Pak. J. Bot., 54(3): 887-896, 2022. DOI: http://dx.doi.org/10.30848/PJB2022-3(1)

GENOMIC ANALYSIS OF C3H ZINC FINGER PROTEINS FAMILY IN *BRASSICA RAPA*

FAIZA MIR[1], SAJIDA BIBI[1], NOREEN ASIM[1] AND ASAD JAN[1*]

[1]*Institute of Biotechnology and Genetic Engineering (IBGE), The University of Agriculture Peshawar, Pakistan*
**Corresponding author's email: janasad@aup.edu.pk*

Abstract

The *C3H* zinc finger proteins play critical roles in a number of biological processes both in animals and plants. These transcription factors are characterized by the presence of a unique motif with 3 cysteine and 1 histidine amino acids. In this study, a wide spread computational investigations of *Brassica rapa* C3H zinc finger proteins (BraC3H) were carried out by describing their phylogenetic analysis, motif distribution, chromosomal locations and duplication events. A total of 106 BraC3H proteins were identified in *Brassica rapa* genome. On the basis of phylogenetic tree, the BraC3H transcription factors were characterized into 17 sub families. All the genes were distributed on the *Brassica rapa* 10 chromosomes with ten segmental duplication events and one tandem duplication event. The non-synonymous to synonymous mutations ratio showed that *BraC3H* gene family had undergone strong negative or purifying selection pressure. Expression patterns of the selected salt and drought stress inducible genes *BraC3H02, BraC3H04* and *BraC3H06* were checked using semi quantitative reverse transcriptase PCR. Expression analysis of the BraC3H genes indicates that they have wide range of expression patterns under salt and drought treatments, signifying their varied functions. The expression and genome analysis of these *BraC3H* genes provide foundation for further functional dissection in *Brassica rapa.*

Key words: Zinc finger protein, Phylogenetic analysis, Duplication events, Abiotic stresses, RT-PCR.

Introduction

Plant developmental processes and crop biomass production is highly effected by abiotic stresses resulting in huge economical losses (Boyer *et al*., 2005). In natural environment plants were drastically affected by different type of Stresses such as fungal attack, high level of salt concentration and drought conditions which finally leads to very low crop yield (Xiong & Yang, 2003). A set of different types of abiotic and biotic stresses acted upon the plants in their natural habitat (Dixit & Dhankher, 2011). The plant reaction to these stresses were dependent upon response produce by the affected plant tissues. Those genes which produce regulatory proteins are stimulated by the abiotic stresses. These regulatory proteins regulate the signal transduction pathways, control gene expression and therefore were involved in stress response (Pereira, 2016). A number of transcriptional factors are reported to play role in transcription process of several regulatory mechanisms during plants stress tolerance (Zhao *et al*., 2016). A huge amount of transcription factors are required to mediate the complex processes in living organisms (Todeschini *et al*., 2014).

Zinc finger proteins are ubiquitous in eukaryotic genomes. Some of the major zinc finger protein families are C3H, C2H2, C2HC5 and C4HC3. One of the most important zinc finger protein family is characterized with motifs comprising of 3 cysteine and 1 histidine residue (C3H-type) which coordinate zinc ions to attain its functional peptide structure (Lin *et al*., 2005). According to the previous reports the C3H zinc finger transcription factors were expected to be RNA-binding proteins involved in metabolism of RNA (Bogamuwa & Jang, 2014; Seo & Choi, 2015). The C3H zinc finger proteins are divided into large number of classes on the basis of spacing and number between histidine and cysteine residues (Peng *et al*., 2012). A variety of C3H proteins are reported to mediate important functions in plant development and abiotic stress responses (Khatun *et al*., 2017). For example, the first C3H gene HUA1 identified *in Arabidopsis thaliana* was reported to play roles in floral morphogenesis (Li *et al*., 2001), *Arabidopsis thaliana* ATSZF2 gene in soybean conferred salt tolerance (Kim *et al*., 2017). Similarly rice gene *OsTZF1* was involved in abiotic stress tolerance (Jan *et al*., 2013). The important biological functions and ubiquitous nature of the C3H proteins encouraged its genome wide analysis in plants like Rice *(o Oryza sativa)* (Wang *et al*., 2008), Banana *(Musa acuminate*) *(*Mazumdar *et al*., 2017), Tomato (*Solanum lycopersicum*) (Xu, 2014) and Alfalfa (*Medicago sativa)* (Zhang *et al*., 2013). However, despite being an important crop grown for oil production worldwide, no work has been done on the genomic characterization of C3H family in *Brassica rapa.* In the present research *C3H* genes from the entire genome of *Brassica rapa* were identified and then characterised into 17 subclasses based on phylogenetic analysis.

Materials and Methods

characterization of *BraC3H* genes: The *BraC3H* zinc finger sequences were retrieved from Brassica database (http://brassicadb.org/brad/) (Cheng *et al*., 2011). All the *BraC3H* genes were translated into protein sequences using Expasy Translate online tool (https://www.expasy.org/) (Gasteiger *et al*., 2003). Basic Local Alignment Search Tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi) with algorithm BLASTP was used to align the protein sequences. The physiochemical properties of the entire C3H protein sequences were analysed through ProtParam (https://web.expasy.org/protparam). The BraC3H protein sequences were subjected to Motif scan in order to identify various motifs and the outputs were then compared with MEME results (Bailey *et al*., 2009).

Phylogenetic analysis: Multiple sequence alignment of full length 146 *Brassica rapa* and *Arabidopsis thaliana* C3H protein sequences were conducted through Muscle. The aligned sequences were further used to mediate the Phylogenetic analysis through MEGA 7.0 software (https://www.megasoftware.net/) (Kumar *et al*., 2016).

Chromosomal localization and duplication events: *Brassica rapa* C3H genes were mapped on the 10 *Barssica rapa* chromosomes according to their locations in the BRAD data base (Wang *et al*., 2011). All the chromosomes were drawn through Mapinspect tool and the genes were manually placed on the *Brassica rapa* chromosomes. Paralogous gene pairs were identified according to the previous research studies (Wu *et al*., 2016). The synonymous and non-synonymous mutations rates in the paralogous gene pairs were calculated using the Dnasp software www.ub.edu/dnasp/Interface.html (Librado & Rozas, 2009). Selection pressure in the *BraC3H* genes were estimated using the Ka/Ks ratios.

Interaction analysis: Interaction network among the BraC3H genes were created using string v10.5 database (https://string-db.org/) (Szklarczyk *et al*., 2015). The orthologous Arabidopsis genes were identified through BLASTp.

Promoter analysis: Promoter regions of the BraC3H02, 04 and 06 were retrieved from Brassica Data Base (BRAD). To find the *cis*-regulatory elements, the 1000bp promoter regions were uploaded to PlantCARE available online at (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Rombauts *et al*., 1999).

Plant materials and growth conditions: *Brassica rapa* seeds were sterilized and grown in green house providing optimal growth conditions. A 35-day old *Brassica rapa* young plants were treated with abiotic stresses. The salinity stress the was applied by watering the seedlings with 250 mM NaCl solution and the drought was applied by withholding water. The untreated plants were used as control. Samples collection were carried out at 0, 1and 2 days after the treatment. For RNA extraction the samples were stored at -80^{o}C after freezing it in liquid nitrogen.

Table 1. Primer sequences of the selected *BraC3H* genes.

