Original Article

TIPS Trends in Phramaceutical Sciences 2016: 2(2): 139-150.

Predicting the functionality of major intrinsic proteins: An *in silico* analysis in *Musa*

Shiva Hemmati[1,2,*]

[1]Department of Pharmaceutical Biotechnology, School of Pharmacy, Shiraz University of Medical Sciences, Shiraz, Iran.
[2]Pharmaceutical Sciences Research Center, Shiraz University of Medical Sciences, Shiraz, Iran.

Abstract

Major intrinsic proteins (MIPs) are tetrameric complexes with six transmembrane domains. MIPs that are involved in water and nutrient permeability, have been called aquaporins and aquaglyceroporins, respectively. Four important subfamilies of MIPs are plasma membrane intrinsic proteins (PIPs), tonoplast intrinsic proteins (TIPs), nodolin-26 like intrinsic proteins (NIPs), and small basic intrinsic proteins (SIPs). *Musa* sp. from the order Zingiberals is not only the largest supply of food for millions of people, but also has tremendous therapeutic properties, such as ameliorating digestive and metabolic disorders. This species is sensitive to any kind of water deficit. Recently the genomic sequence of *Musa acuminata* has been determined. Besides the localization of MaMIP genes on the chromosomes and the localization of MaMIP proteins in the subcellular compartments, the substrate selectivity of MaMIPs has been determined by dual NPA (Asparagine-Prolin-Alanine) motifs, the ar/R (aromatic/Arginine) selectivity filter, and Froger's position. MaPIP subfamilies are transporters of water, boron, carbon dioxide, hydrogen peroxide and urea, while MaTIP subfamilies are transporters of water, hydrogen peroxide, urea and ammonia. MaNIP subfamily has been shown to be transporters of silicon, urea, and boron. The functional prediction of the MaMIP roles provides the opportunity to genetically target these passive transporters for improving the species trait.

Keywords: Aquaporin, Major intrinsic protein, *Musa acuminata*, , Protein function, Substrate selectivity.

1. Introduction

Major intrinsic proteins (MIPs) are pore forming proteins in membrane. Most members of this superfamily named aquaporins increase the water permeability (1). MIPs that transport glycerol, urea, ammonia (NH_3), carbon dioxide (CO_2), silicon (Si), and hydrogen peroxide (H_2O_2) have been called aquaglyceroporins (GlpFs) (2). Identification and characterization of MIPs, which are involved in the transport of small molecules like H_2O_2, boron, Si, NH_3, and urea, would have a vital role in defense against biotic and abiotic stresses. MIPs are divided into five subfamilies including plasma membrane intrinsic proteins (PIPs), tonoplast intrinsic proteins (TIPs), nodolin-26 like intrinsic proteins (NIPs), small basic intrinsic proteins (SIPs), and uncategorized X intrinsic proteins (XIPs) (3).

These 28-30 KDa subfamilies of MIPs are tetrameric complexes consisting of six alpha-helical transmembrane domains (TM1-TM6), connected by five inter-helical loops (LA-LE), with cytoplasmic N and C termini (4). The primary structure can be divided into two similar parts: TM1-3 and TM4-6 as hemipore 1 and 2, respectively (5). At the center of the pore, two highly conserved motifs of NPA (Asparagine-Prolin-Alanine) in the loops B and E are observed. Substrate selectivity of MIPs is defined by the two NPA motifs and the tetrad composed of residues from helices 2, 5, and loop E called ar/R (aromatic/Arginine) constriction (6).

Corresponding Author: Shiva Hemmati, Department of Pharmaceuticals Biotechnology, School of Pharmacy, Shiraz University of Medical Sciences, Shiraz, Iran.
Email: hemmatish@sums.ac.ir

Recieved:06/05/2016; Accepted:27/05//2016

Another signature, called Froger's position, is ascribed to residues designated as P1-P5, which are determinant in discriminating MIP function (7).

