R, Yu S. 2014. Molecular cloning and functional analysis of NAC family genes associated with leaf senescence and stresses in *Gossypium hirsutum* L. *Plant Cell, Tissue and Organ Culture* **117**(2): 167-186.

Shang H, Li W, Zou C, Yuan Y. 2013. Analyses of the NAC transcription factor gene family in *Gossypium raimondii* Ulbr.: Chromosomal location, structure, phylogeny, and expression patterns. *Journal of Integrative Plant Biology* **55**(7): 663-676.

Shen H, Yin Y, Chen F, Xu Y, Dixon RA. 2009. A bioinformatic analysis of *NAC* genes for plant cell wall development in relation to lignocellulosic bioenergy production. *BioEnergy Research* **2**: 217-232.

Shiriga K, Sharma R, Kumar K, Yadav SK, Hossain F, Thirunavukkarasu N. 2014. Genome-wide identification and expression pattern of drought-responsive members of the NAC family in maize. *Meta Gene* **2**: 407-417.

Singh AK, Sharma V, Pal AK, Acharya V, Ahuja PS. 2013. Genome-wide organization and expression profiling of the NAC transcription factor family in potato (*Solanum tuberosum* L.). *DNA Research* **20**(4): 403-423.

Skolmen RG, Ledig TF 1990. *Eucalyptus globulus* Labill. Bluegum Eucalyptus. In: R. Burns B. Honkala eds. *Silvics of North America: 2. Hardwoods*. US Department of Agriculture, Forest Service. Available online at http://www.na.fs.fed.us/pubs/silvics_manual/volume_2/eucalyptus/globulus.htm [accessed 26 September 2014].

Slater GS, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**: 31.

Southerton SG, Marshall H, Mouradov A, Teasdale RD. 1998. Eucalypt MADS-box genes expressed in developing flowers. *Plant Physiology* **118**: 365-372.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**: 2731-2739.

Tibbits WN, White TL, Hodge GR, Borralho NMG. 2006. Genetic variation in frost resistance of *Eucalyptus globulus* ssp. *globulus* assessed by artificial freezing in winter. *Australian Journal of Botany* **54**: 521-529.

Toufighi K, Brady SM, Austin R, Ly E, Provart NJ. 2005. The Botany Array Resource: e-Northerns, Expression Angling, and promoter analyses. *The Plant Journal* **43**: 153-163.

Tran L-SP, Nakashima k, Sakuma Y, Simpson SD, Fujita Y, Maruyama K, Fujita M, Seki M, Shinozaki K, 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. *The Plant Cell* **16**: 2481-2498.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**(9): 1105-1111.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Baren MJv, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* **28**: 511-515.

Voorrips RE. 2002. MapChart: Software for the graphical presentation of linkage maps and QTLs. *The Journal of Heredity* **93**(1): 77-78.

Wang H-Z, Dixon RA. 2011. On–off switches for secondary cell wall biosynthesis. *Molecular Plant* **5**(2): 297-303.

Wang N, Zheng Y, Xin H, Fang L, Li S. 2013. Comprehensive analysis of NAC domain transcription factor gene family in *Vitis vinifera*. *Plant Cell Reports* **32**: 61-75.

Wilkins O, Nahal H, Foong J, Provart NJ, Campbell MM. 2009. Expansion and diversification of the *Populus* R2R3-MYB family of transcription factors. *Plant Physiology* **149**(2): 981-993.

Willemsen V, Bauch M, Bennett T, Campilho A, Wolkenfelt H, Xu J, Haseloff J, Scheres B. 2008. The NAC domain transcription factors FEZ and SOMBRERO control the orientation of cell division plane in *Arabidopsis* root stem cells. *Developmental Cell* **15**(6): 913-922.

Wu Y, Deng Z, Lai J, Zhang Y, Yang C, Yin B, Zhao Q, Zhang L, Li Y, Yang C, Xie Q. 2009. Dual function of *Arabidopsis ATAF1* in abiotic and biotic stress responses. *Cell Research* **19**: 1279-1290.

Xu B, Ohtani M, Yamaguchi M, Toyooka K, Wakazaki M, Sato M, Kubo M, Nakano Y, Sano R, Hiwatashi Y, Murata T, Kurata T, Yoneda A, Kato K, Hasebe M, Demura T. 2014. Contribution of NAC transcription factors to plant adaptation to land. *Science* **343**(6178): 1505-1508.

Yamaguchi M, Demura T. 2010. Transcriptional regulation of secondary wall formation controlled by NAC domain proteins. *Plant Biotechnology* **27**: 237-242.

Yoo SY, Kim Y, Kim SY, Lee JS, Ahn JH. 2007. Control of flowering time and cold response by a NAC-domain protein in *Arabidopsis*. *PLoS ONE* **2**(7): e642.

Zhang H, Jin JP, Tang L, Zhao Y, Gu XC, Gao G, Luo JC. 2011. PlantTFDB 2.0: update and improvement of the comprehensive plant transcription factor database. *Nucleic Acids Research* **39**: D1114-D1117.

Zhang J, S. K, Nei M. 1997. Small-sample tests of episodic adaptive evolution: a case study of primate lysozymes. *Molecular Biology and Evolution* **14**(12): 1335-1338.

Zhao C, Avci U, Grant EH, Haigler CH, Beers EP. 2008. XND1, a member of the NAC domain family in *Arabidopsis thaliana*, negatively regulates lignocellulose synthesis and programmed cell death in xylem. *The Plant Journal* **53**(3): 425-436.

Zhong R, Lee C, Ye Z-H. 2010a. Evolutionary conservation of the transcriptional network regulating secondary cell wall biosynthesis. *Trends in Plant Science* **15**(11): 625-632.

Zhong R, Lee C, Ye Z-H. 2010b. Functional characterization of Poplar wood-associated NAC domain transcription factors. *Plant Physiology* **152**: 1044–1055.

Zhong R, McCarthy RL, Lee C, Ye Z-H. 2011. Dissection of the transcriptional program regulating secondary wall biosynthesis during wood formation in poplar. *Plant Physiology* **157**: 1452–1468.

Zhu T, Nevo E, Sun D, Peng J. 2012. Phylogenetic analyses unravel the evolutionary history of NAC proteins in plants. *Evolution* **66**(6): 1833-1848.

Tables

Table 1. Distribution of amino acid motifs in NAC domain proteins of dicot and monocot genomes.

Colour key (%)[a] 100 90 80 70 60 50 40 30 20 10 0	*Eucalyptus* (190)[b]	*Populus* (192)	*Glycine* (183)	*Carica* (82)	*Arabidopsis* (135)	*Vitis* (142)	*Oryza* (186)	*Brachypodium* (100)	*Zea* (190)
General motifs									
Motif 1	175	162	171	62	116	129	135	66	163
Motif 2	156	119	150	52	97	114	94	44	129
Motif 3	136	119	134	44	90	115	95	43	126
Motif 4	168	129	149	53	105	128	111	48	139
Motif 5	173	127	163	54	107	124	118	53	157
Motif 6	181	148	176	73	118	133	152	67	170
Motif 7	63	70	83	31	61	63	76	32	87
Specific motifs									
Motif 8	15	4	7	2	1	4	5	2	2
Motif 9	10	12	22	4	7	8	15	9	20
Motif 10	15	4	6	2	1	5	0	0	0
Motif 11	21	5	5	2	3	5	2	3	3
Motif 12	12	0	0	0	0	0	0	0	0
Motif 13	14	2	2	0	2	0	0	0	0
Motif 14	16	1	2	0	1	2	0	0	0
Motif 15	13	0	0	0	0	0	0	0	0
Motif 16	7	0	0	0	0	0	0	0	0

[a]The percentage of NAC proteins in each genome containing a particular motif is represented by a heat map.

[b]The total number of NAC proteins in each genome, according to the PlantTFDB (Zhang *et al.*, 2011), is indicated in parenthesis, while the number of NAC domain proteins in each genome with a match to a given motif is indicated in each row.

Figure legends

Fig. 1. Maximum likelihood phylogeny of the *Eucalyptus* NAC family in relation to angiosperm lineages *Arabidopsis*, *Populus*, *Oryza* and *Vitis*. Trees were rooted at the midpoint. (a) Circular phylogram showing subfamilies described by Zhu *et al.* (2012); aLRT branch support values for each subfamily are indicated. (b) Linearized circular representation, normalized with respect to evolutionary distance. aLRT values >70 are indicated and the organism of origin of each respective taxon is indicated with a diamond symbol. A detailed dendrogram is available in Supporting Information Fig. S1.

Fig. 2. Predicted EgrNAC protein phylogeny, transcript abundance profiles, conserved amino acid motifs and gene structure. From left to right: (a) unrooted maximum likelihood phylogeny of EgrNAC proteins, showing subfamily classification and aLRT values greater than 50. (b) RNA-seq transcript abundance of *EgrNAC* genes in shoot tip (ST), young leaf (YL), mature leaf (ML), flowers (Fl), root (Rt), phloem (Ph) and developing xylem (DX) of three field-grown *E. grandis* trees. Values are expressed as the $\log_2$ value of average fragments per kilobase of exon per million fragments mapped (FPKM) per tissue. (c) Composition and distribution of overrepresented amino acid motifs (see Supporting Information Table S8). Grey bars indicate relative protein lengths. (d) Position of exons and introns in the *EgrNAC* gene models.

Fig. 3. Chromosomal locations of *EgrNAC* genes. Tandem duplicates are represented by shaded blocks; red shading indicates blocks with dN/dS < 1.0 ($P < 0.05$; codon-based Z-test). *P*-values for individual pairs of tandem duplicate pairs are available in Supporting Information Table S11.

Fig. 4. Tissue and organ expression data for 33 *EgrNAC* orthologs from *E. globulus*, herein denoted *EglNAC*. Fluidigm RT-qPCR data were hierarchically clustered using a quality threshold (QT) algorithm. PS, primary stem; SS, secondary stem; YX, young xylem; MX, mature xylem; YL, young leaf; ML, mature leaf; Rt, root; FB, flower bud; FC, flower capsule. Branch distances indicate Pearson's correlation.

Fig. 5. *EglNAC* genes showing a positive or negative transcriptional response towards cold treatment in primary stems (a), secondary stems (b), young leaves (c) and roots (d) in *E. globulus*. Error bars indicate the standard deviation of three biological replicates. A Bonferroni-adjusted *P*-value of 0.05 was applied to a two-tailed Student's *t* test for each of 33 *EglNAC* genes. Only genes exhibiting significant responses are shown.

Fig. 6. Transcriptional profile of *EglNAC* genes in xylem tissue from tension or opposite wood, relative to upright control in *E. globulus*. Error bars indicate standard deviation across three genotypes, using the average value of three ramets per genotype. *Significant difference relative to upright control according to two-tailed Student's *t* test, using a Bonferroni-corrected *P*-value ($P^* = 0.05/33$).

Supporting information

Fig. S1.pdf: Dendrogram and subfamily classification of NAC sequences from *Arabidopsis*, *Populus*, *Oryza*, *Vitis* and *Eucalyptus*. The dendrogram is based on the phylogeny shown in Fig. 1b.

Fig. S2. *EgrNAC* transcript abundance in seven *E. grandis* tissues and organs.

Fig. S3. Distribution of EgrNAC protein motifs across nine angiosperm genomes.

Fig. S4. Intron/exon structure of *EgrNAC* genes showing intron phase.

Fig. S5. Comparison of tissue expression patterns of *EgrNAC* paralogs with those of their closest *Arabidopsis*, *Oryza* and/or *Populus* homologs.

Fig. S6. Correlation of xylem/leaf expression ratio calculated with *E. grandis* RNA-seq data and *E. globulus* Fluidigm data for 33 NAC domain genes.

Table S1. Primers pairs used for Fluidigm RT-qPCR analysis in *E. globulus.*

Table S2. Classification of NAC domain genes in the draft *E. grandis* genome assembly (v.1.0, www.phytozome.net).

Table S3.xls: Lists of *Arabidopsis*, *Populus*, *Oryza* and *Vitis* NAC domain proteins used for phylogenetic analysis.

Table S4. Nomenclature, length and coordinates of 189 NAC domain proteins identified in the draft *Eucalyptus grandis* genome assembly (v.1.0, www.phytozome.net).

Table S5. Biological functions of functionally characterized NAC domain proteins occurring in the subfamilies annotated in Fig. 1 of the main manuscript.

Table S6. Putative *E. grandis* homologs of *Arabidopsis* NAC domain proteins known to be involved in regulating secondary cell wall biosynthesis.

Table S7.xls: *EgrNAC* RNA-seq data for developing xylem (DX), flowers (Fl), mature leaf (ML), phloem (Ph), roots (Rt), shoot tips (ST) and young leaves (YL) in three individual ramets. Values are expressed as average number of fragments per kilobase of transcript per million fragments mapped (FPKM).

Table S8. Amino acid sequence logos of sixteen overrepresented motifs identified in EgrNAC proteins using MEME. The E-value, number of proteins containing each motif and, where applicable, the annotation of each motif is indicated.

Table S9. Putative EgrNAC membrane-tethered transcription factors (MTFs) and their corresponding *Arabidopsis* NAC MTF homologs as deduced from Supporting Information Fig. S1.