Name	Sequence
BraC3H02- L	5'-CCG CCG CGA TTT ATG GAT AG-3'
BraC3H02- R	5'-CCT CCT GCT AAG AGA CGG AA-3'
BraC3H04- L	5'-TTG CAC TTT CGC TCA TGG AG-3'
BraC3H04- R	5'-TCA AAT CTG CGG TTC TCC CT-3'
BraC3H 06- L	5'-GCA CTT CTG GCT GTC CTT TT-3'
BraC3H 06- R	5'-AGA CTC TCC TCC TCT CCC TG-3'
BraC3H Actin- L	5'-GGA AGG ATC TGT ACG GTA AC-3'
BraC3H Actin-R	5'-TGT GAA CGA TTC CTG CAC CT-3'

RNA extraction and expression analysis: Trizol reagent method was used to extract RNA from all the samples. The cDNA synthesis was carried out from the purified RNA by using the Superscript Reverse Transcriptase reagent kit (invitrogen). Gene specific primers were designed for *BraC3H02, BraC3H04* and *BraC3H06* through primer 3 input software (Table 1). *BraACTIN* was used as a loading control. Semi quantitative reverse transcriptase PCR was conducted for expression analysis of the targeted genes. The following conditions was used to conduct the PCR cDNA (2 uL), Forward Primer (4 uL), Reverse Primer (4 uL), PCR Master Mix (50 uL), and PCR H_2O (40uL) with following conditions 95^{o}C for 2min, 95^{o}C for 20sec, 56^{o}C for 20 sec, and finally extension at 72^{o}C for 30sec.

Results

Identification of *C3H Brassica rapa* genes: All the gene sequences of C3H transcription factors were retrieved from the Brassica database. Expasy Translate tool was used to translate the entire *C3H* gene sequences into protein sequences. Physicochemical properties of C3H proteins were investigated by using Protparam tool. Chromosome numbers, Molecular weight, theoretical isoelectric point (PI) and Amino acid size of BraC3H proteins are provided in (Table 2). The BraC3H proteins length ranged from 366 amino acids of BraC3H88 to 3006 amino acids of BraC3H51. Their molecular weight ranged from 30570.4 Da (BraC3H88) to 424944.3 Da (BraC3H77). Theoretical pI of BraC3H proteins showed that they are acidic in nature having pI value from 4.75 to 5.29. Interestingly, the number of C3H proteins and number of different CCCH motif variants out number *Arabidopsis thaliana,* chick pea, rice and tomato (Fig. 1).

Phylogenetic tree: The phylogenetic analysis was conducted using the protein sequences of BraC3H transcription factors (Fig. 2). Based on the tree topology and homology with *Arabidopsis thaliana* C3H proteins, BraC3H proteins were divided into 17 subgroups designated from A-Q (Fig. 2). Group G possesses the maximum number of BraC3H proteins (10) and accounts for 10.6% of the BraC3H zinc finger protein family. Group E and K contained 9 members each, whereas Group M contained the least number of BraC3H protein members (2). The C3H zinc finger proteins are known to regulate the post transcriptional modifications and protein-protein interactions due to the presence of other functional domains (Wang *et al*., 2008).

Motifs: To investigate the total number of motifs in *Brassica rapa* C3H proteins, motif scan and pfam software was used. In this research study, a total of 178 BraC3H motifs were recognised (Fig. 1). Fifty-four members contained C3H motif designated as 1, 17 members possessed motif number 4, 13 members contained motif number 2 and 12 BraC3H members contained motif number 3 (Table 3). The comparative analysis of C3H proteins among *Brassica rapa, Arabidopsis thaliana,* tomato chick pea and rice also showed that among all these plants *Brassica rapa* contained the largest number of C3H proteins (Fig. 1). The two most commonly observed C3H motifs were the CX8-CX5-CX3-H and CX7-CX5-CX3-H; the results were similar to the previous research findings (Pradhan *et al*., 2017). We also identified two unique C3H motifs in the BraC3H proteins, including CX17-CX4-CX3-H and CX20-CX5-CX5-H (Table 3). Multiple sequence alignment of the selected BraC3H and AtC3H proteins showed that 18 members contained highly conserved CX8-CX5-CX2-H motif, 3 members possessed CX7-CX12-CX16 motif. Additionally, BraC3H100 contained a CCH in which the Cysteine amino acid was replaced by Glycine (CX12-CX16-H) motif (Fig. 3).

Table 2. Characterization of BraC3H proteins.

Gene IDs	Chromosome	Start (bp)	End (bp)	Gene name	Size	M. wt	Pi
Bra011087	A01	3958590	3959675	*BraC3H01*	1086	89760.9	5.02
Bra013400	A01	5447728	5453553	*BraC3H 02*	4713	391587.1	4.76
Bra013894	A01	8343426	8345497	*BraC3H 03*	1278	105888.3	5.02
Bra023860	A01	19662751	19665912	*BraC3H 04*	1203	96233	5.09
Bra038204	A01	20827687	20828851	*BraC3H 05*	1080	87380.4	5.09
Bra038697	A01	24401217	24403090	*BraC3H 06*	738	58560.4	5.16
Bra040048	A01	25315441	25317159	*BraC3H 07*	930	77008.6	5.08
Bra040220	A01	25737569	25739354	*BraC3H 08*	1308	110092	4.99
Bra021409	A01	26466526	26468044	*BraC3H 09*	1062	88642.2	5.04
Bra028719	A02	1005050	1005601	*BraC3H 10*	471	38547.8	5.23
Bra023379	A02	2058707	2064934	*BraC3H 11*	2736	227464.4	4.85
Bra023683	A02	3482973	3485055	*BraC3H 12*	1302	109940.4	4.99
Bra035636	A02	6047025	6049505	*BraC3H 13*	1929	158219.7	4.95
Bra026613	A02	20074117	20076232	*BraC3H 14*	1872	148828.3	4.98
Bra006465	A03	3604453	3606720	*BraC3H 15*	1269	106289.1	4.99
Bra029149	A03	6580799	6582838	*BraC3H 16*	1317	107504.8	5.02
Bra022919	A03	7764531	7766736	*BraC3H 17*	1392	115337.9	5
Bra022957	A03	7946675	7949427	*BraC3H 18*	1680	137103	4.98
Bra022958	A03	7950789	7955583	*BraC3H 19*	2454	201637.4	4.89
Bra000466	A03	11257522	11259236	*BraC3H 20*	1215	103845.3	4.98
Bra001066	A03	14600683	14602197	*BraC3H 21*	1077	90208.9	5.04
Bra001196	A03	15268764	15271065	*BraC3H 22*	1353	112651.7	4.99
Bra001285	A03	15699786	15701218	*BraC3H 23*	864	70681.1	5.09
Bra001436	A03	16389573	16391292	*BraC3H 24*	708	56925.3	5.15
Bra001675	A03	17819864	17827502	*BraC3H 25*	1329	113941.3	4.99
Bra001745	A03	18238787	18239935	*BraC3H 26*	1068	85796	5.11
Bra013181	A03	20039508	20041731	*BraC3H 27*	1602	133913	4.96
Bra012827	A03	22083079	22085004	*BraC3H 28*	1506	124863.8	4.99
Bra012540	A03	23721983	23726550	*BraC3H 29*	3831	316877.9	4.81
Bra024157	A03	27297460	27298509	*BraC3H 30*	1050	86869.8	5.03
Bra033465	A04	4689720	4691813	*BraC3H 31*	1605	132595.1	4.98
Bra025586	A04	8049537	8051307	*BraC3H 32*	1290	105276.6	5.05
Bra039988	A04	13335254	13336950	*BraC3H 33*	1362	112894.2	4.97
Bra021817	A04	14773130	14775291	*BraC3H 34*	1227	101911.7	5.02
Bra016903	A04	17788426	17794226	*BraC3H 35*	2178	179091.2	4.89
Bra004427	A05	76985	78964	*BraC3H 36*	1365	117208.8	4.95
Bra004438	A05	110827	116185	*BraC3H 37*	3027	254000.4	4.83
Bra004982	A05	2795637	2797307	*BraC3H 38*	1671	136612.8	4.98
Bra005342	A05	4934444	4935400	*BraC3H 39*	810	66388.8	5.09
Bra005468	A05	5660684	5663241	*BraC3H 40*	1875	152825.1	4.97
Bra005543	A05	6132348	6134708	*BraC3H 41*	1437	120148.5	4.98
Bra018446	A05	8261914	8265777	*BraC3H 42*	2244	175516.5	4.98
Bra010155	A05	15203622	15204811	*BraC3H 43*	1008	81521.1	5.06
Bra034744	A05	21971375	21974953	*BraC3H 44*	1602	134938.6	4.92
Bra034776	A05	22095044	22096969	*BraC3H 45*	735	58876.2	5.15
Bra020750	A05	23954969	23956747	*BraC3H 46*	1539	121496.6	5.04
Bra040604	A05	24910338	24912132	*BraC3H47*	1179	98630.3	5.02
Bra018759	A06	2160278	2162251	*BraC3H 48*	1974	159026.8	4.96
Bra019944	A06	3583199	3586420	*BraC3H 49*	1137	96214.2	5
Bra025776	A06	7627423	7629068	*BraC3H 50*	1206	100063.2	5.01
Bra017929	A06	8622748	8631800	*BraC3H 51*	3006	247810.4	4.84
Bra018059	A06	9603459	9605611	*BraC3H 52*	1317	109806.5	5
Bra018163	A06	10452543	10453673	*BraC3H 53*	882	69999.4	5.13

