ANNALS OF PLANT SCIENCES
ISSN: 2287-688X
OPEN ACCESS
www.annalsofplantsciences.com

Research Article

Genome-wide identification of MscS like gene family in *Cicer arietinum* using bioinformatics approach

Noymi Basumatary, Archana Kumari, Jatin Sarmah*
DBT sponsored Bioinformatics Infrastructure Facility, Department of Biotechnology, Bodoland University, Kokrajhar, Assam, India.

Received: 12/30/2017; **Accepted:** 1/20/2018

Abstract: In plants, the mechanosensitive channels have important roles in proprioception, gravity sensing, growth of pollen tube and controlling the shape and size of plastid. In the present study, we performed computational analysis and identified the mechanosensitive channel of small conductance like (MscS) gene family in *Cicer arietinum* (chickpea). Six *C. arietinum* MscS-like (CaMSL) genes were identified. These six genes were distributed on four different chromosomes. A detailed overview of CaMSL genes in *C. arietinum* is explored by using a total of 33107 proteins from chickpea. On the basis of BLAST analysis, presence of complete ORF and evolutionary relationship we identified 6 genes (named as CaMSL1, CaMSL2, CaMSL3, CaMSL4, CaMSL5 and CaMSL6). Further, the transmembrane regions, subcellular localization, physico-chemical properties, gene structure analysis and genome structure annotation of the identified genes confirmed that the identified genes might act as transmembrane proteins. Prediction of their sub-cellular localization demonstrated three genes to be located in the plasma membrane, two in the chloroplast membrane, and one in the mitochondrion. Thus, further analysis of the data obtained from this study will help to make a baseline to increase the understanding of involvement of MSL gene families in plant growth, development and other functions in *C. arietinum*.

Keywords: MscS genes, transmembrane protein, sub-cellular localization, gene structure

Introduction

Chickpea (*Cicer arietinum*) is one of the important legume having family Fabaceae, subfamily Faboideae and genus *Cicer*. Commonly known as Bengal gram, garbanzo bean or Kabuli chana, chickpeas are one among the most widely consumed legumes. *C. arietinum* is the only cultivated species of the *Cicer* genus. It is a nitrogen-fixing legume and can be used to restore soil fertility (Edwards *et al.,* 2005). It reduces the use of pesticides and herbicides. Based on the seed morphology chickpeas are categorized as *Desi* type and *Kabuli*. The seeds of the *Desi* type are small and dark brown while those of *Kabuli* are bold pale brown seeds (Gupta *et al.,* 2016). *C. arietinum* is a highly nutritious food containing a high protein and starch content. The United Nations Food and Agricultural Organization (UNFAO) as well as World Health Organization (WHO) have established that the proteins content are very high in cooked and also in germinated chickpeas. *C. arietinum* is rich in essential amino acids (http://www.fao.org/3/a-i6212e.pdf). According to the geographical distribution in Asian countries India is the largest producer as well as consumer of chickpea, followed by other Asian countries and Canada too. In India, chickpea contributes to about 45% of the total pulses produced. However, in the last few years production of chickpea has fallen down(http://www.commoditiescontrol.com/eagritrader/common/newsdetail.php?type=SPR&itemid=8204&cid1=, 2, &varietyid=, 33). Though good farm management practices affect the crop yield, certain factors such as frost and heavy rain after sowing, flowering stage and seed maturity stage may also results in reduction of chickpea yield (http://www.agrifarming.in/chickpea-farming).

Various genes are playing important role in chickpea; Mechanosensitive channel of small conductance (MscS) gene is one of them. MscS or MSL gene families are a group of membrane proteins that are capable of responding to membrane stimuli including touch, gravity, osmotic pressure and stress (Fasano *et al.,* 2002; Edwards *et al.,* 2012). They are also known as mechanosensitive ion channels or stress-gated ion channels (Sukharev and Sachs 2012; Gottlieb and Sachs 2012). In addition to the mechanical stimuli, they also respond to development process signals such as lateral roots development, damages to the cell wall, pollen tube development and plant-pathogen interactions (Anishkin and Sukharev 2009; Cox *et al.,* 2015).

In plants, the mechanosensitive channels play important role in proprioception, gravity sensing and maintenance of the shape and size of plastid (Perozo 2006; Carlos 2011). MscS function as mechanotransducers. It has the capacity to generate both electrical and ion flux signal. These signals are responsible for the control of external and internal

***Corresponding Author:**
Dr. Jatin Sarmah,
Associate Professor, Department of Biotechnology,
Bodoland University, Kokrajhar, Assam, India.
E-mail: jatinsarmahindia@gmail.com

http://dx.doi.org/10.21746/aps.2018.7.2.4

stimuli. It also allows plants to make a distinction between up to down through sensing the force of gravity (Lucas *et al.,* 2013). MscS genes are not pressure-sensitive. But it is very much sensitive to various local stresses for example tension in the surrounding lipid bilayer (Pliotas *et al.,* 2015). On the other hand, mechanosensitive channels are involved in controlling the membrane tension via altering the confirmation of open and close state. These channels are paving a new pathway to study the interplay between membrane mechanical properties and their proteomics analysis (Monshausen and Haswell 2013; Jayaraman *et al.,* 2014). McsS regulates osmotic pressure changes within the cell without the help of other proteins. It helps to open in response to stretch forces in the membrane lipid bilayer. MscS also has capacity to form an ion channel having conductance capacity approximately 1.0 nanosiemens (Wang *et al.,* 2008). McsS channel is very sensitive to voltage because of the depolarization of membrane, less tension is required to open the channel and vice versa (www.uniprot.org/uniprot/P0C0S1). The first MSs were identified in *Escherichia coli*. MS ion channels were well characterized in model plant system both in dicot *(*9 MSL genes in *A. thaliana;* Haswell *et al.,* 2006*)* and monocot (6 MSL genes in *O. sativa).*

Considering the nutritional value and the present status of *C. arietinum* production, the present study has been aimed to identify and analyze the MscS gene family that may help to understand their roles in the development of the plant (Kawahara *et al.,* 2013). All organisms starting from unicellular to multi-cellular have ability to sense and respond to mechanical stimuli both in their internal and external environments for their proper growth, development and survival (Chang *et al.,* 1998). This action of sensing and responding to such mechanical forces are performed by different types of proteins and their protein complexes (Liu *et al.,* 2011). The mechanosensitive channels alter their conformation between open states as well as in a closed state, allowing solutes to flow across the cell membranes to cop up with the membrane tension (Haswell *et al.,* 2011). The first MscS identified in the plasma membrane of giant *E. coli* spheroplasts has a conductance of 350-950pS, depending on the ionic conditions (Haswell 2007). In plants, the MscS-like (MSL) genes are identified in *Arabidopsis thaliana, Oryza sativa, Zea mays, Solanum lycopersicum, Glycine max, Brachypodium distachyon, Sorghum bicolor, Hordeum vulgare and Populus Trichocarpa*. The monocot plants (*O. sativa, Z. mays, S. bicolor, H. vulgare* and *B. distachyon*) were reported to have 6 to 8 MSL genes, while the dicot plants (*G. max* and *P. trichocarpa)* has 14 MSL genes each and *S. lycopersicum* with 8 MSL genes (Saddhe and Kumar 2015).