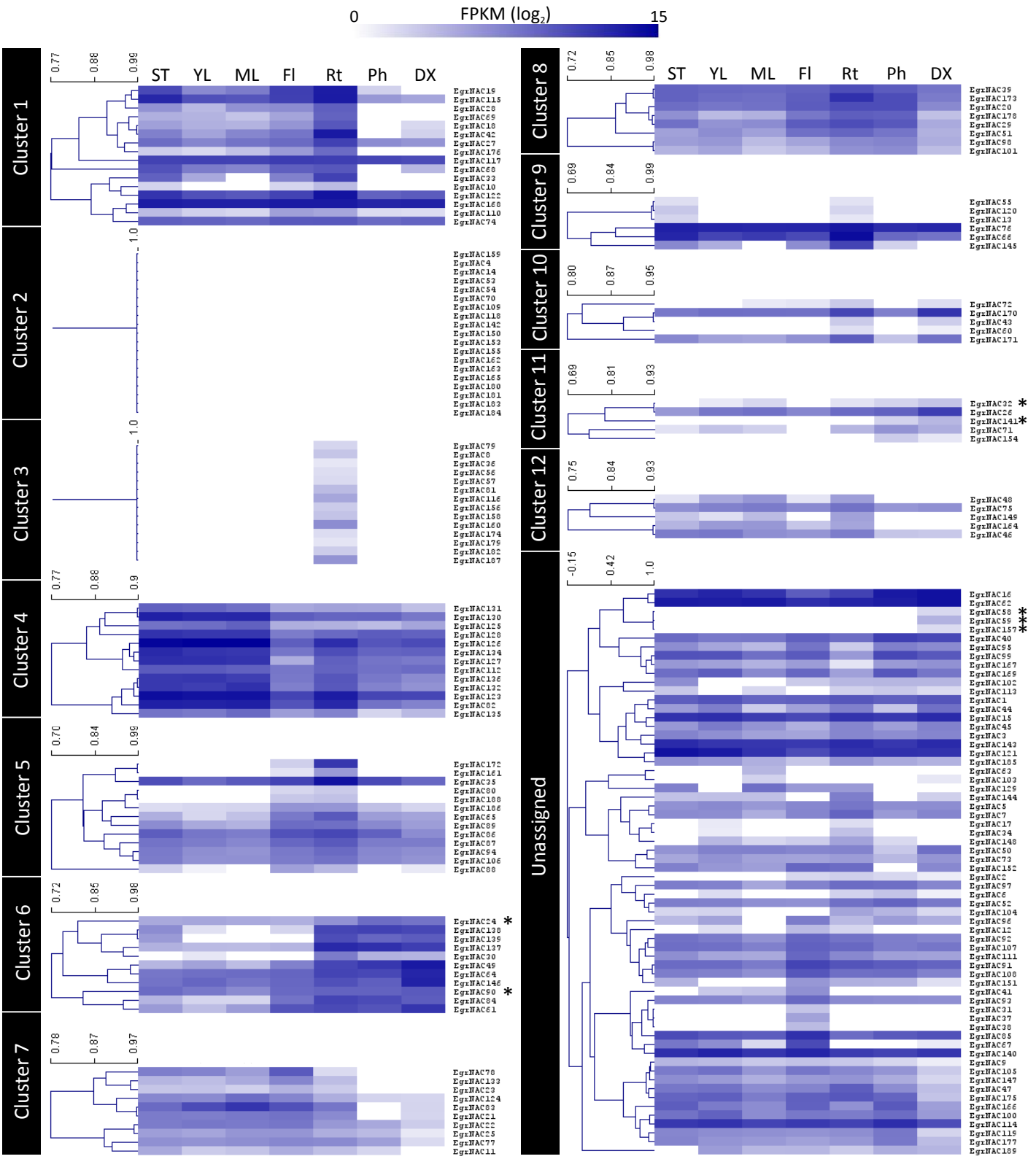
Table S10. *EgrNAC* genes occurring in blocks of tandem duplications (Fig. 3 of main manuscript).

Table S11.xls: Codon-based Z-test for purifying selection between pairwise comparisons of *EgrNAC* genes in each of twenty-three blocks of tandem duplicates. The P-value for each comparison is shown below the diagonal (*P*-values < 0.05 are indicated in bold); the Z-test statistic is shown above the diagonal. Non-expressed genes are indicated in red.

Note S1: Notes of manually curated and discarded EgrNAC gene candidates. Low confidence annotations refer to those included in the v.1.0 annotation (Phytozome v.7) but excluded from the v.1.1 annotation (Phytozome v.8) of the *E. grandis* genome at www.phytozome.net. Evidence for expression was obtained from Eucspresso (eucspresso.bi.up.ac.za/; Mizrachi *et al.* 2010) and EucGenIE (http://eucgenie.bi.up.ac.za; Hefer *et al.*, 2011).

Note S2: Differences in phylogenetic clustering of NAC domain proteins (Fig. 1, main manuscript) compared to that of Zhu *et al.* (2012)

Figures S2-S6, Tables S1, S2, S4-S6 & S8-S10 and Notes S1 & S2



(previous page)

Fig. S2. Quality threshold (QT) clustering of the 189 *EgrNAC* gene transcripts according to their expression profiles across shoot tips (ST), young leaves (YL), mature leaves (ML), flowers (Fl), roots (Rt) phloem (Ph) and developing (secondary) xylem (DX) in *Eucalyptus grandis*. Normalized RNA-seq transcript abundance data is expressed as the $\log_2$ value of fragments per kilobase of exon per million fragments mapped (FPKM). Branch lengths represent Pearson correlation coefficient, as indicated on the scale for each cluster. Asterisks indicate novel candidates potentially involved in the regulation of xylogenesis.

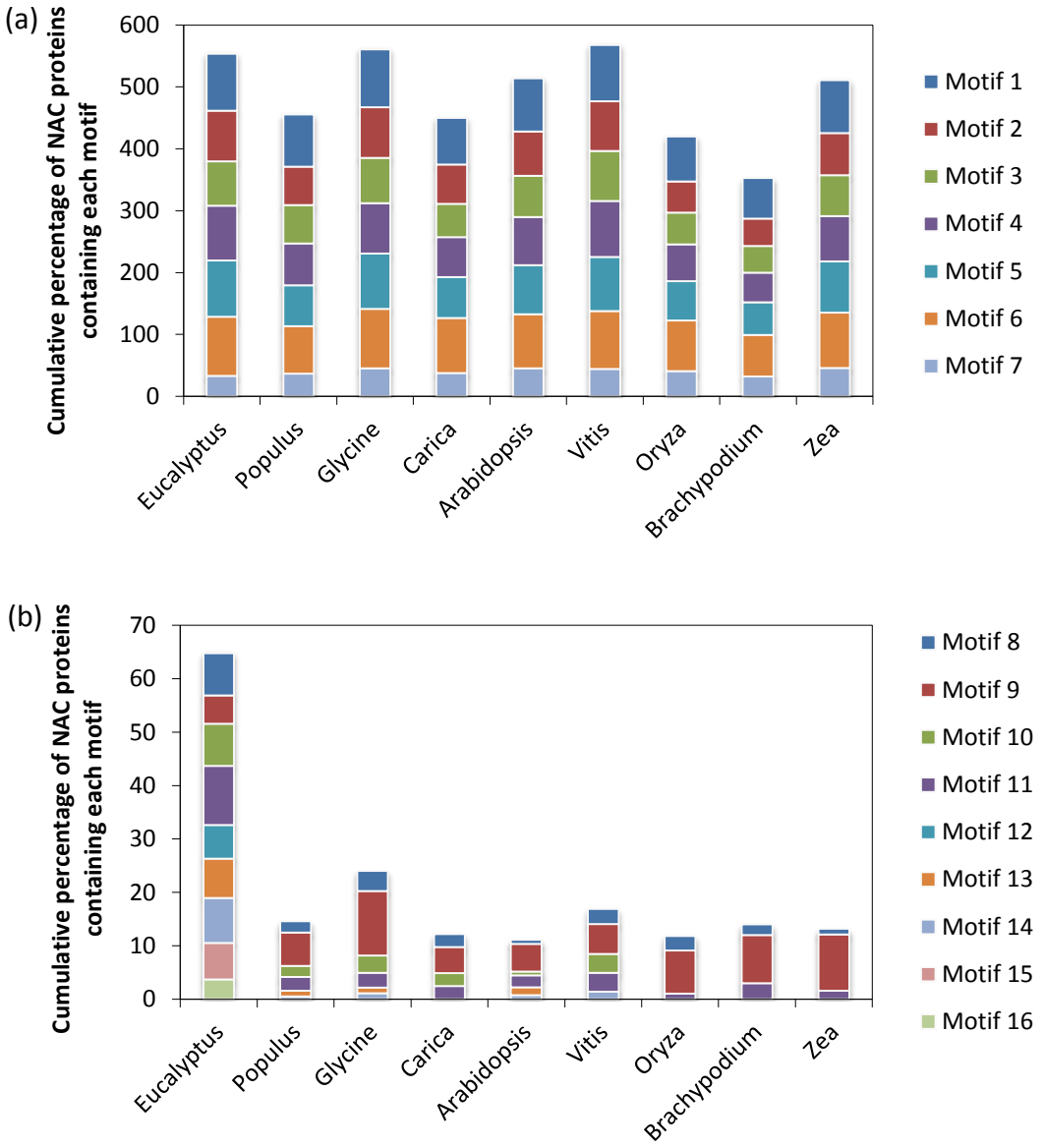


Fig. S3. Distribution of conserved protein motifs in eight plant genomes as identified using Hidden Markov Models built on *Eucalyptus* motif alignments. Motif frequencies are expressed as a percentage of NAC proteins containing a given motif out of the total NAC proteins in each genome. **(a)** General motifs, corresponding to subdomains A through E of the NAC domain, are distributed evenly across the genomes, showing that the Hidden Markov Models are not biased. **(b)** Specific motifs, showing enrichment in, or exclusive occurrence in, *Eucalyptus* NAC proteins with the exception of Motif 9.

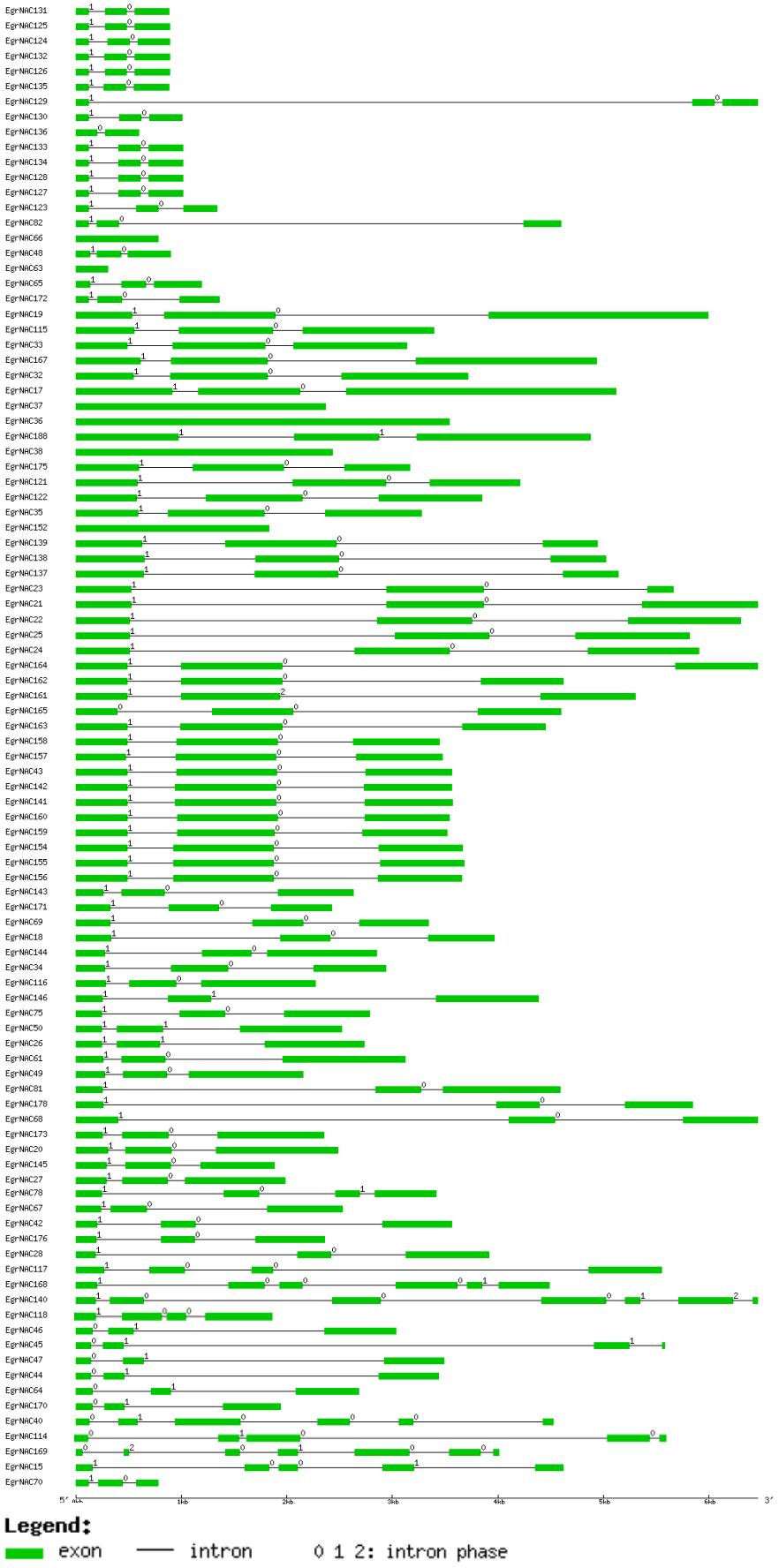


Fig. S4. Intron and exon structure of *EgrNAC* proteins, showing intron phase. The *EgrNAC* genes occur in the same order as those in Fig. 2 of the main manuscript.

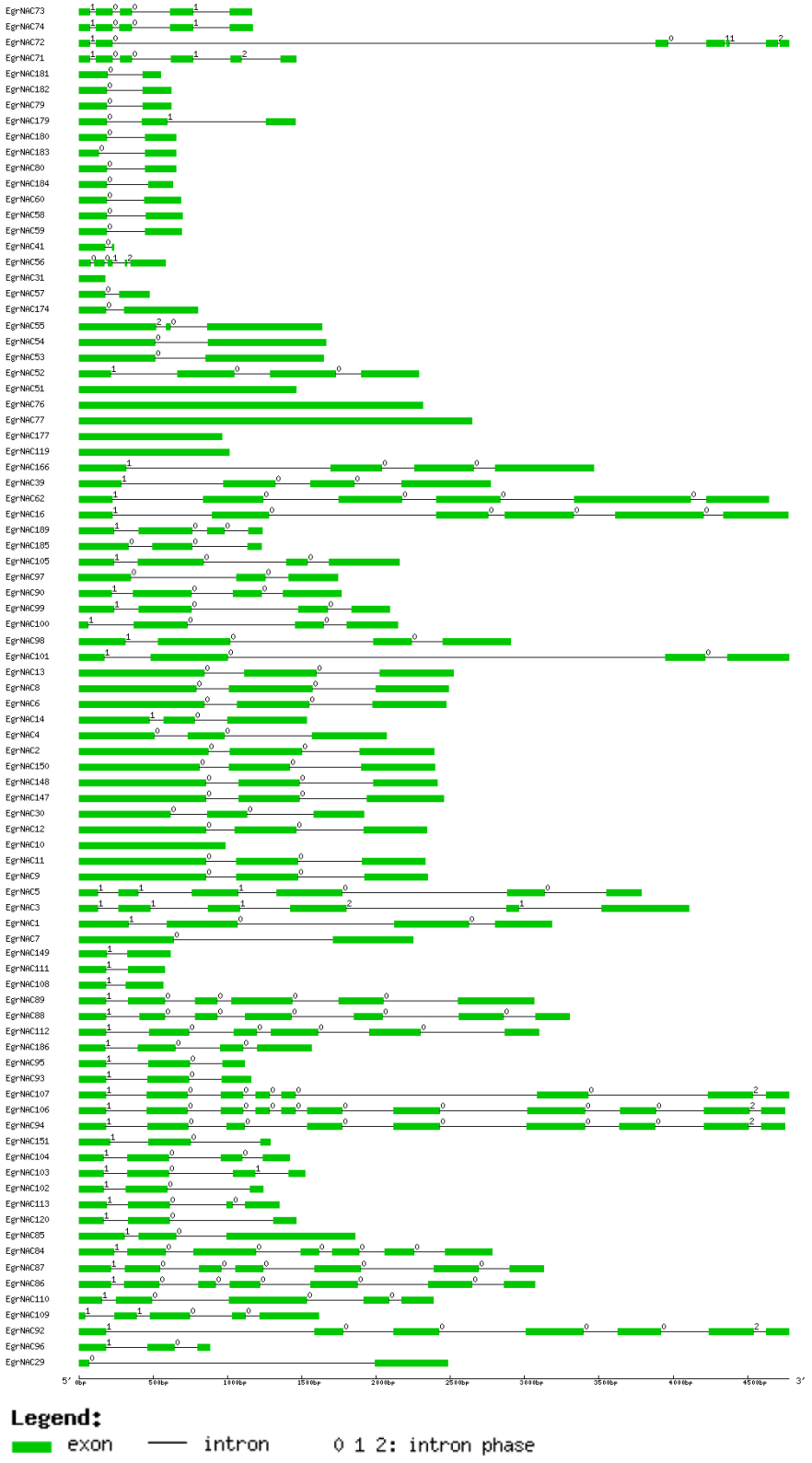


Fig. S4 (*continued from previous page*).

