

# A comprehensive bioinformatics study on Tea (*Camellia sinensis*) Genome for WRKY transcription factor

<sup>1</sup>Moumita Nath, <sup>2</sup>Mousumi Datta, <sup>3</sup>Aditi Nayak

Department of Life Science  
Guru Nanak Institute of Pharmaceutical Science and Technology  
157/F, Nilgunj Rd, Sahid Colony, Panihati, Kolkata, West Bengal-700114, India

**Abstract-** Tea (*Camellia sinensis*; *CS*) is one of the world's most popular aromatic and non-alcoholic beverages, with great economic, health, & cultural importance. It is one of the major evergreen commodities grown in over 50 nations across the world. Tea leaves contain a variety of bioactive substances, particularly caffeine, theophylline, catechins, and flavonoids, which have been proven to have antioxidant, anti-inflammatory, and anti-cancer activities. Nowadays, the major concern is that the quality of tea leaves is getting deteriorated due to various environmental stresses. Fascinatingly, a transcription factor named as 'WRKY' has been found to play a key role in the development and growth of plants in response to these stresses. However, research on the role of WRKY in tea plants is limited. Therefore, we focused on the identification and characterization of this transcription factor (*CsWRKY*) in Tea plants through bioinformatics analyses. Our *in silico* study has found a total of 92 WRKY transcription factors. Out of the 46 were found to be basic in nature, 2 were neutral, and 44 were acidic in nature. As per our chromosomal distribution study, the *CsWRKY* genes were mostly located at chromosome numbers 1 to 10, 12, and 13. The chromosome number 1 was found to contain the maximum number of *CsWRKY* genes and the least were present on Chr. 12 and 13. Further, our phylogenetic study found that most of the *CsWRKY* genes were orthologous to *AtWRKY* (*Arabidopsis thaliana*). Our expression analysis reported that most of these genes are expressed in bud, flower, leaf, ovary, root, seed, and shoot of the Tea plant. This study also revealed the highly conserved motifs and structures of *CsWRKY*. Considering together, this study provides in-depth information regarding the *CsWRKY* gene family, and further research is required to validate the stress-tolerant activity of *CsWRKY*.

**Keywords:** *Camellia sinensis*, *CsWRKY*, Biotic stress, Abiotic stress, Phylogenetic analysis.

## 1. INTRODUCTION

Most of the plant cells respond differently with respect to environmental stresses. At the molecular level, it affects gene expression where some genes are upregulated and some are downregulated. WRKY is one such transcription factor (TFs) that regulates the gene expression against these external stimuli. They are a huge group of regulatory molecules in plants (1). The name is WRKY because of the presence of the conserved amino acid sequence 'WRKYGQK'. They play the role of both activator and repressor in plant signaling. The WRKY domain is 60 amino acid-long with a conserved heptapeptide of WRKYGQK at the N-terminus and unique zinc-finger-like motifs at the C-terminal region. Both of these conserved regions are necessary for the specific binding to the W box (TTGACT/C) cis-acting sites present in the promoter of target genes (2). A majority of proteins that have one WRKY domain belonged to group II, however, only those with two WRKY domains do. Putative zinc ligands for the two groups I and group II protein followed a similar structure (C-X4-5-C-X22-23-H-X1-H). Because WRKY proteins only have one WRKY domain, they are classified as Group III. The C2H2 pattern is replaced with WRKY domains from group III. Group II proteins are further divided into IIa, IIb, IIc, IId, and IIe based on their primary amino acid sequence (3).

The perennial evergreen plant known as tea (*Camellia sinensis* L.) has significant commercial value and is a member of the Theaceae family. There are several different bioactive substances in tea, including polyphenols, catechins, flavonoids, and alkaloid compounds that contribute to its flavor, scent, and possible health advantages. Tea is drunk all over the world. It is rich in various secondary metabolites, which are important for human health. Tea is rich in beneficial metabolites, such as tea polyphenols, theanine, and polysaccharides, and has been widely reported to prevent cancer, cardiovascular, and neurodegenerative diseases (4–8). Catechins (flavan-3-ols), a group of polyphenolic compounds produced via flavonoid biosynthesis, have a main contribution to protect cells against oxidative stress (9,10). Tea plant can grow in moist climate and in acidic soil of tropical, subtropical, and temperate regions (11). However, tea plants have been subjected to varied climates and complex geography that affect tea yield and quality (8). In plants, relevant mechanisms have been established to demonstrate growth and reproduction in diverse environments (12). WRKY proteins are one of the most important transcription factor families that regulate various signaling pathways of several biological processes related to plant stress defenses, development, and metabolism (13). However, the WRKY transcription factor family in tea plant has yet to be elucidated. Therefore, identification and analysis of WRKY family transcription factors in tea plant have a profound significance. Our *in silico* study has provided a thorough analysis of the tea plant (*Camellia sinensis*) WRKY family transcription factors. This research also laid the groundwork for further molecular identification and genetic breeding by providing a comprehensive scan of *CsWRKY* transcription factors.

## 2. MATERIALS AND METHODS

### 2.1. Identification of *CsWRKYs* gene family

The information on the genomic sequences of tea plants and proteins was amassed in a worldwide network of integrated plant databases. The plant transcription factors database, including all 92 sequences of tea proteins, was used to produce the data annotation. The presence of a WRKY domain in an accessed protein sequence was predicted using the online TeaCon database, and the chemical and physical characteristics of putative *CsWRKYs* were confirmed using a pay-program tool that gathered amino acid numbers determining each protein sequence's length, the molecular weight of the protein (MW), and an isoelectric point (pI) (14).

### 2.2. Phylogenetic analysis of *CsWRKY* genes

The MEGA7.0 program was used to create a disconnected evolutionary tree of recognized sequences of protein of *CsWRKY* in the species *C. sinensis* & *A. thaliana*. The following parameters were used to generate the phylogenetic tree: a phylogenetic test using 1000 bootstrap replicates, gap/missing information treatment: partially deletion, the rate across sites: gamma distribution with invariant (G) models/method: LG model, and the maximum likelihood technique (ML), respectively (15).

### 2.3. Analysis of *CsWRKY* Expression of Genes in Various Tissues and Organs

The TeaCon database was used to acquire DNA sequence data regarding the *CsWRKY* families in order to analyze the patterns of gene regulation of the *CsWRKY* families in various tissues and organs. The gene id was used to calculate the genome-wide transcription frequencies of the *CsWRKY* genes.

### 2.4. Exon-intron structure, conservation domains, and motif descriptions are provided

The newly found *CsWRKY* gene structure was investigated by mapping the CDS to the DNA sequences using the online GSDS2.0 database. *C. sinensis* genomic DNA and CDS sequences were downloaded using the Tea genome database (16).

### 2.5. Pathways analysis of *CsWRKY*

KEGG pathway (17) is a collection of manually drawn pathway maps representing our knowledge of the molecular interaction, reaction and relation networks. Here, Organism-specific versions were created for KEGG pathway maps, BRITE hierarchies and KEGG modules through the KEGG Orthology (KO) system, either as static files in the daily database update procedure (for the well-annotated genomes in KEGG GENES) or as temporary files on the fly (for GhostKOALA-annotated MGENES). The organism-specific pathway maps and module maps are colored in green, which is a KEGG convention.

## 3. RESULTS

### 3.1. Identification of WRKY transcription factor in Tea plant

First of all, from the tea plant genome, 92 probable associates of WRKY transcription factor genes were identified after removal of redundant and incomplete protein sequences. The identified 92 WRKY were renamed as TEA015194.1 to TEA020191.1 on the bases of the regularity order of the WRKY TF genes on chromosomes (1–13) of tea. The length of *CsWRKY* protein sequences ranges from 29 (TEA008808.1) to 735 (TEA030233.1) amino acids. The isoelectric point ranged from 4.77 (TEA020545.1) to 10.071 (TEA028765.1), suggesting the operation of the different *CsWRKYs* in various microenvironments (18). Out of them, 46 were found to be basic in nature, 2 were neutral, and 44 were acidic in nature. Interestingly composition of the total proteins (87.09 %) of *CsWRKYs* showed higher proportion of WRKY superfamily protein (average ~715 %) amino acid. The parameters used to describe have been shown in (Table 1).