Along with mechanosensitive ion channel activity MSL genes perform other functions also. Veley *et al.,* (2014) studied the MSL genes of *Arabidopsis* and found that the gene MSL10 has a regulated cell death signaling activity in addition to mechanosensitive ion channel activity. MSL8 is responsible to regulate the osmotic forces especially in the time of pollen hydration and germination. During the delivery of the male germ-line to the female gamete a dramatic change occurs in pollen grains due to presence of cellular water potential. MSL8 responds to the resulting mechanical stress and regulates pollen germination (Hamilton *et al.,* 2015). In *Arabidopsis*, the MSL proteins are found to control the shape and size of plastids. Plastids are cell organelles responsible for photosynthesis and innumerable metabolic reactions and processes. It includes the synthesis of various fatty acids and amino acids, gravity perception including starch production and their storage (Neuhaus and Emes 2000).

Mechanosensitive channels control the tension of the lipid bilayer; in the open state less lipids are present than the closed state (Rasmussen 2016). MscS has four transmembrane helices: TM1, TM2, TM3a and TM3b. TM1 and TM2 are packed together and arranged in an anti-parallel way roughly upright to the plane of the membrane. Although, TM3a lines (the central pore) and the location of TM3b is anticipated as the crossing point between lipids and cytosol (Pliotas *et al.,* 2015). Mechanosentive ion channels also occur in a sub-conducting or partly open state along with conducting (open) and non-conducting (closed) state (Shapovalov and Lester 2004; Edwards *et al.,* 2005; Anishkin and Sukharev 2009; Cox *et al.,* 2013). In the closed state the TM3a forms a hydrophobic seal by packing tightly (Bass *et al.,* 2002). The normal shape and size of the plastids are regulated by the MSL genes – MSL2 and MSL3 (Haswell and Meyerowitz 2006). Therefore, our aim of this study is to identify MSL genes in *C. arietinum* to understand the molecular mechanism of identified genes and also further analysis will help to select candidate genes to understand plant growth and development.

Materials and Methods

Genome wide identification of MSL genes in *C. arietinum*

The protein sequences of MSL genes from model plants (9 MSL genes in *A. thaliana* and 6 MSL genes in *O. sativa)* were downloaded from the TAIR (Berardini *et al.,* 2015) and TIGR (Kawahara *et al.,* 2013) databases respectively. We performed standalone BLAST analysis using blast 2.5.0+ NCBI version. Further, the downloaded sequences were used in the studies of *C. arietinum*, using BlastP and BlastX programs. For BlastP, the protein sequences and for BlastX the CDS nucleotide sequences of MSL genes of both model plants were used as query against *C. arietinum* database (created for standalone BLAST analysis). The resulting sequences were examined for complete ORF using the tool BESTORF. The sequences with complete ORF were taken for further analysis and those with incomplete ones were excluded. Six sequences with complete ORF were obtained. The protein sequences of these six genes were then aligned using MEGA7 (Kumar *et al.,* 2016).

Phylogenetic analysis and analysis of conserved transmembrane motifs

The protein sequences were aligned by MEGA7 program using Maximum-Likelihood (ML) method. The protein sequences were examined for conserved transmembrane motifs. The transmembrane motifs identification was done using MEME with the default parameters (Timothy *et al.,* 2009). The significance of the screened motifs was verified in NCBI and PROSITE (Sigrist *et al.,* 2012). Further, the motifs identified by MEME were analyzed for their function with the help of the tool GOMo (Timothy *et al.,* 2009).

Characteristics of the predicted proteins and their sub-cellular localization

The physico-chemical characteristics of the six MSL genes predicted in *C. arietinum* were calculated using the tool ProtParam (Gasteiger *et al.,* 2005). The parameters calculated were theoretical pI, amino acid composition and molecular weight. The trans-membrane regions of the MSL genes identified in *C. arietinum* using the tool Phyre^2 (Kelley *et al.,* 2015) and represented by drawing diagrams. On the other hand, the subcellular localization of the predicted MSL proteins were analyzed using LOCTREE3 (Goldberg *et al.,* 2014) and WoLF PSORT (Horton *et al.,* 2007) algorithms with default parameters.

Genome structure, Annotation and synteny of MSL genes in *C. arietinum* genome

The gene structure and the position of exons and introns were obtained using the webserver GenomeScan (Yeh *et al.,* 2001). The genomic DNA sequences in addition with the protein sequences of the MSL genes were used to find the gene structure, exons and introns using GenomeScan web server with default parameters. Additionally, the MSL genes were mapped in the *C. arietinum* chromosome with respect to their genome co-ordinates using MapChart 2.3v (Voorrips 2002) software. Further, we used SyMAP 4.2v (Soderlund *et al.,* 2006) software to produce a synteny map for the orthologous MSL genes of *C. arietinum* and *A. thaliana*.

3D structure analysis

I-TASSER server has been used for 3D structure generation. I-TASSER server generates information on the basis of fold recognition method (Yang *et al.,* 2008). The server provided detailed information regarding 3D structure, ligands binding sites, Enzyme Commission (EC) numbers, active sites, top homologous including GO template information. The 3D structure and other information of our six MSL genes predicted in *C. arietinum* were generated by uploading the sequences in the server.

Results

Genome wide identification as well as classification of MSL genes in the chickpea (*C. arietinum*) genome

For identification of MSL gene family in *C. arietinum* genome, MSL genes protein sequences of *O. sativa* and *A. thaliana* downloaded from TAIR and TIGR databases respectively. Further, these sequences were used as query for standalone BlastP analysis against *C. arietinum* protein database. To confirm the results obtained from BlastP analysis, BlastX was completed using the nucleotide sequences of the MSL genes against the same database. On the basis of highest bit score, lowest e-value and highest identity percentage, 6 MSL genes were obtained in *C. arietinum*. Phylogenetic analysis of the MSL gene family members from different plant species demonstrated that plant MSL genes generally clustered into two different classes a) class I and b) class II (Pivetti *et al.,* 2003). Remarkably, members of the class I family were very similar to the prokaryotic MscS channels while those of class II formed independent plant/fungi lineages (Haswell 2007). The MSL genes identified in *C. arietinum* genome were named as CaMSL1, CaMSL2, CaMSL3, CaMSL4, CaMSL5 and CaMSL6 based on the result obtained from phylogeny analysis (Fig. 3). The CaMSL genes were named according to the nearest member of *A. thaliana* and *O. sativa* MSL genes they resembled. The total number of MSL genes identified in *C. arietinum* and other plant species till date has been shown in Fig. 1.

