

Genomics, Molecular and Evolutionary Perspective of NAC Transcription Factors

Tapan Kumar Mohanta^{1*s}, Dhananjay Yadav^{2s}, Adil Khan¹, Abeer Hashem^{3,4}, Baby Tabassum⁵, Abdul Latif Khan¹, Eslayed Fathi Abda_Allah⁶, Ahmed Al-Harrasi^{1*}

¹Natural and Medicinal Plant Sciences Research Center, University of Nizwa, Nizwa, 616, Oman

²Dept. of Medical Biotechnology, Yeungnam University, Gyeongsan, 38541, Republic of Korea.

³Botany and Microbiology Department, College of Science, King Saud University, P.O. Box. 2460, Riyadh 11451, Saudi Arabia.

⁴Mycology and Plant Disease Survey Department, Plant Pathology Research Institute, ARC, Giza 12511, Egypt.

⁵Toxicology laboratory, Department of Zoology, Raza P.G. College, Rampur, Uttar Pradesh, 244091, India

⁶Plant Production Department, College of Food and Agricultural Sciences, King Saud University, P.O. Box. 2460 Riyadh 11451, Saudi Arabia.

s Contributed equally

*Corresponding author

Tapan Kumar Mohanta, E-mail: nostoc.tapan@gmail.com, tapan.mohanta@unizwa.edu.om

Ahmed Al-Harrasi, E-mail: aharrasi@unizwa.edu.om

Abstract

NAC (NAM, ATAF1,2, and CUC2) transcription factors are one of the largest transcription factor families found in plants and are involved in diverse developmental and signalling events. Despite the availability of comprehensive genomic information from diverse plant species, the basic genomic, biochemical, and evolutionary details of NAC TFs have not been established. Therefore, NAC TFs family proteins from 160 plant species were analyzed in the current study. The analysis, among other things, identified the first algal NAC TF in the Charophyte, *Klebsormidium flaccidum*. Furthermore, our analysis revealed that NAC TFs are membrane bound and contain monopartite, bipartite, and multipartite nuclear localization signals. NAC TFs were also found to encode a novel chimeric protein domain and are part of a complex interactome network. Synonymous codon usage is absent in NAC TFs and it appears that they have evolved from orthologous ancestors and undergone significant duplication events to give rise to paralogous NAC TFs.

Keywords: NAC transcription factor, Transmembrane domain, Chimeric NAC transcription factor, Nuclear localization signal, Codon usage, Gene duplication

Introduction

Next-generation sequencing (NGS) has fostered the sequencing of many plant genomes. The availability of so many genomes has allowed researchers to readily identify genes, examine genetic diversity within a species, and gain insight into the evolution of genes and gene families. Gene expression is regulated in part by different families of proteins known as transcription factors (TFs). TFs are involved in inducing the transcription of DNA into RNA. They include numerous and diverse proteins, all of which contain one or more DNA-binding motifs. The DNA-binding domain enables them to bind to the promoter or repressor sequence of DNA that is present either at the upstream, downstream, or within an intron region of a coding gene. Some TFs bind to a DNA promoter region located near the transcription start site of a gene and help to form the transcription initiation complex. Other TFs bind to regulatory enhancer sequences and stimulate or repress transcription of the related genes. Regulating transcription is of paramount importance to controlling gene expression and TFs enable the expression of an individual gene in a unique manner, such as during different stages of development or in response to biotic or abiotic stress. TFs act as a molecular switch for temporal and spatial gene regulation. A considerable portion of a genome consists of genes encoding transcription factors. For example, there are at least 52 TF families in *Arabidopsis thaliana*, and the NAC (no apical meristem (NAM) TF family is one of them.

NAC TFs are characterised by the presence of a conserved N-terminal NAC domain comprising approximately 150 amino acids and a diversified C-terminal end. The DNA binding NAC domain is divided into five sub-domains designated A-E. Sub-domain A is apparently involved in the formation of functional dimers, while sub-domains B and E appear to be responsible for the functional divergence of NAC genes¹⁻⁴. The dimeric architecture of NAC proteins can remain stable even at a concentration of 5M NaCl⁴. The dimerization is established by Leu14-Thr23, and Glu26-Tyr31 amino acid residues. The dimeric form is responsible for the functional unit of stress-responsive SNAC1 and can modulate DNA-binding specificity. Sub-domains C and D contain positively charged amino acids that bind to DNA. The crystal structure of the SNAC1 TF revealed the presence of a central semi- β -barrel formed from seven twisted anti-parallel β -strands with three α -helices⁴. The NAC domain is most responsible for DNA binding activity that lies between amino acids Val119-Ser183, Lys123-Lys126, with Lys79, Arg85, and Arg88 reside within different strands of β -sheets^{2,5,6}. The remaining portion of the NAC domain contains a loop region composed of the amino acids, Gly144-Gly149 and Lys180-Asn183, which are very flexible in nature⁴. The loop region of SNAC1 is quite long and different from the loop region of ANAC, an abscisic-acid-responsive NAC, and could underlie the basis for different biological functions. NAC TFs possesses mono or bipartite nuclear localization signals which contain a Lys residue in sub-domain D^{1,6-8}. In addition, NAC proteins, as part of a mechanism of self-regulation, also modulate the expression of several other proteins^{6,9}. The D subunit of a few NAC TFs contain a hydrophobic negative regulatory domain (NRD), comprised of **L-V-F-Y** amino acids, which is involved in suppressing transcriptional activity¹⁰. For example, the NRD domain can suppress the transcriptional activity of Dof, WRKY, and APETALA 2/dehydration responsive elements (AP2/DRE) TFs¹⁰.

Studies indicate that the diverse C-terminal domain contains a transcription regulatory region (TRR) which has several group-specific motifs that can activate or repress transcription activity^{11–14}. The diverse C-terminal region imparts differences in the function of individual NAC proteins by regulating the interaction of NAC TFs with diverse target proteins. Although the C-terminal region of NAC TFs is diverse, it also contains group-specific conserved motifs¹⁵. Although the diverse aspects of NAC TFs have been studied, most were conducted within individual plant species. A detailed comparative study of the genomic, molecular biology, and evolution of NAC TFs has not been conducted. Therefore, a comprehensive analysis of NAC TFs is presented in the current study.

Results and discussion

NAC transcription factors exhibit diverse genomic and biochemical features

Advancements in genome sequencing technology have enabled the discovery of the genomic details of hundreds of plant species. The availability of this genome sequence data allowed us to characterize the genomic details of NAC TFs in diverse plant species. The presence of NAC TFs in 160 species (18774 NAC sequences) was identified and served as the basis of the conducted analyses. Comparisons of NAC sequences revealed that *Brassica napus* has the highest number (410) of NAC TFs, while the pteridophyte plant, *Marchantia polymorpha*, was found to contain the lowest number (9) (Table 1). On average, monocot plants contain a higher (141.20) number of NAC TFs relative to dicot plants (125.56). Except for *Hordeum vulgare* (76), *Saccharum officinarum* (44), and *Zostera marina* (62) all other monocot species possess more than one hundred NAC TFs each (Table 1). Lower eukaryotic plants, bryophytes and pteridophytes also possess NAC TFs. In addition, the algal species, *Klebsormidium flaccidum*, also contains NAC TFs and this finding represents the first report of NAC TFs in algae (Table 1). A NAC TF in *Trifolium pratense* (Tp57577_TGAC_v2_mRNA14116) was found to be the largest NAC TF, comprising 3101 amino acids, while a NAC TF in *Fragaria x ananassa* (FANhyb_icon00034378_a.1.g00001.1) was found to be the smallest NAC TF, comprising only 25 amino acids. Although it only contains a 25 amino acid sequence, it still encodes a NAC domain. Typically, NAC TFs contain a single NAC domain located near the N-terminal region of the protein. The current analysis, however, also identified NAC TFs with two NAC domains. At least 77 of the 160 studied species were found to contain two NAC domains (Table 1).

Multiple sequence alignment revealed the presence of a conserved consensus sequence at the N-terminus. The major conserved consensus sequences are P-G-F-R-F-H-P-T-D-D/E-L-I/V, Y-L-x₂-K, D-L-x-K-x₂-P-W-x-L-P, E-W-Y-F-F, G-Y-W-K-A/T-T-G-x-D-x₁₋₂-I/V, G-x-K-K-x-L-V-F-Y, and T-x-W-x-M-H-E-Y. Among these consensus sequences, D-D/E-L-I/V, E-W-Y-F-F, G-Y-W-K, and M-H-E-Y are the conserved motifs most observed. The D-D/E-L motif is a characteristic feature of the calcium-binding motifs present in the EF-hand of calcium-dependent protein kinases and the presence of this motif in NAC TFs indicates that they have the potential to regulate Ca²⁺ signalling events in cells¹⁶. The D-D-E/E motif is located in the β' sheet whereas the Y-L-x₂-K motif is in the α1a/b chain. Except for G-F-R-F-H-P-T-D-D/E-L-I/V, the conserved consensus sequences contain the positively charged amino acids Lys (L) and Arg (K) that can bind to negatively charged DNA. Welner et al. (2012) published the crystal structure of

ANAC019 and reported that Y⁹⁴-W-K-A-T-G-T-D in β₃, I¹¹-K-K-A-L-V-F-Y of β₄, K¹²³-A-P-K-G-T-K-T-N-W in the loop between β₄ and β₅, and I¹³³-M-H-E-Y-R of β₅ and Y¹⁶⁰-K-K-Q at the C-terminal end are located close to the bound DNA and are associated with DNA binding activity¹⁷. They reported that Y⁹⁴-W-K-A-T-G-T-D is responsible for the specific recognition of DNA and binds at the major groove within DNA, whereas I¹¹-K-K-A-L-V-F-Y, K¹²³-A-P-K-G-T-K-T-N-W, I¹³³-M-H-E-Y-R, and Y¹⁶⁰-K-K-Q bind to the backbone of the DNA molecule and provide affinity for DNA binding activity¹⁷. In the present analysis of 160 plant species, the identification of the conserved consensus sequences G-Y-W-K-A/T-T-G-x-D-x₁₋₂-I/V, G-x-K-K-x-L-V-F-Y, and T-x-W-x-M-H-E-Y is in agreement with Welner et al (2012); suggesting that NAC TFs contain conserved consensus sequences for specific DNA recognition and increasing the affinity for DNA binding.

Hao et al., (2010) reported that the D subunit of NAC TFs contain a hydrophobic L-V-F-Y amino acid motif that suppresses WRKY, Dof, and APETALA2 transcriptional regulators. This suggests that NAC TFs may also function as a negative regulator of transcription¹⁰. Our sequence alignment, however, revealed that NAC TF family proteins in many different and diverse plant species possess a conserved hydrophobic L-V-F-Y motif. As reported by Hao et al., (2010), all of the NAC TFs have the potential to suppress the transcriptional activity of WRKY, Dof, and APETALA 2/dehydration responsive element TFs; however, it is highly unlikely that an organism could regulate transcriptional events in a specific and sustained manner if NAC will conduct transcriptional repression.

The molecular weight of NAC TFs ranged from 346.46 kilodaltons (kDa) (*Trifolium pratense*_Tp57577_TGAC_v2) to 2.94 kDa (*Fragaria x ananassa*_FANhyb_icon00034378_a.1.g00001.1) (Figure 1). Among the studied NAC TFs, only 10 NAC proteins have a molecular weight (MW) more than 200 kDa and approximately 99 are between 100 to 200 kDa. The MW of the majority of the NAC proteins range between 40 to 55 kDa (Figure 1). The Isoelectric point (pI) of the NAC proteins ranged from 11.47 (Brast01G304500.1.p, (*Brachypodium stacei*) to 3.60 (ObartAA03S_FGP19036, *Oryza barthii*). The majority of the NAC TFs fell within a pI range of 5-8 (Figure 2). Among the 18774 analysed NAC TFs, the pI of 99 of them were ≥ 10. Approximately 69.28% of the NAC TFs had a pI that was in an acidic range, whereas the remaining 30.72% had a pI within in a basic range. A protein with a pH below the pI carries a net positive charge, whereas a protein with a pH above the pI carries a net negative charge. The pI of a protein determines its transport, solubility, and sub-cellular localization¹⁸⁻²⁰. Biomembranes, such as those surrounding the nucleus, are negatively charged; as a result, positively charged (acidic pI) NAC TFs are readily attracted to the nuclear membrane and subsequently transported into the nucleus to function in transcriptional regulation. There are, however, approximately 30.72% NAC TFs that possess a basic Pi; suggesting that they are localized in the cytosol or plasma membrane of the cell. The major role of TFs is to bind to specific DNA sequences to regulate transcription. The majority of proteins have either an acidic or basic pI and those with a neutral pI close to 7.4 are few because proteins tend to be insoluble, unreactive, and unstable at a pH close to its pI. This is the main reason why among the 18774 NAC TFs analysed, only two (XP_010925972.1, *Elaeis guineensis*; Lus10008200, *Linum usitatissimum*) had a pI 7.4. The existence of NAC proteins with a pI above 10 led us to speculate whether these TFs function while attached to a transmembrane domain. Therefore, additional

analyses were conducted to determine if NAC TFs also have the potential to bind to the transmembrane domain or if the NAC TFs with a basic pI remain within the cytosol.

NAC TFs are membrane bound

Transcription factors regulate diverse cellular events at transcriptional, translational, and posttranslational levels. They are also involved in nuclear transport and posttranslational modifications. In several cases, TFs are synthesized but remain inactive in the cytoplasm and are only induced into activity through non-covalent interactions^{21,22}. TFs are able to remain inactive through their physical association with intracellular membranes and are released by proteolytic cleavage. NAC TFs are a family of proteins whose numbers are in the hundreds in the majority of plant species. The fact that NAC TFs are such a large protein family, it is not surprising that NAC TFs have evolved diverse functional roles. Therefore, it is plausible that NAC TFs may be associated with sub-cellular organelle other than the nucleus to fulfil their diverse functional roles. It is essential, however, to confirm if NAC TFs contain signalling sequences for transmembrane localization. Therefore, we analysed the NAC gene sequences to determine if the signalling sequences present in NAC TFs possess a transmembrane domain.

Results indicated that at least 2190 (8.57%) NAC TFs possess a transmembrane domain. Transmembrane domains were found at both the N- and C-terminal ends of NAC proteins. In the majority of the cases, however, the transmembrane domain was located towards the C-terminal end. Seo et al., (2008) indicated the presence of a transmembrane domain in TFs and suggested that transmembrane domain functions through two proteolytic mechanisms, commonly known as regulated ubiquitin/proteasome-dependent (RUP) and regulated intramembrane proteolysis (RIP)^{23,24}. The bZIP plant TF is present as an integral membrane protein associated with stress response in the endoplasmic reticulum (ER)^{25–28}. Studies suggest that the majority of membrane bound TFs are associated with the ER and a membrane bound TF was also found to be involved in cell division^{29,30}. At least 10% of the TFs in *Arabidopsis thaliana* have been reported to be transmembrane bound³⁰. The collective evidence clearly indicates that membrane-mediated transcriptional regulation is a common stress response and that NAC TFs play a vital role in stress resistance in the ER. Therefore, these membrane-bound NAC TFs can be of great importance for the manipulation of stress resistance using biotechnology.

NAC TF contain monopartite, bipartite, non-canonical, and nuclear export signal sequences

The import of NAC TFs into the nucleus is mediated by nuclear membrane-bound importins and exportins that form a ternary complex consisting of importin α , importin β 1, and a cargo molecule. Importin α serve as an adaptor molecule of importin β 1 and recognises the nuclear localization signal (NLS) of the cargo protein needing to be imported. Importin β 1 and β 2, however, also recognize the NLS directly and bind to the cargo protein. Although the NLS of TFs have been widely studied in the animal kingdom, their study in plants has been more restricted. Therefore, the NLS of NAC TFs was examined in the current study. Results indicate that NAC TFs contain diverse NLS. The NLS were found in the N- and C-terminal regions of NAC TF proteins. Some NAC TFs were found to contain only one NLS whereas other contain multiple NLS. At least 3579 of the total NAC TFs analysed were found to contain either one or multiple NLS. More specifically, 2604 NAC TFs were found to possess only one NLS at the N-terminal end of the

NAC protein, whereas 975 were found to possess two NLS, 254 possess three NLS, and 48 were possess four NLS. The NLS were located towards the N-terminal end in the majority of NAC proteins.

NLS motifs are rich in positively charged amino acids and bind to importin α to be imported into the nucleus. The NLS motifs are classified as monopartite or bipartite. A monopartite NLS contains a single cluster of positively charged amino acids and are grouped into two subclasses, class-I and class-II. Class-I possesses four consecutive positively charged amino acids and class-II contains three positively charged amino acids, represented by K(K/R)-x-K/R; where x represents any amino acid that is present after two basic amino acids. Bipartite NLS motifs contain two clusters of positively charged amino acids separated by a 10-12 amino acid linker sequence. Bipartite NLS motifs are characterised by the consensus sequence K-R-P-A-A-T-K-K-A-G-Q-A-K-K-K-K. In addition to monopartite and bipartite NLS motifs, importin α also recognises non-canonical NLS motifs. Non-canonical NLS motifs are longer and considerably variable relative to monopartite and bipartite NLS motifs and are classified as class-III and class-IV NLS. Non-canonical NLS motifs are usually present in the C-terminal end and bind with importin β 2. Class-III and class-IV NLS motifs contain K-R-x(W/F/Y)-x₂-A-F and (P/R)-x₂-K-R-(K/R) consensus sequences, respectively. We identified at least 1702 unique NLS consensus sequences in the N-terminal region of NAC TFs. The monopartite class I NLS motifs were found to contain more than four consecutive basic amino acids with the number of their consecutive basic amino acids ranging from four to fourteen (K-K-K-K-K-K-K-K-K-K-K-K-K-K-K). The bipartite NLS motifs contain two clusters of consecutive basic amino acids separated by up to twenty-four linker amino acids (K-K-K-x₃-R-x₂-R-x₄-K-x₃-K-x₃-K-x-K-x₂-R-K-K).