### 3.2. Chromosomal location of WRKY in tea plant

Identified *CsWRKY* genes were found to be distributed in the thirteen tea chromosomes with uneven distribution (Table 2). The study found that relatively more *CsWRKY* genes are distributed on chromosome 3 (25.80 %), and least on chromosome 5 (8.06 %). This paradox suggested that the mapped *CsWRKY* genes on all the chromosomes had a significantly uneven distribution and non-random. As per our chromosomal distribution study, the *CsWRKY* genes were mostly located at chromosome numbers 1 to 10, 12, and 13. The chromosome number 1 was found to contain the maximum number of *CsWRKY* genes and the least were present on Chr. 12 and 13. The chromosomal location of TEA020545.1, TEA022377.1, TEA006586.1, TEA026188.1, TEA001873.1, and TEA011588.1 were not found in the analysis.

**Table 1.** List of putative *CsWRKY* genes and their physicochemical properties, variants, and number of WRKY domains.

Sl. No.	Gene ID	Protein Name	TF ID	Taxonomic ID	Proteins properties			WRKY Domain	
					Length (aa)	MW (Da)	pI	Start	End
1	TEA015194.1	WRKY DNA-binding protein 72, putative	XP_028051760.1	542762	541	58403.5	7.3916	211	270
2	TEA028099.1	WRKY transcription factor 1-like isoform	XP_028051846.1	542762	468	51752.	6.4626	195 369	252 427

3	TEA002467.1	WRKY DNA-binding protein 33 isoform	XP_028051918.1	542762	557	61469.3	8.8917	221389	278447
4	TEA002471.1	FMN-linked oxidoreductases superfamily protein	XP_028051958.1	542762	346	38234.3	10.1369	275	332
5	TEA008808.1	WRKY DNA-binding protein 70, putative isoform	XP_028052697.1	542762	29	33397.2	5.0936	130	190
6	TEA001162.1	WRKY DNA-binding domain	XP_028053974.1	542762	296	32041.2	6.3139	151	208
7	TEA030998.1	Translation initiation factor 6	XP_028054239.1	542762	165	18114.9	4.8841	128	147
8	TEA007370.1	WRKY DNA-binding protein 70, putative isoform	XP_028055071.1	542762	325	36726.6	6.9482	131	191
9	TEA014384.1	WRKY transcription factor 22-like	XP_028057975.1	542762	328	35845.6	6.2539	161	219
10	TEA012365.1	WRKY18 transcription factor	XP_028059704.1	542762	290	32749.9	7.6868	156	213
11	TEA012360.1	WRKY18 transcription factor	XP_028059727.1	542762	293	32512.2	7.6292	130	188
12	TEA019167.1	WRKY DNA-binding domain	XP_028061059.1	542762	227	25989.1	8.6764	139	197
13	TEA033820.1	WRKY transcription factor	XP_028062068.1	542762	186	20611.7	5.6583	117	175
14	TEA005142.1	WRKY transcription factor 6-like	XP_028064109.1	542762	613	66861	6.6157	352	410
15	TEA029093.1	Dihydrofolate synthase/folylpolyglutamate synthase	XP_028066881.1	542762	117	12659.7	8.4876	9	32
16	TEA031948.1	Putative WRKY1a transcription factor	XP_028067460.1	542762	589	63901.2	7.0207	322	380
17	TEA004305.1	WRKY transcription factor	XP_028071264.1	542762	442	48385.5	5.1083	230	288

18	TEA004132.1	WRKY family transcription factor family protein	XP_028072625.1	542762	475	52431.2	8.7562	197	253
19	TEA023777.1	WRKY transcription factor	XP_028072777.1	542762	269	29692.8	5.6358	76	134
20	TEA005334.1	WRKY family transcription factor	XP_028075202.1	542762	287	30358.1	9.6215	33	91
21	TEA006754.1	Putative WRKY transcription factor 15 - like protein	XP_028077060.1	542762	324	35493.3	9.8054	253	310
22	TEA027312.1	WRKY transcription factor 22- like isoform	XP_028077318.1	542762	330	37498.5	5.2501	162	220
23	TEA028913.1	WRKY transcription factor 5 isoform	XP_028077632.1	542762	280	31314.5	5.8787	75	133
24	TEA011602.1	WRKY DNA-binding protein 30, putative	XP_028077914.1	542762	336	37641	5.0619	120	181
25	TEA027100.1	WRKY transcription factor 55 isoform	XP_028078525.1	542762	388	40986.2	6.3118	182	242
26	TEA027106.1	WRKY transcription factor	XP_028079096.1	542762	523	57335.2	5.7849	211 380	266 436
27	TEA027083.1	WRKY DNA-binding protein 70, putative isoform	XP_028079122.1	542762	313	34468.7	4.7385	135	195
28	TEA027100.1	WRKY transcription factor 55 isoform	XP_028079124.1	542762	220	23669.1	6.6659	14	74
29	TEA008318.1	WRKY transcription factor	XP_028080001.1	542762	576	63926.6	7.959	242 408	298 466
30	TEA004792.1	Plant zinc cluster domain	XP_028081295.1	542762	322	35992.7	10.2627	257	314
31	TEA007721.1	WRKY DNA-binding protein 9, putative isoform	XP_028081368.1	542762	486	53307.1	6.1507	229	288
33	TEA027058.1	WRKY18 transcription factor	XP_028081784.1	542762	312	34821.1	7.4044	153	212

34	TEA0 22377. 1	WRKY DNA- binding protein isoform 57	XP_028082795. 1	542762	303	33595. 8	5.9643	162	219
35	TEA0 08812. 1	WRKY transcription factor 55 isoform	XP_028083209. 1	542762	321	35502. 8	7.9828	145	205
36	TEA0 08808. 1	WRKY DNA- binding protein 70, putative isoform	XP_028083248. 1	542762	296	33397. 2	5.0936	130	190
37	TEA0 27198. 1	WRKY transcription factor, putative	XP_028084020. 1	542762	182	20808. 5	9.9083	101	159
38	TEA0 25613. 1 TEA0 11326. 1	WRKY DNA- binding protein 27, putative	XP_028084310. 1	542762	193	22072. 7	8.1188	70	128
39	TEA0 05213. 1	Putative WRKY transcription factor	XP_028085508. 1	542762	355	40449. 9	4.7801	128	188
40	TEA0 31299. 1	WRKY transcription factor 41 family protein	XP_028087006. 1	542762	336	38060. 1	6.3161	127	187
41	TEA0 07197. 1	Putative WRKY transcription factor	XP_028089465. 1	542762	307	34434. 7	7.15	158	215
42	TEA0 11330. 1	WRKY transcription factor 39 family protein	XP_028089628. 1	542762	337	38034. 4	10.417 8	270	327
43	TEA0 28167. 1	WRKY transcription factor	XP_028091443. 1	542762	522	57343	7.6458	232 408	288 466
44	TEA0 23233. 1	WRKY super family protein	XP_028091585. 1	542762	337	37069. 9	6.6786	187	244
45	TEA0 04132. 1	WRKY family transcription factor family protein	XP_028091903. 1	542762	478	52138	9.0901	199 397	255 454
46	TEA0 30336. 1	WRKY family transcription factor	XP_028092641. 1	542762	222	25341. 7	7.2358	148	206
47	TEA0 30312. 1	P-loop containing nucleoside triphosphate hydrolases	XP_028093134. 1	542762	99	10689. 8	9.2266	71	93

		superfamily protein isoform							
48	TEA006586.1	WRKY transcription factor, putative	XP_028093640.1	542762	186	21087.5	9.6582	107	165
49	TEA023297.1	WRKY super family protein	XP_028094020.1	542762	255	29081.8	4.8493	55	113
50	TEA005358.1 J	WRKY transcription factor 1-like isoform	XP_028094353.1	542762	433	48408.7	9.2572	155329	212387
51	TEA026188.1	WRKY DNA-binding protein 70, putative isoform	XP_028094837.1	542762	331	36724.8	6.8805	125	185
52	TEA001873.1	WRKY super family protein	XP_028100462.1	542762	322	35693.6	6.8496	177	234
53	TEA002329.1	WRKY DNA - binding domain	XP_028100469.1	542762	243	28316.6	9.0231	139187	163227
54	TEA023720.1	Plant zinc cluster domain	XP_028100648.1	542762	319	35482	9.9554	247	304
55	TEA030233.1	WRKY super family protein	XP_028101411.1	542762	715	77797.6	7.1857	248452	305510
56	TEA011588.1	WRKY transcription factor	XP_028103268.1	542762	555	60997.5	6.6413	230383	287441
57	TEA028505.1	WRKY DNA-binding protein 33 isoform	XP_028106608.1	542762	569	63061.7	8.282	227392	284450
58	TEA029100.1	WRKY DNA-binding protein 30, putative	XP_028106713.1	542762	364	39755.8	5.0625	129	190
59	TEA016074.1	WRKY transcription factor	XP_028107295.1	542762	363	40371.9	7.5437	163	221
60	TEA023807.1	WRKY DNA - binding domain	XP_028107356.1	542762	173	19830.3	10.0352	93	151
61	TEA010793.1	WRKY family transcription factor	XP_028107506.1	542762	577	63049.6	6.5784	330	388
62	TEA013771.1	WRKY DNA - binding domain	XP_028109415.1	542762	199	22457.5	10.1077	120	178

