(RESEARCH ARTICLE)

Check for updates

Genome-wide characterization of Major Intrinsic Proteins (MIPs) family in *Momordica charantia*

Shifa Begum [1, *] and Ruhul A. Khan [2]

[1] Department of Genetic Engineering and Biotechnology, University of Chittagong, Chittagong-4331, Bangladesh.
[2] Institute of Radiation and Polymer Technology, Atomic Energy Research Establishment, Savar, Dhaka, Bangladesh.

Abstract

Major intrinsic proteins (MIPs) family include the group of proteins also called channel proteins or water channels. Aquaporins facilitate the transport of water and other uncharged small molecules, forming pores in cellular boundaries which help in plant growth, development, and sometimes in plant adaptation. They usually have six tran*Sm*embrane alpha-helical domains with conserved amino acids residues along with hydrophobic and hydrophilic regions. Though aquaporins were found both in animals and plants, also in bacteria, fungi cells, it comprises the largest and most complex group in plant species. A little is known about the aquaporins family of the plant *Momordica charantia*. About 25 aquaporins of its are characterized in this study. The aquaporins of *M. charantia* are divided into five sub-families here, they are PIPs, TIPs, XIPs, SIPs, and NIPs. There are twelve PIPs, three TIPs, eight NIPs, and one AQPs in both XIP and SIP. In the XIP subfamily, it contains SPI other than the NPA motif. Moreover, SIP and NIP subfamily also contain different motifs both in loop B and loop E other than NPA motifs. *Mc*SIPs contain NPT in loop B and *Mc*NIPs contain NPS in loop B (*Mc*NIP1; 1, *Mc*NIP1; 3 and *Mc*NIP3; 1). Again, *Mc*NIPs contain NPV in loop E (*Mc*NIP2; 2, *Mc*NIP3; 1). The conserved residues and motifs of this plant show its transportation of water and the possibility of other molecules' transportation through those channel proteins.

Keywords: Intrinsic protein; Aquaporin; Amino acid; Protein channel

1. Introduction

Aquaporins are small proteins (~30kDa) that locate in intracellular membranes or cytoplasmic membranes, function as channel proteins in the different types of cells as they form pores. These channel proteins help to transport fluids like water and other small polar molecules [1]. These are homologous water-channel proteins, found in both plants and animals. They are water selective and/or pass small neutral solutes (urea, boric acid, silicic acid) or gases (ammonia, carbon dioxide) in the plant cells but ions or other metabolites cannot pass through the channels [2]. These channel proteins are providing an important role in plant and animal cells to facilitate and regulate the passage of water through their membranes [3]. The MIP family members (Aquaporins are under the superfamily of MIP) generally have a molecular weight of 23–31 kDa, six transmembrane regions, and cytosolic N- and C-termini [4]. Five loops (A to E) connect six membrane-spanning alpha-helices. Two short hydrophobic alpha-helices are on the opposite sides in the loops B and E, they are highly conserved NPA (Asn- Pro-Ala) motifs [5].

The aromatic arginine (ar/Ar) regions are found in helix 2 and helix 5 and the loop E (E1 and E2), generally in the exoplasmic C-terminus. The constrictions of NPA and ar/R regions determine substrate permeability. Furthermore, Froger's residues (Q/M, S, A, F, W) were found to be associated with substrate specificity [6]. Aquaporins are found in many organisms but they are not as large a number as plant's AQPs, such as in mammalsitis 13 and less than this number

---

in other organisms [7]. While there are fifty-five AQPs in *Populus trichocarpa*[8], sixty-six in *Glycine max* [9], forty-seven in *Solanum lycopersicum* [10], etc. Plant AQPs are generally divided into five major sub-families: plasma membrane intrinsic proteins (PIPs), nodulin 26-like intrinsic proteins (NIPs), tonoplast intrinsic proteins (TIPs), small intrinsic proteins (SIPs), and uncharacterized intrinsic proteins (XIPs) [11].

M. charantia is a plant of the *Cucurbitaceae* family that grows well in hot and humid weather and is mostly found in Asia and Africa. It has many varieties depending on its bitterness. From the very ancient time, it is popular for its medicinal uses such as in diabetes and its complications, as an antibacterial and antiviral agent, etc. Recent studies show its efficacy in cancer treatment [12].

Little is known about the AQPs of *M. charantia.* Here 25*Mc*AQPs are assigned and analyzed, grouped into five subfamilies (twelve PIPs, three TIPs, one XIPs and SIPs, eight NIPs). The groups of aquaporins were classified and characterized based on phylogeny analyses and conserved resides. Moreover, gene properties of *M. charantia* AQPs are also included here.

2. Materials and methods

2.1. Data Retrieving Sources

The whole-genome sequences of *M. charantia* were available in NCBI [13] and downloaded from there. A 35 *At*AQPs sequences of the model plant *Arabidopsis thaliana* were collected from Johanson et al. (2001)[14] and 30 AQPs sequences of *Oryza sativa* were collected from Sakurai et al. (2005) [15] and both *At*AQPs and *Os*AQPs sequences were separately used as queries in the tblastn [16] to search *Mc*AQPs. Further, the whole protein sequences of *M. charantia* and *At*AQPs sequences were used in BLASTp tools for searching similarity to find out *Mc*AQPs.

2.2. Characterization of Major Intrinsic Proteins (MIPs)

The retrieved sequences of predicted AQPs of *M. charantia* were submitted to the Conserved Domains Database (CDD) for CD search and to verify AQP domains [17]. Sequences encoding in the family MIP include two NPA motifs with an amphipathic channel which indicates both the region of hydrophilic and hydrophobic were primarily selected as aquaporin. For further analysis, the AQP sequences were submitted to TMHMM Server v.2.0 [18] to assess the trans-membrane helical domain. Again, the sequences were submitted to SOSUI [19] to detect the secondary structure along with the transmembrane domain. Additionally, using ExPASyProtParam [20], molecular weight, iso-electric point (pI), and amino acid length of *Mc*AQPs were calculated. Protein transmembrane helix probability for each sequence was also checked from SOSUI results.

2.3. Sequence alignment and phylogenetic analyses

Aquaporin sequences of *M. charantia* were aligned with *At*AQPs from Arabidopsis (*Arabidopsis thaliana*) [21], *Os*AQPs from *Oryza sativa* [22], *Cs*AQPs from *Cucumis sativus* [23], and *Pp*AQPs from *Physcomitrella patens* [24] using Clustal Omega [25]. The Phylogenetic tree was constructed using Clustal Omega and further illustrated in the Interactive Tree of Life (iTOL) [26]. The tree was generated to classify the AQP subfamily and isoform properly and from the evolutional relationship of that tree, the names were assigned. Further alignment of AQPs of *Physcomitrella patens and M. charantia* was used to identify the conserved region in the AQP sequences and Ar/R filter and Froger's position.

3. Results

3.1. Classification of *M. charantia* AQPs from Phylogeny Analyses and their Properties

Primarily, twenty-eight aquaporin sequences were found from *M.charantia.* According to the conserved domain search result, one AQP sequence (XP_022159758.1) was lack of NPA (Asn-Pro-Ala) motif and amphipathic channel, which was named *Mc*NIP3; 2 from the phylogenetic tree analyses. To exclude *Mc*NIP3; 2, it was further searched for predicted transmembrane domain in TMHMM and SOSUI (SOSUI result is not included in the result section, Appendix- Table 2), it showed four TMD both in TMHMM and SOSUI result. Again, its sequence alignment showed its NPA was truncated (Loop B), moreover, its protein length and molecular weight were too small, so it was excluded from further analysis of *Mc*AQPs. Again *Mc*PIP3; 1 (XP_022159426.1) was lack of sequences in the alignment and it had truncated NPA in loop E, further analysis in TMHMM showed two TMH and SOSUI result showed three TMD of *Mc*PIP3; 1, its protein length and molecular weight were also too small, so it was also excluded. Another *Mc*AQP named PIP2; 7 (XP_022146281.1) contained a large protein length (684) and molecular weight 75.1 kDa which was a bigger amount than other AQPs of

M. charantia and its predicted TMH was 16 according to TMHMM and 13 according to SOSUI, so it was primarily excluded. Finally, 25 *Mc*AQPs were assigned.

M. charantia's aquaporin protein length ranged from 213 to 325 amino acids. The molecular weight of *Mc*AQPs ranged from 22.32 kDa to 34.52 kDa with the pI range 5.78 to 9.91 (Table 1). *Mc*PIPs molecular weight was higher than the other subfamily and subgroup, ranging between 29.32 to 31.36 kDa and they were almost alkaline. *Mc*TIPs pI range was between 5.33 to 6.88. The majority of *Mc*AQPs had six tran*Sm*embrane helices, while *Mc*PIP1: 3, *Mc*PIP1: 4, and *Mc*NIP2: 2 had five TMHs, and *Mc*TIP4: 1 had seven TMH according to TMHMM (Table 1) results.

Sequence alignment of aquaporins of Arabidopsis, Physcomitrella, Cucumber, Rice and their phylogeny analyses led to group them into five subfamilies: PIP, TIP, XIP, SIP, and NIP. Among them twelve PIPs, three TIPs, one XIPs, one SIPs, and eight NIPs. Further PIP subfamily was divided into five PIP1s and seven PIP2s. XIPs and SIPs contained only one member as a subgroup, XIP1, and SIP1 respectively. TIP included three subgroups: TIP1s, TIP4s, and TIP5s. Finally NIP included four subgroups: three NIP1s, two NIP2s, one NIP3s, and two NIP4s. Their phylogeny analysis showed a more precise classification of those MIPs of *M. charantia* (Full illustration depicted on Appendix-Figure1).