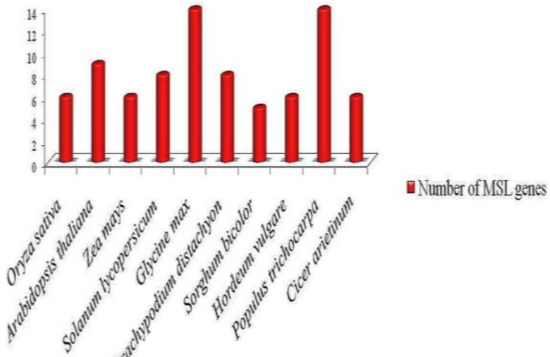

Fig. 1. Number of MSL genes present in different plant species.

Fig. 2. Alignment of MSL genes of *C. arietinum*, *A. thaliana* and *O. sativa*. (A) Class I MSL genes (B) Class II MSL genes

Analysis of evolutionary relationship and identification of conserved transmembrane motifs

A phylogenetic analysis confirmed the evolutionary relationship between the MSL gene family members of *C. arietinum* with the model plants *A. thaliana* and *O. sativa* using MEGA7 software (Kumar *et al.,* 2016). Phylogenetic analysis showed the MSL genes to be grouped into 2 groups, class I and class II. Both classes I and II contain 3 MSL genes each, CaMSL1, CaMSL2 and CaMSL3 in class I whereas CaMSL4, CaMSL5 and CaMSL6 in class II. From the previous studies it has been confirmed that class I members were seen to resemble with the prokaryotic MscS channels and class II formed independent plant/fungi lineages (Haswell 2007). The conserved regions of class I and class II MSL genes are shown in Fig. 2. Further, a phylogenetic tree was constructed among the MSL genes of *O. sativa*, *A. thaliana* and *C. arietinum* to study their evolutionary relationship using MEGA7 software (Fig. 3). Phylogenetic analysis of MSL genes confirmed similar evolutionary relationships with *C. arietinum* with *A. thaliana* and *O. sativa.* Between the two classes, class II organized the larger clade consisting of 12 members and class I consisting of 9 members (Souza *et al.,* 2016). Also, to identify the presence of motifs in CaMSL genes using MEME showed the presence of 10 distinct motifs in all genes (Fig. 4). Only motif 3 is present in all the CaMSL proteins. Moreover, analysis of the motifs using GOMo revealed motifs 2 and 7 to possess transcription factor activity, and motif 5 to act as a cellular component. Pfam analysis confirmed the presence of MS channel domains in all the CaMSL proteins (Table 1). In similar studies of *P. vulgaris* and *O. sativa* MSL genes too, the motifs showed the presence of casein kinase C and protein kinase C phosphorylation site (Saddhe and Kumar 2015).

Table 1. Consensus sequences and other details of the conserved motifs of CaMSL proteins

ID[a]	Consensus sequences	E-values	Sites	Width
1	FGNTAKTIFEAIIFVFVMHPFDVGDRCVIDGVQMTVEEMNI	6.2e-040	3	41
2	WCMLVMVTFCGMLVTKWFMHIVVFLIEMNFLLKKKVLYFVH	5.5e-030	5	41
3	TTIFLRYDNEKIFYPNSVLATKPISNFYRSPEMSDLVEFA	4.6e-031	6	40
4	TLLLKILASNFHVKSFFDRIQESIFHQYLLQTLSGPPLME	1.2e-024	5	40
5	YLESRPQHWRPNHNVIVKDIENVNKMKMALNVTHTINFQNF	1.9e-023	5	41
6	FAYCLSSLIQQAQKFFLEANDNSDARNMGLDFAGKAVYTAVWVAAVSLFM	2.2e-019	2	50
7	MRKVLAKNPQVEQQKLHRRVFLEDISPENQALKILISCFVKTSHFEEYLC	1.1e-018	2	50
8	MNLLPRTLKQDSLALHLLSGVHAPIRHAPSRYSVFVCRSAMIPGGGSGTA	4.2e-017	2	50
9	YEEEIDQRDNEITSEWEAKAAAYRIFRNVAKPGSKYIEKED	1.2e-017	5	41
10	GNEEKDAKIDEPVASDSKSDDNFAGTSTSPHGVNSKDKSKSNSDAQIQNM	3.3e-014	2	50

[a] represents motif ID

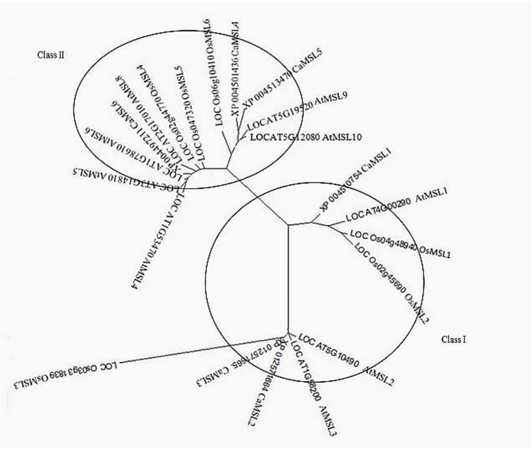

Figure 3. Evolutionary relationship of *C. arietinum* with respect to *A. thaliana* and *O. sativa* MSL genes.

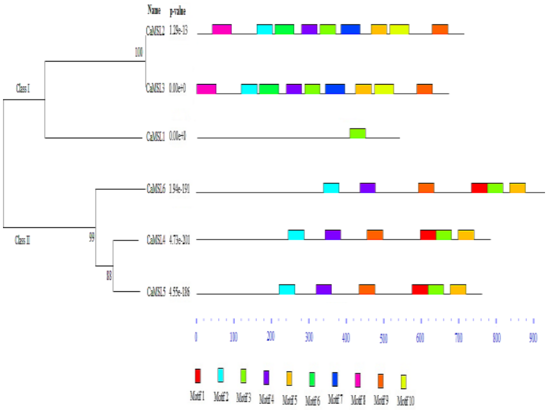

Figure 4. Schematic representation of the motifs identified in *C. arietinum* MSL genes. The motifs identified are shown in coloured boxes.

