Research Article

Genome-wide Identification, Characterization and Expression Profiling of SnRK Gene Family in Wheat: An *in-silico* Approach

Manoj Kumar Sharma[1], Sachin Kumar[1], Manoj Kumar Sharma[2,*]

[1]Department of Bioinformatics, JV College, Baraut (Baghpat), Uttar Pradesh, INDIA.
[2]Department of Botany, JV College, Baraut (Baghpat), Uttar Pradesh, INDIA.

Submission Date: 05-01-2022; Revision Date: 13 -02-2022; Accepted Date: 12-03-2022.

Correspondence:
Dr. Manoj Kumar Sharma,
Assistant Professor,
Department of Botany,
JV College, Baraut (Baghpat),
Uttar Pradesh, INDIA.

Email: mbhardwaj1501@gmail.com

ABSTRACT

Sucrose non-fermenting 1 related protein kinases (SnRK) are well known for their crucial roles in responding to biotic and abiotic stresses through activating protein phosphorylation pathways in plants. So far, members of this important gene family have already been cloned and characterized and well-studied in different plant species. However, there is still a lack of comprehensive genomic information about SnRK gene family in wheat (*Triticum aestivum* L.) on genome-scale. It is vital to perform genome-wide identification and characterization of TaSnRK gene family in the wheat for an opportunity to improve the variety that have more and more abiotic resistance. In the present study, total of 5097 TaSnRK genes were identified in wheat genome using BLAST algorithm against the fully annotated reference genome available in Ensemble Plants 42.0 using AtSnRK2.1 protein from *Arabidopsis thaliana* as a query. Sequence analysis revealed that the coding sequence (CDS) length of inferred TaSnRK genes ranged from 210 to 7721bp, and corresponding protein length ranged from 69 to 2448aa. All the 5097 genes were studied for the stability of their proteins. Out of them 18 *TaSnRK* genes located on 11 different chromosomes with an uneven distribution were finally inferred to have a stable protein structure based on different parameters. The genes are classified into 9 subgroups on the basis of phylogenetic analysis. According to gene structure and motif analysis through MEME, the TaSnRKs showed obvious divergence among subgroups. Finally, the expression analysis revealed differential responses among the *TaSnRK* genes to osmotic stress. Overall, the present genomic and *in silico* analyses of *TaSnRK* genes offer a solid foundation for further investigation of functions of these identified genes, leading to improved variety of wheat against abiotic/osmotic stress.

Key words: Sucrose non-fermentation-related protein kinase (SnRK) Wheat, Functional annotation, Expression profiling, Abiotic Stress, Osmotic stress.

INTRODUCTION

The sucrose non-fermentation-related protein kinase (SnRK) is a Ser/Thr protein kinase that regulates the interconnection of a wide range of signalling pathways by phosphorylating the target protein during plant stress responses.[1] The SnRK gene family has three subfamilies in *Arabidopsis thaliana*: SnRK1, SnRK2, and SnRK3.[2] The functional member of this important gene family, namely, SnRK1 has been found to be associated with higher nitrogen stress tolerance as well as energy sensing and gene regulation in plants.[3,4] Two genes, SnRK2 and SnRK3 have been found to play a part in signalling pathways that control plant responses to nutritional deficiency, drought, cold, salt, and osmotic stress.[1]

The SnRK2 family of serine/threonine kinases are plant-specific serine/threonine kinases implicated in abiotic stress response and ABA-dependent plant development.[5] SnRK2 is involved in the ABA signalling system, as well as osmotic stress and sugar metabolism.

SCAN QR CODE TO VIEW ONLINE

www.ajbls.com

DOI: 10.5530/ajbls.2022.11.15

Eight of the ten SnRK2 genes in *A. thaliana* can be triggered by hyperosmotic and salt stress, while five can be activated by ABA.[6] In particular, stomatal closure and ABA-mediated gene expression in the ABA regulatory pathway of AtSnRK2.6.[7,8] Furthermore, overexpression of AtSnRK2.8 results in increased drought tolerance.[9] This subfamily has been investigated extensively in several plants, including *Oryza sativa,*[10] *Zea mays,*[11] *Glycine max,*[12] *Gossypium hirsutum* L.,[13] *Brachypodium distachyon,*[14] and *Hevea brasiliensis*.[15] Abiotic stresses, particularly salt and osmotic stress, cause all SnRK2 members to be activated, overexpression of SAPK and NtSnRK2.1 boosted salt tolerance of transgenic plants.[16,17] Plant responses to salt stress are positively regulated by PtSnRK2.5 and PtSnRK2.7.[18] AtSnRK2.3, AtSnRK2.4, AtSnRK2.7, and AtSnRK2.8 overexpression in *A. thaliana* has been shown to improve plant tolerance to salt and other stressors.[19,20] TaSnRK2.9 has been proven in recent research to improve salt and drought stress tolerance by detoxifying reactive oxygen species (ROS).[20] Furthermore, overexpression of GhCIPK6 greatly improved transgenic *A. thaliana* tolerance to salt, dehydration, and ABA stress.[21] The salt stress signal transduction route is positively regulated by ZmSnRK2.8, whereas the drought stress signal transduction pathway is negatively regulated by ZmSnRK2.11.[22] Various findings like these at various times have proved that members of the SnRK2 subfamily are involved in salt and osmotic stress with other abiotic stress responses.

Wheat (*Triticum aestivum* L.) is one of the most important cereal crops in the world, yet abiotic stresses have a substantial impact on its growth and development, resulting in a significant drop in grain output.[23] The SnRK2 family of Arabidopsis contains 10 members and, with the exception of SnRK2.9, all of them are activated by saline stress.[6] Furthermore, over-expression of one member of the family, SnRK2.8, has been shown to increase drought tolerance and up-regulate stress-induced genes.[9] Activation of SnRK2 in response to hyperosmotic treatments depends on phosphorylation of a specific serine residue located in the activation loop, although some other sites may also be phosphorylated.[24,25] Other plant species also have members of the SnRK2 family that have been shown to be involved in salt signalling. In rice, for example, the entire SnRK2 family has been shown to be activated by hyperosmotic stress and this activation involves phosphorylation.[10] Interestingly, not all members of the family are activated similarly in response to various salt concentrations. Activation of one member, SAPK1, for example, has been observed at NaCl concentrations higher than 300 mM, whereas another, SAPK2, becomes active at lower concentrations. Domain exchange experiments have revealed that the C-terminal domain is responsible for these responses.[10] The C-terminal domain of SnRK2s is short and contains a characteristic acidic patch. Paradoxically, for understanding the evolution of the family, the acidic patch is highly aspartic acid-rich in some SnRK2s but glutamic acid-rich in others.[26]