Table 1 Structural motifs of different isoforms of major intrinsic proteins in *Momordica charantia* and their properties.

Name	Accession ID	Protein length (aa)	Molecular weight (kDa)	pI	TMH
*Mc*PIP1; 1	XP_022143699.1	282	29.44	8.8	6
*Mc*PIP1; 2	XP_022145561.1	290	31.36	9.55	6
*Mc*PIP1; 3	XP_022145702.1	213	22.32	9.91	5
*Mc*PIP1; 4	XP_022159472.1	286	30.84	8.97	5
*Mc*PIP1; 5	XP_022156922.1	293	31.33	8.90	6
*Mc*PIP2; 1	XP_022143787.1	279	29.59	9.30	6
*Mc*PIP2; 2	XP_022143897.1	279	29.59	8.94.	6
*Mc*PIP2; 3	XP_022143902.1	284	29.83	6.64.	6
*Mc*PIP2; 4	XP_022143819.1	284	30.17	8.84	6
*Mc*PIP2; 5	XP_022134214.1	287	30.48	9.34	6
*Mc*PIP2; 6	XP_022139134.1	288	30.33	9.18	6
*Mc*PIP2; 8	XP_022135869.1	280	29.86	8.93	6
*Mc*TIP1; 1	XP_022133519.1	253	26.38	5.33	6
*Mc*TIP4; 1	XP_022131720.1	247	25.63	5.91	7
*Mc*TIP5; 1	XP_022159882.1	260	26.73	6.88	6
*Mc*XIP1; 1	XP_022146248.1	325	34.52	5.95	6
*Mc*SIP1; 1	XP_022156236.1	249	26.31	9.55	6
*Mc*NIP1; 1	XP_022143298.1	251	26.48	5.88	6
*Mc*NIP1; 2	XP_022142087.1	275	28.99	9.21	6
*Mc*NIP1; 3	XP_022142207.1	267	28.25	8.69	6
*Mc*NIP2; 1	XP_022149787.1	260	27.74	5.78	6
*Mc*NIP2; 2	XP_022149783.1	293	30.74	7.82	5
*Mc*NIP3; 1	XP_022159570.1	298	31.17	7.76	6
*Mc*NIP4; 1	XP_022138205.1	244	25.73	9.1	6
*Mc*NIP4; 2	XP_022138204.1	247	26.17	9.1	6

Table 2 Structural motifs of different isoforms of major intrinsic proteins in *Momordica charantia.*

*Mc*AQPs		*Mc*AQPs		NPA Motifs		Aromatic/R filter		FP(Froger's position)				
Subfamily	Subgroup	Loop B	Loop E	H2	H5	LE1	LE2	P1	P2	P3	P4	P5
PIP	*Mc*PIP1; 1	NPA	NPA	F	N	A	R	Q	S	S	L	W
	*Mc*PIP1; 2	NPA	NPA	F	H	T	R	Q	S	A	L	W
	*Mc*PIP1; 3	NPA	NPA	F	H	T	R	Q	S	A	F	W
	*Mc*PIP1; 4	NPA	NPA	F	H	T	R	Q	S	A	F	W
	*Mc*PIP1; 5	NPA	NPA	F	H	T	R	Q	S	A	F	W
	*Mc*PIP2; 1	NPA	NPA	F	H	T	R	Q	S	A	F	W
	*Mc*PIP2; 2	NPA	NPA	F	H	T	R	Q	S	A	F	W
	*Mc*PIP2; 3	NPA	NPA	F	H	T	R	Q	S	A	F	W
	*Mc*PIP2; 4	NPA	NPA	F	H	T	R	Q	S	A	F	W
	*Mc*PIP2; 5	NPA	NPA	F	H	T	R	Q	S	S	F	W
	*Mc*PIP2; 6	NPA	NPA	F	H	T	R	Q	S	A	F	W
	*Mc*PIP2; 8	NPA	NPA	F	H	T	R	Q	S	A	F	W
TIP	*Mc*TIP1; 1	NPA	NPA	F	I	A	V	Q	S	P	A	W
	*Mc*TIP4; 1	NPA	NPA	F	I	A	R	Q	S	P	L	W
	*Mc*TIP5; 1	NPA	NPA	L	V	G	C	Q	A	S	T	A
XIP	*Mc*XIP1; 1	SPI	NPA	G	V	V	R	Q	C	P	A	W
SIP	*Mc*SIP1; 1	NPT	NPA	M	I	P	N	Q	A	W	F	Y
NIP	*Mc*NIP1; 1	NPS	NPA	M	V	A	R	Q	T	P	V	W
	*Mc*NIP1; 2	NPA	NPA	A	V	A	R	Q	S	P	I	W
	*Mc*NIP1; 3	NPS	NPA	A	V	A	R	Q	S	P	L	W
	*Mc*NIP2; 1	NPA	NPA	S	S	G	R	Q	S	P	L	W
	*Mc*NIP2; 2	NPA	NPV	S	S	G	R	Q	T	P	L	W
	*Mc*NIP3; 1	NPS	NPV	A	I	G	R	Q	T	P	I	W
	*Mc*NIP4; 1	NPA	NPA	M	V	G	R	Q	S	P	I	W
	*Mc*NIP4; 2	NPA	NPA	M	V	G	R	Q	S	P	I	W

3.1.1. Phylogeny of PIPs

The phylogeny of PIPs shows *Mc*AQPs are in the cluster of a clade of other PIPs of Cucumber, Arabidopsis, and Physcomitrella (Figure 1). So, *Mc*AQPs are here classified as *Mc*PIPs. Further, close observation shows *Mc*PIP1 has made a clade with CsPIP1, AtPIP1, and PpPIP1. Among them, *Mc*PIP1; 1, *Mc*PIP1; 2 and *Mc*PIP1; 4 has directly clade with CsPIP1. Again, *Mc*PIP1; 3, *Mc*PIP1; 5 are in the close of AtPIP1, PpPIP1, and CsPIP1. This careful observation from the tree helped to classify the group and subgroup of *Mc*AQPs (Table 2).
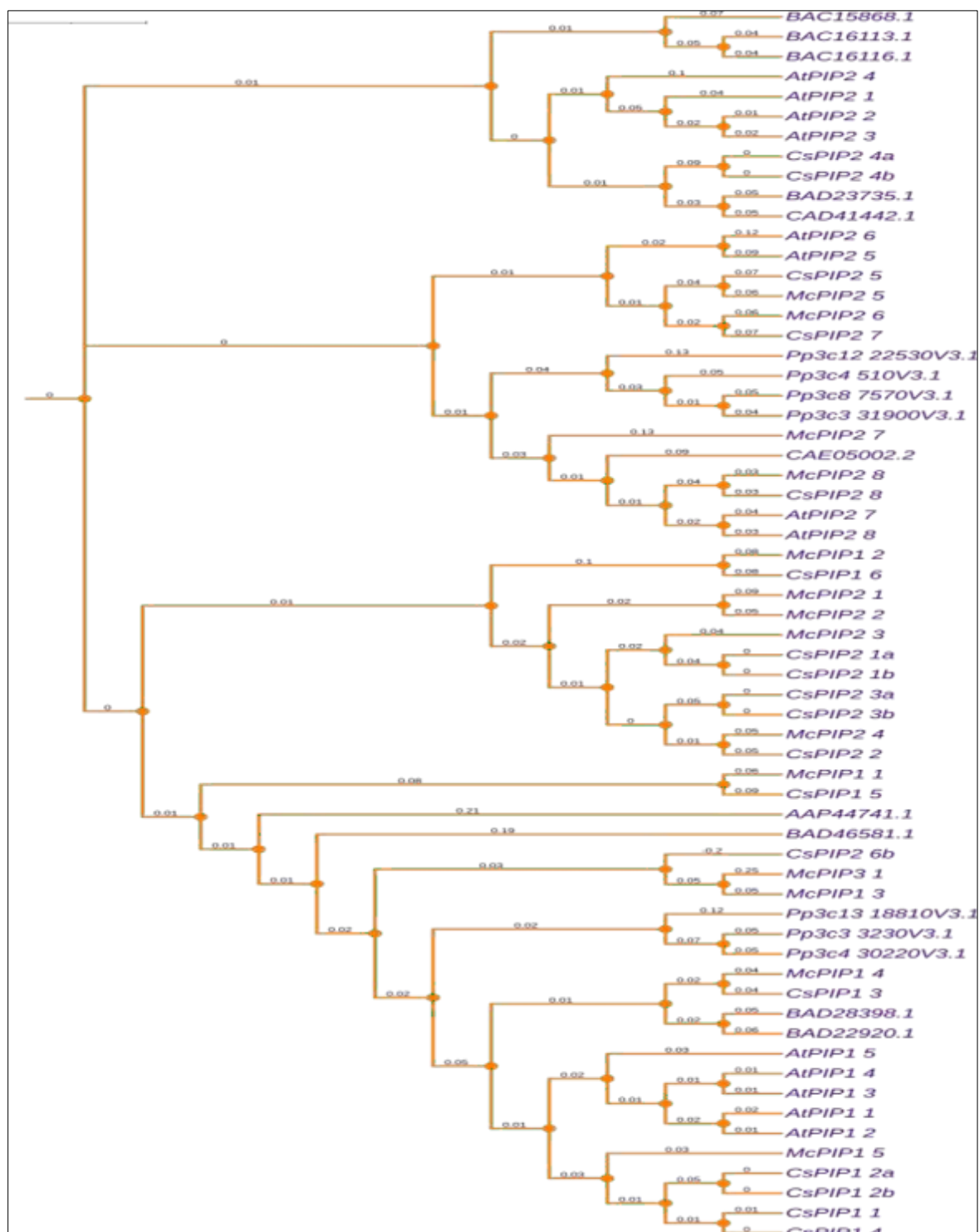
Figure 1 Phylogeny of *Mc*PIPs: The phylogenetic tree of PIPs of Arabidopsis, Physcomitrella, Cucumber, and Rice with the AQPs of *M.charantia.* The AQPs are depicted on the right side. Numbers in the node indicate the branch length.

