

RESEARCH ARTICLE

Genomics, molecular and evolutionary perspective of NAC transcription factors

Tapan Kumar Mohanta¹*, Dhananjay Yadav², Adil Khan¹, Abeer Hashem^{3,4}, Baby Tabassum⁵, Abdul Latif Khan¹, Elsayed Fathi Abd_Allah⁶, Ahmed Al-Harrasi^{1*}

1 Natural and Medicinal Plant Sciences Research Center, University of Nizwa, Nizwa, Oman, **2** Dept. of Medical Biotechnology, Yeungnam University, Gyeongsan, Republic of Korea, **3** Botany and Microbiology Department, College of Science, King Saud University, Riyadh, Saudi Arabia, **4** Mycology and Plant Disease Survey Department, Plant Pathology Research Institute, ARC, Giza, Egypt, **5** Department of Zoology, Toxicology laboratory, Raza P.G. College, Rampur, Uttar Pradesh, India, **6** Plant Production Department, College of Food and Agricultural Sciences, King Saud University, Riyadh, Saudi Arabia

* These authors contributed equally to this work.

* nostoc.tapan@gmail.com, tapan.mohanta@unizwa.edu.om (TKM); aharrasi@unizwa.edu.om (AAH)

OPEN ACCESS

Citation: Mohanta TK, Yadav D, Khan A, Hashem A, Tabassum B, Khan AL, et al. (2020) Genomics, molecular and evolutionary perspective of NAC transcription factors. PLoS ONE 15(4): e0231425. <https://doi.org/10.1371/journal.pone.0231425>

Editor: Serena Aceto, University of Naples Federico II, ITALY

Received: January 2, 2020

Accepted: March 23, 2020

Published: April 10, 2020

Copyright: © 2020 Mohanta et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The author would like to extend their sincere thanks to the Natural and Medical Sciences Research Center, University of Nizwa to facilitate the study. The authors would also like to extend their sincere appreciation to the Researchers Supporting Project Number (RSP-2019/134), King Saud University, Riyadh, Saudi Arabia. The funders had no role in study design, data collection and

Abstract

NAC (NAM, ATAF1,2, and CUC2) transcription factors are one of the largest transcription factor families found in the 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. Study revealed, *Brassica napus* (410) encodes highest number and *Klebsormidium flaccidum* (3) encodes the lowest number of TFs. The study further revealed the presence of NAC TF in the Charophyte algae *K. flaccidum*. On average, the monocot plants encode higher number (141.20) of NAC TFs compared to the eudicots (125.04), gymnosperm (75), and bryophytes (22.66). Furthermore, our analysis revealed that several NAC TFs are membrane bound and contain monopartite, bipartite, and multi-partite nuclear localization signals. NAC TFs were also found to encode several novel chimeric proteins and regulate a complex interactome network. In addition to the presence of NAC domain, several NAC proteins were found to encode other functional signature motifs as well. Relative expression analysis of NAC TFs in *A. thaliana* revealed root tissue treated with urea and ammonia showed higher level of expression and leaf tissues treated with urea showed lower level of expression. The synonymous codon usage is absent in the NAC TFs and it appears that they have evolved from orthologous ancestors and undergone vivid duplications to give rise to paralogous NAC TFs. The presence of novel chimeric NAC TFs are of particular interest and the presence of chimeric NAC domain with other functional signature motifs in the NAC TF might encode novel functional properties in the plants.

analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist

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) [1–4]. The TFs are involved in inducing the transcription of DNA into RNA [5–8]. They include numerous and diverse proteins, all of which contain one or more DNA-binding motifs [8–10]. 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 [11,12]. 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 [13–16]. Other TFs bind to regulatory enhancer sequences and stimulate or repress transcription of the related genes [17–19]. 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 [20–22]. TFs act as a molecular switch for temporal and spatial gene regulation [23,24]. A considerable portion of a genome consists of genes encoding transcription factors. For example, there are at least 52 different TF families in the *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 [25–28]. The dimeric architecture of NAC proteins can remain stable even at a concentration of 5M NaCl [28]. 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 [28–30]. Sub-domains C and D contain positively charged amino acids that bind to DNA [28]. 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 [28]. 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 [26,31,32]. 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 [28]. 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 possess mono or bipartite nuclear localization signals which contain a Lys residue in sub-domain D [25,32–34]. In addition, NAC proteins, as part of a mechanism of self-regulation, also modulate the expression of several other proteins [32,35]. 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 [36]. For example, the NRD domain can suppress the transcriptional activity of Dof, WRKY, and APETALA 2/dehydration responsive elements (AP2/DRE) TFs [36].

Studies indicate that the diverse C-terminal domain contains a transcription regulatory region (TRR) which has several group-specific motifs that can activate or repress transcription activity [37–40]. The C-terminal region imparts differences in the function of individual NAC proteins by regulating the interaction of NAC TFs with various target proteins. Although the C-terminal region of NAC TFs is varied greatly, it also contains group-specific conserved

motifs [41]. Although various aspects of NAC TFs have been studied [42,43], most studies were limited within a few plant species. For example, Zhu et al., (2012) has studied with only 16 species where in few cases they used expressed sequence tag (EST) as well [42] and Pereira-Santana et al., (2015) used 24 land plant species [43] where they were included the genome sequences of unicellular organisms including algae and bacteria. However, Pereira-Santana et al., (2015) did not find any NAC TFs in the algae and bacteria [43]. Therefore, a detailed comparative study of the genomic, molecular biology, and evolution of NAC TFs has across the lineage level of plant kingdom has not been conducted so far. Therefore, a comprehensive analysis of NAC TFs is presented in the current study. We analysed nucleotide and protein data of the NAC TFs to find out the genomic diversity, biochemical, evolutionary, and expression analysis of NAC TFs from 160 plant species.

Materials and methods

Identification of NAC TFs

NAC genes from 160 plant species (9 algae, 3 bryophytes, 1 pteridophyte, 5 gymnosperms, and 142 higher plants) were obtained from searches in the National Centre for Biotechnology Information (<https://www.ncbi.nlm.nih.gov/>), Phytozome, and Plant Genome databases [44,45]. BLASTP (E-value cut-off was $1E-5$) and hidden Markov model were used to identify the NAC TFs in different species using AtNAC1 and AtNAC2 as the query sequences [46]. BLASTP analysis was conducted against the respected proteome of the individual species to find the best hit to minimize the error rate [44]. Protein and CDS sequences of each species were collected and further analysed. Protein sequences of the NAC TFs were subjected to BLASTP analysis against the reference databases NCBI, Phytozome, and Plant Genome Database [44,45] to reconfirm them as a NAC TF of the respective identified species. All of the NAC TF protein sequences in the examined species were also subjected to ScanProsite and InterProScan to confirm the presence of a NAC domain [47,48]. 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 InterProScans [47,48]. The presence of multiple functional sites in NAC TFs were also analysed using ScanProsite software [48].

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 TMHMM server v. 2.0 [49]. 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 [50]. The 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 [50].

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 [51,52]. The interactome network of each of NAC TF were noted and the results were later used to construct the interactome network of *A. thaliana* NAC TFs. The presented interactome network was based on an

experimentally validated network, co-expressed network, and a mined network [52]. These outputs were used to construct the interactome network. The NAC TFs used to construct the interactome network were subjected to GO (gene ontology) and cellular process analyses [52].

Gene expression analysis

Differential gene expression of NAC TFs was analysed to elucidate their role in growth, development, and nitrogen assimilation. *A. thaliana* NAC TFs were used to examine differential gene expression. The transcriptome data from *A. thaliana* treated with ammonia, nitrate, and urea were utilized from the PhytoMine database in Phytozome [44]. The experimental conditions were as follows; the *A. thaliana* seeds were cold stratified in water for 3 days and sown in pots. The pots were placed in the growth chamber (22° C day/20° C night, 14 hrs light with flux density of 350 $\mu\text{mol m}^{-2}\text{s}^{-1}$) and later thinned one plant per pot. When rosette was achieved 7–8 leaves, treatment was conducted. The plants were watered with nutrient solution containing 5mM urea, 10 mM KNO₃ (potassium nitrate), and 10 mM (NH₄)₃PO₄ (ammonium phosphate) for each of individual experiment. The nutrient solutions were supplied at three days interval for four weeks. After four weeks, the leaf, stem, and root tissues were harvested for expression analysis. The expression pattern of NAC TFs for leaf and root tissues in the treated *A. thaliana* plants were analysed separately. The expression was measured in fragments per kilobase of exon per million fragments mapped (FPKM). Transcripts with a zero value were discarded from the study.

Construction of a phylogenetic tree

Two approaches were used to construct the phylogenetic trees. In the first approach, a phylogenetic tree was constructed using the NAC TFs of individual species. In the second approach, the NAC TFs of all of the examined species were combined to construct a phylogenetic tree. The phylogenetic tree for individual species was constructed to determine the deletion and duplication events in NAC TFs within individual species. We excluded the short sequences from the study those resulted in error during the alignment. Prior to the construction of the phylogenetic trees, a model selection was carried out in MEGA6 software [53]. The following parameters were used in the model, analysis, model selection; tree to use, automatic (neighbour joining), statistical method, maximum likelihood; substitution type, nucleotides; gaps/missing data treatment, partial deletion; site coverage cut-off (%), 95; codons included, 1st+2nd+3rd+non-coding. Based on the lowest BIC values of model selection, phylogenetic trees of NAC TFs were carried out using the neighbour joining method, a GTR statistical model, and 1000 bootstrap replicates.

Analysis of transition and transversion rates

Transition and transversion rates in NAC TFs within individual species were analysed using MEGA6 software [53]. The converted MEGA file format of individual species was used to determine the rate of transition and transversion. The following statistical parameters were used to study the transition/transversion rate: estimate transition/transversion bias; maximum composite likelihood estimates of the pattern of nucleotide substitution; substitution type, nucleotides; model/method, Tamura-Nei; gaps/missing data treatment, pairwise deletion; codon position, 1st, 2nd, 3rd, and non-coding sites.

Analysis of gene deletion and duplication

Prior to the analysis of deletion and duplication events in NAC TFs, a species tree was constructed in the NCBI taxonomy browser (<https://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi>). All of the studied species were used to construct the species tree. The resulting phylogenetic trees of individual species in a nwk file format were uploaded in Notung 2.9 software [54] as a gene tree and reconciled as a gene tree with the species tree to obtain duplicated and deleted genes. Deletion and duplication events were analysed in all of the studied species individually.

Results and discussion

NAC transcription factors exhibit diverse genomic and biochemical features

Advancements in genome sequencing technology have enabled the discovery of the genomic details of large number of plant species. The availability the genome sequence data allowed us to study the genomic details of NAC TFs in diverse plant species. The presence of NAC TFs in 160 species (18774 NAC sequences) was identified and served as the basis of the conducted analyses. Comparisons of NAC sequences revealed that *Brassica napus* has the highest number (410) of NAC TFs, while the pteridophyte plant, *Marchantia polymorpha*, was found to contain the lowest number (9) (Table 1). On average, monocot plants contain a higher (141.20) number of NAC TFs relative to dicot plants (125.56). Except for *Hordeum vulgare* (76), *Saccharum officinarum* (44), and *Zostera marina* (62) all other monocot species possess more than one hundred NAC TFs each (Table 1). Lower eukaryotic plants, bryophytes and pteridophytes also possess NAC TFs. In addition, the algal species, *Klebsormidium flaccidum*, also contains NAC TFs and this finding represents the first report of NAC TFs in algae (Table 1). A NAC TF in *Trifolium pratense* (Tp57577_TGAC_v2_mRNA14116) was found to be the largest NAC TF, comprising 3101 amino acids, while a NAC TF in *Fragaria x ananassa* (FANhyb_i-con00034378_a.1.g00001.1) was found to be the smallest NAC TF, comprising only 25 amino acids. Although it only contains a 25 amino acid sequence, it still encodes a NAC domain. Typically, NAC TFs contain a single NAC domain located near the N-terminal region of the protein. The current analysis, however, also identified NAC TFs with two NAC domains. At least 77 of the 160 studied species were found to contain two NAC domains (Table 1).

Multiple sequence alignment revealed the presence of a conserved consensus sequence at the N-terminus. The major conserved consensus sequences are P-G-F-R-F-H-P-T-D-D/E-L-I/V, Y-L-x₂-K, D-L-x-K-x₂-P-W-x-L-P, E-W-Y-F-F, G-Y-W-K-A/T-T-G-x-D-x₁₋₂-I/V, G-x-K-K-x-L-V-F-Y, and T-x-W-x-M-H-E-Y. Among these consensus sequences, D-D/E-L-I/V, E-W-Y-F-F, G-Y-W-K, and M-H-E-Y are the conserved motifs most observed. The D-D/E-L-I/V motif is a characteristic feature of the calcium-binding motifs present in the EF-hand of calcium-dependent protein kinases and the presence of this motif in NAC TFs indicates that they have the potential to regulate Ca²⁺ signalling events in cells [55]. The D-D-E/E motif is located in the β' sheet whereas the Y-L-x₂-K motif is in the α1a/b chain. Except for G-F-R-F-H-P-T-D-D/E-L-I/V, the conserved consensus sequences contain the positively charged amino acids Lys (L) and Arg (K) that can bind to negatively charged DNA. Welner et al. (2012) published the crystal structure of ANAC019 and reported that Y⁹⁴-W-K-A-T-G-T-D in β3, I¹¹-K-K-A-L-V-F-Y of β4, K¹²³-A-P-K-G-T-K-T-N-W in the loop between β4 and β5, and I¹³³-M-H-E-Y-R of β5 and Y¹⁶⁰-K-K-Q at the C-terminal end are located close to the bound DNA and are associated with DNA binding activity [56]. They reported that Y⁹⁴-W-K-A-T-G-T-D is responsible for the specific recognition of DNA and binds at the major groove within DNA,

Table 1. Genomic details of NAC TFs of plants. NAC TFs have not undergone conditional duplication and none of a NAC TF gene has lost. In addition, transfer of NAC TFs was not observed from one species to another.