The non-canonical NLS motifs contain at least six centrally-located, positively charged amino acids (K-x-R-R-R-P-R-R-x₂-R-K) flanked by positively charged amino acids on both sides. Our analysis of the N-terminal NLS of NAC TFs, however, did not identify any NAC TFs containing this consensus sequence. Instead, several new variants of this consensus sequence were identified with multiple clusters of positively charged amino acids. These NLS were designated as multipartite NLS motifs. A few examples of the multipartite NLS include, K-K-K-K-x₇-K-K-K-K-x₇-K-K-K-K, K-K-K-K-x-K-x₅-K-x-K-K-x₇-K-K-K-K-x₂-K-K-K, K-K-K-x₂-K-K-x-K-x₅-K-x₄-K-K-K-R-x-K-R-K-x-K-x₄-K-K-K-R-K-K, K-K-R-x-R-K-x₂-K-x-K-x₂-K-K-K-x-RK-x₂-K-R-R-x₂-K-K-K-x-R, K-K-R-x-R-K-x₂-K-x-K-x₂-K-K-R-x-R-K-x₂-K-x₂-K-x-K-x-R, K-x₂-K-K-K-x₃-K-K-K-K-K-x-K-x₈-K-x₉-K-x₂-K-K-R-x₂-K-K-K-K-x-K, K-x₂-K-K-x₃-K-x-K-K-K-x-K-K-K-x₂-K-K-K-x-K, R-K-R-x-R-x-R-K-K-x₂-K-x-K-K-K-R-x₂-K-x₂-KK-x₂-R-R-K-x₂-K, and R-K-R-x-R-x-R-x₂-K-x-K-K-K-R-x₂-K-x₄-K-R-x₂-R-R-K-x-K-x₂-R. Much of the diversity of NLS motifs is associated with the sequence of the variable linker amino acids. In our analysis, we removed the linker amino acid sequences, represented as x, to obtain a more concise picture of NLS diversity. Removing the linker amino acids present in monopartite, bipartite, and multipartite NLS motifs resulted in the identification of 97 different NLS consensus sequences in the N-terminal region of NAC TFs. The unique NLS signal sequences were R-K-R-R-K, K-K-K, K-R-K, K-K-R, K-R-R, R-R-R, R-K-K, R-K-R, K-K-K-K, R-K-R-K, R-R-K, R-R-R, K-K-R-K, K-K-R-K-R, K-R-K-R, R-K-R-R-R, R-K-R-R, K-K-K-K-K, R-R-K-R, K-R-K-R-R-K, R-R-K-K, R-R-R-K, K-R-R-K, K-K-R-R, R-K-R-K-R, R-R-R-R-R, K-R-K-K, K-R-R-R, K-R-K-R-R-R, K-K-K-R, R-K-K-K, K-R-K-K-K, K-R-K-R-K, R-K-K-R, K-R-R-R-R, R-R-K-R-R-

K, K-K-K-K-R, K-K-K-R-K, K-K-R-K-K, K-K-K-K-K-K, R-K-K-R-K, R-K-R-K-K, R-K-K-K-K, R-K-R-K-R-K, R-R-K-K-K, R-R-K-R-R-R, R-R-R-R-K, K-K-K-K-K-K-K, K-K-K-K-K-K-K, K-K-K-K-R-R, K-K-R-R-R, K-K-R-R-R, R-R-K-R-K-R, R-R-R-K-K-K, R-R-R-R-R-R, R-R-R-R-R-K, K-K-K-K-K-R, K-K-K-K-K-R-K, K-K-K-R-K-K, K-K-K-R-R-R-R-R, K-R-K-K-K-K, K-R-K-R-R, K-R-R-K-R, R-K-K-K-K-R, R-K-R-K-K-K, R-K-R-R-K-R-K, R-R-K-K-K-K, R-R-R-K-K, R-R-R-K-R, R-R-R-R-K-K, R-R-R-R-R-K, K-K-K-K-K-K-K-K, K-K-K-K-K-K-K-K, K-K-K-K-K-K, K-K-K-K-K-K-K-K-K-K-K-K-K-K-K, K-K-K-K-K-R-R, K-K-K-K-R-K-R, K-K-K-R-K-R, K-K-K-R-R-R-R, K-K-K-R-R-R-R-R-R, K-K-R-R-R-R-R-R-R-R, K-R-K-K-R, K-R-K-R-K-R-K-K, K-R-R-K-K, R-K-K-K-K-K-R, R-K-K-R-K-K-R, R-K-K-R-K-R, R-K-R-K-R-R, R-K-R-K-R-R-K, R-K-R-R-K, R-K-R-R-K-R, R-R-K-K-R, R-R-K-K-R-K, R-R-K-R-K, R-R-R-K-R-R, R-R-R-R-K-K-K, R-R-R-R-K-R, R-R-R-R-R-R-R, and R-R-R-R-R-R-R-R. The R-K-R-R-K consensus sequence was found to be present 347 times, K-K-K 297 times, K-R-K 185 times, K-K-R 165 times, K-R-R 153 times, R-R-R 96 times, R-K-K 95 times, R-K-R 83 times, K-K-K 75 times, R-R-K 74 times, R-R-R-R 58 times, K-K-R-K 49 times, K-K-R-K-R 49 times, and K-R-K-R 40 times. At least 27 NLS amino acid consensus sequences were only found once among the 160 studied species.

The C-terminal end of NAC TF proteins also contain monopartite, bipartite, and multipartite NLS motifs. The multipartite NLS motifs found in the C-terminal end of NAC proteins were R-K-R-x-R-x-R-K-K-x₄-K-x-K-K-K-R-x₃-K-x₃-K-K-x₃-R-R-K-x₂-K, R-R-R-x₄-K-x₆-R-x₂-R-x₂-R-R-x₄-R-R-R-x₆-R-x₂-R-R-x₉-R-R-R-R-R-R-R-x₂-R-R, K-K-K-x₄-K-K-x-K-x₅-K-x₄-K-K-K-R-x-K-R-K-x-K-x₄-K-K-K-R-K-K, K-K-R-x₄-K-x₂-K-x-K-x₂-K-K-R-x-R-K-x₄-K-x₂-K-x-K-K-R-x-R-K-x₄-K-x₂-K-x-K-K-R-x-R-K-x₂-K-x-K-x₂-K-K-K-x-R, K-K-R-x-R-K-x₂-K-x-K-x₂-K-K-R-x-R-K-x₂-K-x-K-x₂-K-K-R, R-K-R-x-R-x₃-K-K-R-R-x₂-K-x₉-K-x₄-R-x-K-x₂-R-x-R-R-x₅-K-K-R, R-K-R-x-R-x-R-x₅-K-x-K-K-K-R-x₃-K-x₄-K-R-x₂-R-R-K, R-R-x-R-R-R-x-R-R-x₈-R-x₆-R-R-x₅-R-R-R-x-R-x₅-R-x₈-R-R-R-R, R-R-x-R-R-x-R-x-R-R-R-x₉-R-x₂-R-R-K-R-K-x-R-x₄-R-R-R-R-R-R-x₄-R-K, R-x-R-R-R-R-x₆-R-x₁₁-R-x₈-R-R-x₃-R-R-R-x₂-R-R-x-R-x-R-x₆-R-R-R-R-R-x₄-R-R-x₂-R, R-x-R-R-x₃-K-R-R-R-x₂-R-x-R-R-x-R-x-R-x₇-R-x₃-R-R-R-x₇-R-x₂-R-R-R-R, R-x-R-x-R-R-R-x₃-R-R-R-x₃-R-x-R-x₂-R-x₄-R-R-R-x₅-R-K-x-R-x₃-R-R-x₁₃-R-R-x-K-x₅-R-R-x₆-K-R-R, and others. Removal of the linker amino acids present in between the consecutive basic amino acids, resulted in the identification of 94 unique consensus sequences. These include K-K-K, K-K-R, R-R-R, K-R-K, K-K-R-K-R, R-K-K, K-K-R-K, R-K-R-K, K-K-K-K, K-R-K-R, K-R-R, R-K-R, R-R-K, R-R-R-R, K-R-K-K, R-K-R-R-K, and others. The NLS consensus sequence K-K-K was identified 144 times, K-K-R 83 times, R-R-R 65 times, K-R-K 60 times, K-K-R-K-R 58 times, R-K-K 47 times, K-K-R-K 45 times, R-K-R-K 40 times, K-K-K-K 39 times, K-R-K-R 37 times, K-R-R 36 times, R-K-R 35 times, R-R-K 31 times, R-R-R-R 24 times, K-R-K-K 17 times, and R-K-R-R-K 17 times. A comparison of the 97 NLS consensus sequence present in N-terminal region with the 94 NLS sequences present in the C-terminal region indicated that 84 NLS consensus sequences were shared between the N-terminal and C-terminal regions. This indicates that there is a close relationship between the NLS sequences in these two regions. An analysis of the unique NLS consensus sequence in the N-and C-terminal regions indicated that 13 NLS consensus sequences were unique to the N-terminal region, namely R-K-R-R-K, R-R-R-R-K, K-K-K-K-R-R, K-R-K-K-K-K, R-R-K-K-K-K, K-K-K-K-K-R-R, K-K-K-R-K-R, R-K-K-R-K-K-R, R-K-R-K-R-R, R-K-R-R-K-R, R-R-K-K-R-K, R-R-

R-K-R-R, and R-R-R-R-K-R. Similarly, nine NLS consensus sequences were unique to the C-terminal region, namely K-K-R-R-K, K-K-K-K-R-K, R-R-K-K-K-R-R-R-R-R-R, K-K-R-K-R-K, K-R-R-R-K, R-K-K-R-K-K, R-K-K-R-R, R-R-K-R-R-R-K, and R-R-R-R. Up to six classes of NLS have been reported to be associated with importin α subunit³¹. To the best of our knowledge, this is the first report describing such a high level of diversity and dynamism in the NLS consensus sequences of NAC TFs and plant transcription factors in general. This is also the first report of the presence of unique NLSs in the N- and C-terminal regions of NAC TFs.

Several nuclear-associated proteins contain NLS, as well as nuclear export signals (NESs). Proteins that perform their function within the nucleus need to be exported out of the nucleus and into the cytoplasm to undergo proteasomal degradation. Therefore, a NES is required in addition to an NLS. A Ran-GTP complex binds directly to an NES and mediates the nuclear export process of cargo molecules³². NES sequences contain a hydrophobic, conserved L-V-F-Y (substitute L-V/I-F-M) motif separated by variable linker amino acids at both ends³³. The presence of an L-V-F-Y motif in all NAC proteins, suggests that all NAC proteins have the potential to be exported out of the nucleus. Hao et al. (2010), however, reported that the hydrophobic L-V-F-Y motif functions as a transcriptional repressor of WRKY, Dof, and APETALA TFs. If the L-V-F-Y motif acts as a transcriptional repressor, then the transcriptional activity of these TFs would be lost; resulting in an unstable genome. Therefore, we suggest that the L-V-F-Y motifs do not function as a transcriptional repressor but rather as a NES.

NAC TFs possess a complex interactome network

The interacting partner of a protein can provide significant information about its potential function and an entire protein-protein interactome network can greatly assist in unravelling the signalling cascade of the proteins. Different cascades are interlinked in signalling systems and form intricate constellations that provide information about cell response and function. Thus, the interactome network of NAC TFs in *A. thaliana* were explored. The presence of a dynamic network was revealed and a diverse set of interacting protein partners of NAC TFs were identified (Figure 3, Table 2). Results indicated that NAC TFs interact with RNS1 (ribonuclease 1), ERD14 (early responsive to dehydration 14), VND1 (vascular related NAC domain 1), VND7, GAI (gibberellins inducible), ZF-HD1 (zinc finger homeodomain 1), TCP8 (Teosinte branched, Cycloidea, and Proliferating cell nuclear antigen factor 8), TCP20, CPL1 (C-terminal domain phosphatase-like1), RHA1A (ring H2 finger H1A), RHA2A, SHR (short root), PHB (phabulosa), PLT2 (plethora 2), MYB59, HB23 (homeobox 23), HB30, NAC1, NAC6, NAC19, NAC32, NAC41, NAC45, NAC50, NAC52, NAC76, NAC83, NAC97, NAC101, NAC105, IAA14 (auxin responsive protein indole3-acetic acid), HAI1 (protein phosphatase), ABI1 (ABA insensitive 1), RVE2 (reveille), PYL4 (PYR-like 4), BRM (brahma), HB52, RCD1 (radical-induced cell death 1), JMJ14 (jumonji 14), TPL (TOPELESS), F2P16 (TOPELESS related), TOPELESS, RING/U-box, ZF-domain (zinc finger domain), SRO1 (similar to RCD 1), CUC2 (cup shaped cotyledon 2), PAS1 (pasticcino 1), TI1 (defensin like 1), TSPO (tryptophan rich sensory protein), TIP2.2 (TCV-interacting protein 2.2), TIP3.1, T21F11.18 (TOPELESS related), LRR (leucine rich repeat), RPA2 (replicon protein A2), and VR-NAC (Table 2). The interaction of NAC TF proteins with the diverse number of listed proteins has been experimentally validated in *A. thaliana*. The interactome network includes stress responsive proteins, other transcription factors, hormonal signalling proteins, protein

phosphatases, and defense related proteins. In addition to experimentally-validated interacting proteins, bioinformatic mining indicated that NAC TFs also interact with several other proteins (Table 2). Some of the identified interacting proteins were MYB, NAC, NTL (NTL2-like), UBC30 (ubiquitin conjugating enzyme 30), ATM (Ataxia-Telangiectasia mutated), ATR (serine/threonine kinase ATR), KNAT (knox tail), AOX1A (alternative oxidase 1A), ASG2 (altered seed germination 2), NYE (nonyellowing), CPL, TMO6 (target of monopteros 6), RHA2A (ring H2-finger A2A), XCP (xylem cysteine peptidase), DBP (downstream auxin binding), WAK5 (wall associated kinase 5), RCD, CYP71A25 (cytochrome P71A25), Stay green, ERF (ethylene responsive transcription factor), PPR (pentatricopeptide), WOX (Wuschel related homeobox), PPD6 (psbP-domain protein 6), MFDX (mitochondrial ferredoxin), LEA (late embryogenesis abundant), IRX (irregular xylem), CESA4 (cellulose synthase A4), SCRL20 (SCR-like 20), PIP1-5 (aquaporin), PUP4 (purine permease 4), XERO1 (dehydrin xero 1), SWAP (suppression of white apricot), TIR-NBS, NBS-LRR, PAS, CHI (chitinase), MC5 (metacaspase 5), XCP (xylem cysteine peptidase), RNS, LAC (laccase), TIR-NBS (toll/interleukin receptor-nucleotide binding site), NBS-LRR, DTA4 (downstream target 4), BAG6 (BCL-2 associated anthogene 6), and others as well (Table 2). Some NAC TFs are co-expressed with other proteins. These include DREB2A, XCP1, XCP2, ATM, ATR, MYB63, MYB69, MYB83, IRX1, AOX1A, RCD1, ASG2, ERD1, TMO6, DOF6, SHR, PLT2GRP20, CYP86C4, VND7, NAC6, NAC32, NAC97, NAC19, NAC102, HAI1, WRKY33, WRKY46, WRKY53 and others (Table 2). Some of the NAC TFs directly interact with the interacting partner while others form complexes and appear to play an indirect role. NAC TFs act as a negative regulator of ABA signalling, while they induce JA/ET-associated marker genes³⁴.

The expression of several of NAC genes are either up- or down-regulated by auxin, ethylene, or ABA, suggesting that NAC TFs play a role in plant hormonal signalling³⁵⁻³⁷. One of the most challenging aspects of a protein-protein interactome network is that the interaction can vary depending upon the cell and its environment³⁸. Therefore, it is necessary to investigate the dynamic interactions of proteins in different cells and environmental conditions to completely understand their interacting partner and the cellular function of the TF. NAC TFs regulate *ERD* and *NCED* (ABA biosynthesis) genes through a direct interaction with their promoters^{39,40}. NAC TFs (ANAC019, ANAC055, and ANAC072) interact with ERD1 which encodes a Clp protease regulatory subunit⁴¹. The over expression of one of these three NAC TFs, however, did not induce the up-regulation of ERD1 because the induction of ERD1 depends on the co-expression of a zinc finger homeodomain TF, ZFHD1⁴¹. ANAC019 and ANAC055 interact with ABI (abscisic acid insensitive), and at least five MYB TFs can bind to the NAC TF promoter region^{42,43}. In this case, the NAC DNA binding domain mediates the interaction with RHA2A and ZFHD1⁴³.

NAC TFs encodes chimeric proteins and contain multiple binding sites

NAC TFs are characterised by the presence of a DNA binding domain. Several NAC TFs, however, contain more than one NAC domain. Chimeric NAC TFs have also been identified. At least 45 variants of chimeric NAC TFs were identified in our analysis (Figure 4). Several of the NAC TFs were also found to possess as many as three or four NAC DNA binding domains. Furthermore, the NAC domains were found to be associated with PPR (pentatricopeptide), protein

kinase, PI3_4_kinase_3, EF-hands (elongation factor), CRM, peptidase A1, WRKY, cytochrome B561, OFOF, FFO, Dna_J2, ZF_B, TIR, LRR, CS, F-box, IQ, PPC, ENT, ABC_TM1F, RWP_RK, PB1, PABC, ACT, INTEGRA, RESPO, JMJC, SAM, BRX, G_TR_2, RORP, CHCH, TPR, YJEF_N, HTH, HOMEEO, GH16, ANK_REPEAT_REGION, Peroxidase, LONGIN, V_SNA, RECA_2, KH_TY, APAG, RRM, carrier, and a DCO domain. At least four NAC TFs from *A. thaliana*, ten from *B. napus*, four from *B. rapa*, two from *M. domestica*, four from *P. virgatum*, 17 from *C. sativa*, eight from *D. oligosanthos*, eight from *E. tef*, and five from *L. perrieri* were found to possess 2 NAC domains (Supplementary Table 1). NAC TFs in several other species were also found to contain two NAC domains (Supplementary Table1). When two NAC domains were present, both domains were located towards the N-terminal end. NAC TFs of at least three species, *O. rufipogon*, *B. stacei*, and *Camelina sativa* were found to possess three NAC domains whereas the NAC TFs in *A. lyrata* (gene id: 338342), *C. sativa* (Csa16g052260.1), and *E. tef* (462951506) were found to possess four NAC domains (Figure 4).

Other chimeric domains were also identified in different regions of the NAC protein. PPR domains were found in both the N-terminal or C-terminal region, a protein kinase domain was found upstream to the NAC domain, and a NAC domain was found to be adjacent to a protein kinase and EF-hand domain. Additionally, a protein kinase domain was found to be followed by either a NAC and a CRM domain, a NAC domain was followed by a peptidase_A1 domain, a NAC domain was followed by the presence of a WRKY domain, a cytochrome_B561 domain was followed by either a NAC domain and a CRM domain, a DFDF and a FFD domain were followed by a cytochrome B and NAC domain, a DNAJ_2 domain was followed by a NAC domain, a DNAJ_2 domain was followed by a NAC and ZF_B domain, a NAC domain was followed by TIR, LRR and CS domains, a NAC domain was followed by a TIR domain, an F-box domain was followed by a NAC domain, an IQ domain was followed by a NAC domain, a NAC domain was followed by a ZF_B domain, an EF-hand domain was followed by a NAC domain, a NAC domain was followed by a PPC domain, an ENT domain was followed by a NAC domain, a NAC domain was followed by an ABC_TM1F, a NAC domain was followed by CRM domain, a NAC domain was followed by a RWP_RK and a PB1 domain, a NAC domain was followed by three ACT domains, a NAC domain was followed by a PABC domain, a NAC domain was followed by an INTEGRA domain, a RESPO domain was followed by a NAC domain, a NAC domain was followed by a JMJC and a JMJC domain, a SAM domain was followed by a NAC domain, a BRX domain was followed by a NAC domain, a NAC domain was followed by ZF_B, NAC and ZF_B domains, an F-box and protein kinase domain was followed by a NAC domain, a NAC domain was followed by a G_TR_2 domain, an RDRP domain was followed by a NAC domain, a NAC domain was followed by a CHCH domain, a TPR domain followed by a NAC domain, an F-box domain followed by NAC and F-box domains, a NAC domain was followed by a YJEF_N domain, a NAC domain was followed by a HTH domain, a homeobox domain was followed by a NAC domain, a NAC domain was followed by three GH16_2 domains, an ANK repeat region was followed by a NAC domain, a NAC domain was followed by a peroxidase domain, a NAC domain was followed by a LONGIN and V_SNA domain, a NAC domain was followed by RECA_2 and RECA_3 domains, a KH domain was followed by a NAC domain, a NAC domain was followed by a RAB domain, a JMJC domain was followed by a NAC domain, a NAC domain was followed by an APAG domain, an RRM domain was followed by a NAC domain, a carrier

domain was followed by a NAC domain, and a NAC domain was followed by DCO domain (Figure 4).