3.1.2. Phylogeny of TIPs

The figure below shows *Mc*AQPs make a cluster with the TIPs of references plants (Figure 2). *Cs*TIP5s, *Cs*TIP4s and *Cs*TIP1s make a clear clade with *Mc*AQPs. Arabidopsis and Rice TIPs are also close in the clade. Thus *Mc*TIP1; 1, *Mc*TIP4; 1, and *Mc*TIP5; 1 are classified. *Mc*TIP1; 1 have a direct clade with *Cs*TIP1s, *At*TIP1s, and *Os*TIP1s also in the close cluster of clades. *Mc*TIP4; 1 directly clade with *Cs*TIP4s and *At*TIP4s. Again *Mc*TIP5; 1 same as made the clade like *Mc*TIP4s with *Cs*TIP5s and *At*TIP5s. Thu*Smc*TIPs were classified (Table 2).
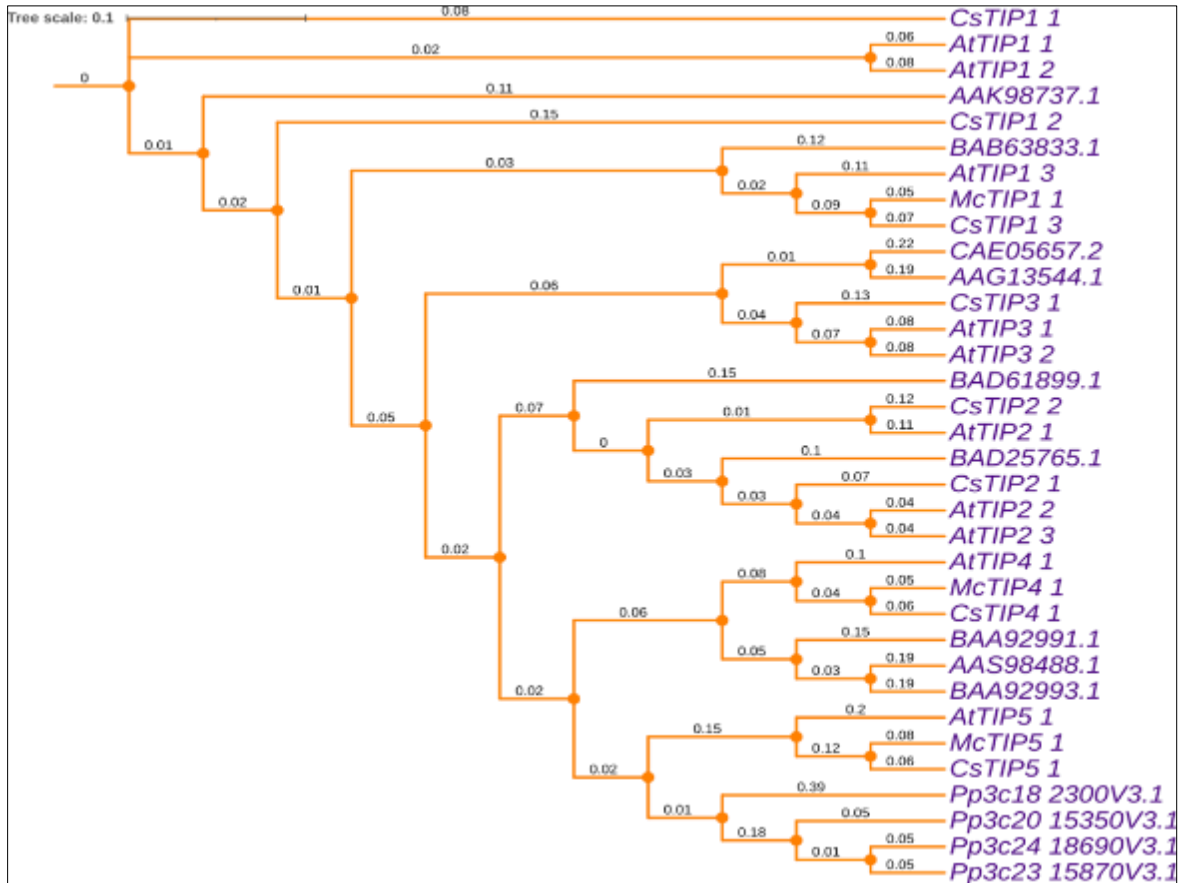
Figure 2 Phylogeny of *Mc*TIPs: The phylogenetic tree of TIPs of Arabidopsis, Physcomitrella, Cucumber, Rice with the AQPs of *M.charantia.* The AQPs are depicted on the right side. Numbers in the node indicate the branch length.

3.1.3. Phylogeny of XIPs and SIPs

This figure includes both XIPs and SIPs of references plants where *Mc*AQPs make clade (Figure 3). So these *Mc*AQPs are *Mc*XIPs and *Mc*SIPs. *Mc*XIP1 directly clade with *Cs*XIP1 and also direct clade formation with *Cs*SIP1 and thus named *Mc*SIP1
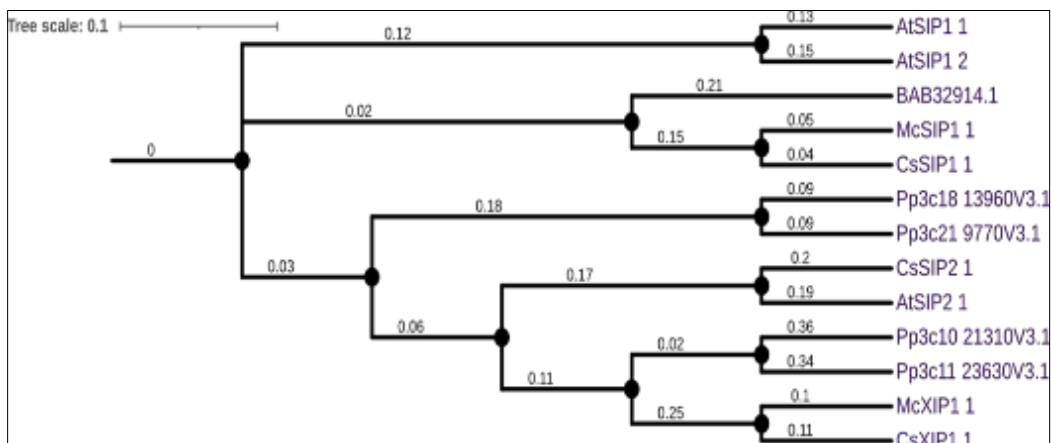

Figure 3 Phylogeny of *Mc*XIPs and *Mc*SIPs: The phylogenetic tree of XIPs and SIPs of Arabidopsis, Physcomitrella, Cucumber, Rice with the AQPs of *M.charantia.* The AQPs are depicted in the right side. Numbers in the node indicate the branchlength