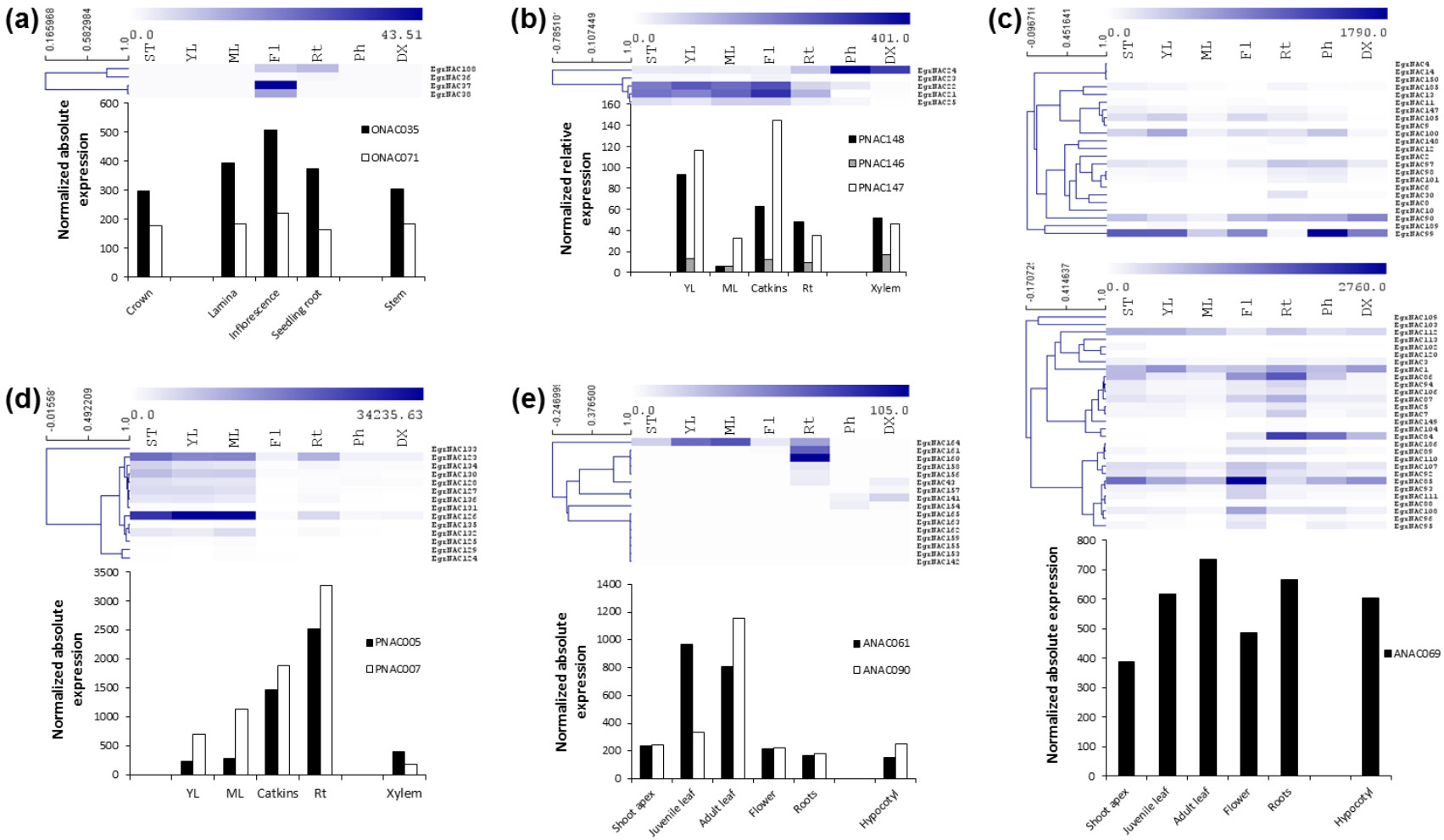


Fig. S5. Comparison of tissue expression patterns of *EgrNAC* paralogs with those of their closest *Arabidopsis*, *Oryza* and/or *Populus* homologs. Absolute transcript levels from the EucGenIE database (http://eucgenie.bi.up.ac.za/) (FPKM; blue heatmap) are shown for groups of *EgrNAC* paralogous genes, arbitrarily labelled A through O. Group C contains two blocks of paralogous genes which are phylogenetically distinct but share the same *Arabidopsis* homolog (see Supporting Information Fig. S1). Expression patterns were hierarchically clustered according to Pearson's correlation. Expression patterns of closest homologs inferred from Supporting

Information Fig. S1 are shown in black and white for corresponding tissues, where available, of each *EgrNAC* paralog panel. These data were obtained from Genevestigator (Hruz *et al.*, 2008) (*Arabidopsis* and *Oryza*) or the Poplar Expression Angler developmental series (Toufighi *et al.*, 2005; Wilkins *et al.*, 2009). ST, shoot tip; YL, young leaf; ML, mature leaf; Fl, flower; Rt, root; DX, developing xylem.

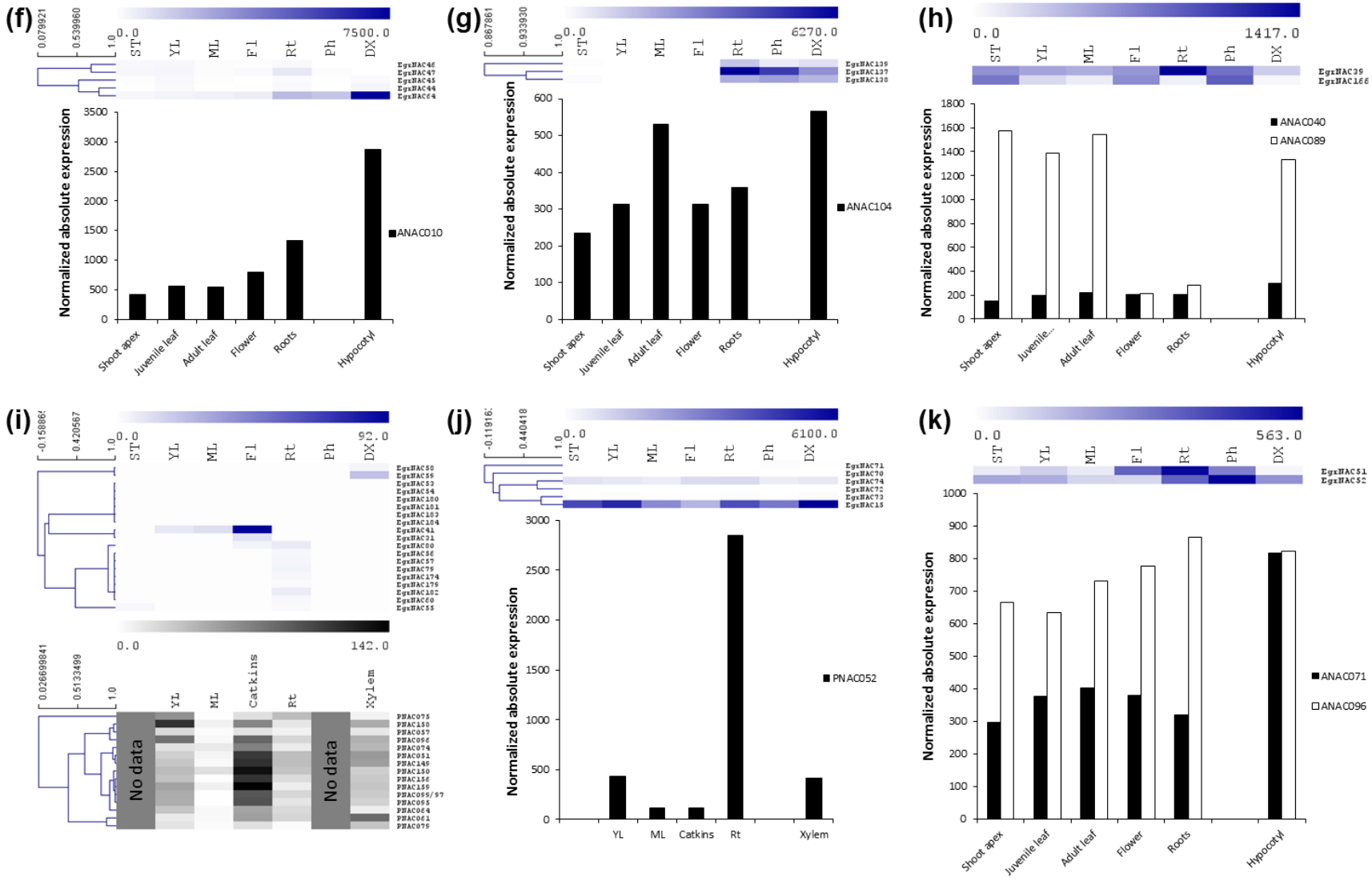


Fig. S5. *(Continued from previous page).*

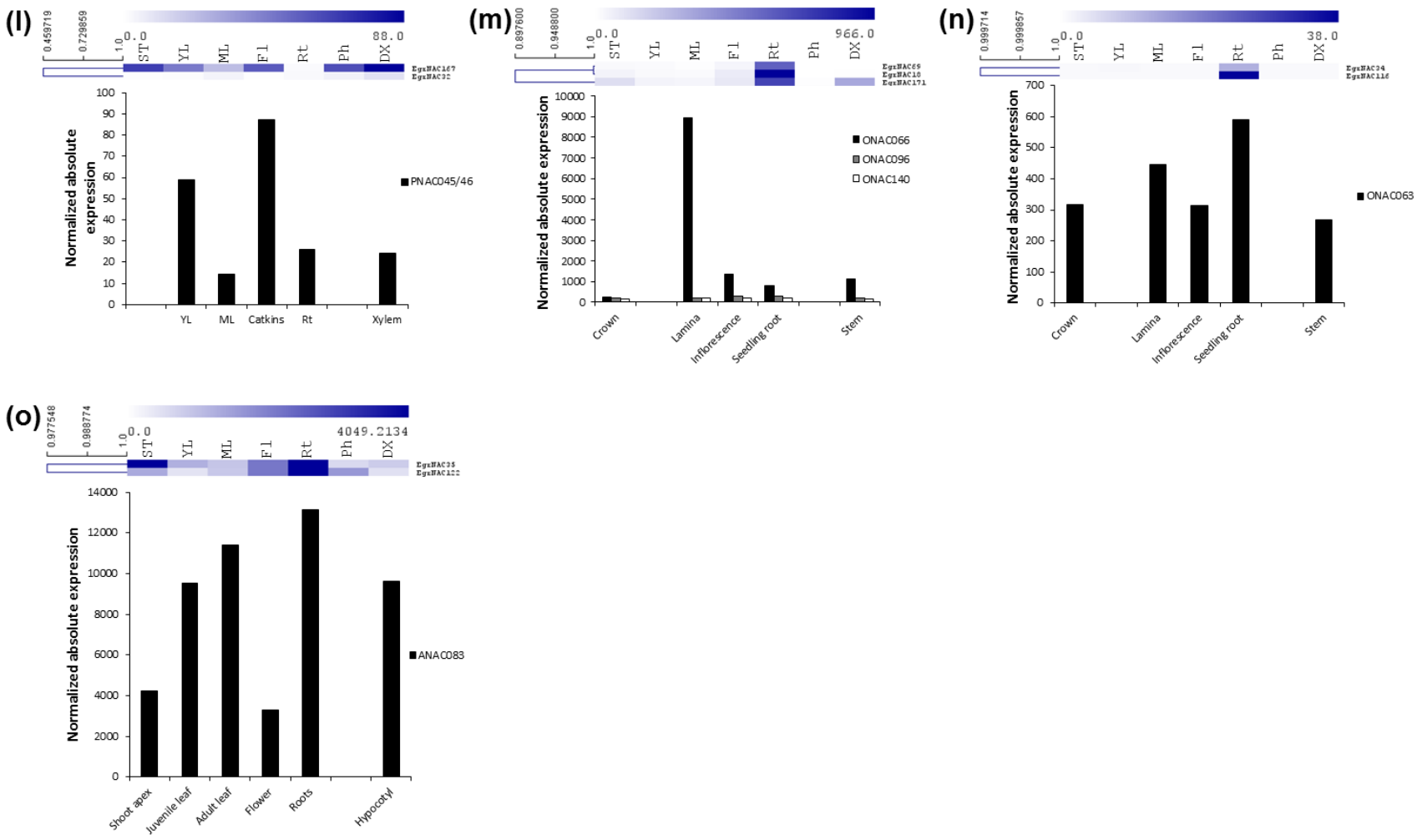


Fig. S5. *(Continued from previous page).*

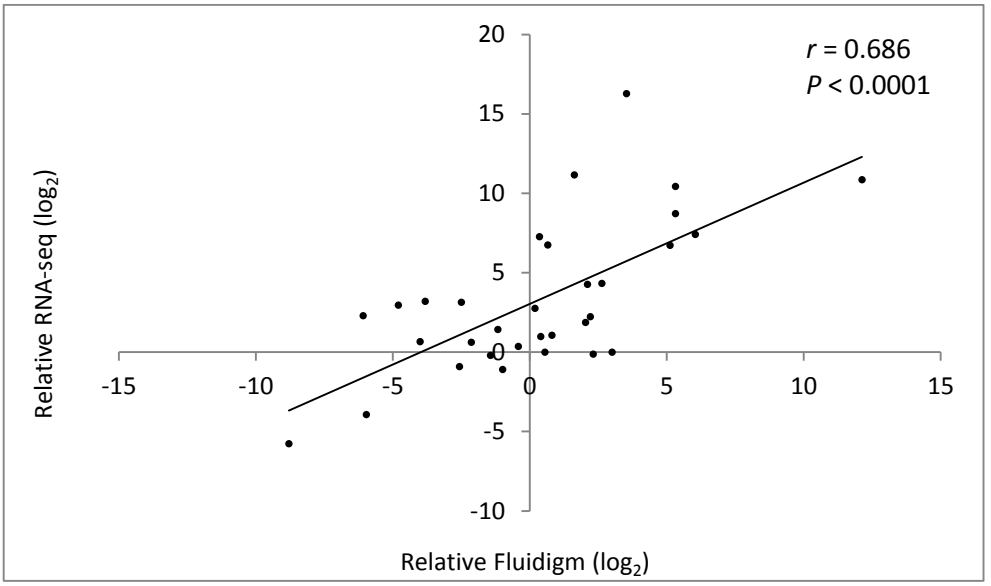


Fig. S6. Correlation of xylem/leaf expression ratio calculated with *E. grandis* RNA-seq data and *E. globulus* Fluidigm data for 33 NAC domain genes. *P*-value represents the two-tailed probability for Pearson's correlation (*r*).