The presence of chimeric domains within NAC TFs is of particular interest, especially for understanding why they are there and how they impact the function of a specific NAC TF. The most common domains, such as PPR, TIR, WRKY, protein kinase, ZF_B, EF-hands, cytochrome B, DNAJ, F-box, peroxidase, and GH16 are involved in diverse cellular processes, including transcriptional regulation of plant development and stress response⁴⁴⁻⁵². The association of a TIR domain with an NBS-LRR domain is an example of the association of TF domains with other domains to form chimeric proteins⁵³. The presence of different domains with the NAC domain could potentially enable the NAC domain to assist in the function of the associated domains and vice versa. For example, NAC TFs could have the potential to regulate peroxidase by possessing a peroxidase domain within the NAC TF, instead of regulating it separately with another TF. The presence of multiple domains can enable the co-regulation of diverse functional sites within the NAC TFs. The presence of chimeric TFs has been recently reported in WRKY TFs as well^{54,55}. Therefore, the presence of chimeric domains in NAC TFs can impart a significant dynamic aspect to the ability of NAC TFs to regulate gene expression.

In addition to the presence of multiple chimeric domains, NAC TFs were also found to contain diverse active/binding motifs for several other proteins. It is possible that NAC TFs may play a dual role as a transcription factor and as an enzyme. At least 404 NAC TFs were found to possess other functional motifs comprising 101 unique functional sequences (Supplementary Table 2). In addition to NAC domains, the other function sequences included a Fe-2S ferredoxin-type iron-sulfur binding region signature, 2-oxo acid dehydrogenase acyltransferase component lipoyl binding site, 4Fe-4S ferredoxin-type iron-sulfur binding domain profile, 7,8-dihydro-6-hydroxymethylepterin-pyrophosphokinase signature (30), ABC transporter family signature (2), adenosine and AMP deaminase signature (2), adipokinetic hormone family signature, aldehyde dehydrogenase cysteine active site (4), aldehyde dehydrogenase glutamic acid active site (28), aldo/keto reductase family putative active site signature (4), alkaline phosphatase active site (5), aminoacyl-transfer RNA synthetase class-I signature, aminotransferase class-II pyridoxal-phosphate attachment site (15), antenna complexes beta subunit signature (2), ArgE/dapE/ACY1/CPG2/yscS family signature 1 (2), aspartate and glutamate racemases signature 1 (3), aspartokinase signature (2), ATP binding site and proton acceptor, ATP synthase alpha and beta subunit signature (19), ATP dependent DNA ligase AMP-binding site (2), bacterial regulatory proteins araC family signature, beta-ketoacyl synthases active site (2), C-5 cytosine-specific DNA methylases active site, cadherin domain signature, carbamoyl-phosphate synthase subdomain signature 2, cysteine protease inhibitor signature (19), cytochrome p450 cysteine heme-iron ligand signature (7), endopeptidase Clp serine active site (3), eukaryotic and viral aspartyl proteases active site (10), FGGY family of carbohydrate kinase signature 2, fumarate lyases signature, GHMP kinases putative ATP-binding domain, glucoamylase active site region signature, glyceraldehyde 3-phosphate dehydrogenase active site, glycoprotease family signature, glycohydrolase family 5 signature, glycosyl hydrolase family 9 active site signature 2, heavy metal associated domain, hemopexin domain signature (2), histone H4 signature (4), HMG-I and HMG-Y DNA binding domain (A+T hook) (9), immunoglobulins and major histocompatibility complex protein signature (4), inorganic pyrophosphate signature (13), iron-containing alcohol

dehydrogenase signature 1, legume lectins beta-chain signature (15), lipocalin signature (23), mannitol dehydrogenase signature, N-6 adenine-specific DNA methylases signature (7), neutral zinc metallopeptidase, zinc binding region signature, Nt-DnaJ domain signature (2), peroxidase active site signature, pfkB family of carbohydrate kinases signature 1 and 2 (4), phospholipase A2 histidine active site (5), phosphopantetheine attachment site (17), polygalacturonase active site (2), polyprenyl synthases signature, PPM-type phosphatase domain signature, prokaryotic membrane lipoprotein lipid attachment site, putative AMP binding domain signature, regulator of chromosome condensation (RCC1) signature 2 (2), ribosomal protein L24e signature (7), ribosome binding factor A signature, rubredoxin signature (2), serine protease, subtilase family aspartic acid active site (21), serine protease, trypsin family, serine active site, serine/threonine protein kinase active-site signature (3), sigma-54 interaction domain ATP-binding site A signature, signal peptidase I serine active site (3), signal peptidase I signature 3 (4), soybean trypsin inhibitor (Kunitz) protease inhibitor family signature, SRP54-type proteins GTP-binding domain signature, sugar transport proteins signature 2, synaptobrevin signature, syntaxin/epimorphin family signature, TonB-dependent receptor proteins signature 1 (7), translationally controlled tumor protein (TCTP) domain signature 2, Trp-Asp (WD) repeats signature (12), tubulin subunit alpha, beta, and gamma signature (2), tubulin-beta mRNA autoregulation signal (2), zinc carboxypeptidases, zinc binding region 2 signature (11), zinc finger BED-type profile, zinc finger C2H2 type domain signature, and a zinc-containing alcohol dehydrogenase signature (2) (Supplementary Table 2). This is the first study to report the presence of such a diverse number of functional sites and signature motifs in NAC TFs. Although the majority of the functional domains are associated with a specific function in plants, the presence of a histocompatibility complex and a translationally controlled tumor protein (TCTP) sequence are of particular interest. These proteins are specifically found in animal systems and the histocompatibility complex is the major contributing factor regulating the binding of antigens. More specifically, TCTP is a highly conserved protein that is involved in microtubule stabilization, calcium binding, and apoptosis and is associated with the early growth phase of tumors ⁵⁶. The presence of MHC and TCTP in association with NAC domains suggests that this combination may be playing a crucial role in the plant immune system and in uncontrolled cell growth. The presence of diverse functional sites in NAC TFs indicates that NAC TFs are involved in diverse cellular functions and metabolic pathways. This statement is supported by the large number of NAC TFs that are present in plant genomes.

NAC TFs are involved in diverse cellular processes

NAC TFs are known to possess diverse chimeric domains, as a result, it is more than likely that NAC TFs are also involved in the regulation of diverse cellular pathways and cellular processes. To help substantiate this premise, the interactome associated with NAC TFs in *A. thaliana* was analysed. Results indicated that NAC TFs are potentially involved in a least 289 different cellular processes and pathways (Supplementary Table 3). The majority are related to cell, tissue, and organ (root, stem, meristem) development, as well as signalling processes. Several NAC TFs also appear to be associated with phytohormone signalling, including auxin, gibberellin, jasmonic acid, and salicylic acid signalling pathways. NAC TFs were also found to be associated with pathways involved in the response to bacterial, fungal, UV, heat and other biotic and abiotic stresses (Supplementary Table 3). At least 202 genes in the NAC TF interactome network were found to

be associated with pathways related to the nucleus, 239 were associated with intracellular membranes, and 241 were associated with intracellular organelles, 20 with the endoplasmic reticulum, and 3 with the nuclear matrix. If the association is designated based on the description of a pathway, 127 genes were found to be associated with transcription factor activity and sequence-specific DNA binding, 143 with DNA binding, 146 with nucleic acid binding, 220 with organic cyclic compound binding, 220 with heterocyclic compound binding, 65 with ATP binding, 49 with macromolecular complex binding, 48 with chromatin binding, 35 with ADP binding, 25 with sequence-specific DNA binding, 18 with transcription regulatory region binding, 8 with structural constituents of the cell wall, 11 with auxin transport activity, 2 with LRR binding, and 2 with bHLH transcription factor binding. These data clearly indicate that NAC TFs are involved in diverse cellular processes. The identification of LRR protein in the pathway description of NAC TFs agrees with the presence of an LRR domain in a chimeric NAC domain of NAC TFs.

NAC TFs are expressed in a spatiotemporal manner

Patterns of NAC TF gene expression were analysed in leaf and root tissues of *A. thaliana* treated with ammonia, nitrate, or urea (Figure 5 and Figure 6). Among a total of 120 NAC TFs, 95, 97, and 98 were differentially expressed in leaf tissue treated with ammonia, nitrate, or urea, respectively. Leaf tissues treated with ammonia, nitrate and urea exhibited 70.14, 117.11, and 58.35 FPKM expression values for *AtNAC1* (AT1G01010.1), *AtNAC4* (AT1G02230.1), and *AtNAC1* (AT1G01010.1), respectively. At least 46 genes in leaves exhibited expression of more than one FPKM in response to ammonia, 54 in response to nitrate, and 44 in response to urea. *AtNAC1* was highly expressed in ammonia and urea treated leaves. At least 24, 26, and 25 NAC TFs did not exhibit any expression in leaf tissues treated with ammonia, nitrate, or urea.

Relative to leaf tissues, the expression of NAC TFs in root tissues was more dynamic. Root tissue treated with urea exhibited the highest expression of NAC TFs relative to leaves treated with ammonia or nitrate (Figure 6). The number of *AtNAC* TFs whose expression was one or more FPKM in response to ammonia, nitrate, or urea were 75, 71, and 70, respectively. *AtNAC8* (AT5G08790.1) was highly expressed in ammonia-treated roots, whereas, *AtNAC91* (AT5G24590.2) was highly expressed in nitrate- and urea-treated roots. Urea, ammonia and nitrate (UAN) commonly serve as a source of nitrogen (N) for plants. Analysis of the levels of gene expression indicate that ammonia and nitrate modulate the expression of NAC TFs more than urea. A study utilizing *Pinus taeda* revealed that fertilization with ammonium, nitrate, or urea produces different effects on growth and drought tolerance⁵⁷. Results of the current analysis indicate that *AtNAC8* and *AtNAC91* are the major NAC TFs involved in nitrogen assimilation during plant growth.

Codon usage in NAC TF is dynamic

Codon usage bias in NAC TFs of the examined species were studied. separately. Among 61 sense codons, only 14 were found in the all species. These included AAG (K), ACU (R), AGA (R), AGG (R), UCU (S), AUC (I), AUG (M), CAA (Q), CCU (P), GAA (E), GCU (A), GGA (G), UGG (O), and UUC (F) (Table 3). The most abundant codon was UCU (S), which was found 30 times in in *Humulus lupulus* NAC TFs (Table 3). The codons CGA (R), CGC (R), CGG (R), CGU

(R) were absent in 127 of the 160 examined species. ACG (T), UCG (S), CAG (Q), CAC (H), CCA (P), CCC (P), CCG (P), and GCG (A) were absent in 126 of the examined species. The highest relative synonymous codon usage bias (RSCU) was found to be 1.35, 1.23, 1.29 for the codon AAA (K) in *Ocimum tenuifolium*, *Picea sitchensis*, and *Ipomea trifida*. Synonymous codon-usage was not observed in NAC TFs. Relative codon usage is determined by dividing the ratio of observed frequency of codons by the expected frequency, provided that all of the synonymous codons for the same amino acids are used equally. Relative Synonymous Codon Usage (RSCU), however, is not related to the usage of amino acids. An RSCU > 1 indicates the occurrence of codons more frequently than expected, while an RSCU < 1 indicates that the codon occurs less frequently than expected^{58,59}. Non-synonymous substitution in organisms is subject to natural selection^{60,61}. Genes with lower non-synonymous selection leads to functional diversity of a gene. The presence of a low level of nonsynonymous codon usage in NAC TFs indicates that they are functional and have evolved from paralogous ancestors.

Rate of transition of NAC TFs is higher than the rate of transversion

Nucleotide mutation is an integral part of the evolution of a genome and leads to the acquisition of required traits and the elimination of detrimental traits from the genome. It is a regular process and hundreds of thousands of nucleotides have undergone addition or deletion events in the evolution of a genome. The alteration or conversion of a nucleotide occurs either through a transition or a transversion. A transition event involves the interchange of two-ring purines (A and G) or of one-ring pyrimidines (C and T). Transversion events the exchange of a purine for a pyrimidine or vice versa. The rate at which these two events occur is important to understanding of the evolution of a gene. Therefore, the rate of nucleotide substitution in NAC TFs was analysed. Results indicated that the rate of transition in NAC TFs is higher than the rate of transversion. The substitution of adenine with guanine was found to be highest in *Linum usitatissimum* (15.82), while the substitution of guanine to adenine was found to be the highest in *Lotus japonicas* (19.07). The lowest rate of substitution from adenine to guanine and vice versa was found in *Trifolium pratense* (9.73) and *Amborella trichopoda* (10.8), respectively (Table 4). The highest rate of substitution from thiamine to cytosine and vice versa was found in *Klebsormidium flaccidum* (7.19) and *Pseudotsuga menziesii* (11.59), respectively. The lowest rate of substitutions from thiamine to cytosine and vice versa was found in *Capsella grandiflora* (2.41) and *Cicer arietinum* (1.62), respectively (Table 4). These data make it evident that the rates of transition of purine (adenine and guanine) nucleotides are higher than the rates of pyrimidines. The highest rate of transversion from adenine to thiamine and vice versa was found in *Capsella grandiflora* (12.34 for adenine to thiamine and 9.91 for thiamine to adenine) (Table 4). The rate of substitution by transversion is slower relative to the rate of substitution by transition.

Capsella grandiflora is a close relative of *Arabidopsis thaliana* and is predicted to be the progenitor of *Capsella bursa-pastoris*. *Capsella grandiflora* is a self-pollinating plant and is used as a model organism in evolutionary studies and the change from self-incompatibility into self-compatibility. The genomic consequences of the evolution of selfing, however, is poorly understood. *Capsella rubella*, a close relative of *Capsella grandiflora*, that evolved self-compatibility 200,000 years ago⁶² also exhibits a high rate of transversion from adenine to thiamine (11.19). Thus, the higher rate of transversion from adenine to thiamine in *Capsella*

grandiflora and *Capsella rubella* may be a possible factor in the evolution of self-pollination. Higher rates of transversion were also found in *Solanum pimpinellifolium* (11.4) and *Castanea mollissima* ((11.31) Chinese chestnut). *Solanum pimpinellifolium* is self-pollinating and exhibits high levels of stress tolerance⁶³. *Castanea mollissima* has evolved over a period of time in coexistence with chestnut blight and is resistant to the pathogen. This indicates that higher rates of transversion from adenine to thiamine and vice versa are associated with self-pollination and stress tolerance in plants. The highest rate of substitution from guanine to cytosine and vice versa was found in *Arachis hypogaea* (11.07), and *Camelina sativa* (11.46), respectively (Table 4). The lowest rate of substitution from adenine to thiamine and vice versa was found in *Linum usitatissimum* (3.72) and *Klebsormidium flaccidum* (6.67), respectively. Notably, the highest rate of substitution from thiamine to cytosine was found in *Klebsormidium flaccidum* and the highest rate of substitution from adenine to guanine was found in *Linum usitatissimum*. This indicates that organisms which exhibit the highest rate of transition possess the lowest rate of transversion.

NAC TFs evolved from orthologous ancestors

A phylogenetic tree of NAC TFs was constructed to understand their evolutionary relationships. A model selection was conducted before constructing the phylogenetic tree using the maximum likelihood statistical method. The phylogenetic tree revealed the presence of at least seven phylogenetic clustered orthologous (COGs) groups originating from a common, orthologous ancestor (Figure 7). Each phylogenetic cluster was further divided into two or more sub-groups. A phylogenetic tree of each individual species was subsequently constructed to examine duplication and loss events in NAC TFs. The phylogenetic tree of each species was independently reconciled with the collective species tree. This analysis indicated that NAC TFs in all of the species were duplicated and no NAC TFs were found to be lost. This suggest that NAC TFs evolved from common ancestors (orthology) and underwent numerous duplication events during speciation (paralogy), which gave rise to diverse gene functions in plant development and growth. We also checked for the presence of potential foreign or homologous sequences (xenologs) in NAC TFs. No primary xenologs, sibling donor xenologs, sibling recipient xenologs, incompatible xenologs, autoxenologs, or paraxenologs were identified in NAC TFs. Although the phylogenetic tree indicates the evolution NAC TFs from common ancestors, none of the NAC genes in the examined species were found to have been transferred from one species to another (Table 1). Previous studies of NAC TFs in six plant species also reported a high level of duplication and divergent evolution⁶⁴. The expansion of TF families was associated with an increase in the structural complexity of the organism⁶⁵. Previous studies reported the lineage-specific grouping of transcription factors^{54,64}. The phylogenetic tree of NAC TFs also revealed the presence of lineage-specific clustering as well. In a few cases, however, order-specific clustering of NAC TFs was also observed. For example, NAC TFs in dicot species of the Brassica lineage, including *A. thaliana*, *A. halleri*, *B. napus*, *B. rapa*, *R. sativus*, *R. raphanistrum*, *C. rubella*, *A. alpine*, and others, grouped together. Similarly, NAC TFs in monocot plant species, including *O. sativa*, *O. nivara*, *B. distachyon*, and others, also grouped together.

Conclusion

NAC TFs are present in higher plants, as well as in a few species of algae. The number of NAC TFs per genome and their structural and functional properties increased with the complexity of the organism. The algae *Klebsormidium flaccidum*, a charophyte, was also found to possess NAC TFs; suggesting that the evolution of NAC TFs was associated with the adaptation of plant life from an aquatic to a terrestrial form. The paralogous evolution of NAC TFs underlies their diverse functional role in plant growth and development. Duplication events in NAC TFs were greater than deletion events and the absence of any loss of NAC TFs in different plant species indicates their evolution in recent times. As NAC TFs play a pivotal role within the nucleus regulating gene expression, the presence of bipartite and multipartite nuclear localization signals is of particular interest and provides the basis for further investigation of their functional roles.

Materials and Methods

Identification of NAC TFs

NAC genes in the studied plant species were obtained from searches in the National Centre for Biotechnology Information (NCBI), Phytozome, and Plant Genome databases⁶⁶. BLASTP and hidden Markov model were used to identify the NAC TFs in different species using AtNAC1 and AtBAC2 as the query sequences⁶⁷. Protein and CDS sequences of each species were collected and further analysed. Protein sequences of the NAC TFs were subjected to BLASTP analysis against a reference database to reconfirm them as a NAC TF of the identified species. All of the NAC TF protein sequences in the examined species were also subjected to ScanProsite and InterPro scans to confirm the presence of a NAC domain^{68,69}. Sequences that were found to contain a NAC domain were considered as NAC TFs. The presence of multiple NAC domains, along with the presence of chimeric NAC domains, were determined through ScanProsite and InterPro scans. The presence of multiple functional sites in NAC TFs were also analysed using ScanProsite software.