Physico-chemical characteristics and subcellular localization of the predicted MSL proteins

The physio-chemical characteristics of the MSL genes predicted in *C. arietinum* were calculated using ProtParam tool (Table 2). This illustrates the physio-chemical properties of the CaMSL genes. The results showed that the amino acid length where the lowest one was 298 with pI 7.65 (CaMSL1) and the highest was 708 with pI 9.45 (CaMSL2), with highest composition of Serine and Leucine respectively. Similarly, the molecular weight of the CaMSL1 was 33.13 and for CaMSL2 was 78.24 kDa. CaMSL2 was found to have the highest molecular weight and amino acid length but lowest pI among all the MSL proteins. Class I MSL1 protein showed smallest amino acid size and lowest molecular weight compared to the class II MSL proteins. Similar results were obtained in *P. vulgaris* and *O. sativa* MSL proteins too (Souza *et al.,* 2016). For predicting the subcellular localization of the MSL genes two online tools LOCTREE3 and WoLF PSORT were used. All class II MSL genes are localized in the plasma membrane, while the class I MSL genes, CaMSL1 is located in the mitochondrion and the other two genes CaMSL2 and CaMSL3 in the chloroplast membrane. Studies in *O. sativa* MSL gene families showed class I genes (OsMSL1, OsMSL2 and OsMSL3) to be localized either in the chloroplast membrane or else in mitochondria and class II genes (OsMSL4, OsMSL5 and OsMSL6) in the plasma membrane (Haswell 2007). Some other studies in *A. thaliana* MSL genes revealed class I genes localized either in the plastid or else in mitochondria and class II genes in the plasma membrane (Haswell 2007; Saddhe and Kumar 2015).

Table 2. Physiochemical characteristics of MSL genes of *C. arietinum*

Class	CaMSL	Gene ID	Nucleotide CDS bp	Amino acid	pI	Mol. Wt. (kDa)	Subcellular localization
I	**CaMSL1**	**XP_004510754.1**	**894**	**298**	**9.40**	**33.13**	**Mitochondrion**
	CaMSL2	**XP_012571664.1**	**2124**	**708**	**9.45**	**78.24**	**Chloroplast membrane**
	CaMSL3	XP_012571665.1	2007	669	9.28	73.74	Chloroplast membrane
II	CaMSL4	XP_004510436.1	1464	488	9.36	55.25	Plasma membrane
	CaMSL5	XP_004513470.1	1395	465	9.22	53.36	Plasma membrane
	CaMSL6	XP_004497211.1	1872	624	7.65	71.24	Plasma membrane

Identification of Transmembrane helix region

The topology of CaMSL genes were predicted using Phyre^2 (Fig. 5). In class I, CaMSL1 gene revealed the presence of three transmembrane (TM) regions whereas CaMSL2 and CaMSL3 revealed to have five TM regions each. All members of class II confirmed the availability of four TM regions. On the other hand, in *A. thaliana* and *O. sativa*, five TM domains were present in class I members and six TM domains in class II gene family members (Haswell 2007; Saddhe and Kumar 2015). These channels sense and respond to different variations in membrane tension intrinsic to plant life cycle, thereby affecting the development of plants (Hamilton *et al.,* 2015).

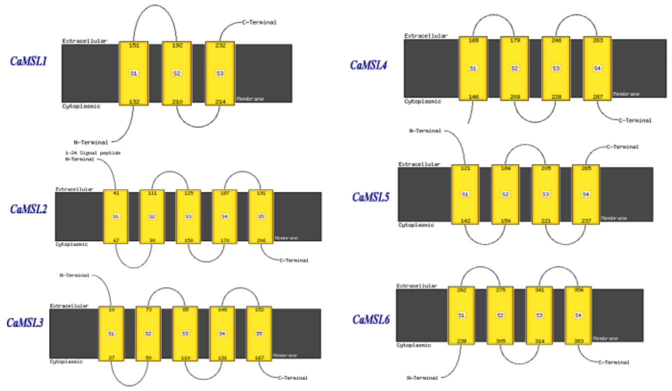

Figure 5. Diagram is showing that the transmembrane (TM) helix topology pointing the linear coordinates for the helices. This diagram is demonstrating the protein's extra-and intercellular regions. Color indicators: Yellow boxes = TM region (labeled as S1, S2, S3, S4 and S5); Black boxes = Membrane regions; Black single line = N and C terminal regions.

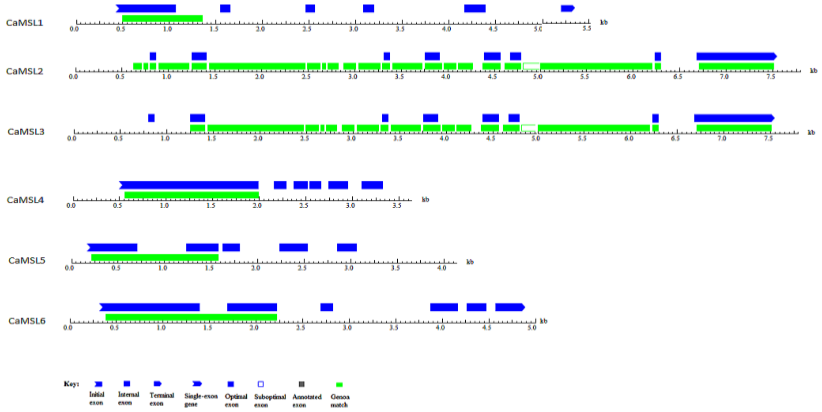

Figure 6. Schematic diagram for the gene structure of CaMSL genes.

Genome structure, annotation and synteny of CaMSL genes

The gene structure of the CaMSL genes was analyzed with the help of GenomeScan webserver. The organization of exon/intron of the *C. arietinum* MSL genes are illustrated in Fig. 6. All CaMSL genes showed the presence of exons, class I MSL genes contained 6-8 exons while class II members contained 5-6 exons. Highest numbers of exons were seen in CaMSL2 and CaMSL3. In *A. thaliana* and *O. sativa* the number, size of the intron and the position were very nearly preserved within classes (Saddhe and Kumar 2015).