Members of the soybean SnRK2 family, SPK1 and SPK2, have also been shown to be activated by NaCl when expressed in yeast, although concentrations higher than 0.5 M are required, and to phosphorylate a soybean phosphatidylinositol transfer protein (Ssh1p) in response to saline stress. Ssh1p might be involved in phosphoinositide metabolism, playing an essential role in hyperosmotic signalling.[27] In wheat, three SnRK2 family members have been shown to be induced by saline treatments. Expression of PKABA1, W55a, and TaSnRK2.4 is stimulated by high salt treatment and expression of W55a and AtSnRK2.4 in *Arabidopsis* has been shown to enhance salt tolerance.[28-30] In the present study, we have identified and annotated the members of SnRK2 subfamily with utilizing wheat reference genome data available in public domain

Experimental Procedure

The whole work is done in different steps sequentially as mentioned below.

Identification of SnRK genes in the wheat genome

The BLAST method was used to identify all possible SnRK domain containing genes in wheat genome. Using the amino acid sequence of 16 SnRK domain-containing protein from *A. thaliana* AtSnRK2.1, a BLASTP search was performed against the fully annotated reference genome [IWGSC] available in EnsemblPlants 42.0 (https://plants.ensembl.org/index.html).[31,32] We looked at and gathered all sequences with an E-value of less than 1.0. Pfam was used to confirm the presence of other domains in SnRK genes discovered by BLAST searches, in addition to the SnRK domain.[33] Gene features *e.g.,* composition and position of exons and introns etc. were investigated through Gene Structure Display Server (GSDS).[34]

Homology modelling, structure evaluation and structure alignment

The protein's 3D structure is more conserved than amino acid sequences, and vital for cellular functions *e.g.,* gene regulation and differentiation. However, many plant SnRK proteins have yet to be solved crystal and NMR structures, according to the Protein Data Bank

(PDB) repository (rcsb.org). Furthermore, structural biologists face a difficult problem in molecular modelling of SnRK proteins, so homology modelling (also known as comparative modelling) is used in the current study to estimate the 3D structure of wheat SnRK proteins based on the resolved structure of homologous proteins.[35]

Using the position-specific iterated BLAST algorithm (PSI -BLAST) against the PDB, the best homologs template structures were found based on a high score and a low *e*-value.[36] For wheat SnRK proteins, the automated Swiss-Model server was employed to predict the necessary 3D structures.[37,38] Furthermore, using PSVS (http://psvs-1 5-dev.nesg.org), Ramachandran plots of modelled wheat SnRK proteins were generated, assessing phi (Φ) and psi (Ψ) torsion angles and covalent bond quality, most favoured regions etc. The quality of protein 3D structure was assessed using ERRAT (http://saves.mbi.ucla.edu), a so-called "overall quality factor" for non-bonded atomic interactions.[39] A higher score denotes a greater level of quality. Projected models of wheat SnRK proteins were extracted in various 3D locations using UCSF CHIMERA 1.10.[40]

Physiochemical properties and subcellular localization

Physiochemical properties such as molecular weight (MW), isoelectric point (pI), aliphatic index (AI), and grand average of hydropathicity (GRAVY) were determined using the ProtParam tool on the Expasy website (https://web.expasy.org/protparam/).[41]

Subcellular localization of TaSnRK proteins was predicted using a novel integrated web-server called BUSCA[42] and Plant-mPLoc.[43]

Analysis of motifs

The MEME Suite is a comprehensive web-based toolkit for analysing sequence motifs in proteins, DNA, and RNA. Many biological functions are encoded by such motifs, and their identification, characterization is very critical in the study of molecular interactions in cells, including gene expression regulation.[44] Many biological functions are encoded by such motifs, and their identification, characterization is very critical in the study of molecular interactions in cells, including gene expression regulation. In the present study, TBtools, a bioinformatics software, was used for comparative analyses, of conserved motifs and their phylogenetic relationships which is important to reveal the conservation and differences among members of a gene family.[45]

Analysis of evolutionary relation among TaSnRK under study

To confirm the evolutionary relationship among TaSnRK proteins in plants, we used MEGA 6.0[46] and ClustalW ver. 2.1 for multiple sequence alignment (MSA) of the full length deduced amino acid sequences of TaSnRK domain-containing proteins in *T. aestivum*.[47] Subsequently, a phylogenetic tree was build using the neighbor-joining approach with the Poisson substitution model, uniform rates, and pairwise deletion,[48] with bootstrap values computed for a percentage of 1000 iterations.[49] The evolutionary relation among TaSnRK between *Arabidopsis thaliana, Oryza sativa*, and *Triticum aestivum* confirmed from ClustalW ver 2.1 and MEGA6.0 was displayed with help of iTOL (Interactive Tree Of Life) *i.e.,* a web-based application for viewing, manipulating, and annotating phylogenetic trees.[50]

Expression profiling

Expression profile of the TaSnRK genes, identified in this study, was examined using Wheat Expression Browser (http://www.wheatexpression.com/), which is powered by the expression Visualization and Integration Platform (expVIP) to retrieve RNA-seq expression data[51] and Plant-mPLoc.[43]

RESULTS

Identification of TaSnRK gene family in wheat

In the present study, initially a total of 5097 different loci encoding TaSnRK proteins in wheat have been identified using genome-wide approach. Thus, identified proteins were filtered, duplicate redundant sequences and various transcripts of the same gene were removed from the identified TaSnRK. The amino acid sequences of all 5097 proteins were checked for the existence of the conserved TaSnRK domain and other associated domains using the InterPro and PROSITE databases to further verify the reliability of these TaSnRK genes. The identified 5097 TaSnRK genes were named TaSnRK1 -1A to TaSnRK5097 -7D according to their chromosomal positions.

As revealed by sequence analysis, the length of inferred SnRK genes ranged from 210bp (TaSnRK1644-3A) to 7721bp (TaSnRK2559-4A), with corresponding protein lengths ranging from 69aa (TaSnRK1644-3A) to 2448aa (TaSnRK2120-3B).