Analysis of membrane attachment and nuclear localization signal sequences

The presence of transmembrane domains in NAC TFs of all of the examined species were identified using the TMHMM server v. 2.0⁷⁰ and default parameters. Nuclear localization signal sequences in NAC TFs were identified using NLStradamus software, which uses a hidden Markov model for the prediction of nuclear localization signals⁷¹. NAC TF protein sequences were uploaded in FASTA format to run the program. The parameters used to run the NLS analysis were; HMM state emission and transition frequencies, 2 state HMM static; prediction type Viterbi and posterior, prediction cut-off 0.4; prediction display, and image and graphic.

Interactome analysis of NAC TFs

A. thaliana NAC TFs were used to examine the complex interactome network of NAC TFs. The individual interaction network of each NAC TF in *A. thaliana* was searched in a string database that contains 9.6 million proteins from 2031 organisms^{72,73}. The interactome network of each of

NAC TF were noted and the results were later used to construct the interactome network of *A. thaliana* NAC TFs. The presented interactome network was based on an experimentally validated network, co-expressed network, and a mined network. These outputs were used to construct the interactome network. The NAC TFs used to construct the interactome network were subjected to GO (gene ontology) and cellular process analyses.

Gene expression analysis

Differential gene expression of NAC TFs was analysed to elucidate their role in growth, development, and nitrogen assimilation. *A. thaliana* NAC TFs were used to examine differential gene expression. Transcriptome data from *A. thaliana* treated with ammonia, nitrate, and urea were utilized from the PhytoMine database in Phytozome. The expression pattern of NAC TFs for leaf and root tissues in the treated *A. thaliana* plants were analysed separately. The expression was measured in fragments per kilobase of exon per million fragments mapped (FPKM). Transcripts with a zero value were discarded from the study.

Construction of a phylogenetic tree

Two approaches were used to construct phylogenetic trees. In the first approach, a phylogenetic tree was constructed using the NAC TFs of individual species. In the second approach, the NAC TFs of all of the examined species were used to construct a phylogenetic tree. The phylogenetic tree for individual species was constructed to determine the deletion and duplication events in NAC TFs within individual species. Prior to construction of the phylogenetic trees, a model selection was carried out in MEGA6 software. The following parameters were used in the model, analysis, model selection; tree to use, automatic (neighbor joining), statistical method, maximum likelihood; substitution type, nucleotides; gaps/missing data treatment, partial deletion; site coverage cut-off (%), 95; codons included, 1st+2nd+3rd+non-coding. Based on the lowest BIC values of model selection, phylogenetic trees of NAC TFs were carried out using the neighbor joining method, a GTR statistical model, and 1000 bootstrap replicates.

Analysis of transition and transversion rates

Transition and transversion rates in NAC TFs within individual species were analysed using MEGA6 software. The converted MEGA file format of individual species was used to determine the rate of transition and transversion. The following statistical parameters were used to study the transition/transversion rate: estimate transition/transversion bias; maximum composite likelihood estimates of the pattern of nucleotide substitution; substitution type, nucleotides; model/method, Tamura-Nei; gaps/missing data treatment, pairwise deletion; codon position, 1st, 2nd, 3rd, and non-coding sites.

Analysis of gene deletion and duplication

Prior to the analysis of deletion and duplication events in NAC TFs, a species tree was constructed in the NCBI taxonomy browser (<https://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi>). All of the studied species were used to construct the species tree. The resulting phylogenetic trees of individual species in a nwk file format were uploaded in Notung2.9 software as a gene tree and reconciled as a gene tree

with the species tree to obtain duplicated and deleted genes. Deletion and duplication events were analysed in all of the studied species individually.

Data availability

All the data used during this study was taken from publicly available genomic databases and details are mentioned in the materials and methods section.

Competing of interest

Authors don't have any competing of interest to declare.

References

1. Puranik, S., Sahu, P. P., Srivastava, P. S. & Prasad, M. NAC proteins: Regulation and role in stress tolerance. *Trends Plant Sci.* **17**, 369–381 (2012).
2. Ernst, H. A., Nina Olsen, A., Skriver, K., Larsen, S. & Lo Leggio, L. Structure of the conserved domain of ANAC, a member of the NAC family of transcription factors. *EMBO Rep.* **5**, 297–303 (2004).
3. Jensen, M. *et al.* The Arabidopsis thaliana NAC transcription factor family: structure-function relationships and determinants of ANAC019 stress signalling. *Biochem. J.* **426**, 183–196 (2009).
4. Chen, Q., Wang, Q., Xiong, L. & Lou, Z. A structural view of the conserved domain of rice stress-responsive NAC1. *Protein Cell* **2**, 55–63 (2011).
5. Duval, M., Hsieh, T.-F., Kim, S. Y. & Thomas, T. L. Molecular characterization of AtNAM: a member of the Arabidopsis NAC domain superfamily. *Plant Mol. Biol.* **50**, 237–248 (2002).
6. Olsen, A. N., Ernst, H. A., Leggio, L. Lo & Skriver, K. NAC transcription factors: structurally distinct, functionally diverse. *Trends Plant Sci.* **10**, 79–87 (2005).
7. Le, D. T. *et al.* Genome-Wide Survey and Expression Analysis of the Plant-Specific NAC Transcription Factor Family in Soybean During Development and Dehydration Stress. *DNA Res. An Int. J. Rapid Publ. Reports Genes Genomes* **18**, 263–276 (2011).
8. Olsen, A. N., Ernst, H. A., Leggio, L. Lo & Skriver, K. DNA-binding specificity and molecular functions of NAC transcription factors. *Plant Sci.* **169**, 785–797 (2005).
9. Greve, K., La Cour, T., Jensen, M. K., Poulsen, F. M. & Skriver, K. Interactions between plant RING-H2 and plant-specific NAC (NAM/ATAF1/2/CUC2) proteins: RING-H2 molecular specificity and cellular localization. *Biochem. J.* **371**, 97–108 (2003).
10. Hao, Y.-J. *et al.* Plant NAC-type transcription factor proteins contain a NARD domain for repression of transcriptional activation. *Planta* **232**, 1033–1043 (2010).
11. Fang, Y., You, J., Xie, K., Xie, W. & Xiong, L. Systematic sequence analysis and identification of tissue-specific or stress-responsive genes of NAC transcription factor family in rice. *Mol. Genet. Genomics* **280**, 547–563 (2008).
12. Yamaguchi, M. *et al.* VND-INTERACTING2, a NAC Domain Transcription Factor, Negatively Regulates Xylem Vessel Formation in Arabidopsis. *Plant Cell* **22**, 1249–1263 (2010).
13. Ho, S. K. *et al.* Identification of a calmodulin-binding NAC protein as a transcriptional repressor in Arabidopsis. *J. Biol. Chem.* **282**, 36292–36302 (2007).
14. Christian, D. *et al.* The transcription factor ATAF2 represses the expression of pathogenesis-related genes in Arabidopsis. *Plant J.* **43**, 745–757 (2005).
15. Shen, H., Yin, Y., Chen, F., Xu, Y. & Dixon, R. A. A Bioinformatic Analysis of NAC Genes for Plant Cell Wall Development in Relation to Lignocellulosic Bioenergy

- 770 Production. *BioEnergy Res.* **2**, 217 (2009).
- 771 16. Mohanta, T., Mohanta, N., Mohanta, Y. & Bae, H. Genome-Wide Identification of Calcium
772 Dependent Protein Kinase Gene Family in Plant Lineage Shows Presence of Novel D-x-D
773 and D-E-L Motifs in EF-Hand Domain. *Front. Plant Sci.* **6**, 1146 (2015).
- 774 17. Welner, D. H. *et al.* DNA binding by the plant-specific NAC transcription factors in crystal
775 and solution: a firm link to WRKY and GCM transcription factors. *Biochem. J.* **444**, 395–
776 404 (2012).
- 777 18. Shaw, K. L., Grimsley, G. R., Yakovlev, G. I., Makarov, A. A. & Pace, C. N. The effect of
778 net charge on the solubility, activity, and stability of ribonuclease Sa. *Protein Sci.* **10**,
779 1206–1215 (2001).
- 780 19. Pergande, M. R. & Cologna, S. M. Isoelectric Point Separations of Peptides and Proteins.
781 *Proteomes* **5**, 4 (2017).
- 782 20. Cunningham, J. *et al.* Intracellular Electric Field and pH Optimize Protein Localization and
783 Movement. *PLoS One* **7**, e36894 (2012).
- 784 21. Hoppe, T., Rape, M. & Jentsch, S. Membrane-bound transcription factors: regulated release
785 by RIP or RUP. *Curr. Opin. Cell Biol.* **13**, 344–348 (2001).
- 786 22. Vik, Å. & Rine, J. Membrane biology: Membrane-regulated transcription. *Curr. Biol.* **10**,
787 R869–R871 (2000).
- 788 23. Hoppe, T. *et al.* Activation of a Membrane-Bound Transcription Factor by Regulated
789 Ubiquitin/Proteasome-Dependent Processing. *Cell* **102**, 577–586 (2000).
- 790 24. Seo, P. J., Kim, S. G. & Park, C. M. Membrane-bound transcription factors in plants.
791 *Trends Plant Sci.* **13**, 550–556 (2008).
- 792 25. Shen, J., Chen, X., Hendershot, L. & Prywes, R. ER Stress Regulation of ATF6
793 Localization by Dissociation of BiP/GRP78 Binding and Unmasking of Golgi Localization
794 Signals. *Dev. Cell* **3**, 99–111 (2002).
- 795 26. Liu, J.-X., Srivastava, R., Che, P. & Howell, S. H. An Endoplasmic Reticulum Stress
796 Response in Arabidopsis Is Mediated by Proteolytic Processing and Nuclear Relocation of
797 a Membrane-Associated Transcription Factor, bZIP28. *Plant Cell* **19**, 4111–4119 (2007).
- 798 27. Liu, J.-X., Srivastava, R., Che, P. & Howell, S. H. Salt stress responses in Arabidopsis
799 utilize a signal transduction pathway related to endoplasmic reticulum stress signaling.
800 *Plant J.* **51**, 897–909 (2007).
- 801 28. Iwata, Y. & Koizumi, N. An Arabidopsis transcription factor, AtbZIP60, regulates the
802 endoplasmic reticulum stress response in a manner unique to plants. *Proc. Natl. Acad. Sci.*
803 *U. S. A.* **102**, 5280–5285 (2005).
- 804 29. Kim, Y.-S. *et al.* A Membrane-Bound NAC Transcription Factor Regulates Cell Division
805 in Arabidopsis. *Plant Cell* **18**, 3132–3144 (2006).
- 806 30. Kim, S.-Y. *et al.* Exploring membrane-associated NAC transcription factors in
807 Arabidopsis: implications for membrane biology in genome regulation. *Nucleic Acids Res.*
808 **35**, 203–213 (2007).

- 849 46. Saha, D., Prasad, A. M. & Srinivasan, R. Pentatricopeptide repeat proteins and their
850 emerging roles in plants. *Plant Physiol. Biochem.* **45**, 521–534 (2007).
- 851 47. Barkan, A. & Small, I. Pentatricopeptide Repeat Proteins in Plants. *Annu. Rev. Plant Biol.*
852 **65**, 415–442 (2014).
- 853 48. Nandety, R. S. *et al.* The role of TIR-NBS and TIR-X proteins in plant basal defense
854 responses. *Plant Physiol.* **162**, 1459–72 (2013).
- 855 49. Wan, H. *et al.* Analysis of TIR- and non-TIR-NBS-LRR disease resistance gene analogous
856 in pepper: characterization, genetic variation, functional divergence and expression
857 patterns. *BMC Genomics* **13**, 502 (2012).
- 858 50. Mohanta, T. K. *et al.* Ginkgo biloba responds to herbivory by activating early signaling and
859 direct defenses. *PLoS One* **7**, e32822 (2012).
- 860 51. Mohanta, T. K., Arora, P. K., Mohanta, N., Parida, P. & Bae, H. Identification of new
861 members of the MAPK gene family in plants shows diverse conserved domains and novel
862 activation loop variants. *BMC Genomics* **16**, (2015).
- 863 52. Mohanta, T. K., Mohanta, N., Mohanta, Y. K. & Bae, H. Genome-Wide Identification of
864 Calcium Dependent Protein Kinase Gene Family in Plant Lineage Shows Presence of
865 Novel D-x-D and D-E-L Motifs in EF-Hand Domain. *Front. Plant Sci.* **6**, 1146 (2015).
- 866 53. Zhang, X. *et al.* Multiple functional self-association interfaces in plant TIR domains. *Proc.*
867 *Natl. Acad. Sci.* **114**, E2046 LP-E2052 (2017).
- 868 54. Mohanta, T. K., Park, Y.-H. & Bae, H. Novel Genomic and Evolutionary Insight of WRKY
869 Transcription Factors in Plant Lineage. *Sci. Rep.* **6**, (2016).
- 870 55. Rinerson, C. I., Rabara, R. C., Tripathi, P., Shen, Q. J. & Rushton, P. J. The evolution of
871 WRKY transcription factors. *BMC Plant Biol.* **15**, 1–18 (2015).
- 872 56. Holger, T., Mario, B., Angela, S. & Bernd-Joachim, T. Expression of the gene and
873 processed pseudogenes encoding the human and rabbit translationally controlled tumour
874 protein (TCTP). *Eur. J. Biochem.* **267**, 5473–5481 (2003).
- 875 57. Faustino, L. I., Moretti, A. P. & Graciano, C. Fertilization with urea, ammonium and nitrate
876 produce different effects on growth, hydraulic traits and drought tolerance in Pinus taeda
877 seedlings. *Tree Physiol.* **35**, 1062–1074 (2015).
- 878 58. Kirkpatrick, C. L. *et al.* Growth control switch by a DNA-damage-inducible toxin–
879 antitoxin system in Caulobacter crescentus. *Nat. Microbiol.* **1**, 16008 (2016).
- 880 59. Xu, X., Liu, Q., Fan, L., Cui, X. & Zhou, X. Analysis of synonymous codon usage and
881 evolution of begomoviruses. *J. Zhejiang Univ. Sci. B* **9**, 667–674 (2008).
- 882 60. Hu, T. & Banzhaf, W. Nonsynonymous to Synonymous Substitution Ratio k_a / k_s :
883 Measurement for Rate of Evolution in Evolutionary Computation. in *PPSNX, LNCS 5199*
884 448–457 (2008).
- 885 61. Tomoko, O. Synonymous and nonsynonymous substitutions in mammalian genes and the
886 nearly neutral theory. *J. Mol. Evol.* **40**, 56–63 (1995).
- 887 62. Slotte, T. *et al.* The Capsella rubella genome and the genomic consequences of rapid

- mating system evolution. *Nat. Genet.* **45**, 831–5 (2013).
63. Bolger, A. *et al.* The genome of the stress-tolerant wild tomato species *Solanum pennellii*. *Nat. Genet.* **46**, 1034 (2014).
64. Jin, X. *et al.* Divergent Evolutionary Patterns of NAC Transcription Factors Are Associated with Diversification and Gene Duplications in Angiosperm. *Front. Plant Sci.* **8**, 1156 (2017).
65. Schmitz, J. F., Zimmer, F. & Bornberg-Bauer, E. Mechanisms of transcription factor evolution in Metazoa. *Nucleic Acids Res.* **44**, 6287–6297 (2016).
66. Goodstein, D. M. *et al.* Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178-86 (2012).
67. Boratyn, G. M. *et al.* Domain enhanced lookup time accelerated BLAST. *Biol. Direct* **7**, 12 (2012).
68. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
69. de Castro, E. *et al.* ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* **34**, W362-5 (2006).
70. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes¹¹Edited by F. Cohen. *J. Mol. Biol.* **305**, 567–580 (2001).
71. Nguyen Ba, A., Pogoutse, A., Provart, N. & Moses, A. NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. *BMC Bioinformatics* **10**, 202 (2009).
72. Szklarczyk, D. *et al.* The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* **45**, D362–D368 (2017).
73. von Mering, C. *et al.* STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res.* **31**, 258–261 (2003).

Author contributions

TKM: conceived the idea, performed the experiments and analysis, drafted and revised the manuscript, AK: performed the analysis, DY: revised the manuscript, AH: drafted and revised the manuscript, BT: revised the manuscript, ALK: analysed the data and revised the manuscript, EFA: revised the manuscript, AAH: revised the manuscript.

Figure legends

Figure 1

The distribution of the molecular weight of NAC TFs. The molecular weight of NAC TFs ranged from 2.94 kDa (*Fragaria x ananassa*, FANhyb_icon00034378_a.1.g00001.1) to 346.46 kDa (*Trifolium pratense*, Tp57577_TGAC_v2_mRNA14116). The average molecular weight of NAC TFs was 38.72 kDa. In total, 17158 NAC TFs were utilized in the analysis of molecular weight. The analysis was conducted using a protein isoelectric point calculator (<http://isoelectric.org/>).

Figure 2

The distribution of the isoelectric point of NAC TFs. The isoelectric point of NAC TFs ranged from pI 3.78 (OB07G17140.1, *Oryza brachyantha*) to pI 11.47 (Sevir.3G242500, *Setaria viridis*). The average isoelectric point of NAC TFs was 6.38. A total of 17158 NAC TFs were utilized in the analysis of the pI of NAC TFs. The analysis of pI was conducted using a protein isoelectric point calculator (<http://isoelectric.org/>).

Figure 3

Interactome network of NAC TFs. The interactome network of NAC TF reflects a diverse complex of interacting proteins. The NAC TFs of *A. thaliana* were utilized in the interactome network analysis. The interactome map of *A. thaliana* was determined using the string database (<https://string-db.org>).

Figure 4

Chimeric NAC domains. NAC TFs possess chimeric NAC domains with at least 34 diverse chimeric NAC domains identified in the studied species. The identification of chimeric NAC domain sequences was determined using the ScanProsite and InterProScan server.

Figure 5

Chimeric NAC domains NAC TFs possess chimeric NAC domains with at least 21 diverse chimeric NAC domains identified in the studied species. The identification of chimeric NAC domain sequences was determined using the ScanProsite and InterProScan server.

Figure 6

Differential expression of NAC TFs in leaves of *A. thaliana* plants treated with ammonia, nitrate, and urea. The expression of *A. thaliana* NAC TFs was analysed to determine their response to different sources of nitrogen. Expression data were obtained from the PhytoMine database in Phytozome and presented as FPKM (Fragments per Kilobase of transcripts per million mapped reads).

Figure 7

Differential expression of NAC TFs in roots of *A. thaliana* plants treated with ammonia, nitrate, and urea. The expression of *A. thaliana* NAC TFs was analysed to determine their response to

different sources of nitrogen. Expression data were obtained from the PhytoMine database in Phytozone and presented as FPKM (Fragments per Kilobase of transcripts per million mapped reads).