3.1.4. Phylogeny of NIPs

The figure below shows *Mc*AQPs make clear clade with Cucumber's NIPs and Rice NIPs (Figure 4). *At*NIPs and *Pp*NIPs are also in close in the clade of *Mc*AQPs. *Mc*NIP4 is connected by node with *Cs*NIP4. *Mc*NIP3; 1 make node connection with *Cs*NIP3; 1 and *Os*NIP3; 1,while*Mc*NIP3; 2 make clear clade with *Cs*NIP3; 2 and *Mc*NIP2 and *Mc*NIP1 clear clade with *Cs*NIP2 and *Cs*NIP1 respectively. *Mc*NIP1; 2 and *Mc*NIP1; 3 have node connection with *At*NIP1 and *Cs*NIP1
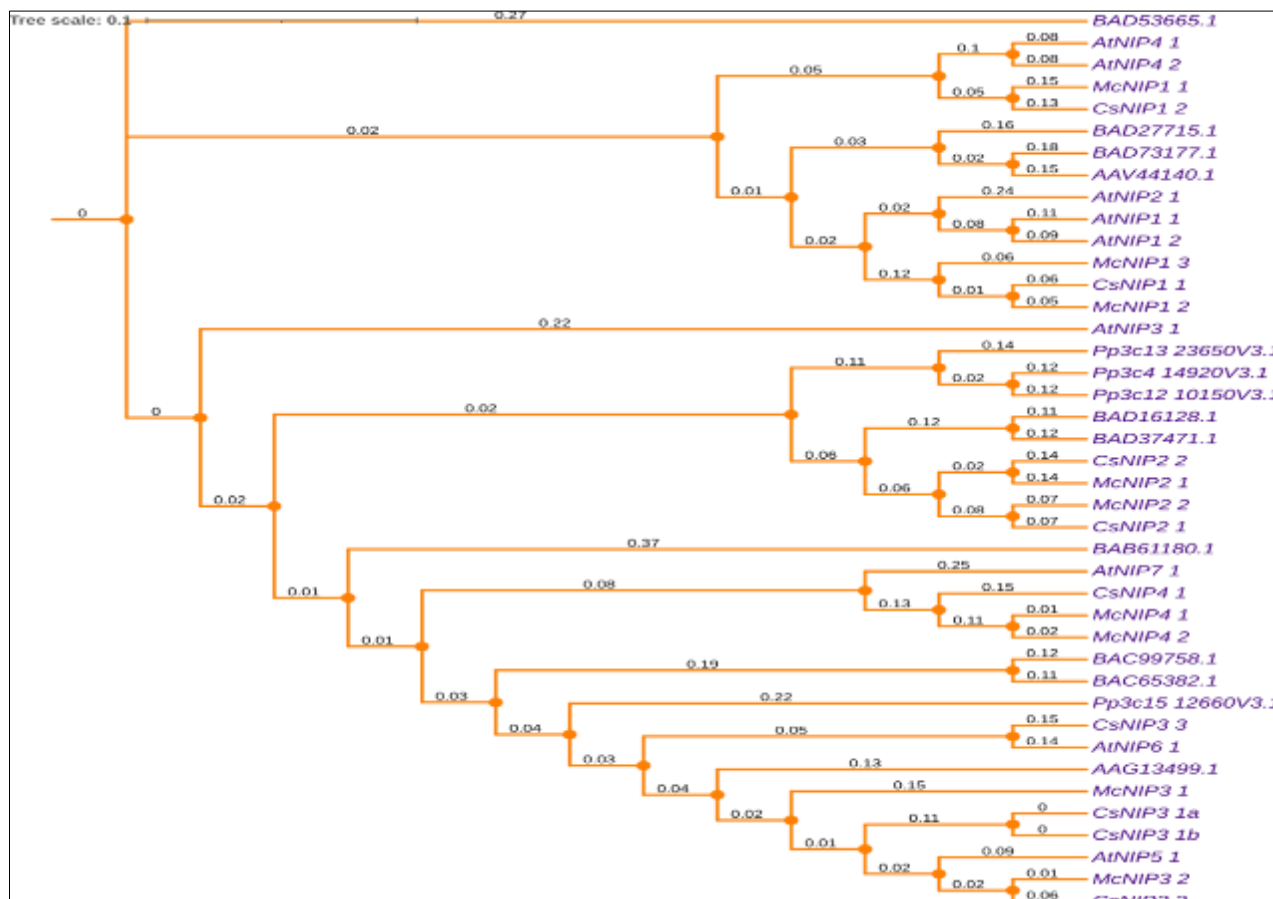

Figure 4 Phylogeny of *Mc*NIPs: The phylogenetic tree of NIPs of Arabidopsis, Physcomitrella, Cucumber, Rice with the AQPs of *M.charantia*. The AQPs are depicted on the right side. Numbers in the node indicate the branch length.

3.2. Distribution of Motifs in Loops (B and E) and Conserved Amino Acid Residues in ar/R filter and Froger's position

*Mc*AQPs have aligned again to assign the conserved position of Asn-Pro-Ala in loop B and E and amino acid reside in aromatic/R filter and Froger's position to find out their role in water and other molecules transportation (Appendix Figure 2). In Table 2 several conserved amino acid positions were reported with a careful visual inspection.

The conserved resides influence substrate specificity through their pore diameter and hydrophobicity [27]. In loops B and E, highly conserved NPA motifs are found. In some cases, different motifs other than NPA were found which indicates transport of other substances along with water. Another set of four conserved resides of Phenylalanine-Histidine-Threonine-Arginine form aromatic/R filter. The first two resides are in the position of H2 and H5 respectively and threonine and arginine are in loop E (E1 and E2). Finally, P1-P2- P3-P4-P5 these positions are said Froger's position as they contain five conserved resides: Q/M-S-A-F-W [28].

All *Mc*PIPs contained the two NPA motifs in loop B and loop E which are characteristic for AQPs. Moreover, all *Mc*PIPs except *Mc*PIP1; 1 showed the typical ar/R selectivity filter for water transporting (F-H-T-R). The P1, P2, and P5 positions were strictly conserved in all *Mc*PIPs and filled with Q-S-W, while the P3 position was more variable and filled by S/A. In *Mc*PIP1; 1, H5 position contained Asparagine (N) instead of Histidine (H) and LE1, Alanine (A) instead of Threonine (T) (Figure 5)
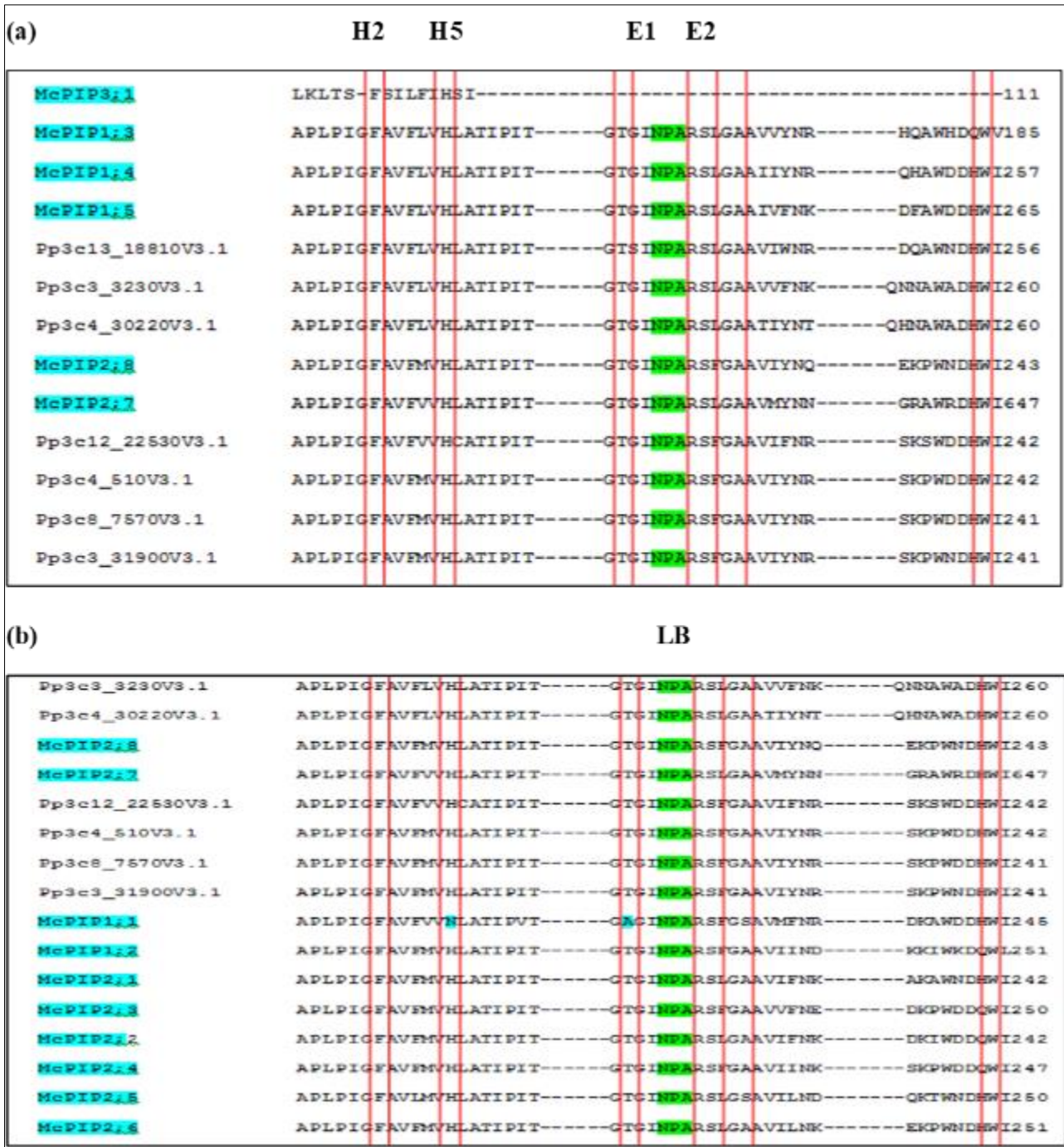
Figure 5 ar/R filter (a) and FPs of *Mc*PIPs (b): the conserved regions are marked, and the motifsare highlighted