63	TEA013772.1	WRKY transcription factor 22-like	XP_028109416.1	542762	316	34951.4	6.5837	151	209
64	TEA028473.1	WRKY DNA-binding domain	XP_028109764.1	542762	319	35605.7	7.3257	156	214
65	TEA009907.1	WRKY family transcription factor	XP_028110002.1	542762	463	51218.6	7.6737	267	325
66	TEA003387.1	WRKY transcription factor	XP_028111113.1	542762	526	57356.3	8.2245	236 411	292 469
67	TEA006755.1	WRKY transcription factor 13 family protein	XP_028112541.1	542762	235	26611.1	9.5994	158	216
68	TEA001484.1	WRKY11 transcription factor	XP_028113091.1	542762	328	35713.4	10.1798	251	309
69	TEA027106.1	WRKY transcription factor	XP_028113356.1	542762	520	57061.3	6.1166	214 384	270 440
70	TEA008513.1	Putative WRKY transcription factor 48-like protein	XP_028115186.1	542762	311	34324.3	7.0322	159	217
71	TEA031299.1	WRKY transcription factor 41 family protein	XP_028115466.1	542762	363	40781	4.9772	130	190
72	TEA007789.1	WRKY transcription factor 23-1	XP_028115574.1	542762	197	22227.3	9.6324	118	176
73	TEA007804.1	WRKY transcription factor 22-like isoform	XP_028115579.1	542762	336	36788.7	4.8849	159	217
74	TEA020545.1	Putative WRKY transcription factor	XP_028116770.1	542762	337	37674	4.7721	116	175
75	TEA030233.1	WRKY super family protein	XP_028117015.1	542762	735	79548.8	6.3654	309 521	366 579
76	TEA017542.1	WRKY DNA-binding protein 35, putative	XP_028117815.1	542762	275	31056.9	4.4842	53	111
77	TEA017544.1	WRKY super family protein	XP_028117834.1	5452762	331	37046.7	8.577	149	212
78	TEA030998.1	Translation initiation factor 6	XP_028118079.1	542762	134	14560.6	8.4804	106	126

79	TEA008132.1	WRKY transcription factor 18-like	XP_028118081.1	542762	235	26582.1	8.1111	100	158
80	TEA008134.1	FMN-linked oxidoreductases superfamily protein, partial	XP_028118126.1	542762	340	37988.8	6.525	313	332
81	TEA023720.1	Plant zinc cluster domain	XP_028119049.1	542762	282	31177.8	9.8933	250	276
82	TEA005334.1	WRKY family transcription factor	XP_028119210.1	542762	559	60702.6	8.162	305	363
83	TEA020387.1	WRKY DNA-binding protein 30, putative	XP_028119435.1	542762	336	37640.1	5.3655	120	181
84	TEA003077.1	WRKY DNA-binding protein 72, putative	XP_028120991.1	542762	611	66726.9	7.4496	263	321
85	TEA026850.1	WRKY DNA-binding protein 72, putative	XP_028121796.1	542762	547	60193	6.4311	247	306
86	TEA011154.1	WRKY transcription factor	XP_028122384.1	542762	318	35125.3	8.0069	164	223
87	TEA022842.1	Plant zinc cluster domain	XP_028124442.1	542762	345	38960.9	10.387	279	337
88	TEA019607.1	Putative WRKY transcription factor	XP_028124717.1	542762	303	33786.8	4.8139	114	170
89	TEA028765.1	WRKY transcription factor, putative	XP_028124928.1	542762	181	20467	10.0718	102	160
90	TEA005255.1	WRKY transcription factor 51 family protein	XP_028125074.1	542762	192	21654.8	6.504	104	162
91	TEA020117.1	Transcription termination factor MTERF9, chloroplastic	XP_028127034.1	542762	604	65595.4	6.6741	244 417	301 475
92	TEA020191.1	WRKY transcription factor	XP_028127360.1	542762	320	35075.3	8.8922	160	219

Table 2. Chromosomal distribution of CsWRKYs.



Sl no.	Gene ID	Protein Name	TF ID	Taxonomic ID	Chromosomal location
1	TEA015194.1	WRKY DNA-binding protein 72, putative	XP_028051760.1	542762	Chr05:776392-781485(-)
2	TEA028099.1	WRKY transcription factor 1-like isoform	XP_028051846.1	542762	Chr01:1882448-1885976(-)
3	TEA002467.1	WRKY DNA-binding protein 33 isoform	XP_028051918.1	542762	Chr03:299883-302756(+)
4	TEA002471.1	FMN-linked oxidoreductases superfamily protein	XP_028051958.1	542762	Chr03:1688743-1716217(+)
5	TEA008808.1	WRKY DNA-binding protein 70, putative isoform	XP_028052697.1	542762	Chr04:430848-433334(-)
6	TEA001162.1	WRKY DNA - binding domain	XP_028053974.1	542762	Chr03:686293-690246(-)
7	TEA030998.1	Translation initiation factor 6	XP_028054239.1	542762	Chr06:1079774-1096574(-)
8	TEA007370.1	WRKY DNA-binding protein 70, putative isoform	XP_028055071.1	542762	Chr01:307023-309236(+)
9	TEA014384.1	WRKY transcription factor 22-like	XP_028057975.1	542762	Chr02:1272608-1274342(+)
10	TEA012365.1	WRKY18 transcription factor	XP_028059704.1	542762	Chr01:732103-745499(-)
11	TEA012360.1	WRKY18 transcription factor	XP_028059727.1	542762	Chr01:823531-826145(-)
12	TEA019167.1	WRKY DNA - binding domain	XP_028061059.1	542762	Chr01:1575674-1578863(-)
13	TEA033820.1	WRKY transcription factor	XP_028062068.1	542762	Chr01:1844883-1846475(-)
14	TEA005142.1	WRKY transcription factor 6-like	XP_028064109.1	542762	Chr02:601835-605813(-)
15	TEA029093.1	Dihydrofolate synthase/folylpoly glutamate synthase	XP_028066881.1	542762	Chr07:723455-732491(-)
16	TEA031948.1	Putative WRKY1a transcription factor	XP_028067460.1	542762	Chr01:638294-642060(+)
17	TEA004305.1	WRKY transcription factor	XP_028071264.1	542762	Chr03:997322-1000207(-)
18	TEA004132[.1	WRKY family transcription factor family protein	XP_028072625.1	542762	Chr05:714276-718356(+)

