Genomics, Molecular and Evolutionary Perspective of NAC Transcription Factors

Tapan Kumar Mohanta[1]*[s], Dhananjay Yadav[2s], Adil Khan[1], Abeer Hashem[3,4], Baby Tabassum[5], Abdul Latif Khan[1], Eslayed Fathi Abda_Allah[6], Ahmed Al-Harrasi[1]*

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

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

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

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

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

[6]Plant 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 genes [1–4]. The dimeric architecture of NAC proteins can remain stable even at a concentration of 5M NaCl [4]. 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-helices [4]. 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 nature [4]. 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 activity [10]. For example, the NRD domain can suppress the transcriptional activity of Dof, WRKY, and APETALA 2/dehydration responsive elements (AP2/DRE) TFs [10].

78   Studies indicate that the diverse C-terminal domain contains a transcription regulatory region
79   (TRR) which has several group-specific motifs that can activate or repress transcription activity
80   [11–14]. The diverse C-terminal region imparts differences in the function of individual NAC
81   proteins by regulating the interaction of NAC TFs with diverse target proteins. Although the C-
82   terminal region of NAC TFs is diverse, it also contains group-specific conserved motifs [15].
83   Although the diverse aspects of NAC TFs have been studied, most were conducted within
84   individual plant species. A detailed comparative study of the genomic, molecular biology, and
85   evolution of NAC TFs has not been conducted. Therefore, a comprehensive analysis of NAC TFs
86   is presented in the current study.

Results and discussion

NAC transcription factors exhibit diverse genomic and biochemical features

89   Advancements in genome sequencing technology have enabled the discovery of the genomic
90   details of hundreds of plant species. The availability this genome sequence data allowed us to
91   characterize the genomic details of NAC TFs in diverse plant species. The presence of NAC TFs
92   in 160 species (18774 NAC sequences) was identified and served as the basis of the conducted
93   analyses. Comparisons of NAC sequences revealed that *Brassica napus* has the highest number
94   (410) of NAC TFs, while the pteridophyte plant, *Marchantia polymorpha,* was found to contain
95   the lowest number (9) (Table 1). On average, monocot plants contain a higher (141.20) number of
96   NAC TFs relative to dicot plants (125.56). Except for *Hordeum vulgare* (76), *Saccharum*
97   *officinarum* (44), and *Zostera marina* (62) all other monocot species possess more than one
98   hundred NAC TFs each (Table 1). Lower eukaryotic plants, bryophytes and pteridophytes also
99   possess NAC TFs. In addition, the algal species, *Klebsormidium flaccidum,* also contains NAC
100   TFs and this finding represents the first report of NAC TFs in algae (Table 1). A NAC TF in
101   *Trifolium pratense* (Tp57577_TGAC_v2_mRNA14116) was found to be the largest NAC TF,
102   comprising 3101 amino acids, while a NAC TF in *Fragaria x ananassa*
103   (FANhyb_icon00034378_a.1.g00001.1) was found to be the smallest NAC TF, comprising only
104   25 amino acids. Although it only contains a 25 amino acid sequence, it still encodes a NAC
105   domain. Typically, NAC TFs contain a single NAC domain located near the N-terminal region of
106   the protein. The current analysis, however, also identified NAC TFs with two NAC domains. At
107   least 77 of the 160 studied species were found to contain two NAC domains (Table 1).

108   Multiple sequence alignment revealed the presence of a conserved consensus sequence at
109   the N-terminus. The major conserved consensus sequences are P-G-F-R-F-H-P-T-D-D/E-L-I/V,
110   Y-L-x_2-K, D-L-x-K-x_2-P-W-x-L-P, E-W-Y-F-F, G-Y-W-K-A/T-T-G-x-D-x $_{1-2}$-I/V, G-x-K-K-x-
111   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,
112   G-Y-W-K, and M-H-E-Y are the conserved motifs most observed. The D-D/E-L motif is a
113   characteristic feature of the calcium-binding motifs present in the EF-hand of calcium-dependent
114   protein kinases and the presence of this motif in NAC TFs indicates that they have the potential to
115   regulate Ca^{2+} signalling events in cells [16]. The D-D-E/E motif is located in the β' sheet whereas
116   the Y-L-x_2-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
117   consensus sequences contain the positively charged amino acids Lys (L) and Arg (K) that can
118   bind to negatively charged DNA. Welner et al. (2012) published the crystal structure of

119 ANAC019 and reported that Y^{94}-W-K-A-T-G-T-D in β3, I^{11}-K-K-A-L-V-F-Y of β4, K^{123}-A-P-K-
120 G-T-K-T-N-W in the loop between β4 and β5, and I^{133}-M-H-E-Y-R of β5 and Y^{160}-K-K-Q at the
121 C-terminal end are located close to the bound DNA and are associated with DNA binding activity
122 [17]. They reported that Y^{94}-W-K-A-T-G-T-D is responsible for the specific recognition of DNA
123 and binds at the major groove within DNA, whereas I^{11}-K-K-A-L-V-F-Y, K^{123}-A-P-K-G-T-K-T-
124 N-W, I^{133}-M-H-E-Y-R, and Y^{160}-K-K-Q bind to the backbone of the DNA molecule and provide
125 affinity for DNA binding activity [17]. In the present analysis of 160 plant species, the identification
126 of the conserved consensus sequences G-Y-W-K-A/T-T-G-x-D-x_{1-2}-I/V, G-x-K-K-x-L-V-F-Y,
127 and T-x-W-x-M-H-E-Y is in agreement with Welner et al (2012); suggesting that NAC TFs
128 contain conserved consensus sequences for specific DNA recognition and increasing the affinity
129 for DNA binding.

130 Hao et al., (2010) reported that the D subunit of NAC TFs contain a hydrophobic L-V-F-Y
131 amino acid motif that suppresses WRKY, Dof, and APETALA2 transcriptional regulators. This
132 suggests that NAC TFs may also function as a negative regulator of transcription [10]. Our
133 sequence alignment, however, revealed that NAC TF family proteins in many different and
134 diverse plant species possess a conserved hydrophobic L-V-F-Y motif. As reported by Hao et al.,
135 (2010), all of the NAC TFs have the potential to suppress the transcriptional activity of WRKY,
136 Dof, and APETALA 2/dehydration responsive element TFs; however, it is highly unlikely that an
137 organism could regulate transcriptional events in a specific and sustained manner if NAC will
138 conduct transcriptional repression.

139 The molecular weight of NAC TFs ranged from 346.46 kilodaltons (kDa) (*Trifolium*
140 *pratense*_Tp57577_TGAC_v2) to 2.94 kDa (*Fragaria* x
141 *ananassa*_FANhyb_icon00034378_a.1.g00001.1) (Figure 1). Among the studied NAC TFs, only
142 10 NAC proteins have a molecular weight (MW) more than 200 kDa and approximately 99 are
143 between 100 to 200 kDa. The MW of the majority of the NAC proteins range between 40 to 55
144 kDa (Figure 1). The Isoelectric point (pI) of the NAC proteins ranged from 11.47
145 (Brast01G304500.1.p, (*Brachypodium stacei*) to 3.60 (ObartAA03S_FGP19036, *Oryza barthii*).
146 The majority of the NAC TFs fell within a pI rage of 5-8 (Figure 2). Among the 18774 analysed
147 NAC TFs, the pI of 99 of them were ≥ 10. Approximately 69.28% of the NAC TFs had a pI that
148 was in an acidic range, whereas the remaining 30.72% had a pI within in a basic range. A protein
149 with a pH below the pI carries a net positive charge, whereas a protein with a pH above the pI
150 carries a net negative charge. The pI of a protein determines its transport, solubility, and sub-
151 cellular localization [18–20]. Biomembranes, such as those surrounding the nucleus, are negatively
152 charged; as a result, positively charged (acidic pI) NAC TFs are readily attracted to the nuclear
153 membrane and subsequently transported into the nucleus to function in transcriptional regulation.
154 There are, however, approximately 30.72% NAC TFs that possess a basic Pi; suggesting that they
155 are localized in the cytosol or plasma membrane of the cell. The major role of TFs is to bind to
156 specific DNA sequences to regulate transcription. The majority of proteins have either an acidic or
157 basic pI and those with a neutral pI close to 7.4 are few because proteins tend to be insoluble,
158 unreactive, and unstable at a pH close to its pI. This is the main reason why among the 18774
159 NAC TFs analysed, only two (XP_010925972.1, *Elaeis guineensis*; Lus10008200, *Linum*
160 *usitatissimum)* had a pI 7.4. The existence of NAC proteins with a pI above 10 led us to speculate
161 whether these TFs function while attached to a transmembrane domain. Therefore, additional

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

164 ### *NAC TFs are membrane bound*

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

177 Results indicated that at least 2190 (8.57%) NAC TFs possess a transmembrane domain.
178 Transmembrane domains were found at both the N- and C-terminal ends of NAC proteins. In the
179 majority of the cases, however, the transmembrane domain was located towards the C-terminal
180 end. Seo et al., (2008) indicated the presence of a transmembrane domain in TFs and suggested
181 that transmembrane domain functions through two proteolytic mechanisms, commonly known as
182 regulated ubiquitin/proteasome-dependent (RUP) and regulated intramenbrane proteolysis (RIP)
183 [23,24]. The bZIP plant TF is present as an integral membrane protein associated with stress response
184 in the endoplasmic reticulum (ER) [25–28]. Studies suggest that the majority of membrane bound TFs
185 are associated with the ER and a membrane bound TF was also found to be involved in cell
186 division [29,30]. At least 10% of the TFs in *Arabidopsis thaliana* have been reported to be
187 transmembrane bound [30]. The collective evidence clearly indicates that membrane-mediated
188 transcriptional regulation is a common stress response and that NAC TFs play a vital role in stress
189 resistance in the ER. Therefore, these membrane-bound NAC TFs can be of great importance for
190 the manipulation of stress resistance using biotechnology.

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

192 The import of NAC TFs into the nucleus is mediated by nuclear membrane-bound importins and
193 exportins that form a ternary complex consisting of importin α, importin β1, and a cargo molecule.
194 Importin α serve as an adaptor molecule of importin β1 and recognises the nuclear localization
195 signal (NLS) of the cargo protein needing to be imported. Importin β1 and β2, however, also
196 recognize the NLS directly and bind to the cargo protein. Although the NLS of TFs have been
197 widely studied in the animal kingdom, their study in plants has been more restricted. Therefore,
198 the NLS of NAC TFs was examined in the current study. Results indicate that NAC TFs contain
199 diverse NLS. The NLS were found in the N- and C-terminal regions of NAC TF proteins. Some
200 NAC TFs were found to contain only one NLS whereas other contain multiple NLS. At least 3579
201 of the total NAC TFs analysed were found to contain either one or multiple NLS. More
202 specifically, 2604 NAC TFs were found to possess only one NLS at the N-terminal end of the

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

206   NLS motifs are rich in positively charged amino acids and bind to importin α to be
207 imported into the nucleus. The NLS motifs are classified as monopartite or bipartite. A
208 monopartite NLS contains a single cluster of positively charged amino acids and are grouped into
209 two subclasses, class-I and class-II. Class-I possesses four consecutives positively charged amino
210 acids and class-II contains three positively charged amino acids, represented by K(K/R)-x-K/R;
211 where x represents any amino acid that is present after two basic amino acids. Bipartite NLS
212 motifs contain two clusters of positively charged amino acids separated by a 10-12 amino acid
213 linker sequence. Bipartite NLS motifs are characterised by the consensus sequence K-R-P-A-A-T-
214 K-K-A-G-Q-A-K-K-K-K. In addition to monopartite and bipartite NLS motifs, importin α also
215 recognises non-canonical NLS motifs. Non-canonical NLS motifs are longer and considerably
216 variable relative to monopartite and bipartite NLS motifs and are classified as class-III and class-
217 IV NLS. Non-canonical NLS motifs are usually present in the C-terminal end and bind with
218 importin β2. Class-III and class-IV NLS motifs contain K-R-x(W/F/Y)-x_2-A-F and (P/R)-x_2-K-R-
219 (K/R) consensus sequences, respectively. We identified at least 1702 unique NLS consensus
220 sequences in the N-terminal region of NAC TFs. The monopartite class I NLS motifs were found
221 to contain more than four consecutive basic amino acids with the number of their consecutive
222 basic amino acids ranging from four to fourteen (K-K-K-K-K-K-K-K-K-K-K-K-K-K). The
223 bipartite NLS motifs contain two clusters of consecutive basic amino acids separated by up to
224 twenty-four linker amino acids (K-K-K-x_3-R- x_2-R- x_4-K- x_3-K- x_3-K-x-K- x_2-R-K-K).