Out of the 5097 different wheat loci encoding SnRK proteins, the gene ID, length, corresponding protein

Table 1: List of some SnRK*genes* identified in wheat genome in the study along with their length, corresponding protein length, chromosome location.

Gene	EnsemblPlants ID	Chromosome	Coordinates	Length (bp)	Length (aa)	Splice variants
TaSnRK97-1A	TraesCS1A02G203400	1A	365064927-365071233	2160	524	2
TaSnRK302-1B	TraesCS1B02G217500	1B	394405839-394411967	2111	526	2
TaSnRK623-1D	TraesCS1D02G450800	1D	493025001-493027243	1296	340	-
TaSnRK1261-2D	TraesCS2D02G020100	2D	9625899-9629201	1962	653	-
TaSnRK2024-3B	TraesCS3B02G334400	3B	540490811-540498364	2616	604	2
TaSnRK2687-4B	TraesCS4B02G053500	4B	42161234-42163491	1977	658	-
TaSnRK2792-4B	TraesCS4B02G328400	4B	619363496-619367707	2352	686	-
TaSnRK2949-4D	TraesCS4D02G325200	4D	485520708-485526115	2310	521	-
TaSnRK2950-4D	TraesCS4D02G325500	4D	485536885-485539970	2608	675	-
TaSnRK2979-5A	TraesCS5A02G052400	5A	32705951-32708571	2346	781	-
TaSnRK3163-5A	TraesCS5A02G487000	5A	657481959-657486706	1858	490	3
TaSnRK3170-5A	TraesCS5A02G500000	5A	666301440-666306856	2162	524	-
TaSnRK3209-5B	TraesCS5B02G056000	5B	62027233-62029396	2076	691	-
TaSnRK3222-5B	TraesCS5B02G064400	5B	72097141-72099708	2568	704	-
TaSnRK3419-5B	TraesCS5B02G500800	5B	668186318-668191395	2339	490	2
TaSnRK3488-5D	TraesCS5D02G063800	5D	59043061-59045181	2121	706	-
TaSnRK3694-5D	TraesCS5D02G501100	5D	529346790-529351653	2189	490	-
TaSnRK4746-7B	TraesCS7B02G324500	7B	578716034-578717257	759	252	-

(-) means no splice variants

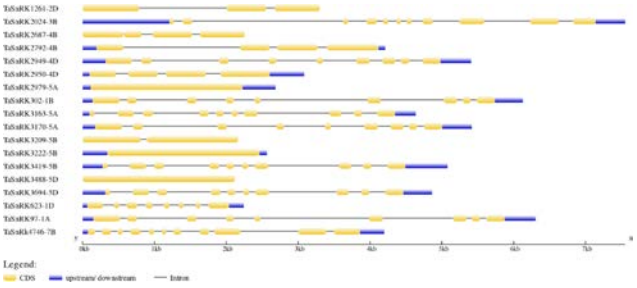

Figure 1: Gene features of TaSnRK genes as visualized with GSDS server.

length, chromosome location, splice variants, and coordinates of some selected TaSnRK, which found important in further investigations are listed in Table 1. The Gene features of some important genes as visualized with GSDS is shown in Figure 1.

Homology modelling, structure evaluation and structure alignment

All the 5097 simulated SnRK proteins encoded by the identified wheat loci were studied for their homology with typical SnRKs. When superimposed on homolog templates, the simulated 3D structures of typical TaSnRK proteins indicate less than 1 RMSD. Ramachandran plots for phi (Φ) and psi (Ψ) torsion angles study demonstrated that the simulated 3D structures of

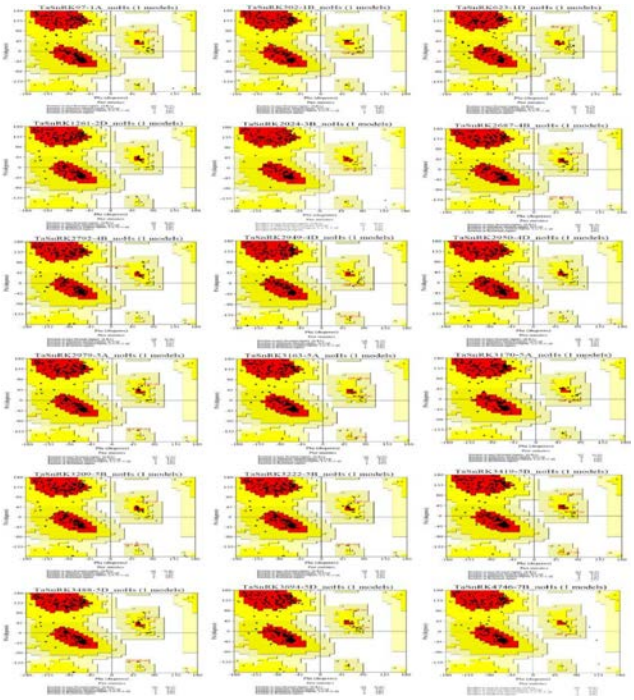

Figure 2: Ramachandran plots of modelled wheat bZIP proteins were computed using PSVS tool.

representative proteins have excellent geometry (Figure 2). The simulated structure revealed that up to 94.1 percent of residues fell in the most favored regionas seen in the plots (Table 2). A good quality model,

Table 2: List of 18 of the total 5097 identified TaSnRK genes in wheat genome along with the results of other studies e.g., homology modelling, %similarity, corresponding SMTL ID, properties of Ramachandran plot and ERRAT score (representing overall structure quality), with 0 Error.