Musa spp. perennial monocotyledonous herb of the order Zingiberales, as the world's second largest fruit crop, is an important supply of food for millions of people worldwide especially in tropical and subtropical countries (8). In addition to nutritional importance, numerous medicinal properties have been attributed to this plant, both in modern and traditional medicine. Several parts of the plant have been used in digestive disorders (9). Banana as a prebiotic is used as a food source for colonal probiotics. Banana is rich of non-digestible fibers and pectins, which normalizes bowel movement. This plant not only ameliorates gastric ulcers, but also lowers cholesterol and regulates blood glucose in diabetes. *Musa* is a rich source of vitamins and potassium, which controls high blood pressure and refines renal function. Banana extract reduces the risk of neurodegenerative afflictions (9). It has been shown that peels and pulps are effective in bacterial and fungal infections (10). Flowers have been used for the treatment of dysentery and bronchitis. Saps have shown astringent properties.

Musa needs ample supply of water. Drought stress limits banana production (11). *Musa* is one of the 132 crops in over 80 countries that encounter boron deficiency (12). On the other hand, it has been revealed that high boron concentration damaged the leaf margins of banana (13). The contribution of channel proteins in boron transport in banana is poorly understood. Identification and manipulation of MIPs involved in boron transport would alleviate symptoms of boron deficiency and toxicity. Banana plants also have a very high nitrogen demand, and frequent nitrogen addition is needed (14). Most banana crops in tropical countries are fertilized with high rates of nitrogen. Half of nitrogen fertilizers are urea. Si helps in preventing the pathogen penetration and in the activation of defense mechanisms, like the production of phenolics, phytoalexins, and lignin (15). Identification of CO_2 transporters, which affect photosynthetic efficiency, would increase banana tolerance to abiotic stresses (16). H_2O_2 which is known as a reactive oxygen species (ROS), initiates a signaling process to provide adaptation to biotic and abiotic stresses at low concentrations (17). The genomic sequence of DH-Pahang, a doubled-haploid *M. acuminata* subspecies malaccensis has been determined (18). Identification of MIP channels in banana provides the opportunity to improve the traits by genetic manipulation.

In this study, bioinformatic analysis has been employed to get deep insights into the functional properties of banana AQPs. Forty-seven *AQP* genes in *M. acuminata* were distributed in four subfamilies by comprehensive homology and phylogenetic analysis. Chromosomal distribution, subcellular localization, and TM domains were determined. Residues that are determinant in substrate specificity were identified. The substrate selectivity/specificity for CO_2, H_2O_2, urea, NH_3, boron, and Si in banana MIPs has been predicted.

2. Materials and Methods

2.1. Identification of MIP homologues in M. acuminata (MaMIPs)

To comprehensively identify putative banana MIPs, the analysis of the *M. acuminata* DH Pahang genome (v. 1) available at banana genome hub (http://banana-genome.cirad.fr/) (19) was conducted by multiple BLAST tools, using rice, maize, *Arabidopsis*, and potato AQPs as a query sequence with cut-off *E*-value of 1e-10. From the 523 Mb size of the Pahang genome, 472 Mb has been assembled (19). This assembly consists of 11 pseudomolecules correlated to the 11 chromosomes of banana and an extra pseudomolecule not yet designated to any chromosome. The chromosomal location of banana *AQP* genes were determined using gene ID in the genome browser of banana genome hub.

2.2. In silico prediction of TM helical domains

Sequences of MaMIPs were submitted to Pfam (http://pfam.xfam.org/) (20) and CDD (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) to confirm the family of retrieved proteins (21). The Interpro (http://www.ebi.ac.uk/interpro/), was used to re-confirm the protein family and TM domains (22). The TMHMM server v.2.0 (http://www.cbs.dtu.dk/services/TMHMM/) was used to predict transmembrane topology of AQPs

(23). In addition to TMHHM, TMpred (http://www.ch.embnet.org/software/TMPRED_form.html), which suggests strong models for TM topology was used. The minimum inquiry for the TMpred software to consider a domain as TM are helices with 18 or larger amino acids (24). Scores above 500 are considered as significant. Although success rates of TM prediction tools are usually more than 90%, a notable number of false positive hits is predicted (25). Servers like Phobius (http://phobius.sbc.su.se/) result in significant false negative results. A modified dense alignment surface method called DAS-TMfilter (http://www.enzim.hu/DAS/DAS.html) was used, in which false positive TM segments could be excluded (26). Higher accuracy achieved using the PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/), which detects zero false positive and has a 0.4% false negative response rate (27). Using Phyre2, the 3D model of MaMIPs was constructed for further quality check of the proteins. Sequences that did not contain six TM domains were identified by these servers.