225   The non-canonical NLS motifs contain at least six centrally-located, positively charged
226 amino acids (K-x-R-R-R-P-R-R-x_2-R-K) flanked by positively charged amino acids on both sides.
227 Our analysis of the N-terminal NLS of NAC TFs, however, did not identify any NAC TFs
228 containing this consensus sequence. Instead, several new variants of this consensus sequence were
229 identified with multiple clusters of positively charged amino acids. These NLS were designated as
230 multipartite NLS motifs. A few examples of the multipartite NLS include, K-K-K-K-x_7-K-K-K-
231 K-x_7-K-K-K-K, K-K-K-K-x-K-x_5-K-x-K-K-x_7-K-K-K-K-x_2-K-K-K, K-K-K-x_2-K-K-x-K-x_5-K-
232 x_4-K-K-K-R-x-K-R-K-x-K-x_4-K-K-K-R-K-K, K-K-R-x-R-K-x_2-K-x-K-x_2-K-K-K-x-RK-x_2-K-R-
233 R-x_2-K-K-K-x-R, K-K-R-x-R-K-x_2-K-x-K- x_2-K-K-R-x-R-K-x_2-K-x-K-x_2-K-K-R-x-R-K-x_2-K-
234 x_2-K-x-K-x-R, K-x_2-K-K-K-x_3-K-K-K-K-K-x-K-x_8-K-x$_9$-K-x_2-K-K-R-x_2-K-K-K-K-x-K, K-x_2-K-
235 K-K-x_3-K-x-K-K-K-x-K-K-K-x_2-K-K-K-x-K, R-K-R-x-R-x-R-K-K-x_2-K-x-K-K-K-R-x_2-K-x_2-
236 KK-x_2-R-R-K-x_2-K, and R-K-R-x-R-x-R-x_2-K-x-K-K-K-R-x_2-K-x4-K-R-x_2-R-R-K-x-K-x_2-R.
237 Much of the diversity of NLS motifs is associated with the sequence of the variable linker amino
238 acids. In our analysis, we removed the linker amino acid sequences, represented as x, to obtain a
239 more concise picture of NLS diversity. Removing the linker amino acids present in monopartite,
240 bipartite, and multipartite NLS motifs resulted in the identification of 97 different NLS consensus
241 sequences in the N-terminal region of NAC TFs. The unique NLS signal sequences were R-K-R-
242 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-
243 R-R, K-K-R-K, K-K-R-K-R, K-R-R-K, R-K-R-R-R, R-K-R-R, K-K-K-K-K, R-R-K-R, K-R-K-R-
244 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-
245 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-

6

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, R-R-R-R-R-R, R-R-R-R-R-K, K-K-K-K-K-R, K-K-K-K-K-K, K-K-K-K-K-K, K-K-K-R-R-R-R, K-K-K-K, K-R-K-R-R, K-R-R-K-R, R-K-K-K-R, R-K-K-K-K, R-K-R-R-K-R, R-R-K-K-K, R-R-K-K, R-R-R-K-R, R-R-R-R-K-K, 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-R-R, K-K-K-K-R-K-R, K-K-K-R-K-R, K-K-K-R-R-R, K-K-K-R-R-R-R-R, K-R-R-R-R-R-R-R-R, K-R-K-K-R, K-R-K-R-K-R-K, K-R-R-K-K, R-K-K-K-K-R, R-K-K-R-K-R, R-K-K-R-K-R, R-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-R-K, R-R-K-R-K, R-R-R-K-R, R-R-R-R-K-K, R-R-R-R-K-R, R-R-R-R-R-R, and 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-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-x_4-K-x-K-K-K-R-x_3-K-x_3-K-K-x_3-R-R-K-x_2-K, R-R-R-x_4-K-K-x_6-R-x_2-R-x_2-R-R-x_4-R-R-R-x_6-R-x_2-R-R-x_9-R-R-R-R-R-R-R-x_2-R-R, K-K-K-x_4-K-K-x-K-x_5-K-x_4-K-K-K-R-x-K-R-K-x-K-x_4-K-K-K-R-K-K, K-K-R-x_4-K-x_2-K-x-K-x_2-K-K-R-x-R-K-x_4-K-x_2-K-x-K-K-R-x-R-K-x_4-K-x_2-K-x-K-x-R, K-K-R-x-R-K-x_2-K-x-K-x_2-K-K-K-x-R-K-x_2-K-R-R-x_2-K-K-K-x-R, K-K-R-x-R-K-x_2-K-x-K-x_2-K-K-R-x-R-K-x_2-K-x-K-x_2-K-K-R, R-K-R-x-R-x_3-K-K-R-R-x_2-K-x_9-K-x_4-R-x-K-x_2-R-x-R-R-x_5-K-K-R, R-K-R-x-R-x-R-x_5-K-x-K-K-K-R-x_3-K-x_4-K-R-x_2-R-R-K, R-R-x-R-R-R-x-R-R-x_8-R-x_6-R-R-x_5-R-R-R-x-R-x_5-R-x_8-R-R-R-R, R-R-x-R-R-x-R-x-R-R-R-x_9-R-x_2-R-R-K-R-K-x-R-x_4-R-R-R-R-R-R-x_4-R-K, R-x-R-R-R-R-x_6-R-x_{11}-R-x_8-R-R-x_3-R-R-R-x_2-R-R-x-R-x-R-x_6-R-R-R-R-R-x_4-R-R-x_2-R, R-x-R-R-x-x_3-K-R-R-R-x_2-R-x-R-R-x-R-x-R-x_7-R-x_3-R-R-R-x_7-R-x_2-R-R-R-R, R-x-R-x-R-R-R-x_3-R-R-R-x_3-R-x-R-x_2-R-x_4-R-R-R-x_5-R-K-x-R-x_3-R-R- x_{13}-R-R-x-K-x_5-R-R-x_6-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-

289 R-K-R-R, and R-R-R-R-K-R. Similarly, nine NLS consensus sequences were unique to the C-
290 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-
291 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
292 NLS have been reported to be associated with importin α subunit [31]. To the best of our knowledge,
293 this is the first report describing such a high level of diversity and dynamism in the NLS
294 consensus sequences of NAC TFs and plant transcription factors in general. This is also the first
295 report of the presence of unique NLSs in the N-and C-terminal regions of NAC TFs.

296 Several nuclear-associated proteins contain NLS, as well as nuclear export signals (NESs).
297 Proteins that perform their function within the nucleus need to be exported out of the nucleus and
298 into the cytoplasm to undergo proteosomal degradation. Therefore, a NES is required in addition
299 to an NLS. A Ran-GTP complex binds directly to an NES and mediates the nuclear export process
300 of cargo molecules [32]. NES sequences contain a hydrophobic, conserved L-V-F-Y (substitute L-
301 V/I-F-M) motif separated by variable linker amino acids at both ends [33]. The presence of an L-V-
302 F-Y motif in all NAC proteins, suggests that all NAC proteins have the potential to be exported
303 out of the nucleus. Hao et al. (2010), however, reported that the hydrophobic L-V-F-Y motif
304 functions as a transcriptional repressor of WRKY, Dof, and APETALA TFs. If the L-V-F-Y motif
305 acts as a transcriptional repressor, then the transcriptional activity of these TFs would be lost;
306 resulting in an unstable genome. Therefore, we suggest that the L-V-F-Y motifs do not function as
307 a transcriptional repressor but rather as a NES.

308 *NAC TFs possess a complex interactome network*

309 The interacting partner of a protein can provide significant information about its potential function
310 and an entire protein-protein interactome network can greatly assist in unravelling the signalling
311 cascade of the proteins. Different cascades are interlinked in signalling systems and form intricate
312 constellations that provide information about cell response and function. Thus, the interactome
313 network of NAC TFs in *A. thaliana* were explored. The presence of a dynamic network was
314 revealed and a diverse set of interacting protein partners of NAC TFs were identified (Figure 3,
315 Table 2). Results indicated that NAC TFs interact with RNS1 (ribonuclease 1), ERD14 (early
316 responsive to dehydration 14), VND1 (vascular related NAC domain 1), VND7, GAI (gibberellins
317 inducible), ZF-HD1 (zinc finger homeodomain 1), TCP8 (Teosinte branched, Cycloidea, and
318 Proliferating cell nuclear antigen factor 8), TCP20, CPL1 (C-terminal domain phosphatise-like1),
319 RHA1A (ring H2 finger H1A), RHA2A, SHR (short root), PHB (phabulosa), PLT2 (plethora 2),
320 MYB59, HB23 (homeobox 23), HB30, NAC1, NAC6, NAC19, NAC32, NAC41, NAC45,
321 NAC50, NAC52, NAC76, NAC83, NAC97, NAC101, NAC105, IAA14 (auxin responsive protein
322 indole3-acetic acid), HAI1 (protein phosphatase), ABI1 (ABA insensitive 1), RVE2 (reveille),
323 PYL4 (PYR-like 4), BRM (brahma), HB52, RCD1 (radical-induced cell death 1), JMJ14 (jumonji
324 14), TPL (TOPLESS), F2P16 (TOPLESS related), TOPLESS, RING/U-box, ZF-domain (zinc
325 finger domain), SRO1 (similar to RCD 1), CUC2 (cup shaped cotyledon 2), PAS1 (pasticcino 1),
326 TI1 (defensin like 1), TSPO (tryptophan rich sensory protein), TIP2.2 (TCV-interacting protein
327 2.2), TIP3.1, T21F11.18 (TOPLESS related), LRR (leucine rich repeat), RPA2 (replicon protein
328 A2), and VR-NAC (Table 2). The interaction of NAC TF proteins with the diverse number of
329 listed proteins has been experimentally validated in *A. thaliana*. The interactome network includes
330 stress responsive proteins, other transcription factors, hormonal signalling proteins, protein

331 phosphatases, and defense related proteins. In addition to experimentally-validated interacting
332 proteins, bioinformatic mining indicated that NAC TFs also interact with several other proteins
333 (Table 2). Some of the identified interacting proteins were MYB, NAC, NTL (NTL2-like),
334 UBC30 (ubiquitin conjugating enzyme 30), ATM (Ataxia-Telangiectasia mutated), ATR
335 (serine/threonine kinase ATR), KNAT (knox tail), AOX1A (alternative oxidase 1A), ASG2
336 (altered seed germination 2), NYE (nonyellowing), CPL, TMO6 (target of monopteros 6),
337 RHA2A (ring H2-finger A2A), XCP (xylem cysteine peptidase), DBP (downstream auxin
338 binding), WAK5 (wall associated kinase 5), RCD, CYP71A25 (cytochrome P71A25), Stay green,
339 ERF (ethylene responsive transcription factor), PPR (pentatricopeptide), WOX (Wuschel related
340 homeobox), PPD6 (psbP-domain protein 6), MFDX (mitochondrial ferredoxin), LEA (late
341 embryogenesis abundant), IRX (irregular xylem), CESA4 (cellulose synthase A4), SCRL20
342 (SCR-like 20), PIP1-5 (aquaporin), PUP4 (purine permease 4), XERO1 (dehydrin xero 1), SWAP
343 (suppression of white apricot), TIR-NBS, NBS-LRR, PAS, CHI (chitinase), MC5 (metacaspase
344 5), XCP (xylem cysteine peptidase), RNS, LAC (laccase), TIR-NBS (toll/interleukin receptor-
345 nucleotide binding site), NBS-LRR, DTA4 (downstream target 4), BAG6 (BCL-2 associated
346 anthogene 6), and others as well (Table 2). Some NAC TFs are co-expressed with other proteins.
347 These include DREB2A, XCP1, XCP2, ATM, ATR, MYB63, MYB69, MYB83, IRX1, AOX1A,
348 RCD1, ASG2, ERD1, TMO6, DOF6, SHR, PLT2GRP20, CYP86C4, VND7, NAC6, NAC32,
349 NAC97, NAC19, NAC102, HAI1, WRKY33, WRKY46, WRKY53 and others (Table 2). Some of
350 the NAC TFs directly interact with the interacting partner while others form complexes and
351 appear to play an indirect role. NAC TFs act as a negative regulator of ABA signalling, while
352 they induce JA/ET-associated marker genes [34].

353   The expression of several of *NAC* genes are either up- or down-regulated by auxin,
354 ethylene, or ABA, suggesting that NAC TFs play a role in plant hormonal signalling [35–37]. One of
355 the most challenging aspects of a protein-protein interactome network is that the interaction can
356 vary depending upon the cell and its environment [38]. Therefore, it is necessary to investigate the
357 dynamic interactions of proteins in different cells and environmental conditions to completely
358 understand their interacting partner and the cellular function of the TF. NAC TFs regulate *ERD*
359 and *NCED* (ABA biosynthesis) genes through a direct interaction with their promoters [39,40]. NAC
360 TFs (ANAC019, ANAC055, and ANAC072) interact with ERD1 which encodes a Clp protease
361 regulatory subunit [41]. The over expression of one of these three NAC TFs, however, did not
362 induce the up-regulation of ERD1 because the induction of ERD1 depends on the co-expression
363 of a zinc finger homeodomain TF, ZFHD1 [41]. ANAC019 and ANAC055 interact with ABI
364 (abscisic acid insensitive), and at least five MYB TFs can bind to the NAC TF promoter region
365 [42,43]. In this case, the NAC DNA binding domain mediates the interaction with RHA2A and
366 ZFHD1 [43].