Gene	SMTL ID	Percent Identity	Most Favoured Region	Errat	Verify 3D	Prove	Procheck
TaSnRK97-1A	4i94.1.A	68.42%	92.10%	97.41	P	W	8 eval, E:0, W:5, P:3
TaSnRK302-1B	4i93.1.A	68.35%	92.80%	98.14	P	W	8 eval, E:0, W:6, P:2
TaSnRK623-1D	5w4w.2.A	56.01%	90.70%	96.00	P	W	8 eval, E:0, W:6, P:2
TaSnRK1261-2D	6cth.1.A	60.80%	92.00%	96.47	P	W	8 eval, E:0, W:5, P:3
TaSnRK2024-3B	3tl8.2.A	66.46%	94.10%	93.92	P	W	9 eval, E:0, W:7, P:2
TaSnRK2687-4B	6cth.1.A	61.13%	90.70%	94.34	P	W	8 eval, E:0, W:6, P:2
TaSnRK2792-4B	6cth.1.A	58.86%	91.90%	94.22	P	W	9 eval, E:0, W:6, P:3
TaSnRK2949-4D	4i92.1.A	68.71%	92.70%	96.28	P	W	8 eval, E:0, W:5, P:3
TaSnRK2950-4D	6cth.1.A	58.72%	91.50%	91.54	P	W	8 eval, E:0, W:5, P:3
TaSnRK2979-5A	6cth.1.A	61.20%	91.20%	97.00	P	W	8 eval, E:0, W:6, P:2
TaSnRK3163-5A	4i94.2.A	72.47%	92.50%	98.92	P	W	8 eval, E:0, W:6, P:2
TaSnRK3170-5A	4i93.1.A	69.06%	92.30%	95.62	P	W	8 eval, E:0, W:6, P:2
TaSnRK3209-5B	6cth.1.A	62.21%	91.80%	92.28	P	W	8 eval, E:0, W:6, P:2
TaSnRK3222-5B	6cth.1.A	61.54%	91.10%	95.91	P	W	8 eval, E:0, W:6, P:2
TaSnRK3419-5B	4i94.2.A	72.47%	92.50%	98.92	P	W	8 eval, E:0, W:6, P:2
TaSnRK3488-5D	6cth.1.A	61.54%	90.70%	95.91	P	W	8 eval, E:0, W:6, P:2
TaSnRK3694-5D	4i94.1.A	72.82%	92.90%	98.19	P	W	8 eval, E:0, W:5, P:3
TaSnRK4746-7B	3tl8.2.A	65.02%	93.90%	95.10	P	W	9 eval, E:0, W:7, P:2

P stand for Pass, W- stand for warning, Eval- Evaluations, E- Errors

Figure 3: 3D structure images of TaSnRK proteins, as generated using the UCSF Chimera program from the resource for Biocomputing, Visualization, and Informatics (http://www.cgl.ucsf.edu/ chimera).

according to the PROCHECK algorithm, should have over 90% residues in the most favoured regions.[52] The Ramachandran plot is a very well recognized stage in the three-dimensional structure verification of proteins.[53,54] Modelled 3D structures of wheat TaSnRK proteins provides a structural foundation for understanding the complex mechanism of TaSnRKs. A total of 544 TaSnRK protein were investigated in the study to have a similarity of 50% or more.

Protein structures, modelled through homology modelling using automated Swiss-Model server that were having a most favoured region and ERRAT score more than 90, were further visualized in CPK by UCSF CHIMERA are shown in Figure 3. The structure generated using CHIMERA is found comparable to the established SnRKs.

Physiochemical properties and subcellular localization

The selected TaSnRKs' physicochemical parameters and subcellular localization as determined by the ProtParam tool and BUSCA, respectively are listed in Table 3. Almost all the proteins under study are discovered to be subcellularly localised in the nucleus, as indicated by studies using ProtParam and BUSCA, which supports the assumption that they have role to play in signal transduction.

Analysis of motifs

10 motifs were found among the gene (SnRK) being shared among *Arabidopsis thaliana, Oryza sativa*, and *Triticum aestivum*.

Table 3: Predicted physicochemical properties and subcellular localization of TaSnRK proteins.

Protein	Molecular weight	Isoelectric point	Aliphatic index	GRAVY	Subcellular Localization
TaSnRK97-1A	58571.49	6.26	70.86	-0.510	N
TaSnRK302-1B	58556.44	6.30	70.59	-0.509	N
TaSnRK623-1D	38958.28	6.93	73.18	0.532	N
TaSnRK1261-2D	72131.70	5.81	91.45	-0.032	N
TaSnRK2024-3B	66876.20	5.37	98.11	-0.154	Memb
TaSnRK2687-4B	72370.65	7.45	87.02	0.091	N
TaSnRK2792-4B	75933.32	6.92	81.63	-0.234	Cyt\N
TaSnRK2949-4D	57921.72	6.26	69.42	-0.472	N
TaSnRK2950-4D	73386.11	6.05	79.19	-0.205	N / Chl / Cyt
TaSnRK2979-5A	75370.57	7.83	81.93	-0.128	N
TaSnRK3163-5A	53872.93	5.83	81.49	-0.271	N
TaSnRK3170-5A	58431.28	6.19	68.45	-0.511	N
TaSnRK3209-5B	133069.82	6.44	79.25	-0.261	N / Chl
TaSnRK3222-5B	75316.87	7.14	83.31	-0.058	N / Memb
TaSnRK3419-5B	53872.93	5.83	81.49	-0.271	N
TaSnRK3488-5D	75801.38	8.34	81.84	-0.132	N / Memb
TaSnRK3694-5D	53828.83	5.75	81.69	-0.265	N
TaSnRK4746-7B	69138.88	6.71	95.26	-0.073	Memb

* N – stands for Nucleus, Memb- Cell Membrane, Chl- for Chloroplast, Cyt- for Cytoplasm

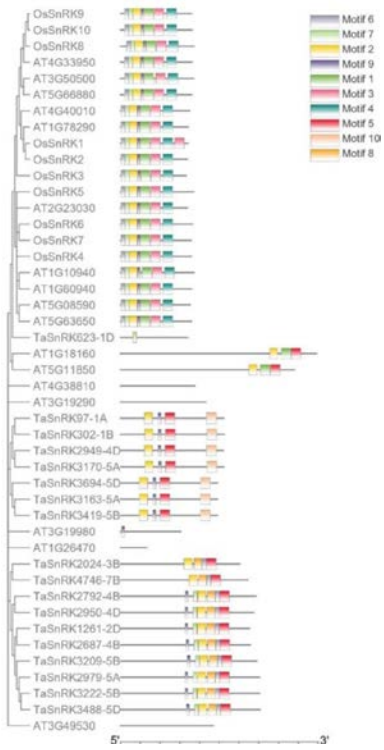

Figure 4: Motifs shared by the genes among *Arabidopsis thaliana, Oryza sativa,* and *Triticum aestivum* and indicating their evolutionary relations.