Table 2. (Cont'd).

Gene IDs	Chromosome	Start (bp)	End (bp)	Gene name	Size	M. wt	Pi
Bra038619	A06	14824780	14826984	*BraC3H 54*	1251	102480	5.02
Bra025134	A06	22223896	22225238	*BraC3H 55*	1227	99705	5.08
Bra033731	A06	24304000	24305136	*BraC3H 56*	1467	116447.4	5.04
Bra036459	A07	276380	278230	*BraC3H 57*	1110	87147.9	5.11
Bra039022	A07	539319	540392	*BraC3H 58*	1074	88673.2	5.04
Bra015163	A07	5123703	5125645	*BraC3H 59*	1608	134651.6	4.97
Bra015109	A07	5640630	5642147	*BraC3H 60*	489	40690.8	5.19
Bra014881	A07	7844293	7848696	*BraC3H 61*	1743	141065	4.95
Bra030151	A07	8429784	8432384	*BraC3H 62*	1506	115959	5.07
Bra030150	A07	8435435	8436641	*BraC3H 63*	618	48737.5	5.23
Bra030149	A07	8440173	8442439	*BraC3H 64*	1029	81201	5.1
Bra030148	A07	8453181	8454936	*BraC3H 65*	765	60008.9	5.18
Bra030057	A07	9143620	9144703	*BraC3H 66*	771	61310.9	5.13
Bra011947	A07	13272192	13275797	*BraC3H 67*	2358	193282.7	4.89
Bra004045	A07	19785229	19790809	*BraC3H 68*	2163	177957.5	4.92
Bra004189	A07	20611221	20612232	*BraC3H 69*	936	79051.4	5
Bra004288	A07	21079477	21081074	*BraC3H 70*	930	75601.4	5.05
Bra015851	A07	23800220	23802065	*BraC3H 71*	1197	100669.7	4.99
Bra038415	A08	9000835	9002934	*BraC3H 72*	1329	111090.6	4.99
Bra010334	A08	13141837	13142895	*BraC3H 73*	1059	87280.6	5.05
Bra010767	A08	15536953	15541334	*BraC3H74*	1893	152834.3	4.94
Bra010926	A08	16383927	16384787	*BraC3H 75*	861	67803.7	5.12
Bra016385	A08	17408455	17410120	*BraC3H 76*	1122	93339.5	5.02
Bra016416	A08	17578133	17584669	*BraC3H 77*	5160	424944.3	4.75
Bra016491	A08	17928864	17930519	*BraC3H 78*	1194	98609.6	5.02
Bra030679	A08	20068437	20070768	*BraC3H 79*	1443	119644.3	4.97
Bra030542	A08	20629078	20630277	*BraC3H 80*	1200	101174.4	5
Bra036592	A09	2676283	2677244	*BraC3H 81*	366	30570.4	5.24
Bra036689	A09	6034763	6036694	*BraC3H 82*	984	78763	5.12
Bra027459	A09	10934010	10934885	*BraC3H 83*	876	69465.3	5.09
Bra027537	A09	11621071	11622204	*BraC3H 84*	1134	95568.2	5
Bra023260	A09	20190564	20192029	*BraC3H 85*	1086	87497.4	5.06
Bra032327	A09	22601319	22602458	*BraC3H 86*	546	42840.8	5.26
Bra032328	A09	22602478	22602955	*BraC3H 87*	444	35142.4	5.29
Bra032326	A09	22614922	22619970	*BraC3H 88*	1749	134808.3	5.04
Bra006885	A09	27527793	27529786	*BraC3H 89*	1572	130132.7	4.99
Bra007205	A09	29494980	29496617	*BraC3H 90*	1637	133105.4	4.99
Bra007830	A09	32658186	32659960	*BraC3H 91*	1774	109762.7	4.99
Bra031114	A09	34159729	34161926	*BraC3H 92*	1104	86809.5	5.11
Bra032634	A09	38696406	38699660	*BraC3H 93*	726	59911	5.14
Bra015376	A10	1736510	1738599	*BraC3H 94*	1212	101824.9	5.01
Bra015377	A10	1738911	1739909	*BraC3H 95*	681	57017.1	5.13
Bra015264	A10	2332308	2333480	*BraC3H 96*	1173	99391.5	5
Bra033268	A10	3238963	3240160	*BraC3H 97*	1014	81306.7	5.08
Bra028289	A10	4674792	4677289	*BraC3H 98*	1623	131556.4	5
Bra002785	A10	7663378	7666337	*BraC3H 99*	1815	151390.3	4.95
Bra002632	A10	8543428	8545146	*BraC3H 100*	1719	141970.5	4.94
Bra002167	A10	11193687	11195740	*BraC3H 101*	1311	110143.7	4.98
Bra008877	A10	13048154	13050097	*BraC3H 102*	1944	163444	4.91
Bra008887	A10	13097031	13099271	*BraC3H 103*	1839	151988.5	4.95
Bra009257	A10	14627965	14628705	*BraC3H 104*	741	60009.9	5.12
Bra009465	A10	15581281	15587226	*BraC3H 105*	1083	90334.1	5.06
Bra009466	A10	15609616	15611180	*BraC3H106*	1266	101491.6	5.05