NAC TFs encodes chimeric proteins and contain multiple binding sites

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

373 kinase, PI3_4_kinase_3, EF-hands (elongation factor), CRM, peptidase A1, WRKY, cytochrome
374 B561, OFOF, FFO, Dna_J2, ZF_B, TIR, LRR, CS, F-box, IQ, PPC, ENT, ABC_TM1F,
375 RWP_RK, PB1, PABC, ACT, INTEGRA, RESPO, JMJC, SAM, BRX, G_TR_2, RORP, CHCH,
376 TPR, YJEF_N, HTH, HOMEO, GH16, ANK_REP_REGION, Peroxidase, LONGIN, V_SNA,
377 RECA_2, KH_TY, APAG, RRM, carrier, and a DCO domain. At least four NAC TFs from *A.*
378 *thaliana*, ten from *B. napus*, four from *B. rapa*, two from *M. domestica*, four from *P. virgatum*, 17
379 from *C. sativa*, eight from *D. oligosanthes*, eight from *E. tef*, and five from *L. perrieri* were found
380 to possess 2 NAC domains (Supplementary Table 1). NAC TFs in several other species were also
381 found to contain two NAC domains (Supplementary Table1). When two NAC domains were
382 present, both domains were located towards the N-terminal end. NAC TFs of at least three
383 species, *O. rufipogon*, *B. stacei*, and *Camelina sativa* were found to possess three NAC domains
384 whereas the NAC TFs in *A. lyrata* (gene id: 338342), *C. sativa* (Csa16g052260.1), and *E. tef*
385 (462951506) were found to possess four NAC domains (Figure 4).

386   Other chimeric domains were also identified in different regions of the NAC protein. PPR
387 domains were found in both the N-terminal or C-terminal region, a protein kinase domain was
388 found upstream to the NAC domain, and a NAC domain was found to be adjacent to a protein
389 kinase and EF-hand domain. Additionally, a protein kinase domain was found to be followed by
390 either a NAC and a CRM domain, a NAC domain was followed by a peptidase_A1 domain, a
391 NAC domain was followed by the presence of a WRKY domain, a cytochrome_B561 domain was
392 followed by either a NAC domain and a CRM domain, a DFDF and a FFD domain were followed
393 by a cytochrome B and NAC domain, a DNAJ_2 domain was followed by a NAC domain, a
394 DNAJ_2 domain was followed by a NAC and ZF_B domain, a NAC domain was followed by
395 TIR, LRR and CS domains, a NAC domain was followed by a TIR domain, an F-box domain was
396 followed by a NAC domain, an IQ domain was followed by a NAC domain, a NAC domain was
397 followed by a ZF_B domain, an EF-hand domain was followed by a NAC domain, a NAC domain
398 was followed by a PPC domain, an ENT domain was followed by a NAC domain, a NAC domain
399 was followed by an ABC_TM1F, a NAC domain was followed by CRM domain, a NAC domain
400 was followed by a RWP_RK and a PB1 domain, a NAC domain was followed by three ACT
401 domains, a NAC domain was followed by a PABC domain, a NAC domain was followed by an
402 INTEGRA domain, a RESPO domain was followed by a NAC domain, a NAC domain was
403 followed by a JMJN and a JMJC domain, a SAM domain was followed by a NAC domain, a BRX
404 domain was followed by a NAC domain, a NAC domain was followed by ZF_B, NAC and ZF_B
405 domains, an F-box and protein kinase domain was followed by a NAC domain, a NAC domain
406 was followed by a G_TR_2 domain, an RDRP domain was followed by a NAC domain, a NAC
407 domain was followed by a CHCH domain, a TPR domain followed by a NAC domain, an F-box
408 domain followed by NAC and F-box domains, a NAC domain was followed by a YJEF_N
409 domain, a NAC domain was followed by a HTH domain, a homeobox domain was followed by a
410 NAC domain, a NAC domain was followed by three GH16_2 domains, an ANK repeat region
411 was followed by a NAC domain, a NAC domain was followed by a peroxidase domain, a NAC
412 domain was followed by a LONGIN and V_SNA domain, a NAC domain was followed by
413 RECA_2 and RECA_3 domains, a KH domain was followed by a NAC domain, a NAC domain
414 was followed by a RAB domain, a JMJN domain was followed by a NAC domain, a NAC domain
415 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 response [44–52]. 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 proteins [53]. 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 phosphatise 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

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

491   *NAC TFs are involved in diverse cellular processes*

492   NAC TFs are known to possess diverse chimeric domains, as a result, it is more than likely that
493   NAC TFs are also involved in the regulation of diverse cellular pathways and cellular processes.
494   To help substantiate this premise, the interactome associated with NAC TFs in *A. thaliana* was
495   analysed. Results indicated that NAC TFs are potentially involved in a least 289 different cellular
496   processes and pathways (Supplementary Table 3). The majority are related to cell, tissue, and
497   organ (root, stem, meristem) development, as well as signalling processes. Several NAC TFs also
498   appear to be associated with phytohormone signalling, including auxin, gibberellin, jasmonic acid,
499   and salicylic acid signalling pathways. NAC TFs were also found to be associated with pathways
500   involved in the response to bacterial, fungal, UV, heat and other biotic and abiotic stresses
501   (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 tolerance [57]. 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 (0), 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

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

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

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

578 *Capsella grandiflora* is a close relative of *Arabidopsis thaliana* and is predicted to be the
579 progenitor of *Capsella bursa-pastoris. Capsella grandiflora* is a self-pollinating plant and is used
580 as a model organism in evolutionary studies and the change from self-incompatibility into self-
581 compatibility. The genomic consequences of the evolution of selfing, however, is poorly
582 understood. *Capsella rubella*, a close relative of *Capsella grandiflora,* that evolved self-
583 compatibility 200,000 years ago [62] also exhibits a high rate of transversion from adenine to
584 thiamine (11.19). Thus, the higher rate of transversion from adenine to thiamine in *Capsella*

585 *grandiflora* and *Capsella rubella* may be a possible factor in the evolution of self-pollination.
586 Higher rates of transversion were also found in *Solanum pimpinellifolium* (11.4) and *Castanea*
587 *mollissima* ((11.31) Chinese chestnut). *Solanum pimpinellifolium* is self-pollinating and exhibits
588 high levels of stress tolerance [63]. *Castanea mollissima* has evolved over a period of time in
589 coexistence with chestnut blight and is resistant to the pathogen. This indicates that higher rates of
590 transversion from adenine to thiamine and vice versa are associated with self-pollination and
591 stress tolerance in plants. The highest rate of substitution from guanine to cytosine and vice versa
592 was found in *Arachis hypogaea* (11.07), and *Camelina sativa* (11.46), respectively (Table 4). The
593 lowest rate of substitution from adenine to thiamine and vice versa was found in *Linum*
594 *usitatissimum* (3.72) and *Klebsormidium flaccidum* (6.67), respectively. Notably, the highest rate
595 of substitution from thiamine to cytosine was found in *Klebsormidium flaccidum* and the highest
596 rate of substitution from adenine to guanine was found in *Linum usitatissimum*. This indicates that
597 organisms which exhibit the highest rate of transition possess the lowest rate of transversion.

598 ### *NAC TFs evolved from orthologous ancestors*

599 A phylogenetic tree of NAC TFs was constructed to understand their evolutionary relationships. A
600 model selection was conducted before constructing the phylogenetic tree using the maximum
601 likelihood statistical method. The phylogenetic tree revealed the presence of at least seven
602 phylogenetic clustered orthologous (COGs) groups originating from a common, orthologous
603 ancestor (Figure 7). Each phylogenetic cluster was further divided into two or more sub-groups.
604 A phylogenetic tree of each individual species was subsequently constructed to examine
605 duplication and loss events in NAC TFs. The phylogenetic tree of each species was independently
606 reconciled with the collective species tree. This analysis indicated that NAC TFs in all of the
607 species were duplicated and no NAC TFs were found to be lost. This suggest that NAC TFs
608 evolved from common ancestors (orthology) and underwent numerous duplication events during
609 speciation (paralogy), which gave rise to diverse gene functions in plant development and growth.
610 We also checked for the presence of potential foreign or homologous sequences (xenologs) in
611 NAC TFs. No primary xenologs, sibling donor xenologs, sibling recipient xenologs, incompatible
612 xenologs, autoxenologs, or paraxenologs were identified in NAC TFs. Although the phylogenetic
613 tree indicates the evolution NAC TFs from common ancestors, none of the NAC genes in the
614 examined species were found to have been transferred from one species to another (Table 1).
615 Previous studies of NAC TFs in six plant species also reported a high level of duplication and
616 divergent evolution [64]. The expansion of TF families was associated with an increase in the
617 structural complexity of the organism [65]. Previous studies reported the lineage-specific grouping
618 of transcription factors [54,64]. The phylogenetic tree of NAC TFs also revealed the presence of
619 lineage-specific clustering as well. In a few cases, however, order-specific clustering of NAC TFs
620 was also observed. For example, NAC TFs in dicot species of the Brassica lineage, including *A.*
621 *thaliana, A. halleri, B. napus, B. rapa, R. sativus, R. raphanistrum, C. rubella, A. alpine,* and
622 others, grouped together. Similarly, NAC TFs in monocot plant species, including *O. sativa*, *O.*
623 *nivara, B. distachyon*, and others, also grouped together.

624

625

15

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 databases [66]. BLASTP and hidden Markov model were used to identify the NAC TFs in different species using AtNAC1 and AtBAC2 as the query sequences [67]. 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. 2.0 [70] 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 signals [71]. 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

16

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

Gene expression analysis

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

Construction of a phylogenetic tree

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

Analysis of transition and transversion rates

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

Analysis of gene deletion and duplication

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

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

705   **Data availability**

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

708   **Competing of interest**

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

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

18

731

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).

809 31. Kosugi, S. *et al.* Six classes of nuclear localization signals specific to different binding
810   grooves of importin α. *J. Biol. Chem.* **284,** 478–485 (2009).

811 32. Mattaj, I. W. & Englmeier, L. Nucleocytoplasmic Transport: The Soluble Phase. *Annu.*
812   *Rev. Biochem.* **67,** 265–306 (1998).

813 33. Kosugi, S., Hasebe, M., Tomita, M. & Yanagawa, H. Nuclear export signal consensus
814   sequences defined using a localization-based yeast selection system. *Traffic* **9,** 2053–2062
815   (2008).

816 34. Jensen, M. *et al.* Transcriptional regulation by an NAC (NAM–ATAF1,2–CUC2)
817   transcription factor attenuates ABA signalling for efficient basal defence towards Blumeria
818   graminis f. sp. hordei in Arabidopsis. *Plant J.* **56,** 867–880 (2008).

819 35. Xie, Q., Frugis, G., Colgan, D. & Chua, N.-H. Arabidopsis NAC1 transduces auxin signal
820   downstream of TIR1 to promote lateral root development. *Genes Dev.* **14,** 3024–3036
821   (2000).

822 36. He, X. *et al.* AtNAC2, a transcription factor downstream of ethylene and auxin signaling
823   pathways, is involved in salt stress response and lateral root development. *Plant J.* **44,** 903–
824   916 (2005).

825 37. Sperotto, R. A. *et al.* Identification of up-regulated genes in flag leaves during rice grain
826   filling and characterization of OsNAC5, a new ABA-dependent transcription factor. *Planta*
827   **230,** 985–1002 (2009).

828 38. Bonetta, L. Interactome under construction. *Nature* **468,** 851 (2010).

829 39. Tran, L.-S. P. *et al.* Isolation and Functional Analysis of Arabidopsis Stress-Inducible NAC
830   Transcription Factors That Bind to a Drought-Responsive cis-
831   Element in the early responsive to dehydration stress 1 Promoter.
832   *Plant Cell* **16,** 2481 LP-2498 (2004).

833 40. Jensen, M. K. *et al.* ATAF1 transcription factor directly regulates abscisic acid biosynthetic
834   gene NCED3 in Arabidopsis thaliana. *FEBS Open Bio* **3,** 321–327 (2013).

835 41. Kazuo, N., Tomohiro, K., Kazuko, Y.-S. & Kazuo, S. A nuclear gene, erd1, encoding a
836   chloroplast-targeted Clp protease regulatory subunit homolog is not only induced by water
837   stress but also developmentally up-regulated during senescence in Arabidopsis thaliana.
838   *Plant J.* **12,** 851–861 (2003).

839 42. Reeves, W. M., Lynch, T. J., Mobin, R. & Finkelstein, R. R. Direct targets of the
840   transcription factors ABA-Insensitive(ABI)4 and ABI5 reveal synergistic action by ABI4
841   and several bZIP ABA response factors. *Plant Mol. Biol.* **75,** 347–363 (2011).

842 43. Jensen, M. K. & Skriver, K. NAC transcription factor gene regulatory and protein-protein
843   interaction networks in plant stress responses and senescence. *IUBMB Life* **66,** 156–166
844   (2014).

845 44. Jingjing, J. *et al.* WRKY transcription factors in plant responses to stresses. *J. Integr. Plant*
846   *Biol.* **59,** 86–101 (2016).

847 45. Pandey, S. P. & Somssich, I. E. The Role of WRKY Transcription Factors in Plant
848   Immunity. *Plant Physiol.* **150,** 1648–1655 (2009).

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

888   mating system evolution. *Nat. Genet.* **45,** 831–5 (2013).

889   63.   Bolger, A. *et al.* The genome of the stress-tolerant wild tomato species Solanum pennellii.
890   *Nat. Genet.* **46,** 1034 (2014).

891   64.   Jin, X. *et al.* Divergent Evolutionary Patterns of NAC Transcription Factors Are Associated
892   with Diversification and Gene Duplications in Angiosperm. *Front. Plant Sci.* **8,** 1156
893   (2017).