19	TEA023777.1	WRKY transcription factor	XP_028072777.1	542762	Chr05:262202-263401(+)
20	TEA005334.1	WRKY family transcription factor	XP_028075202.1	542762	Chr05:366521-369321(+)
21	TEA006754.1	Putative WRKY transcription factor 15 -like protein	XP_028077060.1	542762	Chr05:311583-318065(-)
22	TEA027312.1	WRKY transcription factor 22-like isoform	XP_028077318.1	542762	Chr02:404328-405790(+)
23	TEA028913.1	WRKY transcription factor 5 isoform	XP_028077632.1	542762	Chr05:811788-812900(+)
24	TEA011602.1	WRKY DNA-binding protein 30, putative	XP_028077914.1	542762	Chr13:64541-66364(+)
25	TEA027100.1	WRKY transcription factor 55 isoform	XP_028078525.1	542762	Chr01:1666955-1670001(-)
26	TEA027106.1	WRKY transcription factor	XP_028079096.1	542762	Chr01:242198-254828(-)
27	TEA027083.1	WRKY DNA-binding protein 70, putative isoform	XP_028079122.1	542762	Chr01:1641047-1643714(+)
28	TEA027100.1	WRKY transcription factor 55 isoform	XP_028079124.1	542762	Chr01:1666955-1670001(-)
29	TEA008318.1	WRKY transcription factor	XP_028080001.1	542762	Chr01:783601-794873(-)
30	TEA004792.1	Plant zinc cluster domain	XP_028081295.1	542762	
31	TEA007721.1	WRKY DNA-binding protein 9, putative isoform	XP_028081368.1	542762	Chr01:1338565-1340946(-)
33	TEA027058.1	WRKY18 transcription factor	XP_028081784.1	542762	Chr01:1830700-1833870(-)
34	TEA022377.1	WRKY DNA-binding protein 57 isoform	XP_028082795.1	542762	-
35	TEA008812.1	WRKY transcription factor 55 isoform	XP_028083209.1	542762	Chr04:314891-322381(+)
36	TEA008808.1	WRKY DNA-binding protein 70, putative isoform	XP_028083248.1	542762	Chr04:430848-433334(-)
37	TEA027198.1	WRKY transcription factor, putative	XP_028084020.1	542762	Chr01:795490-797056(+)
38	TEA025613.1 TEA011326.1	WRKY DNA-binding protein 27, putative	XP_028084310.1	542762	Chr02:2048873-2050150(+)

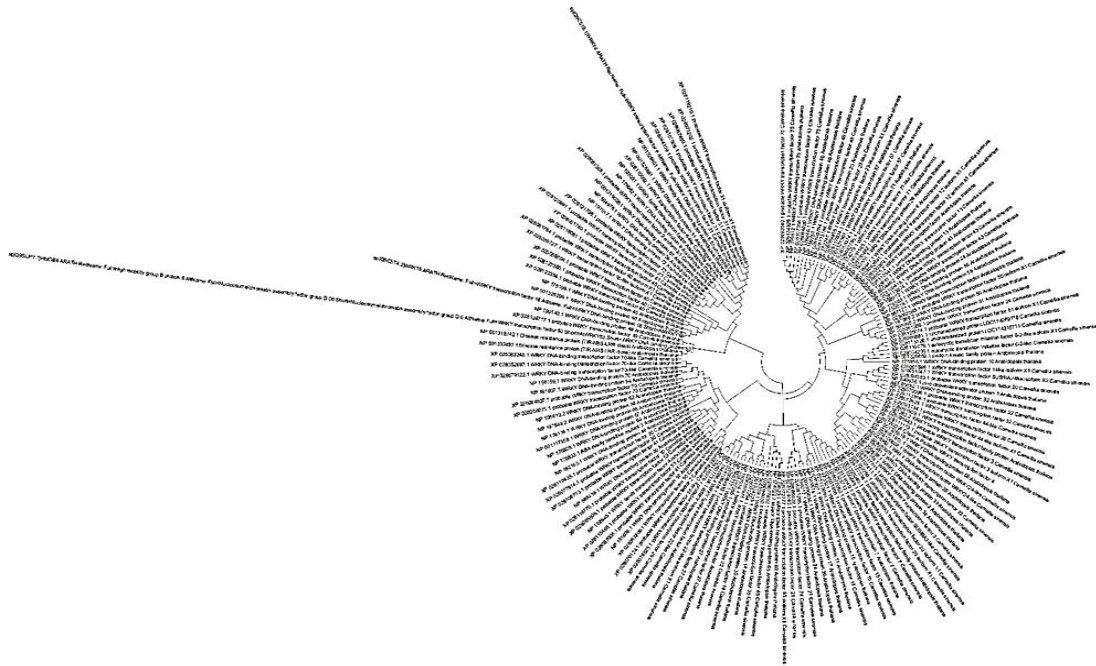
39	TEA005213.1	Putative WRKY transcription factor	XP_028085508.1	542762	Chr08:766533-770719(+)
40	TEA031299.1	WRKY transcription factor 41 family protein	XP_028087006.1	542762	Chr05:552276-554060(-)
41	TEA007197.1	Putative WRKY transcription factor	XP_028089465.1	542762	Chr12:255889-258533(-)
42	TEA011330.1	WRKY transcription factor 39 family protein	XP_028089628.1	542762	Chr02:272636-275687(-)
43	TEA028167.1	WRKY transcription factor	XP_028091443.1	542762	Chr01:676453-680728(+)
44	TEA023233.1	WRKY super family protein	XP_028091585.1	542762	Chr05:355249-356906(-)
45	TEA004132.1	WRKY family transcription factor family protein	XP_028091903.1	542762	Chr05:714276-718356(+)
46	TEA030336.1	WRKY family transcription factor	XP_028092641.1	542762	Chr04:1006727-1009288(+)
47	TEA030312.1	P-loop containing nucleoside triphosphate hydrolases superfamily protein isoform	XP_028093134.1	542762	Chr01:508284-529724(-)
48	TEA006586.1	WRKY transcription factor, putative	XP_028093640.1	542762	
49	TEA023297.1	WRKY super family protein	XP_028094020.1	542762	Chr07:78925-95783(+)
50	TEA005358.1 J	WRKY transcription factor 1-like isoform	XP_028094353.1	542762	Chr03:180957-185253(+)
51	TEA026188.1	WRKY DNA-binding protein 70, putative isoform	XP_028094837.1	542762	
52	TEA001873.1	WRKY super family protein	XP_028100462.1	542762	
53	TEA002329.1	WRKY DNA - binding domain	XP_028100469.1	542762	Chr02:152376-155299(-)
54	TEA023720.1	Plant zinc cluster domain	XP_028100648.1	542762	Chr01:1329877-1331672(+)
55	TEA030233.1	WRKY super family protein	XP_028101411.1	542762	Chr03:856995-860560(+)
56	TEA011588.1	WRKY transcription factor	XP_028103268.1	542762	
57	TEA028505.1	WRKY DNA-binding protein 33 isoform	XP_028106608.1	542762	Chr01:408528-411725(+)

58	TEA029100.1	WRKY DNA-binding protein 30, putative	XP_028106713.1	542762	Chr07:650747-653443(+)
59	TEA016074.1	WRKY transcription factor	XP_028107295.1	542762	Chr09:1649386-1650694(+)
60	TEA023807.1	WRKY DNA - binding domain	XP_028107356.1	542762	Chr04:59114-62001(+)
61	TEA010793.1	WRKY family transcription factor	XP_028107506.1	542762	Chr01:1434099-1440574(+)
62	TEA013771.1	WRKY DNA - binding domain	XP_028109415.1	542762	Chr02:892500-906216(-)
63	TEA013772.1	WRKY transcription factor 22-like	XP_028109416.1	542762	Chr02:916853-919668(-)
64	TEA028473.1	WRKY DNA - binding domain	XP_028109764.1	542762	Chr02:2772424-2773965(+)
65	TEA009907.1	WRKY family transcription factor	XP_028110002.1	542762	Chr08:538218-540713(-)
66	TEA003387.1	WRKY transcription factor	XP_028111113.1	542762	Chr01:3039767-3046181(+)
67	TEA006755.1	WRKY transcription factor 13 family protein	XP_028112541.1	542762	Chr07:816877-821297(+)
68	TEA001484.1	WRKY11 transcription factor	XP_028113091.1	542762	Chr04:2844474-2845743(+)
69	TEA027106.1	WRKY transcription factor	XP_028113356.1	542762	Chr01:242198-254828(-)
70	TEA008513.1	Putative WRKY transcription factor 48 -like protein	XP_028115186.1	542762	Chr02:1085977-1087276(-)
71	TEA031299.1	WRKY transcription factor 41 family protein	XP_028115466.1	542762	Chr05:552276-554060(-)
72	TEA007789.1	WRKY transcription factor 23-1	XP_028115574.1	542762	Chr04:1024979-1027228(-)
73	TEA007804.1	WRKY transcription factor 22-like isoform	XP_028115579.1	542762	Chr04:1057597-1059281(-)
74	TEA020545.1	Putative WRKY transcription factor	XP_028116770.1	542762	-
75	TEA030233.1	WRKY super family protein	XP_028117015.1	542762	Chr03:856995-860560(+)
76	TEA017542.1	WRKY DNA-binding protein 35, putative	XP_028117815.1	542762	Chr03:1315202-1316756(+)
77	TEA017544.1	WRKY super family protein	XP_028117834.1	5452762	Chr03:1777201-1779660(+)