The CaMSL genes were plotted in the *C. arietinum* chromosomes according to their genome co-ordinates using MapChart 2.3v (Voorrips *et al.,* 2002) software. On the basis of available sequences in NCBI, their BLAST and phylogenetic analysis the chromosomal locations of CaMSLs were determined. The MSL genes are located in four different chromosomes, out of which chromosome 4 contained three MSL genes and the other three chromosomes contained single MSL gene (Fig. 7). In similar study of *P. vulgaris*, PvMSL genes were distributed randomly in 5 out of 11 chromosomes, with highest number of MSL genes in chromosome 9 (Souza *et al.,* 2016). Similar studies in rice revealed rice MSL genes to be distributed in chromosomes 2, 3, 4 and 6 (Saddhe and Kumar 2015). Gene family expansions in plants have been suggested to be mainly due to tandem and segmental duplications (Cannon *et al.*, 2004). If more than two genes are located in the same chromosome it is a sort of confirmation of tandem duplication event (Cai *et al.,* 2013) whereas gene duplication between different chromosomes labeled as a segmental duplication event (Liu *et al.,* 2011).

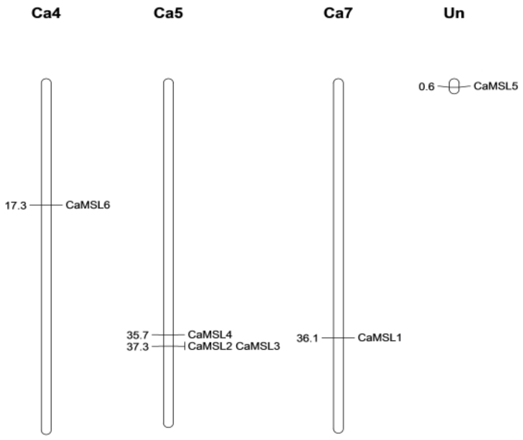

Figure 7. Chromosomal location of CaMSL genes. The CaMSL genes were mapped according to their genome co-ordinates.

In our research work, a comparative syntenic map of *C. arietinum* related with the model plant i.e. *A. thaliana* (Fig. 7) was built. Remarkably, we got a three pairs of syntenic relations i.e. AtMSL genes (AtMSL1, AtMSL4 and AtMSL5) and all CaMSL genes (CaMSL1, CaMSL2, CaMSL3, CaMSL4, CaMSL5 and CaMSL6).

Discussion

As, it has been proved by various research that the mechanosensitive channels mostly controlled the gravity sensing, pollen tube growth including the shape and size of plastid in plants. In our study, we performed in silico analysis for the identification of the mechanosensitive channel of small conductance like (MscS) gene family in *C. arietinum* (chickpea). In our result we found six *C. arietinum* MscS-like (CaMSL) genes. Similar work related to evolutionary relationship of the MSL gene family members from different plant species revealed that plant MSL genes generally clustered into two different classes i.e. class I and class II (Pivetti *et al.,* 2003). In our work six genes were found to be related with synteny events. These synteny events indicate that a number of MSL genes ascended in the past as a sustainable divergence in *Arabidopsis* and chickpea lineages. As a consequence, our result may indicate that CaMSL genes in *C. arietinum* share similar function as well as structure with *Arabidopsis* MSL genes (Figure S1).

Table S1: Top models different scores

MSL gene	TM-score	RMSD^a	IDEN^a	Cov
CaMSL1	0.5491	4.4	0.08	0.73
CaMSL2	0.2697	8.31	0.03	0.43
CaMSL3	0.2871	8.34	0.04	0.48
CaMSL4	0.4766	5.6	0.04	0.65
CaMSL5	0.4994	5.68	0.06	0.69
CaMSL6	0.4584	5.93	0.07	0.62

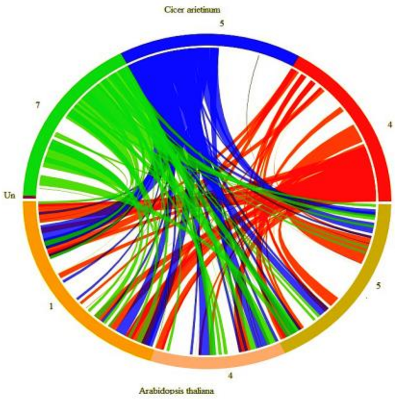

Figure S1. This synteny map represents the orthologous gene positions of MSL genes with respect to *C. arietinum* and *A. thaliana*. Synteny map demonstrate the highly conserved synteny between *C. arietinum* (Ca04, Ca05. Ca07 and CaUn) and *Arabidopsis* chromosomes (At02, At04 and At05). In the above diagram each block represents individual chromosome whereas the connecting red lines between two chromosomal regions represent syntenic regions of genomes.

3D structure analysis using I-TASSER

As, I-TASSER provided top 10 threading templates, top 5 models of structures, top 10 identified structural analog, top 5 PDB hit on ligand binding sites, top 5 enzyme commission number and active sites including top 10 GO terms. For our analysis we have taken the best top hits only (Table S1 and S2). In each case only the sequence with the highest C-score value was selected (Table 3). Actually, C-Score is the confidence score calculated on the basis of threading template alignments to estimate the quality of predicted models. Similar work was carried out by Mishra *et al.,* (2005), where he built 3D structure with the help of I-TASSER. 3D structure labeled using Chimera; a similar type of work on SsMYB2R protein has been done where they used I-TASSER and Pymol for molecular graphics which help to determine the binding specificity (Kulkarni *et al.*, 2014).

Table S2. GO term for all six MSL genes

MSL gene	Molecular function	Biological process	Cellular component
CaMSL1	GO:0005488, GO:0008565	GO:0050658, GO:0051168, GO:0006605	GO:0031967, GO:0044444, GO:0000775, GO:0032991, GO:0042175, GO:0016604
CaMSL2	GO:0003899, GO:0003677, GO:0005524, GO:0008381, GO:0019843, GO:0004642, GO:0003746, GO:0005515, GO:0016887	GO:0006351, GO:0009992, GO:0055085, GO:0006811, GO:0006541, GO:0006414, GO:0006189, GO:0006415, GO:0006200	GO:0005886, GO:0016021, GO:0022626
CaMSL3	No information available	No information available	No information available
CaMSL4	GO:0022892, GO:0005515	GO:0051168, GO:0006605, GO:0016458	GO:0031981, GO:0043232, GO:0044444
CaMSL5	GO:0005515, GO:0022892	GO:0046907, GO:0034613, GO:0045184	GO:0005634
CaMSL6	GO:0008565, GO:0005515	GO:0051168, GO:0006605	GO:0044444, GO:0032991, GO:0048475, GO:0044425, GO:0005634

Table 3. Best model predicted by I-TASSER with respect to their ligand and their binding sites

MSL gene	C-score	Cluster Size	PDB hit	Ligand Name	Ligand Binding Site Residues
CaMSL1	0.12	8	4xriA	SO4	248, 249, 250, 253
CaMSL2	0.03	1	1izlM	CLA	210, 217
CaMSL3	0.06	2	3cmvG	ANP	337, 340, 341, 342, 343, 377
CaMSL4	0.07	3	3jcuH	CLA	336, 343
CaMSL5	0.1	6	2h6bA	3C4	271, 274
CaMSL6	0.07	5	4bntD	36E	351, 355