Table S1. Primers pairs used for Fluidigm RT-qPCR analysis in *E. globulus*

Gene Name	Gene Symbol	Forward primer sequence [5'-3']	Reverse primer sequence [5'-3']	Amplicon length (bp)
Eucgr.A00357	EglNAC1	TTCCTCAAGTGCTGCAACTGTC	CCTGGCTAGGAAGTTGTTTGACTG	84
Eucgr.A00363	EglNAC6	GATCCACCAAACGGTCAAAC	GCCAGGTAAATCCCAAGATG	142
Eucgr.A00494	EglNAC15	GGGAAACAACAACCTGTCAACTGC	AACAGCCTGGTGGTTGCTTGTG	88
Eucgr.A00969	EglNAC16	TCAGGAAGCGACTGAAAGACAGAG	TGGAAGTTTCCTCGAGCCATCC	105
Eucgr.A02638	EglNAC24	AAGAGATCCCAGTGGTGCCAAG	TGATTTCTCCCATCGGTCTTCGG	102
Eucgr.A02887	EglNAC26	TGAGAACGGAACTGGTCAGGAAG	ATCCATGCTCGGTCATCTTGCG	92
Eucgr.B01624	EglNAC31	TCATGTTCGGTGACAAGTTCGG	GCCGCTTGAGTACTTGTGCTTC	74
Eucgr.C00958	EglNAC40	AGATTTGCGGATCCAAACAGTGC	TGCTGGTAGACCTAACCAATCCG	77
Eucgr.D00591	EglNAC44	CCAGAGAGACTTCCAGGAGTAAGC	TCTTGTGCCACCTTGTCTCAGTC	142
Eucgr.D00595	EglNAC47	TGATGTGGCCAAAGCAAGATGC	CCCTCTGATCGAACATTGGGAACC	144
Eucgr.D01671	EglNAC49	GAGCCATGGGATATCCAAGAGAGG	TTGTGGCTAAAGAAGTACCAGTCG	74
Eucgr.D02027	EglNAC50	TCAGGAGGAAGGATGGGTTGTG	AGGGCTGAGAAATCCTCCTTGG	149
Eucgr.E00574	EglNAC59	AGCATCCTCGCAACGAAACG	AGTGTCTGTGCCTTCTTTGCTC	148
Eucgr.E00575	EglNAC60	GGCGAACATGCAAGAAGATACCTG	TTCTTCAACCGATCCGCCCATTC	129
Eucgr.E01053	EglNAC61	TCGACTTGGACGTGATTCGTGAG	CCAATCCTGCACTTCTCTTGGATG	76
Eucgr.E03226	EglNAC64	AGAGGGAGAGAATGGGATCTGC	CCCATCTTTGCTCACTCCTGGTAG	64
Eucgr.F01091	EglNAC65	ATTCCCTGAGGCTGGATGACTG	TTCATGCCGTGAGGGTTCATCG	148
Eucgr.F02615	EglNAC75	AGAAACGACCACATCCCACTCC	AGGCGGTACACACGATTCCAAC	60
Eucgr.F04341	EglNAC82	AAGAACAGCTTGAGGCTTGACGAC	TTCTTCTCGATCGCGCCCTTCTTG	71
Eucgr.G01047	EglNAC84	GCATCCTGATGATACGGGCTTC	ACAGCTGCATAGTTATCTGCCTTC	78
Eucgr.G01061	EglNAC90	CGAGTACTTTGGCCAATTCAAGC	TTTCTTCCTCAGATGCCTGTGC	99
Eucgr.G01063	EglNAC91	GCCAGATGGCCTTTGTTCTCTGTC	TTCGTCGGATCGTCCATTTCCG	134
Eucgr.G01066	EglNAC93	TCAATGAACTGTTTCCACGTGCTC	CCTTGGTGCTTTAAGTGACAGACG	64
Eucgr.I00191	EglNAC137	GCTCCAACAGGTCAAGAGACAAAC	CCGGGTTCTTGTGTTGAATTAGCC	89
Eucgr.I00192	EglNAC138	CTACCACCTGGATTTCGGTTCTC	GGATGCAGAAAGTGAAGGACGAG	62
Eucgr.I00193	EglNAC139	TGTCTTCGCTCTATGCTCACTTGG	TCCATTCCACAGTGCCTTTCCG	89
Eucgr.I00583	EglNAC141	TGAACTCTCGCCGACCAATCTC	ATGCGAATTCACGCCTTAGCTC	74
Eucgr.I00587	EglNAC142	TGAGAACGGAACTGGTCAGGAAG	ATCCATGCTCGGTCATCTTGCG	92
Eucgr.I01494	EglNAC143	TGACTCGTCGCCCAAGGAAATG	TGGCGGCCTTATTCATGCCTTC	111
Eucgr.I02366	EglNAC146	ATTGCACCGAGTCTGCAAGC	TACACACGACTCCATCGTCTCC	66
Eucgr.I02695	EglNAC152	TATGATCCGTGGGAGCTTGAAGGG	ATAGGCTTCCAGTACCCGTTGC	110
Eucgr.J01038	EglNAC168	AAGGCTGGAATTCCGCAAGATG	GTTCTTTGGGCCAGAACCACTC	72
Eucgr.K01228	EglNAC171	TCCCTGGGATCTCCATGATGTTAG	CCGGATCCAGTCACTCTATTTGGC	114
Reference gene primers				
Eucgr.B02473	EF-1α	ATGCGTCAGACTGTGGCTGTTG	ATGCGTCAGACTGTGGCTGTTG	74
Eucgr.F02901	IDH	AATCGACCTGCTTCGACCCTTC	TCGACCTTGATCTTCTCGAAACCC	68
Eucgr.B03386	PP2A1	TCGAGCTTTGGACCGCATACAAG	ACCACAAGAGGTCACACATTGGC	62
Eucgr.B03031	PP2A3	CAGCGGCAAACAACTTGAAGCG	ATTATGTGCTGCATTGCCCAGTC	67
Eucgr.B02502	SAND	TTGATCCACTTGCGGACAAGGC	TCACCCATTGACATACACGATTGC	63
gDNA contamination assessment				
Intergenic	3' of Eucgr.H02589	GCGGCTTTTAAGTCTCTTGCGAA	TTCGAAGCATAGCTTCGCCATATG	150

Table S2. Classification of NAC domain genes identified in the v.1.0 *E. grandis* genome assembly (www.phytozome.net)

Manually curated *EgrNAC* gene models					
Eucgr.A00357.1	Eucgr.B03208.1	Eucgr.E03226.1	Eucgr.G01070.1	Eucgr.I00059.4	Eucgr.J00517.1
Eucgr.A00359.1	Eucgr.B03439.1	Eucgr.F01091.1	Eucgr.G01071.1	Eucgr.I00060.1	Eucgr.J00518.1
Eucgr.A00360.1	Eucgr.B03537.1	Eucgr.F01093.1	Eucgr.G01074.1	Eucgr.I00060.2	Eucgr.J00519.1
Eucgr.A00361.1	Eucgr.B03693.1	Eucgr.F01170.1	Eucgr.G01075.1	Eucgr.I00095.1	Eucgr.J00520.1
Eucgr.A00362.1	Eucgr.B03703.1	Eucgr.F01449.1	Eucgr.G01077.1	Eucgr.I00099.1	Eucgr.J00521.1
Eucgr.A00363.1	Eucgr.B03704.1	Eucgr.F01463.1	Eucgr.G01078.1	Eucgr.I00100.1	Eucgr.J00531.1
Eucgr.A00364.1	Eucgr.B03823.1	Eucgr.F01535.1	Eucgr.G01081.1	Eucgr.I00101.1	Eucgr.J00940.1
Eucgr.A00365.1	Eucgr.C00958.1	Eucgr.F01536.1	Eucgr.G01082.1	Eucgr.I00102.1	Eucgr.J01038.1
Eucgr.A00368.1	Eucgr.C01264.1	Eucgr.F01537.1	Eucgr.G01083.1	Eucgr.I00191.1	Eucgr.J02254.1
Eucgr.A00369.1	Eucgr.C02105.1	Eucgr.F01538.1	Eucgr.G01507.1	Eucgr.I00192.1	Eucgr.K01061.1
Eucgr.A00370.1	Eucgr.C02446.1	Eucgr.F01539.1	Eucgr.G01548.1	Eucgr.I00193.1	Eucgr.K01228.1
Eucgr.A00371.1	Eucgr.D00591.1	Eucgr.F02615.1	Eucgr.G01550.1	Eucgr.I00213.1	Eucgr.K01471.1
Eucgr.A00435.1	Eucgr.D00592.1	Eucgr.F02771.1	Eucgr.G01551.1	Eucgr.I00583.1	Eucgr.K01472.1
Eucgr.A00437.1	Eucgr.D00593.1	Eucgr.F02910.1	Eucgr.G01553.1	Eucgr.I00587.1	Eucgr.K01845.1
Eucgr.A00494.1	Eucgr.D00594.1	Eucgr.F03588.1	Eucgr.G01554.1	Eucgr.I01494.1	Eucgr.K02205.1
Eucgr.A00969.1	Eucgr.D00595.1	Eucgr.F03962.1	Eucgr.G01555.1	Eucgr.I01940.1	Eucgr.K02225.1
Eucgr.A01272.1	Eucgr.D00665.1	Eucgr.F03963.1	Eucgr.G01758.1	Eucgr.I01958.1	Eucgr.K02303.1
Eucgr.A02028.1	Eucgr.D01671.1	Eucgr.F04097.1	Eucgr.G01984.1	Eucgr.I02366.1	Eucgr.K03256.1
Eucgr.A02070.1	Eucgr.D02027.1	Eucgr.F04341.1	Eucgr.G02349.1	Eucgr.I02571.1	Eucgr.K03356.1
Eucgr.A02074.1	Eucgr.D02182.1	Eucgr.G00054.1	Eucgr.G02486.1	Eucgr.I02573.1	Eucgr.K03357.1
Eucgr.A02635.1	Eucgr.E00298.1	Eucgr.G01047.1	Eucgr.G02506.1	Eucgr.I02574.1	Eucgr.K03358.1
Eucgr.A02636.1	Eucgr.E00541.1	Eucgr.G01049.1	Eucgr.G02740.1	Eucgr.I02576.1	Eucgr.K03359.1
Eucgr.A02637.1	Eucgr.E00542.1	Eucgr.G01052.1	Eucgr.G02742.1	Eucgr.I02578.1	Eucgr.K03360.1
Eucgr.A02638.1	Eucgr.E00543.1	Eucgr.G01053.1	Eucgr.H00614.1	Eucgr.I02695.1	Eucgr.K03361.1
Eucgr.A02639.1	Eucgr.E00545.1	Eucgr.G01058.1	Eucgr.H00826.1	Eucgr.J00505.1	Eucgr.L00819.1
Eucgr.A02887.1	Eucgr.E00551.1	Eucgr.G01060.1	Eucgr.H03362.1	Eucgr.J00508.1	Eucgr.L01867.1
Eucgr.B00529.1	Eucgr.E00573.1	Eucgr.G01061.1	Eucgr.H03387.1	Eucgr.J00509.1	Eucgr.L02267.1
Eucgr.B00724.1	Eucgr.E00574.1	Eucgr.G01063.1	Eucgr.H05089.1	Eucgr.J00511.1	Eucgr.L02674.1

Table S2. *(continued from previous page).*

Manually curated *EgrNAC* gene models *(continued)*					
Eucgr.B01567.1	Eucgr.E00575.1	Eucgr.G01064.1	Eucgr.I00056.1	Eucgr.J00512.1	Eucgr.L03347.1
Eucgr.B01593.1	Eucgr.E01053.1	Eucgr.G01066.1	Eucgr.I00057.1	Eucgr.J00513.1	
Eucgr.B01624.1	Eucgr.E01095.1	Eucgr.G01067.1	Eucgr.I00058.1	Eucgr.J00514.1	
Eucgr.B02485.1	Eucgr.E03225.1	Eucgr.G01069.1	Eucgr.I00059.1	Eucgr.J00516.1	
Alternative splice variants					
Eucgr.A00357.2	Eucgr.D00593.2	Eucgr.F02771.5	Eucgr.G01548.2	Eucgr.I00213.2	Eucgr.L01924.2
Eucgr.A00494.2	Eucgr.E01095.2	Eucgr.F04341.2	Eucgr.G02486.2	Eucgr.I00213.3	Eucgr.L02268.2
Eucgr.A02028.2	Eucgr.E03226.2	Eucgr.G01047.2	Eucgr.G02740.2	Eucgr.I00213.4	
Eucgr.A02887.2	Eucgr.F01463.2	Eucgr.G01067.2	Eucgr.I00059.2	Eucgr.I00213.5	
Eucgr.B03537.2	Eucgr.F02771.2	Eucgr.G01067.3	Eucgr.I00059.3	Eucgr.I00213.6	
Eucgr.C00958.2	Eucgr.F02771.3	Eucgr.G01067.4	Eucgr.I00100.2	Eucgr.I02366.2	
Eucgr.C00958.3	Eucgr.F02771.4	Eucgr.G01067.5	Eucgr.I00191.2	Eucgr.K01228.2	
Putative alleles					
Eucgr.L01840.1	Eucgr.L02201.1	Eucgr.L02268.1	Eucgr.L02673.1	Eucgr.L02696.1	Eucgr.L03434.1
Eucgr.L01924.1	Eucgr.L02202.1	Eucgr.L02499.1	Eucgr.L02683.1	Eucgr.L02867.1	
Eucgr.L01925.1	Eucgr.L02266.1	Eucgr.L02501.1	Eucgr.L02695.1	Eucgr.L03094.1	
Failed manual curation					
Eucgr.A01274.1	Eucgr.G01265.1	Eucgr.H03391.1	Eucgr.L02177.1		
Eucgr.A01885.1	Eucgr.G01267.1	Eucgr.I02577.1			
Eucgr.G01073.1	Eucgr.G01448.1	Eucgr.J01735.1			
Collapsed into a single gene model					
Eucgr.I00097.1	Eucgr.I00098.1				

Table S4. Nomenclature, lengths and coordinates of 189 NAC domain proteins identified in the draft *E. grandis* genome assembly (V1.0, www.phytozome.net).