As shown in the Figure 4, Motif 2 is shared by 39 genes in total, among which 17 belongs to *T. aestivum*, 12 belongs to *A. thaliana* while 10 belongs to *O. sativa*. Motif 9 is shared by 36 genes in total, among which 15 belongs to *T. aestivum*, 11 belongs to *A. thaliana* while 10 belongs to *O. sativa*. Motif 6 is shared by 30 genes in total, among which 10 belongs to *T. aestivum*, 10 belongs to *O. sativa*, while 10 belongs to *A. thaliana*. Motif 7 is shared by 29 genes in total, among which 9 belongs to *T. aestivum*, 10 belongs to *O. sativa*, while 10 belongs to *A. thaliana*. Motif 1 is shared by 22 genes in total, among which 17 belongs to *T. aestivum*, 10 belongs to *O. sativa*, while 12 belongs to *A. thaliana*. Motif 3 is shared by 20 genes in total, among which 10 belongs to *O. sativa*, while 10 belongs to *A. thaliana*, Motif 3 is not found in genes of *T. aestivum*. Motif 4 is shared by 20 genes in total, among which 10 belongs to *O. sativa*, while 10 belongs to *A. thaliana*, Motif 4 is not found in genes of *T. aestivum*. Motif 5 is shared by 19 genes in total, among which 17 belongs to *T. aestivum* while 2 belongs to *A. thaliana*. Motif 8 and Motif 10 is shared by 10 genes and 7 genes of *T. aestivum* only respectively.

Evolutionary relation among TaSnRKs of A. thaliana, T. aestivum and O. sativa

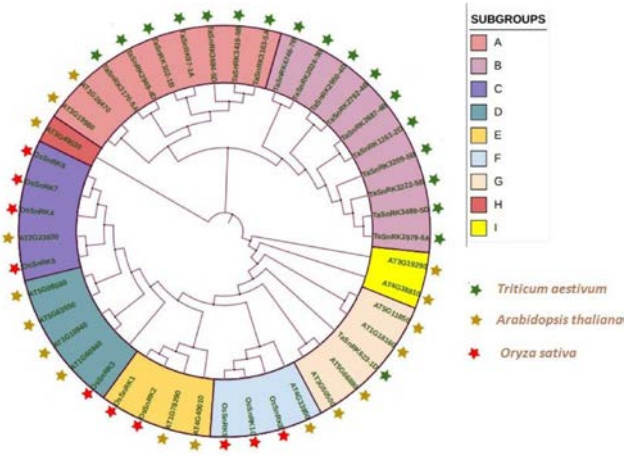

Figure 5: Evolutionary relationship among TaSnRKs, based on the sequence alignment of the different TaSnRKs proteins of *A. thaliana, T. aestivum and O. sativa*.

A phylogenetic tree was constructed to confirm the evolutionary relationship between *T. aestivum, A. thaliana,* and *O. sativa*. The unrooted phylogeny tree was generated using ClustalW 2.1 and MEGA 6.0 program and displayed with iTOL (Figure 5). TaSnRKs were mainly classified into 9 subgroups. Subgroups A has 9 genes in total, among which 7 belongs to *T. aestivum* and 2 belongs to *A. thaliana*. Subgroups B has 10 genes belongs to *T. aestivum*. Subgroups C has 5 genes in total, among which 4 belongs to *O. sativa* and 1 belongs to *A. thaliana*. Subgroups D has 5 genes in total, among which 4 belongs to *A. thaliana* and 1 belongs to *O. sativa*. Subgroups E has 4 genes in total, among which 2 belongs to *O. sativa* and 2 belongs to *A. thaliana*. Subgroups F has 4 genes in total, among which 3 belongs to *O. sativa* and 1 belongs to *A. thaliana*. Subgroups G has 5 genes in total, among which 4 belongs to *A. thaliana* 10 genes 1 gene belongs to *T. aestivum*. Subgroups H has 1 gene which 4 belongs to *A. thaliana*. Subgroups I has 2 genes which 4 belongs to *A. thaliana*. Phylogeny analysis revealed presence of divergent sub-groups of SnRK candidate genes in *T. aestivum* and on comparison exhibited similarity with *A. thaliana* and *O. sativa*. So, indicating that the SnRK candidate genes encoding proteins are derived from common ancestor *A. thaliana* and *O. sativa*.

Expression profile

The expression profile under osmotic stress revealed the expression of seven genes TaSnRK2024-3B, TaSnRK2792-4B, TaSnRK2949-4D, TaSnRK2950-4D, TaSnRK302-1B, TaSnRK3170-5A, and TaSnRK97-1A (Figure 6). Mainly TaSnRK2024-3B, TaSnRK302-1B and TaSnRK97-1A exhibited the higher expression

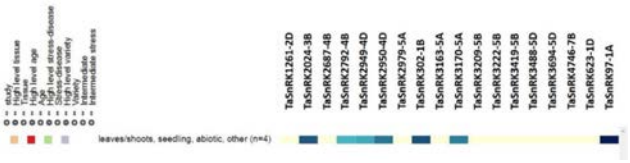

Figure 6: Expression profiles of TaSnRK genes. The heatmap was constructed using the expVIP, and the FPKM (fragments per kilobase of transcript per million fragments) value of TaSnRK genes were transformed by log2. The yellow and blue colours represent the lower and higher relative abundance of the transcript, respectively.

under stress condition as revealed by the higher relative abundance of the transcript.

Studies on management of abiotic stress has been documented to increase wheat production,[54] in the present study TaSnRK2024-3B, TaSnRK302-1B and TaSnRK97-1A are supposed to play most important role in signal transduction of abiotic stress among the total of 5097 loci identified through genome wide search.

DISCUSSION

As mentioned earlier in introduction, the sucrose non-fermentation-related protein kinase (SnRK) is a Ser/Thr protein kinase that regulates the interconnection of a wide range of signalling pathways by phosphorylating the target protein during plant stress responses.[1] In *Arabidopsis thaliana*: SnRK gene family has been well studied and found to have three subfamilies SnRK1, SnRK2, and SnRK3.[2] Before our study, there had been no genome-wide, in-depth study of the SnRK family reported in wheat.

In this study 18 TaSnRK genes located on 11 different chromosomes with an uneven distribution were finally inferred to have a stable protein structure based on different parameters using tools and parameters used by different workers in past.[31,32,35] We also identified instances of both the conservation and divergence of the gene expression, and protein evolution in wheat. Expression data analysis revealed wheat SnRK genes are potentially intricate participants in regulating the pathways of abiotic stress. The genome-wide identification and characterization of SnRK gene in wheat is an essential starting point for further exploring the function of this gene in depth. It is believed that, as the accumulation and extension of data on genomes and transcriptomes continues, there will be a much better understanding of the SnRK gene in wheat.