894   65.   Schmitz, J. F., Zimmer, F. & Bornberg-Bauer, E. Mechanisms of transcription factor
895   evolution in Metazoa. *Nucleic Acids Res.* **44,** 6287–6297 (2016).

896   66.   Goodstein, D. M. *et al.* Phytozome: a comparative platform for green plant genomics.
897   *Nucleic Acids Res.* **40,** D1178-86 (2012).

898   67.   Boratyn, G. M. *et al.* Domain enhanced lookup time accelerated BLAST. *Biol. Direct* **7,** 12
899   (2012).

900   68.   Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics*
901   **30,** 1236–1240 (2014).

902   69.   de Castro, E. *et al.* ScanProsite: detection of PROSITE signature matches and ProRule-
903   associated functional and structural residues in proteins. *Nucleic Acids Res.* **34,** W362-5
904   (2006).

905   70.   Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. L. Predicting transmembrane
906   protein topology with a hidden markov model: application to complete genomes11Edited
907   by F. Cohen. *J. Mol. Biol.* **305,** 567–580 (2001).

908   71.   Nguyen Ba, A., Pogoutse, A., Provart, N. & Moses, A. NLStradamus: a simple Hidden
909   Markov Model for nuclear localization signal prediction. *BMC Bioinformatics* **10,** 202
910   (2009).

911   72.   Szklarczyk, D. *et al.* The STRING database in 2017: quality-controlled protein–protein
912   association networks, made broadly accessible. *Nucleic Acids Res.* **45,** D362–D368 (2017).

913   73.   von Mering, C. *et al.* STRING: a database of predicted functional associations between
914   proteins. *Nucleic Acids Res.* **31,** 258–261 (2003).

915

916

917

918

919

920

921

922

23

923 **Author contributions**

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

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

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

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

989 **Figure 8**

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

994

995

996

997

998

999

1000

1001

1002

1003

1004

1005

1006

1007

1008

1009

1010

1011

1012

1013

1014

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	*Aegilops tauschii*		4	117	114	0	0	0	114	0
2	*Brachypodium distachyon*	2	1	137	135	0	0	0	135	0
3	*Brachypodium stacei*	1	1	128	127	0	0	0	127	0
4	*Hordeum vulgare*			76	76	0	0	0	76	0
5	*Leersia perrieri*	5	2	163	162	0	0	0	162	0
6	*Oropetium thomaeum*	1		118	103	0	0	0	103	0
7	*Oryza barthii*		4	134	138	0	0	0	138	0
8	*Oryza brachyantha*	1	1	118	110	0	0	0	110	0
9	*Oryza glaberrima*	1		116	110	0	0	0	110	0
10	*Oryza glumipatula*	2		140	139	0	0	0	139	0
11	*Oryza longistaminata*	1	6	125	98	0	0	0	98	0
12	*Oryza meridionalis*	2	2	127	123	0	0	0	123	0
13	*Oryza nivara*	4	1	146	130	0	0	0	130	0
14	*Oryza punctata*	6	1	135	133	0	0	0	133	0
15	*Oryza rufipogon*	4	3	136	129	0	0	0	129	0
16	*Oryza sativa subsp. indica*	1	3	157	156	0	0	0	156	0
17	*Oryza sativa subsp. japonica*	1		139	138	0	0	0	138	0
18	*Panicum hallii*	3	6	139	126	0	0	0	126	0
19	*Panicum virgatum*	9	6	310	309	0	0	0	309	0
20	*Phoenix dactylifera*	3	1	124	123	0	0	0	123	0
21	*Phyllostachys edulis*			125	124	0	0	0	124	0
22	*Phyllostachys heterocycla*	2	2	125	124	0	0	0	124	0
23	*Saccharum officinarum*			44	33	0	0	0	33	0
24	*Setaria italica*	4		139	134	0	0	0	134	0
25	*Setaria viridis*	1		135	118	0	0	0	118	0
26	*Sorghum bicolor*	1		141	134	0	0	0	134	0
27	*Spirodela polyrhiza*			55	48	0	0	0	48	0
28	*Triticum aestivum*	2	2	263	209	0	0	0	209	0
29	*Triticum urartu*		1	103	74	0	0	0	74	0
30	*Zea mays*	1	1	130	119	0	0	0	119	0
31	*Zostera marina*	1		62	55	0	0	0	55	0
32	*Zoysia japonica*		4	176	160	0	0	0	160	0
33	*Zoysia matrella*	1	3	313	230	0	0	0	230	0
34	*Zoysia pacifica*	1	2	205	183	0	0	0	183	0
				Dicots						
35	*Actinidia chinensis*	1	5	167	166	0	0	0	166	0
36	*Aethionema arabicum*	3		85	84	0	0	0	84	0

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

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

135	*Theobroma cacao*			132	131	0	0	0	131	0
136	*Trifolium pratense*	2	2	97	76	0	0	0	76	0
137	*Utricularia gibba*		1	74	73	0	0	0	73	0
138	*Vigna angularis*			98	97	0	0	0	97	0
139	*Vigna radiata*	2		82	81	0	0	0	81	0
140	*Vigna unguiculata*			20	19	0	0	0	19	0
141	*Ziziphus jujuba*			101	100	0	0	0	100	0
142	*Vitis vinifera*	1		70	79	0	0	0	79	0
	Gymnosperms									
143	*Picea abies*	1		100	73	0	0	0	73	0
144	*Picea glauca*			32	31	0	0	0	31	0
145	*Picea sitchensis*			16	15	0	0	0	15	0
146	*Pinus taeda*			31	27	0	0	0	27	0
147	*Pseudotsuga menziesii*	5	3	196	195	0	0	0	195	0
	Pteridophyte									
148	*Selaginella moellendorffii*			22	21	0	0	0	21	0
	Bryophytes									
149	*Marchantia polymorpha*			9						
150	*Physcomitrella patens*			33	32	0	0	0	32	0
151	*Sphagnum fallax*			26	25	0	0	0	25	0
	Algae									
152	*Bathycoccus prasinos*			0	0	0	0	0	0	0
153	*Chlamydomonas reinhardtii*			0	0	0	0	0	0	0
154	*Chlorella sp.* NC64A			0	0	0	0	0	0	0
155	*Coccomyxa sp.*			0	0	0	0	0	0	0
156	*Dunaliella salina*			0	0	0	0	0	0	0
157	*Klebsormidium flaccidum*			3	0	0	0	0	0	0
158	*Micromonas pusilla*			0	0	0	0	0	0	0
159	*Ostreococcus lucimarinus*			0	0	0	0	0	0	0
160	*Volvox carteri*			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,	
NAC38	BRM	MYB69, CIPK4, ABCA8	AT4G29770, AIP2, SDE3
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, JMJ14	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

32

		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

33

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	*Glycine soja*
AAG (K)	146	0	10.75	24.2	*Sphagnum fallax*
AAC (N)	144	2	3.66	14.2	*Beta vulgaris*
AAU (N)	127	19	9.25	20.5	*Spinacia oleracea*
ACA (T)	139	7	2.33	15.2	*Citrus sinensis*
ACC (T)	137	9	2.4	17	*Amborella trichopoda*
ACG (T)	20	126	5.91	13	*Dorcoceras hygrometricum*
ACU (T)	146	0	7.42	16.6	*Sesamum indicum*
AGA (R)	146	0	10.92	24.3	*Klebsormidium flaccidum*
AGG (R)	146	0	4.12	18.8	*Amborella trichopoda*
CGA (R)	19	127	5.22	13.9	*Linum usitatissimum*
CGC (R)	19	127	2.47	6	*Linum usitatissimum*
CGG (R)	19	127	3.93	8.6	*Citrullus lanatus*
CGU (R)	19	127	2.06	4.7	*Linum usitatissimum*
AGC (S)	143	3	3.54	24.2	*Beta vulgaris*
AGU (S)	144	2	1.83	5.2	*Dorcoceras hygrometricum*
UCC (S)	141	5	4.51	12.3	*Aegilops tauschii*
UCG (S)	20	126	2.64	6.4	*Dorcoceras hygrometricum*
UCU (S)	146	0	4.65	30.5	*Humulus lupulus*
UCA (S)	139	7	5.09	15.1	*Morus notabilis*
AUA (I)	124	22	4.80	15.3	*Sphagnum fallax*
AUC (I)	146	0	5.10	16.7	*Sphagnum fallax*
AUU (I)	126	20	8.71	15.9	*Spinacia oleracea*
AUG (M)	146	0	7.81	22.8	*Sphagnum fallax*
CAA (Q)	146	0	5.31	15.4	*Fragaria vesca*
CAG (Q)	20	126	13.3	22.6	*Linum usitatissimum*
CAC (H)	20	126	6.64	10.9	*Beta vulgaris*
CAU (H)	144	2	4.45	9.7	*Setaria viridis*
CCA (P)	20	126	11.09	16.3	*Dorcoceras hygrometricum*
CCC (P)	20	126	14.18	19.2	*Amborella trichopoda*
CCG (P)	20	126	5.10	11.1	*Dorcoceras hygrometricum*
CCU (P)	146	0	8.00	24.7	*Klebsormidium flaccidum*
CUA (L)	143	3	5.83	28.3	*Sphagnum fallax*
CUC (L)	123	23	5.74	23.6	*Sphagnum fallax*
CUG (L)	142	4	5.87	43.9	*Sphagnum fallax*
CUU (L)	145	1	5.94	32.6	*Sphagnum fallax*
UUG (L)	125	21	5.94	24.4	*Sphagnum fallax*
UAA (L)	124	22	5.37	17.2	*Sphagnum fallax*
GAA (E)	146	0	4.62	27	*Klebsormidium flaccidum*
GAG (E)	145	1	5.54	18.1	*Sphagnum fallax*

34

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

1037

1038

1039

1040

1041

1042

1043

1044

1045

1046

1047

1048

1049

1050

1051

35

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
	Actinidia chinensis				*Aegilops tauschi*				*Aethionema arabicum*				*Amaranthus hypochondriacus*			
A	-	10.62	4.26	**10.83**	-	10.36	4.71	**12.32**	-	11.43	4.3	**10.15**	-	11.69	4.83	**10.45**
T	8.34	-	**4.34**	5.38	7.78	-	**4.37**	5.96	9.01	-	**3.54**	5.7	8.54	-	**3.45**	5.16
C	8.34	**10.83**	-	5.38	7.78	**9.62**	-	5.96	9.01	**9.39**	-	5.7	8.54	**8.35**	-	5.16
G	**16.79**	10.62	4.26	-	**16.09**	10.36	4.71	-	**16.04**	11.43	4.3	-	**17.32**	11.69	4.83	-
	Amborella trichopoda				*Ananas comosus*				*Aquilegia coerulea*				*Arabidopsis halleri*			
A	-	3.99	10.98	**14.14**	-	3.73	10.95	**15.36**	-	11	4.33	**11.11**	-	10.93	4.15	**11.07**
T	8.17	-	**5.41**	10.7	7.66	-	**5.91**	10.32	8.67	-	**3.44**	5.32	8.98	-	**3.3**	5.73
C	8.17	**1.97**	-	10.7	7.66	**2.01**	-	10.32	8.67	**8.73**	-	5.32	8.98	**8.71**	-	5.73
G	**10.8**	3.99	10.98	-	**11.39**	3.73	10.95	-	**18.09**	11	4.33	-	**17.34**	1.93	4.15	-
	Arabidopsis lyrata				*Arabidopsis thaliana*				*Arabis alpina*				*Arachis duranensis*			
A	-	4.03	10.6	**14.35**	-	4.06	10.4	**14.86**	-	11.04	4.25	**10.62**	-	11.17	4.43	**10.3**
T	8.63	-	**4.51**	10.64	8.41	-	**4.65**	10.52	8.88	-	**3.63**	5.67	8.76	-	**3.82**	5.37
C	8.63	**1.71**	-	10.64	8.41	**1.82**	-	10.52	8.88	**9.42**	-	5.67	8.76	**9.62**	-	5.37
G	**11.63**	4.03	10.6	-	**11.89**	4.06	10.4	-	**16.63**	11.04	4.25	-	**16.79**	11.17	4.43	-
	Arachis hypogaea				*Arachis ipaensis*				*Artemisia annua*				*Azadirachta indica*			
A	-	4	11.07	**14.21**	-	3.93	10.85	**14.63**	-	10.85	4.08	**10.93**	-	10.98	4.22	**10.34**
T	8.16	-	**5.53**	10.24	8.39	-	**4.95**	10	8.24	-	**4.22**	5.62	8.91	-	**3.82**	5.6
C	8.16	**2**	-	10.24	8.39	**1.79**	-	10	8.24	**11.24**	-	5.62	8.91	**9.95**	-	5.6
G	**11.32**	4	11.07	-	**12.28**	3.93	10.85	-	**16.01**	10.85	4.08	-	**16.44**	10.98	4.22	-
	Beta vulgaris				*Boechera stricta*				*Brachypodium distachyon*				*Brachypodium stacei*			
A	-	3.85	10.95	**14.38**	-	11.02	4.23	**10.86**	-	10.6	4.93	**12.57**	-	10.52	4.96	**12.65**
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	10.95	-	**17.09**	11.02	4.23	-	**15.74**	10.6	4.93	-	**15.63**	10.52	4.96	-
	Brassica napus				*Brassica oleracea*				*Brassica rapa*				*Cajanus cajan*			
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**