In all *Mc*TIPs, NPA motifs were conserved but variance was found in the ar/R filter. P1 position was conserved and P5 was almost conserved but variance occurred in the P2, P3, and P4 positions (Figure 6)
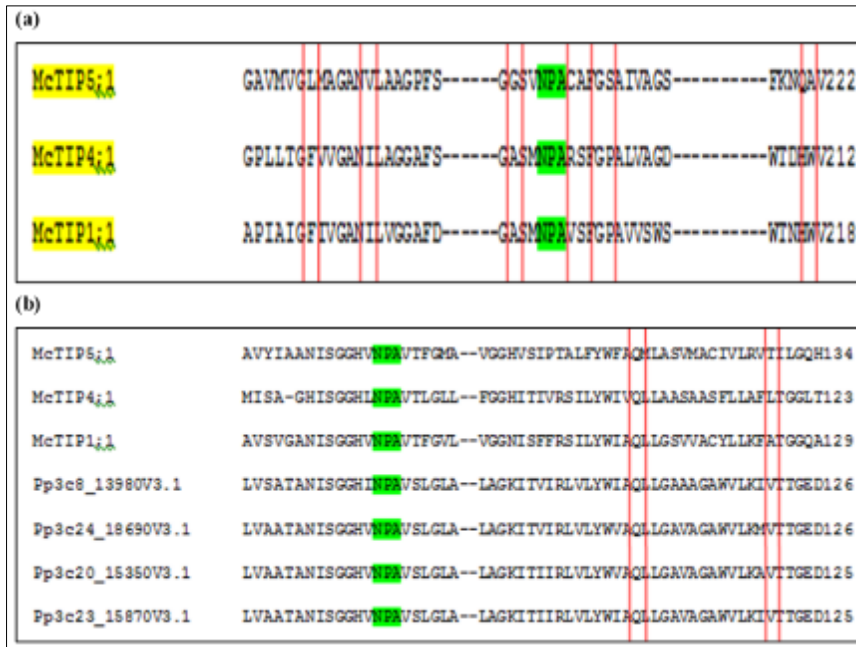
Figure 6 ar/R filter (a) and Froger's position of *Mc*TIPs (b). The motifs are highlighted, and conserved regions are marked

NPA motif in loop B was different in both *Mc*XIPs and *Mc*SIPs, SPI and NPT respectively. They were highly variable in ar/R filter resides and Froger's position excepting P1 position (Fig-7).

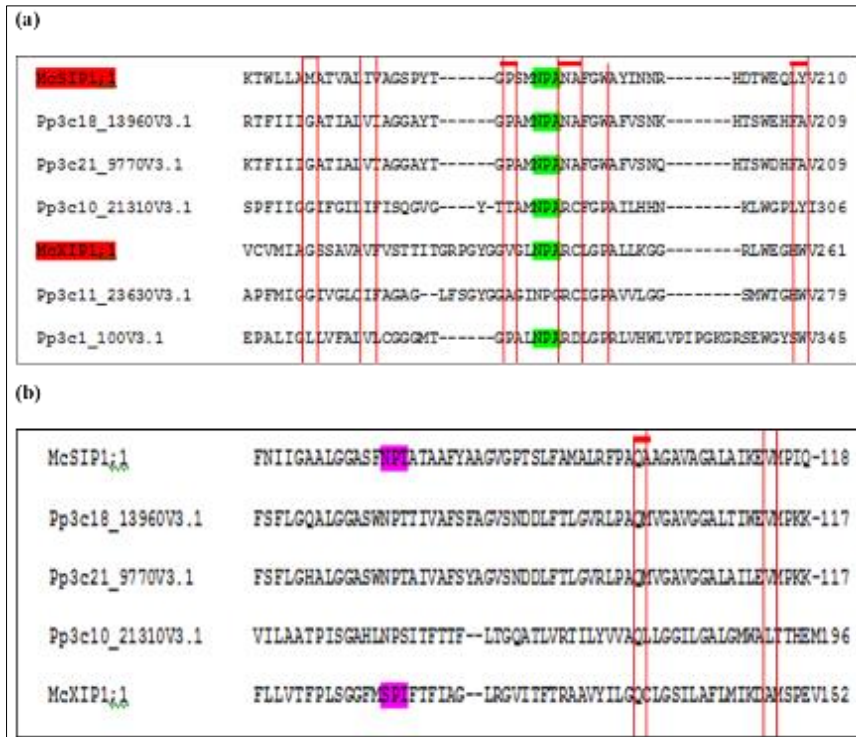**Figure 7** ar/R filter (a) and FPs of *Mc*XIPs and *Mc*SIPs (b)

*Mc*NIPs, NPA had a variance for both loopB and loop E (NPS/NPV). Highly conserved in LE2 position and P1 and P5 positions and variance came in the reside of other positions (Figure 8)
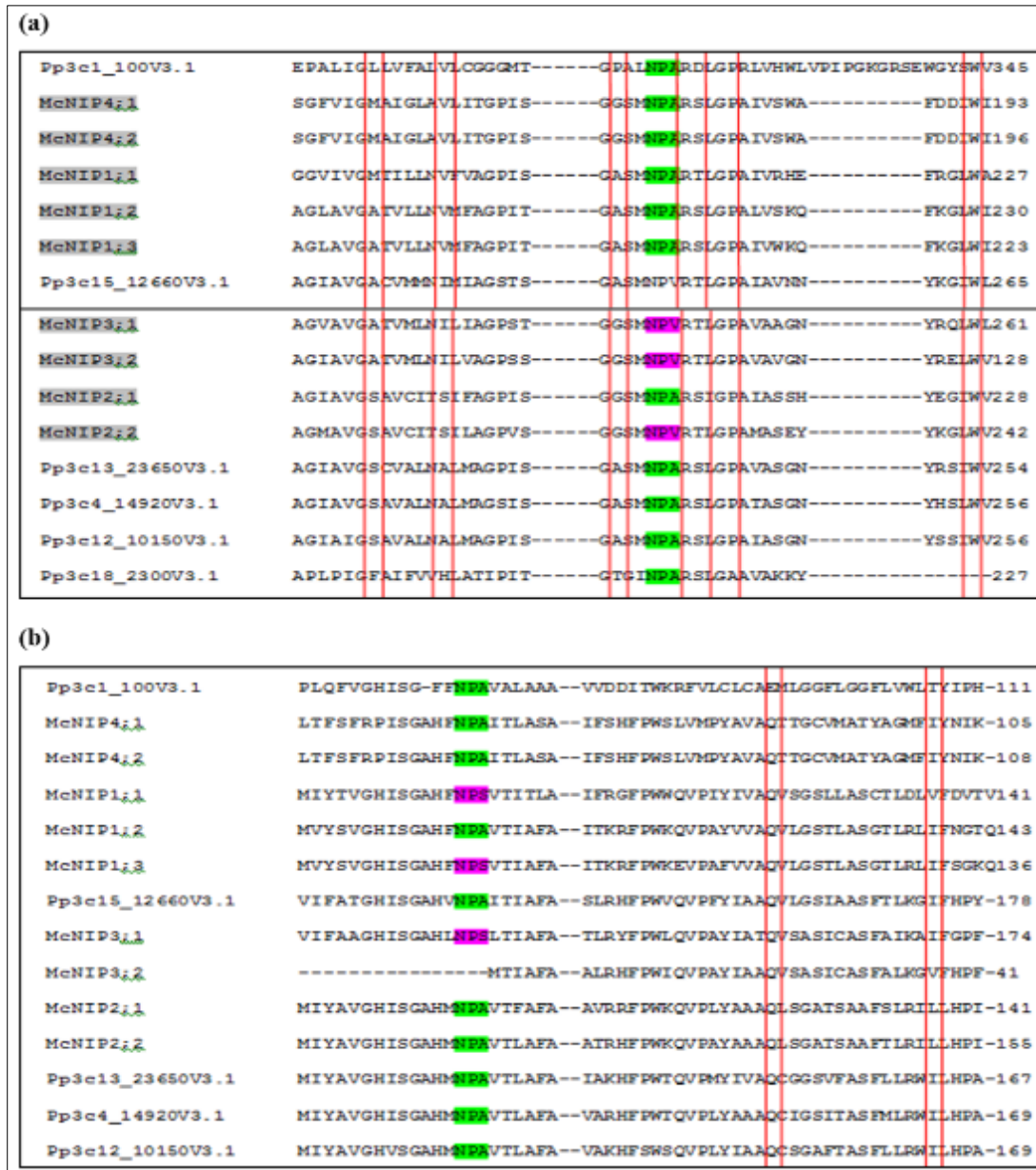
**Figure 8** ar/R filter (a) and FPs of *Mc*NIPs (b)

4. Discussion

Aquaporin as channel protein transport water or other small molecules which help in plant growth and development that is analyzed and proved from different studies of plant and animal aquaporins. In the case of plant aquaporin to assess their function and classification and other properties, model plants are used mostly like Arabidopsis, Rice, Zea mays, etc. Very little information or still nothing is known about the aquaporins of the plant *Momordica charantia*. *Mc*AQPs were classified and characterized here. After screening the 28 *Mc*AQPs, finally, twenty-five were selected, among them *Mc*PIPs were large in number, in total twelve isoforms distributed in *Mc*PIP1s and *Mc*PIP2s. Again, three TIPs, one XIPs, one SIPs, and eight NIPs. Most of the groups and subgroups found in other land plants were present in *M. charantia*. TIP1; 1, NIP4; 1, NIP4; 2,PIP2; 2,PIP2; 4,PIP2; 1,PIP1; 1,NIP1; 1,NIP1; 2,NIP1; 3,PIP1; 5,XIP1; 1,TIP5; 1,NIP3; 1these14*Mc*AQPshad six transmembrane domains according to both the result of TMHMM and SOSUI (result included in appendix Table 2). There were some similarities and dissimilarities of the result of TMHMM and SOSUI. But the excluded aquaporins didn't have identical characteristics of aquaporin according to both the result of TMHMM and SOSUI. *M.charantia*is a plant of the Cucurbitaceae family which is a dicot. It had a close relation with AQPs of *Cucumis sativus*, it is also a plant of the same family of Cucurbitaceae. Moreover, its AQPs had also a similarity with the AQPs of Arabidopsis, Rice, Pyscomitrella. *Mc*AQPs were less in number comparing the AQPs of cucumber. *Mc*XIPs were less than

the other dicots like Tomato (*Solanum lycopersicum*) (4) [10] but similar with Cucumber (one XIPs) [29], while monocots like rice, sorghum were seen mostly not contain XIPs [30] [31]. *Mc*SIPs were also fewer than other dicots, while cucumber had two SIPs, the tomato had four SIPs. Indeed, *Mc*AQPs were also smaller than the other higher plants AQPs. But *Mc*AQPs were more than the AQPs of spike-moss *Selaginella moellendorfii* (19 *Sm*AQPs) [32].