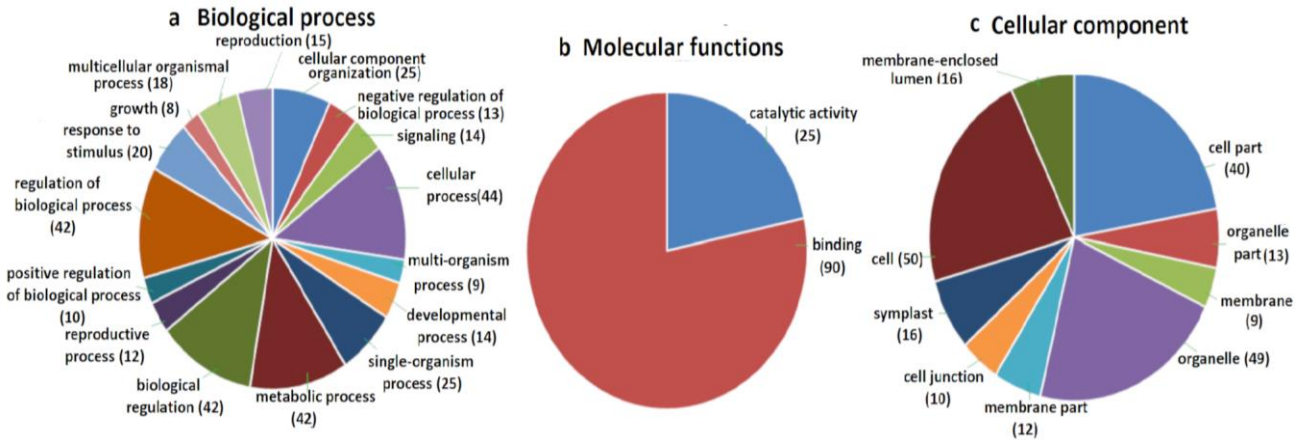

Fig. 1. Comparison of C3H numbers and C3H motifs among *Brassica rapa, Arabidopsis thaliana,* Chick pea, Rice and Tomato.

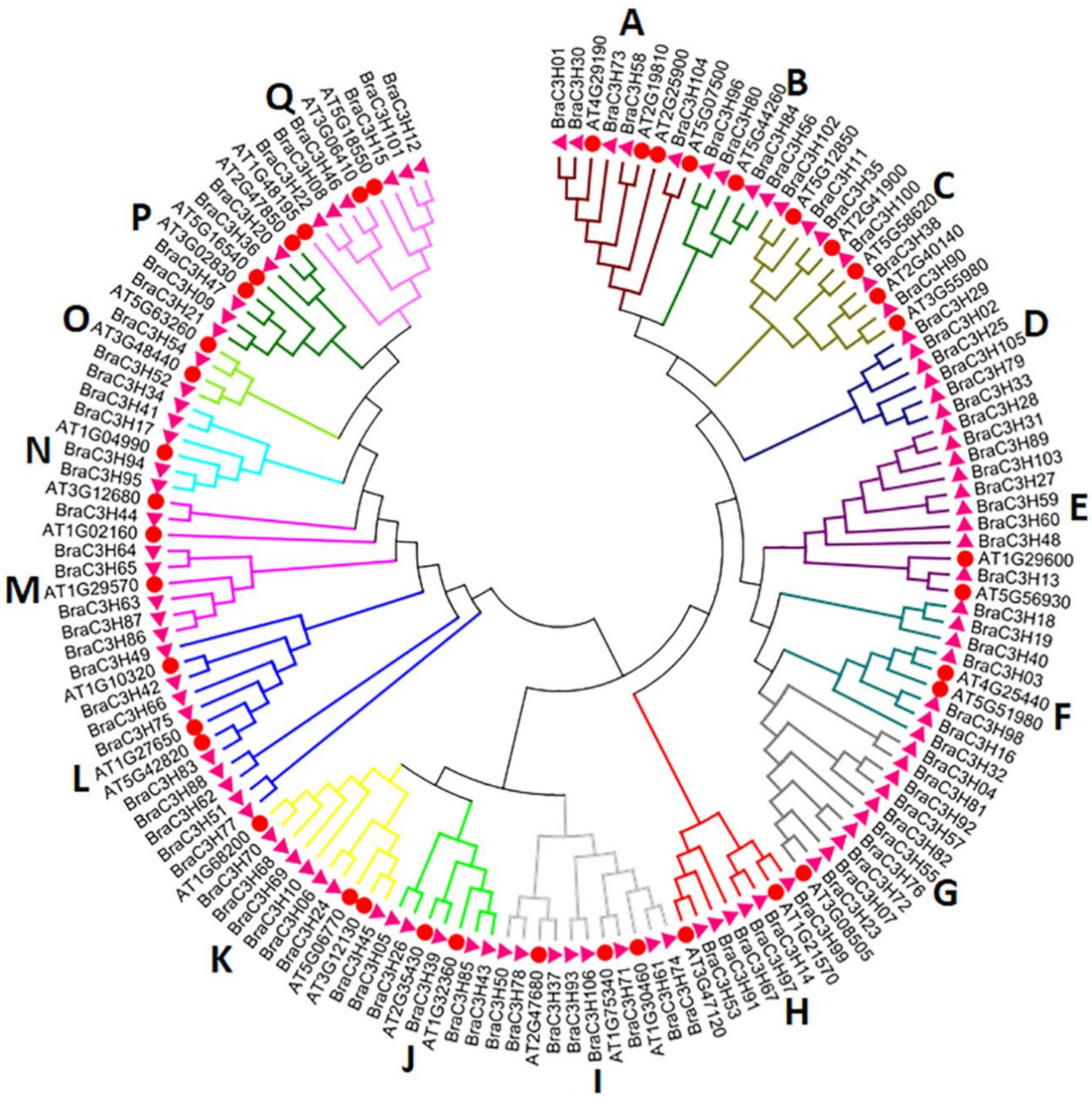

Fig. 2. Phylogenetic tree of C3H proteins among *Brassica rapa and Arabidopsis thaliana.* The phylogenetic analysis was conducted through MEGA using neighbour joining method. The bootstrap replicates were set to 100. The different families are represented by different colours.

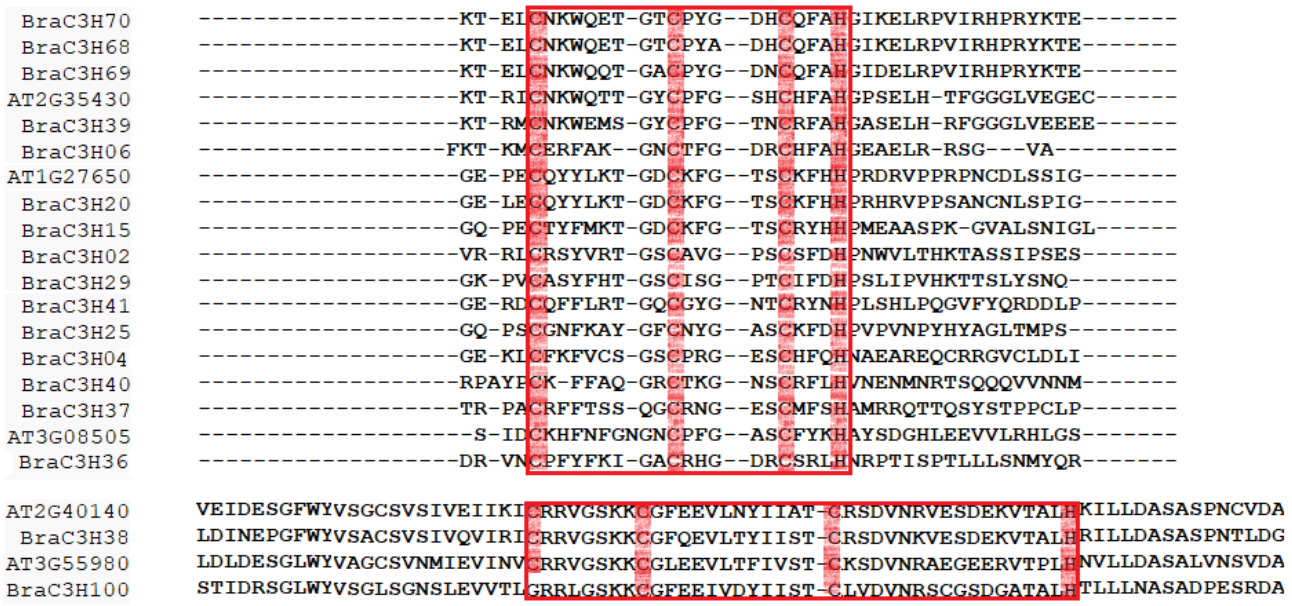

Fig. 3. MSA (Multiple sequence alignment) of some *Brassica rapa* C3H proteins with the *Arabidopsis thaliana* C3H proteins. The conserved cysteine and Histidine residues are highlighted with red colour. The alignment was conducted using Clustal X software.