Further the protein's 3D structure is more conserved than amino acid sequences, and vital for cellular functions e.g., gene regulation and differentiation. However, many plant SnRK proteins have yet to be

solved crystal and NMR structures, according to the Protein Data Bank (PDB) repository (rcsb.org). The best homologs template structures were discovered based on a high score and a low *e*-value.[36] The automated Swiss-Model server was used to anticipate the essential 3D structures for wheat SnRK proteins.[37,38]

Furthermore, Ramachandran plots of modelled wheat SnRK proteins were created and highly favourable areas were assessed. The quality of protein 3D structure for non-bonded atomic interactions is assessed.[39] Projected models of wheat SnRK proteins were extracted in various 3D locations.[40] Up to 94.1 percent of residues fell in the most favourable locations, according to the modelled structure. The study looked at a total of 544 TaSnRK proteins to see if they were comparable by 50% or more. The molecular weight (MW), isoelectric point (pI), aliphatic index (AI), and grand average of hydropathicity (GRAVY) were determined.[41] Subcellular localization of TaSnRK proteins was predicted.[43] Almost all the proteins under study are discovered to be subcellularly localised in the nucleus, which supports the assumption that they have role to play in signal transduction.

Biological motifs encode a wide range of biological functions, and their identification and definition are crucial in the study of molecular interactions in cells, including gene expression regulation.[44] There were ten motifs shared by *Arabidopsis thaliana, Oryza sativa*, and *Triticum aestivum* in the gene (SnRK). The deduced amino acid sequences of TaSnRK domain-containing proteins in *T. aestivum, Arabidopsis thaliana*, and *Oryza sativa* were used to confirm the evolutionary link among TaSnRK proteins in plants. In *T. aestivum,* phylogeny analysis revealed the presence of divergent 9 subgroups of SnRK candidate genes, which were compared to *A. thaliana* and *O. sativa* and found to be identical. As a result, the SnRK candidate genes that encode proteins are descended from *A. thaliana* and *O. sativa's* common ancestor. To collect RNA-seq expression data, the expression profile of the TaSnRK genes identified in this work was evaluated.[51] In the current study, TaSnRK2024-3B, TaSnRK302-1B, and TaSnRK97-1A are assumed to play the most essential role in signal transduction of abiotic stress among the total of 5097 loci identified through genome wide search.

CONCLUSION

Though certain SnRK2 genes of *Arabidopsis* and rice has been well defined and their function has clearly been demonstrated, SnRK genes in wheat are still subtle. In the present study, a genome-wide study using bioinformatics-based techniques and algorithms, provides a systematic identification and functional annotation of the wheat SnRK gene family (TaSnRK). Total 18 TaSnRK genes were inferred to have a stable protein structure based on different parameters, located on different chromosome with an uneven distribution, out of a total of 5097 TaSnRK loci initially found in the wheat genome. Finally, the expression analysis revealed differential responses among the TaSnRK genes to osmotic stress. Overall, the present extensive bioinformatics analysis of TaSnRK genes offer a solid foundation for further investigation of TaSnRK2 gene functions, leading to development of wheat cultivars resistant to abiotic stress.

ACKNOWLEDEGMENT

The authors hereby acknowledge Dr. Anuj Kumar, Post-Doctoral Fellow, Department of Microbiology and Immunology, Dalhousie University, Halifax, Nova Scotia, Canada for his assistance in finalizing this manuscript.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Coello P, Hey SJ, Halford NG. The sucrose non-fermenting-1-related (SnRK) family of protein kinases: Potential for manipulation to improve stress tolerance and increase yield. J Exp Bot. 2011 Jan;62(3):883-93. doi: 10.1093/jxb/erq331, PMID 20974737.
2. Hrabak EM, Chan CW, Gribskov M, Harper JF, Choi JH, Halford N, *et al*. The Arabidopsis CDPK-SnRK superfamily of protein kinases. Plant Physiol. 2003 Jun;132(2):666-80. doi: 10.1104/pp.102.011999, PMID 12805596.
3. Wang X, Peng F, Li M, Yang L, Li G. Expression of a heterologous SnRK1 in tomato increases carbon assimilation, nitrogen uptake and modifies fruit development. J Plant Physiol. 2012 Aug;169(12):1173-82. doi: 10.1016/j.jplph.2012.04.013, PMID 22727046.
4. Robaglia C, Thomas M, Meyer C. Sensing nutrient and energy status by SnRK1 and TOR kinases. Curr Opin Plant Biol. 2012 Jun;15(3):301-7. doi: 10.1016/j.pbi.2012.01.012, PMID 22305521.
5. Kulik A, Wawer I, Krzywińska E, Bucholc M, Dobrowolska G. SnRK2 protein kinases--key regulators of plant response to abiotic stresses. Omics. 2011 Dec;15(12):859-72. doi: 10.1089/omi.2011.0091, PMID 22136638.
6. Boudsocq M, Barbier-Brygoo H, Laurière C. Identification of nine sucrose nonfermenting 1-related protein kinases 2 activated by hyperosmotic and saline stresses in *Arabidopsis thaliana*. J Biol Chem. 2004 Oct;279(40):41758-66. doi: 10.1074/jbc.M405259200, PMID 15292193.
7. Mustilli AC, Merlot S, Vavasseur A, Fenzi F, Giraudat J. Arabidopsis OST1 protein kinase mediates the regulation of stomatal aperture by abscisic acid and acts upstream of reactive oxygen species production. Plant Cell. 2002 Dec;14(12):3089-99. doi: 10.1105/tpc.007906, PMID 12468729.
8. Yoshida R, Umezawa T, Mizoguchi T, Takahashi S, Takahashi F, Shinozaki K. The regulatory domain of SRK2E/OST1/SnRK2.6 interacts with ABI1 and integrates abscisic acid (ABA) and osmotic stress signals controlling stomatal closure in Arabidopsis. J Biol Chem. 2006 Feb 24;281(8):5310-8. doi: 10.1074/jbc.M509820200, PMID 16365038.