78	TEA030998.1	Translation initiation factor 6	XP_028118079.1	542762	Chr06:1079774-1096574(-)
79	TEA008132.1	WRKY transcription factor 18-like	XP_028118081.1	542762	Chr01:1171624-1180172(-)
80	TEA008134.1	FMN-linked oxidoreductases superfamily protein, partial	XP_028118126.1	542762	Chr01:426612-433580(+)
81	TEA023720.1	Plant zinc cluster domain	XP_028119049.1	542762	Chr01:1329877-1331672(+)
82	TEA005334.1	WRKY family transcription factor	XP_028119210.1	542762	Chr05:366521-369321(+)
83	TEA020387.1	WRKY DNA-binding protein 30, putative	XP_028119435.1	542762	Chr01:625156-626983(-)
84	TEA003077.1	WRKY DNA-binding protein 72, putative	XP_028120991.1	542762	Chr10:1022993-1044947(-)
85	TEA026850.1	WRKY DNA-binding protein 72, putative	XP_028121796.1	542762	Chr03:1407223-1412123(+)
86	TEA011154.1	WRKY transcription factor	XP_028122384.1	542762	Chr03:741152-750066(+)
87	TEA022842.1	Plant zinc cluster domain	XP_028124442.1	542762	Chr03:997123-999071(-)
88	TEA019607.1	Putative WRKY transcription factor	XP_028124717.1	542762	Chr06:746509-749310(-)
89	TEA028765.1	WRKY transcription factor, putative	XP_028124928.1	542762	Chr10:168505-172993(+)
90	TEA005255.1	WRKY transcription factor 51 family protein	XP_028125074.1	542762	Chr11:669585-674439(-)
91	TEA020117.1	Transcription termination factor MTERF9, chloroplastic	XP_028127034.1	542762	Chr02:534002-544254(-)
92	TEA020191.1	WRKY transcription factor	XP_028127360.1	542762	Chr06:2601588-2609578(+)

### 3.3. Phylogenetic analysis of *CsWRKY*

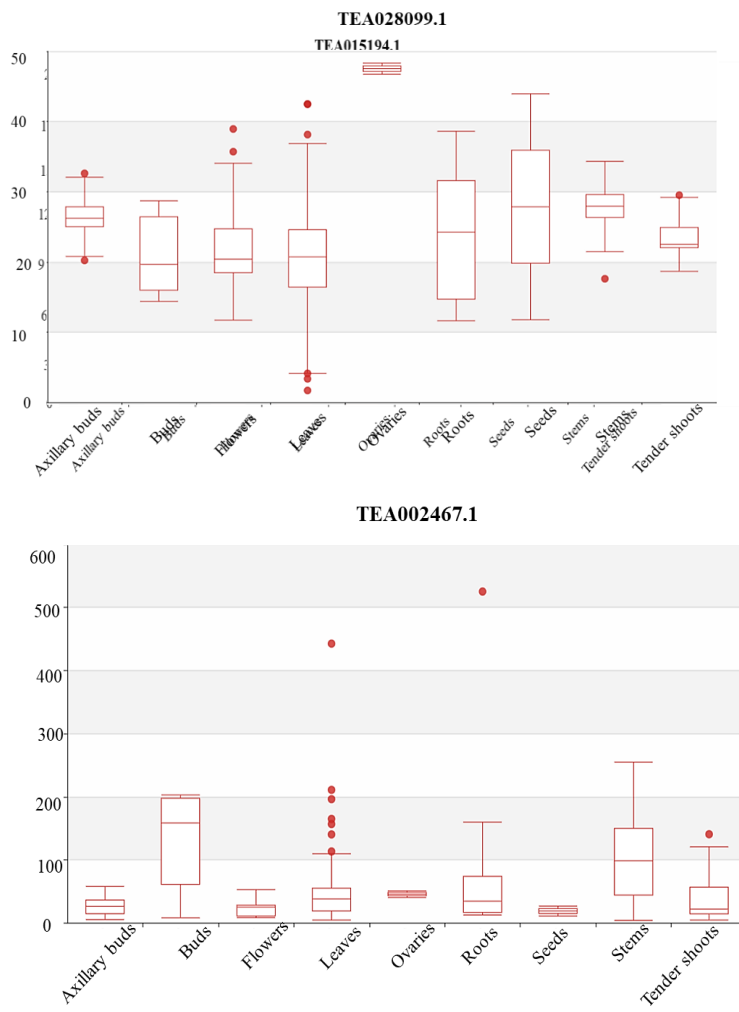
An evolutionary tree is a schematic diagram or tree that depicts the predicted evolutionary connections and phylogeny of distinct biological organisms or other entities based on genetic similarities and differences. The taxa connected in the tree are thought to be descended from an ancestral lineage. The multiple sequence alignment was performed for all sequences of *CsWRKY* proteins and the construction of phylogenetic tree was done using the maximum-likelihood (ML) mode in MEGA 7.0 with 1,000 bootstrap replicates. The tree of phylogeny was built using neighbor-joining. For 1,000 encounters, the bootstrap parameters were determined.

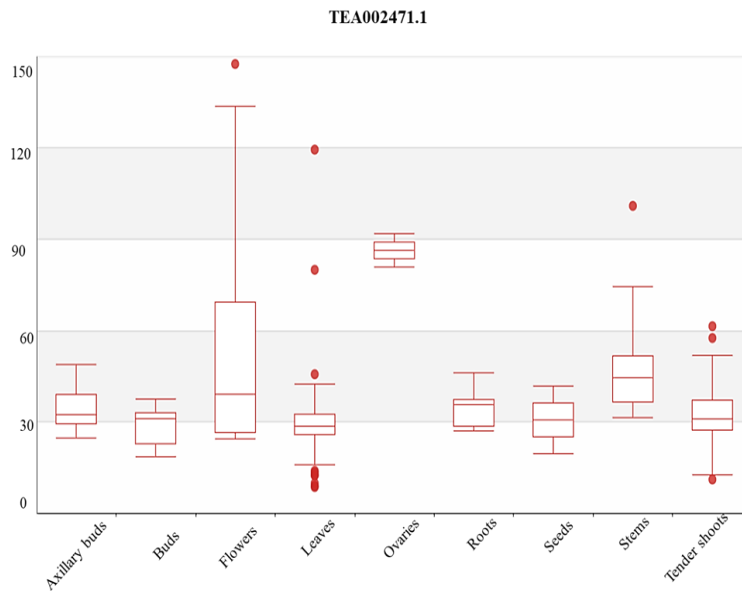


**Figure 1. Phylogenetic tree demonstrating interactions among Tea WRKY proteins.** The multiple sequence alignment was performed for all sequences of *CsWRKY* proteins and the construction of phylogenetic tree was done using the maximum-likelihood (ML) mode in MEGA 7.0 with 1,000 bootstrap replicates.

**3.4. Expression of WRKY in different plant parts**

*CsWRKY* gene expression levels in Tea plant tissues (Axillary buds, buds, leaves, flowers, ovaries, roots, seeds, & stems) were studied. Interestingly, TEA015194.1 was found to be overexpressed in roots.