	Camelina sativa				Cannabis sativa				Capsella grandiflora				Capsella rubella			
	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	**10.18**	-	5.57	8.7	**9.87**	-	5.57	8.48	**1.72**	-	10.6	8.63	**9.73**	-	5.53
G	**16.57**	10.83	4.26	-	**16.73**	10.83	4.29	-	**11.58**	3.98	10.73	-	**16.07**	11.31	4.53	-
A	-	3.88	10.44	**14.9**	-	11.17	4.35	**10.49**	-	12.34	4.92	**9.49**	-	11.19	4.35	**10.48**
T	8.41	-	**4.9**	10.46	9.25	-	**3.24**	5.72	9.91	-	**2.41**	6.34	8.81	-	**3.69**	5.65
C	8.41	**1.82**	-	10.46	9.25	**8.33**	-	5.72	9.91	**6.2**	-	6.34	8.81	**9.48**	-	5.65
G	**11.99**	3.88	10.44	-	**16.97**	11.17	4.35	-	**14.83**	12.34	4.92	-	**16.33**	11.19	4.35	-

1053

	Capsicum annuum				Carica papaya				Castanea mollissima				Catharanthus roseus			
A	-	10.95	3.95	**10.8**	-	10.43	4.05	**11.49**	-	11.31	4.53	**10.03**	-	10.54	4.13	**11.35**
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	**10.24**	-	5.47	8.98	**9.37**	-	5.73	8.6	**9.76**	-	5.69
G	**17.51**	10.95	3.95	-	**17.63**	10.43	4.05	-	**15.72**	11.31	4.53	-	**17.14**	10.54	4.13	-
	Chenopodium quinoa				Cicer arietinum				Citrullus lanatus				Citrus clementina			
A	-	11.2	4.48	**10.46**	-	4.05	10.91	**13.8**	-	3.95	10.82	**14.62**	-	3.9	10.56	**14.72**
T	8.77	-	**3.76**	5.67	8.93	-	**4.36**	10.17	8.25	-	**5.19**	10.24	8.33	-	**5.26**	10.29
C	8.77	**9.4**	-	5.67	8.93	**1.62**	-	10.17	8.25	**1.89**	-	10.24	8.33	**1.94**	-	10.29
G	**16.16**	11.2	4.48	-	**12.11**	4.05	10.91	-	**11.77**	3.95	10.82	-	**11.92**	3.9	10.56	-
	Citrus sinensis				Coffea canephora				Cucumis melo				Cucumis sativus			
A	-	3.93	10.49	**14.37**	-	11.22	4.64	**10.3**	-	11.31	4.53	**10.21**	-	4.18	10.97	**14.12**
T	8.39	-	**5.72**	10.12	8.63	-	**4.03**	5.76	8.5	-	**4.07**	5.39	8.45	-	**4.75**	10.2
C	8.39	**2.14**	-	10.12	8.63	**9.75**	-	5.76	8.5	**10.17**	-	5.39	8.45	**1.81**	-	10.2
G	**11.91**	3.93	10.49	-	**15.42**	11.22	4.64	-	**16.09**	11.31	4.53	-	**11.69**	4.18	10.97	-
	Daucus carota				Dianthus caryophyllus				Dichanthelium oligosanthes				Dorcoceras hygrometricum			
A	-	4.04	10.99	**13.8**	-	3.93	10.74	**14.26**	-	10.55	4.97	**12.66**	-	4.04	10.85	**14.14**
T	8.52	-	**5.16**	10.34	8.57	-	**5.02**	10.25	7.49	-	**4.32**	5.97	8.47	-	**5.04**	10.3

C	8.52	**1.89**	-	10.34	8.57	**1.84**	-	10.25	7.49	**9.16**	-	5.97	8.47	**1.88**	-	10.3
G	**11.37**	4.04	10.99	-	**11.91**	3.93	10.74	-	**15.88**	10.55	4.97	-	**11.62**	4.04	10.85	-
	Elaeis guineensis				*Eragrostis tef*				*Eucalyptus camaldulensis*				*Eucalyptus grandis*			
A	-	3.85	10.79	**15.28**	-	10.72	4.7	**11.48**	-	10.84	4.28	**10.81**	-	3.95	10.68	**14.94**
T	7.88	-	**5.23**	10.63	8.34	-	**3.99**	6.09	8.47	-	**4.13**	5.7	7.87	-	**5.87**	10.31
C	7.88	1.86	-	10.63	8.34	9.12	-	6.09	8.47	10.44	-	5.7	7.87	2.17	-	10.31
G	**11.33**	3.85	10.79	-	**15.72**	10.72	4.7	-	**16.05**	10.84	4.28	-	**11.4**	3.95	10.68	-
	Eutrema salsugineum				*Fragaria vesca*				*Fragaria x ananassa*				*Genlisea aurea*			
A	-	11.04	4.27	**10.67**	-	10.85	4.49	**10.67**	-	10.22	4.07	**11.19**	-	10.5	3.98	**11.58**
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	11.24	-	5.34	9.08	8.79	-	6.52
G	**16.7**	11.04	4.27	-	**17.06**	10.85	4.49	-	**17.32**	10.22	4.07	-	**16.12**	10.5	3.98	-
	Glycine max				*Glycine soja*				*Gossypium arboreum*				*Gossypium hirsutum*			
A	-	11.43	4.62	**10.18**	-	11.53	4.72	**10.07**	-	11.05	4.39	**10.56**	-	11.06	4.4	**10.51**
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	10.02	-	5.55
G	**16.01**	11.43	4.62	-	**15.92**	11.53	4.72	-	**16.58**	11.05	4.39	-	**16.27**	11.06	4.4	-
	Gossypium raimondii				*Helianthus annuus*				*Hordeum vulgare*				*Humulus lupulus*			
A	-	11.28	4.52	**10.22**	-	10.39	4.31	**10.79**	-	10.74	4.81	**11.64**	-	11.21	4.63	**10.52**
T	8.85	-	**3.77**	5.62	8.23	-	**4.74**	5.46	8.41	-	**3.82**	6.48	8.7	-	**3.68**	5.37
C	8.85	9.39	-	5.62	8.23	11.42	-	5.46	8.41	8.53	-	6.48	8.7	8.92	-	5.37
G	**16.09**	11.28	4.52	-	**16.28**	10.39	4.31	-	**15.12**	10.74	4.81	-	**17.06**	11.21	4.63	-
	Ipomoea trifida				*Jatropha curcas*				*Juglans regia*				*Kalanchoe laxiflora*			
A	-	10.79	4.47	**10.94**	-	11.19	4.34	**10.66**	-	11.09	4.45	**10.82**	-	11.52	4.93	**9.82**
T	8.33	-	**4.22**	5.6	8.68	-	**3.66**	5.5	8.45	-	**3.93**	5.64	8.72	-	**4.03**	5.74
C	8.33	**10.19**	-	5.6	8.68	9.44	-	5.5	8.45	9.79	-	5.64	8.72	9.4	-	5.74
G	**16.27**	10.79	4.47	-	**16.83**	11.19	4.34	-	**16.22**	11.09	4.45	-	**14.92**	11.52	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**

(continued from previous page)

	A	T	C	G
A	-		9	8
T	9.37	-	2.86	6.19
C	9.37	6.59	-	6.19
G	14.93	12.08	5.24	-

	A	T	C	G
A	-			8
T	6.67	-	7.19	5.04
C	6.67	13.78	-	5.04
G	18.28	7.73	4.03	-

	A	T	C	G
A	-	9	7	
T	9.13	-	3.67	5.91
C	9.13	9.16	-	5.91
G	15.28	11.39	4.57	-

	A	T	C	G
A	-	1		7
T	7.33	-	4.28	5.82
C	7.33	9.66	-	5.82
G	16.73	10.31	4.57	-

Linum usitatissimum

	A	T	C	G
A	-	3.72	10.39	**15.82**
T	7.89	-	**5.6**	10.09
C	7.89	**2.01**	-	10.09
G	**12.37**	3.72	10.39	-

Lotus japonicus

	A	T	C	G
A	-	10.31	3.72	**12.64**
T	8.2	-	**3.43**	5.44
C	8.2	**9.51**	-	5.44
G	**19.07**	10.31	3.72	-

Malus domestica

	A	T	C	G
A	-	10.65	4.23	**11.21**
T	8.18	-	**4.17**	5.24
C	8.18	**10.5**	-	5.24
G	**17.5**	10.65	4.23	-

Manihot esculenta

	A	T	C	G
A	-	11.22	4.54	**10.48**
T	8.65	-	**3.82**	5.57
C	8.65	**9.45**	-	5.57
G	**16.29**	11.22	4.54	-

Medicago truncatula

	A	T	C	G
A	-	11.24	4.3	**10.19**
T	8.9	-	**3.68**	5.4
C	8.9	**9.62**	-	5.4
G	**16.79**	11.24	4.3	-

Mimulus guttatus

	A	T	C	G
A	-	10.72	4.02	**10.51**
T	8.91	-	**3.91**	5.67
C	8.91	**10.42**	-	5.67
G	**16.51**	10.72	4.02	-

Morus notabilis

	A	T	C	G
A	-	11.16	4.54	**10.64**
T	8.54	-	**3.8**	5.32
C	8.54	**9.34**	-	5.32
G	**17.1**	11.16	4.54	-

Musa acuminata

	A	T	C	G
A	-	10.93	4.62	**10.89**
T	8.14	-	**4.42**	5.9
C	8.14	**10.47**	-	5.9
G	**15.05**	10.93	4.62	-

Nelumbo nucifera

	A	T	C	G
A	-	10.88	4.4	**10.9**
T	8.43	-	**4.04**	5.56
C	8.43	**9.99**	-	5.56
G	**16.55**	10.88	4.4	-

Nicotiana benthamiana

	A	T	C	G
A	-	10.88	4.11	**10.39**
T	8.82	-	**3.89**	5.38
C	8.82	**10.29**	-	5.38
G	**17.03**	10.88	4.11	-

Nicotiana sylvestris

	A	T	C	G
A	-	10.88	4.26	**10.4**
T	8.87	-	**3.86**	5.44
C	8.87	**9.87**	-	5.44
G	**16.98**	10.88	4.26	-

Nicotiana tabacum

	A	T	C	G
A	-	11.2	4.31	**10.21**
T	9.07	-	**3.57**	5.55
C	9.07	**9.28**	-	5.55
G	**16.69**	11.2	4.31	-

Nicotiana tomentosiformis

	A	T	C	G
A	-	11.05	4.3	**10.43**
T	8.93	-	**3.64**	5.43
C	8.93	**9.36**	-	5.43
G	**17.16**	11.05	4.3	-

Ocimum tenuiflorum

	A	T	C	G
A	-	11.02	4.32	**10.59**
T	8.44	-	**4.15**	5.68
C	8.44	**10.58**	-	5.68
G	**15.74**	11.02	4.32	-

Oropetium thomaeum

	A	T	C	G
A	-	10.52	4.82	**12.41**
T	7.46	-	**4.56**	5.93
C	7.46	**9.97**	-	5.93
G	**15.61**	10.52	4.82	-

Oryza brachyantha

	A	T	C	G
A	-	10.51	4.85	**12.98**
T	7.3	-	**4.37**	5.92
C	7.3	**9.47**	-	5.92
G	**16.01**	10.51	4.85	-

1055

Oryza glaberrima

	A	T	C	G
A	-	10.65	5.01	**12.86**
T	7.52	-	**4.08**	6.14

Oryza glumaepatula

	A	T	C	G
A	-	10.7	5	**12.47**
T	7.61	-	**4.18**	6.16

Oryza longistaminata

	A	T	C	G
A	-	10.69	4.74	**12.11**
T	7.96	-	**4.03**	6.09

Oryza meridionalis

	A	T	C	G
A	-	10.4	4.87	**13.32**
T	7.43	-	**4.1**	6.12

	A	T	C	G	A	T	C	G	A	T	C	G	A	T	C	G
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	**15.73**	10.65	5.01	-	**15.42**	10.71	5	-	**15.82**	10.69	4.74	-	**16.14**	10.42	4.87	-

	Oryza nivara				*Oryza punctata*				*Oryza rufipogon*				*Oryza sativa*			
A	-	10.64	5.01	**12.49**	-	10.71	4.94	**12.37**	-	10.67	5.04	**12.84**	-	10.86	5.12	**12.22**
T	7.6	-	**4.25**	6.17	7.66	-	**4.2**	6.23	7.47	-	**4.14**	6.18	7.64	-	**4.21**	6.33
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	**15.39**	10.6	5.01	-	**15.22**	10.71	4.94	-	**15.53**	10.67	5.04	-	**17.73**	10.86	5.12	-

	Panicum hallii				*Panicum virgatum*				*Petunia axillaris*				*Petunia inflata*			
A	-	10.66	5	**12.63**	-	10.4	5.02	**12.72**	-	10.03	4.18	**10.07**	-	11.13	4.28	**10.28**
T	7.75	-	**3.97**	6.32	7.53	-	**4.39**	6.14	8.77	-	**4.05**	5.35	8.83	-	**3.79**	5.41
C	7.75	**8.47**	-	6.32	7.53	**9.12**	-	6.14	8.77	**10.68**	-	5.35	8.83	**9.86**	-	5.41
G	**15.48**	10.66	5	-	**15.6**	10.4	5.02	-	**16.52**	10.03	4.18	-	**16.77**	11.13	4.28	-