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 paralogous genes
Monocots						
1	<i>Aegilops tauschii</i>		4	117	114	114
2	<i>Brachypodium distachyon</i>	2	1	137	135	135
3	<i>Brachypodium stacei</i>	1	1	128	127	127
4	<i>Hordeum vulgare</i>			76	76	76
5	<i>Leersia perrieri</i>	5	2	163	162	162
6	<i>Oropetium thomaeum</i>	1		118	103	103
7	<i>Oryza barthii</i>		4	134	138	138
8	<i>Oryza brachyantha</i>	1	1	118	110	110
9	<i>Oryza glaberrima</i>	1		116	110	110
10	<i>Oryza glumipatula</i>	2		140	139	139
11	<i>Oryza longistaminata</i>	1	6	125	98	98
12	<i>Oryza meridionalis</i>	2	2	127	123	123
13	<i>Oryza nivara</i>	4	1	146	130	130
14	<i>Oryza punctata</i>	6	1	135	133	133
15	<i>Oryza rufipogon</i>	4	3	136	129	129
16	<i>Oryza sativa subsp. indica</i>	1	3	157	156	156
17	<i>Oryza sativa subsp. japonica</i>	1		139	138	138
18	<i>Panicum hallii</i>	3	6	139	126	126
19	<i>Panicum virgatum</i>	9	6	310	309	309
20	<i>Phoenix dactylifera</i>	3	1	124	123	123
21	<i>Phyllostachys edulis</i>			125	124	124
22	<i>Phyllostachys heterocycla</i>	2	2	125	124	124
23	<i>Saccharum officinarum</i>			44	33	33
24	<i>Setaria italica</i>	4		139	134	134
25	<i>Setaria viridis</i>	1		135	118	118
26	<i>Sorghum bicolor</i>	1		141	134	134
27	<i>Spirodela polyrhiza</i>			55	48	48
28	<i>Triticum aestivum</i>	2	2	263	209	209
29	<i>Triticum urartu</i>		1	103	74	74
30	<i>Zea mays</i>	1	1	130	119	119
31	<i>Zostera marina</i>	1		62	55	55
32	<i>Zoysia japonica</i>		4	176	160	160
33	<i>Zoysia matrella</i>	1	3	313	230	230
34	<i>Zoysia pacifica</i>	1	2	205	183	183
Dicots						
35	<i>Actinidia chinensis</i>	1	5	167	166	166
36	<i>Aethionema arabicum</i>	3		85	84	84
37	<i>Amaranthus hypochondriacus</i>	1		44	37	37
38	<i>Amborella trichopoda</i>			46	45	45
39	<i>Ananas comosus</i>		1	73	72	72
40	<i>Aquilegia coerulea</i>			80	79	79
41	<i>Arabidopsis halleri</i>	2		94	93	93
42	<i>Arabidopsis lyrata</i>	4	1	122	121	121

(Continued)

Table 1. (Continued)

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 paralogous genes
43	<i>Arabidopsis thaliana</i>	5		113	112	112
44	<i>Arabis alpina</i>	1		82	81	81
45	<i>Arachis duranensis</i>			82	81	81
46	<i>Arachis hypogaea</i>			162	161	161
47	<i>Arachis ipaensis</i>			83	81	81
48	<i>Artemisia annua</i>			28	27	27
49	<i>Azadirachta indica</i>			183	182	182
50	<i>Beta vulgaris</i>			53	52	52
51	<i>Boechnera stricta</i>	2		123	122	122
52	<i>Brassica napus</i>	10	7	410	409	409
53	<i>Brassica oleracea</i>	4	3	271	270	270
54	<i>Brassica rapa</i>	4	2	256	255	255
55	<i>Cajanus cajan</i>			96	95	95
56	<i>Camelina sativa</i>	17	3	341	330	330
57	<i>Cannabis sativa</i>			58	57	57
58	<i>Capsella grandiflora</i>	2		95	94	94
59	<i>Capsella rubella</i>	5		119	118	118
60	<i>Capsicum annum</i>			96	95	95
61	<i>Carica papaya</i>			82	81	81
62	<i>Castanea mollissima</i>	4		91	78	78
63	<i>Catharanthus roseus</i>		2	121	120	120
64	<i>Chenopodium quinoa</i>		1	96	95	95
65	<i>Cicer arietinum</i>			96	95	95
66	<i>Citrullus lanatus</i>			80	79	79
67	<i>Citrus clementina</i>			129	128	128
68	<i>Citrus sinensis</i>	2		145	143	143
69	<i>Coffea canephora</i>			63	62	62
70	<i>Cucumis melo</i>			92	91	91
71	<i>Cucumis sativus</i>			83	80	80
72	<i>Daucus carota</i>		2	96	95	95
73	<i>Dianthus caryophyllus</i>			79	77	77
74	<i>Dichanthelium oligosanthes</i>	8	2	131	100	100
75	<i>Doroceras hygrometricum</i>		2	83	76	76
76	<i>Elaeis guineensis</i>	2	1	170	167	167
77	<i>Eragrostis tef</i>	8	3	172	165	165
78	<i>Eucalyptus camaldulensis</i>			200	124	124
79	<i>Eucalyptus grandis</i>			164	150	150
80	<i>Eutrema salsugineum</i>	2		122	104	104
81	<i>Fragaria vesca</i>	3	6	127	123	123
82	<i>Fragaria x ananassa</i>	2	1	98	97	97
83	<i>Genlisea aurea</i>		1	45	42	42
84	<i>Glycine max</i>			180	175	175
85	<i>Glycine soja</i>		1	173	166	166
86	<i>Gossypium arboreum</i>			150	146	146
87	<i>Gossypium hirsutum</i>	1	2	306	296	296
88	<i>Gossypium raimondii</i>			153	145	145

(Continued)

Table 1. (Continued)

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

(Continued)

Table 1. (Continued)

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

<https://doi.org/10.1371/journal.pone.0231425.t001>

whereas I¹¹-K-K-A-L-V-F-Y, K¹²³-A-P-K-G-T-K-T-N-W, I¹³³-M-H-E-Y-R, and Y¹⁶⁰-K-K-Q bind to the backbone of the DNA molecule and provide affinity for DNA binding activity [56]. In the present analysis of 160 plant species, the identification of the conserved consensus sequences G-Y-W-K-A/T-T-G-x-D-x₁₋₂-I/V, G-x-K-K-x-L-V-F-Y, and T-x-W-x-M-H-E-Y is in agreement with Welner et al (2012); suggesting that NAC TFs contain conserved consensus sequences for specific DNA recognition and increasing the affinity for DNA binding.

Hao et al., (2010) reported that the D subunit of NAC TFs contain a hydrophobic L-V-F-Y amino acid motif that partially suppresses the WRKY, Dof, and APETALA2 transcriptional regulators [36]. This suggests that NAC TFs function as a negative regulator of transcription for WRKY, Dof, and APETALA 2/ dehydration responsive element. The sequence alignment, however, revealed the presence L-V-F-Y transcriptional repressor motif in NAC TF family proteins in diverse plant species. If all the NAC TF with L-V-F-Y motif will suppress the

transcriptional activity of WRKY, Dof, and APETALA 2, it will be challenging for the plants to sustain its cellular and biological activities.

The molecular weight of NAC TFs ranged from 2.94 kDa (*Fragaria x ananassa*_FANhyb_i-con00034378_a.1.g00001.1) to 346.46 kilodaltons (kDa) (*Trifolium pratense*_Tp57577_T-GAC_v2) (Fig 1). Among the studied NAC TFs, only 10 NAC proteins have a molecular weight (MW) more than 200 kDa and 99 are between 100 to 200 kDa. The MW of the majority of the NAC proteins range between 40 to 55 kDa (Fig 1). The average molecular weight of the plant proteins falls in the same range (average 48.256 kDa) as found in the case of *A. thaliana* proteome) [57].

The Isoelectric point (pI) of the NAC proteins ranged from 11.47 (Brast01G304500.1.p, (*Brachypodium stacei*) to 3.60 (ObartAA03S_FGP19036, *Oryza barthii*). The majority of the NAC TFs fell within a pI range of 5–8 (Fig 2). Among the 18774 analysed NAC TFs, the pI of 99 proteins were ≥ 10 . Approximately 69.28% of the NAC TFs had a pI that was in an acidic range, whereas the remaining 30.72% had a pI within a basic range. A protein with a pI below the pI carries a net positive charge, whereas a protein with a pI above the pI carries a net negative charge. The pI of a protein determines its transport, solubility, and sub-cellular localization [57–60]. Biomembranes, such as those surrounding the nucleus, are negatively charged; as a result, positively charged (acidic pI) NAC TFs are readily attracted to the nuclear membrane and subsequently transported into the nucleus to function in transcriptional regulation. There are, however, approximately 30.72% NAC TFs that possess a basic pI; suggesting that they are localized in the cytosol or plasma membrane of the cell. The major role of the TFs is to bind to specific DNA sequences to regulate transcription. The majority of the proteins have either an acidic or basic pI and those with a neutral pI close to 7.4 are few because proteins tend to be insoluble, unreactive, and unstable at a pH close to its pI. This is the main reason why among the 18774 NAC TFs analysed, only two (XP_010925972.1, *Elaeis guineensis*; Lus10008200, *Linum usitatissimum*) had a pI 7.4. The existence of NAC proteins with a pI above 10 led us to speculate whether these TFs function while attached to a transmembrane domain. Therefore, additional analyses were conducted to determine if NAC TFs also have the potential to bind to the transmembrane domain or if the NAC TFs with a basic pI remain within the cytosol.

NAC TF proteins are membrane bound

Transcription factors regulate diverse cellular events at transcriptional, translational, and post-translational levels. They are also involved in nuclear transport and posttranslational modifications. In several cases, TFs are synthesized but remain inactive in the cytoplasm and are only induced into activity through non-covalent interactions [61,62]. TFs are able to remain inactive through their physical association with intracellular membranes and are released by proteolytic cleavage. NAC TFs are a family of proteins whose numbers are in the hundreds in the majority of plant species. The fact that NAC TFs are such a large protein family, it is not surprising that NAC TFs have evolved diverse functional roles. Therefore, it is plausible that NAC TFs may be associated with sub-cellular organelle other than the nucleus to fulfil their diverse functional roles. It is essential, however, to confirm if NAC TFs contain signalling sequences for transmembrane localization. Therefore, we analysed the NAC gene sequences to determine if the signalling sequences present in NAC TFs possess a transmembrane domain.

Results indicated that at least 2190 (8.57%) NAC TFs possess a transmembrane domain (S1 Fig, S1 File). Transmembrane domains were found at both the N- and C-terminal ends of NAC proteins. In the majority of the cases, however, the transmembrane domain was located towards the C-terminal end. Seo et al., (2008) indicated the presence of a transmembrane

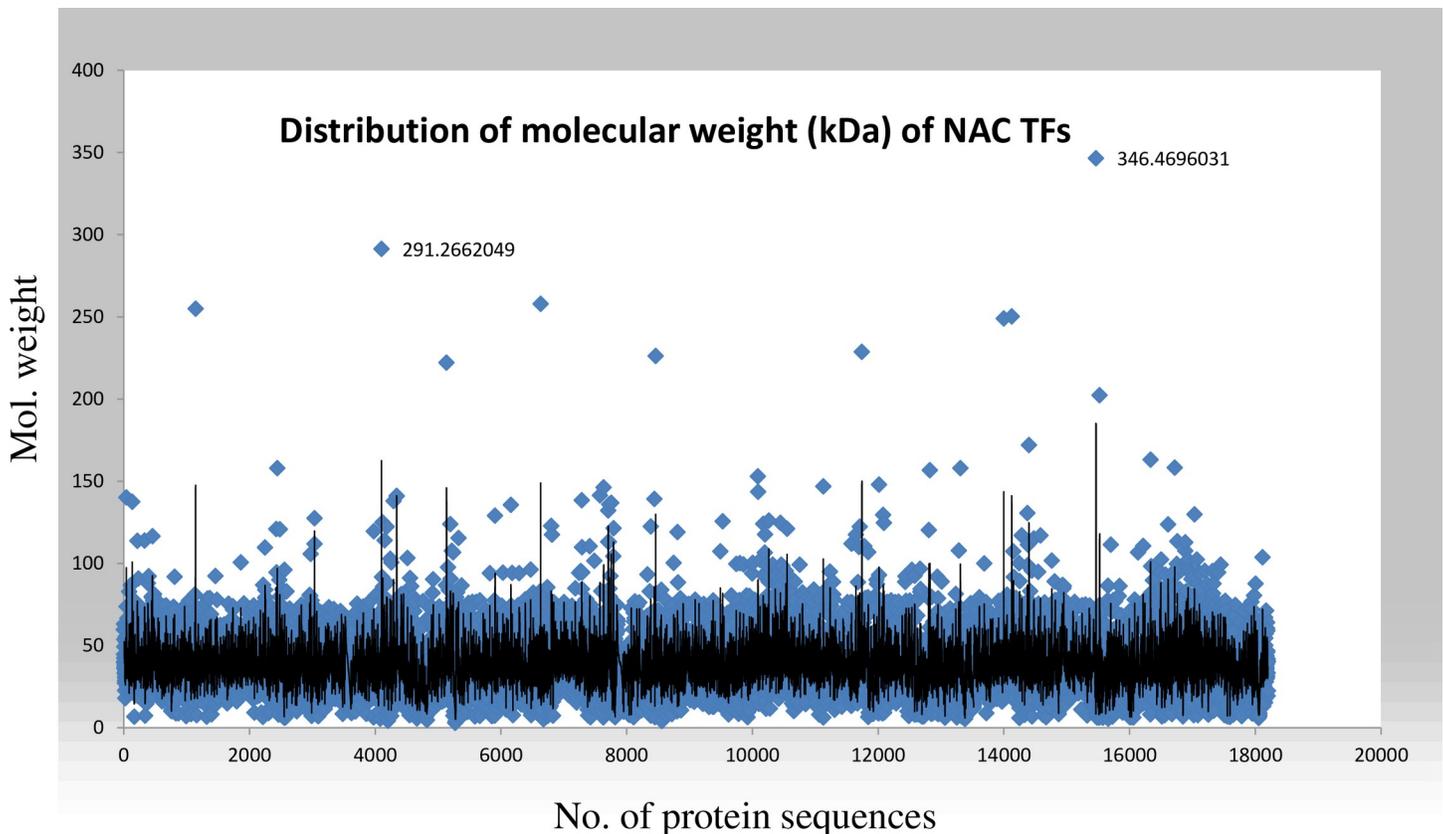


Fig 1. The distribution of the molecular weight of NAC TFs. The molecular weight of NAC TFs ranged from 2.94 kDa (*Fragaria x ananassa*, FANhyb_Icon00034378_a.1.g00001.1) to 346.46 kDa (*Trifolium pratense*, Tp57577_TGAC_v2_mRNA14116). The average molecular weight of NAC TFs was 38.72 kDa. In total, 17158 NAC TFs were utilized in the analysis of molecular weight. The analysis was conducted using a protein isoelectric point calculator (<http://isoelectric.org/>).

<https://doi.org/10.1371/journal.pone.0231425.g001>

domain in TFs and suggested that transmembrane domain functions through two proteolytic mechanisms, commonly known as regulated ubiquitin/proteasome-dependent (RUP) and regulated intramembrane proteolysis (RIP) [63,64]. The bZIP plant TF is present as an integral membrane protein associated with stress response in the endoplasmic reticulum (ER) [65–68]. Studies suggest that the majority of membrane bound TFs are associated with the ER and a membrane bound TF was also found to be involved in cell division [69,70]. At least 10% of the TFs in *Arabidopsis thaliana* have been reported to be transmembrane bound [70]. The collective evidence clearly indicates that membrane-mediated transcriptional regulation is a common stress response and that NAC TFs play a vital role in stress resistance in the ER. Therefore, these membrane-bound NAC TFs can be of great importance for the manipulation of stress resistance using biotechnology.