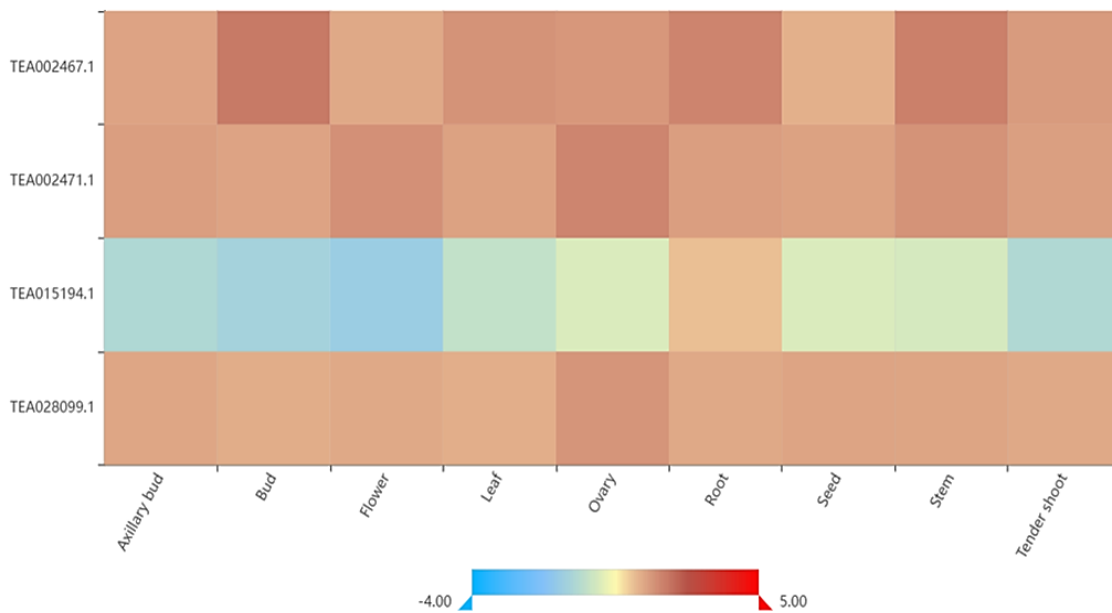




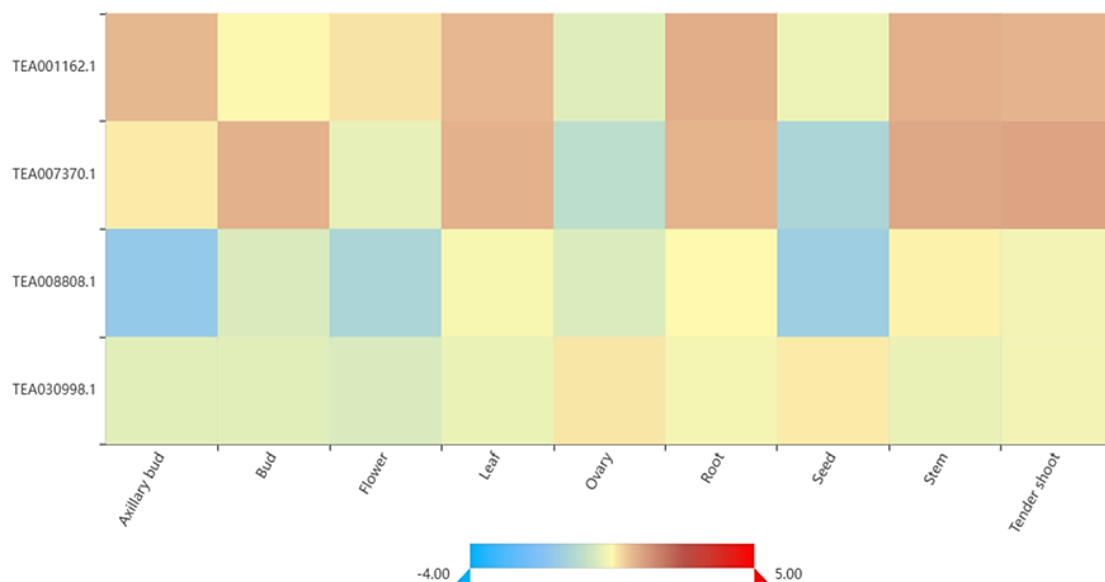
**Figure 2. WRKY expression profile in different parts of Tea plant.**

**3.5. Co-expression analysis of different CsWRKY in various regions of the Tea Plant**

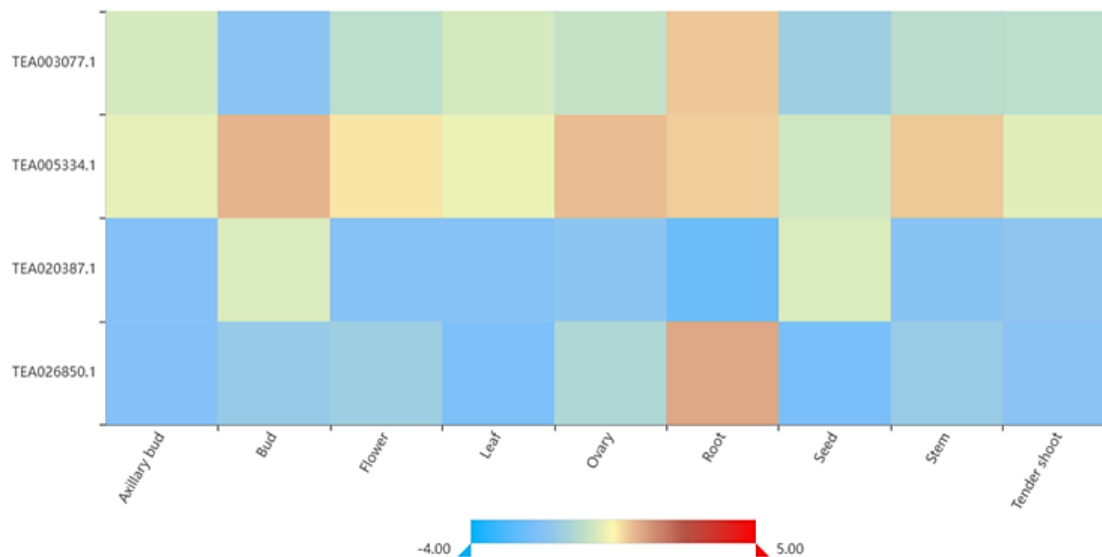
DNA-seq data of CsWRKY families were obtained from the TeaCon databases, which include the expression levels across root, leaf, blossom, stem, and seed to analyze the expression profile, of the CsWRKY. The overall expression level of CsWRKY transcripts was estimated utilizing standardized analysis and visualization of gene expression information for each gene id of a transcript was performed (Figure 3, 4, and 5). Red (dark) denotes strong expression, yellow shows medium expression, and blue indicates lower expression of gene in the figure 3, 4 and 5.



**Figure 3. Co-expression of CsWRKY genes (TEA002467.1, TEA002471.1, TEA015194.1, and TEA028099.1) in different parts of Tea plant. Red (dark) denotes strong expression, yellow shows medium expression, and blue indicates lower expression**



**Figure 4. Co-expression of *CsWRKY* genes (TEA001162.1, TEA007370.1, TEA008808.1, and TEA030998.1) in different parts of Tea plant.** Red (dark) denotes strong expression, yellow shows medium expression, and blue indicates lower expression. Almost all the *CsWRKY* genes are expressed in all parts of the plant except TEA008808.1.



**Figure 5. Co-expression of *CsWRKY* genes (TEA003077.1, TEA005334.1, TEA020387.1, and TEA026850.1) in different parts of Tea plant.** Red (dark) denotes strong expression, yellow shows medium expression, and blue indicates lower expression. Expression of TEA026850.1, TEA020387.1, and TEA003077.1 are negligible.

### 3.6. WRKY mediated signalling Pathway in Tea Plant

The MAP kinase pathway is related to stress adaptive pathways (19). Interestingly, WRKY is found to be associated with the MAPK pathway representing its role in stress responses (Figure 6).



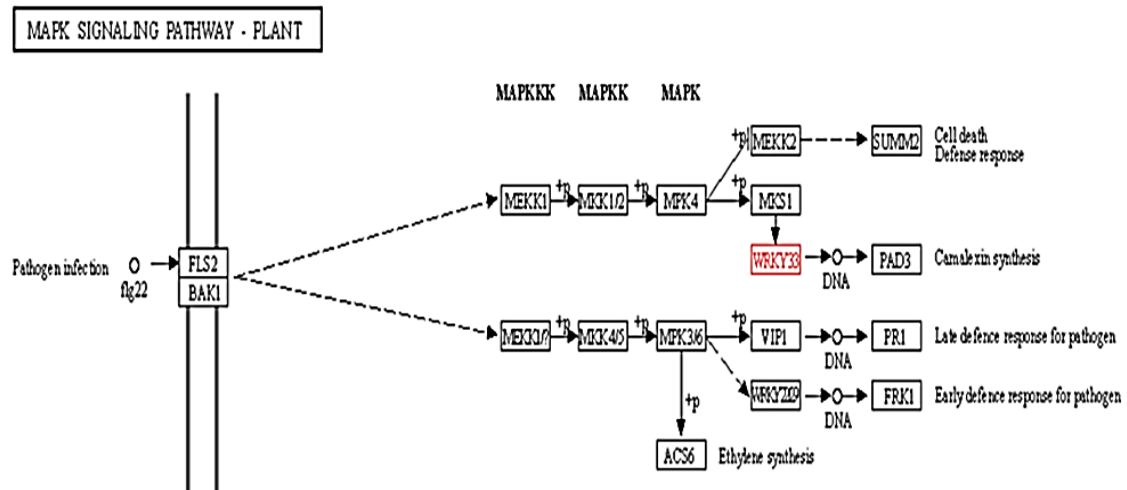


Figure 6. CsWRKYs have a role in controlling stress responses through the MAP kinase pathway.