9. Umezawa T, Yoshida R, Maruyama K, Yamaguchi-Shinozaki K, Shinozaki K. SRK2C, a SNF1-related protein kinase 2, improves drought tolerance by controlling stress-responsive gene expression in *Arabidopsis thaliana*. Proc Natl Acad Sci U S A. 2004 Dec;101(49):17306-11. doi: 10.1073/pnas.0407758101, PMID 15561775.
10. Kobayashi Y, Yamamoto S, Minami H, Kagaya Y, Hattori T. Differential activation of the rice sucrose nonfermenting1-related protein kinase2 family by hyperosmotic stress and abscisic acid. Plant Cell. 2004 May;16(5):1163-77. doi: 10.1105/tpc.019943, PMID 15084714.
11. Huai J, Wang M, He J, Zheng J, Dong Z, Lv H, *et al*. Cloning and characterization of the SnRK2 gene family from *Zea mays*. Plant Cell Rep. 2008 Dec;27(12):1861-8. doi: 10.1007/s00299-008-0608-8, PMID 18797872.
12. Zhao W, Cheng YH, Zhang C, Shen XJ, You QB, Guo W, *et al*. Genome-wide identification and characterization of the GmSnRK2 family in soybean. Int J Mol Sci. 2017 Aug 23;18(9):1834. doi: 10.3390/ijms18091834, PMID 28832544.
13. Liu Z, Ge X, Yang Z, Zhang C, Zhao G, Chen E, *et al*. Genome-wide identification and characterization of SnRK2 gene family in cotton (*Gossypium hirsutum* L.). BMC Genet. 2017 Jun;18(1):54. doi: 10.1186/s12863-017-0517-3, PMID 28606097.
14. Wang L, Hu W, Sun J, Liang X, Yang X, Wei S, *et al*. Genome-wide analysis of SnRK gene family in *Brachypodium distachyon* and functional characterization of BdSnRK2.9. Plant Sci. 2015 Aug;237:33-45. doi: 10.1016/j.plantsci.2015.05.008, PMID 26089150.
15. Guo D, Li HL, Zhu JH, Wang Y, An F, Xie GS, *et al*. Genome-wide identification, characterization, and expression analysis of SnRK2 family in *Hevea brasiliensis*. Tree Genetics and Genomes. 2017 Aug;13(4):86. doi: 10.1007/s11295-017-1168-2.
16. Diédhiou CJ, Popova OV, Dietz KJ, Golldack D. The SNF1-type serine-threonine protein kinase SAPK4 regulates stress-responsive gene expression in rice. BMC Plant Biol. 2008 Apr 28;8:49. doi: 10.1186/1471-2229-8-49, PMID 18442365.
17. Zhang H, Jia H, Liu G, Yang S, Zhang S, Yang Y, *et al*. Cloning and characterization of SnRK2 subfamily II genes from *Nicotiana tabacum*. Mol Biol Rep. 2014 Sep;41(9):5701-9. doi: 10.1007/s11033-014-3440-y, PMID 24919756.
18. Song X, Yu X, Hori C, Demura T, Ohtani M, Zhuge Q. Heterologous overexpression of poplar SnRK2 genes enhanced salt stress tolerance in Arabidopsis thaliana. Front Plant Sci. 2016 May;7:612. doi: 10.3389/fpls.2016.00612, PMID 27242819.
19. Zhang H, Mao X, Jing R, Chang X, Xie H. Characterization of a common wheat (*Triticum aestivum* L.) TaSnRK2.7 gene involved in abiotic stress responses. J Exp Bot. 2011 Jan;62(3):975-88. doi: 10.1093/jxb/erq328, PMID 21030389.
20. Feng J, Wang L, Wu Y, Luo Q, Zhang Y, Qiu D, *et al*. TaSnRK2.9, a sucrose non-fermenting 1-related protein kinase gene, positively regulates plant response to drought and salt stress in transgenic tobacco. Front Plant Sci. 2018;9:2003. doi: 10.3389/fpls.2018.02003, PMID 30693013.
21. He L, Yang X, Wang L, Zhu L, Zhou T, Deng J, *et al*. Molecular cloning and functional characterization of a novel cotton CBL-interacting protein kinase gene (GhCIPK6) reveals its involvement in multiple abiotic stress tolerance in transgenic plants. Biochem Biophys Res Commun. 2013 May;435(2):209-15. doi: 10.1016/j.bbrc.2013.04.080, PMID 23660187.
22. Zhang F, Chen XJ, Wang JH, Zheng J. Overexpression of a maize SNF-related protein kinase gene, ZmSnRK2.11, reduces salt and drought tolerance in Arabidopsis. J Integr Agric. 2015 Jul;14(7):1229-41. doi: 10.1016/S2095-3119(14)60872-8.
23. Miao L, Mao X, Wang J, Liu Z, Zhang B, Li W, *et al*. Elite haplotypes of a protein kinase gene TaSnRK2.3 associated with important agronomic traits in common wheat. Front Plant Sci. 2017 Mar;8:368. doi: 10.3389/fpls.2017.00368, PMID 28400774.
24. Burza AM, Pekala I, Sikora J, Siedlecki P, Małagocki P, Bucholc M, *et al*. *Nicotiana tabacum* osmotic stress-activated kinase is regulated by phosphorylation on Ser-154 and Ser-158 in the kinase activation loop. J Biol Chem. 2006 Nov;281(45):34299-311. doi: 10.1074/jbc.M601977200, PMID 16980311.
25. Boudsocq M, Droillard MJ, Barbier-Brygoo H, Laurière C. Different phosphorylation mechanisms are involved in the activation of sucrose non-fermenting 1 related protein kinases 2 by osmotic stresses and abscisic acid. Plant Mol Biol. 2007 Mar;63(4):491-503. doi: 10.1007/s11103-006-9103-1, PMID 17103012.
26. Halford NG, Hardie DG. SNF1-related protein kinases: Global regulators of carbon metabolism in plants? Plant Mol Biol. 1998 Jul;37(5):735-48. doi: 10.1023/a:1006024231305, PMID 9678569.
27. Monks DE, Aghoram K, Courtney PD, DeWald DB, Dewey RE. Hyperosmotic stress induces the rapid phosphorylation of a soybean phosphatidylinositol transfer protein homolog through activation of the protein kinases SPK1 and SPK2. Plant Cell. 2001 May;13(5):1205-19. doi: 10.1105/tpc.13.5.1205, PMID 11340192.
28. Holappa LD, Walker-Simmons MK. The Wheat Abscisic Acid-Responsive Protein Kinase mRNA, PKABA1, Is Up-Regulated by Dehydration, Cold Temperature, and Osmotic Stress. Plant Physiol. 1995 Jul;108(3):1203-10. doi: 10.1104/pp.108.3.1203, PMID 12228537.
29. Xu ZS, Liu L, Ni ZY, Liu P, Chen M, Li LC, *et al*. W55a encodes a novel protein kinase that is involved in multiple stress responses. J Integr Plant Biol. 2009 Jan;51(1):58-66. doi: 10.1111/j.1744-7909.2008.00776.x, PMID 19166495.
30. Mao X, Zhang H, Tian S, Chang X, Jing R. TaSnRK2.4, an SNF1-type serine/threonine protein kinase of wheat (*Triticum aestivum* L.), confers enhanced multistress tolerance in Arabidopsis. J Exp Bot. 2010 Mar;61(3):683-96. doi: 10.1093/jxb/erp331, PMID 20022921.
31. Bolser D, Staines DM, Pritchard E, Kersey P. Ensembl plants: Integrating tools for visualizing, mining, and analyzing plant genomics data. Methods Mol Biol. 2016;1374:115-40. doi: 10.1007/978-1-4939-3167-5_6, PMID 26519403.
32. Kumar A, Batra R, Gahlaut V, Gautam T, Kumar S, Sharma M, *et al*. Genome-wide identification and characterization of gene family for RWP-RK transcription factors in wheat (*Triticum aestivum* L.). PLOS ONE. 2018 Dec;13(12):e0208409. doi: 10.1371/journal.pone.0208409, PMID 30540790.
33. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, *et al*. InterPro in 2017-beyond protein family and domain annotations. Nucleic Acids Res. 2017 Jan;45(D1):D190-9. doi: 10.1093/nar/gkw1107, PMID 27899635.
34. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics. 2015 Apr;31(8):1296-7. doi: 10.1093/bioinformatics/btu817, PMID 25504850.
35. Kumar A, Mishra DC, Rai A, Sharma MK, Gajula MNP. *In-silico* analysis of protein–protein interaction between resistance and virulence protein during leaf rust disease in wheat (*Triticum aestivum* L.). World Res J Pept Protein. 2013;2(1):52-8.
36. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, *et al*. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 1997 Sep;25(17):3389-402. doi: 10.1093/nar/25.17.3389, PMID 9254694.
37. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. Bioinformatics. 2006 Jan;22(2):195-201. doi: 10.1093/bioinformatics/bti770. PMID 16301204.
38. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, *et al*. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 2014 Jul;42(Web Server issue); Web Server Issue:W252-8. doi: 10.1093/nar/gku340, PMID 24782522.
39. Colovos C, Yeates TO. Verification of protein structures: Patterns of nonbonded atomic interactions. Protein Sci. 1993 Sep;2(9):1511-9. doi: 10.1002/pro.5560020916, PMID 8401235.
40. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, *et al*. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 2004 Oct;25(13):1605-12. doi: 10.1002/jcc.20084, PMID 15264254.
41. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, *et al*. Protein identification and analysis tools in the ExPASy server. Methods Mol Biol. 1999;112:531-52. doi: 10.1385/1-59259-584-7:531. PMID 10027275.
42. Savojardo C, Martelli PL, Fariselli P, Profiti G, Casadio R. BUSCA: An integrative web server to predict subcellular localization of proteins. Nucleic Acids Res. 2018 Jul;46(W1):W459-66. doi: 10.1093/nar/gky320, PMID 29718411.