Table 3. Comparison of *Brassica rapa* C3H motif pattern and numbers with reported C3H motifs in *Brassica rapa*, Tomato, Rice, Arabidopsis and Chickpea.

Motifs	*Brassica rapa*	Tomato	Rice	Arabidopsis	Chickpea
CX5-CX4-CX3-H	10	7	9	12	5
CX7-CX4-CX3-H	0	10	6	5	6
CX7-CX5-CX3-H	46	53	35	42	44
CX8-CX4-CX3-H	0	1	1	2	0
CX8-CX5-CX3-H	103	82	36	44	70
CX9-CX5-CX3-H	0	7	6	4	4
CX10-CX5-CX3-H	2	0	3	1	1
CX12-CX10-CX3-H	2	0	0	0	1
CX11-CX5-CX3-H	1	1	1	1	0
CX17-CX4-CX3-H	10	0	0	0	0
CX17-CX6-CX3-H	2	0	0	0	1
CX20-CX5-CX5-H	1	0	0	0	0

Chromosomal positions and duplication events: In order to identify the chromosomal positions of *BraC3H* genes, they were mapped on the *Brassica rapa* chromosomes. A total of 106 *BraC3H* genes were anchored on the *Brassica rapa* 10 chromosomes (Fig. 4). The highest number of genes (16) were located on chromosome 3 and 7 whereas chromosome 2 and 4 carried the least number of genes (5). Chromosome 1, 6 and 8 contained 9 genes each, whereas chromosome 9 and 10 possessed 13 *BraC3H* genes each. Twelve genes were assigned to chromosome 5. The exact location of each BraC3H gene on the *Brassica rapa* chromosome is listed in Table 1. The *BraC3H* genes are distributed irregularly on the *Brassica rapa* chromosomes. Maximum number of genes were distributed at the top of chromosome 3 whereas the bottom of chromosome 2 did not possess any *BraC3H* gene (Fig. 4).

Majority of the plant gene families expanded through genome duplication events like the NAC and bHLH (Song *et al*., 2014). To find evolutionary mechanisms among the *C3H* genes in *Brassica rapa*, we analysed the tandem and segmental duplication events in this gene family. We observed 11 duplicated gene pairs, out them 10 gene pairs were segmentally duplicated whereas *BraC3H18/BraC3H19* was a tandem event (Fig. 4). We calculated Ks (synonymous), Ka (non synonymous) and the Ka/Ks ratios in order to recognize impact of selective pressure on the paralogous genes pairs (Table 4). The results indicated that the 11 paralogous gene pairs showed <1 Ka/Ks values, which suggest that *BraC3H* gene family were evolved under purifying selection.

Table 4. Details of *BraC3H* duplicated gene pairs and their ka/ks values.

Paralogous gene pairs	Ka	Ks	Ka/Ks
BraC3H06-BraC3H45	0.70704	2.1449	0.3296
BraC3H09-BraC3H21	0.65677	1.5637	0.42
BraC3H18-BraC3H19	0.28248	0.3545	0.7968
BraC3H14-BraC3H21	0.74587	3.9022	0.1911
BraC3H28-BraC3H29	0.72668	2.6031	0.2791
BraC3H31-BraC3H89	0.74582	3.8927	0.1915
BraC3H57-BraC3H94	0.56294	1.0415	0.5405
BraC3H27-BraC3H59	0.72105	2.4409	0.2954
BraC3H82-BraC3H92	0.74315	3.5219	0.211
BraC3H22-BraC3H46	0.73918	3.179	0.2325
BraC3H24-BraC3H42	0.74593	3.9124	0.1906

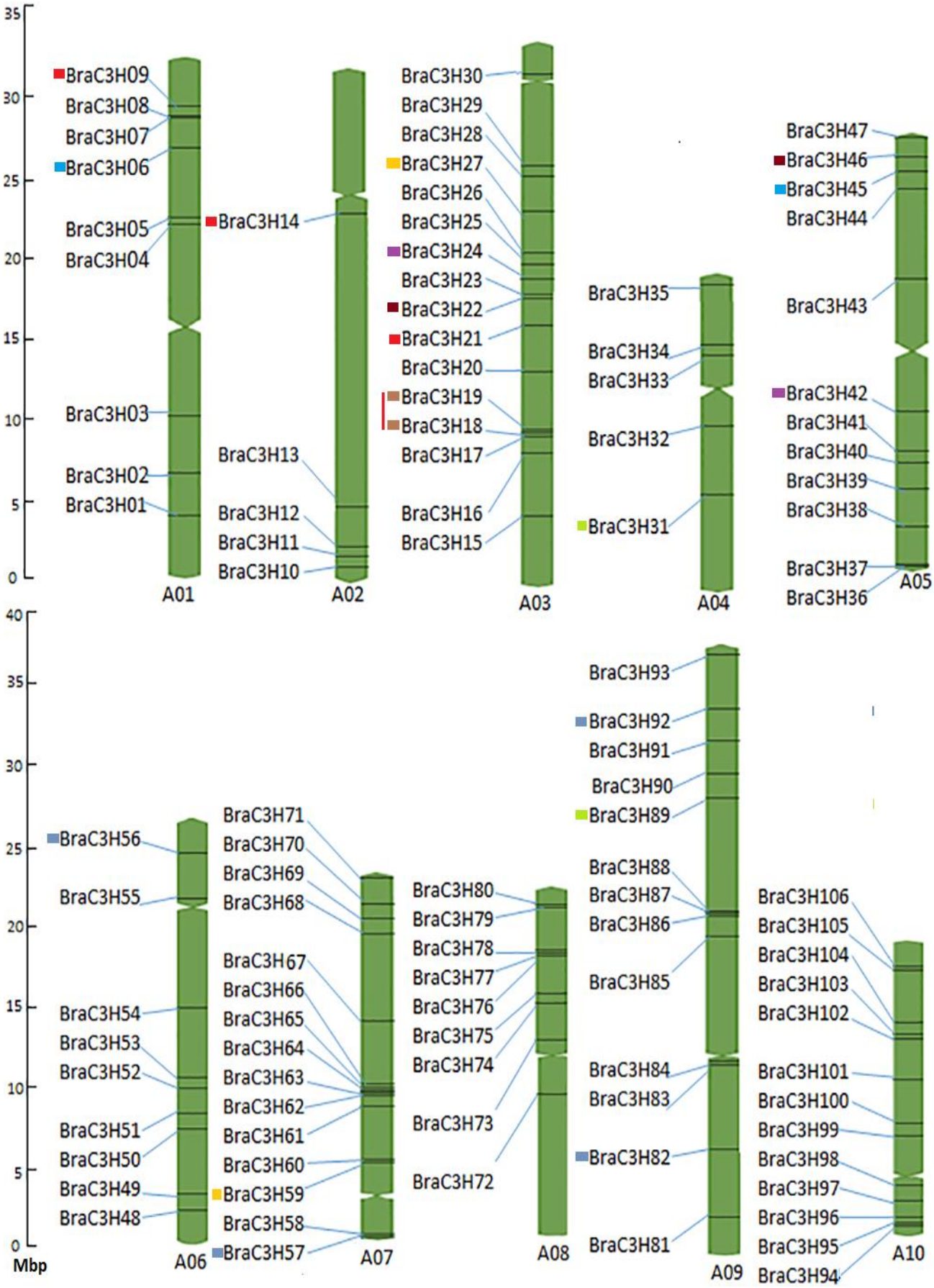

Fig. 4. Chromosomal positions of *C3H* zinc finger genes on *Brassica rapa* 10 chromosomes. The paralogous gene pairs were represented by different symbols. The tandemly duplicated gene pair was joined by red line. The chromosome numbers are mentioned at the bottom on each chromosome.