3.7. Analysis of Tea plant motifs

The motifs present in *CsWRKY* protein were also analysed (Figure 7). Different coloured boxes represent the motif conservation in the *CsWRKY* protein. The size of the box indicates the overall length of the motif, whereas grey lines represent non-conserved sequences. Motifs 1-8 were identified as a total of 10 distinct motifs. Motifs one to three contain the WRKYGQK sequence. The length of the motif is indicated by the length of the box.

3.8. *CsWRKY* gene structure analysis

The *CsWRKY* gene's exon-intron distribution from many main groups and their related subgroups are shown by boxes in figure 8. There are no upstream or downstream sequences in the majority of WRKY.

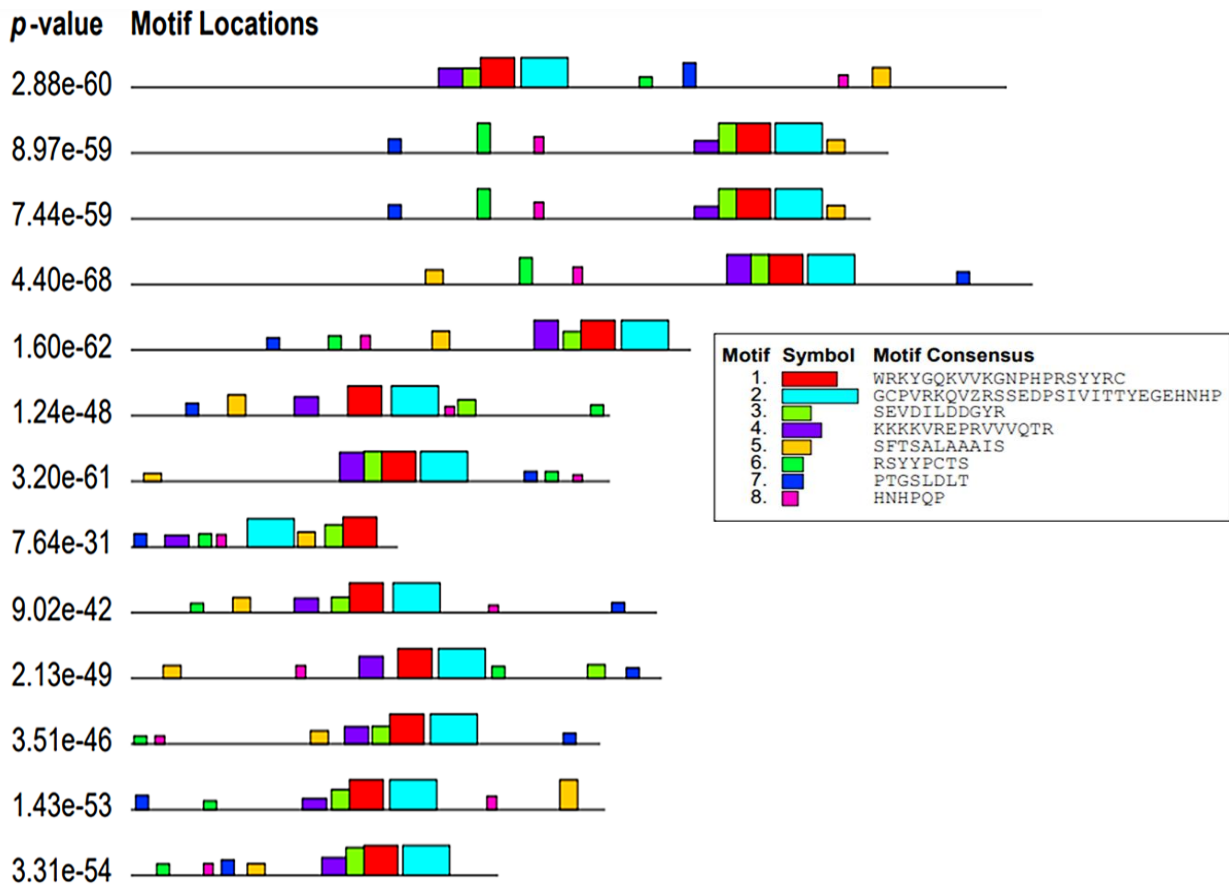
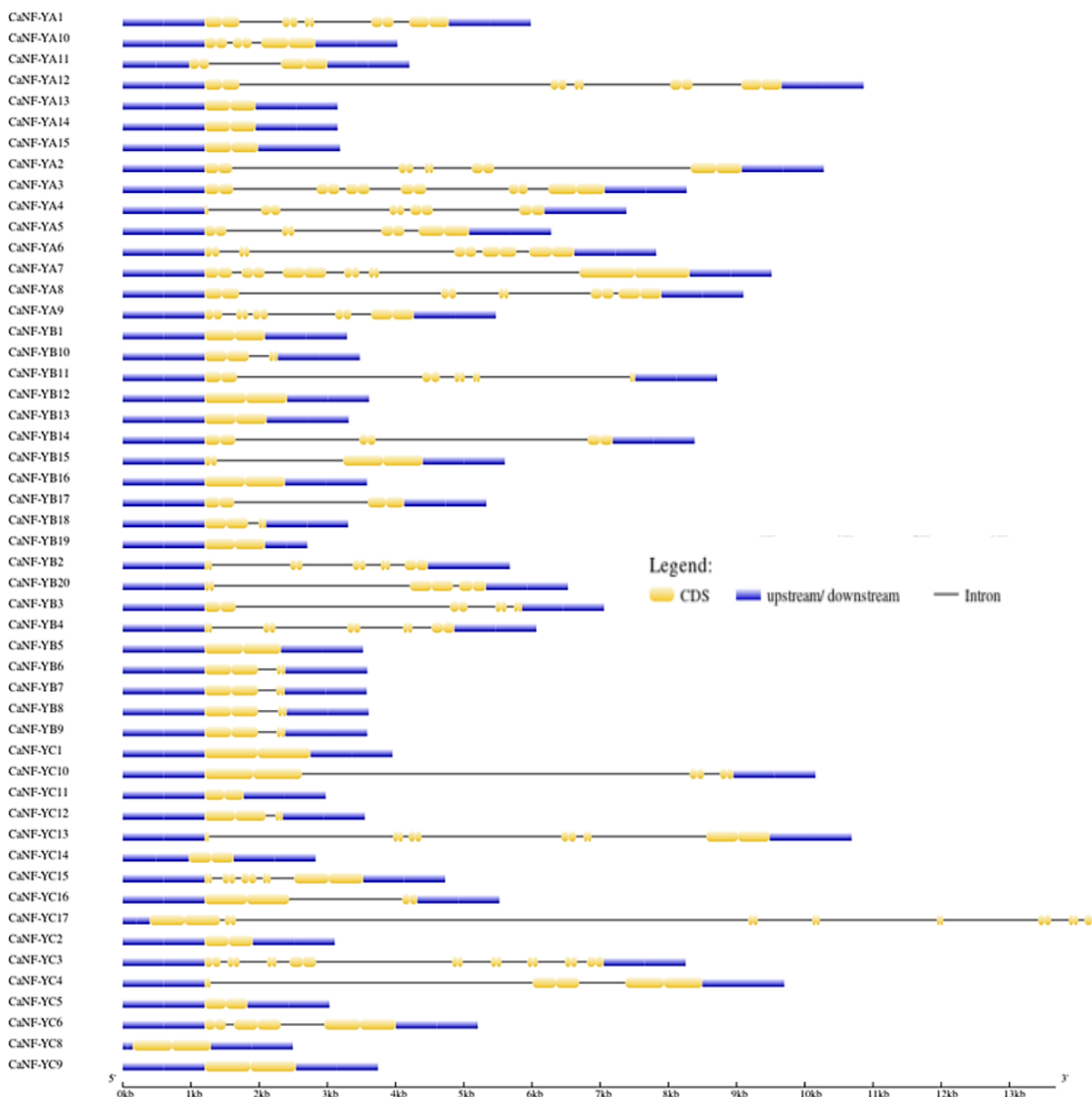


Figure 7. The conserved motif in *CsWRKY* in Tea plant protein is represented by different coloured boxes using MEME software. The length of the motif is represented by the size of the box, and non-conserved sequences are indicated by grey lines.