	Phaselous vulgaris				*Phoenix dactylifera*				*Phyllostachys heterocycla*				*Physcomitrella patens*			
A	-	11.4	4.62	**10.24**	-	10.71	4.44	**11.16**	-	11.03	4.97	**11.6**	-	10.61	4.64	**10.77**
T	8.75	-	**3.78**	5.69	8.2	-	**4.31**	5.71	8.1	-	**3.94**	6.27	8.1	-	**4.77**	5.53
C	8.75	**9.32**	-	5.69	8.2	**10.39**	-	5.71	8.1	**8.73**	-	6.27	8.1	**10.92**	-	5.53
G	**15.75**	11.4	4.62	-	**16.0**	10.71	4.44	-	**14.99**	11.03	4.97	-	**15.77**	10.61	4.64	-

	Picea abies				*Picea glauca*				*Picea sitchensis*				*Pinus taeda*			
A	-	10.82	4.54	**10.29**	-	10.85	4.61	**10.35**	-	10.35	4.45	**10.58**	-	10.08	4.37	**11.67**
T	8.53	-	**4.42**	5.48	8.42	-	**4.47**	5.42	8.02	-	**5.11**	5.12	8.18	-	**4.39**	5.35
C	8.53	**10.53**	-	5.48	8.42	**10.52**	-	5.42	8.02	**11.87**	-	5.12	8.18	**10.11**	-	5.35
G	**16**	10.82	4.54	-	**16.08**	10.85	4.61	-	**16.56**	10.35	4.45	-	**17.86**	10.08	4.37	-

	Populus euphratica				*Populus trichocarpa*				*Prunus mume*				*Prunus persica*			
A	-	10.86	4.53	**10.85**	-	10.98	4.52	**10.68**	-	11.12	4.43	**10.65**	-	11.38	4.53	**10.23**
T	8.33	-	**4.15**	5.31	8.47	-	**4.01**	5.35	8.91	-	**3.49**	5.55	8.95	-	**3.55**	5.6
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	-	**16.93**	10.98	4.52	-	**17.07**	11.12	4.43	-	**16.36**	11.38	4.53	-

	Pseudotsuga menziesii				*Pyrus bretschneideri*				*Raphanus raphanistrum*				*Raphanus sativus*			
A	-	10.74	4.44	**9.98**	-	11.1	4.51	**10.78**	-	11.07	4.4	**10.51**	-	10.96	4.35	**10.71**

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	-

1056

	Ricinus communis				*Saccharum officinarum*				*Salix purpurea*				*Salvia miltiorrhiza*			
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	-

	Selaginella moellendorffii				*Sesamum indicum*				*Setaria italica*				*Setaria viridis*			
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	-

	Sisymbrium irio				*Solanum lycopersicum*				*Solanum melongena*				*Solanum pennellii*			
A	-	10.74	4.29	**11.05**	-	11.17	4.12	**10.13**	-	10.72	4.04	**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	-	**16.9**	11.17	4.12	-	**17.46**	10.72	4.04	-	**17.03**	11.07	4.11	-

	Solanum pimpinellifolium				*Solanum tuberosum*				*Sorghum bicolor*				*Sphagnum fallax*			
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	-

	Spinacia oleracea				*Spirodela polyrhiza*				*Tarenaya hassleriana*				*Thellungiella parvula*			
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	**16.51**	11.37	4.57	-	**14.16**	11.01	4.84	-	**16.02**	11.05	4.52	-	**15.78**	11.34	4.46	-

	Theobroma cacao				*Trifolium pratense*				*Triticum aestivum*				*Triticum urartu*			
A	-	11.04	4.45	**10.52**	-	11.21	4.26	**9.73**	-	10.65	4.85	**11.2**	-	10.78	4.94	**11.4**
T	8.53	-	**4.04**	5.43	9.05	-	**3.89**	5.43	8.28	-	**4.34**	6.16	8.12	-	**4.27**	6.19
C	8.53	**10.02**	-	5.43	9.05	**10.24**	-	5.43	8.28	**9.54**	-	6.16	8.12	**9.32**	-	6.19
G	**16.53**	11.04	4.45	-	**16.24**	11.21	4.26	-	**15.04**	10.65	4.85	-	**14.95**	10.78	4.94	-

	Utricularia gibba				*Vigna angularis*				*Vigna radiata*				*Vigna unguiculata*			
A	-	10.97	4.56	**10.67**	-	11.18	4.45	**10.4**	-	10.76	4.33	**10.64**	-	10.21	3.95	**11.11**
T	8.41	-	**4.26**	6	8.87	-	**3.72**	5.74	8.58	-	**4.16**	5.59	8.54	-	**4.35**	5.86
C	8.41	**10.23**	-	6	8.87	**9.34**	-	5.74	8.58	**10.34**	-	5.59	8.54	**11.24**	-	5.86
G	**14.96**	10.97	4.56	-	**16.08**	11.18	4.45	-	**16.33**	10.76	4.33	-	**16.19**	10.21	3.95	-

1057

	Vitis vinifera				*Zea mays*				*Ziziphus jujuba*				*Zoysia japonica*			
A	-	11.27	4.64	**10.21**	-	10.77	5.14	**12.1**	-	11.03	5.54	**10.71**	-	10.83	5.02	**12.08**
T	8.69	-	**3.98**	5.74	7.82	-	**4.15**	6.34	8.6	-	**3.84**	5.47	7.87	-	**4.04**	6.18
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	**15.45**	11.27	4.64	-	**14.91**	10.77	5.14	-	**16.84**	11.03	5.54	-	**15.38**	10.83	5.02	-

	Zoysia matrella				*Zoysia pacifica*			
A	-	10.65	5	**12.29**	-	10.41	4.88	**12.67**
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	**15.54**	10.65	5	-	**18.89**	10.41	4.88	-

1058

1059

1060

1061

1062

Supplementary Data

Supplementary Table 1

42

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 molecular weight (kDa) of NAC TFs
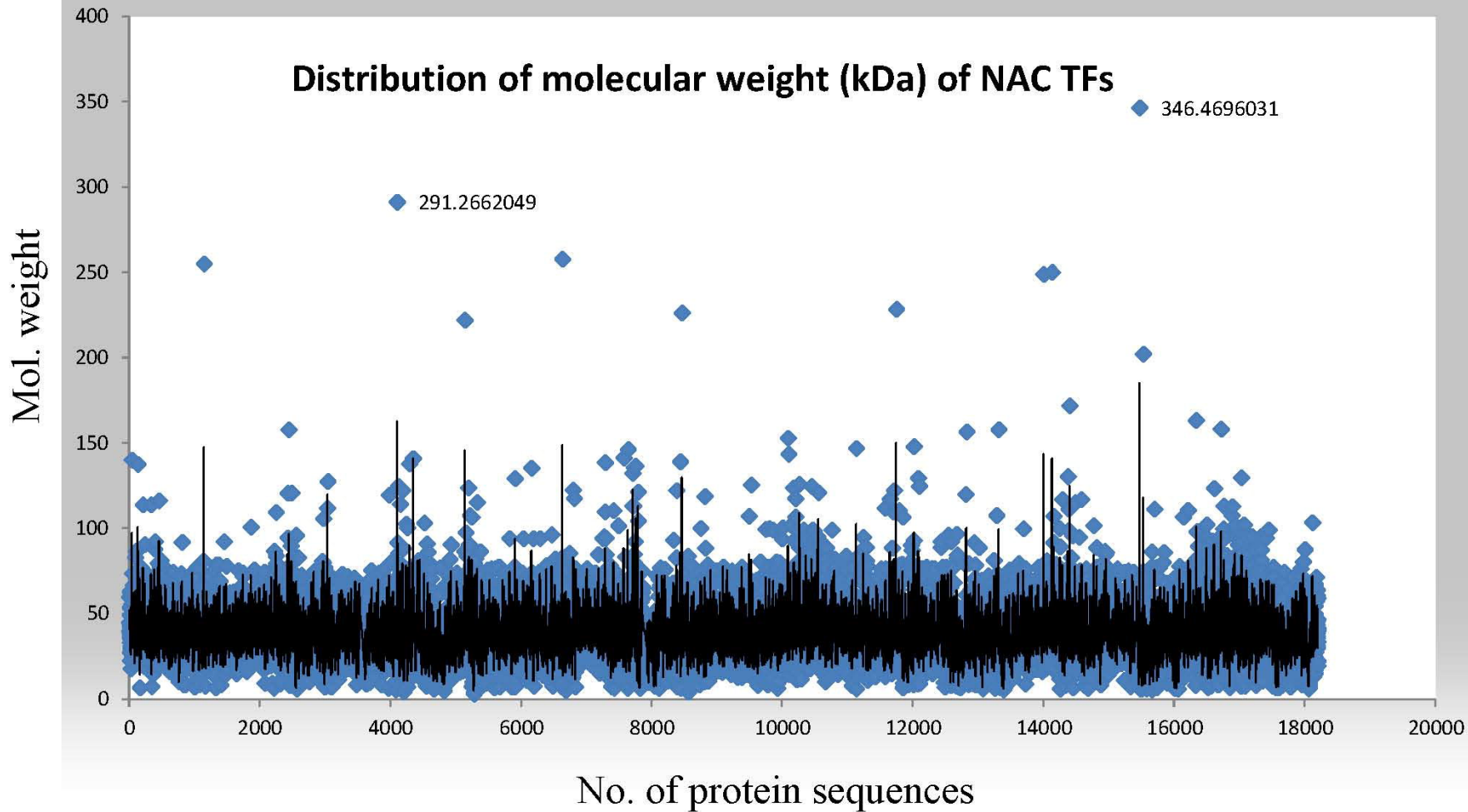
Distribution of isoelectric points (*pI*) of NAC TFs

pI (vertical axis, ranging from 0 to 14)

No. of protein sequences (horizontal axis, ranging from 0 to 18000)