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

The import of NAC TFs into the nucleus is mediated by nuclear membrane-bound importins and exportins that form a ternary complex consisting of importin α , importin β 1, and a cargo molecule. Importin α serve as an adaptor molecule of importin β 1 and recognises the nuclear localization signal (NLS) of the cargo protein needing to be imported. Importin β 1 and β 2, however, also recognize the NLS directly and bind to the cargo protein. Although the NLS of

Distribution of isoelectric points (pI) of NAC TFs

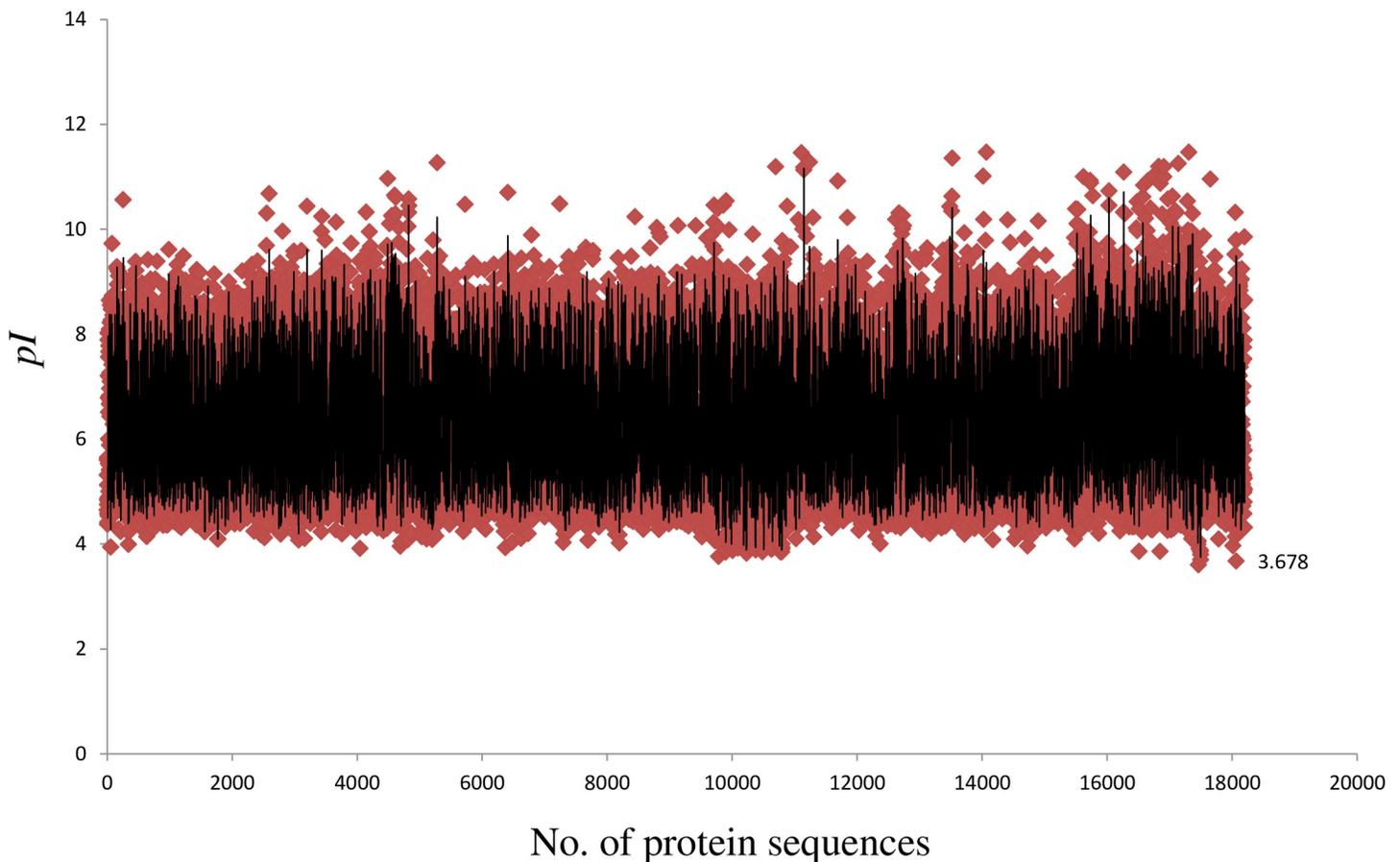


Fig 2. The distribution of the isoelectric point of NAC TFs. The isoelectric point of NAC TFs ranged from pI 3.78 (OB07G17140.1, *Oryza brachyantha*) to pI 11.47 (Sevir.3G242500, *Setaria viridis*). The average isoelectric point of NAC TFs was 6.38. A total of 17158 NAC TFs were utilized in the analysis of the pI of NAC TFs. The analysis of pI was conducted using a protein isoelectric point calculator (<http://isoelectric.org/>).

<https://doi.org/10.1371/journal.pone.0231425.g002>

TFs have been widely studied in the animal kingdom, their study in plants has been more restricted. Therefore, the NLS of NAC TFs was examined in the current study. Results indicate that NAC TFs contain diverse NLS. The NLS were found in the N- and C-terminal regions of NAC TF proteins. Some NAC TFs were found to contain only one NLS whereas other contain multiple NLS. At least 3579 of the total NAC TFs analysed were found to contain either one or multiple NLS (S2 Fig, S2 File). More specifically, 2604 NAC TFs were found to possess only one NLS at the N-terminal end of the NAC protein, whereas 975 were found to possess two NLS, 254 possess three NLS, and 48 were possess four NLS. The NLS were located towards the N-terminal end in the majority of NAC proteins.

NLS motifs are rich in positively charged amino acids and bind to importin α to be imported into the nucleus. The NLS motifs are classified as monopartite or bipartite. A monopartite NLS contains a single cluster of positively charged amino acids and are grouped into two subclasses, class-I and class-II. Class-I possesses four consecutive positively charged amino acids and class-II contains three positively charged amino acids, represented by K(K/R)-x-K/R; where x represents any amino acid that is present after two basic amino acids. Bipartite NLS motifs contain two clusters of positively charged amino acids separated by a 10–12

amino acid linker sequence. Bipartite NLS motifs are characterised by the consensus sequence K-R-P-A-A-T-K-K-A-G-Q-A-K-K-K-K. In addition to monopartite and bipartite NLS motifs, importin α also recognises non-canonical NLS motifs. Non-canonical NLS motifs are longer and considerably variable relative to monopartite and bipartite NLS motifs and are classified as class-III and class-IV NLS. Non-canonical NLS motifs are usually present in the C-terminal end and bind with importin β 2. Class-III and class-IV NLS motifs contain K-R-x(W/F/Y)-x₂-A-F and (P/R)-x₂-K-R-(K/R) consensus sequences, respectively. We identified at least 1702 unique NLS consensus sequences in the N-terminal region of NAC TFs. The monopartite class I NLS motifs were found to contain more than four consecutive basic amino acids with the number of their consecutive basic amino acids ranging from four to fourteen (K-K-K-K-K-K-K-K-K-K-K-K-K). The bipartite NLS motifs contain two clusters of consecutive basic amino acids separated by up to twenty-four linker amino acids (K-K-K-x₃-R-x₂-R-x₄-K-x₃-K-x₃-K-x-K-x₂-R-K-K).

The non-canonical NLS motifs contain at least six centrally-located, positively charged amino acids (K-x-R-R-R-P-R-R-x₂-R-K) flanked by positively charged amino acids on both sides. Our analysis of the N-terminal NLS of NAC TFs, however, did not identify any NAC TFs containing this consensus sequence. Instead, several new variants of this consensus sequence were identified with multiple clusters of positively charged amino acids. These NLS were designated as multipartite NLS motifs (Table 2, S2 Fig, S2 File). Much of the diversity of NLS motifs is associated with the sequence of the variable linker amino acids. In our analysis, we removed the linker amino acid sequences, represented as x, to obtain a more concise

Table 2. Putative multipartite nuclear localization signal sequences of NAC transcription factor proteins. The underlined amino acids are designated as NLS and letter x denoted as any amino acid.

C-terminal multipartite NLS	N-terminal multipartite NLS
<u>R-K-R-x-R-x-R-K-K-x₄-K-x-K-K-R-x₃-K-x₃-K-K-x₃-R-R-K-x₂-K</u>	<u>K-K-K-K-x₇-K-K-K-K-x₇-K-K-K-K</u>
<u>R-R-R-x₄-K-K-x₆-R-x₂-R-x₂-R-R-x₄-R-R-R-x₆-R-x₂-R-R-x₉-R-R-R-R-R-R-x₂-R-R</u>	<u>K-K-K-K-x-K-x₅-K-x-K-K-x₇-K-K-K-K-x₂-K-K-K</u>
<u>K-K-K-x₄-K-K-x-K-x₅-K-x₄-K-K-K-R-x-K-R-K-x-K-x₄-K-K-K-R-K-K</u>	<u>K-K-K-x₂-K-K-x-K-x₅-K-x₄-K-K-K-R-x-K-R-K-x-K-x₄-K-K-K-R-K-K</u>
<u>K-K-R-x₄-K-x₂-K-x-K-x₂-K-K-R-x-R-K-x₄-K-x₂-K-x-K-K-R-x-R-K-x₄-K-x₂-K-x-K-x-R</u>	<u>K-K-R-x-R-K-x₂-K-x-K-x₂-K-K-K-x-RK-x₂-K-R-R-x₂-K-K-K-x-R</u>
<u>K-K-R-x-R-K-x₂-K-x-K-x₂-K-K-K-x-R-K-x₂-K-R-R-x₂-K-K-K-x-R</u>	<u>K-K-R-x-R-K-x₂-K-x-K-x₂-K-K-R-x-R-K-x₂-K-x-K-x₂-K-K-R-x-R-K-x₂-K-x-K-x-R</u>
<u>K-K-R-x-R-K-x₂-K-x-K-x₂-K-K-R-x-R-K-x₂-K-x-K-x₂-K-K-R</u>	<u>K-x₂-K-K-K-x₃-K-K-K-K-K-x-K-x₈-K-x₉-K-x₂-K-K-R-x₂-K-K-K-K-x-K</u>
<u>R-K-R-x-R-x₃-K-K-R-R-x₂-K-x₉-K-x₄-R-x-K-x₂-R-x-R-R-x₅-K-K-R</u>	<u>K-x₂-K-K-K-x₃-K-x-K-K-K-x-K-K-K-x₂-K-K-K-x-K</u>
<u>R-K-R-x-R-x-R-x₅-K-x-K-K-K-R-x₃-K-x₄-K-R-x₂-R-R-K</u>	<u>R-K-R-x-R-x-R-K-K-x₂-K-x-K-K-K-R-x₂-K-x₂-KK-x₂-R-R-K-x₂-K</u>
<u>R-R-x-R-R-R-x-R-R-x₈-R-x₆-R-R-x₅-R-R-R-x-R-x₅-R-x₈-R-R-R-R</u>	<u>R-K-R-x-R-x-R-x₂-K-x-K-K-K-R-x₂-K-x₄-K-R-x₂-R-R-K-x-K-x₂-R</u>
<u>R-R-x-R-R-R-x-R-R-x₉-R-x₂-R-R-K-R-K-x-R-x₄-R-R-R-R-R-R-x₄-R-K</u>	
<u>R-x-R-R-R-R-x₆-R-x₁₁-R-x₈-R-R-x₃-R-R-R-x₂-R-R-x-R-x-R-x₆-R-R-R-R-R-x₄-R-R-x₂-R</u>	
<u>R-x-R-R-x₃-K-R-R-R-x₂-R-x-R-R-x-R-x-R-x₇-R-x₃-R-R-R-x₇-R-x₂-R-R-R-R</u>	
<u>R-x-R-x-R-R-R-x₃-R-R-R-x₃-R-x-R-x₂-R-x₄-R-R-R-x₅-R-K-x-R-x₃-R-R-x₁₃-R-R-x-K-x₅-R-R-x₆-K-R-R</u>	

<https://doi.org/10.1371/journal.pone.0231425.t002>

picture of NLS diversity. Removing the linker amino acids present in monopartite, bipartite, and multipartite NLS motifs resulted in the identification of 97 different NLS consensus sequences in the N-terminal region of NAC TFs (S2 Fig, S2 File). 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 (S2 File).

The C-terminal end of NAC TF proteins also contain monopartite, bipartite, and multipartite NLS motifs (Table 2, S2 Fig, S2 File). Removal of the linker amino acids present in between the consecutive basic amino acids, resulted in the identification of 94 unique consensus sequences. Some of the important NLS found in the C-terminal end were K-K-K (144), K-K-R (83), R-R-R (65), K-R-K (60), K-K-R-K-R (58) and others (Table 2, S2 Fig, S2 File). 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 whereas nine NLS consensus sequences were unique to the C-terminal region (Table 2, S2 Fig, S2 File). Up to six classes of NLS have been reported to be associated with importin α subunit [71]. To the best of our knowledge, this is the first report describing such a high level of diversity and dynamism in the NLS consensus sequences of NAC TFs and plant transcription factors in general. This is also the first report of the presence of unique NLSs in the N- and C-terminal regions of NAC TFs.

Several nuclear-associated proteins contain NLS, as well as nuclear export signals (NESs). Proteins that perform their function within the nucleus need to be exported out of the nucleus and into the cytoplasm to undergo proteosomal degradation. Therefore, a NES is required in addition to an NLS. A Ran-GTP complex binds directly to an NES and mediates the nuclear export process of the cargo molecules [72]. NES sequences contain a hydrophobic, conserved L-V-F-Y (substitute L-V/I-F-M) motif separated by variable linker amino acids at both ends [73]. The presence of an L-V-F-Y motif in all NAC proteins, suggests that all NAC proteins have the potential to be exported out of the nucleus. Hao et al. (2010), however, reported that the hydrophobic L-V-F-Y motif functions as a transcriptional repressor of WRKY, Dof, and APETALA TFs. If the L-V-F-Y motif (S3 Fig) acts as a transcriptional repressor, then the transcriptional activity of these TFs would be affected; resulting suppressed transcriptional activities. Therefore, we feel that the L-V-F-Y motifs might not function as a transcriptional repressor for WRKY, Dof, and APETALA 2 transcription factor. Instead it act as a nuclear export signal sequence as reported by Kosugi et al. (2008) [73].

NAC TFs possess a complex interactome network

The interacting partner of a protein can provide significant information about its potential function and an entire protein-protein interactome network can greatly assist in unravelling the signalling cascade of the proteins. Different cascades are interlinked in signalling systems and form intricate constellations that provide information about cell response and function. Thus, the interactome network of NAC TFs in *A. thaliana* were explored. The presence of a dynamic network was revealed, and a diverse set of interacting protein partners of NAC TFs were identified (Fig 3, Table 3). The NAC TFs frequently interact with *ABI* (*ABSCISIC ACID INSENSITIVE*), *VND7* (*VASCULAR RELATED NAC DOMAIN*), *MYB* (*MYELOBLASTOSIS*),

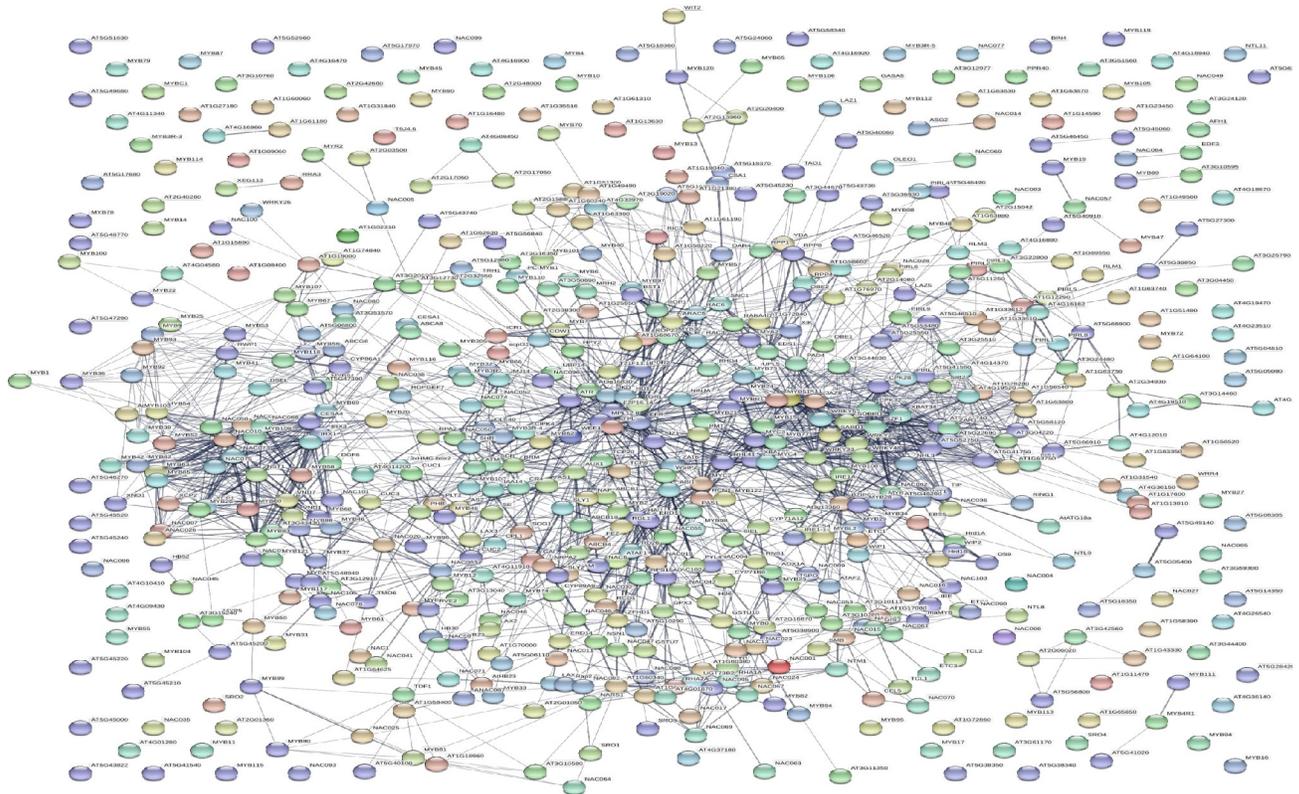


Fig 3. Interactome network of NAC TFs. The interactome network of NAC TF reflects a diverse complex of interacting proteins. The NAC TFs of *A. thaliana* were utilized in the interactome network analysis. The interactome map of *A. thaliana* was determined using the string database (<https://string-db.org>).