43. Chou KC, Shen HB. Plant-mPLoc: A top-down strategy to augment the power for predicting plant protein subcellular localization. PLOS ONE. 2010 Jun 28;5(6):e11335. doi: 10.1371/journal.pone.0011335, PMID 20596258.
44. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME suite. Nucleic Acids Res. 2015 Jul;43(W1):W39-49. doi: 10.1093/nar/gkv416, PMID 25953851.
45. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, *et al*. TBtools: An integrative toolkit developed for interactive analyses of big Biological Data. Mol Plant. 2020 Aug;13(8):1194-202. doi: 10.1016/j.molp.2020.06.009, PMID 32585190.
46. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013 Dec;30(12):2725-9. doi: 10.1093/molbev/mst197, PMID 24132122.
47. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, *et al*. Clustal W and Clustal X version 2.0. Bioinformatics. 2007 Nov;23(21):2947-8. doi: 10.1093/bioinformatics/btm404, PMID 17846036.
48. Saitou N, Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987 Jul;4(4):406-25. doi: 10.1093/oxfordjournals.molbev.a040454, PMID 3447015.
49. Felsenstein J. Confidence Limits on Phylogenies: An Approach using the Bootstrap. Evolution. 1985 Jul;39(4):783-91. doi: 10.1111/j.1558-5646.1985.tb00420.x, PMID 28561359.
50. Letunic I, Bork P. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. Nucleic Acids Res. 2019 Jul;47(W1):W256-9. doi: 10.1093/nar/gkz239, PMID 30931475.
51. Borrill P, Ramirez-Gonzalez R, Uauy C. expVIP: A customizable RNA-seq data analysis and visualization platform. Plant Physiol. 2016 Apr;170(4):2172-86. doi: 10.1104/pp.15.01667, PMID 26869702.
52. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: A program to check the stereochemical quality of protein structures. J Appl Crystallogr. 1993 Apr;26(2):283-91. doi: 10.1107/S0021889892009944.
53. Kumar A, Kumar S, Kumar A, Sharma N, Sharma M, Singh KP, *et al*. Homology modeling, molecular docking and Molecular Dynamics based functional insights into rice urease bound to urea. Proc Natl Acad Sci India Sect B Biol Sci. 2018;88(4):1539-48. doi: 10.1007/s40011-017-0898-0.
54. Kumar A, Kumar S, Kumar U, Suravajhala P, Gajula MNVP. Functional and structural insights into novel DREB1A transcription factors in common wheat (*Triticum aestivum* L.): A molecular modeling approach. Comput Biol Chem. 2016 Oct;64:217-26. doi: 10.1016/j.compbiolchem.2016.07.008, PMID 27471160.

Cite this article: Sharma MK, Kumar S, Sharma MK. Genome-wide Identification, Characterization and Expression Profiling of SnRK Gene Family in Wheat: An *in-silico* Approach. Asian J Biol Life Sci. 2022;11(1):107-16.