Figure 8

Phylogenetic tree of NAC TFs. A phylogenetic tree of NAC TF reveals the presence of seven clustered orthologous groups (COGs). Each group also possesses two or more sub-groups. The phylogenetic tree shows lineage (monocot/dicot) specific grouping of NAC TFs. The phylogenetic tree was constructed using the neighbor joining method with 1000 bootstrap replicates.

1015 Table 1. Genomic details of NAC TFs of plants

Sl. No	Name of the species	No. of double domain NAC TF	No. of Novel chimeric NAC TFs	Total No. of NAC TFs	No. of duplicated genes	No. of conditional duplicated genes	No. of lost genes	No. of orthologous genes	No. of paralogous genes	No. of transfer genes
Monocots										
1	<i>Aegilops tauschii</i>		4	117	114	0	0	0	114	0
2	<i>Brachypodium distachyon</i>	2	1	137	135	0	0	0	135	0
3	<i>Brachypodium stacei</i>	1	1	128	127	0	0	0	127	0
4	<i>Hordeum vulgare</i>			76	76	0	0	0	76	0
5	<i>Leersia perrieri</i>	5	2	163	162	0	0	0	162	0
6	<i>Oropetium thomaeum</i>	1		118	103	0	0	0	103	0
7	<i>Oryza barthii</i>		4	134	138	0	0	0	138	0
8	<i>Oryza brachyantha</i>	1	1	118	110	0	0	0	110	0
9	<i>Oryza glaberrima</i>	1		116	110	0	0	0	110	0
10	<i>Oryza glumipatula</i>	2		140	139	0	0	0	139	0
11	<i>Oryza longistaminata</i>	1	6	125	98	0	0	0	98	0
12	<i>Oryza meridionalis</i>	2	2	127	123	0	0	0	123	0
13	<i>Oryza nivara</i>	4	1	146	130	0	0	0	130	0
14	<i>Oryza punctata</i>	6	1	135	133	0	0	0	133	0
15	<i>Oryza rufipogon</i>	4	3	136	129	0	0	0	129	0
16	<i>Oryza sativa subsp. indica</i>	1	3	157	156	0	0	0	156	0
17	<i>Oryza sativa subsp. japonica</i>	1		139	138	0	0	0	138	0
18	<i>Panicum hallii</i>	3	6	139	126	0	0	0	126	0
19	<i>Panicum virgatum</i>	9	6	310	309	0	0	0	309	0
20	<i>Phoenix dactylifera</i>	3	1	124	123	0	0	0	123	0
21	<i>Phyllostachys edulis</i>			125	124	0	0	0	124	0
22	<i>Phyllostachys heterocycla</i>	2	2	125	124	0	0	0	124	0
23	<i>Saccharum officinarum</i>			44	33	0	0	0	33	0
24	<i>Setaria italica</i>	4		139	134	0	0	0	134	0
25	<i>Setaria viridis</i>	1		135	118	0	0	0	118	0
26	<i>Sorghum bicolor</i>	1		141	134	0	0	0	134	0
27	<i>Spirodela polyrhiza</i>			55	48	0	0	0	48	0
28	<i>Triticum aestivum</i>	2	2	263	209	0	0	0	209	0
29	<i>Triticum urartu</i>		1	103	74	0	0	0	74	0
30	<i>Zea mays</i>	1	1	130	119	0	0	0	119	0
31	<i>Zostera marina</i>	1		62	55	0	0	0	55	0
32	<i>Zoysia japonica</i>		4	176	160	0	0	0	160	0
33	<i>Zoysia matrella</i>	1	3	313	230	0	0	0	230	0
34	<i>Zoysia pacifica</i>	1	2	205	183	0	0	0	183	0
Dicots										
35	<i>Actinidia chinensis</i>	1	5	167	166	0	0	0	166	0
36	<i>Aethionema arabicum</i>	3		85	84	0	0	0	84	0

37	<i>Amaranthus hypochondriacus</i>	1		44	37	0	0	0	37	0
38	<i>Amborella trichopoda</i>			46	45	0	0	0	45	0
39	<i>Ananas comosus</i>		1	73	72	0	0	0	72	0
40	<i>Aquilegia coerulea</i>			80	79	0	0	0	79	0
41	<i>Arabidopsis halleri</i>	2		94	93	0	0	0	93	0
42	<i>Arabidopsis lyrata</i>	4	1	122	121	0	0	0	121	0
43	<i>Arabidopsis thaliana</i>	5		113	112	0	0	0	112	0
44	<i>Arabis alpina</i>	1		82	81	0	0	0	81	0
45	<i>Arachis duranensis</i>			82	81	0	0	0	81	0
46	<i>Arachis hypogaea</i>			32	31	0	0	0	31	0
47	<i>Arachis ipaensis</i>			83	81	0	0	0	81	0
48	<i>Artemisia annua</i>			28	27	0	0	0	27	0
49	<i>Azadirachta indica</i>			183	182	0	0	0	182	0
50	<i>Beta vulgaris</i>			53	52	0	0	0	52	0
51	<i>Boechera stricta</i>	2		123	122	0	0	0	122	0
52	<i>Brassica napus</i>	10	7	410	409	0	0	0	409	0
53	<i>Brassica oleracea</i>	4	3	271	270	0	0	0	270	0
54	<i>Brassica rapa</i>	4	2	256	255	0	0	0	255	0
55	<i>Cajanus cajan</i>			96	95	0	0	0	95	0
56	<i>Camelina sativa</i>	17	3	341	330	0	0	0	330	0
57	<i>Cannabis sativa</i>			58	57	0	0	0	57	0
58	<i>Capsella grandiflora</i>	2		95	94	0	0	0	94	0
59	<i>Capsella rubella</i>	5		119	118	0	0	0	118	0
60	<i>Capsicum annum</i>			96	95	0	0	0	95	0
61	<i>Carica papaya</i>			82	81	0	0	0	81	0
62	<i>Castanea mollissima</i>	4		91	78	0	0	0	78	0
63	<i>Catharanthus roseus</i>		2	121	120	0	0	0	120	0
64	<i>Chenopodium quinoa</i>		1	96	95	0	0	0	95	0
65	<i>Cicer arietinum</i>			96	95	0	0	0	95	0
66	<i>Citrullus lanatus</i>			80	79	0	0	0	79	0
67	<i>Citrus clementina</i>			129	128	0	0	0	128	0
68	<i>Citrus sinensis</i>	2		145	143	0	0	0	143	0
69	<i>Coffea canephora</i>			63	62	0	0	0	62	0
70	<i>Cucumis melo</i>			92	91	0	0	0	91	0
71	<i>Cucumis sativus</i>			83	80	0	0	0	80	0
72	<i>Daucus carota</i>		2	96	95	0	0	0	95	0
73	<i>Dianthus caryophyllus</i>			79	77	0	0	0	77	0
74	<i>Dichanthelium oligosanthes</i>	8	2	131	100	0	0	0	100	0
75	<i>Dorcoceras hygrometricum</i>		2	83	76	0	0	0	76	0
76	<i>Elaeis guineensis</i>	2	1	170	167	0	0	0	167	0
77	<i>Eragrostis tef</i>	8	3	172	165	0	0	0	165	0
78	<i>Eucalyptus camaldulensis</i>			200	124	0	0	0	124	0
79	<i>Eucalyptus grandis</i>			164	150	0	0	0	150	0
80	<i>Eutrema salsugineum</i>	2		122	104	0	0	0	104	0
81	<i>Fragaria vesca</i>	3	6	127	123	0	0	0	123	0
82	<i>Fragaria x ananassa</i>	2	1	98	97	0	0	0	97	0
83	<i>Genlisea aurea</i>		1	45	42	0	0	0	42	0
84	<i>Glycine max</i>			180	175	0	0	0	175	0
85	<i>Glycine soja</i>		1	173	166	0	0	0	166	0

86	<i>Gossypium arboreum</i>			150	146	0	0	0	146	0
87	<i>Gossypium hirsutum</i>	1	2	306	296	0	0	0	296	0
88	<i>Gossypium raimondii</i>			153	145	0	0	0	145	0
89	<i>Helianthus annuus</i>			21	20	0	0	0	20	0
90	<i>Humulus lupulus</i>			74	68	0	0	0	68	0
91	<i>Ipomoea trifida</i>	1	2	131	123	0	0	0	123	0
92	<i>Jatropha curcas</i>		1	97	93	0	0	0	93	0
93	<i>Juglans regia</i>	3		92	81	0	0	0	81	0
94	<i>Kalanchoe laxiflora</i>			166	165	0	0	0	165	0
95	<i>Kalanchoe marnieriana</i>			179	178	0	0	0	178	0
96	<i>Lactuca sativa</i>			54	52	0	0	0	52	0
97	<i>Linum usitatissimum</i>	1	1	191	187	0	0	0	187	0
98	<i>Lotus japonicus</i>	2		98	92	0	0	0	92	0
99	<i>Malus domestica</i>	2	9	253	232	0	0	0	232	0
100	<i>Manihot esculenta</i>			130	128	0	0	0	128	0
101	<i>Medicago truncatula</i>	1		97	90	0	0	0	90	0
102	<i>Mimulus guttatus</i>			114	113	0	0	0	113	0
103	<i>Morus notabilis</i>		2	78	77	0	0	0	77	0
104	<i>Musa acuminata</i>	1	1	170	164	0	0	0	164	0
105	<i>Nelumbo nucifera</i>			88	79	0	0	0	79	0
106	<i>Nicotiana benthamiana</i>	2	2	227	185	0	0	0	185	0
107	<i>Nicotiana glauca</i>			156	149	0	0	0	149	0
108	<i>Nicotiana tabacum</i>			280	279	0	0	0	279	0
109	<i>Nicotiana tomentosiformis</i>			172	162	0	0	0	162	0
110	<i>Ocimum tenuiflorum</i>	2	1	110	82	0	0	0	82	0
111	<i>Petunia axillaris</i>	3		131	108	0	0	0	108	0
112	<i>Petunia inflata</i>			157	147	0	0	0	147	0
113	<i>Phaseolus vulgaris</i>			85	84	0	0	0	84	0
114	<i>Populus euphratica</i>	2	3	155	149	0	0	0	149	0
115	<i>Populus trichocarpa</i>		1	169	149	0	0	0	149	0
116	<i>Prunus mume</i>	1		129	128	0	0	0	128	0
117	<i>Prunus persica</i>	1	1	115	114	0	0	0	114	0
118	<i>Pyrus bretschneideri</i>	1	5	185	183	0	0	0	183	0
119	<i>Raphanus raphanistrum</i>	4	3	207	206	0	0	0	206	0
120	<i>Raphanus sativus</i>	5	1	217	197	0	0	0	197	0
121	<i>Ricinus communis</i>			95	87	0	0	0	87	0
122	<i>Salix purpurea</i>			175	152	0	0	0	152	0
123	<i>Salvia miltiorrhiza</i>	1	2	87	81	0	0	0	81	0
124	<i>Sesamum indicum</i>			105	104	0	0	0	104	0
125	<i>Sisymbrium irio</i>	2	2	121	118	0	0	0	118	0
126	<i>Solanum lycopersicum</i>			101	94	0	0	0	94	0
127	<i>Solanum melongena</i>	1	3	95	85	0	0	0	85	0
128	<i>Solanum pennellii</i>		2	102	98	0	0	0	98	0
129	<i>Solanum pimpinellifolium</i>			97	90	0	0	0	90	0
130	<i>Solanum tuberosum</i>	1		129	115	0	0	0	115	0
131	<i>Spinacia oleracea</i>			45	43	0	0	0	43	0
132	<i>Tarenaya hassleriana</i>	1		178	177	0	0	0	177	0
133	<i>Thellungiella halophila</i>	2		122	121	0	0	0	121	0
134	<i>Thellungiella parvula</i>	1		92	91	0	0	0	91	0

135	<i>Theobroma cacao</i>			132	131	0	0	0	131	0
136	<i>Trifolium pratense</i>	2	2	97	76	0	0	0	76	0
137	<i>Utricularia gibba</i>		1	74	73	0	0	0	73	0
138	<i>Vigna angularis</i>			98	97	0	0	0	97	0
139	<i>Vigna radiata</i>	2		82	81	0	0	0	81	0
140	<i>Vigna unguiculata</i>			20	19	0	0	0	19	0
141	<i>Ziziphus jujuba</i>			101	100	0	0	0	100	0
142	<i>Vitis vinifera</i>	1		70	79	0	0	0	79	0
Gymnosperms										
143	<i>Picea abies</i>	1		100	73	0	0	0	73	0
144	<i>Picea glauca</i>			32	31	0	0	0	31	0
145	<i>Picea sitchensis</i>			16	15	0	0	0	15	0
146	<i>Pinus taeda</i>			31	27	0	0	0	27	0
147	<i>Pseudotsuga menziesii</i>	5	3	196	195	0	0	0	195	0
Pteridophyte										
148	<i>Selaginella moellendorffii</i>			22	21	0	0	0	21	0
Bryophytes										
149	<i>Marchantia polymorpha</i>			9						
150	<i>Physcomitrella patens</i>			33	32	0	0	0	32	0
151	<i>Sphagnum fallax</i>			26	25	0	0	0	25	0
Algae										
152	<i>Bathycoccus prasinus</i>			0	0	0	0	0	0	0
153	<i>Chlamydomonas reinhardtii</i>			0	0	0	0	0	0	0
154	<i>Chlorella</i> sp. NC64A			0	0	0	0	0	0	0
155	<i>Coccomyxa</i> sp.			0	0	0	0	0	0	0
156	<i>Dunaliella salina</i>			0	0	0	0	0	0	0
157	<i>Klebsormidium flaccidum</i>			3	0	0	0	0	0	0
158	<i>Micromonas pusilla</i>			0	0	0	0	0	0	0
159	<i>Ostreococcus lucimarinus</i>			0	0	0	0	0	0	0
160	<i>Volvox carteri</i>			0	0	0	0	0	0	0

1016

1017

1018

1019

1020

1021

1022

1023

1024

1025

1026 Table 2

1027 Interactome partners of NAC TFs in plants.

1028

NAC TFs	Experimental Interactions	Co-expression	Text mining Interactions
NAC1	RNS1, AT3G10260, AT1G17080		NAC024, NAC095, ARV1, AT2G01410, AT1G60380, AT1G60340
NAC2	ERD14	NAC32, NAC102, DREB2A	NAC32, NAC102
NAC3	***	****	NTL
NAC4	***	****	UBC30
NAC5	****	****	CYP96A2, MYB
NAC7	VND7	XCP1, XCP2	VND7, MYB46
NAC8	***	ATM, ATR	ATM, ATR
NAC10	***	MYB83, MYB63	MYB83, MYB85, MYB46, MY63, MYB58, MYB52, MYB69, KNAT
NAC11	****	****	NAC95
NAC12	*	IRX1	MYB46, MYB83, MYB58, MYB63, IRX9, APL, KNAT7
NAC13	RCD1	AOX1A, RCD1	AOX1A, RCD1, NAC88
NAC14		ASG2	ASG2, ASG8, AT1G61900
NAC16			NYE, NACA5,
NAC17			NAC88, AT5G13610
NAC18	GAI		NAM, NAC
NAC19	ZFHD1, TCP20, CPL1, TCP8, NAC32, RHA1A, RHA2A	NAC32, ERD1	ZFHD1, TCP20, CPL1, TCP8, NAC32, RHA1A, RHA2A, ERD1
NAC20	AT3G43430, SHR, PHB, PLT2, MYB59, HB23, HB30	TMO6,DOF6, SHR, PLT2	TMO6,DOF6, SHR, PLT2, AT1G64620, AT3G43430
NAC23	****	*****	NAC95, AT3G01030, AT5G27880, AT5G01860, MYB64
NAC24	****	*****	NAC95, NAC47
NAC25	****	At1g75910, GRP20, CYP86C4	At1g75910, GRP20, CYP86C4
NAC26	VND7	VND7, MYB83, XCP1, AT4G08160	VND7, MYB46, MYB85, MYB83, XCP1
NAC028	*****	*****	TOM2B, DBP1, PDLP2, TOM2A
NAC29	NAC6, GRL, IAA14,	NAC6, HAI1	NAC6, HAI1, SAG12, PI
NAC32	HAI1, NAC019, ABI1, NAM, RVE2, PYL4	ATAF1, HAI1, NAC019, GSTU7,	NAC102, NAM, NAC19ATAF1
NAC36	*****	AT5G52760,	AT5G42050

		XBAT34, AT5G52750, SOBIR1, RING1, WRKY53, WRKY46, SARD1, MYB69, CIPK4, ABCA8	AT4G29770, AIP2, SDE3
NAC38	BRM		
NAC41	NAC83	NAC83, AT1G12810	NAC83, GSTF3, AT1G12810
NAC42	****	CYP71A12, GSTU10, AT5G38900, CYP71B6	CYP71A12, GSTU10, AT5G38900,
NAC44	****	****	AT1G54890, NAC90
NAC45	HB52, NAC97	NAC97	CYP71B34, WAK5, NAC97
NAC46	RCD1, BRM	CYP89A9, AT4G11910	RCD1, AT1G78040, bHLH11,
NAC47	***	HAI1, Rap2.6L, NAC6	NAC5, NAC24, HAI1, AT1G60380
NAC48	****	*****	CYP89A9, STAY-GREEN2
NAC49	****	*****	ERF115, WOX5, LBD19
NAC50	JMJ14, NAC052, GAI, TPL	NAC52, MJM14	JMJ14, PPR, NAC52, AT5G41650, CYP71A25
NAC52	JMJ14, NAC50	JMJ14, PPR, UBP14	JMJ14, NAC50, PPR, CRCK2, PPD6, MFDX1, CYP71A25
NAC53	****	AT5G25930, AT3G25610, UGT73B5,	NTL, AT5G25930,
NAC55	ZFHD1, HAI1, F2P16.14	ERD1, AT2G31945, MYB2	ZFHD1, ERD1, HAI1, ABF2, bZIP, MYC2
NAC57	*****	*****	MYB19, AT3G58090, AT1G07730, AT4G13580, AT3G13650
NAC58	*****	RWP1, ABCG6, CYP86A1	PPR, RWP1, ABCG6, MYB86, MYB26
NAC60	****	OLEO1	NACA5, AT3G52350, RINL, AT1G65240, NTL
NAC61	****	NAC90, ACS4,	NAC44, LEA, NAC85, NAC95, NAC90,
NAC62	****	BZIP60, WRKY33, TIP, SZF1, CPK32, CPK28, NHL3	BZIP60, WRKY33, TIP
NAC63	*****	*****	LRR, NAC95, ATPMEPCRD,
NAC64	*****	*****	AT3G59880, AT5G50540, AT2G44010, sks16, SKS6
NAC66	*****	*****	MYB26, MYB46, MYB83, MYB85, MYB63, MYB58, KNAT7, WRKY12
NAC67	*****	****	NAM, AT1G78040, NAC95
NAC69	****	*****	NAC95, CYP96A2, NAM
NAC71	****	WNK, TM6, AT1G64625	Rap2.6L, AT2G41870, RAP2.4
NAC73	****	MYB46, MYB83,	MYB46, MYB83, IRX1, IRX3, MYB63, CESA4