PIPs are considered as transporting water through the channel more than the other AQPs. *Mc*PIPs here almost had that water transporting conserved region in ar/R selectivity filer these conserved regions of residues are F-H- T-R and also conserved almost in Froger's position (Q-S-A-F-W). But *Mc*PIP1; 1 in H5 position contained Asparagine (N) other than Histidine (H). In Froger's position *Mc*PIP1; 1 contained Serine (S) and Leucine (L) instead of Alanine (A) and Phenylalanine (F) respectively in P3 and P4 position. *Mc*PIP1; 2 also contains Leucine (L) instead of Phenylalanine in the P4 position. In the P3 position *Mc*PIP2; 5 also contains Serine (S) other than Alanine (A). But all the *Mc*PIPs contained two NPA motifs. As most of them contained NPA (Asn-Pro-Ala) motifs and amphipathic channel and also those identical water transporting residues, PIPs of *M. charantia were* selective for transporting water as from different studies it was observed that most of the PIPs help in transporting water through the plasma membrane.

Among all AQPs of plants, it is considered that PIPs and TIPs transport more water than others. Here, *Mc*TIPs were divided into three groups: *Mc*TIP1; 1, *Mc*TIP4; 1 and *Mc*TIP5; 1 which was more in other dicot and land plants like tomato, cucumber. *Mc*TIPs also had two conserved NPA motifs in the B and E loop. In the phylogenetic tree *Mc*TIP5; 1 was in the cluster of the clade of AtTIP5; 1 and CsTIP5; 1. AtTIP5; 1 can transport both water and urea [33]. TIP1s are found in roots generally. Analysis from model plants it is seen that TIPs help in water transportation along with urea, NH3 (ammonia), Hydrogen peroxide. Different types of metal and non-metal elements like boric acid (H3BO3), arsenite (NaAsO2) can also be transported through NIPs [34]. The emergence and second highest diversity of NIPs in *M. charantia* indicate that the plant is adaptive as they would be able to transport metal-nonmetal elements so they could respond against disease, stress, and other environmental changes/ defects. There were three NIP1s and two of NIP2 and NIP4 and one NIP3 in *M. charantia* and they had high diversity in NPA motifs. In the case of *Mc*NIP1; 1, *Mc*NIP1; 3 and *Mc*NIP3; 1, loop B contained NPS (Asn-Pro-Ser) instead of NPA(Asn-Pro-Ala), and in *Mc*NIP2; 2, *Mc*NIP3; 1 they had NPV(Asn-Pro-Val) in loop E other than NPA, whereas *Mc*NIP1; 2, *Mc*NIP2; 1, *Mc*NIP4; 1 and *Mc*NIP4; 2 contained NPA motifs in both loops (B and E). Their ar/R filter and FP were also highly diversified.

Appendix
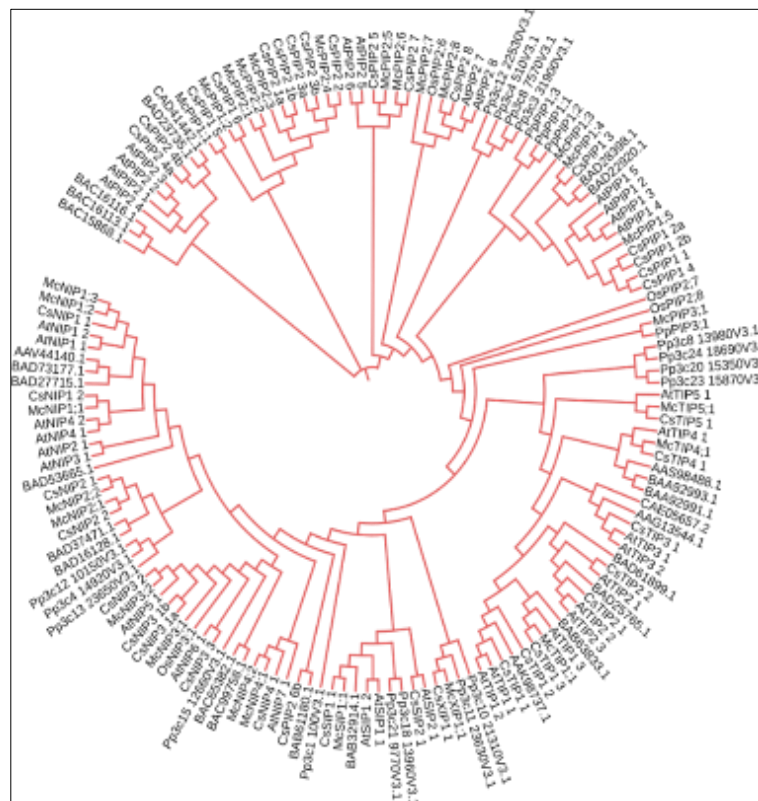
Appendix Figure 1 Phylogenetic tree of *Mc*AQPs.

```
McSIP1;1          FNIIGAALGGASFNPTATAAFYAAGVGPTSLFAMALRFPAQAAGAVAGALAIKEVMPIQ-118
Pp3c18_13960V3.1  FSFLGQALGGASWNPTTIVAFSFAGVSNDDLFTLGVRLPAQMVGAVGGALTIWEVMPKK-117
Pp3c21_9770V3.1   FSFLGHALGGASWNPTAIVAFSYAGVSNDDLFTLGVRLPAQMVGAVGGALAILEVMPKK-117
Pp3c10_21310V3.1  VILAATPISGAHLNPSITFTTF--LTGQATLVRTILYVVAQLLGGILGALGMWALTTHEM196
McXIP1;1          FLLVTFPLSGGFMSPTFTFIAG--LRGVITFTRAAVYILGQCLGSILAFLMIKDAMSPEV152
Pp3c11_23630V3.1  CIFAAAPATGGHVNPCITWTEM--LTGHISPVRGVLYIIGQILGSIVGSFMAKIVVGNAL168
Pp3c1_100V3.1     PLQFVGHISG-FFNPAVALAAA--VVDDITWKRFVLCLCADMLGGFLGGFLVWLTYIPH-111
McNIP4;1          LTFSFRPISGAHFNPAITLASA--IFSHFPWSLVMPYAVAQTTGCVMATYAGMFIYNIK-105
McNIP4;2          LTFSFRPISGAHFNPAITLASA--IFSHFPWSLVMPYAVAQTTGCVMATYAGMFIYNIK-108
McNIP1;1          MIYTVGHISGAHFNPSVTITLA--IFRGFPWWQVPIYIVAQVSGSLLASCTLDLVFDVTV141
McNIP1;2          MVYSVGHISGAHFNPAVTIAFA--ITKRFPWKQVPAYVVAQVLGSTLASGTLRLIFNGTQ143
McNIP1;3          MVYSVGHISGAHFNPSVTIAFA--ITKRFPWKEVPAFVVAQVLGSTLASGTLRLIFSGKQ136
Pp3c16_12660V3.1  VIFATGHISGAHVNPAITIAFA--SLRHFPWVQVPFYIAAQVLGSIAASFTLKGIFHPY-178
McNIP3;1          VIFAAGHISGAHLNPSLTIAFA--TLRYFPWLQVPAYIATQVSASICASFAIKAIFGPF-174
McNIP3;2          ---------------MTIAFA--ALRHFPWIQVPAYIAAQVSASICASFALKGVFHPF-41
McNIP2;1          MIYAVGHISGAHMNPAVTFAFA--AVRRFPWKQVPLYAAAQLSGATSAAFSLRILLHPI-141
McNIP2;2          MIYAVGHISGAHMNPAVTLAFA--ATRHFPWKQVPAYAAAQLSGATSAAFTLRILLHPI-155
Pp3c13_23650V3.1  MIYAVGHISGAHMNPAVTLAFA--IAKHFPWTQVPMYIVAQCGGSVFASFLLRWILHPA-167
Pp3c4_14920V3.1   MIYAVGHISGAHMNPAVTLAFA--VARHFPWTQVPLYAAAQCIGSITASFMLRWILHPA-169
Pp3c12_10150V3.1  MIYAVGHVSGAHMNPAVTLAFA--VAKHFSWSQVPLYIAAQCSGAFTASFLLRWILHPA-169
Pp3c18_2300V3.1   LVYCTAGISGGHINPAVTFGLF--LAQQVTLPRASAYIVAQCLGAIVGAAIARGVQEGGE135
```

```
McPIP3;1          VHLATIPITGTGINPARS----------------------------LGAAVVRSFQKGE-44
McPIP1;3          LVYCTAGISGGHINPAVTFGLL--LARKLSLTRAVFYVIMQCLGAVCGAAVVRSFQKGE-85
McPIP1;4          LVYCTAGISGGHINPAVTFGLF--LARKLSLTRAIFYIIMQCLGAICGAGVVKGFEKSI-157
McPIP1;5          LVYCTAGISGGHINPAVTFGLF--LARKLSLTRAVFYMVMQCLGAIAGAGVVKGFQPKP-165
Pp3c13_18810V3.1  LVYCTAGISGGHINPAVTFGLF--LARKVTFPRTVLYIVCQCLGAICGAGAVKGFQPDF-156
Pp3c3_3230V3.1    LVYCTAGISGGHINPAVTFGLF--LARKVSLNRALFYMIMQCLGAMCGAEIVKGFQPNF-159
Pp3c4_30220V3.1   LVYCTAGISGGHINPAVTFGLF--LARKVSLNRALYYMIMQCLGAMAGAGIVKGFQPDF-159
McPIP2;8          LVYCTAGISGGHINPAVTFGLF--LARKVSLFRAVGYMLAQCAGAIVGVGLVKAFMKHD-143
McPIP2;7          LVYCTAGISGGHINPAVTFGLL--LARKVSVVRAVAYMAAQCLGAIVGVALVKSFMKHA-547
Pp3c12_22530V3.1  LVYCTAGISGGHINPAVTFGLL--LARKISLTRSLAYMVAQCLGAICGAGLVKEFQHSF-142
```

```
Pp3c4_510V3.1     LVYCTAGISGGHINPAVTFGLL--LARKISLPRALAYMIAQCLGAICGAGLVKGFQQSF-142
Pp3c8_7570V3.1    LVYCTAGVSGGHINPAVTFGLL--MARKISLPRALTYMIAQCLGAICGAGLAKGFQTAF-141
Pp3c3_31900V3.1   LVYCTAGISGGHINPAVTFGLL--LARKISLPRALAYMIAQCLGAICGAGLVKGFQTAF-141
McPIP1;1          LVYCTAGISGGHINPAVTFGLF--LGRKVSLVRAVLYIAAQCLGAICGCGLVKSLKKPN-145
McPIP1;2          LVYCIAGISGGHINPAVTFGMF--LARKISLVRALLYIIAQCVGAICGCALVKTLQRDQ-151
McPIP2;1          LVYCTAGISGGHINPAVTFGLF--LARKISLARAVLYMGVQCLGAIWGCALAKSFQKAY-142
McPIP2;3          LVYCTAGISGGHINPAVTVGLL--LARKVSLVRAVLYIVAQCLGAICGCALVKSFQNAH-150
McPIP2;2          LVYCTAGISGGHINPAVTFGLF--LARKVSLVRAVLYMARQCLGAICGCALVKSFQKAH-142
McPIP2;4          LVYCTAGISGGHINPAVTFGLL--LARKVSLVRAILYMVAQSLGAICGCALVKSFQKGL-147
McPIP2;5          LVYCTAGISGGHINPAVTFGLL--LARKVSLVRAVMYMVAQCLGAISGVGLVKAFQKAH-150
```