<https://doi.org/10.1371/journal.pone.0231425.g003>

DREB2A (DEHYDRATION RESPONSIVE ELEMENT BINDING), *DREB2G*, *WRKY*, *JMJ* (*JUMONJI*), *LEA* (LATE EMBRYOGENESIS ABUNDANT), *KNAT* (*KNOX TAIL*), *CUC* (*CUP SHAPED COTYLEDON*), *MC5* (*METACASPASES 5*) and other important genes involved in plant growth, development, and stress responses (Table 3). In addition, NAC TFs were also found to interact with other NAC TFs as well (Table 3).

The expression of several of NAC genes are either up- or down-regulated by auxin, ethylene, or ABA, suggesting that NAC TFs play a role in plant hormonal signalling [74–76]. One of the most challenging aspects of a protein-protein interactome network is that the interaction can vary depending upon the cell and its environment [77]. Therefore, it is necessary to investigate the dynamic interactions of proteins in different cells and environmental conditions to completely understand their interacting partner and the cellular function of the TF. NAC TFs regulate *ERD* and *NCED* (ABA biosynthesis) genes through a direct interaction with their promoters [78,79]. NAC TFs (*ANAC019*, *ANAC055*, and *ANAC072*) interact with *ERD1* which encodes a Clp protease regulatory subunit [80]. The overexpression of one of these three NAC TFs, however, did not induce the up-regulation of *ERD1* because the induction of *ERD1* depends on the co-expression of a zinc finger homeodomain TF, *ZFHD1* [80]. *ANAC019* and *ANAC055* interact with *ABI* (abscisic acid insensitive), and at least five MYB TFs can bind to the NAC TF promoter region [81,82]. In this case, the NAC DNA binding domain mediates the interaction with *RHA2A* and *ZFHD1* [82].

Table 3. Interacting partners of NAC TFs in plants. *A. thaliana* NAC TFs was used to construct the interactome network. Asterisk indicates no interaction.

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	***	****	NTL, PLP transferase
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	HB4, LZFI, NTL, BZIP61, MYB30, RSW3
NAC16			NYE, NYC1, EEL, ABF2, PAP20, UTR1, TAG1
NAC17			TAG1, UTR1, UTR3, WRKY15, RGF6, FRU, AOX1A, NTL
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	*****	*****	TOM2A, TOM2B, TOM3, ARLA1C, ARLA1D, DBP1, PDLP2, OBE2
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, XBAT34, AT5G52750, SOBIR1, RING1, WRKY53, WRKY46, SARD1,	AT5G42050
NAC38	BRM	MYB69, CIPK4, ABCA8	AT4G29770, AIP2, SDE3
NAC40			NTL, MEE59, NPX, SCP2, SCO1, PUB18, PUB19, LB20
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, WOXY5, LBD19
NAC50	JMJ14, NAC052, GAI, TPL	NAC52, MJM14	JMJ14, PPR, NAC52, AT5G41650, CYP71A25
NAC52	JMJ14, NAC50	JMJ14, PPR, UBP14	JMJ14, NAC50, PPR, CRCK2, PPD6, MFDX1, CYP71A25
NAC53	****	BZIP60, UGT73B, DREB2A, MYB27	NTL, PUM4, MYB103,

(Continued)

Table 3. (Continued)

NAC TFs	Experimental Interactions	Co-expression	Text mining Interactions
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	****	ABI4, DREB2G, WOX12	NACA5, NTL, SCP2, SCO1, ZFP3, GRF7
NAC61	****	NAC90, ACS4,	NAC44, LEA, NAC85, NAC95, NAC90,
NAC62	****	BZIP60, CZF, WRKY33, TIP, SZF1, CPK32, CPK28, TET8,	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
NAC68	****	BZIp60, NAC62	NTL, LPP gamma, LINC2, DEG9, S1P, ENODL17, RPL23AB
NAC69	****	NAC95	NTL, IAA30, RIN3, SPT16, RLP18
NAC71	****	WNK, TM6, AT1G64625	Rap2.6L, AT2G41870, RAP2.4
NAC73	****	MYB46, MYB83, IRX1, IRX3, CESA4	MYB46, MYB83, IRX1, IRX3, MYB63, CESA4
NAC74	F2P16.14, TOPLESS, BRM	DSEL, scpl31, HXXXD type	SCRL20, F-ox/LLR, sks11
NAC75	****	RING/U-box	GATA5, LBD15, GATA12, JLO, scpl48, RNS3, EIF3E, SHM7
NAC76	VND7, NAC83	****	VND7, NAC83, UBP, MYB46
NAC77	*****	*****	DOT5, NAC23, LBD10, NF-YB7, MYB84, GRF5, GRF7, RR8
NAC78	*****	PIP-3	NTL, MAYB27, MYB103, PUM4, KNAT2, KNAT6, SUF4, GH9B8
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	VAP27-1, TSPO, TI1,	****	BZIP28, BZIP60, MC5
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, UBP, RNS3
NAC102	****	ATAF1, tolB, NAC32, RHL41, ZAT6, UGT73B2	ATAF1, NAC32
NAC103	****	****	BZIP60, BZIP28, D111, CLPTM1, NAC44
NAC105	VND7, NAC83,	****	VND7, GH, NAC83, UBP, LAC1, MYB46, RIC4

<https://doi.org/10.1371/journal.pone.0231425.t003>

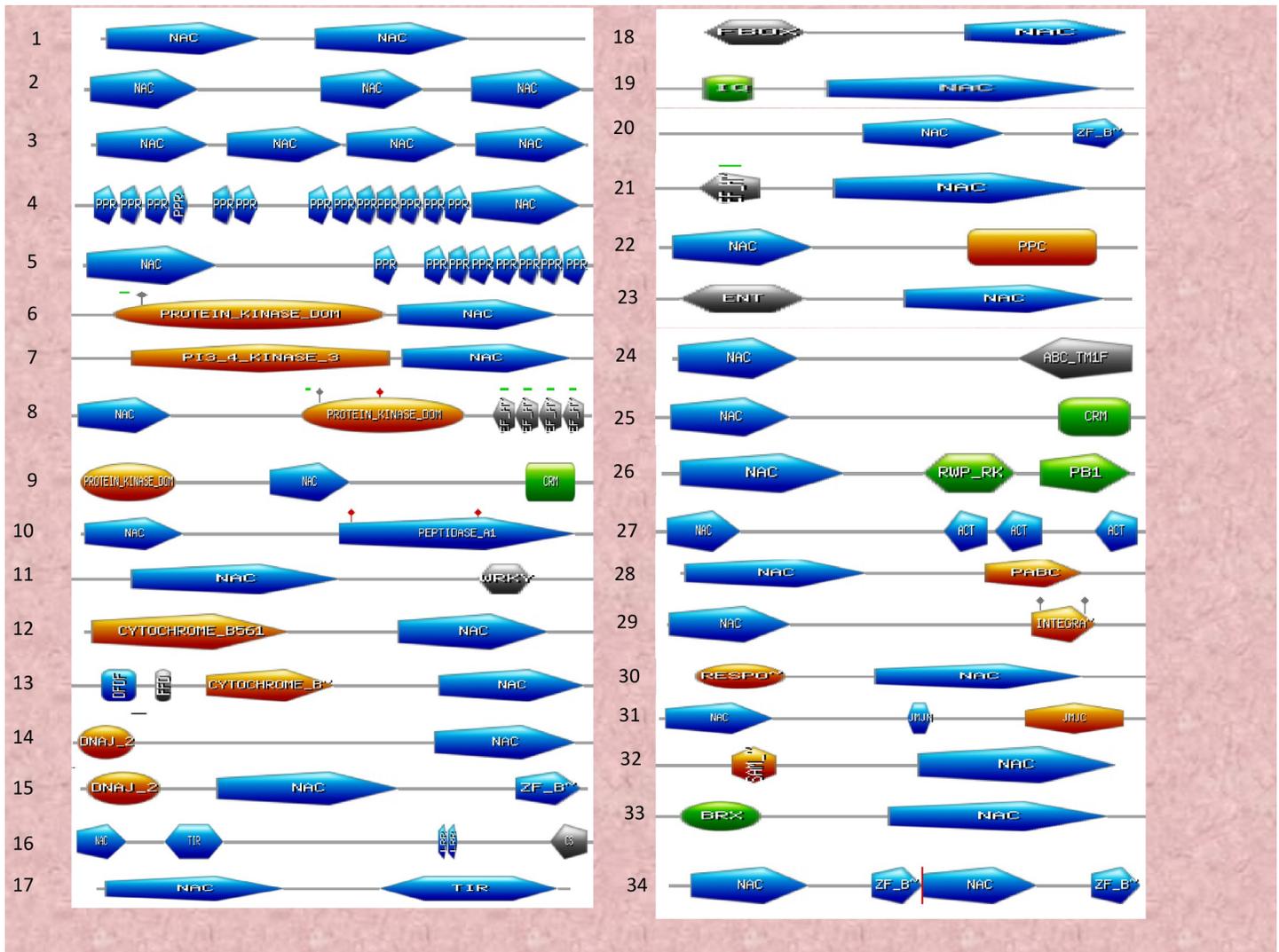


Fig 4. Chimeric NAC domains. NAC TFs possess chimeric NAC domains with at least 34 diverse chimeric NAC domains identified in the studied species. (1) two NAC domain (2) three NAC domain (3) four NAC domain (4) 13 PPR repeats followed by a NAC (5) NAC domain followed by eight PPR repeats (6) protein kinase domain followed by NAC (7) PI3_kinase_3 domain followed by NAC (8) NAC domain followed by kinase and EF-hand domain (9) protein kinase domain followed by NAC and CRM domain (10) NAC domain followed by peptidase A1 domain (11) NAC domain followed by WRKY domain (12) cytochrome B561 domain followed by NAC (13) two DFDF domain followed by cytochrome B and NAC (14) DNA_J2 domain followed by NAC (15) DNA_J2 domain followed by NAC and ZF_B domain (16) NAC domain followed by a TIR, two LRR and a CS domain (17) NAC followed by TIR domain (18) F-box domain followed by NAC (19) IQ domain followed by NAC (20) NAC domain followed by ZF_B domain (21) EF-hand domain followed by NAC (22) NAC domain followed by PPC domain (23) ENT domain followed by NAC (24) NAC domain followed by ABC_TM1F domain (25) NAC domain followed by CRM domain (26) NAC domain followed by RWP_RK and PB1 domain (27) NAC domain followed by three ACT domain (28) NAC domain followed by PABC domain (29) NAC domain followed by INTEGRA domain (30) RESPO domain followed by NAC (31) NAC domain followed by JMJN and JMJC domain (32) SAM domain followed by NAC (33) BRX domain followed by NAC and (34) repeat of NAC and ZF_domain. The identification of chimeric NAC domain sequences was determined using the ScanProsite and InterProScan server. The details regarding the presence of chimeric NAC TF in different taxa can be found in [S1 Table](#).

<https://doi.org/10.1371/journal.pone.0231425.g004>

NAC TFs encodes chimeric proteins and contain multiple binding sites

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

protein kinase, PI3_4_kinase_3, EF-hands (elongation factor), CRM, peptidase A1, WRKY, cytochrome B561, OFOF, FFO, Dna_J2, ZF_B, TIR, LRR, CS, F-box, IQ, PPC, ENT, ABC_TM1F, RWP_RK, PB1, PABC, ACT, INTEGRA, RESPO, JMJC, SAM, BRX, G_TR_2, RORP, CHCH, TPR, YJEF_N, HTH, HOMEO, GH16, ANK_REP_REGION, Peroxidase, LONGIN, V_SNA, RECA_2, KH_TY, APAG, RRM, carrier, and a DCO domain. At least four NAC TFs from *A. thaliana*, ten from *B. napus*, four from *B. rapa*, two from *M. domestica*, four from *P. virgatum*, 17 from *C. sativa*, eight from *D. oligosanthos*, eight from *E. tef*, and five from *L. perrieri* were found to possess 2 NAC domains (S1 Table). NAC TFs in several other species were also found to contain two NAC domains (S1 Table). When two NAC domains were present, both domains were located towards the N-terminal end. NAC TFs of at least three species, *O. rufipogon*, *B. stacei*, and *Camelina sativa* were found to possess three NAC domains whereas the NAC TFs in *A. lyrata* (gene id: 338342), *C. sativa* (Csa16g052260.1), and *E. tef* (462951506) were found to possess four NAC domains (Fig 4). Other chimeric domains were also identified in different regions of the NAC protein (Fig 4). The F-box and protein kinase domain was followed by a NAC domain and the NAC domain was followed by a G_TR_2 domain (Fig 5).