		IRX1, IRX3, CESA4	
NAC74	F2P16.14, TOPLESS, BRM	DSEL, scpl31, HXXXD type	SCRL20, F-ox/LLR, sks11
NAC75	RING/U-box, ZF domain	*****	ERF16, UTr7, CML, TPL/TPR,
NAC76	VND7, NAC83	****	VND7, NAC83, UBQ, MYB46
NAC77	*****	*****	PIP1-5
NAC80	BRM	*****	PPR, TT7, 4CL3, BRM
NAC82	SRO1, RCD1	*****	UBX, WW
NAC83	VND7, NAC41, CUC2, VND1, NAC105, NAC76, NAC101, NAC1	*****	VND7, NAC41, CUC2, VND1, NAC105, NAC76, MYB83, MYB46
NAC84	****	EDF3	ZFP10, Delta9, EDF3, SPT16, GS1
NAC85	****	****	LEA, PUP4, NAC90, NAC61, XERO1
NAC87	****	****	SWAP, WRKY36, TIR-NBS, NBS-LRR, BHLH11
NAC88	****	****	UBC18, NAC17, NAC13, NAC53
NAC89	PAS1, MYB, TI1, TSPO, TIP2.2, TIP3.1	*****	PAS1, maMYB, BZIP60, BZIP28
NAC90	*****	AT3G57460, MPK11	DTA4, CHI, NAC44, NAC85, LEA
NAC94	*****	*****	MC5, D111, RML, BAG6, LCAT3, AATP1, BZIP28
NAC95	*****	NAC24, NAM	NAC23, NAM, NAC24, MAY64, NAC69
NAC96	T21F11.18	*****	ABF2, Dna-J, TOPLESS,
NAC97	NAC45, LRR, BRM	*****	*****
NAC100	*****	*****	AT4G27850, AT1G26410, GRP20, TT7, 4CL3,
NAC101	RPA2, VND7, VR- NAC, NAC83	*****	NVD7, NAC83, XCP1, UBQ, RNS3
NAC102	****	ATAF1, tolB, NAC32, RHL41, ZAT6, UGT73B2	ATAF1, NAC32
NAC103	****	*****	BZIP60, BZIP28, D111, CLPTM1, NAC44
NAC105	VND7, NAC83,	*****	VND7, GH, NAC83, UBQ, LAC1, MYB46, RIC4

1029

1030

1031

1032

1033

1034 Table 3

1035 Codon usage of NAC TFs in plants.

1036

Codons	Codon present in No. of species	Codon absent in No. of species	Average abundance of codons	Highest no. of codons	Name of the species with highest no. of codons
AAA (K)	126	20	4.77	9.9	<i>Glycine soja</i>
AAG (K)	146	0	10.75	24.2	<i>Sphagnum fallax</i>
AAC (N)	144	2	3.66	14.2	<i>Beta vulgaris</i>
AAU (N)	127	19	9.25	20.5	<i>Spinacia oleracea</i>
ACA (T)	139	7	2.33	15.2	<i>Citrus sinensis</i>
ACC (T)	137	9	2.4	17	<i>Amborella trichopoda</i>
ACG (T)	20	126	5.91	13	<i>Doroceras hygrometricum</i>
ACU (T)	146	0	7.42	16.6	<i>Sesamum indicum</i>
AGA (R)	146	0	10.92	24.3	<i>Klebsormidium flaccidum</i>
AGG (R)	146	0	4.12	18.8	<i>Amborella trichopoda</i>
CGA (R)	19	127	5.22	13.9	<i>Linum usitatissimum</i>
CGC (R)	19	127	2.47	6	<i>Linum usitatissimum</i>
CGG (R)	19	127	3.93	8.6	<i>Citrullus lanatus</i>
CGU (R)	19	127	2.06	4.7	<i>Linum usitatissimum</i>
AGC (S)	143	3	3.54	24.2	<i>Beta vulgaris</i>
AGU (S)	144	2	1.83	5.2	<i>Doroceras hygrometricum</i>
UCC (S)	141	5	4.51	12.3	<i>Aegilops tauschii</i>
UCG (S)	20	126	2.64	6.4	<i>Doroceras hygrometricum</i>
UCU (S)	146	0	4.65	30.5	<i>Humulus lupulus</i>
UCA (S)	139	7	5.09	15.1	<i>Morus notabilis</i>
AUA (I)	124	22	4.80	15.3	<i>Sphagnum fallax</i>
AUC (I)	146	0	5.10	16.7	<i>Sphagnum fallax</i>
AUU (I)	126	20	8.71	15.9	<i>Spinacia oleracea</i>
AUG (M)	146	0	7.81	22.8	<i>Sphagnum fallax</i>
CAA (Q)	146	0	5.31	15.4	<i>Fragaria vesca</i>
CAG (Q)	20	126	13.3	22.6	<i>Linum usitatissimum</i>
CAC (H)	20	126	6.64	10.9	<i>Beta vulgaris</i>
CAU (H)	144	2	4.45	9.7	<i>Setaria viridis</i>
CCA (P)	20	126	11.09	16.3	<i>Doroceras hygrometricum</i>
CCC (P)	20	126	14.18	19.2	<i>Amborella trichopoda</i>
CCG (P)	20	126	5.10	11.1	<i>Doroceras hygrometricum</i>
CCU (P)	146	0	8.00	24.7	<i>Klebsormidium flaccidum</i>
CUA (L)	143	3	5.83	28.3	<i>Sphagnum fallax</i>
CUC (L)	123	23	5.74	23.6	<i>Sphagnum fallax</i>
CUG (L)	142	4	5.87	43.9	<i>Sphagnum fallax</i>
CUU (L)	145	1	5.94	32.6	<i>Sphagnum fallax</i>
UUG (L)	125	21	5.94	24.4	<i>Sphagnum fallax</i>
UAA (L)	124	22	5.37	17.2	<i>Sphagnum fallax</i>
GAA (E)	146	0	4.62	27	<i>Klebsormidium flaccidum</i>
GAG (E)	145	1	5.54	18.1	<i>Sphagnum fallax</i>

GAC (D)	145	1	5.05	14.9	<i>Beta vulgaris</i>
GAU (D)	144	2	5.86	21.7	<i>Spinacia oleracea</i>
GCA (A)	135	11	5.49	18.5	<i>Citrus sinensis</i>
GCC (A)	130	16	5.05	15	<i>Amborella trichopoda</i>
GCG (A)	20	126	4.64	11.2	<i>Doroceras hygrometricum</i>
GCU (A)	146	0	4.65	31.1	<i>Setaria viridis</i>
GGA (G)	146	0	4.63	27.5	<i>Setaria viridis</i>
GGC (G)	141	5	5.41	17.1	<i>Amborella trichopoda</i>
GGG (G)	145	1	2.7	6.7	<i>Elaeis guineensis</i>
GGU (G)	145	1	2.40	5.9	<i>Elaeis guineensis</i>
GUA (V)	140	6	1.46	3.5	<i>Sphagnum fallax</i>
GUC (V)	123	23	0.93	2	<i>Morus notabilis</i>
GUG (V)	142	4	4.35	11.8	<i>Beta vulgaris</i>
GUU (V)	143	3	5.38	16.6	<i>Klebsormidium flaccidum</i>
UAC (Y)	138	8	3.84	10.1	<i>Morus notabilis</i>
UAU (Y)	126	20	6.23	14.8	<i>Solanum melongena</i>
UGG (W)	147	0	3.89	14.5	<i>Vitis vinifera</i>
UGC (C)	143	3	5.14	15.6	<i>Oropetium thomaeum</i>
UGU (C)	145	1	3.9	9.6	<i>Zoysia matrella</i>
UUC (F)	146	0	4.60	25.4	<i>Picea glauca</i>
UUU (F)	126	20	10.67	19.2	<i>Sphagnum fallax</i>

1037

1038

1039

1040

1041

1042

1043

1044

1045

1046

1047

1048

1049

1050

1051

1052 Table 4. Substitution rate of NAC TFs of plants.

	A	T	C	G	A	T	C	G	A	T	C	G	A	T	C	G
	<i>Actinidia chinensis</i>				<i>Aegilops tauschii</i>				<i>Aethionema arabicum</i>				<i>Amaranthus hypochondriacus</i>			
A	-	10.6	4.26	10.8	-	10.3	4.7	12.3	-	11.4	4.3	10.1	-	11.6	4.83	10.4
		2		3		6	1	2		3		5		9		5
T	8.34	-	4.34	5.38	7.78	-	4.3	5.96	9.01	-	3.54	5.7	8.54	-	3.45	5.16
				7			7					7				
C	8.34	10.8	-	5.38	7.78	9.62	-	5.96	9.01	9.39	-	5.7	8.54	8.35	-	5.16
		3														
G	16.79	10.6	4.26	-	16.09	10.3	4.7	-	16.0	11.4	4.3	-	17.3	11.6	4.83	-
		2				6	1		4	3			2	9		
	<i>Amborella trichopoda</i>				<i>Ananas comosus</i>				<i>Aquilegia coerulea</i>				<i>Arabidopsis halleri</i>			
A	-	3.99	10.9	14.1	-	3.73	10.9	15.3	-	11	4.33	11.1	-	10.9	4.15	11.0
			8	4			5	6				1		3		7
T	8.17	-	5.41	10.7	7.66	-	5.91	10.3	8.67	-	3.44	5.32	8.98	-	3.3	5.73
								2								
C	8.17	1.97	-	10.7	7.66	2.01	-	10.3	8.67	8.73	-	5.32	8.98	8.71	-	5.73
								2								
G	10.8	3.99	10.9	-	11.3	3.73	10.9	-	18.0	11	4.33	-	17.3	1.93	4.15	-
			8		9		5		9				4			
	<i>Arabidopsis lyrata</i>				<i>Arabidopsis thaliana</i>				<i>Arabis alpina</i>				<i>Arachis duranensis</i>			
A	-	4.03	10.6	14.3	-	4.06	10.4	14.8	-	11.0	4.25	10.6	-	11.1	4.43	10.3
				5				6		4		2		7		
T	8.63	-	4.51	10.6	8.41	-	4.65	10.5	8.88	-	3.63	5.67	8.76	-	3.82	5.37
				4				2								
C	8.63	1.71	-	10.6	8.41	1.82	-	10.5	8.88	9.42	-	5.67	8.76	9.62	-	5.37
				4				2								
G	11.6	4.03	10.6	-	11.8	4.06	10.4	-	16.6	11.0	4.25	-	16.7	11.1	4.43	-
	3				9				3	4			9	7		
	<i>Arachis hypogaea</i>				<i>Arachis ipaensis</i>				<i>Artemisia annua</i>				<i>Azadirachta indica</i>			
A	-	4	11.0	14.2	-	3.93	10.8	14.6	-	10.8	4.08	10.9	-	10.9	4.22	10.3
			7	1			5	3		5		3		8		4
T	8.16	-	5.53	10.2	8.39	-	4.95	10	8.24	-	4.22	5.62	8.91	-	3.82	5.6
				4												
C	8.16	2	-	10.2	8.39	1.79	-	10	8.24	11.2	-	5.62	8.91	9.95	-	5.6
				4						4						
G	11.3	4	11.0	-	12.2	3.93	10.8	-	16.0	10.8	4.08	-	16.4	10.9	4.22	-
	2		7		8		5		1	5			4	8		
	<i>Beta vulgaris</i>				<i>Boechera stricta</i>				<i>Brachypodium distachyon</i>				<i>Brachypodium stacei</i>			
A	-	3.85	10.9	14.3	-	11.0	4.23	10.8	-	10.6	4.93	12.5	-	10.5	4.96	12.6
			5	8		2		6				7		2		5
T	8.62	-	4.72	10	8.77	-	3.57	5.58	7.75	-	4.05	6.19	7.68	-	4.16	6.21
C	8.62	1.66	-	10	8.77	9.29	-	5.58	7.75	8.71	-	6.19	7.68	8.83	-	6.21
G	12.4	3.85	10.9	-	17.0	11.0	4.23	-	15.7	10.6	4.93	-	15.6	10.5	4.96	-
			5		9	2			4				3	2		
	<i>Brassica napus</i>				<i>Brassica oleracea</i>				<i>Brassica rapa</i>				<i>Cajanus cajan</i>			
A	-	10.8	4.26	10.7	-	10.8	4.29	10.7	-	3.98	10.7	14.4	-	11.3	4.53	10.3

		3		1		3		2		3	7		1			
T	8.61	-	4.01	5.57	8.7	-	3.91	5.57	8.48	-	4.63	10.6	8.63	-	3.9	5.53
C	8.61	10.1	-	5.57	8.7	9.87	-	5.57	8.48	1.72	-	10.6	8.63	9.73	-	5.53
		8														
G	16.5	10.8	4.26	-	16.7	10.8	4.29	-	11.5	3.98	10.7	-	16.0	11.3	4.53	-
	7	3			3	3			8		3		7	1		
Camelina sativa				Cannabis sativa				Capsella grandiflora				Capsella rubella				
A	-	3.88	10.4	14.9	-	11.1	4.35	10.4	-	12.3	4.92	9.49	-	11.1	4.35	10.4
			4			7		9		4				9		8
T	8.41	-	4.9	10.4	9.25	-	3.24	5.72	9.91	-	2.41	6.34	8.81	-	3.69	5.65
				6												
C	8.41	1.82	-	10.4	9.25	8.33	-	5.72	9.91	6.2	-	6.34	8.81	9.48	-	5.65
				6												
G	11.9	3.88	10.4	-	16.9	11.1	4.35	-	14.8	12.3	4.92	-	16.3	11.1	4.35	-
	9		4		7	7			3	4			3	9		

1053

Capsicum annuum				Carica papaya				Castanea mollissima				Catharanthus roseus				
A	-	10.9	3.95	10.8	-	10.4	4.05	11.4	-	11.3	4.53	10.0	-	10.5	4.13	11.3
		5				3		9		1		3		4		5
T	8.91	-	3.47	5.49	8.39	-	3.98	5.47	8.98	-	3.75	5.73	8.6	-	3.82	5.69
C	8.91	9.62	-	5.49	8.39	10.2	-	5.47	8.98	9.37	-	5.73	8.6	9.76	-	5.69
						4										
G	17.5	10.9	3.95	-	17.6	10.4	4.05	-	15.7	11.3	4.53	-	17.1	10.5	4.13	-
	1	5			3	3			2	1			4	4		
Chenopodium quinoa				Cicer arietinum				Citrullus lanatus				Citrus clementina				
A	-	11.2	4.48	10.4	-	4.05	10.9	13.8	-	3.95	10.8	14.6	-	3.9	10.5	14.7
				6			1				2	2			6	2
T	8.77	-	3.76	5.67	8.93	-	4.36	10.1	8.25	-	5.19	10.2	8.33	-	5.26	10.2
								7				4				9
C	8.77	9.4	-	5.67	8.93	1.62	-	10.1	8.25	1.89	-	10.2	8.33	1.94	-	10.2
								7				4				9
G	16.1	11.2	4.48	-	12.1	4.05	10.9	-	11.7	3.95	10.8	-	11.9	3.9	10.5	-
	6				1		1		7		2		2		6	
Citrus sinensis				Coffea canephora				Cucumis melo				Cucumis sativus				
A	-	3.93	10.4	14.3	-	11.2	4.64	10.3	-	11.3	4.53	10.2	-	4.18	10.9	14.1
			9	7		2				1		1			7	2
T	8.39	-	5.72	10.1	8.63	-	4.03	5.76	8.5	-	4.07	5.39	8.45	-	4.75	10.2
				2												
C	8.39	2.14	-	10.1	8.63	9.75	-	5.76	8.5	10.1	-	5.39	8.45	1.81	-	10.2
				2						7						
G	11.9	3.93	10.4	-	15.4	11.2	4.64	-	16.0	11.3	4.53	-	11.6	4.18	10.9	-
	1		9		2	2			9	1			9		7	
Daucus carota				Dianthus caryophyllus				Dichanthelium oligosanthes				Dorcoceras hygrometricum				
A	-	4.04	10.9	13.8	-	3.93	10.7	14.2	-	10.5	4.97	12.6	-	4.04	10.8	14.1
			9				4	6		5		6			5	4
T	8.52	-	5.16	10.3	8.57	-	5.02	10.2	7.49	-	4.32	5.97	8.47	-	5.04	10.3
				4				5								

C	8.52	1.89	-	10.3 4	8.57	1.84	-	10.2 5	7.49	9.16	-	5.97	8.47	1.88	-	10.3
G	11.3 7	4.04	10.9 9	-	11.9 1	3.93	10.7 4	-	15.8 8	10.5 5	4.97	-	11.6 2	4.04	10.8 5	-
Elaeis guineensis				Eragrostis tef				Eucalyptus camaldulensis				Eucalyptus grandis				
A	-	3.85	10.7 9	15.2 8	-	10.7 2	4.7	11.4 8	-	10.8 4	4.28	10.8 1	-	3.95	10.6 8	14.9 4
T	7.88	-	5.23	10.6 3	8.34	-	3.99	6.09	8.47	-	4.13	5.7	7.87	-	5.87	10.3 1
C	7.88	1.86	-	10.6 3	8.34	9.12	-	6.09	8.47	10.4 4	-	5.7	7.87	2.17	-	10.3 1
G	11.3 3	3.85	10.7 9	-	15.7 2	10.7 2	4.7	-	16.0 5	10.8 4	4.28	-	11.4	3.95	10.6 8	-
Eutrema salsugineum				Fragaria vesca				Fragaria x ananassa				Genlisea aurea				
A	-	11.0 4	4.27	10.6 7	-	10.8 5	4.49	10.6 7	-	10.2 2	4.07	11.1 9	-	10.5	3.98	11.5 8
T	8.8	-	3.67	5.62	8.64	-	3.96	5.4	8.26	-	4.48	5.34	9.08	-	3.34	6.52
C	8.8	9.49	-	5.62	8.64	9.56	-	5.4	8.26	11.2 4	-	5.34	9.08	8.79	-	6.52
G	16.7 4	11.0	4.27	-	17.0 6	10.8 5	4.49	-	17.3 2	10.2 2	4.07	-	16.1 2	10.5	3.98	-
Glycine max				Glycine soja				Gossypium arboreum				Gossypium hirsutum				
A	-	11.4 3	4.62	10.1 8	-	11.5 3	4.72	10.0 7	-	11.0 5	4.39	10.5 6	-	11.0 6	4.4	10.5 1
T	8.78	-	3.73	5.59	8.98	-	3.55	5.68	8.64	-	3.89	5.5	8.6	-	3.98	5.55
C	8.78	9.24	-	5.59	8.98	8.67	-	5.68	8.64	9.8	-	5.5	8.6	10.0 2	-	5.55
G	16.0 1	11.4 3	4.62	-	15.9 2	11.5 3	4.72	-	16.5 8	11.0 5	4.39	-	16.2 7	11.0 6	4.4	-
Gossypium raimondii				Helianthus annuus				Hordeum vulgare				Humulus lupulus				
A	-	11.2 8	4.52	10.2 2	-	10.3 9	4.31	10.7 9	-	10.7 4	4.8	11.6 4	-	11.2 1	4.63	10.5 2
T	8.85	-	3.77	5.62	8.23	-	4.74	5.46	8.41	-	3.8 2	6.48	8.7	-	3.68	5.37
C	8.85	9.39	-	5.62	8.23	11.4 2	-	5.46	8.41	8.53	-	6.48	8.7	8.92	-	5.37
G	16.0 9	11.2 8	4.52	-	16.2 8	10.3 9	4.31	-	15.1 2	10.7 4	4.8 1	-	17.0 6	11.2 1	4.63	-
Ipomoea trifida				Jatropha curcas				Juglans regia				Kalanchoe laxiflora				
A	-	10.7 9	4.47	10.9 4	-	11.1 9	4.34	10.6 6	-	11.0 9	4.4 5	10.8 2	-	11.5 2	4.93	9.82
T	8.33	-	4.22	5.6	8.68	-	3.66	5.5	8.45	-	3.9 3	5.64	8.72	-	4.03	5.74
C	8.33	10.1 9	-	5.6	8.68	9.44	-	5.5	8.45	9.79	-	5.64	8.72	9.4	-	5.74
G	16.2 7	10.7 9	4.47	-	16.8 3	11.1 9	4.34	-	16.2 2	11.0 9	4.4 5	-	14.9 2	11.5 2	4.93	-
Kalanchoe marnieriana				Klebsormidium flaccidum				Lactuca sativa				Leersia perrieri				
A	-	12.0	5.24	9.86	-	7.73	4.03	13.0	-	11.3	4.5	9.9	-	10.3	4.57	13.2