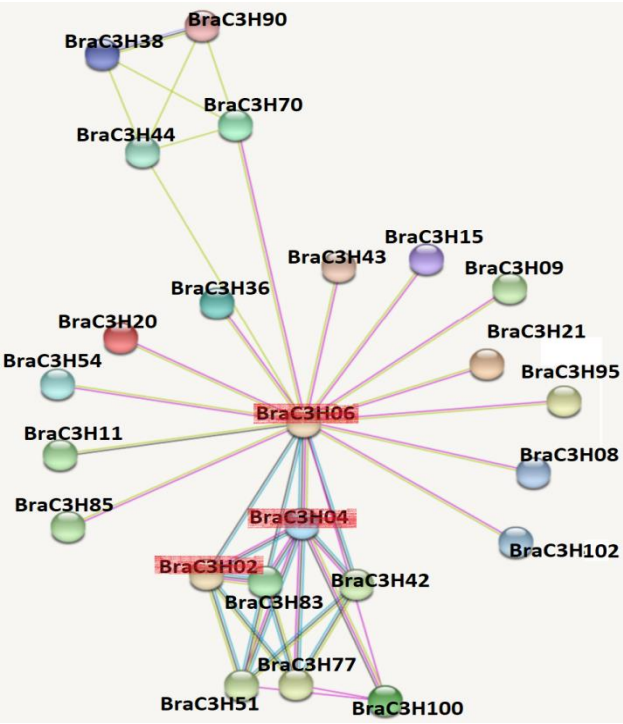

Fig. 5. Protein- protein interaction network among BraC3H proteins according to orthologous Arabidopsis. The nodes represent the BraC3H protein and the edges represent the interaction among proteins. The colour of the edges represents the edge score.

Interaction network: Biological processes like plant defence against stress conditions, signal transduction pathways and plant developmental processes were due to Protein-protein interaction (Zhang *et al*., 2010). The functional analysis and phylogenetic tree revealed that *BraC3H* genes were involved in abiotic stress tolerance. To get insight into the molecular interactions of C3H family in *Brassica rapa*, we conducted the protein – protein interaction network using the string software. Orthologous gene pairs among the *Brassica rapa* and *Arabidopsis thaliana* were identified according to the previous research studies (Song *et al*., 2014). The results showed that 28 *BraC3H* genes were involved in the protein network analysis (Fig. 5). For instance, the *BraC3H02*, *BraC3H04* and *BraC3H06* were identified to interact with large number of proteins, implying that they play important role in gene regulatory pathways. On the basis of phylogenetic and protein-protein interaction network analysis, three genes including BraC3H02, BraC3H04 and BraC3H06 were selected for further study.

Promoter analysis: The 1000bp putative promoter regions of the selected *BraC3H* genes were investigated for diffierent *cis*-regulatory elements. In these promoters, at least 6 type of *cis*- regulatory elements were recognised comprising of Abscisic Acid Responsive Element (involved in abiotic stress responses), Auxin Responsive Element (ARE), Gibberellin Responsive Element (GARE), Myb Binding Sites (MBS), TC- rich repeats (involved in drought stresses) and TGACG. The promoter regions of *BraC3H02*, *BraC3H04* and *BraC3H06* carried TGACG and TC- rich repeats (Fig. 6). Similarly, the promoter regions of *BraC3H02* and *BraC3H06* contained ABRE *cis*-element, whereas the promoter region of *BraC3H04* possesses ARE and GARE *cis*- regulatory elements. The MBS *cis*-element was detected in the *BraC3H04* and *BraC3H06*. Among all the genes *BraC3H04* contained the maximum number of *cis*-elements (11) followed by *BraC3H02* and *BraC3H06.*

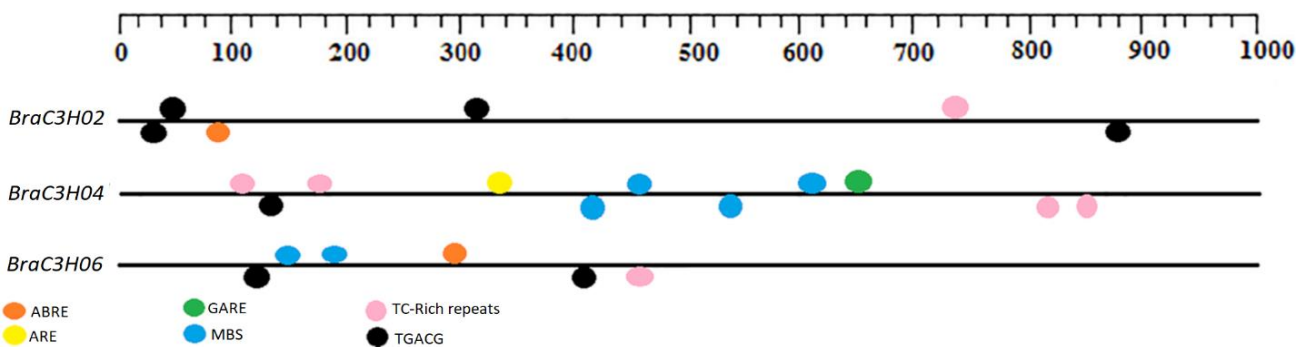

Fig. 6. *In silico* study of the *Cis*- regulatory elements present in 1kb promoter sequence upstream of the start codon in the stress responsive genes. The analysis was performed using plantCARE database. The binding sites are represented by coloured circles.

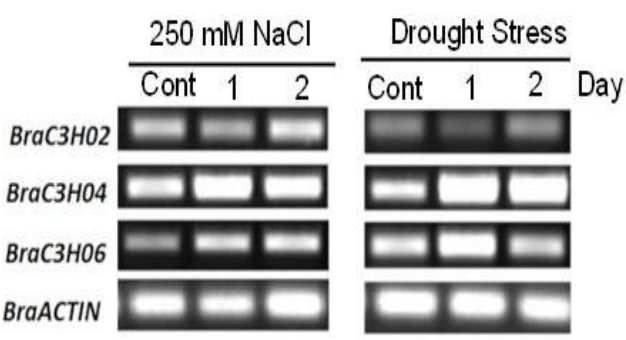

Fig. 7. Expression analysis of selected salt and drought stress responsive *BraC3H* genes. *BraActin* was used as a loading control.

Expression analysis: For the expression analysis of *BraC3H02, BraC3H04* and *BraC3H06* genes Semi quantitative reverse transcriptase PCR was conducted, using cDNA derived from salt and drought treated *Brassica rapa* plants (Fig. 7). The results showed that the *BraC3H04* was highly expressed under salt stress conditions. The *BraC3H02* and *BraC3H06* were expressed strongly at 2-day after salt treatment. Under drought stress the *BraC3H02* was weekly expressed whereas the *BraC3H04* was highly expressed. The *BraC3H06* was upregulated after 1 and slightly decreased 2 day drought treatment (Fig. 7).