Name	Gene	Protein length[a]	Locus[b]	Strand
EgrNAC1	Eucgr.A00357	308	scaffold_1:4972108-4973849	-
EgrNAC2	Eucgr.A00359	338	scaffold_1:4977347-4978655	-
EgrNAC3	Eucgr.A00360	292	scaffold_1:5011350-5013598	-
EgrNAC4	Eucgr.A00361	228	scaffold_1:5067016-5068149	-
EgrNAC5	Eucgr.A00362	273	scaffold_1:5073131-5075204	-
EgrNAC6	Eucgr.A00363	333	scaffold_1:5078319-5079673	-
EgrNAC7	Eucgr.A00364	213	scaffold_1:5102051-5103282	-
EgrNAC8	Eucgr.A00365	335	scaffold_1:5107090-5108452	-
EgrNAC9	Eucgr.A00368	308	scaffold_1:5180298-5181584	-
EgrNAC10	Eucgr.A00369	179	scaffold_1:5197166-5197706	-
EgrNAC11	Eucgr.A00370	308	scaffold_1:5211733-5213010	-
EgrNAC12	Eucgr.A00371	308	scaffold_1:5228116-5229398	-
EgrNAC13	Eucgr.A00435	333	scaffold_1:6414247-6415628	-
EgrNAC14	Eucgr.A00437	221	scaffold_1:6448797-6449635	-
EgrNAC15	Eucgr.A00494	374	scaffold_1:7719332-7723951	-
EgrNAC16	Eucgr.A00969	597	scaffold_1:15385239-15388740	-
EgrNAC17	Eucgr.A01272	466	scaffold_1:20607436-20609052	-
EgrNAC18	Eucgr.A02028	312	scaffold_1:31019047-31021637	-
EgrNAC19	Eucgr.A02070	385	scaffold_1:31501255-31503148	+
EgrNAC20	Eucgr.A02074	410	scaffold_1:31544167-31545788	+
EgrNAC21	Eucgr.A02635	266	scaffold_1:36895028-36897068	-
EgrNAC22	Eucgr.A02636	266	scaffold_1:36937941-36939979	-
EgrNAC23	Eucgr.A02637	177	scaffold_1:36965867-36967655	-
EgrNAC24	Eucgr.A02638	264	scaffold_1:36976003-36977913	-
EgrNAC25	Eucgr.A02639	268	scaffold_1:37001186-37003068	-
EgrNAC26	Eucgr.A02887	348	scaffold_1:39264076-39265859	+
EgrNAC27	Eucgr.B00529	361	scaffold_2:6904675-6905972	+
EgrNAC28	Eucgr.B00724	357	scaffold_2:9040994-9044251	+
EgrNAC29	Eucgr.B01567	184	scaffold_2:26103161-26105643	+
EgrNAC30	Eucgr.B01593	221	scaffold_2:26870281-26871330	-
EgrNAC31	Eucgr.B01624	127	scaffold_2:27726776-27727160	+
EgrNAC32	Eucgr.B02485	279	scaffold_2:47008154-47009327	-
EgrNAC33	Eucgr.B03208	255	scaffold_2:57054023-57055014	+
EgrNAC34	Eucgr.B03439	326	scaffold_2:59177735-59179654	+
EgrNAC35	Eucgr.B03537	253	scaffold_2:60000406-60001441	+
EgrNAC36	Eucgr.B03693	372	scaffold_2:61396803-61398876	-
EgrNAC37	Eucgr.B03703	248	scaffold_2:61459361-61460515	-
EgrNAC38	Eucgr.B03704	255	scaffold_2:61468791-61470297	-
EgrNAC39	Eucgr.B03823	371	scaffold_2:62322277-62324308	-
EgrNAC40	Eucgr.C00958	486	scaffold_3:14868676-14873196	+
EgrNAC41	Eucgr.C01264	135	scaffold_3:19938845-19939356	-
EgrNAC42	Eucgr.C02105	326	scaffold_3:38063370-38066334	+
EgrNAC43	Eucgr.C02446	242	scaffold_3:46604381-46605534	+
EgrNAC44	Eucgr.D00591	300	scaffold_4:10891985-10895422	-
EgrNAC45	Eucgr.D00592	229	scaffold_4:10923473-10929053	-
EgrNAC46	Eucgr.D00594	295	scaffold_4:10995874-10998399	-
EgrNAC47	Eucgr.D00595	300	scaffold_4:11010756-11014245	-
EgrNAC48	Eucgr.D00665	343	scaffold_4:12112914-12114131	+
EgrNAC49	Eucgr.D01671	383	scaffold_4:30694508-30695914	+
EgrNAC50	Eucgr.D02027	355	scaffold_4:34311770-34313414	+
EgrNAC51	Eucgr.D02182	357	scaffold_4:36060390-36062440	+
EgrNAC52	Eucgr.E00298	346	scaffold_5:2804334-2806009	-
EgrNAC53	Eucgr.E00541	318	scaffold_5:5139951-5141158	-
EgrNAC54	Eucgr.E00542	318	scaffold_5:5162361-5163579	-
EgrNAC55	Eucgr.E00543	322	scaffold_5:5171768-5172969	-
EgrNAC56	Eucgr.E00545	297	scaffold_5:5184896-5186151	-
EgrNAC57	Eucgr.E00551	269	scaffold_5:5278815-5279837	-
EgrNAC58	Eucgr.E00573	312	scaffold_5:5446847-5448354	+
EgrNAC59	Eucgr.E00574	312	scaffold_5:5454671-5456169	+
EgrNAC60	Eucgr.E00575	312	scaffold_5:5463114-5464594	+

EgrNAC61	Eucgr.E01053	399	scaffold_5:11289968-11292006	-
EgrNAC62	Eucgr.E01095	656	scaffold_5:11688351-11691755	-
EgrNAC63	Eucgr.E03225	136	scaffold_5:54730393-54730804	+
EgrNAC64	Eucgr.E03226	313	scaffold_5:54751853-54754536	-
EgrNAC65	Eucgr.F01091	366	scaffold_6:14069990-14071611	-
EgrNAC66	Eucgr.F01093	353	scaffold_6:14089859-14091139	-
EgrNAC67	Eucgr.F01170	357	scaffold_6:14978729-14980833	-
EgrNAC68	Eucgr.F01449	334	scaffold_6:18724670-18728883	-
EgrNAC69	Eucgr.F01463	316	scaffold_6:18838114-18840295	-
EgrNAC70	Eucgr.F01535	185	scaffold_6:19715905-19716689	-
EgrNAC71	Eucgr.F01536	420	scaffold_6:19771807-19774964	-
EgrNAC72	Eucgr.F01537	377	scaffold_6:19785158-19795459	-
EgrNAC73	Eucgr.F01538	401	scaffold_6:19808330-19810843	-
EgrNAC74	Eucgr.F01539	402	scaffold_6:19823490-19826009	-
EgrNAC75	Eucgr.F02615	320	scaffold_6:35802610-35804425	-
EgrNAC76	Eucgr.F02771	565	scaffold_6:37242090-37245705	+
EgrNAC77	Eucgr.F02910	645	scaffold_6:38768675-38772557	+
EgrNAC78	Eucgr.F03588	383	scaffold_6:44293211-44296055	-
EgrNAC79	Eucgr.F03962	271	scaffold_6:47920826-47922168	-
EgrNAC80	Eucgr.F03963	282	scaffold_6:47930910-47932324	-
EgrNAC81	Eucgr.F04097	386	scaffold_6:49244683-49247676	+
EgrNAC82	Eucgr.F04341	302	scaffold_6:52357968-52364215	+
EgrNAC83	Eucgr.G00054	432	scaffold_7:561581-563631	+
EgrNAC84	Eucgr.G01047	566	scaffold_7:18148061-18150813	-
EgrNAC85	Eucgr.G01049	468	scaffold_7:18173499-18175338	-
EgrNAC86	Eucgr.G01052	528	scaffold_7:18187867-18190938	-
EgrNAC87	Eucgr.G01053	540	scaffold_7:18203111-18206238	-
EgrNAC88	Eucgr.G01058	505	scaffold_7:18290552-18293821	-
EgrNAC89	Eucgr.G01060	592	scaffold_7:18328319-18331355	-
EgrNAC90	Eucgr.G01061	291	scaffold_7:18350722-18352015	-
EgrNAC91	Eucgr.G01063	208	scaffold_7:18362215-18363154	-
EgrNAC92	Eucgr.G01064	599	scaffold_7:18367448-18372224	-
EgrNAC93	Eucgr.G01066	214	scaffold_7:18382203-18383350	-
EgrNAC94	Eucgr.G01067	726	scaffold_7:18399300-18404004	-
EgrNAC95	Eucgr.G01069	200	scaffold_7:18414360-18415464	-
EgrNAC96	Eucgr.G01070	148	scaffold_7:18422025-18422909	-
EgrNAC97	Eucgr.G01071	215	scaffold_7:18430028-18431574	-
EgrNAC98	Eucgr.G01074	276	scaffold_7:18459516-18461108	-
EgrNAC99	Eucgr.G01075	253	scaffold_7:18464202-18465735	-
EgrNAC100	Eucgr.G01077	241	scaffold_7:18472658-18474231	-
EgrNAC101	Eucgr.G01078	249	scaffold_7:18479499-18482113	-
EgrNAC102	Eucgr.G01081	173	scaffold_7:18489388-18490615	-
EgrNAC103	Eucgr.G01082	229	scaffold_7:18494774-18496279	-
EgrNAC104	Eucgr.G01083	252	scaffold_7:18501578-18502984	-
EgrNAC105	Eucgr.G01507	314	scaffold_7:26090034-26091615	+
EgrNAC106	Eucgr.G01548	799	scaffold_7:26977231-26981938	+
EgrNAC107	Eucgr.G01550	525	scaffold_7:26995211-26999940	+
EgrNAC108	Eucgr.G01551	142	scaffold_7:27102229-27102788	+
EgrNAC109	Eucgr.G01553	314	scaffold_7:27116820-27118435	+
EgrNAC110	Eucgr.G01554	433	scaffold_7:27156206-27158590	+
EgrNAC111	Eucgr.G01555	142	scaffold_7:27167842-27168414	+
EgrNAC112	Eucgr.G01758	490	scaffold_7:32483475-32486540	+
EgrNAC113	Eucgr.G01984	241	scaffold_7:35939865-35941201	-
EgrNAC114	Eucgr.G02349	429	scaffold_7:41845203-41851658	-
EgrNAC115	Eucgr.G02486	282	scaffold_7:43382045-43383118	+
EgrNAC116	Eucgr.G02506	390	scaffold_7:43532994-43534477	-
EgrNAC117	Eucgr.G02740	412	scaffold_7:45473081-45477702	-
EgrNAC118	Eucgr.G02742	383	scaffold_7:45483146-45484914	-
EgrNAC119	Eucgr.H00614	246	scaffold_8:8324740-8327663	-
EgrNAC120	Eucgr.H00826	196	scaffold_8:10366720-10368168	-
EgrNAC121	Eucgr.H03362	243	scaffold_8:49228861-49230191	+
EgrNAC122	Eucgr.H03387	259	scaffold_8:49527394-49528610	-
EgrNAC123	Eucgr.H05089	288	scaffold_8:72636728-72638548	-
EgrNAC124	Eucgr.I00056	281	scaffold_9:932401-933614	-
EgrNAC125	Eucgr.I00057	293	scaffold_9:945378-946589	-
EgrNAC126	Eucgr.I00058	294	scaffold_9:960712-961926	-

EgrNAC127	Eucgr.I00059	293	scaffold_9:982712-984095	-
EgrNAC128	Eucgr.I00059.4	293	scaffold_9:982711-984095	-
EgrNAC129	Eucgr.I00060	294	scaffold_9:1010872-1012226	-
EgrNAC130	Eucgr.I00060.2	286	scaffold_9:1010871-1012247	-
EgrNAC131	Eucgr.I00095	293	scaffold_9:1990554-1991758	-
EgrNAC132	Eucgr.I00097(8)[c]	294	scaffold_9:2015283-2016494	-
EgrNAC133	Eucgr.I00099	293	scaffold_9:2029621-2031006	-
EgrNAC134	Eucgr.I00100	293	scaffold_9:2037886-2039269	-
EgrNAC135	Eucgr.I00101	294	scaffold_9:2054503-2055704	-
EgrNAC136	Eucgr.I00102	231	scaffold_9:2062222-2063030	-
EgrNAC137	Eucgr.I00191	204	scaffold_9:3901232-3902855	+
EgrNAC138	Eucgr.I00192	205	scaffold_9:3933513-3935099	+
EgrNAC139	Eucgr.I00193	231	scaffold_9:3942824-3944386	+
EgrNAC140	Eucgr.I00213	628	scaffold_9:4485737-4491109	-
EgrNAC141	Eucgr.I00583	244	scaffold_9:11983778-11984934	-
EgrNAC142	Eucgr.I00587	244	scaffold_9:12090444-12091597	-
EgrNAC143	Eucgr.I01494	301	scaffold_9:25146958-25148674	+
EgrNAC144	Eucgr.I01940	386	scaffold_9:29288937-29290797	+
EgrNAC145	Eucgr.I01958	305	scaffold_9:29495058-29496284	+
EgrNAC146	Eucgr.I02366	353	scaffold_9:34184572-34187433	-
EgrNAC147	Eucgr.I02571	324	scaffold_9:37055006-37056349	+
EgrNAC148	Eucgr.I02573	307	scaffold_9:37072424-37073743	+
EgrNAC149	Eucgr.I02574	156	scaffold_9:37077815-37078425	+
EgrNAC150	Eucgr.I02576	312	scaffold_9:37088684-37089995	+
EgrNAC151	Eucgr.I02578	184	scaffold_9:37097835-37099113	+
EgrNAC152	Eucgr.I02695	192	scaffold_9:38086280-38087232	+
EgrNAC153	Eucgr.J00505	240	scaffold_10:5382997-5384192	+
EgrNAC154	Eucgr.J00508	240	scaffold_10:5407603-5408789	+
EgrNAC155	Eucgr.J00509	240	scaffold_10:5438076-5439267	+
EgrNAC156	Eucgr.J00511	240	scaffold_10:5480784-5481969	+
EgrNAC157	Eucgr.J00512	241	scaffold_10:5500331-5501455	+
EgrNAC158	Eucgr.J00513	242	scaffold_10:5565803-5566917	+
EgrNAC159	Eucgr.J00514	237	scaffold_10:5644545-5645683	+
EgrNAC160	Eucgr.J00516	240	scaffold_10:5675996-5677141	+
EgrNAC161	Eucgr.J00517	249	scaffold_10:5699131-5700846	+
EgrNAC162	Eucgr.J00518	239	scaffold_10:5741072-5742566	+
EgrNAC163	Eucgr.J00519	240	scaffold_10:5770016-5771456	+
EgrNAC164	Eucgr.J00520	239	scaffold_10:5800427-5802517	+
EgrNAC165	Eucgr.J00521	209	scaffold_10:5822540-5824027	+
EgrNAC166	Eucgr.J00531	420	scaffold_10:5899713-5902252	+
EgrNAC167	Eucgr.J00940	340	scaffold_10:10271423-10272983	+
EgrNAC168	Eucgr.J01038	538	scaffold_10:11338856-11342587	-
EgrNAC169	Eucgr.J02254	430	scaffold_10:28401471-28405482	+
EgrNAC170	Eucgr.K01061	296	scaffold_11:13333715-13335656	-
EgrNAC171	Eucgr.K01228	297	scaffold_11:15492835-15494419	-
EgrNAC172	Eucgr.K01471	329	scaffold_11:17825556-17827411	+
EgrNAC173	Eucgr.K01472	365	scaffold_11:17839379-17840912	+
EgrNAC174	Eucgr.K01845	487	scaffold_11:23025448-23027173	+
EgrNAC175	Eucgr.K02205	218	scaffold_11:29289726-29290726	+
EgrNAC176	Eucgr.K02225	322	scaffold_11:29515773-29517736	+
EgrNAC177	Eucgr.K02303	235	scaffold_11:30215285-30218757	-
EgrNAC178	Eucgr.K03256	283	scaffold_11:41341485-41345296	+
EgrNAC179	Eucgr.K03356	397	scaffold_11:42561958-42565099	+
EgrNAC180	Eucgr.K03357	281	scaffold_11:42567237-42568645	+
EgrNAC181	Eucgr.K03358	226	scaffold_11:42572579-42573772	-
EgrNAC182	Eucgr.K03359	271	scaffold_11:42582884-42584228	+
EgrNAC183	Eucgr.K03360	245	scaffold_11:42604336-42605747	+
EgrNAC184	Eucgr.K03361	249	scaffold_11:42608150-42609513	+
EgrNAC185	Eucgr.L00819	168	scaffold_69:6684-7584	-
EgrNAC186	Eucgr.L01867	310	scaffold_423:7584-9133	-
EgrNAC187	Eucgr.L02267	82	scaffold_741:6537-7111	+
EgrNAC188	Eucgr.L02674	359	scaffold_1217:9229-10769	+
EgrNAC189	Eucgr.L03347	197	scaffold_2771:26-933	+