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 [83–91]. 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 [92]. 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 [93,94]. 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 (S2 Table). Some of the highly abundant functional motifs of NAC TFs were 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase signature, aldehyde dehydrogenase glutamic acid active site, lipocalin signature, phosphopantetheine attachment site, cysteine protease inhibitor, ATP synthase alpha and beta subunit signature, aminotransferase class II-pyridoxal-phosphate attachment site and others (S2 Table). This is the first study to report the presence of such a diverse number of functional sites and signature motifs in NAC TFs. Although majority of the functional domains are associated with a specific function in plants, the presence of a histocompatibility complex and a translationally controlled tumour protein (TCTP) sequence are of very interesting. These proteins are specifically found in animal systems and the histocompatibility complex is the major contributing factor regulating the binding of antigens. More specifically, TCTP is a highly conserved protein that is involved in microtubule stabilization, calcium binding, and apoptosis and is associated with the early growth phase of tumours [95]. The presence of MHC and TCTP in association with NAC domains suggests that this combination may be playing a crucial role in the plant immune system and in uncontrolled cell growth. The presence of diverse functional sites in NAC TFs indicates that NAC TFs are

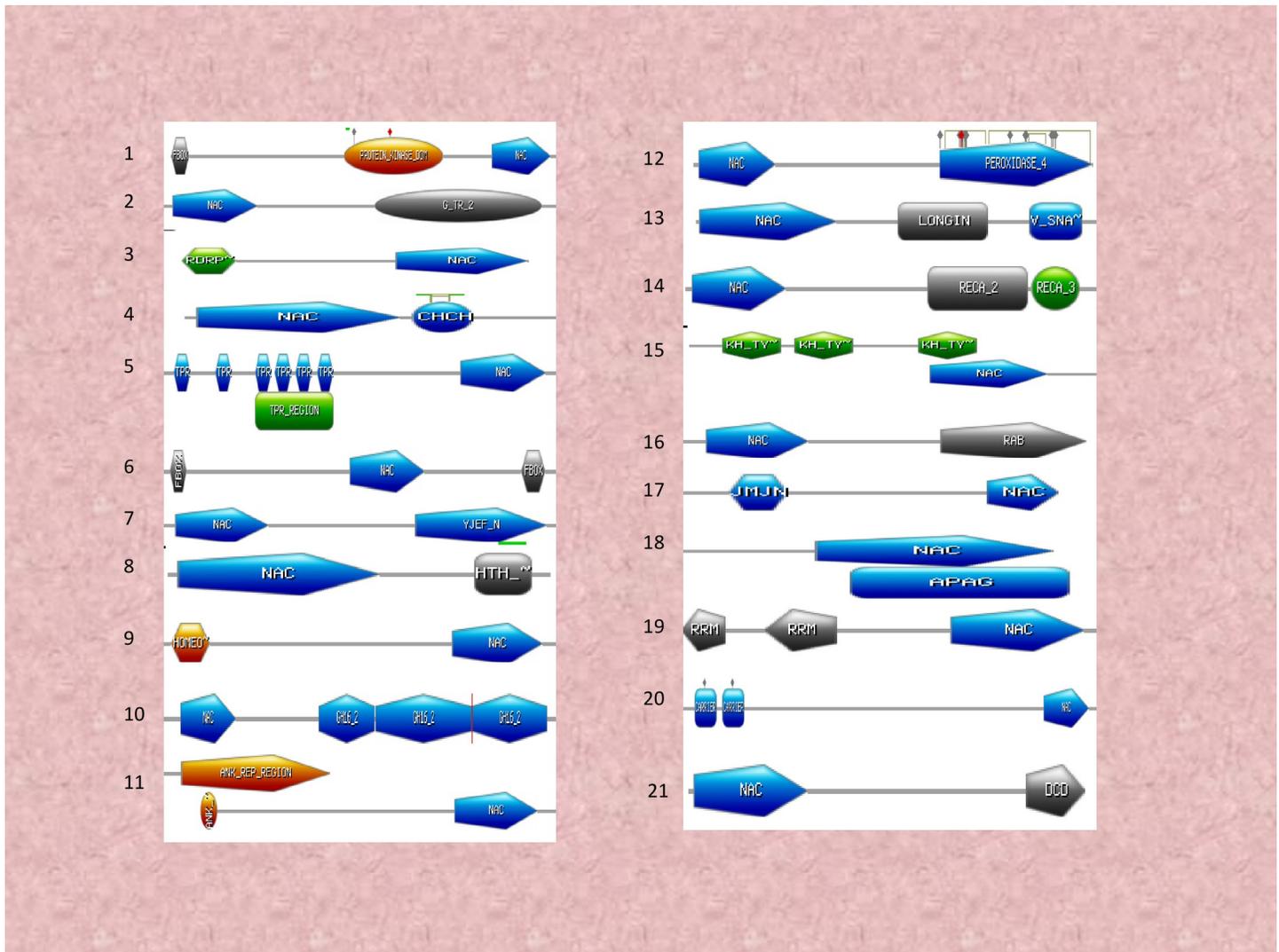


Fig 5. Chimeric NAC domains NAC TFs possess chimeric NAC domains with at least 21 diverse chimeric NAC domains identified in the studied species. (1) F-box domain followed by protein kinase and NAC domain (2) NAC domain followed by G_TR_2 domain (3) RDRP domain followed by NAC (4) NAC domain followed by CHCH domain (5) TPR repeats followed by NAC domain (6) F-box domain followed by NAC and F-box domain (7) NAC domain followed by YJEF_N domain (8) NAC domain followed by HTH domain (9) Homeobox domain followed by NAC domain (10) NAC domain followed by three GH6.2 domain (11) ANK repeat domain followed by NAC domain (12) NAC domain followed by peroxidase domain (13) NAC domain followed by LONGIN and V_SNA domain (14) NAC domain followed by RECA_2 and RECA_3 domain (15) KH_TY repeats followed by NAC domain (16) NAC domain followed by RAB domain (17) JMJN domain followed by NAC domain (18) NAC domain followed by APAG domain (19) two RRM domain followed by NAC domain (20) carrier domain followed by NAC domain and (21) NAC domain followed by DCO domain. The identification of chimeric NAC domain sequences was determined using the ScanProsite and InterProScan server. The details regarding the presence of chimeric NAC TF in different taxa can be found in [S1 Table](#).

<https://doi.org/10.1371/journal.pone.0231425.g005>

involved in diverse cellular functions and metabolic pathways. This statement is supported by the large number of NAC TFs that are present in plant genomes.

NAC TFs are involved in diverse cellular processes

NAC TFs are known to possess diverse chimeric domains, as a result, it is more than likely that NAC TFs are also involved in the regulation of diverse cellular pathways and cellular processes. To help substantiate this premise, the interactome associated with NAC TFs in *A. thaliana* was analysed. Results indicated that NAC TFs are potentially involved in a least 289 different

cellular processes and pathways (S3 Table). The majority are related to cell, tissue, and organ (root, stem, meristem) development, as well as signalling processes. Several NAC TFs also appear to be associated with phytohormone signalling, including auxin, gibberellin, jasmonic acid, and salicylic acid signalling pathways. NAC TFs were also found to be associated with pathways involved in the response to bacterial, fungal, UV, heat and other biotic and abiotic stresses (S3 Table). 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

Plant uses ammonia, nitrate, and urea as the source of nitrogen for its growth and development. Nitrogen is also associated with an increased rate of photosynthesis. Therefore, the role of ammonia source in the growth and development of the plants is very important. Nitrate is readily available as nitrogen source for plants and the uptake of nitrate is high in the acidic pH whereas the uptake of ammonia is high in the neutral pH. Studying the expression pattern of NAC TFs in nitrate and ammonia treated plant can explain how different nitrogen sources modulate the expression of NAC TFs and give the glimpse of their role in plants growing in the acidic and neutral pH soil. Urea is applied as an artificial nitrogen source for the plants when there is a lack of nitrate or ammonia in the soil. Therefore, patterns of NAC TF gene expression were analysed in leaf and root tissues of *A. thaliana* treated with ammonia, nitrate, or urea (Fig 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. The *AtNAC1* is involved in auxin signaling and modulates lateral root formation [74,96,97]. The higher expression of *AtNAC1* with response to treatment of nitrogenous compound reflects its role in plant development. *AtNAC4* is reported to be involved in nitrate transport and its higher expression in nitrate treated plant directly indicates its active role in nitrogen transport and assimilation [98].

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 (Fig 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*

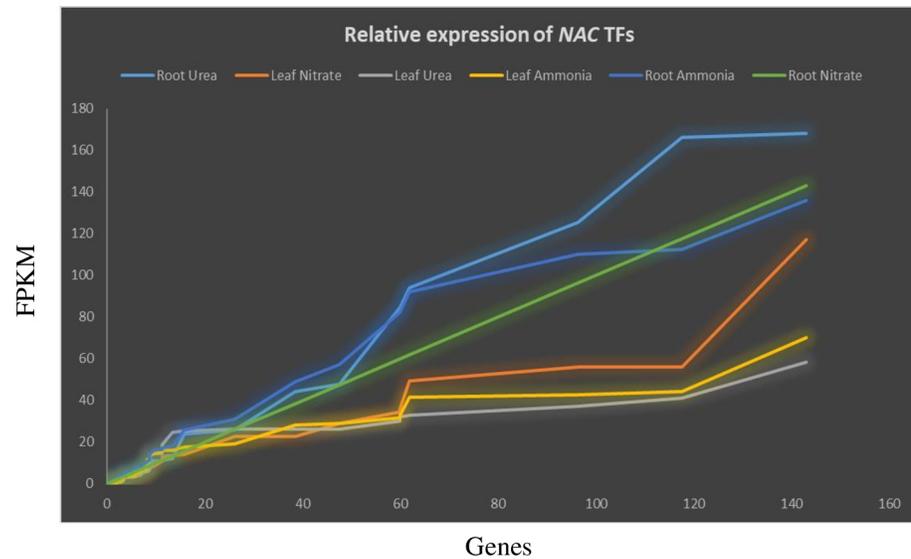


Fig 6. Differential expression of NAC TFs in leaves and roots of *A. thaliana* plants treated with ammonia, nitrate, and urea. The expression of *A. thaliana* NAC TFs was analysed to determine their response to different sources of nitrogen. Urea and ammonia in root tissue show higher expression level whereas urea treated leaf tissue showed low level of NAC expression. The expression data were obtained from the PhytoMine database in Phytozome and presented as FPKM (Fragments per Kilobase of transcripts per million mapped reads). The X-axis represents the NAC TF genes and Y-axis represent the Fragments per Kilobase of transcripts per million mapped reads.

<https://doi.org/10.1371/journal.pone.0231425.g006>

(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 [99]. Results of the current analysis indicate that *AtNAC8* and *AtNAC91* are the major NAC TFs involved in nitrogen assimilation during plant growth. The TaNAC8 was reported to be associated with strip rust and abiotic stress responses [100,101].

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 4). The most abundant codon was UCU (S), which was found 30 times in in *Humulus lupulus* NAC TFs (Table 4). The codons CGA (R), CGC (R), CGG (R), CGU (R) were absent in 127 of the 160 examined species. ACG (T), UCG (S), CAG (Q), CAC (H), CCA (P), CCC (P), CCG (P), and GCG (A) were absent in 126 of the examined species (S4 File). The highest relative synonymous codon usage bias (RSCU) was found to be 1.35, 1.23, 1.29 for the codon AAA (K) in *Ocimum tenuifolium*, *Picea sitchensis*, and *Ipomea trifida*. Synonymous codon-usage was not observed in NAC TFs. Relative codon usage is determined by dividing the ratio of observed frequency of codons by the expected frequency, provided that all of the synonymous codons for the same amino acids are used equally. Relative Synonymous Codon Usage (RSCU), however, is not related to the usage of amino acids. An RSCU > 1 indicates the occurrence of codons more frequently than expected, while an RSCU < 1 indicates that the codon occurs less frequently than expected [102,103]. Non-

Table 4. Codon usage of NAC TFs in plants.

Codons	Codon present in No. of species	Codon absent in No. of species	Average abundance of codons	Highest no. of codons	Name of the species with highest no. of codons
AAA (K)	126	20	4.77	9.9	<i>Glycine soja</i>
AAG (K)	146	0	10.75	24.2	<i>Sphagnum fallax</i>
AAC (N)	144	2	3.66	14.2	<i>Beta vulgaris</i>
AAU (N)	127	19	9.25	20.5	<i>Spinacia oleracea</i>
ACA (T)	139	7	2.33	15.2	<i>Citrus sinensis</i>
ACC (T)	137	9	2.4	17	<i>Amborella trichopoda</i>
ACG (T)	20	126	5.91	13	<i>Doroceras hygrometricum</i>
ACU (T)	146	0	7.42	16.6	<i>Sesamum indicum</i>
AGA (R)	146	0	10.92	24.3	<i>Klebsormidium flaccidum</i>
AGG (R)	146	0	4.12	18.8	<i>Amborella trichopoda</i>
CGA (R)	19	127	5.22	13.9	<i>Linum usitatissimum</i>
CGC (R)	19	127	2.47	6	<i>Linum usitatissimum</i>
CGG (R)	19	127	3.93	8.6	<i>Citrullus lanatus</i>
CGU (R)	19	127	2.06	4.7	<i>Linum usitatissimum</i>
AGC (S)	143	3	3.54	24.2	<i>Beta vulgaris</i>
AGU (S)	144	2	1.83	5.2	<i>Doroceras hygrometricum</i>
UCC (S)	141	5	4.51	12.3	<i>Aegilops tauschii</i>
UCG (S)	20	126	2.64	6.4	<i>Doroceras hygrometricum</i>
UCU (S)	146	0	4.65	30.5	<i>Humulus lupulus</i>
UCA (S)	139	7	5.09	15.1	<i>Morus notabilis</i>
AUA (I)	124	22	4.80	15.3	<i>Sphagnum fallax</i>
AUC (I)	146	0	5.10	16.7	<i>Sphagnum fallax</i>
AUU (I)	126	20	8.71	15.9	<i>Spinacia oleracea</i>
AUG (M)	146	0	7.81	22.8	<i>Sphagnum fallax</i>
CAA (Q)	146	0	5.31	15.4	<i>Fragaria vesca</i>
CAG (Q)	20	126	13.3	22.6	<i>Linum usitatissimum</i>
CAC (H)	20	126	6.64	10.9	<i>Beta vulgaris</i>
CAU (H)	144	2	4.45	9.7	<i>Setaria viridis</i>
CCA (P)	20	126	11.09	16.3	<i>Doroceras hygrometricum</i>
CCC (P)	20	126	14.18	19.2	<i>Amborella trichopoda</i>
CCG (P)	20	126	5.10	11.1	<i>Doroceras hygrometricum</i>
CCU (P)	146	0	8.00	24.7	<i>Klebsormidium flaccidum</i>
CUA (L)	143	3	5.83	28.3	<i>Sphagnum fallax</i>
CUC (L)	123	23	5.74	23.6	<i>Sphagnum fallax</i>
CUG (L)	142	4	5.87	43.9	<i>Sphagnum fallax</i>
CUU (L)	145	1	5.94	32.6	<i>Sphagnum fallax</i>
UUG (L)	125	21	5.94	24.4	<i>Sphagnum fallax</i>
UAA (L)	124	22	5.37	17.2	<i>Sphagnum fallax</i>
GAA (E)	146	0	4.62	27	<i>Klebsormidium flaccidum</i>
GAG (E)	145	1	5.54	18.1	<i>Sphagnum fallax</i>
GAC (D)	145	1	5.05	14.9	<i>Beta vulgaris</i>
GAU (D)	144	2	5.86	21.7	<i>Spinacia oleracea</i>
GCA (A)	135	11	5.49	18.5	<i>Citrus sinensis</i>
GCC (A)	130	16	5.05	15	<i>Amborella trichopoda</i>
GCG (A)	20	126	4.64	11.2	<i>Doroceras hygrometricum</i>
GCU (A)	146	0	4.65	31.1	<i>Setaria viridis</i>

(Continued)

Table 4. (Continued)

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
GGA (G)	146	0	4.63	27.5	<i>Setaria viridis</i>
GGC (G)	141	5	5.41	17.1	<i>Amborella trichopoda</i>
GGG (G)	145	1	2.7	6.7	<i>Elaeis guineensis</i>
GGU (G)	145	1	2.40	5.9	<i>Elaeis guineensis</i>
GUA (V)	140	6	1.46	3.5	<i>Sphagnum fallax</i>
GUC (V)	123	23	0.93	2	<i>Morus notabilis</i>
GUG (V)	142	4	4.35	11.8	<i>Beta vulgaris</i>
GUU (V)	143	3	5.38	16.6	<i>Klebsormidium flaccidum</i>
UAC (Y)	138	8	3.84	10.1	<i>Morus notabilis</i>
UAU (Y)	126	20	6.23	14.8	<i>Solanum melongena</i>
UGG (W)	147	0	3.89	14.5	<i>Vitis vinifera</i>
UGC (C)	143	3	5.14	15.6	<i>Oropetium thomaeum</i>
UGU (C)	145	1	3.9	9.6	<i>Zoysia matrella</i>
UUC (F)	146	0	4.60	25.4	<i>Picea glauca</i>
UUU (F)	126	20	10.67	19.2	<i>Sphagnum fallax</i>

<https://doi.org/10.1371/journal.pone.0231425.t004>

synonymous substitution in organisms is subject to natural selection [104,105]. Genes with lower non-synonymous selection leads to functional diversity of a gene. The presence of a low level of nonsynonymous codon usage in NAC TFs indicates that they are functional and have evolved from paralogous ancestors.