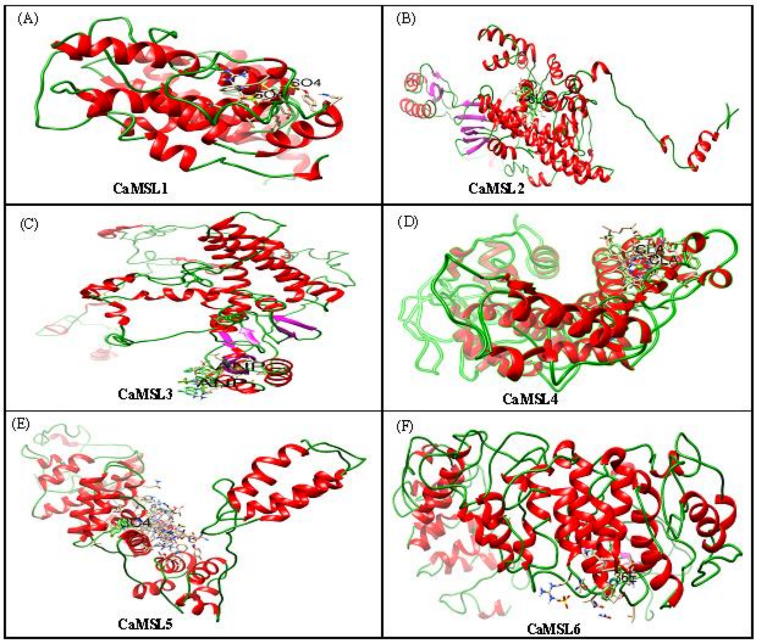

Figure 8. Ligand binding sites of the six CaMSL genes predicted by ITASSER.

There is no information available for GO term of CaMSL3 gene. The molecular 3D modeling of Tomato beta—acetyl 12 hexosaminidase (-Hex SI) using I-TASSER. SPDB viewer and Chimera were used to prepare the images (Hossain *et al.,* 2014). Proteins interact with other molecules, called ligands to perform their biological functions. For understanding their functions identification of specific ligand binding site of the protein molecule is the first important step (Greer *et al.,* 1994). The ligand binding sites of the six predicted CaMSL genes and their ligands are represented in Fig. 8.

From the results obtained from the present study, we can conclude that the MSL gene family in *C. arietinum* consists of 6 genes (CaMSL1, CaMSL2, CaMSL3, CaMSL4, CaMSL5 and CaMSL6), grouped into Class I and Class II. Subcellular localization analysis concluded class I CaMSL genes are located either in the mitochondrion or chloroplast membrane and all class II CaMSL gene family members in the plasma membrane. Topological analysis of the CaMSL genes revealed the existence of 3 transmembrane regions in CaMSL1 and 5 in CaMSL2 and CaMSL3. Class II MSL gene family members contained 4 transmembrane regions. Also, the 6 CaMSL genes were present in 4 different chromosomes, with the highest number of MSL genes in Ca5. Therefore, the data presented in our research study will be significant in selecting candidate genes that play important role in regulating plant growth and their development.

Conclusion

Mechanosensitive channels are proposed to play major roles in the adaptation to osmotic stress, most importantly in the management of transitions from high to low osmolarity environments. Our study is one of the first studies of MSL family genes in *C. arietinum* using bioinformatics approach. CaMSL genes have an essential role in the plant growth and their development. Our analysis will help to increase the understanding of the genetic bases of mechanosensitive ion channels. This analysis will also pave a new pathway to select top quality candidate genes for future structural and functional analyses of CaMSL family in chickpea. Ultimately, it will increase the opportunity to get genetically engineered novel traits which may have great importance in the field of agriculture and food production.

Acknowledgement

We gratefully acknowledge the Department of Biotechnology, Government of India sponsored Bioinformatics Infrastructure Facility for providing traineeship to Noymi Basumatary. All authors have contributed equally to this work.

References

1. Agriculture farming- the information related to agricultural farming in India. Available at: http://www.agrifarming.in/chickpea-farming, Accessed 8 June 2017.

2. Anishkin, Andriy, and Sergei Sukharev. "State-stabilizing interactions in bacterial mechanosensitive channel gating and adaptation." *Journal of Biological Chemistry* 284.29 (2009): 19153-19157. Print.

3. Bass, Randal B, Pavel Strop, Margaret Barclay, and Douglas C Rees. "Crystal structure of *Escherichia coli* MscS, a voltage-modulated and mechanosensitive channel." *Science* 298.5598 (2002): 1582–1587. Print.

4. Berardini, Tanya Z, Leonore Reiser, Donghui Li, Yarik Mezheritsky, Robert Muller, Emily Strait, and Eva Huala. "The *Arabidopsis* Information Resource: Making and mining the "gold standard" annotated reference plant genome." *Genesis* 53.8 (2015): 474-485. Print.

5. Cai, Xiaofeng, Yuyang Zhang, Chanjuan Zhang, Tingyan Zhang, Tixu Hu, Jie Ye, Junhong Zhang, Taotao Wang, Hanxia Li, and Zhibiao Ye. "Genome-wide analysis of plant-specific Dof transcription factor family in tomato." *Journal of Integrative Plant Biology* 55.6 (2013): 552-566. Print.

6. Cannon, Steven B, Arvind Mitra, Andrew Baumgarten, Nevin D Young, and Georgiana May. "The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana." BMC Plant Biology* 4: 10. Print.

7. Carlos, López-Larrea. "Sensing in Nature." In: Advances in Experimental Medicine and Biology, Carlos, López-Larrea (Ed.). Springer, Germany. (2011).

8. Chang, Geoffrey, Robert H Spencer, Allen T Lee, Margaret T Barclay, and Douglas C Rees. "Structure of the MsCL homolog from *Mycobacterium tuberculosis*: a gated mechanosensitive ion channel." *Science* 282.5397 (1998): 2220-2226. Print.

9. Commodities control- all information related to chickpea production in India with respect to their state. Available at: http://www.commoditiescontrol.com/eagritrader/common/ newsdetail.php?type=SPR&itemid=8204&cid1=,2,&varietyid=,33, Accessed 9 June 2017

10. Cox, Charles D, Yoshitaka Nakayama, Takeshi Nomura, and Boris Martinac. "The evolutionary 'tinkering' of MscS-like channels: generation of structural and functional diversity." *Pflugers Arch- European Journal of Physiology* 467.1 (2015): 3-13. Print.