[a]In amino acids

[b]Excluding untranslated regions

[c]Collapsed gene models Eucgr.I00097 and Eucgr.I00098

Table S5. Biological functions of functionally characterized NAC domain proteins occurring in the subfamilies annotated in Fig. 1 of the main manuscript (detailed dendrogram available in Supporting Information Fig. S1).

Protein	Subfamily	General function	Specific function	References
ANAC054/CUC1	Ia	Development	Organ separation, gynoecium development	Aida *et al.*, 1997; Ishida *et al.*, 2000
ANAC098/CUC2			Organ separation, leaf development, axillary meristem formation	Aida *et al.*, 1997; Nikovics *et al.*, 2006; Peaucelle *et al.*, 2007; Raman *et al.*, 2008
ANAC031/CUC3			Organ separation, meristem initiation	Vroemen *et al.*, 2003; Hibara *et al.*, 2006
ANAC092			Leaf senescence	Oh *et al.*, 1997; Kim *et al.*, 2009; Balazadeh *et al.*, 2010
ANAC021	Ib	Development	Lateral root development, apical meristem specification	He *et al.*, 2005
ANAC030/VND7	Ic	Cell wall development	Secondary cell wall biosynthesis in xylem vessels	Kubo *et al.*, 2005; Yamaguchi *et al.*, 2010a; Zhong *et al.*, 2010; Yamaguchi *et al.*, 2011
ANAC101/VND6			Secondary cell wall biosynthesis in xylem vessels	Kubo *et al.*, 2005; Ohashi-Ito *et al.*, 2010; Yamaguchi *et al.*, 2010a
ANAC070/BRN2			Regulation of cell wall modification in root cap	Bennett *et al.*, 2010
ANAC015/BRN1			Regulation of cell wall modification in root cap	Bennett *et al.*, 2010
ANAC033/SMB			Regulation of cell wall modification in root cap	Bennett *et al.*, 2010
ANAC012/SND1			Secondary cell wall biosynthesis in fibres, endothecium, replum	Zhong *et al.*, 2006; Mitsuda *et al.*, 2007; Zhong *et al.*, 2007; Mitsuda & Ohme-Takagi, 2008; Zhong *et al.*, 2008
ANAC066/NST2			Secondary cell wall biosynthesis	Mitsuda *et al.*, 2005
ANAC043/NST1			Secondary cell wall biosynthesis	Mitsuda *et al.*, 2005; Mitsuda *et al.*, 2007; Zhong *et al.*, 2007; Mitsuda & Ohme-Takagi, 2008; Zhong *et al.*, 2008
ANAC008/SOG1	II	Cell wall development, response to DNA damage	Response to DNA damage	Preuss & Britt, 2003; Yoshiyama *et al.*, 2009
ANAC010/SND3			Regulation of secondary cell wall biosynthesis	Zhong *et al.*, 2008
ANAC073/SND2			Regulation of secondary cell wall biosynthesis	Zhong *et al.*, 2008; Hussey *et al.*, 2011

Table S5. *(continued from previous page)*

Protein	Subfamily	General function	Specific function	References
ANAC013	IIIa/b	Response to stress, development	Response to red light and UV-B	Safrany *et al.*, 2008
ANAC016			Response to chitin	Libault *et al.*, 2007
ANAC040/NTL8			Salt regulation of seed germination	Kim *et al.*, 2008
ANAC053/NTL4			Drought-induced leaf senescence	Lee *et al.*, 2012
ANAC062/NTL6			Defence response, response to cold treatment	Libault *et al.*, 2007; Seo *et al.*, 2010
ANAC078			Regulation of flavonoid biosynthesis	Morishita *et al.*, 2009
ANAC089			Regulation of flower development	Li *et al.*, 2010
ONAC054/RIM1	IIIc		Response to biotic stress	Yoshii *et al.*, 2009
ANAC069/NTM2	IVa		Salt and auxin signalling pathways	Park *et al.*, 2011
ANAC035/LOV1	IVb	Development	Regulation of cold response and flowering time	Yoo *et al.*, 2007
ANAC036			Regulation of leaf cell growth	Kato *et al.*, 2010
ANAC068/NTM1			Cytokinin signalling during cell division	Kim *et al.*, 2006
ANAC009/FEZ	IVd		Regulation of periclinal cell division in root cap	Willemsen *et al.*, 2008
ONAC063	IVd		Response to salt stress	Yokotani *et al.*, 2009
ANAC029/NAP	Va(1)		Leaf senescence	Guo & Gan, 2006
ONAC010	Va(2)		Anther dehiscence	Distelfeld *et al.*, 2012
ANAC081/ATAF2	Vb	Stress response	Repression of *PR* genes	Delessert *et al.*, 2005
ANAC019			Abiotic stress response; regulation of jasmonic acid-induced gene expression	Tran *et al.*, 2004; Bu *et al.*, 2008; Jiang *et al.*, 2009
ANAC055				
ANAC002/ATAF1			Drought response	Lu *et al.*, 2007
ANAC083/VNI2	VIa		Negative regulator of xylem vessel development	Yamaguchi *et al.*, 2010b
ANAC104/XND1	VIc		Regulation of secondary cell wall biosynthesis	Zhao *et al.*, 2008

Table S6. Putative *E. grandis* homologs of *Arabidopsis* NAC domain proteins known to be involved in regulating secondary cell wall biosynthesis

ANAC protein	Synonym	Putative *Eucalyptus* (co-)ortholog
ANAC012	SND1	EgrNAC61
ANAC073	SND2	EgrNAC170
ANAC010	SND3	EgrNAC44 EgrNAC45 EgrNAC46 EgrNAC47 EgrNAC64
ANAC043	NST1	EgrNAC49
ANAC066	NST2	-
ANAC083	VNI2	EgrNAC122
ANAC104	XND1	EgrNAC137 EgrNAC138 EgrNAC139 EgrNAC152
ANAC037 ANAC076	VND1 VND2	EgrNAC146
ANAC105	VND3	-
ANAC007 ANAC026	VND4 VND5	EgrNAC50
ANAC101	VND6	EgrNAC26
ANAC030	VND7	EgrNAC75

Table S8. Amino acid sequence logos of sixteen overrepresented motifs identified in EgrNAC proteins using MEME. The E-value, number of proteins containing each motif and, where applicable, the annotation of each motif is indicated.

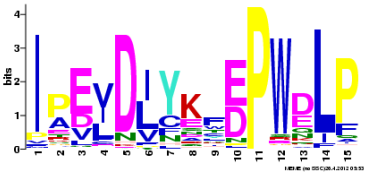
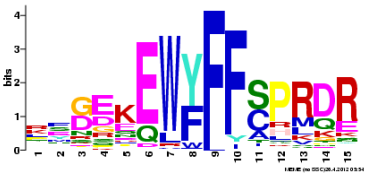
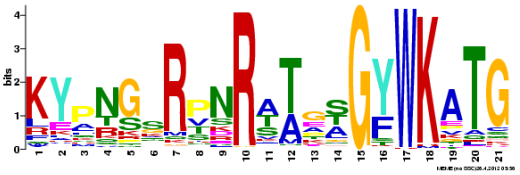
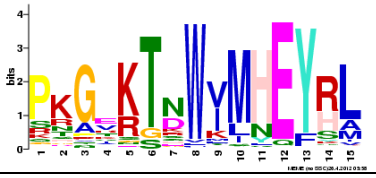
Motif name	Number of sites	Length	E-value	HMM annotation
Motif 1	173	15	2.7e-1645	NAC Subdomain A
				
Motif 2	159	15	1.3e-1161	NAC Subdomain B
				
Motif 3	122	15	5.6e-701	NAC Subdomain C
				
Motif 4	162	21	3.8e-1683	NAC Subdomain C
				
Motif 5	172	15	7.4e-1202	NAC Subdomain D
				
Motif 6	182	15	3.9e-1478	NAC Subdomain D
				

Table S8. (*continued from previous page*)

Motif name	Number of sites	Length	E-value	HMM annotation
Motif 7	165	11	2.0e-553	NAC Subdomain E
				
Motif 8	15	29	5.20E-260	-
				
Motif 9	9	50	4.50E-271	-
				
Motif 10	12	33	3.80E-245	-
				
Motif 11	20	21	4.20E-193	-
				
Motif 12	11	31	1.80E-162	-
				

Table S8. (*continued from previous page*)

Motif name	Number of sites	Length	E-value	HMM annotation
Motif 13	12	41	3.70E-306	-
				
Motif 14	11	50	2.10E-288	-
				
Motif 15	10	50	5.00E-216	-
				
Motif 16	7	50	9.60E-183	-
				

Table S9. Putative EgrNAC membrane-tethered transcription factors (MTFs) and their corresponding *Arabidopsis* NAC MTF homologs as deduced from Supporting Information Fig. S1.

Putative EgrNAC MTF	Homologous *Arabidopsis* MTFs (Kim *et al.*, 2010)
EgrNAC16	ANAC062
	ANAC091
EgrNAC39, EgrNAC166	ANAC040
	ANAC060
	ANAC089
EgrNAC62	ANAC014
	ANAC062
	ANAC091
EgrNAC76	ANAC016
	ANAC017
EgrNAC168	ANAC053
	ANAC078
EgrNAC112	ANAC068
	ANAC069

Table S10. *EgrNAC* genes occurring in blocks of tandem duplications (Fig. 3 of main manuscript). The subfamily classification of each gene (Fig. 1 of main manuscript) is also indicated.

Gene	Subfamily
Block 1	
EgrNAC1	IVa
EgrNAC2	
EgrNAC3	
EgrNAC4	
EgrNAC5	
EgrNAC6	
EgrNAC7	
EgrNAC8	
EgrNAC9	
EgrNAC10	
EgrNAC11	
EgrNAC12	
Block 2	
EgrNAC13	IVa
EgrNAC14	
Block 3	
EgrNAC19	Va
EgrNAC20	Ia
Block 4	
EgrNAC21	IVc
EgrNAC22	
EgrNAC23	
EgrNAC24	
EgrNAC25	
Block 5	
EgrNAC36	VIIIa
EgrNAC37	
EgrNAC38	
Block 6	
EgrNAC44	II
EgrNAC45	
EgrNAC46	
EgrNAC47	
Block 7	
EgrNAC53	VII
EgrNAC54	
EgrNAC55	
EgrNAC56	
EgrNAC57	
Block 8	
EgrNAC58	VII
EgrNAC59	
EgrNAC60	
Block 9	
EgrNAC63	Va(2)
EgrNAC64	II
Block 10	
EgrNAC65	Va
EgrNAC66	Vb
Block 11	
EgrNAC70	XI
EgrNAC71	
EgrNAC72	
EgrNAC73	
EgrNAC74	

Gene	Subfamily
Block 12	
EgrNAC79	VII
EgrNAC80	
Block 13	
EgrNAC84	IVa
EgrNAC85	
EgrNAC86	
EgrNAC87	
EgrNAC88	
EgrNAC89	
EgrNAC90	
EgrNAC91	
EgrNAC92	
EgrNAC93	
EgrNAC94	
EgrNAC95	
EgrNAC96	
EgrNAC97	
EgrNAC98	
EgrNAC99	
EgrNAC100	
EgrNAC101	
EgrNAC102	
EgrNAC103	
EgrNAC104	
Block 14	
EgrNAC106	IVa
EgrNAC107	
EgrNAC108	
EgrNAC109	
EgrNAC110	
EgrNAC111	
Block 15	
EgrNAC117	IIIa/b
EgrNAC118	
Block 16	
EgrNAC124	Vb
EgrNAC125	
EgrNAC126	
EgrNAC127	
EgrNAC129	
Block 17	
EgrNAC131	Vb
EgrNAC132	
EgrNAC133	
EgrNAC134	
EgrNAC135	
EgrNAC136	
Block 18	
EgrNAC137	IVc
EgrNAC138	
EgrNAC139	
Block 19	
EgrNAC141	IVc
EgrNAC142	

Gene	Subfamily
Block 20	
EgrNAC147	IVa
EgrNAC148	
EgrNAC149	
EgrNAC150	
EgrNAC151	Unassigned
Block 21	
EgrNAC153	IVc
EgrNAC154	
EgrNAC155	
EgrNAC156	
EgrNAC157	
EgrNAC158	
EgrNAC159	
EgrNAC160	
EgrNAC161	
EgrNAC162	
EgrNAC163	
EgrNAC164	
EgrNAC165	
EgrNAC166	IIIa/b
Block 22	
EgrNAC172	Va(1)
EgrNAC173	Ia
Block 23	
EgrNAC179	VII
EgrNAC180	
EgrNAC181	
EgrNAC182	
EgrNAC183	
EgrNAC184	

Note S1. Discussion of manually curated and discarded EgrNAC gene candidates. Low confidence annotations refer to those included in the v.1.0 annotation (Phytozome v.7) but excluded from the v.1.1 annotation (Phytozome v.8) of the *E. grandis* genome at www.phytozome.net. Evidence for expression was obtained from Eucspresso (eucspresso.bi.up.ac.za/; Mizrachi *et al.* 2010) and EucGenIE (eucgenie.bi.up.ac.za; Hefer *et al.* 2011).