Gene Expression of NAC TFs in leaf

Gene expression of NAC TF in roots

1_Actinidiachinensis_Achn000661
7115_Gossypiumraimondii_Gorai_012G007000_1
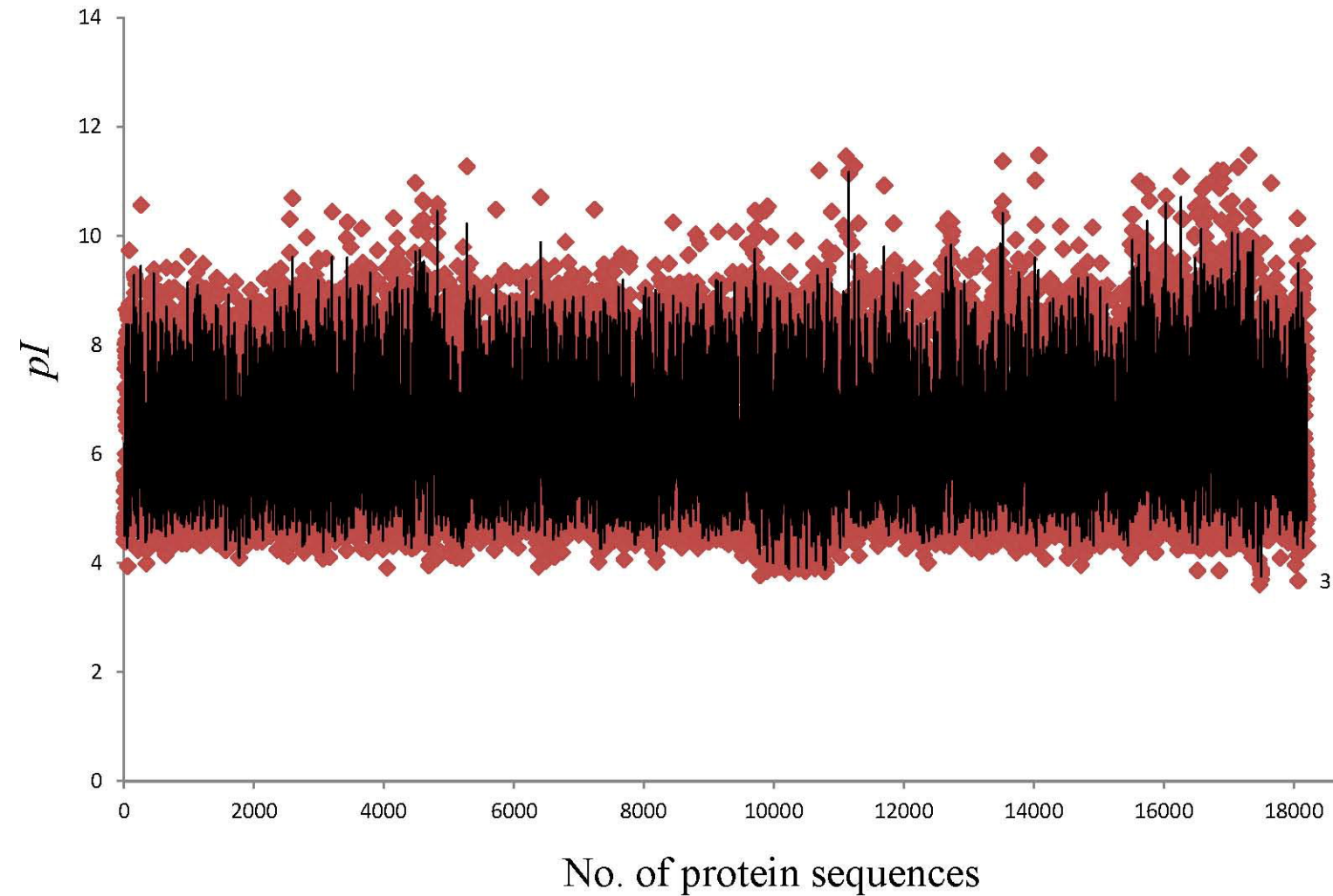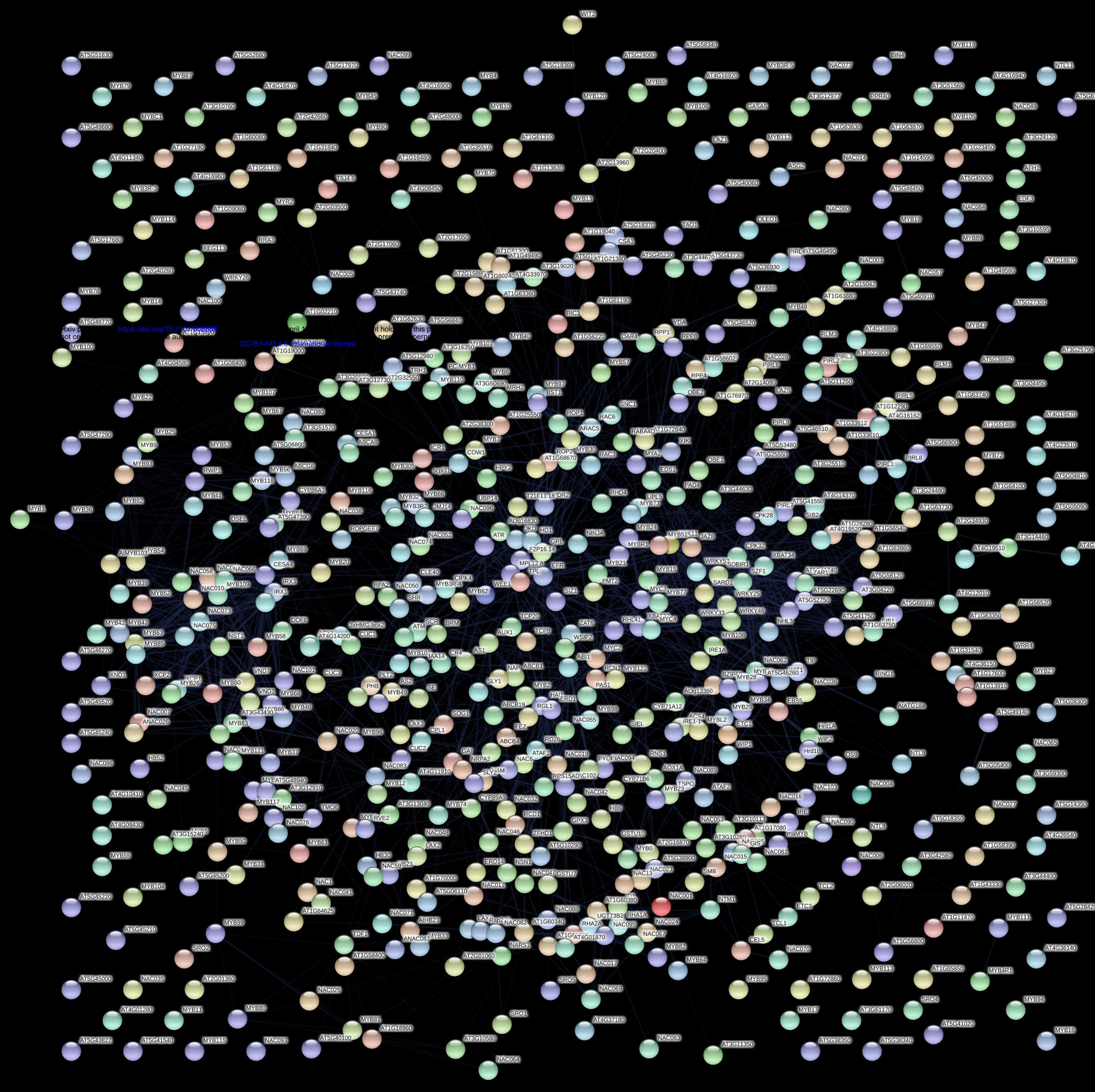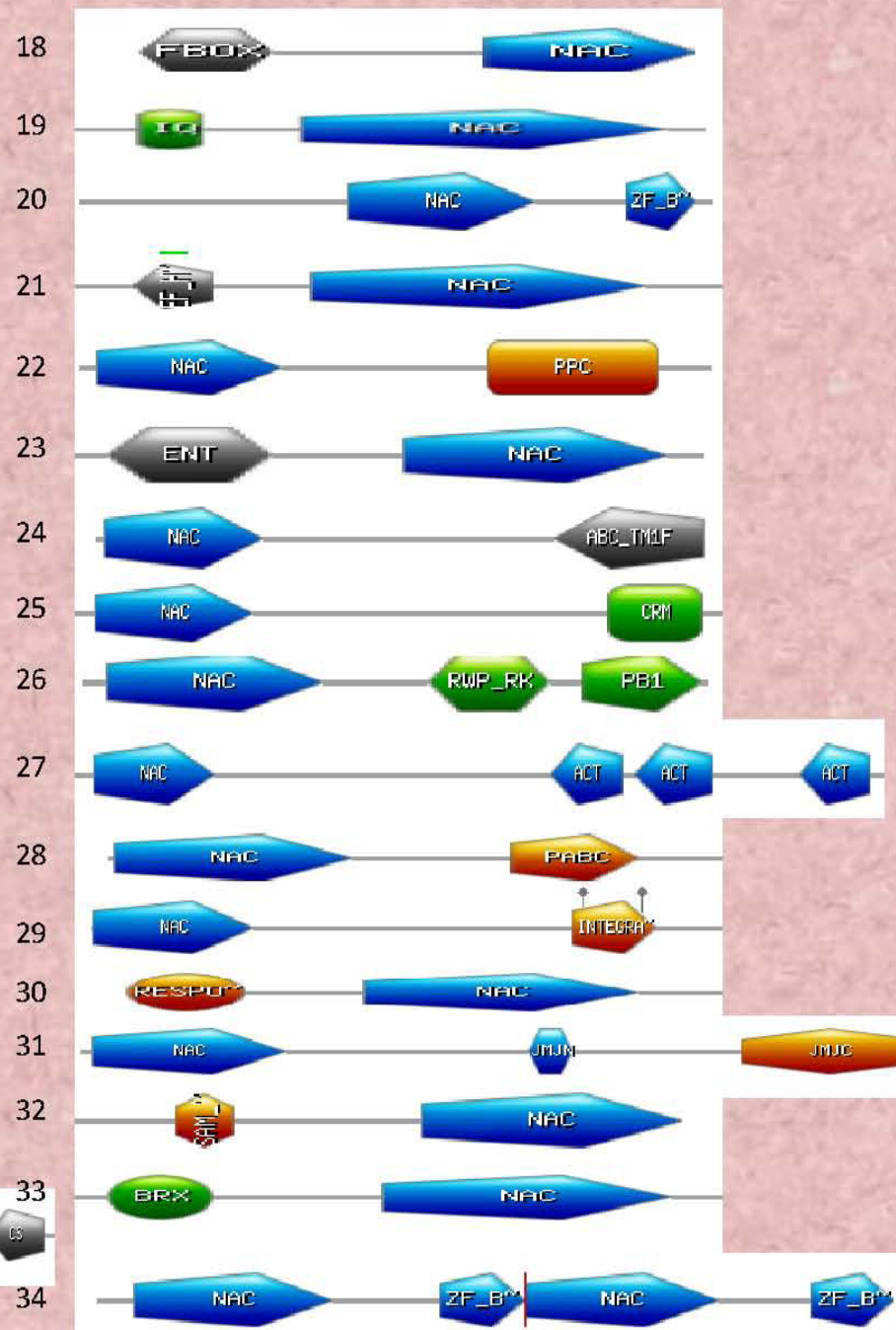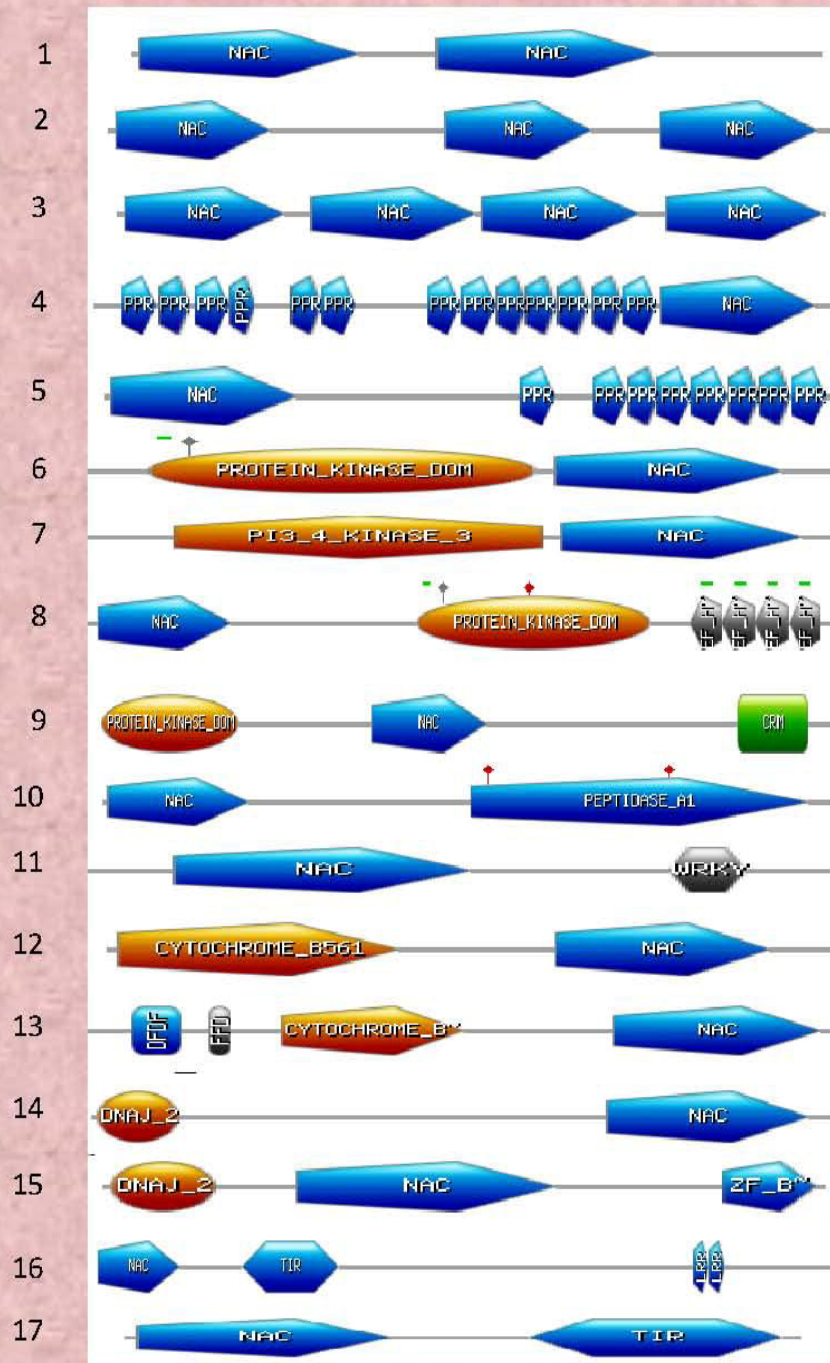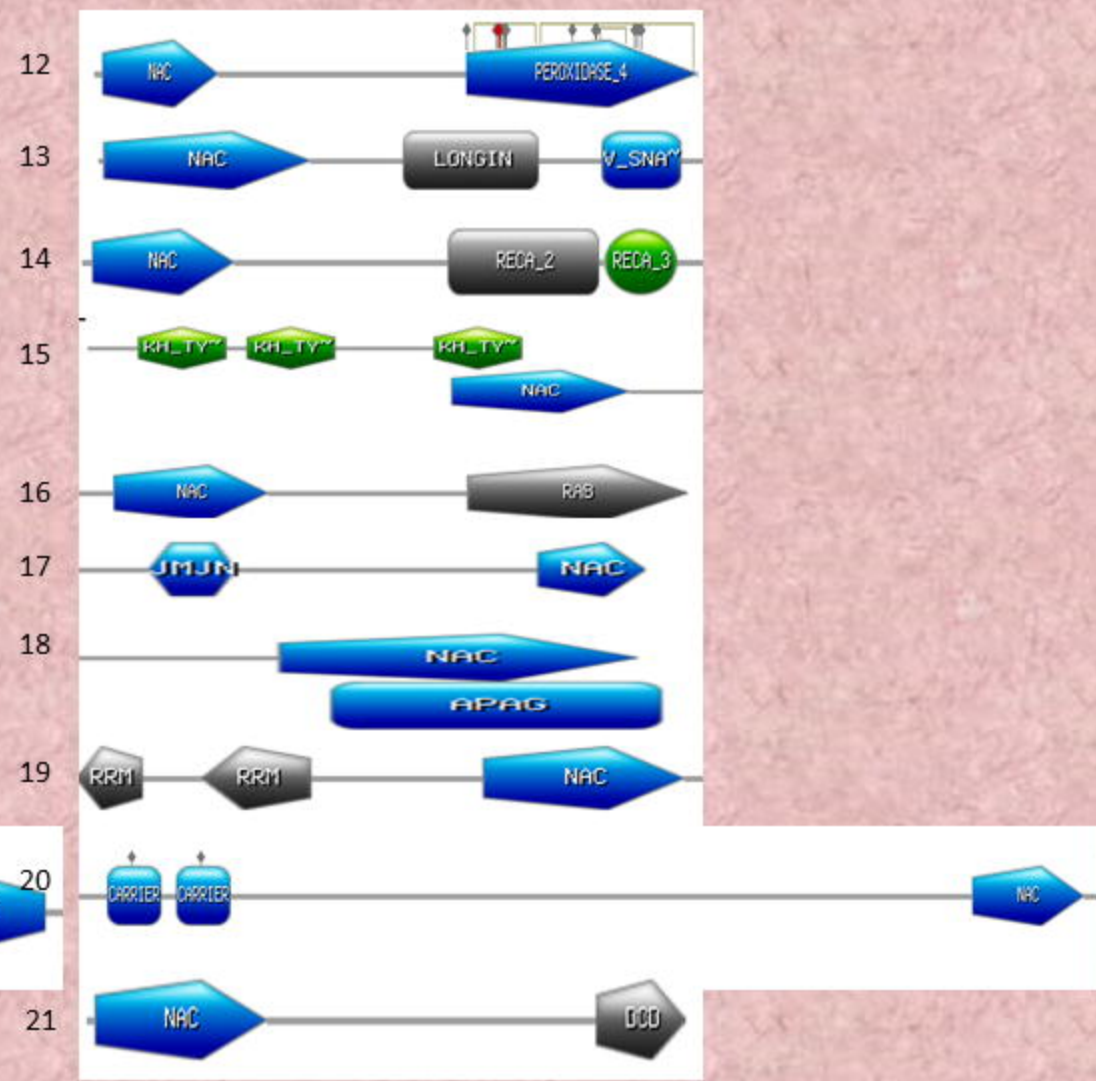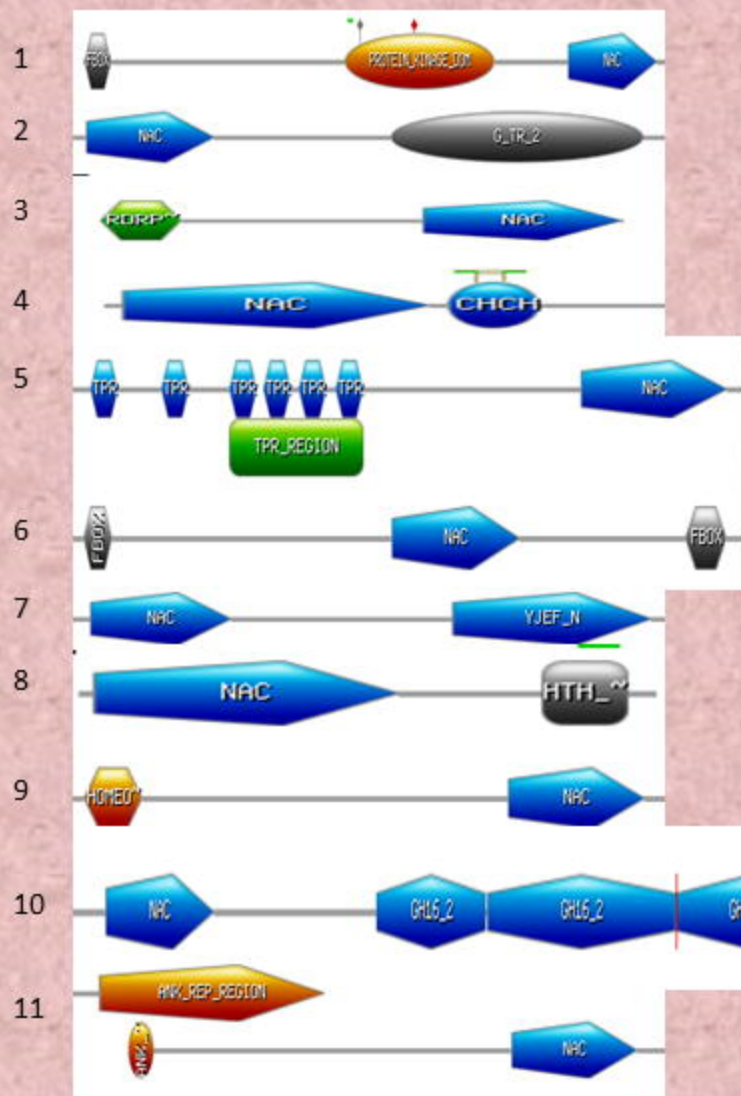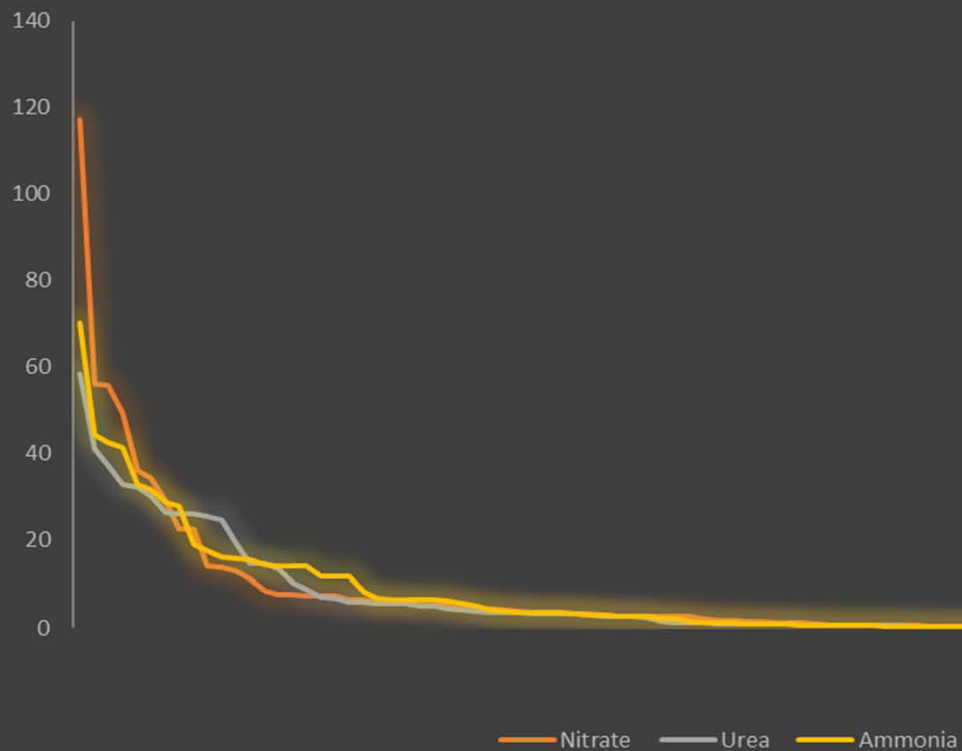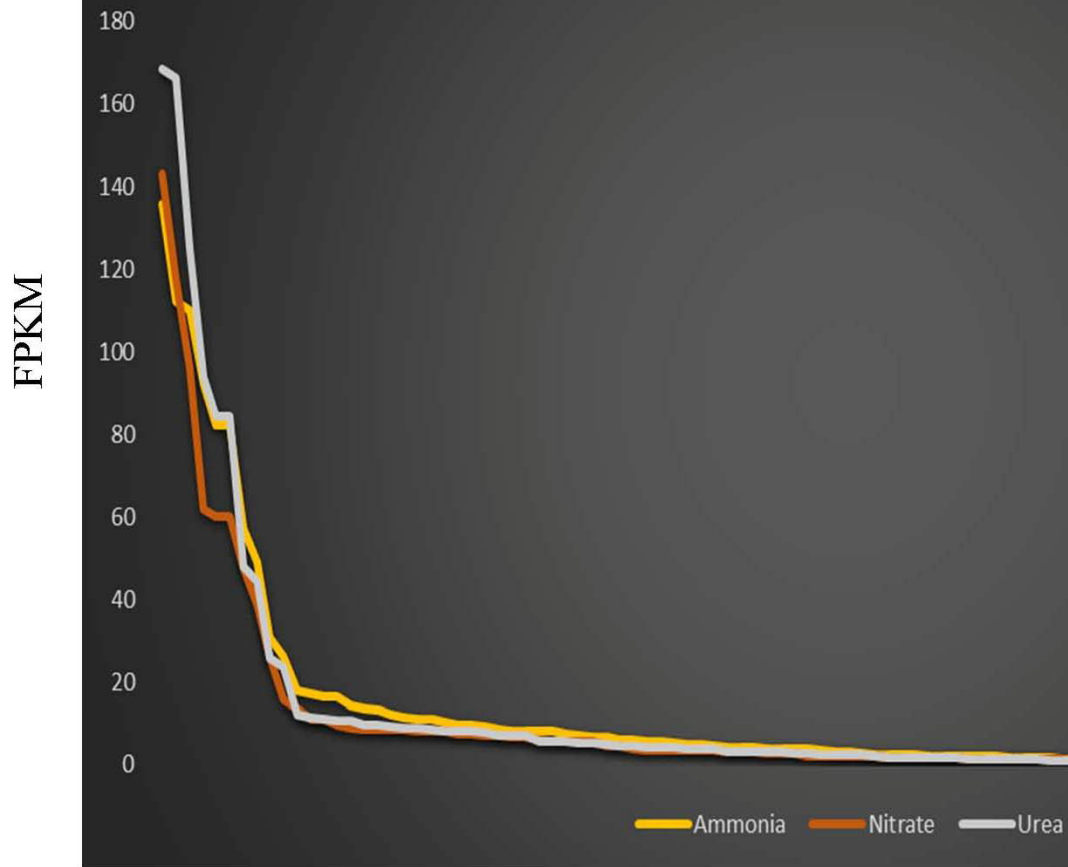10778_Oryzaglumaepatula_OGLUM06G29360_1
10199_Nicotianatomentosiformis_XP_009625395_1
11418_Oryzarufipogon_ORUFI03G42200_1
10550_Oryzabrachyantha_OB08G23420_1
9559_Nicotianabenthamiana_Niben101Scf06245g01015_1
13757_Pseudotsugamenziesii_PSME_00030678-RA
5212_Elaeisguineensis_XP_010936653_1
13368_Populustrichocarpa_Potri_014G075900_1
11849_Panicumhallii_Pahal_C04569_1
13931_Pyrusxbretschneideri_Pbr019570_1
14960_Sesamumindicum_XP_011099807_1
15970_Sorghumbicolor_Sobic_007G003000_1_p
12398_Petuniainflata_Peinf101Scf00071g21008_1
12899_Physcomitrellapatens_Pp1s42_248V6_1
4933_Dichantheliumoligosanthes_Do013815_1
6590_Gossypiumarboreum_Cotton_A_13395_BGI-A2_v1_0
2334_Brassicaoleracea_XP_013595819_1