Appendix Figure 2 Alignment for aromatic/R selectivity filter and Froger's position.

Appendix Table 1 AQPs of plants and their corresponding accession number.

Plant species	Taxa	Accession No
	Name	**GenBank/Phytozome**
Arabidopsis thaliana	*At*PIP1;1 *At*PIP1;2 *At*PIP1;3	CAB71073 AAC28529 AAF81320
	*At*PIP1;4	AAF02782
	*At*PIP1;5	CAA20461
	*At*PIP2;1	CAB67649

	*At*PIP2;2	AAD18142
	*At*PIP2;3	AAD18141
	*At*PIP2;4	BAB09839
	*At*PIP2;5	CAB41102
	*At*PIP2;6	AAC79629
	*At*PIP2;7	CAA17774
	*At*PIP2;8	AAC64216
	*At*TIP1;1	AAD31569
	*At*TIP1;2	BAB01832
	*At*TIP1;3	AAC62778
	*At*TIP2;1	BAB01264
	*At*TIP2;2	CAB10515
	*At*TIP2;3	BAB09071
	*At*TIP3;1	AAG52132
	*At*TIP3;2	AAF97261
	*At*TIP4;1	AAC42249
	*At*TIP5;1	CAB51216
	*At*NIP1;1	CAA16760
	*At*NIP1;2	CAA16748
	*At*NIP2;1	AAC26712
	*At*NIP3;1	AAG50717
	*At*NIP4;1	BAB10360
	*At*NIP4;2	BAB10361
	*At*NIP5;1	CAB39791
	*At*NIP6;1	AAF14664
	*At*NIP7;1	AAF30303
	*At*SIP1;1	AAF26804
	*At*SIP1;2	BAB09487
	*At*SIP2;1	CAB72165
Oryza sativa	OsPIP1;1 OsPIP1;3 OsPIP2;1 OsPIP2;2 OsPIP2;3 OsPIP2;4 OsPIP2;5 OsPIP2;6 OsPIP2;7 OsPIP2;8 OsTIP1;1 OsTIP1;2 OsTIP2;1 OsTIP2;2	BAD28398 BAD22920 BAC15868 BAD23735 CAD41442 BAC16113 BAC16116 CAE05002 BAD46581 AAP44741 AAK98737 BAB63833 BAD25765 BAD61899
	OsTIP3;1	AAG13544
	OsTIP3;2	CAE05657
	OsTIP4;1	AAS98488
	OsTIP4;2	BAA92993

	OsTIP4;3	BAA92991
	OsNIP1;1	BAD27715
	OsNIP1;2	BAD73177
	OsNIP1;3	AAV44140
	OsNIP1;4	BAD53665
	OsNIP2;1	BAD16128
	OsNIP2;2	BAD37471
	OsNIP3;1	AAG13499
	OsNIP3;2	BAC99758
	OsNIP3;3	BAC65382
	OsNIP4;1	BAB61180
	OsSIP1;1	BAB32914
Physcomitrellapatens	*Pp*PIP1;2 *Pp*PIP1;3	Pp1s102_107V6.1 Pp1s305_12V6.1
	*Pp*PIP2;1	Pp1s8_151V6.1
	*Pp*PIP2;2	Pp1s55_301V6.2
	*Pp*PIP2;3	Pp1s267_61V6.1
	*Pp*PIP2;4	Pp1s118_199V6.1
	*Pp*PIP3;1	Pp1s17_281V6.1c
	*Pp*TIP6;1	Pp1s44_31V6.1
	*Pp*TIP6;2	Pp1s156_153V6.1
	*Pp*TIP6;3	Pp1s101_226V6.1
	*Pp*TIP6;4	Pp1s184_96V6.1
	*Pp*NIP3;1	Pp1s258_69V6.1
	*Pp*NIP5;1	Pp1s13_445V6.1
	*Pp*NIP5;2	Pp1s91_35V6.1
	*Pp*NIP5;3	Pp1s37_249V6.1
	*Pp*SIP1;1	Pp1s3_429V6.1
	*Pp*SIP1;2	Pp1s475_9V6.3
	*Pp*XIP1;1	Pp1s31_73V6.1
	*Pp*XIP1;2	Pp1s32_353V6.1
	*Pp*GIP1;1	Pp1s283_16V6.1
Cucumis sativus		http://cucurbitgenomics.org/organism/2
		Gene ID
	CsPIP1;1	Csa5M199270.1
	CsPIP1;2a	Csa5M198770.1
	CsPIP1;2b	Csa5M198770.2
	CsPIP1;3	Csa5M153020.1
	CsPIP1;4	Csa5M199280.1

	CsPIP1;5	Csa6M445090.1
	CsPIP1;6	Csa3M739030.1
	CsPIP2;1a	Csa6M445130.1
	CsPIP2;1b	Csa6M445130.2
	CsPIP2;2	Csa6M445120.1
	CsPIP2;3a	Csa6M445140.1
	CsPIP2;3b	Csa6M445140.2
	CsPIP2;4a	Csa6M140850.1
	CsPIP2;4b	Csa6M140850.2
	CsPIP2;5	Csa6M405320.1
	CsPIP2;6a	Csa6M445150.1
	CsPIP2;6b	Csa6M445150.2
	CsPIP2;7	Csa5M623360.1
	CsPIP2;8	Csa7M014450.1
	CsTIP1;1	Csa6M448110.1
	CsTIP1;1	Csa6M448110.1
	CsTIP1;2	Csa3M743400.1
	CsTIP1;3	Csa5M505790.1
	CsTIP2;1	Csa5M162580.1
	CsTIP2;2	Csa7M447100.1
	CsTIP3;1	Csa1M043290.1
	CsTIP4;1	Csa2M374630.1
	CsTIP5;1	Csa5M168860.1
	CsNIP1;1	Csa6M520340.1
	CsNIP1;2	Csa3M345890.1
	CsNIP2;1	Csa3M826640.1
	CsNIP2;2	Csa3M826650.1
	CsNIP3;1a	Csa5M146200.1
	CsNIP3;1b	Csa5M146200.2
	CsNIP3;2	Csa5M146190.1
	CsNIP3;3	Csa4M007030.1
	CsNIP4;1	Csa3M149960.1
	CsSIP1;1	Csa4M192210.1
	CsSIP2;1	Csa3M816140.1
	CsXIP1;1	Csa2M263850.1

Appendix Table 2 SOSUI result of 28 McAQPs.