**Figure 8.** Diagram of the *Camellia sinensis* WRKY gene organisation. Distribution of exon-intron of *CsWRKY* genes from dissimilar major groups boxes using the GSDS software the blue boxes, yellow boxes, and black lines indicate upstream of initiation codon/downstream of stop codon, CDS, and introns, respectively.

#### 4. CONCLUSION

Studies on Tea WRKYs were carried out on progression, characterization and functional explanation of *CsWRKYs* (20). Our present studies improved the understanding the WRKY genes in Tea and provide a comprehensive awareness into the tea genome. The *in silico* investigation for the WRKY transcription factor in the tea species (*Camellia sinensis*) genome is a crucial step toward identifying the molecular processes driving tea plant growth, advancement, and stress response (21). The evolutionary categorization of WRKY genes alongside Arabidopsis WRKY genes gives a foundation for studying the functional variety of WRKY genes in tea. In the current study, 92 *CsWRKY* genes were acknowledged from the complete genome of Tea plant. Phylogenetic relationships, classification, structure of gene, physical and chemical properties of proteins, chromosomal location, conserved motif of proteins and cis-acting element composition, were thoroughly investigated (22). Considering together, our outcomes will also offer a more widespread understanding on *CsWRKY* gene family and also contributes to screen suitable gene candidates for additional study on functional description of WRKY genes under growth, developmental and stress environments (1).

#### Acknowledgement

This study was partially supported by the Guru Nanak Institute of Pharmaceutical Science and Technology in Kolkata. M.N., M.D., and A.N. acknowledge the Director and Principal, GNIPST, Kolkata for lab facilities. M.N., M.D. received an Institutional fellowship from GNIPST, Kolkata.

**Author contributions**

M.N and A.N conceived the study and designed the approach for execution. M.N and A.N. accomplished the analyses and construed the data. M.N and M.D prepared the figures and tables. A.N. transcribed the main manuscript text with M.N. and all the authors finally reviewed this.

**Competing interests**

The author(s) state no competing interests.

**REFERENCES:**

- Govardhana M, Kumudini BS. In-silico analysis of cucumber (*Cucumis sativus* L.) Genome for WRKY transcription factors and cis-acting elements. *Comput Biol Chem.* 2020 Apr;85:107212.
- Rushton PJ, Somssich IE, Ringler P, Shen QJ. WRKY transcription factors. *Trends Plant Sci.* 2010 May;15(5):247–58.
- Zhao H, Mallano AI, Li F, Li P, Wu Q, Wang Y, et al. Characterization of *CsWRKY29* and *CsWRKY37* transcription factors and their functional roles in cold tolerance of tea plant. *Beverage Plant Res.* 2022;2(1):1–13.
- Bordoni A, Hrelia S, Angeloni C, Giordano E, Guarnieri C, Caldarera CM, et al. Green tea protection of hypoxia/reoxygenation injury in cultured cardiac cells. *J Nutr Biochem.* 2002 Feb;13(2):103–11.
- Levites Y, Amit T, Youdim MBH, Mandel S. Involvement of Protein Kinase C Activation and Cell Survival/ Cell Cycle Genes in Green Tea Polyphenol (–)-Epigallocatechin 3-Gallate Neuroprotective Action. *J Biol Chem.* 2002 Aug;277(34):30574–80.
- Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, et al. Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci.* 2002 Apr 16;99(8):5261–6.
- Lambert DM, García-Dastugue SJ, Croxton KL. AN EVALUATION OF PROCESS-ORIENTED SUPPLY CHAIN MANAGEMENT FRAMEWORKS. *J Bus Logist.* 2005 Mar;26(1):25–51.
- Nan H, Li W, Lin Y, Gao L. Genome-Wide Analysis of WRKY Genes and Their Response to Salt Stress in the Wild Progenitor of Asian Cultivated Rice, *Oryza rufipogon*. *Front Genet.* 2020 May 8;11.
- Fan F-Y, Sang L-X, Jiang M. Catechins and Their Therapeutic Benefits to Inflammatory Bowel Disease. *Molecules.* 2017 Mar 19;22(3):484.
- Elbling L, Weiss R-M, Teufelhofer O, Uhl M, Knasmueller S, Schulte-Hermann R, et al. Green tea extract and (–)-epigallocatechin-3-gallate, the major tea catechin, exert oxidant but lack antioxidant activities. *FASEB J.* 2005 May 28;19(7):1–26.
- Li H, Ma Y, Liu W, Liu W. Soil Changes Induced by Rubber and Tea Plantation Establishment: Comparison with Tropical Rain Forest Soil in Xishuangbanna, SW China. *Environ Manage* [Internet]. 2012 Nov 18;50(5):837–48. Available from: <http://link.springer.com/10.1007/s00267-012-9942-2>
- Larson JE, Funk JL. Regeneration: an overlooked aspect of trait-based plant community assembly models. Whitney K, editor. *J Ecol* [Internet]. 2016 Sep 23;104(5):1284–98. Available from: <https://onlinelibrary.wiley.com/doi/10.1111/1365-2745.12613>
- Jiang J, Ma S, Ye N, Jiang M, Cao J, Zhang J. WRKY transcription factors in plant responses to stresses. *J Integr Plant Biol* [Internet]. 2017 Feb;59(2):86–101. Available from: <https://onlinelibrary.wiley.com/doi/10.1111/jipb.12513>
- Zhang R, Ma Y, Hu X, Chen Y, He X, Wang P, et al. TeaCoN: a database of gene co-expression network for tea plant (*Camellia sinensis*). *BMC Genomics.* 2020 Dec 3;21(1):461.
- Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* [Internet]. 2016 Jul 1;33(7):1870–4. Available from: <https://academic.oup.com/mbe/article/33/7/1870/2579089>
- Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* [Internet]. 2015 Apr 15;31(8):1296–7. Available from: <https://academic.oup.com/bioinformatics/article/31/8/1296/213025>
- Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* [Internet]. 2016 Jan 4;44(D1):D457–62. Available from: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkv1070>
- Wu Z-J, Li X-H, Liu Z-W, Li H, Wang Y-X, Zhuang J. Transcriptome-wide identification of *Camellia sinensis* WRKY transcription factors in response to temperature stress. *Mol Genet Genomics* [Internet]. 2016 Feb 26;291(1):255–69. Available from: <http://link.springer.com/10.1007/s00438-015-1107-6>
- Morrison DK. MAP Kinase Pathways. *Cold Spring Harb Perspect Biol* [Internet]. 2012 Nov 1;4(11):a011254–a011254. Available from: <http://cshperspectives.cshlp.org/lookup/doi/10.1101/cshperspect.a011254>
- Ling J, Jiang W, Zhang Y, Yu H, Mao Z, Gu X, et al. Genome-wide analysis of WRKY gene family in *Cucumis sativus*. *BMC Genomics* [Internet]. 2011 Dec 28;12(1):471. Available from: <https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-12-471>
- Yu Y, Xing Y, Liu F, Zhang X, Li X, Zhang J, et al. The Laccase Gene Family Mediate Multi-Perspective Trade-Offs during Tea Plant (*Camellia sinensis*) Development and Defense Processes. *Int J Mol Sci* [Internet]. 2021 Nov 21;22(22):12554. Available from: <https://www.mdpi.com/1422-0067/22/22/12554>
- Xi D, Yin T, Han P, Yang X, Zhang M, Du C, et al. Genome-Wide Identification of Sweet Orange WRKY Transcription Factors and Analysis of Their Expression in Response to Infection by *Penicillium digitatum*. *Curr Issues Mol Biol* [Internet]. 2023 Feb 3;45(2):1250–71. Available from: <https://www.mdpi.com/1467-3045/45/2/82>