976_Arabisalpina_KFK33194_1
12512_Petuniainflata_Peinf101Scf02138g03033_1
16998_Utriculariagibba_Scf00098_g8363_t1
16917_Triticumurartu_EMS51610
10309_Ocimumtenuiflorum_Ote100193880041
14402_Raphanussativus_Rsa1_0_03990_1_g00005_1
5287_Eragrostistef_462875202
8793_Manihotesculenta_cassava4_1_023969m
7265_Humuluslupulus_HL_SW_v1_0_G020386_1
17766_Zoysiamatrella_Zmw_sc00400_1_g00450_1
1692_Brachypodiumdistachyon_Bradi3g56080_1_p
843_Arabidopsisthaliana_AT1G32870_1
14206_Raphanusraphanistrum_RrC6770_p4
14708_Salixpurpurea_SapurV1A_0894s0120_
13198_Populuseuphratica_CCG025464_1
4788_Dianthuscaryophyllus_Dca11233_1
12306_Petuniaaxillaris_Peaxi162Scf00224g00514_1
5688_Eucalyptusgrandis_Eucgr_F03962_1_p
10877_Oryzalongistaminata_KN538938_1_FGP015
4433_Citrussinensis_orange1_1g045641m
2977_Camelinasativa_Csa05g047230_1
16742_Triticumaestivum_Traes_3DL_0BF080E4B_2
4564_Cucumismelo_MELO3C015427P1
4895_Dichantheliumoligosanthes_Do005286_1
15_Actinidiachinensis_Achn014031
17529_Ziziphusjujuba_XP_015893680_1
3223_Camelinasativa_Csa20g012360_1
15208_Setariaviridis_Sevir_9G038400_
10590_Oryzaglaberrima_ORGLA01G0288400_1
13573_Prunuspersica_Prupe_2G202600_1_p
8130_Leersiaperrieri_LPERR07G23450_1