T	9.37	8 -	2.86	6.19	6.67	-	7.19	8 5.04	9.13	9 -	7 3.6	5.91	7.33	1 -	4.28	7 5.82
C	9.37	6.59	-	6.19	6.67	13.7 8	-	5.04	9.13	9.16	-	5.91	7.33	9.66	-	5.82
G	14.9 3	12.0 8	5.24	-	18.2 8	7.73	4.03	-	15.2 8	11.3 9	4.5 7	-	16.7 3	10.3 1	4.57	-
Linum usitatissimum				Lotus japonicus				Malus domestica				Manihot esculenta				
A	-	3.72	10.3 9	15.8 2	-	10.3 1	3.72	12.6 4	-	10.6 5	4.2 3	11.2 1	-	11.2 2	4.54	10.4 8
T	7.89	-	5.6	10.0 9	8.2	-	3.43	5.44	8.18	-	4.1 7	5.24	8.65	-	3.82	5.57
C	7.89	2.01	-	10.0 9	8.2	9.51	-	5.44	8.18	10.5	-	5.24	8.65	9.45	-	5.57
G	12.3 7	3.72	10.3 9	-	19.0 7	10.3 1	3.72	-	17.5	10.6 5	4.2 3	-	16.2 9	11.2 2	4.54	-
Medicago truncatula				Mimulus guttatus				Morus notabilis				Musa acuminata				
A	-	11.2 4	4.3	10.1 9	-	10.7 2	4.02	10.5 1	-	11.1 6	4.5 4	10.6 4	-	10.9 3	4.62	10.8 9
T	8.9	-	3.68	5.4	8.91	-	3.91	5.67	8.54	-	3.8	5.32	8.14	-	4.42	5.9
C	8.9	9.62	-	5.4	8.91	10.4 2	-	5.67	8.54	9.34	-	5.32	8.14	10.4 7	-	5.9
G	16.7 9	11.2 4	4.3	-	16.5 1	10.7 2	4.02	-	17.1	11.1 6	4.5 4	-	15.0 5	10.9 3	4.62	-
Nelumbo nucifera				Nicotiana benthamiana				Nicotiana sylvestris				Nicotiana tabacum				
A	-	10.8 8	4.4	10.9	-	10.8 8	4.11	10.3 9	-	10.8 8	4.2 6	10.4	-	11.2	4.31	10.2 1
T	8.43	-	4.04	5.56	8.82	-	3.89	5.38	8.87	-	3.8 6	5.44	9.07	-	3.57	5.55
C	8.43	9.99	-	5.56	8.82	10.2 9	-	5.38	8.87	9.87	-	5.44	9.07	9.28	-	5.55
G	16.5 5	10.8 8	4.4	-	17.0 3	10.8 8	4.11	-	16.9 8	10.8 8	4.2 6	-	16.6 9	11.2	4.31	-
Nicotiana tomentosiformis				Ocimum tenuiflorum				Oropetium thomaeum				Oryza brachyantha				
A	-	11.0 5	4.3	10.4 3	-	11.0 2	4.32	10.5 9	-	10.5 2	4.8 2	12.4 1	-	10.5 1	4.85	12.9 8
T	8.93	-	3.64	5.43	8.44	-	4.15	5.68	7.46	-	4.5 6	5.93	7.3	-	4.37	5.92
C	8.93	9.36	-	5.43	8.44	10.5 8	-	5.68	7.46	9.97	-	5.93	7.3	9.47	-	5.92
G	17.1 6	11.0 5	4.3	-	15.7 4	11.0 2	4.32	-	15.6 1	10.5 2	4.8 2	-	16.0 1	10.5 1	4.85	-
Oryza glaberrima				Oryza glumaepatula				Oryza longistaminata				Oryza meridionalis				
A	-	10.65	5.01	12.8 6	-	10.7 1	5	12.4 7	-	10.6 9	4.7 4	12.1 1	-	10.4 2	4.87	13.3
T	7.52	-	4.08	6.14	7.61	-	4.18	6.16	7.96	-	4.0 3	6.09	7.43	-	4.1	6.12

1055

C	7.52	8.69	-	6.14	7.61	8.96	-	6.16	7.96	9.09	-	6.09	7.43	8.79	-	6.12
G	15.7	10.65	5.01	-	15.4	10.7	5	-	15.8	10.6	4.7	-	16.1	10.4	4.87	-
	3				2	1			2	9	4		4	2		
<i>Oryza nivara</i>				<i>Oryza punctata</i>				<i>Oryza rufipogon</i>				<i>Oryza sativa</i>				
A	-	10.64	5.01	12.4	-	10.7	4.94	12.3	-	10.6	5.0	12.8	-	10.8	5.12	12.2
				9		1		7		7	4	4		6		2
T	7.6	-	4.25	6.17	7.66	-	4.2	6.23	7.47	-	4.1	6.18	7.64	-	4.21	6.33
											4					
C	7.6	9.01	-	6.17	7.66	9.12	-	6.23	7.47	8.76	-	6.18	7.64	8.94	-	6.33
G	15.3	10.6	5.01	-	15.2	10.7	4.94	-	15.5	10.6	5.0	-	17.7	10.8	5.12	-
	9				2	1			3	7	4		3	6		
<i>Panicum hallii</i>				<i>Panicum virgatum</i>				<i>Petunia axillaris</i>				<i>Petunia inflata</i>				
A	-	10.66	5	12.6	-	10.4	5.02	12.7	-	10.0	4.1	10.0	-	11.1	4.28	10.2
				3				2		3	8	7		3		8
T	7.75	-	3.97	6.32	7.53	-	4.39	6.14	8.77	-	4.0	5.35	8.83	-	3.79	5.41
											5					
C	7.75	8.47	-	6.32	7.53	9.12	-	6.14	8.77	10.6	-	5.35	8.83	9.86	-	5.41
										8						
G	15.4	10.66	5	-	15.6	10.4	5.02	-	16.5	10.0	4.1	-	16.7	11.1	4.28	-
	8								2	3	8		7	3		
<i>Phaselous vulgaris</i>				<i>Phoenix dactylifera</i>				<i>Phyllostachys heterocycla</i>				<i>Physcomitrella patens</i>				
A	-	11.4	4.62	10.2	-	10.7	4.44	11.1	-	11.0	4.9	11.6	-	10.6	4.64	10.7
				4		1		6		3	7			1		7
T	8.75	-	3.78	5.69	8.2	-	4.31	5.71	8.1	-	3.9	6.27	8.1	-	4.77	5.53
											4					
C	8.75	9.32	-	5.69	8.2	10.3	-	5.71	8.1	8.73	-	6.27	8.1	10.9	-	5.53
						9								2		
G	15.7	11.4	4.62	-	16.0	10.7	4.44	-	14.9	11.0	4.9	-	15.7	10.6	4.64	-
	5					1			9	3	7		7	1		
<i>Picea abies</i>				<i>Picea glauca</i>				<i>Picea sitchensis</i>				<i>Pinus taeda</i>				
A	-	10.82	4.54	10.2	-	10.8	4.61	10.3	-	10.3	4.4	10.5	-	10.0	4.37	11.6
				9		5		5		5	5	8		8		7
T	8.53	-	4.42	5.48	8.42	-	4.47	5.42	8.02	-	5.1	5.12	8.18	-	4.39	5.35
											1					
C	8.53	10.53	-	5.48	8.42	10.5	-	5.42	8.02	11.8	-	5.12	8.18	10.1	-	5.35
						2				7				1		
G	16	10.82	4.54	-	16.0	10.8	4.61	-	16.5	10.3	4.4	-	17.8	10.0	4.37	-
					8	5			6	5	5		6	8		
<i>Populus euphratica</i>				<i>Populus trichocarpa</i>				<i>Prunus mume</i>				<i>Prunus persica</i>				
A	-	10.86	4.53	10.8	-	10.9	4.52	10.6	-	11.1	4.4	10.6	-	11.3	4.53	10.2
				5		8		8		2	3	5		8		3
T	8.33	-	4.15	5.31	8.47	-	4.01	5.35	8.91	-	3.4	5.55	8.95	-	3.55	5.6
											9					
C	8.33	9.95	-	5.31	8.47	9.74	-	5.35	8.91	8.76	-	5.55	8.95	8.92	-	5.6
G	17	10.86	4.53	-	16.9	10.9	4.52	-	17.0	11.1	4.4	-	16.3	11.3	4.53	-
					3	8			7	2	3		6	8		
<i>Pseudotsuga menziesii</i>				<i>Pyrus bretschneideri</i>				<i>Raphanus raphanistrum</i>				<i>Raphanus sativus</i>				
A	-	10.74	4.44	9.98	-	11.1	4.51	10.7	-	11.0	4.4	10.5	-	10.9	4.35	10.7
								8		7		1		6		1

1056

T	8.49	-	4.79	5.68	8.57	-	3.77	5.47	8.72	-	3.85	5.6	8.47	-	4.06	5.52
C	8.49	11.59	-	5.68	8.57	9.26	-	5.47	8.72	9.7	-	5.6	8.47	10.23	-	5.52
G	14.93	10.74	4.44	-	16.89	11.1	4.51	-	16.36	11.07	4.4	-	16.44	10.96	4.35	-
<i>Ricinus communis</i>				<i>Saccharum officinarum</i>				<i>Salix purpurea</i>				<i>Salvia miltiorrhiza</i>				
A	-	10.87	4.28	11.05	-	9.95	4.48	12.78	-	10.85	4.5	10.98	-	11.27	4.64	10.63
T	8.56	-	3.76	5.48	8.25	-	3.94	6.61	8.24	-	4.16	5.32	8.49	-	3.89	5.7
C	8.56	3.76	-	5.48	8.25	8.76	-	6.61	8.24	10.04	-	5.32	8.49	9.44	-	5.7
G	17.27	10.87	4.28	-	15.95	9.95	4.48	-	17.01	10.85	4.5	-	15.82	11.27	4.64	-
<i>Selaginella moellendorffii</i>				<i>Sesamum indicum</i>				<i>Setaria italica</i>				<i>Setaria viridis</i>				
A	-	10.76	4.13	11.65	-	11.3	4.82	10.17	-	10.82	5.13	12.29	-	10.6	5.05	12.45
T	8.65	-	3.52	6.3	8.63	-	4.03	5.65	7.69	-	4.12	6.22	7.67	-	4.22	6.1
C	8.65	9.15	-	6.3	8.63	9.45	-	5.65	7.69	8.67	-	6.22	7.67	8.87	-	6.1
G	16.01	10.76	4.13	-	15.54	11.3	4.82	-	15.19	10.82	5.13	-	15.65	10.6	5.05	-
<i>Sisymbrium irio</i>				<i>Solanum lycopersicum</i>				<i>Solanum melongena</i>				<i>Solanum pennellii</i>				
A	-	10.74	4.29	11.05	-	11.17	4.12	10.13	-	10.72	4.0	10.67	-	11.07	4.11	10.25
T	8.45	-	3.96	5.37	9.1	-	3.58	5.45	8.67	-	3.95	5.3	8.96	-	3.7	5.39
C	8.45	9.91	-	5.37	9.1	9.7	-	5.45	8.67	10.48	-	5.3	8.96	9.95	-	5.39
G	17.37	10.74	4.29	-	16.97	11.1	4.12	-	17.46	10.72	4.0	-	17.03	11.07	4.11	-
<i>Solanum pimpinellifolium</i>				<i>Solanum tuberosum</i>				<i>Sorghum bicolor</i>				<i>Sphagnum fallax</i>				
A	-	11.4	4.23	9.8	-	10.96	4.13	10.34	-	10.64	5.07	12.24	-	10.95	4.92	10.65
T	9.3	-	3.48	5.58	8.83	-	3.86	5.38	7.76	-	4.23	6.21	8.03	-	4.69	5.7
C	9.3	9.38	-	5.58	8.83	10.24	-	5.38	7.76	8.88	-	6.21	8.03	10.44	-	5.7
G	16.33	11.4	4.23	-	16.98	10.96	4.13	-	15.31	10.64	5.07	-	15	10.95	4.92	-
<i>Spinacia oleracea</i>				<i>Spirodela polyrhiza</i>				<i>Tarenaya hassleriana</i>				<i>Thellungiella parvula</i>				
A	-	11.37	4.57	10.2	-	11.01	4.84	10.69	-	11.05	4.52	10.66	-	11.34	4.46	10.21
T	8.74	-	3.76	5.4	8.45	-	4.21	6.38	8.73	-	3.81	5.81	8.69	-	3.9	5.62
C	8.74	9.35	-	5.4	8.45	9.59	-	6.38	8.73	9.31	-	5.81	8.69	9.9	-	5.62

G	16.5	11.37	4.57	-	14.1	11.0	4.84	-	16.0	11.0	4.5	-	15.7	11.3	4.46	-	
	1				6	1			2	5	2		8	4			
<i>Theobroma cacao</i>				<i>Trifolium pratense</i>				<i>Triticum aestivum</i>				<i>Triticum urartu</i>					
A	-	11.04	4.45	10.5	-	11.2	4.26	9.73	-	10.6	4.8	11.2	-	10.7	4.94	11.4	
				2		1				5	5			8			
T	8.53	-	4.04	5.43	9.05	-	3.89	5.43	8.28	-	4.3	6.16	8.12	-	4.27	6.19	
											4						
C	8.53	10.02	-	5.43	9.05	10.2	-	5.43	8.28	9.54	-	6.16	8.12	9.32	-	6.19	
						4											
G	16.5	11.04	4.45	-	16.2	11.2	4.26	-	15.0	10.6	4.8	-	14.9	10.7	4.94	-	
	3				4	1			4	5	5		5	8			
<i>Utricularia gibba</i>				<i>Vigna angularis</i>				<i>Vigna radiata</i>				<i>Vigna unguiculata</i>					
A	-	10.97	4.56	10.6	-	11.1	4.45	10.4	-	10.7	4.3	10.6	-	10.2	3.95	11.1	
				7		8				6	3	4		1		1	
T	8.41	-	4.26	6	8.87	-	3.72	5.74	8.58	-	4.1	5.59	8.54	-	4.35	5.86	
											6						
C	8.41	10.23	-	6	8.87	9.34	-	5.74	8.58	10.3	-	5.59	8.54	11.2	-	5.86	
										4				4			
G	14.9	10.97	4.56	-	16.0	11.1	4.45	-	16.3	10.7	4.3	-	16.1	10.2	3.95	-	
	6				8	8			3	6	3		9	1			
<i>Vitis vinifera</i>				<i>Zea mays</i>				<i>Ziziphus jujuba</i>				<i>Zoysia japonica</i>					
A	-	11.27	4.64	10.2	-	10.7	5.14	12.1	-	11.0	5.5	10.7	-	10.8	5.02	12.0	
				1		7				3	4	1		3		8	
T	8.69	-	3.98	5.74	7.82	-	4.15	6.34	8.6	-	3.8	5.47	7.87	-	4.04	6.18	
											4						
C	8.69	9.68	-	5.74	7.82	8.69	-	6.34	8.6	9.33	-	5.47	7.87	8.72	-	6.18	
G	15.4	11.27	4.64	-	14.9	10.7	5.14	-	16.8	11.0	5.5	-	15.3	10.8	5.02	-	
	5				1	7			4	3	4		8	3			
<i>Zoysia matrella</i>				<i>Zoysia pacifica</i>													
A	-	10.65	5	12.2	-	10.4	4.88	12.6									
				9		1		7									
T	7.73	-	4.21	6.11	7.57	-	4.35	6.04									
C	7.73	8.98	-	6.11	7.57	9.28	-	6.04									
G	15.5	10.65	5	-	18.8	10.4	4.88	-									
	4				9	1											