Discussion

Brassica rapa is cultivated as oil and vegetable crop worldwide (Rakow, 2004). The growth rate of Brassica rapa plants are highly effected by abiotic environmental factors (Cheng *et al*., 2015). Abiotic stresses stimulate regulatory proteins which in turn control the expression of genes under stress conditions (Ghosh & Xu, 2014). The C3H motif was found in a large number of zinc finger proteins (Wang *et al*., 2015). Recent research studies showed that C3H zinc finger transcription factors were involved in RNA processing during post transcription modifications (Mazumdar *et al*., 2017). In the *Brassica rapa* genome 106 *C3H* genes were identified (Table 2). It was observed that the number of C3H transcription factors in any plant species is independent of its genome size. The total number of C3H genes identified *Brassica rapa* were higher than rice (69) (Wang *et al*., 2008) and chickpea (58) (Pradhan *et al*., 2017), though both rice and chickpea have larger genome size than *Brassica rapa*. Conserved motifs analysis revealed that *Brassica rapa* has higher number of C3H motifs as compared to other plants (Fig. 1). Among the 178 C3H motifs identified in *Brassica rapa*, the two most abundant motifs identified were CX8-CX5-CX3-H and CX7-CX5-CX3-H (Table 3a & Fig. 1). Plants adaptation to various environmental conditions resulted differences in the C3H proteins among different plant species. (Khatun *et al*., 2017). The unique motifs (CX17-CX4-CX3-H and CX20-CX5-CX5-H) identified in the *Brassica rapa* may be involved in specific biological activities. The CX8-CX4-CX3-H motif was found in the C3H zinc finger proteins of Arabidopsis thaliana and rice but was found missing in *Brassica rapa,* implying that this C3H motif was lost in the evolutionary process.

The phylogenetic analysis was carried using *Brassica rapa* and *Arabidopsis thaliana* C3H proteins in order to investigate their evolutionary relationships. The BraC3H proteins were divided into 17 subclasses on the basis of tree topology (Fig. 2). Members under the same clade display similar characteristics. Putative homologs of BraC3H in *Arabidopsis thaliana* were identified in the comparative phylogenetic analysis. For example BraC3H38 clustered closely with the ATSZF1 (AT2G40140, group C), was reported to be involved in salinity tolerance (Sun *et al*., 2007), similarly BraC3H61 was a close homolog of (AT1G30460, group P) which was involved in endonuclease activity during RNA biogenesis (Addepalli & Hunt, 2007).

The inheritance of beneficial characteristics in order to ensure the survival of species is a result of many evolutionary forces (Jin *et al*., 2017). Both the purifying and diversifying selection pressure were involved in gene families evolution. Positive (diversifying) selection involves the fixation of beneficial mutations, whereas negative (purifying) selection involves the elimination of deleterious mutations (Wang *et al*., 2008). The Ks and Ka mutations analysis were used to investigation the evolutionary mechanisms in plant gene families. The Ka/Ks values >1 leads to positive selection, Ka/Ks <1 is considered as purifying selection pressure and Ka/Ks=1 involves neutral selection. The Ka/Ks analysis of the *BraC3H* genes family showed that the duplicated gene pairs experienced purifying selection pressure, implying that *C3H* genes conserved their biological functions through the removal of deleterious mutations in *Brassica rapa*. Majority of the paralogous gene pairs showed remarkable similarities in the protein structures, however some of the duplicated pairs showed variable structure. The paralogous gene pairs having similar proteins structures perform similar functions. For example, BraC3H82/92 with similar conserved motifs must perform similar functions, whereas the BraC3H56/83 paralogous pairs will perform different functions due to the presence of variable motifs.

Transcription factors mediated the regulation of genes under stress conditions through several mechanisms that depends on the coordinated network of *cis*-regulatory elements present in the upstream promoter sequences of biotic and abiotic stress inducible genes (Khatun *et al*., 2017). In the putative 1kb promoter regions upstream of start codon 6 *cis*-regulatory elements GARE, MBS, TC, TGACG, ABRE and ARE were identified in BraC3H02, BraC3H04 and BraC3H06 (Fig. 6). The consensus sequences of the *cis* elements are TCTGTTG, CGGTCA, and CGTACGTGCA, respectively. All the identified *cis*-elements in these promoter regions are involved in plant growth and stress tolerance. The *Cis*-regulatory elements were analysed in the upstream regions of the *Brassica rapa C3H* genes indicated that high number of *cis* elements were found in a range of 100bp to 640bp upstream translation start codon.

Transcription factors containing DNA binding motifs (like bHLH and zinc finger motif) were induced by abiotic stress signals under harsh environmental conditions (Kazuo & Yamaguchi, 2000). The C3H proteins contained a DNA and RNA binding motif, implying that they mediate adoptive responses to various abiotic stresses (Bray, 1997). The *C3H* genes from chickpea, Arabidopsis and rice were recently reported to mediate several important functions in abiotic stress resistance including cold, salt, drought (Wang *et al*., 2008). Semi quantitative reverse transcriptase PCR results indicated that the *BraC3H* genes analysed in this study were induced under salt stress. Under drought stress the BraC3H02 was weekly upregulated, whereas *BraC3H04* was highly expressed (Fig. 7). Expression analysis of the *Brassica rapa C3H* genes suggest that these genes might be involved in abiotic stress tolerance. Further research studies are suggested to decipher the exact roles these genes play in the abiotic stress tolerance and growth and development of *Brassica rapa*.

References

Addepalli, B. and A.G. Hunt. 2007. A Novel endonuclease activity associated with the Arabidopsis Ortholog of the 30-kDa subunit of cleavage and polyadenylation specificity factor. *Nucl. Acids Res.,* 35(13): 4453-63.

Bailey, T.L., M. Boden, F.A. Buske, M. Frith, C.E. Grant, L. Clementi, J. Ren, W.W. Li and W.S. Noble. 2009. MEME Suite: Tools for Motif Discovery and Searching. *Nucl. Acids Res.,* 37(2): 202-8.

Bogamuwa, S.P. and J.C. Jang. 2014. Tandem CCCH Zinc finger proteins in plant growth, development and stress response. *Plant & Cell Physiol*., 55(8): 1367.

Boyer, T.P., S. Levitus, J.I. Antonov, R.A. Locarnini and H.E. Garcia. 2005. Linear trends in salinity for the world ocean. *Geophys. Res*., 32: 1-4.

Bray, E.A. 1997. Plant responses to water deficit. *Trends Plant Sci*., 2(2): 48-54.

Cheng, F., J. Wu, J. Liang and X. Wang. 2015. Genome triplication drove the diversification of brassica plants. *Hort. Res.*, 2014(1): 115-20.

Cheng, F., S. Liu, J. Wu, L. Fang, S. Sun, B. Liu, P. Li, W. Hua and X. Wang. 2011. BRAD, the genetics and genomics database for *Brassica* plants. *BMC Plant Biol*., 11(1): 136.

Dixit, A.R. and O.P. Dhankher. 2011. A Novel stress-associated protein AtSAP10' from arabidopsis thaliana confers tolerance to nickel, Manganese, Zinc, and High Temperature Stress. *PLoS One.,* 6(6): e20921.

Gasteiger, E., A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel and A. Bairoch. 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucl. Acids Res*., 31(13): 3784-88.

Ghosh, D. and J. Xu. 2014. Abiotic stress responses in plant Roots: A Proteomics Perspective. *Front. Plant Sci..,* 5: 1-13.