Eucgr.A01272.1 Original gene model (426 AA) was N-terminally truncated. FGENESH analysis of scaffold_1:20605400..20609799 produced a 466 AA sequence. No matching transcript was found in the Eucspresso de novo assembly and no detectable expression was detected in the Eucgenie tissue set.

Eucgr.H03391.1 Gene model removed. Low-confidence annotation. Original 108 AA model appeared N-terminally truncated. No matches to *de novo* assembled transcripts were found on Eucspresso. Although EucGenIE RNA-seq reads mapped to the locus, they did not correspond to the predicted gene model. FGENESH annotation of the locus (scaffold_8:49595300..49597899) produced a 345 AA protein with high homology to a serine/threonine protein kinase in *Arabidopsis*.

Eucgr.G01554.1 Original gene model (168 AA) was N-terminally truncated. FGENES prediction using scaffold_7:27153700..27159199 yielded a 433 AA protein. Although no *de novo* assembled transcript represented this gene model in Eucspresso, there was some RNA-seq support for the longer gene model in EucGenIE.

Eucgr.G01267.1 Gene model removed. Low-confidence annotation. Original annotation (200 AA) appeared to be N- and C-terminally truncated. No matches to *de novo* assembled transcripts were found on Eucspresso. The gene model had some RNA-seq coverage based on EucGenIE, however FGENESH annotation of that region (scaffold_7:21324200..21327299) produced a longer (518 AA) protein with no significant homology to *Arabidopsis* proteins across most of its length.

Eucgr.A01274.1 Gene model removed. Low-confidence annotation. Original annotation (123 AA) appeared to be N- and C-terminally truncated. No matches to *de novo* assembled transcripts were found on Eucspresso. No RNA-seq coverage was detected in EucGenIE. FGENESH annotation of the locus (scaffold_1:20617700..20619999) produced a 287 AA protein with no significant homology to *Arabidopsis* proteins.

Eucgr.A00362.1 Original gene model (132 AA) lacked an initiation codon. No matching transcript was found in the Eucspresso de novo assembly. The FGENESH prediction for the region (scaffold_1:5072500..5077299) yielded a 273 AA protein that appeared to be full-length, with a BLAST result to Arabidopsis NAC proteins, and contained the full NAC domain.

Eucgr.A00361.1 Original gene model (200 AA) lacked an initiation codon. FGENESH annotation of the locus (scaffold_1:5065600..5068799) yielded a 228 AA model with initiation

codon and significant homology to *Arabidopsis* NAC proteins along most of its length. No matching transcript in the Eucspresso de novo assembly and no expression in the EucGenIE RNA-seq tissue set was found.

Eucgr.J01735.1 Gene model removed. Low-confidence annotation (68 AA) that lacked a start codon. No matches to *de novo* assembled transcripts were found on Eucspresso. No RNA-seq coverage in EucGenIE was evident. FGENESH annotation of the locus (scaffold_10:22614000..22616899) could not predict a full-length gene model.

Eucgr.E03225.1 Original gene model (90 AA) appeared C-terminally truncated. Although this is a low-confidence gene model with no significant RNA-seq coverage, FGENESH annotation of the locus (scaffold_5:54729700..54735699) yielded a 136 AA protein with significant BLASTP homology to *Arabidopsis* NAC domain proteins along its length.

Eucgr.G01265.1 Gene model removed. Low-confidence annotation (101 AA) with no RNA-seq coverage in EucGenIE and no matches to *de novo* assembled Eucgenie transcripts. Original gene model appeared C-terminally truncated. FGENESH annotation of the locus (scaffold_7:21243000..21246099) did not yield a full-length gene model.

Eucgr.B01593.1 Original gene model (101 AA) appeared C-terminally truncated. Although there was no significant RNA-seq coverage for the locus, FGENESH annotation of the region (scaffold_2:26870200..26872399) yielded a full-length 221 AA protein with significant homology to *Arabidopsis* NAC domain proteins along its full length.

Eucgr.A00360.1 Original annotation (106 AA) appeared to have a premature termination codon. Although there was no matching *de novo* assembled Eucspresso transcript, there was good EucGenIE RNA-seq coverage. FGENESH annotation of the locus (scaffold_1:5010900..5013799) yielded a longer 292 AA gene model with homology to *Arabidopsis* NAC proteins along most of its length.

Eucgr.G01069.1 Original annotation (121AA) appeared N-terminally truncated. Although there was no matching *de novo* assembled Eucspresso transcript, there was good RNA-seq coverage in EucGenIE. FGENESH annotation of the region yielded a 200 AA gene model with significant homology to *Arabidopsis* NAC proteins throughout its length

Eucgr.K01845.1 Original annotation (127AA) appeared to be C-terminally truncated. Although there was no supporting RNA-seq data for this model, FGENESH annotation of the region (scaffold_11:23024800..23029799) yielded a 487 AA protein with significant homology to *Arabidopsis* NAC domain proteins.

Eucgr.K03356.1 Original annotation (134 AA) lacked a stop codon. Although there was no supporting RNA-seq data for this model, FGENESH annotation of the locus

(scaffold_11:42561400..42565899) yielded a 397 AA protein with significant homology to *Arabidopsis* NAC domain proteins along most of its length.

Eucgr.A00363.1 The original annotation (235 AA) lacked a stop codon. Although there was no RNA-seq support for this model, FGENESH annotation of the locus (scaffold_1:5077300..5080999) yielded a 333 AA sequence with significant homology to *Arabidopsis* NAC proteins along most of the extended length.

Eucgr.I02571.1 Original annotation (141 AA) lacked a stop codon. Although there was no matching *de novo* assembled transcript in Eucspresso, a relatively large number of RNA-seq reads in the EucGenIE data mapped immediately downstream of the original gene model. FGENESH annotation of the locus (scaffold_9:37053800..37057199) yielded a 324 AA sequence with significant homology to *Arabidopsis* NAC domain proteins along its length.

Eucgr.I02573.1 Original annotation (148 AA) lacked a stop codon. Although there was no *de novo* assembled Eucspresso transcript, there was some RNA-seq coverage of the downstream region in EucGenIE. FGENESH annotation of the locus (scaffold_9:37071400..37074999) yielded a 307 AA sequence with significant homolog to *Arabidopsis* NAC domain proteins along its entire length.

Eucgr.I02576.1 Original annotation (148 AA) lacked a stop codon. Although there was no supporting RNA-seq for this gene model, FGENESH annotation of the locus (scaffold_9:37087700..37091499) yielded a longer 312 AA sequence with stop codon and significant homology to *Arabidopsis* NAC domain proteins along most of its length.

Eucgr.G01507.1 Original annotation (196 AA) lacked a stop codon. FGENESH annotation of the locus (scaffold_7:26088500..26093199) yielded a longer (314 AA) sequence supported by good RNA-seq coverage downstream of the original gene model. The extended gene model had significant homology to *Arabidopsis* NAC proteins along most of its length.

Eucgr.G01073.1 Gene model removed. The Phytozome annotation contained the full NAC domain; however there was a long stretch of N's 3' of the gene model. Because of this, the full-length gene model could not be determined. There was good evidence of expression of the locus in EucGenIE RNA-seq data but not in the Eucspresso data.

Eucgr.G01077.1 Original annotation (221 AA) lacked an annotated NAC domain on the N-terminal end, and good RNA-seq coverage supporting an alternative first exon structure. FGENESH annotation of the locus (scaffold_7:18471400..18476099) extended the N-terminal and provided a complete NAC domain.

Eucgr.E00573.1 Original annotation (217 AA) lacked a stop codon. There was no supporting RNA-seq coverage for this transcript. However, FGENESH annotation of the region (scaffold_5:5446600..5452699) produced a 314 AA model with stop codon and significant homology to *Arabidopsis* NAC domains proteins along most of its length.

Eucgr.E00574.1 Original annotation (288 AA) lacked a stop codon. FGENESH annotation of the locus (scaffold_5:5454300..5459999) yielded a 312 AA model with stop codon (the second exon was lengthened). There was supporting evidence of the longer gene model for some tissues in the EucGenIE RNA-seq data.

Eucgr.E00575.1 Original annotation (288 AA) lacked a stop codon. FGENESH predicted a 312 AA model for the locus (scaffold_5:5462800..5466499) with significant homology to *Arabidopsis* NAC domain proteins along most of its length, although there was no RNA-seq support for the transcript.

Eucgr.F03962.1 Original annotation (141 AA) had a stop codon but only a single exon, whereas FGENESH annotation of the locus (scaffold_6:47916200..47923199) yielded a two-exon 271 AA model with significant homology to *Arabidopsis* for around 2/3 of its length. There was no RNA-seq support for the transcript.

Eucgr.K03359.1 Original annotation (257 AA) lacked a stop codon. Although there was no RNA-seq support for the transcript, FGENESH annotation of the locus (scaffold_11:42582600..42585799) extended the second exon to a total of 271 AA.

Eucgr.F03963.1 Original annotation (136 AA) consisted of a single exon. FGENESH annotation of the locus (scaffold_6:47927800..47932699) yielded a 282 AA sequence with significant homology to *Arabidopsis* NAC domain protein along most of its length. There was no RNA-seq support for this transcript.

Eucgr.K03357.1 Original annotation (133 AA) lacked a stop codon. FGENESH annotation of the locus (scaffold_11:42567000..42571999) identified a second exon and a longer 281 AA sequence, with significant homology to *Arabidopsis* NAC domain sequences along most of its length. No evidence of expression was found in Eucspresso or EucGenIE.

Eucgr.K03361.1 Original annotation (183 AA) lacked a stop codon. FGENESH annotation of the locus (scaffold_11:42607700..42613299) yielded a 249 AA sequence with homology to *Arabidopsis* NAC domain proteins. No evidence of expression was found in Eucspresso or EucGenIE.

Eucgr.K03360.1 Although the original annotation (170 AA) had a stop codon, the gene model appeared unusually short. Since FGENESH annotation of the locus (scaffold_11:42603900..42607999) yielded a longer prediction (245 AA) with significant homology to *Arabidopsis* NAC domain proteins, the longer gene model was accepted. No evidence of expression was found in Eucspresso or EucGenIE.

Eucgr.B03439.1 The original annotation (250 AA) lacked a stop codon. FGENESH prediction of the locus (scaffold_2:59176400..59180399) extended the last exon until the stop codon to a 316 AA sequence with significant homology along most of its length to *Arabidopsis* NAC domain proteins. No evidence of expression was found in Eucspresso or EucGenIE.

Eucgr.K02225.1 Original annotation (309 AA) lacked a stop codon. FGENESH annotation of the locus (scaffold_11:29513300..29518299) extended the gene model to 322 AA, which had significant homology to *Arabidopsis* NAC domain proteins along most of its length. No evidence of expression was found in Eucspresso or EucGenIE.

Eucgr.D02027.1 (336 AA) Original annotation (335 AA) seemed full-length, but *E. grandis* RNA-seq data from bulked tissue mapped clearly to a longer second exon. FGENESH annotation of scaffold_4:34308200..34315399 corrected the second exon, making it longer by 19 AA.

Eucgr.J00521.1 Original annotation (169 AA) had a stop codon but the model was unusually short compared to close homologs. FGENESH annotation of the locus (scaffold_10:5821300..5826999) yielded an earlier initiation codon, yielding 40 extra amino acids with acceptable homology to *Arabidopsis* NAC61. The exon-intron structure was also in better agreement with close homologs (Eucgr.J00519.1) in the FGENESH correction.

Eucgr.J00940.1 Original annotation (233 AA) lacked a stop codon. Eucspresso data showed evidence of transcript coverage 3' of the existing gene model (below). FGENESH annotation of the locus (scaffold_10:10269000..10273599) extended the second exon (233 AA in total) in agreement with the RNA-seq profile.

Eucgr.G02486.1 Two alternative transcripts were predicted on Phytozome. The shorter of the two had a shorter third exon. The Eucspresso and EucGenIE RNA-seq data supported the shorter (282 AA) gene model, and there was no evidence of expression of the longer (primary) splice variant in other tissues.

Eucgr.I00060.1 and Eucgr.I00060.2 are not only alternative splice variants. Aside from a shared first and partial second exon, the rest of the gene models were derived from different open reading frames. Therefore, the two alternative splicing variants were treated as independent gene models.

Eucgr.I00101.1 Original annotation (359 AA) seemed to be usually long at the N-terminus. Inspection of EucGenIE RNA-seq bulk tissue revealed no expression of the first predicted exon, and FGENESH annotation of scaffold_9:2051600..2056899 also did not predict the first exon (model corrected to 294 AA).

Eucgr.I00097.1 (239 AA) and Eucgr.I00098.1 (146 AA) corrected to a single gene model (294 AA)

These two gene models appeared adjacent to each other and had similar RNA-seq transcript coverage. Eucgr.I00098.1 is a gene fragment. FGENESH annotation of the region encompassing both models (scaffold_9:2013500..2018199) created a hybrid gene model of 294 AAs that had complete BLAST coverage with ANAC002. The corrected gene model structure was in excellent agreement with close homologs (e.g. Eucgr.I00101.1).

Eucgr.I00059.1 (293 AA) and Eucgr.I00059.4 (293 AA) are not only alternative splice variants

There were four alternative splice variants of these gene models on Phytozome, but variants Eucgr.I00059.1 and Eucgr.I00059.4 appeared to be different genes altogether. The first 188 AAs were identical between the gene models Eucgr.I00059.1 and Eucgr.I00059.4, probably due to very recent tandem gene duplication. Both gene models showed good RNA-seq transcript coverage in EucGenIE. The two gene models had only 91.84% AA identity and were therefore unlikely to be allelic.

Eucgr.I00193.1 The RNA-seq profile for the original (158 AA) gene model clearly showed transcription of a third exon in EucGenIE, and FGENESH annotation of scaffold_9:3941600..3945799 resulted in a 231 AA model that encompassed the transcribed region.

Eucgr.L01867.1 (151 AA) FGENESH annotation of scaffold scaffold_423:7300..10099 yielded a 310 AA sequence.

Eucgr.L03347 (151 AA) FGENESH annotation of scaffold_2771:-249..2899 yielded a 197 AA sequence by extension the C-terminus. Mapped Eucspresso reads supported the longer gene model.

Eucgr.A01885.1 Gene model removed
BLASTS to VOZ1 (AT1G28520.2)

Eucgr.G01448.1 Gene model removed
BLASTS to VOZ1 (AT1G28520.2)

Eucgr.I02577.1 Gene model removed

This gene model appeared to be part of a larger transcript with good expression. The original annotation showed BLASTP hits to TAIR with low homology to NAC proteins (E > 0.001). The FGENESH annotation of scaffold_9:37091400..37098299, which incorporates the RNA-seq coverage for the locus, yielded a slightly longer protein, but also with low homology to NAC-like protein NTL9 (4e-08).