2.3. Multiple sequence alignment, and phylogenetic analysis

To classify banana AQPs and to establish phylogenetic relationship, the amino acid sequences of *AQP* proteins along with *AQP* proteins of rice, maize, *Arabidopsis*, and potato were aligned by Clustal W program, using blosum protein weight matrix with an open gap penalty of 10 and extended gap penalty of 0.05. Alignments were inspected using Bioedit version 7.2.5 to avoid gaps in functionally important features, like the NPA boxes and TM regions. Truncated protein sequences, sequences lacking any of the NPA motifs, and sequences which were stopped in frame were excluded from further analysis. Dual NPA constricts, ar/R selectivity filters, and Forger's positions were determined by multiple sequence alignment of MaMIPs with functionally known MIP sequences of rice, maize and *Arabidopsis*. Using MEGA6 package a phylogenetic tree was constructed with neighbor-joining (NJ) method and 1000 bootstrap replicates (28). Annotation of banana AQPs were performed based on sequence similarities and phylogenetic relationship with orthologous genes in rice, maize, and *Arabidopsis*.

2.4. Functional characterization of MaMIPs for non-aqua substrates

For substrate specificity characterization, the protein sequences of plant AQPs, whose ability to transport solutes have been demonstrated experimentally, were retrieved from NCBI according to the corresponding references reported by Hove (4). These identified proteins have known function by one of the methods like yeast or, xenopus oocyte assay, using transgenic plants or in the presence of MIP inhibitors. Each subfamily of the mentioned sequences (with known residues and known function) was aligned with the corresponding subfamily in *Musa* to find homologous residue and predict the function of MaMIPs.

According to (4) the consensus sequence flanking the NPA filters for MIP boron transporters are as follows: SG(G/A)H(I/L/M)NP(A/S)(V/L)T and (G/S)(T/G/A)(G/S)(I/M)NP(A/V)R(S/T)LG. The ar/R selectivity filter of boron transporter MIPs are F/A/G at the TM2 domain, H/I/S at the TM5 domain, T/G/A at LE1 and R at LE2. The Froger's position would be Q/F/I, T/S, A, F/Y, and L/W/F at P1-P5, respectively. For ammonia transporters, the two NPA consensus sequences are SGGH(V/L/F)NPAVT and G(G/A)SMNPARS(F/L)G. H/W, I/V, A/G, and R residues are located at four positions of ar/R selectivity filter. The P1-P5 positions for MIP ammonia transporters are as follows: F/T, S, A, Y, L/W. In CO_2 transporters, SGGHINPAVT and GTGINPARSLG are the signatures of the first and the second NPA. F-H-T-R residues at ar/R selectivity filter are observed. The P1-P5 positions for CO_2 transporters are Q/F/I, T/S, A, F/Y, and L/W/F, respectively. H_2O_2 transporters have SG(G/A)H(V/L/I/F)NPAVT and G(A/G/T)(S/G)(M/I)NP(A/G)(V/R)(A/S)(F/L)G at the first and second NPA motifs. The ar/R filter for H_2O_2 transporters is H/F/W, I/H/V, A/T/G, and V/R. The Froger's position at P1-P5 are T/Q/F, A/S, A, Y/F, and W/I. SG(G/A)H(I/L/V/M)NPAVT and (G/S)(A/T/G)(G/S)(I/M)NP(A/V)(R/V/C)(S/A/T)(L/F)G are the two flanking regions of the NPA consensus sequence in MIP urea transporters. At the 2nd and 5th TM domain, any of the F/H/G/A/N and H/I/S/V residues are present, respectively. T/A/G and R/V/C are observed at LE1 and LE2.

The P1-P5 position in urea transporters are as follows: M/T/Q/L/F/V/I at P1, S/A/T at P2, A, F/Y, and W/F/L at P3-P5. Finally, for silicon transporters SGAHMNPAVT, GGSMNPARTL(G/A) at the first and second NPA, G-S-G-R at ar/R selectivity filter and I/L-T-A-Y-F at P1-P5 are typified.

2.5. Analysis of protein characteristics and subcellular localization

Molecular weight (MW) and isoelectric point (pI) of different families of MaMIPs were calculated using ProtParam (http://web.expasy.org/protparam/). Subcellular localization of banana AQPs was predicted by the WOLFPSORT algorithm available at http://www.genscript.com/psort/wolf_psort.html.