11. Cox, Charles D, Toshifumi Nomura, Chandra Ziegler, Anthony Campbell, Kenneth T Wann, and Boris Martinac. "Selectivity mechanism of the mechanosensitive channel MscS revealed by probing channel subconducting states." *Nature Communications* 4. (2013): 1-11. Print.

12. Edwards, Michelle D, Susan Black, Tim Rasmussen, Akiko Rasmussen, Neil R Stokes, Terri-Leigh Stephen, Samantha Miller, and Ian R Booth. "Characterization of three novel mechanosensitive channel activities in *Escherichia coli*." *Channels* 6.4 (2012): 272-281. Print.

13. Edwards, Michelle, Yuezhou Li, Sanguk Kim, Samantha Miller, Wendy Bartlett, Susan Black, Sally Dennison, Irene Iscla, Paul Blount, James U Bowie, and Ian R Booth. "Pivotal role of the glycine-rich TM3 helix in gating the MscS mechanosensitive channel." *Nature Structural & Molecular Biology* 12, 113–119. Print.

14. FAO- the information related to Food and agricultural organization of united nation. Available at: http://www.fao.org/3/a-i6212e.pdf, Accessed 8 June 2017

15. Fasano, Jeremiah M, Gioia D Massa, and Simon Gilroy. "Ionic signaling in plant responses to gravity and touch." *Journal of Plant Growth Regulation* 21 (2002): 71-88. Print.

16. Gustier, Elisabeth, Christine Hoogland, Alexandre Gattiker, S'verine Duvaud, Marc R Wilkins, Ron D Appel, and Amos Bairoch. "Protein Identification and Analysis Tools on the ExPASy Server." In: The Proteomics Protocols Handbook, John M. Walker (Ed.). Humana Press, Switzerland, (2005) 571-607. Print.

17. Goldberg, Tatyana, Maximilian Hecht, Tobias Hamp, Timithy Karl, Guy Yachdav, Nadeem Ahmed, Uwe Altermann, Philipp Angerer, Sonja Ansorge, Kinga Balasz, Michael Bernhofer, Alexander Betz, Laura Cizmadija, Kieu Trinh Do, Julia Gerke, Robert Greil, Vadim Joerdens, Maximilian Hastreiter,

Katharina Hembach, Max Herzog, Maria Kalemanov, Michael Kluge, Alice Meier, Hassan Nasir, Ulrich Neumaier, Verena Prade, Jonas Reeb, Aleksandr Sorokoumov, Ilira Troshani, Susann Vorberg, Sonja Waldraff, Jonas Zierer, Henrik Nielsen, and Burkhard Rost. "LocTree3 prediction of localization." *Nucleic Acids Research* 42(W1). (2014): W350-W355. Print.

18. Gottlieb, Philip A, and Frederick Sachs. "The sensation of stretch." *Nature* 483.7388 (2012): 163–164. Print.

19. Greer, Jonathan, John W Erickson, John J Baldwin, and Michael D Varney. "Application of the 3-dimensional structures of protein target molecules in structure-based drug design." *Journal of Medicinal Chemistry* 37.8 (1994): 1035–1054. Print.

20. Gupta, Sonal, Kashif Nawaz, Sabiha Parween, Riti Roy, Kamlesh Sahu, Anil Kumar Pole, Hitaishi Khandal, Rishi Srivastava, Swarup Kumar Parida, and Debasis Chattopadhyay. "Draft genome sequence of *Cicer reticulatum* L., the wild progenitor of chickpea provides a resource for agronomic trait improvement." *Dna Research* 24.1 (2016): 1-10. Print.

21. Hamilton, Eric S, Gregory S Jensen, Grigory Maksaev, Andrew Katims, Ashley M Sherp, and Elizabeth S Haswell. "Mechanosensitive channel MSL8 regulates osmotic forces during pollen hydration and germination." *Science* 350.6259 (2015): 438-441. Print.

22. Haswell, Elizabeth S. "MscS-Like proteins in plants." *Current Topics in Membranes* 58.1 (2007): 329-359. Print.

23. Haswell, Elizabeth S, and Elliot M Meyerowitz. "MscS-like proteins control plastid size and shape in *Arabidopsis thaliana*." *Current Biology* 16.1 (2006): 1-11. Print.

24. Haswell, Elizabeth S, Rob Phillips, and Douglas C Rees. "Mechanosensitive channels: what can they do and how do they do it?" *Structure* 19.10 (2011): 1356–1369. Print.

25. Horton, Paul, Keun-Joon Park, Takeshi Obayashi, Naoya Fujita, Hajime Harada, CJ Adams-Collier, and Kenta Nakai. "WoLF PSORT: protein localization predictor." *Nucleic Acids Research*, 35.Web Server issue (2007): 585–587. Print.

26. Hossain, Md. Anwar, and Hairul Azman Roslan. "Molecular Phylogeny and Predicted 3D Structure of Plant Beta-D-N-Acetylhexosaminidase." *The Scientific World Journal* l14. (2014):1-14. Print.

27. Jayaraman, Dhileepkumar, Simon Gilroy, and Jean-Michel Ané. "Staying in touch: mechanical signals in plant-microbe interactions." *Current Opinion in Plant Biology* 20. (2014): 104-109. Print.

28. Kawahara, Yoshihiro, Melissa de la Bastide, John P Hamilton, Hiroyuki Kanamori, W Richard McCombie, Shu Ouyang, David C Schwartz, Tsuyoshi Tanaka, Jianzhong Wu, Shiguo Zhou, Kevin L Childs, Rebecca M Davidson, Haining Lin, Lina Quesada-Ocampo, Brieanne Vaillancourt, Hiroaki Sakai, Sung Shin Lee, Jungsok Kim, Hisataka Numa, Takeshi Itoh, C Robin Buell, and Takashi Matsumoto. "Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data." *Rice* 6.4 (2013): 1-10. Print.

29. Kelley, Lawrence A, Stefans Mezulis, Christopher M Yates, Mark N Wass, and Michael JE Sternberg. "The Phyre2 web portal for protein modeling, prediction and analysis." *Nature Protocols* 10.6 (2015): 845-858. Print.

30. Kulkarni, Pranali A, and Rachayya M Devarumath. "In silico 3D-structure prediction of SsMYB2R: A novel MYB transcription factor from *Saccharum sponteneum*." *Indian Journal of Biotechnology* 13.4 (2014): 437-447. Print.

31. Kumar, Sudhir, Glen Stecher, and Koichiro Tamura K. "MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets." *Molecular Biology and Evolution* 33.7 (2016): 1870-1874. Print.