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

Nucleotide mutation is an integral part of the evolution of a genome and leads to the acquisition of required traits and the elimination of detrimental traits from the genome. It is a regular process and hundreds of thousands of nucleotides have undergone addition or deletion events in the evolution of a genome. The alteration or conversion of a nucleotide occurs either through a transition or a transversion. A transition event involves the interchange of two-ring purines (A and G) or of one-ring pyrimidines (C and T). Transversion events the exchange of a purine for a pyrimidine or vice versa. The rate at which these two events occur is important to understanding of the evolution of a gene. Therefore, the rate of nucleotide substitution in NAC TFs was analysed. Results indicated that the rate of transition in NAC TFs is higher than the rate of transversion. The substitution of adenine with guanine was found to be highest in *Linum usitatissimum* (15.82), while the substitution of guanine to adenine was found to be the highest in *Lotus japonicas* (19.07). The lowest rate of substitution from adenine to guanine and vice versa was found in *Trifolium pratense* (9.73) and *Amborella trichopoda* (10.8), respectively (S4 Table). The highest rate of substitution from thiamine to cytosine and vice versa was found in *Klebsormidium flaccidum* (7.19) and *Pseudotsuga menziesii* (11.59), respectively. The lowest rate of substitutions from thiamine to cytosine and vice versa was found in *Capsella grandiflora* (2.41) and *Cicer arietinum* (1.62), respectively (S4 Table). These data make it evident that the rates of transition of purine (adenine and guanine) nucleotides are higher than the rates of pyrimidines. The highest rate of transversion from adenine to thiamine and vice versa was found in *Capsella grandiflora* (12.34 for adenine to thiamine and 9.91 for thiamine to adenine) (S4 Table). The rate of substitution by transversion is slower relative to the rate of substitution by transition.

Capsella grandiflora is a close relative of *Arabidopsis thaliana* and is predicted to be the progenitor of *Capsella bursa-pastoris*. *Capsella grandiflora* is a self-pollinating plant and is used as a model organism in evolutionary studies and the change from self-incompatibility into self-compatibility. The genomic consequences of the evolution of selfing, however, is poorly understood. *Capsella rubella*, a close relative of *Capsella grandiflora*, that evolved self-compatibility 200,000 years ago [106] also exhibits a high rate of transversion from adenine to thiamine (11.19). Thus, the higher rate of transversion from adenine to thiamine in *Capsella grandiflora* and *Capsella rubella* may be a possible factor in the evolution of self-pollination. Higher rates of transversion were also found in *Solanum pimpinellifolium* (11.4) and *Castanea mollissima* (11.31) Chinese chestnut). *Solanum pimpinellifolium* is self-pollinating and exhibits high levels of stress tolerance [107]. *Castanea mollissima* has evolved over a period of time in coexistence with chestnut blight and is resistant to the pathogen. This indicates that higher rates of transversion from adenine to thiamine and vice versa are associated with self-pollination and stress tolerance in plants. The highest rate of substitution from guanine to cytosine and vice versa was found in *Arachis hypogaea* (11.07), and *Camelina sativa* (11.46), respectively (S4 Table). The lowest rate of substitution from adenine to thiamine and vice versa was found in *Linum usitatissimum* (3.72) and *Klebsormidium flaccidum* (6.67), respectively. Notably, the highest rate of substitution from thiamine to cytosine was found in *Klebsormidium flaccidum* and the highest rate of substitution from adenine to guanine was found in *Linum usitatissimum*. This indicates that organisms which exhibit the highest rate of transition possess the lowest rate of transversion.

NAC TFs evolved from orthologous ancestors

A phylogenetic tree of NAC TFs was constructed to understand their evolutionary relationships. A model selection was conducted before constructing the phylogenetic tree using the maximum likelihood statistical method. The phylogenetic tree revealed the presence of at least seven phylogenetic clustered orthologous groups (COGs) originating from a common, orthologous ancestor (Fig 7). Each phylogenetic cluster was further divided into two or more subgroups. A phylogenetic tree of each individual species was subsequently constructed to examine the duplication and loss events in NAC TFs. The phylogenetic tree of each species was independently reconciled with the collective species tree. This analysis indicated that NAC TFs in all of the species were duplicated and none of a NAC TFs was found to be lost. This suggests that NAC TFs evolved from common ancestors (orthology) and underwent numerous duplication events during the divergence and speciation (paralogy) events, which gave rise to diverse gene functions in plant development and growth. The NAC TFs of *K. flaccidum* might be the most possible common ancestors of some plant species and the NAC TFs of other algal species could have contributed towards the evolution of other NAC TFs in plants. If the duplication would have disrupted the normal functioning of the cell, the organism might have reduced its reproductive fitness and would have been died. However, the duplication of NAC TFs possesses beneficial character thus providing the fitness advantage. Gene duplication contribute to the evolution that provides new genetic content for mutation, selection, and drift to act and to create new evolutionary opportunities [108]. Genome duplication is a common event in plants and multiple event of genome duplication have occurred during the diversification of angiosperms [109]. Genome duplication sometimes followed by the increased rate of evolution of some important genes [109]. The duplicated genes is responsible for the functional divergence and may play role in escaping the extinction [109,110]. In addition, duplication can lead to decreased probabilities of extinction, increase genetic variation, mutational robustness, and tolerance to changing environmental conditions [109]. The genetic variation

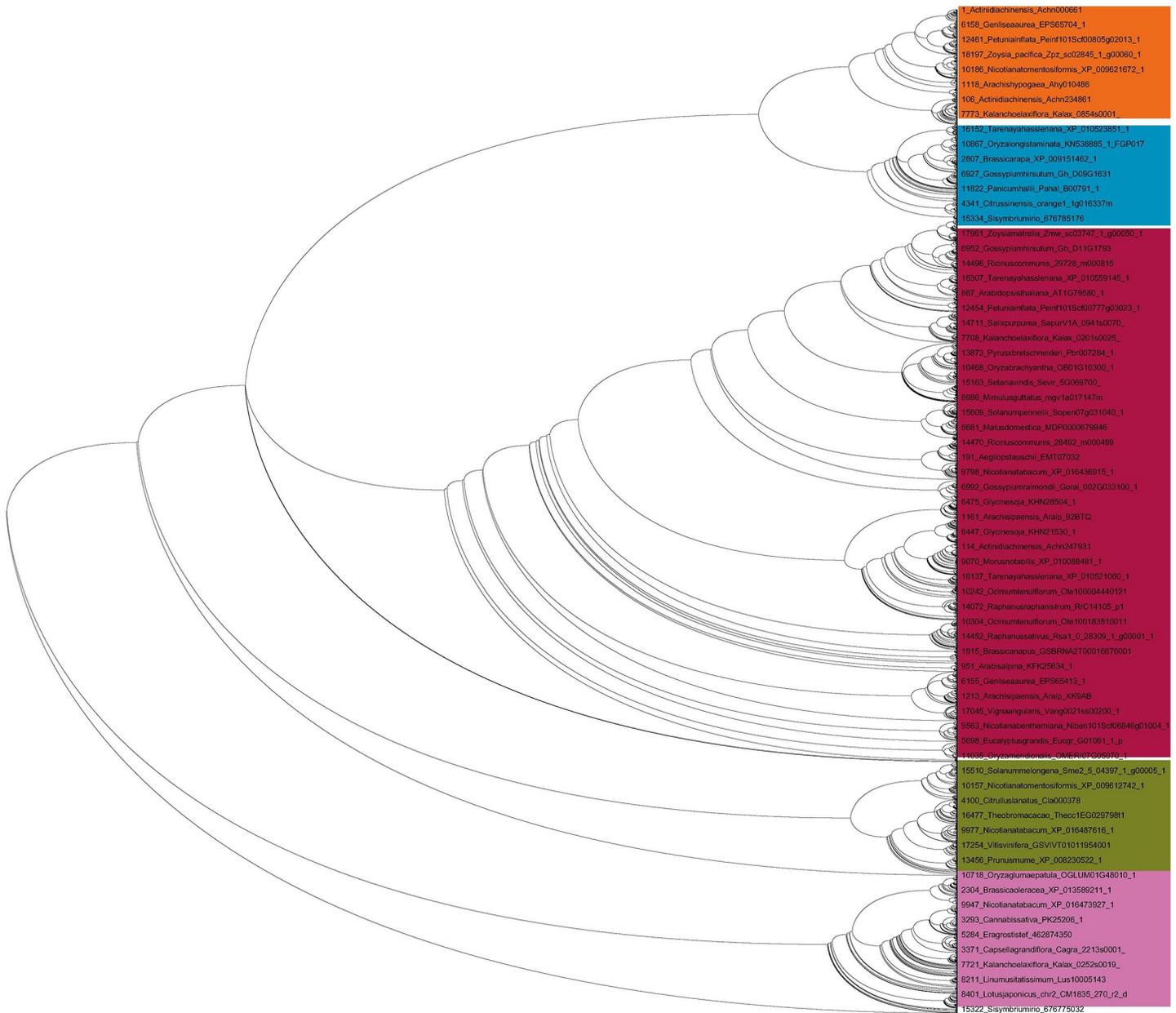


Fig 7. Phylogenetic tree of NAC TFs. A phylogenetic tree of NAC TF reveals the presence of seven clustered orthologous groups (COGs). Each group also possesses two or more sub-groups. The phylogenetic tree shows lineage (monocot/dicot) specific grouping of NAC TFs. The phylogenetic tree was constructed using the neighbour-joining method with 1000 bootstrap replicates.

<https://doi.org/10.1371/journal.pone.0231425.g007>

incurred by duplication contribute to selection pressure and provide the opportunities for survival diverse environmental stress. Being, NAC TFs are highly duplicated, they might be providing such genetic variability in the plant kingdom to evade diverse environmental responses.

We also checked for the presence of potential foreign or homologous sequences (xenologs) in NAC TFs. No primary xenologs, sibling donor xenologs, sibling recipient xenologs, incompatible xenologs, autoxenologs, or paraxenologs were identified in NAC TFs. Although the phylogenetic tree indicates the evolution NAC TFs from common ancestors, none of the NAC genes in the examined species were found to have been transferred from one species to

another. Previous studies of NAC TFs in six plant species also reported a high level of duplication and divergent evolution [111]. The expansion of TF families was associated with an increase in the structural complexity of the organism [112]. Previous studies reported the lineage-specific grouping of transcription factors [93,111]. The phylogenetic tree of NAC TFs also revealed the presence of lineage-specific clustering as well. In a few cases, however, order-specific clustering of NAC TFs was also observed. For example, NAC TFs in dicot species of the Brassica lineage, including *A. thaliana*, *A. halleri*, *B. napus*, *B. rapa*, *R. sativus*, *R. raphanistrum*, *C. rubella*, *A. alpine*, and others, grouped together. Similarly, NAC TFs in monocot plant species, including *O. sativa*, *O. nivara*, *B. distachyon*, and others, also grouped together.

Conclusion

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

Supporting information

S1 Table. Supplementary table showing different chimeric domains of NAC TFs.
(DOCX)

S2 Table. NAC TFs showing the presence of novel functional domain along with NAC domains.
(PDF)

S3 Table. NAC TFs showing their involvement in different pathways and biological process.
(PDF)

S4 Table. Substitution rate of NAC TFs of plants.
(DOCX)

S1 File. Accession number of transmembrane domains containing NAC TF proteins.
(XLSX)

S2 File. Nuclear localization signal sequences of NAC TFs. Sheet 1 of the file show all the raw N-terminal NLS consensus sequences, unique NLS with linker amino acids, and unique NLS post removal of linker amino acids. Sheet 2 represents the number of occurrences of N-terminal NLS and sheet 3 represents C-terminal NLS, number of occurrences, C-terminal unique NLS, and N- and C-terminal unique NLS.
(XLSX)

S3 File. Accession number and species details of NAC TF proteins containing multi-functional binding sites.
(XLSX)

S4 File. Details of codon usage of NAC TFs in plants.
(XLSX)

S1 Fig. Transmembrane bound NAC TF proteins.
(XZ)

S2 Fig. Graphical presentation of nuclear localization signal sequences of NAC TF proteins.
(XZ)

S3 Fig. The presence of L-V-F-Y/H conserved motif in NAC TFs of plants (*A. thaliana*).
(RAR)

Acknowledgments

The author would like to extend their sincere thanks to the Natural and Medical Sciences Research Center, University of Nizwa to facilitate the study.

Author Contributions

Conceptualization: Tapan Kumar Mohanta.

Data curation: Tapan Kumar Mohanta.

Formal analysis: Tapan Kumar Mohanta.

Investigation: Tapan Kumar Mohanta, Adil Khan.

Validation: Tapan Kumar Mohanta.

Writing – original draft: Tapan Kumar Mohanta.

Writing – review & editing: Tapan Kumar Mohanta, Dhananjay Yadav, Abeer Hashem, Baby Tabassum, Abdul Latif Khan, Elsayed Fathi Abd_Allah, Ahmed Al-Harrasi.