Eucgr.L02177.1 Gene model removed. Low-confidence annotation with apparently truncated C-terminal sequence. No improved predictions were obtained.

Note S2. Differences in phylogenetic clustering of NAC domain proteins (Fig. 1, main manuscript) compared to that of Zhu *et al.* (2012)

In our analysis (detailed dendrogram available in Supporting Information Fig. S1), subfamilies IIIa and IIIb could not be reliably dissociated and were combined into a single subfamily, IIIa/b. Four proteins in subfamily VIa (Zhu *et al.*, 2012) (ANAC084, PNAC134, PNAC135, VvNAC097) clustered with subfamily VIb in our phylogeny, and four rice genes previously assigned to VIb (ONAC001, ONAC005, ONAC139, ONAC041) clustered in VIa in our study. Three *Arabidopsis* proteins previously assigned to subfamily VIII (ANAC063, ANAC064, ANAC093; Zhu *et al.*, 2012) formed a well-supported clade with two *Arabidopsis* and three *Populus* NAC sequences that were unassigned to a subfamily by Zhu *et al.* (2012), allowing us to subdivide subfamily VIII into VIIIa and VIIIb. Eleven proteins unassigned by Zhu *et al.* (2012) (ANAC023, ANAC024, PNAC077, PNAC139, PNAC140, PNAC141, PNAC143, ONAC080, ONAC135, ONAC137, ONAC138) were incorporated into subfamily X. Finally, we defined an additional subfamily XI, from proteins unassigned by Zhu *et al.* (2012). Twenty-three proteins (~3%) remained unassigned.

References

Aida M, Ishida T, Fukaki H, Fujisawa H, Tasaka M. 1997. Genes involved in organ separation in *Arabidopsis*: an analysis of the *cup-shaped cotyledon* mutant. *The Plant Cell* **19**(6): 841-857.

Balazadeh S, Siddiqui H, Allu AD, Matallana-Ramirez LP, Caldana C, Mehrnia M, Zanor M-I, Kohler B, Mueller-Roeber B. 2010. A gene regulatory network controlled by the NAC transcription factor ANAC092/AtNAC2/ORE1 during salt-promoted senescence. *The Plant Journal* **62**(2): 250-264.

Bennett T, Toorn Avd, Sanchez-Perez GF, Campilho A, Willemsen V, Snel B, Scheres B. 2010. SOMBRERO, BEARSKIN1, and BEARSKIN2 regulate root cap maturation in *Arabidopsis*. *The Plant Cell* **22**: 640-654.

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

Delessert C, Kazan K, Wilson IW, Straeten DVD, Manners J, Dennis ES, Dolferus R. 2005. The transcription factor ATAF2 represses the expression of pathogenesis-related genes in *Arabidopsis*. *The Plant Journal* **43**: 745–757.

Distelfeld A, Pearce SP, Avni R, Scherer B, Uauy C, Piston F, Slade A, Zhao R, Dubcovsky J. 2012. Divergent functions of orthologous NAC transcription factors in wheat and rice. *Plant Molecular Biology* **78**: 515–524.

Guo Y, Gan S. 2006. AtNAP, a NAC family transcription factor, has an important role in leaf senescence. *The Plant Journal* **46**(4): 601-612.

He X-J, Mu R-L, Cao W-H, Zhang Z-G, Zhang J-S, Chen S-Y. 2005. AtNAC2, a transcription factor downstream of ethylene and auxin signaling pathways, is involved in salt stress response and lateral root development. *The Plant Journal* **44**: 903–916.

Hefer C, Mizrachi E, Joubert F, Myburg A. 2011. The *Eucalyptus* genome integrative explorer (EucGenIE): a resource for *Eucalyptus* genomics and transcriptomics. *BMC Proceedings* **5**(Suppl 7): O49.

Hibara K-i, Karim MR, Takada S, Taoka K-i, Furutani M, Aida M, Tasaka M. 2006. *Arabidopsis CUP-SHAPED COTYLEDON3* regulates postembryonic shoot meristem and organ boundary formation. *The Plant Cell* **18**: 2946-2957.

Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L, Widmayer P, Gruissem W, Zimmermann P. 2008. Genevestigator V3: A reference expression database for the meta-analysis of transcriptomes. *Advances in Bioinformatics* **2008**: 420747.

Hussey SG, Mizrachi E, Spokevicius AV, Bossinger G, Berger DK, Myburg AA. 2011. *SND2*, a NAC transcription factor gene, regulates genes involved in secondary cell wall development in *Arabidopsis* fibres and increases fibre cell area in *Eucalyptus*. *BMC Plant Biology* **11**: 173.

Ishida T, Aida M, Takada S, Tasaka M. 2000. Involvement of *CUP-SHAPED COTYLEDON* genes in gynoecium and ovule development in *Arabidopsis thaliana*. *Plant and Cell Physiology* **41**(1): 60-67.

Jiang H, Li H, Bu Q, Li C. 2009. The RHA2a-interacting proteins ANAC019 and ANAC055 may play a dual role in regulating ABA response and jasmonate response. *Plant Signalling and Bahaviour* **4**(5): 464-466.

Kato H, Motomura T, Komeda Y, Saito T, Kato A. 2010. Overexpression of the NAC transcription factor family gene *ANAC036* results in a dwarf phenotype in *Arabidopsis thaliana*. *Journal of Plant Physiology* **167**: 571–577.

Kim JH, Woo HR, Kim J, Lim PO, Lee IC, Choi SH, Hwang D, Nam HG. 2009. Trifurcate feed-forward regulation of age-dependent cell death involving *miR164* in *Arabidopsis*. *Science* **323**(5917): 1053-1057.

Kim S-G, Lee A-K, Yoon H-K, Park C-M. 2008. A membrane-bound NAC transcription factor NTL8 regulates gibberellic acid-mediated salt signaling in *Arabidopsis* seed germination. *The Plant Journal* **55**: 77–88.

Kim S-G, Lee S, Seo PJ, Kim S-K, Kim J-K, Park C-M. 2010. Genome-scale screening and molecular characterization of membrane-bound transcription factors in *Arabidopsis* and rice. *Genomics* **95**: 56-65.

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

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

Lee S, Seo PJ, Lee H-J, Park C-M. 2012. A NAC transcription factor NTL4 promotes reactive oxygen species production during drought-induced leaf senescence in *Arabidopsis*. *The Plant Journal* **70**: 831–844.

Li J, Zhang J, Wang X, Chen J. 2010. A membrane-tethered transcription factor ANAC089 negatively regulates floral initiation in *Arabidopsis thaliana*. *Science China. Life Sciences* **53**(11): 1299-1306.

Libault M, Wan J, Czechowski T, Udvardi M, Stacey G. 2007. Identification of 118 *Arabidopsis* transcription factor and 30 ubiquitin-ligase genes responding to chitin, a plant-defense elicitor. *Molecular Plant-Microbe Interactions* **20**(8): 900-911.

Lu P-L, Chen N-Z, An R, Su Z, Qi B-S, Ren F, Chen J, 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 Molecular Biology* **63**: 289–305.

Mitsuda N, Iwase A, Yamamoto H, Yoshida M, Seki M, Shinozaki K, Ohme-Takagi M. 2007. NAC transcription factors, NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of *Arabidopsis*. *The Plant Cell* **19**: 270–280.

Mitsuda N, Ohme-Takagi M. 2008. NAC transcription factors NST1 and NST3 regulate pod shattering in a partially redundant manner by promoting secondary wall formation after the establishment of tissue identity. *The Plant Journal* **56**(5): 768 - 778.

Mitsuda N, Seki M, Shinozaki K, 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.

Morishita T, Kojima Y, Maruta T, Nishizawa-Yokoi A, Yabuta Y, Shigeoka S. 2009. *Arabidopsis* NAC transcription factor, ANAC078, regulates flavonoid biosynthesis under high-light. *Plant and Cell Physiology* **50**(12): 2210-2222.

Nikovics K, Blein T, Peaucelle A, Ishida T, Morin H, Aida M, Laufs P. 2006. The balance between the *MIR164A* and *CUC2* genes controls leaf margin serration in *Arabidopsis*. *The Plant Cell* **18**: 2929–2945.

Oh SA, Park J-H, Lee GI, Paek KH, Park SK, Nam HG. 1997. Identification of three genetic loci controlling leaf senescence in *Arabidopsis thaliana*. *The Plant Journal* **12**(3): 527–535.

Ohashi-Ito K, Oda Y, Fukuda H. 2010. *Arabidopsis* VASCULAR-RELATED NAC-DOMAIN6 directly regulates the genes that govern programmed cell death and secondary wall formation during xylem differentiation. *The Plant Cell* **22**: 3461–3473.

Park J, Kim Y-S, Kim S-G, Jung J-H, Woo J-C, Park C-M. 2011. Integration of auxin and salt signals by the NAC transcription factor NTM2 during seed germination in *Arabidopsis*. *Plant Physiology* **156**: 537–549.

Peaucelle A, Morin H, Traas J, Laufs P. 2007. Plants expressing a *miR164*-resistant *CUC2* gene reveal the importance of post-meristematic maintenance of phyllotaxy in *Arabidopsis*. *Development* **134**: 1045-1050.

Preuss SB, Britt AB. 2003. A DNA-damage-induced cell cycle checkpoint in *Arabidopsis*. *Genetics* **164**: 323-334.

Raman S, Greb T, Peaucelle A, Blein T, Laufs P, Theres K. 2008. Interplay of *miR164*, *CUP-SHAPED COTYLEDON* genes and *LATERAL SUPPRESSOR* controls axillary meristem formation in *Arabidopsis thaliana*. *The Plant Journal* **55**: 65–76.

Safrany J, Haasz V, Mate Z, Ciolfi A, Feher B, Oravecz A, Stec A, Dallmann G, Morelli G, Ulm R, Nagy F. 2008. Identification of a novel cis-regulatory element for UV-B-induced transcription in *Arabidopsis*. *The Plant Journal* **54**(3): 402–414.

Seo PJ, Kim MJ, Park J-Y, Kim S-Y, Jeon J, Lee Y-H, Kim J, Park C-M. 2010. Cold activation of a plasma membrane-tethered NAC transcription factor induces a pathogen resistance response in *Arabidopsis*. *The Plant Journal* **61**(4): 661–671.

Toufighi K, Brady SM, Austin R, Ly E, Provart NJ. 2005. The Botany Array Resource: e-Northerns, Expression Angling, and promoter analyses. *The Plant Journal* **43**: 153-163.

Tran L-SP, Nakashima k, Sakuma Y, Simpson SD, Fujita Y, Maruyama K, Fujita M, Seki M, Shinozaki K, 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. *The Plant Cell* **16**: 2481-2498.

Vroemen CW, Mordhorst AP, Albrecht C, Kwaaitaal MACJ, Vries SCd. 2003. The *CUP-SHAPED COTYLEDON3* gene is required for boundary and shoot meristem formation in *Arabidopsis*. *The Plant Cell* **15**(7): 1563–1577.

Wilkins O, Nahal H, Foong J, Provart NJ, Campbell MM. 2009. Expansion and diversification of the *Populus* R2R3-MYB family of transcription factors. *Plant Physiology* **149**(2): 981-993.

Willemsen V, Bauch M, Bennett T, Campilho A, Wolkenfelt H, Xu J, Haseloff J, Scheres B. 2008. The NAC domain transcription factors FEZ and SOMBRERO control the orientation of cell division plane in *Arabidopsis* root stem cells. *Dev Cell* **15**(6): 913-922.

Yamaguchi M, Mitsuda N, Ohtani M, Ohme-Takagi M, Kato K, Demura T. 2011. VASCULAR-RELATED NAC-DOMAIN 7 directly regulates the expression of a broad range of genes for xylem vessel formation. *The Plant Journal* **66**: 579–590.

Yamaguchi M, Nadia G, Igarashi H, Ohtani M, Nakano Y, Mortimer JC, Nishikubo N, Kubo M, Katayama Y, Kakegawa K, Dupree P, Demura T. 2010a. VASCULAR-

RELATED NAC-DOMAIN6 (VND6) and VND7 effectively induce transdifferentiation into xylem vessel elements under control of an induction system. *Plant Physiology* **153**: 906-914.

Yamaguchi M, Ohtani M, Mitsuda N, Kubo M, Ohme-Takagi M, Fukuda H, Demura T. 2010b. VND-INTERACTING2, a NAC domain transcription factor, negatively regulates xylem vessel formation in *Arabidopsis*. *Plant Cell* **22**(4): 1249-1263

Yokotani N, Ichikawa T, Kondou Y, Matsui M, Hirochika H, Iwabuchi M, Oda K. 2009. Tolerance to various environmental stresses conferred by the salt-responsive rice gene *ONAC063* in transgenic *Arabidopsis*. *Planta* **229**(5): 1065-1075.

Yoo SY, Kim Y, Kim SY, Lee JS, Ahn JH. 2007. Control of flowering time and cold response by a NAC-domain protein in *Arabidopsis*. *PLoS ONE* **2**(7): e642.

Yoshii M, Shimizu T, Yamazaki M, Higashi T, Miyao A, Hirochika H, Omura T. 2009. Disruption of a novel gene for a NAC-domain protein in rice confers resistance to *Rice dwarf virus*. *The Plant Journal* **57**(4): 615–625.

Yoshiyama K, Conklin PA, Huefner ND, Britt AB. 2009. Suppressor of gamma response 1 (*SOG1*) encodes a putative transcription factor governing multiple responses to DNA damage. *Proceedings of the National Academy of Sciences of the USA* **106**(31): 12843-12848.

Zhao C, Avci U, Grant EH, Haigler CH, Beers EP. 2008. XND1, a member of the NAC domain family in *Arabidopsis thaliana*, negatively regulates lignocellulose synthesis and programmed cell death in xylem. *The Plant Journal* **53**(3): 425-436.

Zhong R, Demura T, Ye ZH. 2006. SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of *Arabidopsis*. *Plant Cell* **18**(11): 3158-3170.

Zhong R, Lee C, Ye Z-H. 2010. Global analysis of direct targets of secondary wall NAC master switches in *Arabidopsis*. *Molecular Plant* **3**(6): 1087-1103.

Zhong R, Lee C, Zhou J, McCarthy RL, Ye Z-H. 2008. A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in *Arabidopsis*. *The Plant Cell* **20**: 2763-2782.

Zhong R, Richardson EA, Ye Z-H. 2007. Two NAC domain transcription factors, SND1 and NST1, function redundantly in regulation of secondary wall synthesis in fibers of *Arabidopsis*. *Planta* **225**(6): 1603-1611.

Zhu T, Nevo E, Sun D, Peng J. 2012. Phylogenetic analyses unravel the evolutionary history of NAC proteins in plants. *Evolution* **66**(6): 1833-1848