32. Liu, Yan, HaiYang Jiang, Wenjuan Chen, Yexiong Qian, Qing Ma, Beijiu Cheng, and Suwen Zhu. "Genome-wide analysis of the auxin response factor (ARF) gene family in maize (*Zea mays*)." *Plant Growth Regulation* 63.3 (2011): 225-234. Print.

33. Lucas, Mikael, Kim Kenobi, Daniel von Wangenheim, Ute Voβ, Kamal Swarup, Ive De Smet, Daniel Van Damme, Tara Lawrence, Benjamin Péret, Eric Moscardi, Daniel Barbeau, Christophe Godin, David Salt, Soazig Guyomarc'h, Enrst HK Stelzer, Alexis Maizel, Laurent Laplaze, and Malcolm J Bennett. "Lateral root morphogenesis is dependent on the mechanical properties of the overlaying tissues." *Proceedings of the National Academy of Sciences* 110.13 (2013): 5229-5234. Print.

34. Mishra, Anil K, Gitanjali Tandon, Rajendra Sharma, Hariharan Chandrasekharan, and

Punyavrat S Pandey. "In silico structural and functional analysis of protein encoded by Whet early salt-stress response gene (WESR3)." *Indian Journal of Biochemistry and Biophysics* 52. (2015): 97-100. Print.

35. Monshausen, Gabriele B, and Elizabeth S Haswell. "A force of nature: Molecular mechanisms of mechanoperception in plants." *Journal of Experimental Botany* 64.15 (2013): 4663-4680. Print.

36. Neuhaus, Ekkehard, and Michael J Emes. "Nonphotosynthetic metabolism in plastids." *Annual Review of Plant Physiology and Plant Molecular Biology* 51.1 (2000): 111-140. Print.

37. ORF finder-all information related to ORF. Available at: http://www.ncbi.nlm.nih.gov/gorf/ gorf.html, Accessed 10 May 2017

38. Perozo, Eduardo. "Gating prokaryotic mechanosensitive channels." *Nature Reviews* 7.2 (2006): 109-119. Print.

39. Pivetti, Christopher D, Ming-Ren Yen, Samantha Miller, Wolfgang Busch, Yi-Hsiung Tseng, Ian R Booth, and Milton H Saier Jr. "Two families of mechanosensitive channel proteins." *Microbioliology and Molecular Biology Reviews* 67.1 (2003): 66-85. Print.

40. Pliotas, Christos, A Caroline E Dahl, Tim Rasmussen, Kozhinjampara R Mahendran, Terry K Smith, Phedra Marius, Joseph Gault, Thandiwe Banda, Akiko Rasmussen, Samantha Miller, Carol V Robinson, Hagan Bayley, Mark SP Sansom, Ian R Booth, and James H Naismith. "The role of lipids in mechanosensation." *Nature Structure and Molecular Biology* 22.12 (2015): 991-998. Print.

41. Rasmussen, Tim. "How do mechanosensitive channels sense membrane tension?" *Biochemical Society Transactions* 44.4 (2016): 1019-1025. Print.

42. Saddhe, Ankush, and Kundan Kumar. "In silico identification and expression analysis of MscS like gene family in rice." *Plant Gene* 1. (2015): 8-17. Print.

43. Shapovalov, George, and Henry A Lester. "Gating transitions in bacterial ion channels measured at 3 microns resolution." *The Journal of General Physiology* 124.2 (2004): 151-161. Print.

44. Sigrist, Christian J, Edouard de Castro, Lorenzo Cerutti, Beatrice A Cuche, Nicolas Hulo, Alan Bridge, Lydie Bougueleret, and Ioannis Xenarios. "New and continuing developments at PROSITE." *Nucleic Acids Research* 41. Database issue (2012): 344-347. Print.

45. Soderlund, Carol, William Nelson, Austin Shoemaker, and Andrew Paterson. "SyMAP: A system for discovering and viewing syntenic regions of FPC maps." *Genome Research* 16.8 (2006): 1159-1168. Print.

46. Souza, Silvia Graciele Hülse de, Adriana Pereira da Silva, Tania Mayum Ito, Eduardo Goiano da Silva, and Talita Cantú. "Genome-wide analysis of mechanosensitive channel of small conductance (MscS)-like gene family in common bean." *African Journal of Biotechnology* 15.15 (2016): 580-592. Print.

47. Sukharev, Sergei, and Frederick Sachs. "Molecular Force Transduction by Ion Channels: diversity and unifying principles." *Journal of Cell Science* 125.13 (2012): 3075-3083. Print.

48. Bailey, Timothy L, Mikael Bodén, Fabian A Buske, Martin Frith, Charles E Grant, Luca Clementi, Jingyuan Ren, Wilfred W Li, and William S Noble. "MEME SUITE: tools for motif discovery and searching." *Nucleic Acids Research* 37.Web server issue (2009): 202–208. Print.

49. Uniprot-all information related to proteins. Available at: www.uniprot.org, Accessed 11 May, 2017

50. Veley, Kira M, Grigory Maksaev, Elizabeth M Frick, Emma January, Sarah C Kloepper, and Elizabeth S Haswel. "*Arabidopsis* MSL10 Has a Regulated Cell Death Signaling Activity That Is Separable from Its Mechanosensitive Ion Channel Activity." *The Plant Cell* 26.7 (2014): 3115–3131. Print.

51. Voorrips, Roeland. "MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs." *Journal of Heredity* 93.1 (2002): 77-78. Print.

52. Wang, Wenjian, Susan S Black, Michelle D Edwards, Samantha Miller, Emma L Morrison, Wendy Bartlett, Changjiang Dong, James H Naismith, and Ian R Booth. "The structure of an open form of an *E. coli* mechanosensitive channel at 3.45 Å resolution." *Science* 321.5893 (2008): 1179-1183. Print.

53. Yang, Yufeng, Yingshi Ouyang, Lichuan Yang, M Flint Beal, Angus McQuibban, Hannes Vogel, and Bingwei Lu. "Pink1 regulates mitochondrial dynamics through interaction with the fission/fusion machinery." *Proceedings*

of the National Academy of Sciences of the United States 105.19 (2008): 7070--7075. Print.

54. Yeh, Ru-Fang, Lee P Lim, and Christopher B Burge. “Computational inference of homologous gene structures in the human genome.” *Genome Research* 11.5 (2001): 803-816. Print.

Cite this article as:

Noymi Basumatary, Archana Kumari, Jatin Sarmah. Genome-wide identification of MscS like gene family in *Cicer arietinum* using bioinformatics approach. *Annals of Plant Sciences 7.2 (2018) pp. 2002-2012.*

http://dx.doi.org/10.21746/aps.2018.7.2.4

Source of support: Department of Biotechnology, Government of India

Conflict of interest: Nil