References

1. Yanagisawa S. Transcription factors in plants: Physiological functions and regulation of expression. *J Plant Res* [Internet]. 1998; 111:363–71. Available from: <https://doi.org/10.1007/BF02507800>
2. Nuruzzaman M, Sharoni AM, Kikuchi S. Roles of NAC transcription factors in the regulation of biotic and abiotic stress responses in plants. *Front Microbiol* [Internet]. 2013; 4:248. Available from: <https://www.frontiersin.org/article/10.3389/fmicb.2013.00248> PMID: 24058359
3. Guan X, Stege J, Kim M, Dahmani Z, Fan N, Heifetz P, et al. Heritable endogenous gene regulation in plants with designed polydactyl zinc finger transcription factors. *Proc Natl Acad Sci* [Internet]. National Academy of Sciences; 2002; 99:13296–301. Available from: <https://www.pnas.org/content/99/20/13296> <https://doi.org/10.1073/pnas.192412899> PMID: 12271125
4. Ortigosa A, Fonseca S, Franco-Zorrilla JM, Fernández-Calvo P, Zander M, Lewsey MG, et al. The JA-pathway MYC transcription factors regulate photomorphogenic responses by targeting HY5 gene expression. *Plant J* [Internet]. John Wiley & Sons, Ltd; 2019;n/a. Available from: <https://doi.org/10.1111/tbj.14618>
5. Franco-Zorrilla JM, López-Vidriero I, Carrasco JL, Godoy M, Vera P, Solano R. DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proc Natl Acad Sci* [Internet]. National Academy of Sciences; 2014; 111:2367–72. Available from: <https://www.pnas.org/content/111/6/2367> <https://doi.org/10.1073/pnas.1316278111> PMID: 24477691
6. Todeschini A-L, Georges A, Veitia RA. Transcription factors: specific DNA binding and specific gene regulation. *Trends Genet* [Internet]. Elsevier; 2014; 30:211–9. Available from: <https://doi.org/10.1016/j.tig.2014.04.002> PMID: 24774859
7. Geertz M, Maerkl SJ. Experimental strategies for studying transcription factor–DNA binding specificities. *Brief Funct Genomics* [Internet]. 2010; 9:362–73. Available from: <https://doi.org/10.1093/bfpg/elq023> PMID: 20864494

8. Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA, et al. Diversity and Complexity in DNA Recognition by Transcription Factors. *Science* (80-) [Internet]. American Association for the Advancement of Science; 2009; 324:1720–3. Available from: <https://science.sciencemag.org/content/324/5935/1720>
9. Burley SK. DNA-binding motifs from eukaryotic transcription factors. *Curr Opin Struct Biol* [Internet]. 1994; 4:3–11. Available from: <http://www.sciencedirect.com/science/article/pii/S0959440X94900531>
10. Zamanighomi M, Lin Z, Wang Y, Jiang R, Wong WH. Predicting transcription factor binding motifs from DNA-binding domains, chromatin accessibility and gene expression data. *Nucleic Acids Res* [Internet]. Oxford University Press; 2017; 45:5666–77. Available from: <https://doi.org/10.1093/nar/gkx358> PMID: 28472398
11. Morett E, Cannon W, Buck M. The DNA-binding domain of the transcriptional activator protein NifA resides in its carboxy terminus, recognises the upstream activator sequences of nif promoters and can be separated from the positive control function of NifA. *Nucleic Acids Res* [Internet]. 1988; 16:11469–88. Available from: <https://pubmed.ncbi.nlm.nih.gov/3062575> <https://doi.org/10.1093/nar/16.24.11469> PMID: 3062575
12. Mizutani A, Tanaka M. Regions of GAL4 critical for binding to a promoter in vivo revealed by a visual DNA-binding analysis. *EMBO J* [Internet]. Oxford University Press; 2003; 22:2178–87. Available from: <https://pubmed.ncbi.nlm.nih.gov/12727884> <https://doi.org/10.1093/emboj/cdg220> PMID: 12727884
13. Plaschka C, Hantsche M, Dienemann C, Burzinski C, Plitzko J, Cramer P. Transcription initiation complex structures elucidate DNA opening. *Nature* [Internet]. 2016; 533:353–8. Available from: <https://doi.org/10.1038/nature17990> PMID: 27193681
14. Nikolov DB, Burley SK. RNA polymerase II transcription initiation: A structural view. *Proc Natl Acad Sci* [Internet]. National Academy of Sciences; 1997; 94:15–22. Available from: <https://www.pnas.org/content/94/1/15> <https://doi.org/10.1073/pnas.94.1.15> PMID: 8990153
15. Pufall MA, Kaplan CD. Mechanisms of eukaryotic transcription. *Genome Biol* [Internet]. 2013; 14:311. Available from: <https://doi.org/10.1186/gb-2013-14-9-311> <https://doi.org/10.1186/gb-2013-14-9-311> PMID: 24079829
16. Pugh BF, Tjian R. Transcription from a TATA-less promoter requires a multisubunit TFIID complex. *Genes Dev*. 1991; 5:1935–45. <https://doi.org/10.1101/gad.5.11.1935> PMID: 1657708
17. Dao LTM, Spicuglia S. Transcriptional regulation by promoters with enhancer function. *Transcription* [Internet]. 2018/06/25. Taylor & Francis; 2018; 9:307–14. Available from: <https://pubmed.ncbi.nlm.nih.gov/29889606> <https://doi.org/10.1080/21541264.2018.1486150> PMID: 29889606
18. Zabidi MA, Stark A. Regulatory Enhancer–Core–Promoter Communication via Transcription Factors and Cofactors. *Trends Genet* [Internet]. 2016; 32:801–14. Available from: <http://www.sciencedirect.com/science/article/pii/S0168952516301214> <https://doi.org/10.1016/j.tig.2016.10.003> PMID: 27816209
19. Rahnamoun H, Orozco P, Lauberth SM. The role of enhancer RNAs in epigenetic regulation of gene expression. *Transcription* [Internet]. Taylor & Francis; 2020; 11:19–25. Available from: <https://doi.org/10.1080/21541264.2019.1698934> <https://doi.org/10.1080/21541264.2019.1698934> PMID: 31823686
20. Kaufmann K, Airoldi CA. Master Regulatory Transcription Factors in Plant Development: A Blooming Perspective. In: Yamaguchi N, editor. *Plant Transcr Factors Methods Protoc* [Internet]. New York, NY: Springer New York; 2018. p. 3–22. Available from: https://doi.org/10.1007/978-1-4939-8657-6_1
21. Singh KB. Transcriptional Regulation in Plants: The Importance of Combinatorial Control. *Plant Physiol* [Internet]. American Society of Plant Biologists; 1998; 118:1111–20. Available from: <http://www.plantphysiol.org/content/118/4/1111> <https://doi.org/10.1104/pp.118.4.1111> PMID: 9847085
22. Roy S. Function of MYB domain transcription factors in abiotic stress and epigenetic control of stress response in plant genome. *Plant Signal Behav* [Internet]. Taylor & Francis; 2016; 11:e1117723. Available from: <https://doi.org/10.1080/15592324.2015.1117723> PMID: 26636625
23. Leng P, Zhao J. Transcription factors as molecular switches to regulate drought adaptation in maize. *Theor Appl Genet* [Internet]. 2019; Available from: <https://doi.org/10.1007/s00122-019-03494-y>
24. Pajerowska-Mukhtar KM, Wang W, Tada Y, Oka N, Tucker CL, Fonseca JP, et al. The HSF-like transcription factor TBF1 is a major molecular switch for plant growth-to-defense transition. *Curr Biol* [Internet]. 2012/01/12. 2012; 22:103–12. Available from: <https://pubmed.ncbi.nlm.nih.gov/22244999> <https://doi.org/10.1016/j.cub.2011.12.015> PMID: 22244999
25. Puranik S, Sahu PP, Srivastava PS, Prasad M. NAC proteins: Regulation and role in stress tolerance. *Trends Plant Sci*. Elsevier Ltd; 2012; 17:369–81. <https://doi.org/10.1016/j.tplants.2012.02.004> PMID: 22445067
26. Ernst HA, 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*. 2004; 5:297–303. <https://doi.org/10.1038/sj.embor.7400093> PMID: 15083810

27. Jensen M, Kjaersgaard T, Nielsen M, Galberg P, Petersen K, O' Shea C, et al. The Arabidopsis thaliana NAC transcription factor family: structure-function relationships and determinants of ANAC019 stress signalling. *Biochem J*. 2009; 426:183–96.
28. Chen Q, Wang Q, Xiong L, Lou Z. A structural view of the conserved domain of rice stress-responsive NAC1. *Protein Cell*. Beijing: Higher Education Press; 2011; 2:55–63. <https://doi.org/10.1007/s13238-011-1010-9> PMID: 21337010
29. Liu G, Li X, Jin S, Liu X, Zhu L, Nie Y, et al. Overexpression of rice NAC gene SNAC1 improves drought and salt tolerance by enhancing root development and reducing transpiration rate in transgenic cotton. *PLoS One* [Internet]. Public Library of Science; 2014; 9:e86895–e86895. Available from: <https://pubmed.ncbi.nlm.nih.gov/24489802> <https://doi.org/10.1371/journal.pone.0086895> PMID: 24489802
30. An X, Liao Y, Zhang J, Dai L, Zhang N, Wang B, et al. Overexpression of rice NAC gene SNAC1 in ramie improves drought and salt tolerance. *Plant Growth Regul* [Internet]. 2015; 76:211–23. Available from: <https://doi.org/10.1007/s10725-014-9991-z>
31. Duval M, Hsieh T-F, Kim SY, Thomas TL. Molecular characterization of AtNAM: a member of the Arabidopsis NAC domain superfamily. *Plant Mol Biol*. 2002; 50:237–48. <https://doi.org/10.1023/a:1016028530943> PMID: 12175016
32. Olsen AN, Ernst HA, Leggio L Lo, Skriver K. NAC transcription factors: structurally distinct, functionally diverse. *Trends Plant Sci*. Elsevier; 2005; 10:79–87. <https://doi.org/10.1016/j.tplants.2004.12.010> PMID: 15708345
33. Le DT, Nishiyama R, Watanabe Y, Mochida K, Yamaguchi-Shinozaki K, Shinozaki K, 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*. Oxford University Press; 2011; 18:263–76.
34. Olsen AN, Ernst HA, Leggio L Lo, Skriver K. DNA-binding specificity and molecular functions of NAC transcription factors. *Plant Sci*. 2005; 169:785–97.
35. Greve K, La Cour T, Jensen MK, Poulsen FM, 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*. 2003; 371:97–108. <https://doi.org/10.1042/BJ20021123> PMID: 12646039
36. Hao Y-J, Song Q-X, Chen H-W, Zou H-F, Wei W, Kang X-S, et al. Plant NAC-type transcription factor proteins contain a NARD domain for repression of transcriptional activation. *Planta*. 2010; 232:1033–43. <https://doi.org/10.1007/s00425-010-1238-2> PMID: 20683728
37. 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*. 2008; 280:547–63. <https://doi.org/10.1007/s00438-008-0386-6> PMID: 18813954
38. Yamaguchi M, Ohtani M, Mitsuda N, Kubo M, Ohme-Takagi M, Fukuda H, et al. VND-INTERACTING2, a NAC Domain Transcription Factor, Negatively Regulates Xylem Vessel Formation in Arabidopsis. *Plant Cell*. American Society of Plant Biologists; 2010; 22:1249–63. <https://doi.org/10.1105/tpc.108.064048> PMID: 20388856
39. Ho SK, Byung OP, Jae HY, Mi SJ, Sang ML, Hay JH, et al. Identification of a calmodulin-binding NAC protein as a transcriptional repressor in Arabidopsis. *J Biol Chem*. 2007; 282:36292–302. <https://doi.org/10.1074/jbc.M705217200> PMID: 17947243
40. Christian D, Kemal K, WI W., Van Der SD, John M, DE S., et al. The transcription factor ATAF2 represses the expression of pathogenesis-related genes in Arabidopsis. *Plant J*. Wiley/Blackwell (10.1111); 2005; 43:745–57. <https://doi.org/10.1111/j.1365-313X.2005.02488.x> PMID: 16115070
41. Shen H, Yin Y, Chen F, Xu Y, Dixon RA. A Bioinformatic Analysis of NAC Genes for Plant Cell Wall Development in Relation to Lignocellulosic Bioenergy Production. *BioEnergy Res*. 2009; 2:217.
42. Zhu T, Nevo E, Sun D, Peng J. PHYLOGENETIC ANALYSES UNRAVEL THE EVOLUTIONARY HISTORY OF NAC PROTEINS IN PLANTS. *Evolution (N Y)* [Internet]. John Wiley & Sons, Ltd (10.1111); 2012; 66:1833–48. Available from: <https://doi.org/10.1111/j.1558-5646.2011.01553.x>
43. Pereira-Santana A, Alcaraz LD, Castaño E, Sanchez-Calderon L, Sanchez-Teyer F, Rodriguez-Zapata L. Comparative Genomics of NAC Transcriptional Factors in Angiosperms: Implications for the Adaptation and Diversification of Flowering Plants. *PLoS One* [Internet].; 2015; 10:e0141866. Available from: <https://doi.org/10.1371/journal.pone.0141866> PMID: 26569117
44. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*. 2012; 40:D1178–86. <https://doi.org/10.1093/nar/gkr944> PMID: 22110026
45. Dong Q, Schlueter SD, Brendel V. PlantGDB, plant genome database and analysis tools. *Nucleic Acids Res* [Internet]. Oxford University Press; 2004; 32:D354–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/14681433> <https://doi.org/10.1093/nar/gkh046> PMID: 14681433

46. Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL. Domain enhanced lookup time accelerated BLAST. *Biol Direct. BioMed Central*; 2012; 7:12. <https://doi.org/10.1186/1745-6150-7-12> PMID: 22510480
47. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics. Oxford University Press*; 2014; 30:1236–40. <https://doi.org/10.1093/bioinformatics/btu031> PMID: 24451626
48. de Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, et al. Scan-Prosite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* 2006; 34:W362–5. <https://doi.org/10.1093/nar/gkl124> PMID: 16845026
49. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J Mol Biol.* 2001; 305:567–80. <https://doi.org/10.1006/jmbi.2000.4315> PMID: 11152613
50. Nguyen Ba A, Pogoutse A, Provart N, Moses A. NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. *BMC Bioinformatics. BioMed Central*; 2009; 10:202. <https://doi.org/10.1186/1471-2105-10-202> PMID: 19563654
51. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res. Oxford University Press*; 2017; 45:D362–8. <https://doi.org/10.1093/nar/gkw937> PMID: 27924014
52. von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res. Oxford, UK: Oxford University Press*; 2003; 31:258–61. <https://doi.org/10.1093/nar/gkg034> PMID: 12519996
53. Tamura K, Filipowski A, Peterson D, Stecher G, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol Biol Evol [Internet]*. 2013; 30:2725–9. Available from: <https://doi.org/10.1093/molbev/mst197> PMID: 24132122
54. Chen K, Durand D, Farach-Colton M. NOTUNG: A Program for Dating Gene Duplications and Optimizing Gene Family Trees. *J Comput Biol [Internet]*. Mary Ann Liebert, Inc., publishers; 2000; 7:429–47. Available from: <https://doi.org/10.1089/106652700750050871> PMID: 11108472
55. Mohanta T, Mohanta N, Mohanta Y, Bae H. Genome-Wide Identification of Calcium Dependent Protein Kinase Gene Family in Plant Lineage Shows Presence of Novel D-x-D and D-E-L Motifs in EF-Hand Domain. *Front Plant Sci. Frontiers Media S.A.*; 2015; 6:1146. <https://doi.org/10.3389/fpls.2015.01146> PMID: 26734045
56. Welner DH, Lindemose S, Grossmann JG, Møllegaard NE, Olsen AN, Helgstrand C, et al. DNA binding by the plant-specific NAC transcription factors in crystal and solution: a firm link to WRKY and GCM transcription factors. *Biochem J.* 2012; 444:395–404. <https://doi.org/10.1042/BJ20111742> PMID: 22455904
57. Mohanta TK, Khan AL, Hashem A, Abd_Allah EF, Al-Harrasi A. The Molecular Mass and Isoelectric Point of Plant Proteomes. *BMC Genomics [Internet]*. 2019; 20:631. Available from: <http://biorxiv.org/content/early/2019/02/10/546077.abstract> <https://doi.org/10.1186/s12864-019-5983-8> PMID: 31382875
58. Shaw KL, Grimsley GR, Yakovlev GI, Makarov AA, Pace CN. The effect of net charge on the solubility, activity, and stability of ribonuclease Sa. *Protein Sci. Cold Spring Harbor Laboratory Press*; 2001; 10:1206–15. <https://doi.org/10.1110/ps.440101> PMID: 11369859
59. Pergande MR, Cologna SM. Isoelectric Point Separations of Peptides and Proteins. *Coorssen JR, Yergey AL, Wisniewski JR, editors. Proteomes. MDPI*; 2017; 5:4.
60. Cunningham J, Estrella V, Lloyd M, Gillies R, Frieden BR, Gatenby R. Intracellular Electric Field and pH Optimize Protein Localization and Movement. *Csermely P, editor. PLoS One. San Francisco, USA: Public Library of Science*; 2012; 7:e36894. <https://doi.org/10.1371/journal.pone.0036894> PMID: 22623963
61. Hoppe T, Rape M, Jentsch S. Membrane-bound transcription factors: regulated release by RIP or RUP. *Curr Opin Cell Biol.* 2001; 13:344–8. [https://doi.org/10.1016/s0955-0674\(00\)00218-0](https://doi.org/10.1016/s0955-0674(00)00218-0) PMID: 11343906
62. Vik Å, Rine J. Membrane biology: Membrane-regulated transcription. *Curr Biol. Elsevier*; 2000; 10: R869–71.
63. Hoppe T, Matuschewski K, Rape M, Schlenker S, Ulrich HD, Jentsch S. Activation of a Membrane-Bound Transcription Factor by Regulated Ubiquitin/Proteasome-Dependent Processing. *Cell. Elsevier*; 2000; 102:577–86.
64. Seo PJ, Kim SG, Park CM. Membrane-bound transcription factors in plants. *Trends Plant Sci.* 2008; 13:550–6. <https://doi.org/10.1016/j.tplants.2008.06.008> PMID: 18722803