1057

1058

1059

1060

1061

1062

1063 **Supplementary Data**

1064 **Supplementary Table 1**

1065 Supplementary table showing different chimeric domains of NAC TFs.

1066 **Supplementary Table 2**

1067 NAC TFs showing the presence of novel functional domain along with NAC domains.

1068 **Supplementary Table 3**

1069 NAC TFs showing their involvement in different pathways and biological process.

1070

1071

1072

1073

1074

1075

1076

1077

1078

1079

1080

1081

1082

1083

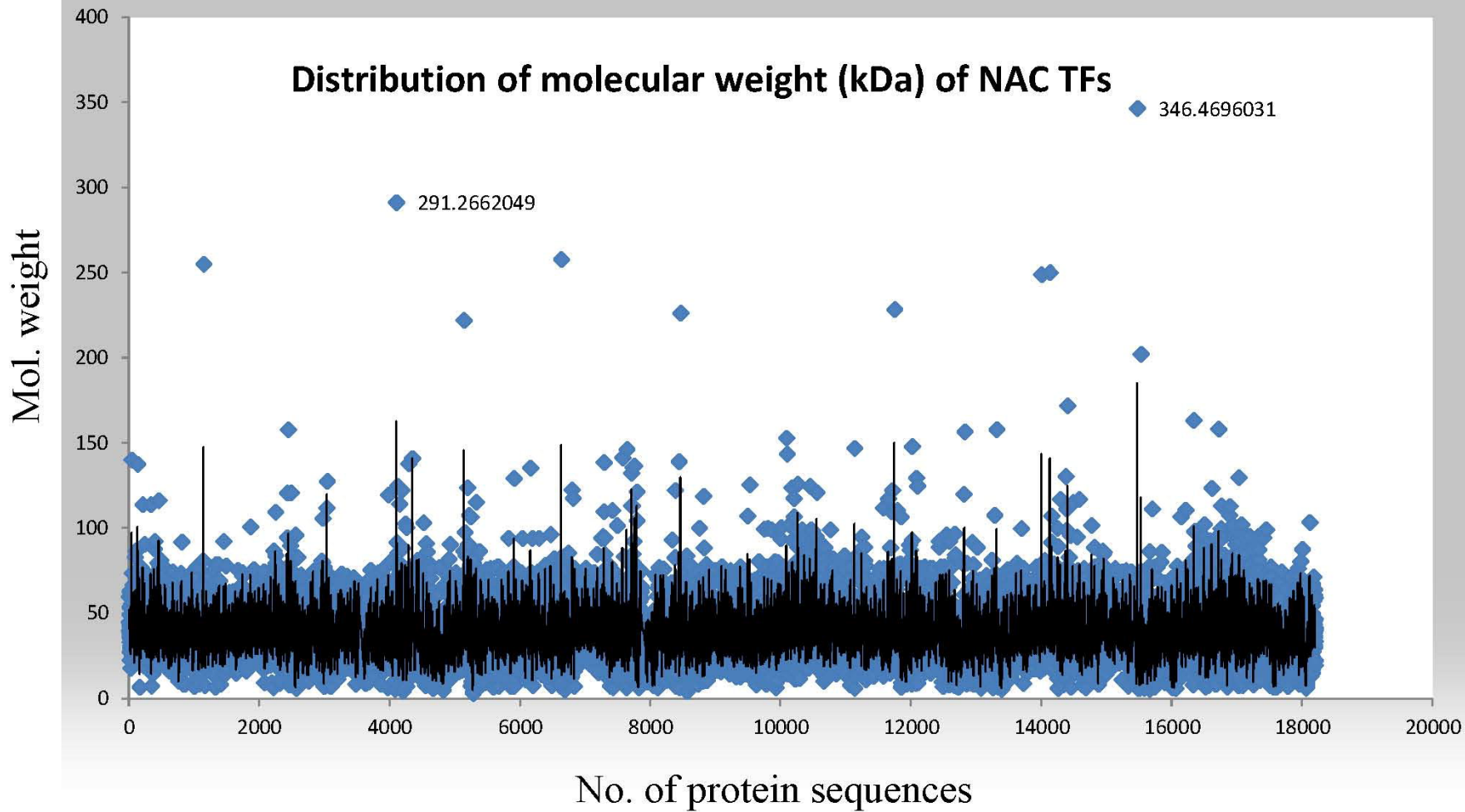
1084

1085

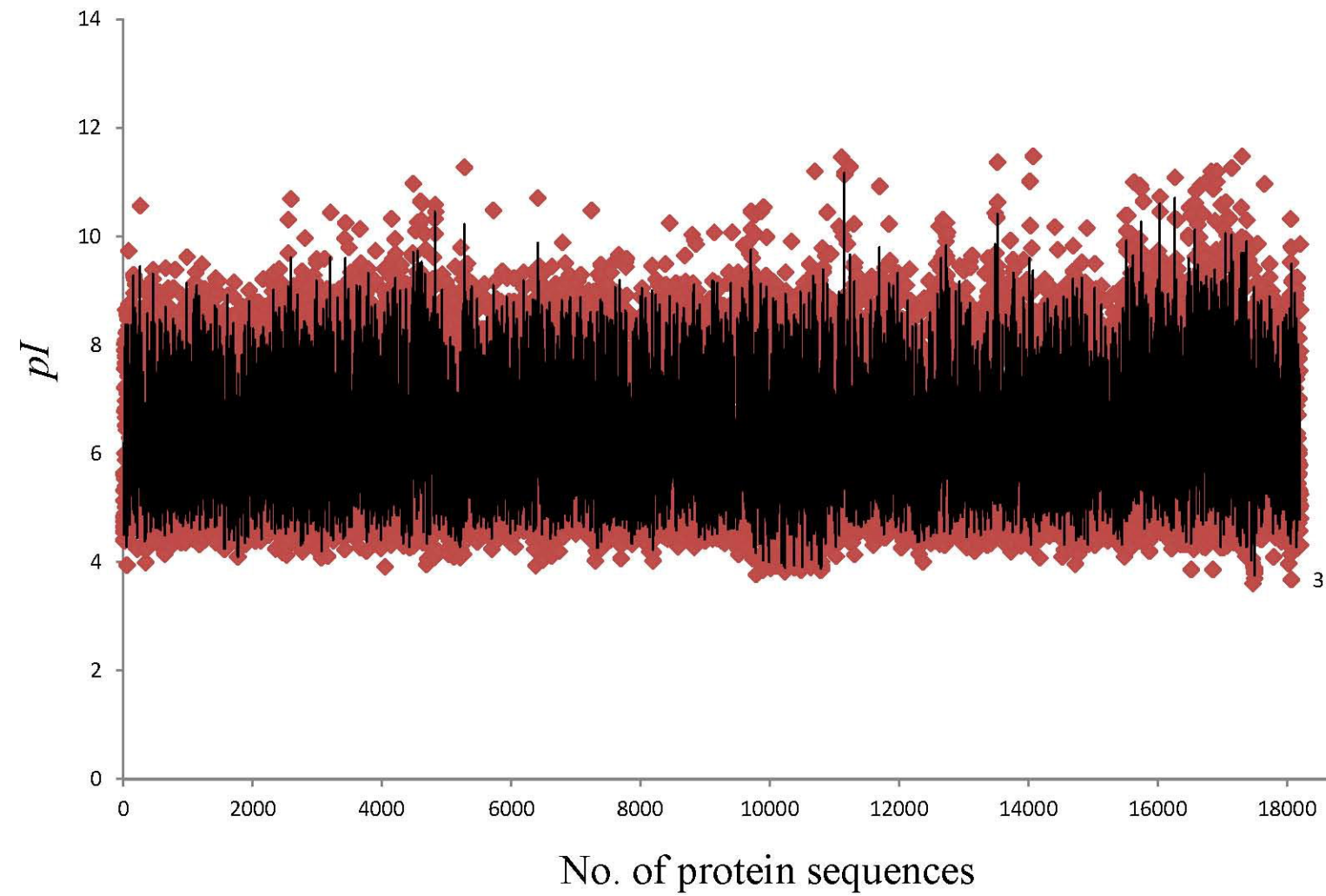
1086

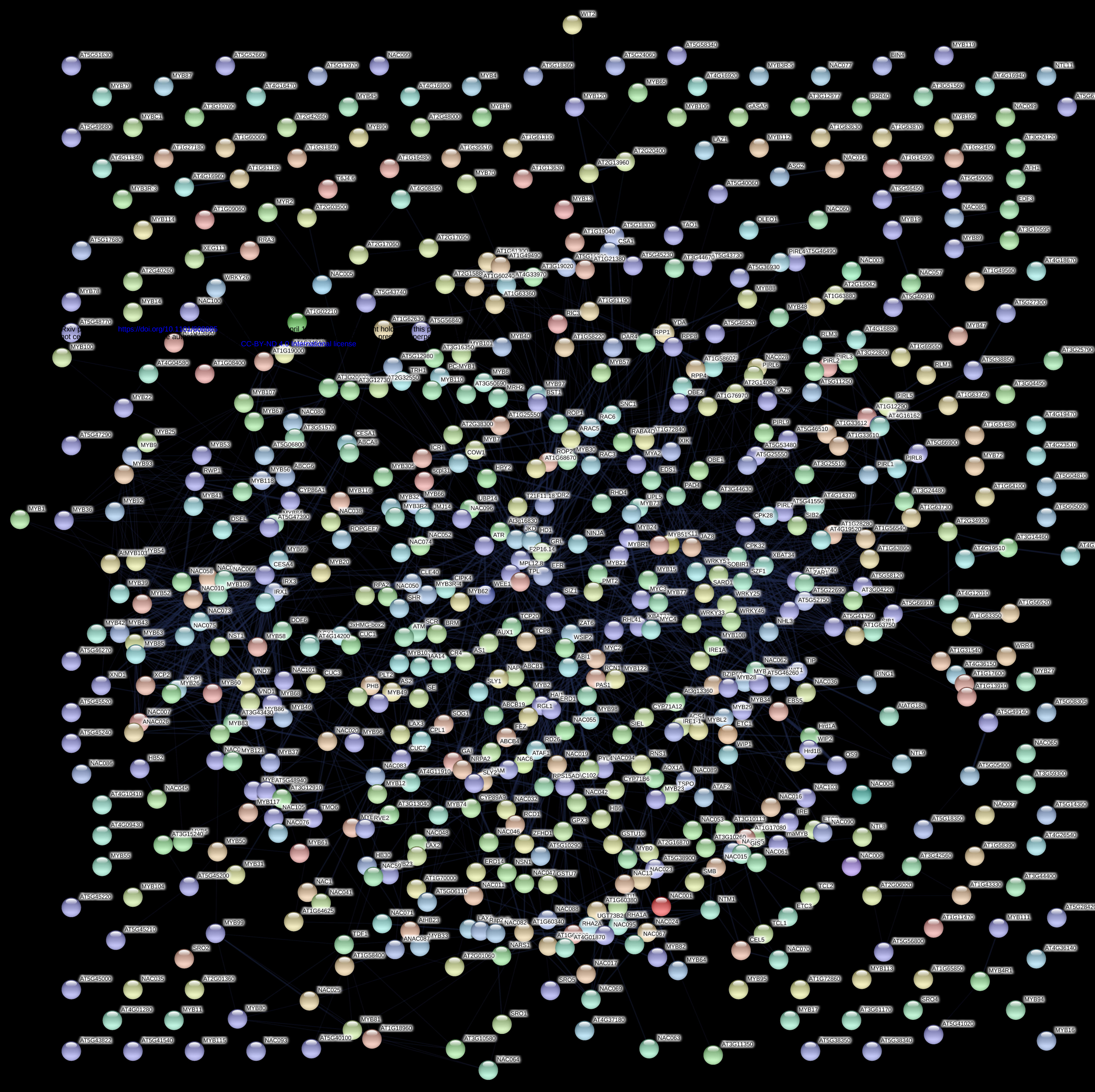
1087

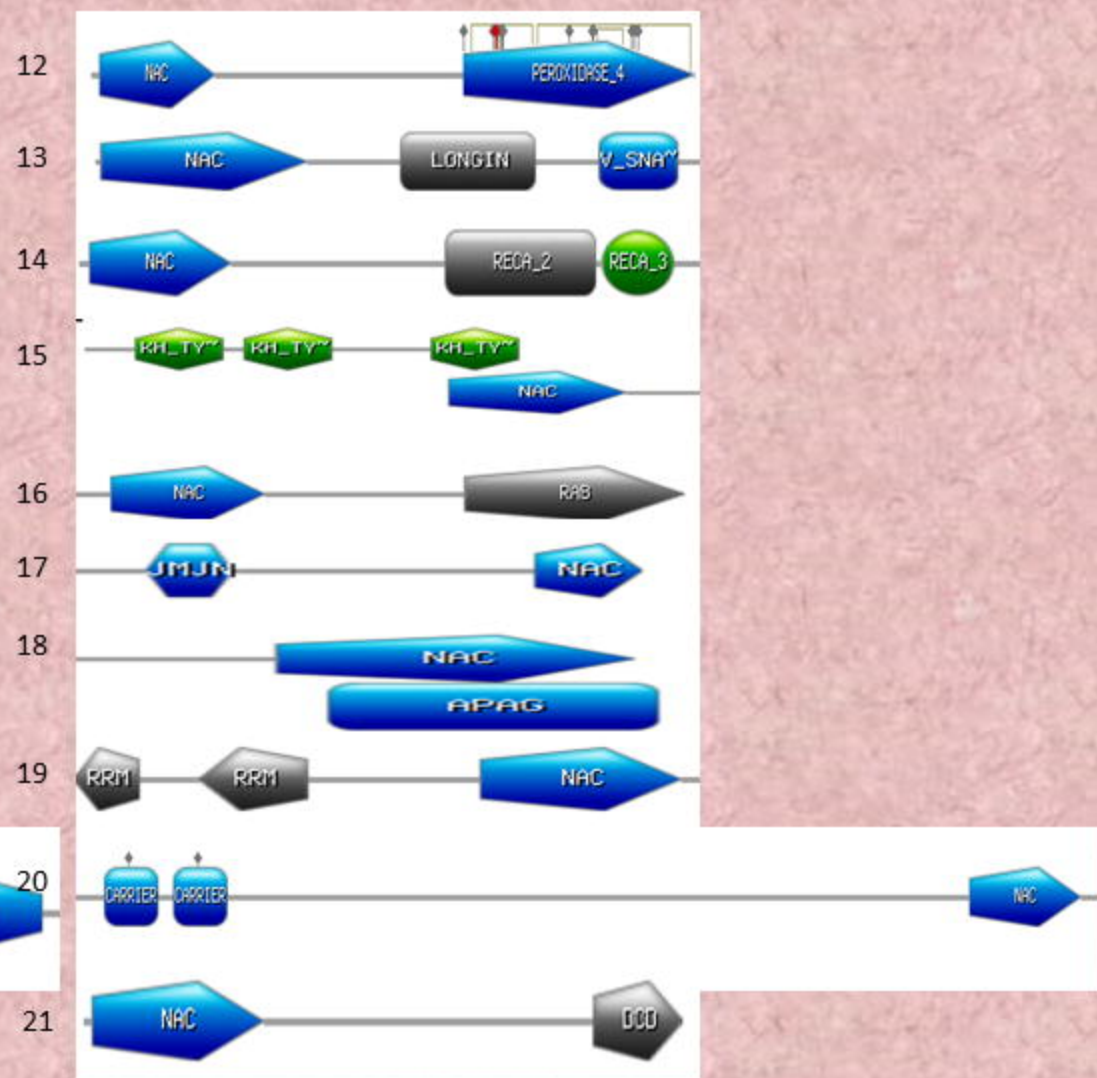
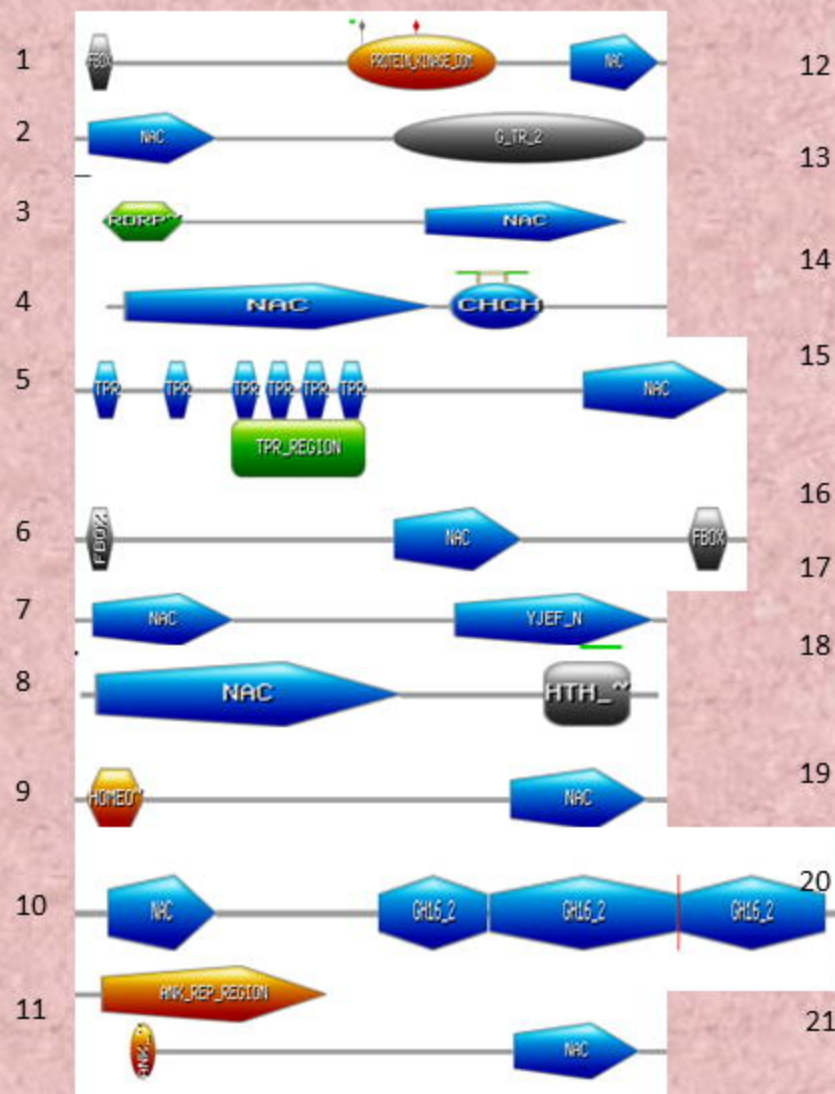
1088



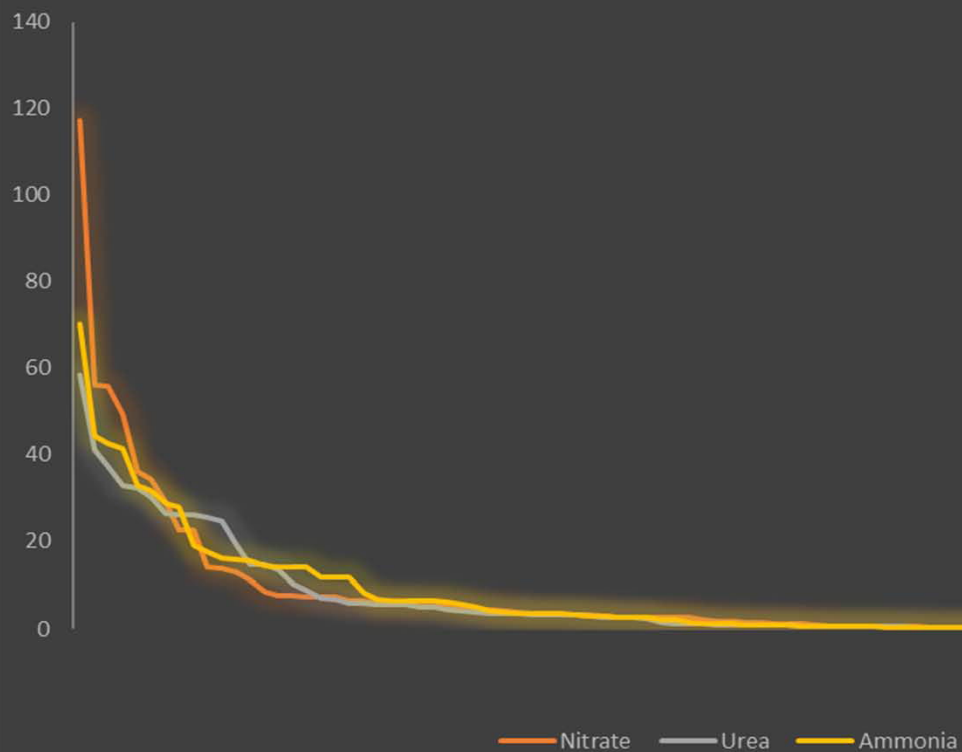
Distribution of isoelectric points (pI) of NAC TFs







Gene Expression of NAC TFs in leaf



Gene expression of NAC TF in roots

FPKM

180
160
140
120
100
80
60
40
20
0

Ammonia Nitrate Urea

Genes