Jan, A., K. Maruyama, D. Todaka, S. Kidokoro, M. Abo, E. Yoshimura, K. Shinozaki, K. Nakashima and K.Y. Shinozaki. 2013. OsTZF1, a CCCH-Tandem zinc finger protein, confers Delayed senescence and stress tolerance in rice by regulating stress-related genes. *Plant Physiol.,* 161(3): 1202-16.

Jin, X., J. Ren, E. Nevo, X. Yin, D. Sun and J. Peng. 2017. Divergent evolutionary patterns of nac transcription factors are associated with diversification and gene duplications in angiosperm. *Front. Plant Sci.,* 8: 1-15.

Kazuko, S. and K. Yamaguchi. 2000. Molecular responses to dehydration and low temperature: Differences and cross-talk between two stress signaling pathways. *Curr. Opin. Plant Biol*., 3(3): 217-223.

Khatun, K., U.K. Nath, A.H.K. Robin, J.I. Park, D.J. Lee, M.B. Kim, C.K. Kim, K.B. Lim, I. Nou and M.Y. Chung. 2017. Genome-Wide analysis and expression profiling of zinc finger homeodomain (zhd) family genes reveal likely roles in organ development and stress responses in tomato. *BMC Genomics.,* 18(1): 695.

Kim, M.J., H.J. Kim, J.H. Pak, H.S. Cho, H.K. Choi, H.W. Jung and D.H. Lee. 2017. Overexpression of AtSZF2 from arabidopsis showed enhanced tolerance to salt stress in soybean. *Plant Breed. Biotech.*, 2017(1): 1-15.

Kumar, S., G. Stecher and K. Tamura. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. & Evol..,* 33(7): 1870-74.

Li, J.D. Jia. and X. Chen. 2001. HUA1, a Regulator of stamen and carpel identities in arabidopsis, codes for a nuclear rna binding protein. *The Plant Cell* ., 13(10): 2269-81.

Librado, P. and J. Rozas. 2009. DnaSP v5: A Software for Comprehensive Analysis of DNA Polymorphism Data. *Bioinformatics,* 25(11): 1451-52.

Lin, Y., J.B. Robbins, E.K.D. Nyannor, Y.H. Chen and I.K.O. Cann. 2005. A CCCH Zinc finger conserved in a replication protein a homolog found in diverse euryarchaeotes. *J. Bacteriol.*., 187(23): 7881-89.

Mazumdar, P., S.E. Lau, W.Y. Wee, P. Singh and J.A. Harikrishna. 2017. Genome-Wide Analysis of the CCCH Zinc-Finger gene family in banana (*Musa acuminata*): An Insight Into motif and gene structure arrangement, evolution and salt stress responses. *Trop. Plant Biol.,* 10(4): 177-93.

Peng, X., Y. Zhao, J. Cao, W. Zhang, H. Jiang, X. Li, Q. Ma, S. Zhu and B. Cheng. 2012. CCCH-Type Zinc finger family in maize: genome-wide identification, classification and expression profiling under abscisic acid and drought treatments. *PLoS One*., 7(7): e40120.

Pereira, A. 2016. Plant Abiotic stress challenges from the changing environment. *Front. Plant Sci.,* 7: 2013-15.

Pradhan, S., C. Kant, S. Verma and S. Bhatia. 2017. Genome-Wide analysis of the CCCH zinc finger family identifies tissue specific and stress responsive candidates in chickpea (*Cicer arietinum* L). *PLoS One*, 12(7): e0180469.

Rakow, G.F.W. 2004. Species, origin and economic importance of brassica. *Biotechn Agricult. & Forestry,* 54: 3-11.

Rombauts, S., P. Déhais, M.V. Montagu and Rouzé. 1999. PlantCARE, a plant Cis-acting regulatory element database. *Nucl. Acids Res.,* 27(1): 29596.

Seo, E. and D. Choi. 2015. Functional studies of transcription factors involved in plant defenses in the genomics Era. *Briefings in Funct. Gen.,* 14(4): 260-67.

Song, X.M., Z.N. Huang, W.K. Duan, J. Ren, T.K. Liu, Y. Li and X.L. Hou. 2014. Genome-Wide Analysis of the bHLH Transcription factor family in chinese cabbage (*Brassica rapa* ssp. Pekinensis). *Mol. Gen. & Genomics.,* 289(1): 77-91.

Sun, J., H. Jiang, Y. Xu, H. Li, X. Wu, Q. Xie and C. Li. 2007. The CCCH-Type Zinc finger proteins AtSZF1 and AtSZF2 regulate salt stress responses in arabidopsis. *Plant & Cell Physiol.,* 48(8): 1148-58.

Szklarczyk, D., A. Franceschini, S. Wyder, K. Forslund, D. Heller, J.H. Cepas and M. Simonovic. 2015. STRING v10: Protein-Protein interaction networks, integrated over the Tree of Life. *Nucl. Acids Res.,* 43 (1): 447-52.

Todeschini, A.L., A. Georges and R.A. Veitia. 2014. Transcription factors: specific DNA binding and specific gene regulation. *Trends Gen.,* 30(6): 211-19.

Wang, D., Y. Guo, C. Wu, G. Yang, Y. Li and C. Zheng. 2008. Genome-Wide analysis of CCCH Zinc finger family in arabidopsis and rice. *BMC Genomics*, 9: 1-20.

Wang, L., Y. Xu, C. Zhang, Q. Ma, S.H. Joo, S.K. Kim, Z. Xu and K. Chong. 2008. OsLIC, a Novel CCCH-Type zinc finger protein with transcription activation, mediates rice architecture via brassinosteroids signaling. *PLoS One.,* 3(10): e3521.

Wang, Q., J. Liu, Y. Wang, Y. Zhao, H. Jiang and B. Cheng. 2015. Systematic analysis of the maize phd-finger gene family reveals a subfamily involved in abiotic stress response. *Int. J. Mol. Sci.,* 16(10): 23517.

Wang, X., H. Wang, J. Wang, R. Sun, J. Wu, S. Liu and Y. Bai. 2011. The genome of the mesopolyploid crop species *Brassica rapa*. *Nature Gen..,* 43(10): 1035-39.

Wu, S., M. Wu, Q. Dong, H. Jiang, R. Cai and Y. Xiang. 2016. Genome-Wide identification, classification and expression analysis of the PHD-Finger Protein family in populus trichocarpa. *Gene.,* 575(1) : 75-89.

Xiong, L. and Y. Yang. 2003. Disease resistance and abiotic stress tolerance in rice are inversely modulated by an abscisic acid – inducible mitogen-activated protein kinase. *The Plant Cell*., 15: 745-59.

Xu, R. 2014. Genome-Wide analysis and identification of stress-responsive genes of the CCCH Zinc finger family in solanum lycopersicum. *Mol. Genet. & Genomics*, 289(5): 965-79.

Zhang, C., H. Zhang, Y. Zhao, H. Jiang, S. Zhu, B. Cheng and Y. Xiang. 2013. Genome-Wide analysis of the CCCH zinc finger gene family in medicago truncatula. *Plant Cell Reports,* 32(10): 1543-55.

Zhang, Y., P. Gao and J. Yuan. 2010. Plant protein-protein interaction network and interactome. *Curr. Genom.,* 11(1): 40-46.

Zhao, P., P. Liu, G. Yuan, J. Jia, X. Li, D. Qi, S. Chen, T. Ma, G. Liu and L. Cheng. 2016. New insights on drought stress response by global investigation of gene expression changes in sheepgrass (*Leymus chinensis*). *Front. Plant Sci*., 7: 1-18.

(Received for publication 19 June 2019)