65. Shen J, Chen X, Hendershot L, Prywes R. ER Stress Regulation of ATF6 Localization by Dissociation of BiP/GRP78 Binding and Unmasking of Golgi Localization Signals. *Dev Cell*. Elsevier; 2002; 3:99–111. [https://doi.org/10.1016/s1534-5807\(02\)00203-4](https://doi.org/10.1016/s1534-5807(02)00203-4) PMID: 12110171
66. Liu J-X, Srivastava R, Che P, Howell SH. An Endoplasmic Reticulum Stress Response in Arabidopsis Is Mediated by Proteolytic Processing and Nuclear Relocation of a Membrane-Associated Transcription Factor, bZIP28. *Plant Cell*. American Society of Plant Biologists; 2007; 19:4111–9. <https://doi.org/10.1105/tpc.106.050021> PMID: 18156219
67. Liu J-X, Srivastava R, Che P, Howell SH. Salt stress responses in Arabidopsis utilize a signal transduction pathway related to endoplasmic reticulum stress signaling. *Plant J*. Blackwell Publishing Ltd; 2007; 51:897–909. <https://doi.org/10.1111/j.1365-313X.2007.03195.x> PMID: 17662035
68. Iwata Y, Koizumi N. An Arabidopsis transcription factor, AtbZIP60, regulates the endoplasmic reticulum stress response in a manner unique to plants. *Proc Natl Acad Sci U S A*. National Academy of Sciences; 2005; 102:5280–5. <https://doi.org/10.1073/pnas.0408941102> PMID: 15781873
69. Kim Y-S, Kim S-G, Park J-E, Park H-Y, Lim M-H, Chua N-H, et al. A Membrane-Bound NAC Transcription Factor Regulates Cell Division in Arabidopsis. *Plant Cell*. American Society of Plant Biologists; 2006; 18:3132–44. <https://doi.org/10.1105/tpc.106.043018> PMID: 17098812
70. Kim S-Y, Kim S-G, Kim Y-S, Seo P-J, Bae M, Yoon H-K, et al. Exploring membrane-associated NAC transcription factors in Arabidopsis: implications for membrane biology in genome regulation. *Nucleic Acids Res*. Oxford University Press; 2007; 35:203–13. <https://doi.org/10.1093/nar/gkl1068> PMID: 17158162
71. Kosugi S, Hasebe M, Matsumura N, Takashima H, Miyamoto-Sato E, Tomita M, et al. Six classes of nuclear localization signals specific to different binding grooves of importin α . *J Biol Chem*. 2009; 284:478–85. <https://doi.org/10.1074/jbc.M807017200> PMID: 19001369
72. Mattaj JW, Englmeier L. Nucleocytoplasmic Transport: The Soluble Phase. *Annu Rev Biochem*. Annual Reviews; 1998; 67:265–306. <https://doi.org/10.1146/annurev.biochem.67.1.265> PMID: 9759490
73. Kosugi S, Hasebe M, Tomita M, Yanagawa H. Nuclear export signal consensus sequences defined using a localization-based yeast selection system. *Traffic*. 2008; 9:2053–62. <https://doi.org/10.1111/j.1600-0854.2008.00825.x> PMID: 18817528
74. Xie Q, Frugis G, Colgan D, Chua N-H. Arabidopsis NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. *Genes Dev*. Cold Spring Harbor Laboratory Press; 2000; 14:3024–36. <https://doi.org/10.1101/gad.852200> PMID: 11114891
75. He X, Mu R, Cao W, Zhang Z, Zhang J, Chen S. AtNAC2, a transcription factor downstream of ethylene and auxin signaling pathways, is involved in salt stress response and lateral root development. *Plant J*. Wiley/Blackwell (10.1111); 2005; 44:903–16. <https://doi.org/10.1111/j.1365-313X.2005.02575.x> PMID: 16359384
76. Sperotto RA, Ricachenevsky FK, Duarte GL, Boff T, Lopes KL, Sperb ER, et al. Identification of up-regulated genes in flag leaves during rice grain filling and characterization of OsNAC5, a new ABA-dependent transcription factor. *Planta*. 2009; 230:985–1002. <https://doi.org/10.1007/s00425-009-1000-9> PMID: 19697058
77. Bonetta L. Interactome under construction. *Nature*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2010; 468:851. <https://doi.org/10.1038/468851a> PMID: 21150998
78. Tran L-SP, Nakashima K, Sakuma Y, Simpson SD, Fujita Y, Maruyama K, et al. Isolation and Functional Analysis of Arabidopsis Stress-Inducible NAC Transcription Factors That Bind to a Drought-Responsive γ -Element in the γ -Element Promoter. *Plant Cell*. 2004; 16:2481 LP– 2498.
79. Jensen MK, Lindemose S, Masi F de, Reimer JJ, Nielsen M, Perera V, et al. ATAF1 transcription factor directly regulates abscisic acid biosynthetic gene NCED3 in Arabidopsis thaliana. *FEBS Open Bio*. Elsevier; 2013; 3:321–7. <https://doi.org/10.1016/j.fob.2013.07.006> PMID: 23951554
80. Kazuo N, Tomohiro K, Kazuko Y-S, Kazuo S. A nuclear gene, *erd1*, encoding a chloroplast-targeted Clp protease regulatory subunit homolog is not only induced by water stress but also developmentally up-regulated during senescence in Arabidopsis thaliana. *Plant J*. Wiley/Blackwell (10.1111); 2003; 12:851–61.
81. Reeves WM, Lynch TJ, Mobin R, Finkelstein RR. Direct targets of the transcription factors ABA-Insensitive (ABI)4 and ABI5 reveal synergistic action by ABI4 and several bZIP ABA response factors. *Plant Mol Biol*. Dordrecht: Springer Netherlands; 2011; 75:347–63. <https://doi.org/10.1007/s11103-011-9733-9> PMID: 21243515

82. Jensen MK, Skriver K. NAC transcription factor gene regulatory and protein-protein interaction networks in plant stress responses and senescence. *IUBMB Life*. 2014; 66:156–66. <https://doi.org/10.1002/iub.1256> PMID: 24659537
83. Jingjing J, Shenghui M, Nenghui Y, Ming J, Jiashu C, Jianhua Z. WRKY transcription factors in plant responses to stresses. *J Integr Plant Biol*. Wiley/Blackwell (10.1111); 2016; 59:86–101.
84. Pandey SP, Somssich IE. The Role of WRKY Transcription Factors in Plant Immunity. *Plant Physiol*. American Society of Plant Biologists; 2009; 150:1648–55. <https://doi.org/10.1104/pp.109.138990> PMID: 19420325
85. Saha D, Prasad AM, Srinivasan R. Pentatricopeptide repeat proteins and their emerging roles in plants. *Plant Physiol Biochem*. 2007; 45:521–34. <https://doi.org/10.1016/j.plaphy.2007.03.026> PMID: 17560114
86. Barkan A, Small I. Pentatricopeptide Repeat Proteins in Plants. *Annu Rev Plant Biol*. Annual Reviews; 2014; 65:415–42. <https://doi.org/10.1146/annurev-arplant-050213-040159> PMID: 24471833
87. Nandety RS, Caplan JL, Cavanaugh K, Perroud B, Wroblewski T, Michelmore RW, et al. The role of TIR-NBS and TIR-X proteins in plant basal defense responses. *Plant Physiol*. 2013; 162:1459–72. <https://doi.org/10.1104/pp.113.219162> PMID: 23735504
88. Wan H, Yuan W, Ye Q, Wang R, Ruan M, Li Z, et al. Analysis of TIR- and non-TIR-NBS-LRR disease resistance gene analogous in pepper: characterization, genetic variation, functional divergence and expression patterns. *BMC Genomics*. 2012; 13:502. <https://doi.org/10.1186/1471-2164-13-502> PMID: 22998579
89. Mohanta TK, Occhipinti A, Atsbaha Zebelo S, Foti M, Fliegmann J, Bossi S, et al. Ginkgo biloba responds to herbivory by activating early signaling and direct defenses. *PLoS One*. 2012; 7:e32822. <https://doi.org/10.1371/journal.pone.0032822> PMID: 22448229
90. Mohanta TK, Arora PK, Mohanta N, Parida P, Bae H. Identification of new members of the MAPK gene family in plants shows diverse conserved domains and novel activation loop variants. *BMC Genomics*. 2015;16. <https://doi.org/10.1186/s12864-014-1191-8>
91. Mohanta TK, Mohanta N, Mohanta YK, Bae H. Genome-Wide Identification of Calcium Dependent Protein Kinase Gene Family in Plant Lineage Shows Presence of Novel D-x-D and D-E-L Motifs in EF-Hand Domain. *Front Plant Sci*. 2015; 6:1146. <https://doi.org/10.3389/fpls.2015.01146> PMID: 26734045
92. Zhang X, Bernoux M, Bentham AR, Newman TE, Ve T, Casey LW, et al. Multiple functional self-association interfaces in plant TIR domains. *Proc Natl Acad Sci*. 2017; 114:E2046 LP–E2052.
93. Mohanta TK, Park Y-H, Bae H. Novel Genomic and Evolutionary Insight of WRKY Transcription Factors in Plant Lineage. *Sci Rep*. 2016;6. <https://doi.org/10.1038/s41598-016-0015-2>
94. Rinerson CI, Rabara RC, Tripathi P, Shen QJ, Rushton PJ. The evolution of WRKY transcription factors. *BMC Plant Biol*. 2015; 15:1–18. <https://doi.org/10.1186/s12870-014-0410-4>
95. Holger T, Mario B, Angela S, Bernd-Joachim T. Expression of the gene and processed pseudogenes encoding the human and rabbit translationally controlled tumour protein (TCTP). *Eur J Biochem*. Wiley/Blackwell (10.1111); 2003; 267:5473–81.
96. Takada S, Hibara KI, Ishida T, Tasaka M. The CUP-SHAPED COTYLEDON1 gene of Arabidopsis regulates shoot apical meristem formation. *Development*. 2001; 128:1127–35. PMID: 11245578
97. Guo H-S, Xie Q, Fei J-F, Chua N-H. MicroRNA Directs mRNA Cleavage of the Transcription Factor NAC1 to Downregulate Auxin Signals for Arabidopsis Lateral Root Development. *Plant Cell* [Internet]. American Society of Plant Biologists; 2005; 17:1376–86. Available from: <http://www.plantcell.org/content/17/5/1376> <https://doi.org/10.1105/tpc.105.030841> PMID: 15829603
98. Vidal EA, Álvarez JM, Gutiérrez RA. Nitrate regulation of AFB3 and NAC4 gene expression in Arabidopsis roots depends on NRT1.1 nitrate transport function. *Plant Signal Behav* [Internet]. Taylor & Francis; 2014; 9:e28501. Available from: <https://doi.org/10.4161/psb.28501> PMID: 24642706
99. Faustino LI, Moretti AP, Graciano C. Fertilization with urea, ammonium and nitrate produce different effects on growth, hydraulic traits and drought tolerance in Pinus taeda seedlings. *Tree Physiol*. 2015; 35:1062–74. <https://doi.org/10.1093/treephys/tpv068> PMID: 26232784
100. Xia N, Zhang G, Sun Y-F, Zhu L, Xu L-S, Chen X-M, et al. TaNAC8, a novel NAC transcription factor gene in wheat, responds to stripe rust pathogen infection and abiotic stresses. *Physiol Mol Plant Pathol* [Internet]. 2010; 74:394–402. Available from: <http://www.sciencedirect.com/science/article/pii/S0885576510000470>
101. Chen L, Ren J, Shi H, Chen X, Zhang M, Pan Y, et al. Physiological and Molecular Responses to Salt Stress in Wild Emmer and Cultivated Wheat. *Plant Mol Biol Report* [Internet]. 2013; 31:1212–9. Available from: <https://doi.org/10.1007/s11105-013-0584-1>

102. Kirkpatrick CL, Martins D, Redder P, Frandi A, Mignolet J, Chapalay JB, et al. Growth control switch by a DNA-damage-inducible toxin–antitoxin system in *Caulobacter crescentus*. *Nat Microbiol*. Macmillan Publishers Limited; 2016; 1:16008. <https://doi.org/10.1038/nmicrobiol.2016.8> PMID: 27572440
103. Xu X, Liu Q, Fan L, Cui X, Zhou X. Analysis of synonymous codon usage and evolution of begomoviruses. *J Zhejiang Univ Sci B*. Hangzhou: Zhejiang University Press; 2008; 9:667–74. <https://doi.org/10.1631/jzus.B0820005> PMID: 18763298
104. Hu T, Banzhaf W. Nonsynonymous to Synonymous Substitution Ratio k_a / k_s : Measurement for Rate of Evolution in Evolutionary Computation. PPSNX, LNCS 5199. 2008. p. 448–57.
105. Tomoko O. Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. *J Mol Evol*. 1995; 40:56–63. <https://doi.org/10.1007/bf00166595> PMID: 7714912
106. Slotte T, Hazzouri KM, Ågren JA, Koenig D, Maumus F, Guo Y-L, et al. The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat Genet*. Nature Publishing Group; 2013; 45:831–5.
107. Bolger A, Scossa F, Bolger ME, Lanz C, Maumus F, Tohge T, et al. The genome of the stress-tolerant wild tomato species *Solanum pennellii*. *Nat Genet*. The Author(s); 2014; 46:1034. <https://doi.org/10.1038/ng.3046> PMID: 25064008
108. Zhang J. Evolution by gene duplication: an update. *Trends Ecol Evol* [Internet]. 2003; 18:292–8. Available from: <http://www.sciencedirect.com/science/article/pii/S0169534703000338>
109. Crow KD, Wagner GP. What Is the Role of Genome Duplication in the Evolution of Complexity and Diversity? *Mol Biol Evol* [Internet]. 2005; 23:887–92. Available from: <https://doi.org/10.1093/molbev/msj083>
110. Zhang J, Zhang Y, Rosenberg HF. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nat Genet* [Internet]. 2002; 30:411–5. Available from: <https://doi.org/10.1038/ng852> PMID: 11925567
111. Jin X, Ren J, Nevo E, Yin X, Sun D, Peng J. Divergent Evolutionary Patterns of NAC Transcription Factors Are Associated with Diversification and Gene Duplications in Angiosperm. *Front Plant Sci*. 2017; 8:1156. <https://doi.org/10.3389/fpls.2017.01156> PMID: 28713414
112. Schmitz JF, Zimmer F, Bornberg-Bauer E. Mechanisms of transcription factor evolution in Metazoa. *Nucleic Acids Res*. Oxford University Press; 2016; 44:6287–97. <https://doi.org/10.1093/nar/gkw492> PMID: 27288445