McAQPs	Protein length (aa)	TMD
*Mc*PIP1;1	287	6
*Mc*PIP1;2	295	4
*Mc*PIP1;3	218	5
*Mc*PIP1;4	291	6
*Mc*PIP1;5	298	6
*Mc*PIP2;1	284	6
*Mc*PIP2;2	284	6
*Mc*PIP2;3	289	5
*Mc*PIP2;4	289	6
*Mc*PIP2;5	292	5
*Mc*PIP2;6	293	5
*Mc*PIP2;7	689	13
*Mc*PIP2;8	285	5
*Mc*PIP3;1	116	3
*Mc*TIP1;1	258	6
*Mc*TIP4;1	251	6
*Mc*TIP5;1	265	6
*Mc*XIP1;1	328	6
*Mc*SIP1;1	254	7
*Mc*NIP1;1	256	6
*Mc*NIP1;2	280	6
*Mc*NIP1;3	272	6
*Mc*NIP2;1	265	5
*Mc*NIP2;2	298	5
*Mc*NIP3;1	303	6
*Mc*NIP3;2	170	4
*Mc*NIP4;1	249	6
*Mc*NIP4;2	252	6

5. Conclusion

Plants aquaporins play important role in plant growth and development. *M. charantia*'s AQPs are little known here. The *Mc*AQPs here were identified, characterized and their phylogeny analysis helps in this classification. Further sequence alignment of the AQPs gives a representation of amino acid residues of the 25 AQPs of *M. charantia.* Its whole profile shows an important role in its diversity of amino acids residue distributions and thus the possibility of water and other molecule's transportation. *M. charantia* is a medicinal and edible fruit plant, as its aquaporins are little known so more analysis of these aquaporins may help in their physiological properties' explanation.

Compliance with ethical standards

Disclosure of conflict of interest

All authors state that there is no conflict of interest.

References

[1] Verkman AS, Mitra AK. Structure and function of aquaporin water channels. American Journal of Physiology-Renal Physiology. 2000 Jan 1;278(1):F13-28.

[2] Maurel C, Verdoucq L, Luu DT, Santoni V. Plant aquaporins: membrane channels with multiple integrated functions. Annu. Rev. Plant Biol. 2008 Jun 2; 59:595-624.

[3] Chrispeels MJ, Agre P. Aquaporins: water channel proteins of plant and animal cells. Trends in biochemical sciences. 1994 Oct 1;19(10):421-5.

[4] Johansson I, Karlsson M, Johanson U, Larsson C, Kjellbom P. The role of aquaporins in cellular and whole plant water balance. Biochimica et Biophysica Acta (BBA)-Biomembranes. 2000 May 1; 1465(1-2):324-42.

[5] Chaumont F, Tyerman SD. Aquaporins: highly regulated channels controlling plant water relations. Plant physiology. 2014 Apr;164(4):1600-18.

[6] Shivaraj SM, Deshmukh RK, Rai R, Bélanger R, Agrawal PK, Dash PK. Genome-wide identification, characterization, and expression profile of aquaporin gene family in flax (*Linumusitatissimum*). Scientific reports. 2017 Apr 27;7(1):1-7.

[7] Gomes D, Agasse A, Thiébaud P, Delrot S, Gerós H, Chaumont F. Aquaporins are multifunctional water and solute transporters highly divergent in living organisms. Biochimica et Biophysica Acta (BBA)-Biomembranes. 2009 Jun 1; 1788(6):1213-28.

[8] Gupta AB, Sankararamakrishnan R. Genome-wide analysis of major intrinsic proteins in the tree plant *Populus trichocarpa*: characterization of XIP subfamily of aquaporins from evolutionary perspective. BMC plant biology. 2009 Dec;9(1):1-28.

[9] Zhang DY, Ali Z, Wang CB, Xu L, Yi JX, Xu ZL, Liu XQ, He XL, Huang YH, Khan IA, Trethowan RM. Genome-wide sequence characterization and expression analysis of major intrinsic proteins in soybean (Glycine max L.). PLoS one. 2013 Feb 20;8(2):e56312.

[10] Di Pasquale G, Salignon M, Le Conte Y, Belzunces LP, Decourtye A, Kretzschmar A, Suchail S, Brunet JL, Alaux C. Influence of pollen nutrition on honey bee health: do pollen quality and diversity matter?. PloS one. 2013 Aug 5; 8(8):e72016.

[11] Deshmukh RK, Sonah H, Bélanger RR. Plant Aquaporins: genome-wide identification, transcriptomics, proteomics, and advanced analytical tools. Frontiers in plant science. 2016 Dec 20;7:1896.

[12] Grover JK, Yadav SP. Pharmacological actions and potential uses of Momordica charantia: a review. Journal of ethnopharmacology. 2004 Jul 1;93(1):123-32.

[13] Fleminger J, Goldacre B. https://www. ncbi. nlm. nih. gov/pubmed/23743517.

[14] Johanson U, Karlsson M, Johansson I, Gustavsson S, Sjövall S, Fraysse L, Weig AR, Kjellbom P. The complete set of genes encoding major intrinsic proteins in Arabidopsis provides a framework for a new nomenclature for major intrinsic proteins in plants. Plant physiology. 2001 Aug 1;126(4):1358-69.

[15] Sakurai J, Ishikawa F, Yamaguchi T, Uemura M, Maeshima M. Identification of 33 rice aquaporin genes and analysis of their expression and function. Plant and Cell Physiology. 2005 Sep 1;46(9):1568-77.

[16] Blast NC. Available online: https://blast. ncbi. nlm. nih. gov. Blast. cgi (accessed on 31 December 2015).

[17] Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ. CDD: NCBI's conserved domain database. Nucleic acids research. 2015 Jan 28;43(D1):D222-6.

[18] Server TM. v. 2.0 http://www. cbs. dtu. dk/services. TMHMM/(accessed Oct 9, 2013).

[19] Tamura K, Goto C, Hara-Nishimura I. Recent advances in understanding plant nuclear envelope proteins involved in nuclear morphology. Journal of experimental botany. 2015 Mar 1;66(6):1641-7.

[20] da Silva ES, Huber S, Alcantara-Neves NM, Asam C, Silveira EF, de Andrade Belitardo EM, Aglas L, Wallner M, Gadermaier G, Briza P, Karner I. N-terminal peptide deletion influences immunological and structural features of Blo t 5. Allergy. 2020 Jan 8;10.

[21] Johanson U, Karlsson M, Johansson I, Gustavsson S, Sjövall S, Fraysse L, Weig AR, Kjellbom P. The complete set of genes encoding major intrinsic proteins in Arabidopsis provides a framework for a new nomenclature for major intrinsic proteins in plants. Plant physiology. 2001 Aug 1;126(4):1358-69.

[22] Sakurai J, Ishikawa F, Yamaguchi T, Uemura M, Maeshima M. Identification of 33 rice aquaporin genes and analysis of their expression and function. Plant and Cell Physiology. 2005 Sep 1;46(9):1568-77.

[23] Zhu YX, Yang L, Liu N, Yang J, Zhou XK, Xia YC, He Y, He YQ, Gong HJ, Ma DF, Yin JL. Genome-wide identification, structure characterization, and expression pattern profiling of aquaporin gene family in cucumber. BMC Plant Biology. 2019 Dec;19(1):1-23.

[24] Danielson JÅ, Johanson U. Unexpected complexity of the aquaporin gene family in the moss Physcomitrella patens. BMC plant biology. 2008 Dec;8(1):1-5.

[25] Tools B. Multiple Sequence Alignment: Clustal Omega[Program]. URL: https://www. ebi. ac. uk/Tools/msa/clustalo/(accessed 20.10. 2018).

[26] Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic acids research. 2019 Jul 2;47(W1):W256-9.

[27] Hove RM, Bhave M. Plant aquaporins with non-aqua functions: deciphering the signature sequences. Plant molecular biology. 2011 Mar 1;75(4-5):413-30.

[28] Murata K, Mitsuoka K, Hirai T, Walz T, Agre P, Heymann JB, Engel A, Fujiyoshi Y. Structural determinants of water permeation through aquaporin-1. Nature. 2000 Oct;407(6804):599-605.

[29] Zhu YX, Yang L, Liu N, Yang J, Zhou XK, Xia YC, He Y, He YQ, Gong HJ, Ma DF, Yin JL. Genome-wide identification, structure characterization, and expression pattern profiling of aquaporin gene family in cucumber. BMC Plant Biology. 2019 Dec;19(1):1-23.

[30] Reddy PS, Rao TS, Sharma KK, Vadez V. Genome-wide identification and characterization of the aquaporin gene family in Sorghum bicolor (L.). Plant Gene. 2015 Mar;1:18-28.

[31] Sakurai J, Ishikawa F, Yamaguchi T, Uemura M, Maeshima M. Identification of 33 rice aquaporin genes and analysis of their expression and function. Plant and Cell Physiology. 2005 Sep 1;46(9):1568-77.

[32] Anderberg HI, Kjellbom P, Johanson U. Annotation of *Selaginella moellendorffii* major intrinsic proteins and the evolution of the protein family in terrestrial plants. Frontiers in plant science. 2012 Feb 20;3:33.

[33] Soto G, Fox R, Ayub N, Alleva K, Guaimas F, Erijman EJ, Mazzella A, Amodeo G, Muschietti J. TIP5; 1 is an aquaporin specifically targeted to pollen mitochondria and is probably involved in nitrogen remobilization in Arabidopsis thaliana. The Plant Journal. 2010 Dec;64(6):1038-47.

[34] Fox AR, Maistriaux LC, Chaumont F. Toward understanding of the high number of plant aquaporin isoforms and multiple regulation mechanisms. Plant Science. 2017 Nov 1;264:179-87.