3. Results

3.1. Identification, nomenclature, and distribution of AQP genes

By mining the banana genome database, a total of 56 putative MIP genes were identified based on sequence homology analysis. Nine MIP genes including chr1G07460, chr1G26390, chrUn_randomG20380, chrUn_randomG20370, chrUn_randomG20350, chr4G10230, chrUn_randomG27360, chr10G27780, and chr2G01970, which were truncated, stopped in frame, missed start codon, lacked any of the conserved NPA motifs or showed incomplete TM helices were considered nonfunctional pseudogenes and were excluded from further analysis. The amino acid sequences of the remaining 47 full length genes encoding MIP proteins were nomenclatured using

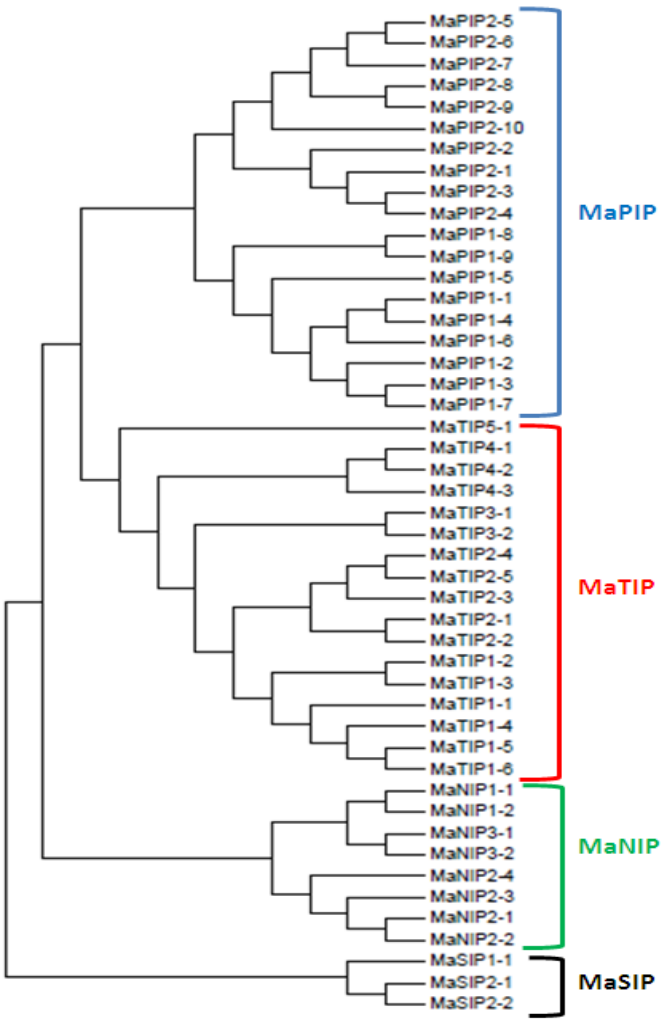

Figure 1. Phylogenetic analysis of the amino acid sequences of banana MIPs using MEGA6 package. The name of each subfamily is indicated.

phylogenetic relationships and amino acid similarity with monocot and dicot species like rice, maize, and *Arabidopsis*. As observed in the tree resulted from the phylogenetic analysis of banana MIPs (Figure 1), MaMIPs are clustered into four separate subfamilies called MaPIPs, MaTIPs, MaNIPs, and MaSIPs.

3.2. Chromosomal distribution of MaMIPs

A schematic distribution of banana *AQP* genes is represented in Figure 2, using MaTIP1-1 in the banana genome hub as the query. MIP genes were located in nine out of the eleven chromosomes (pseudomolecules) in the banana genome v. 1. No loci were detected on chromosomes three and seven using even all MaMIPs to figure out chromosomal distribution of MIP genes in the database.

3.3. Banana AQP protein structure and function

Prediction of TM domain revealed the presence of six TM domains in almost all identified putative MIPs in banana. The TMHMM algorithm did not identify the correct number of TMs in all cases. It was corrected by screening the amino acid sequence alignments and predicted plots of the software. DAS-TMfilter was used, which exclude the false positive TM segments. Higher accuracy achieved using the PSIPRED. By manual inspection of multiple alignments of the MaMIP sequences (Figure 3), residues that determine substrate specificity of AQPs in banana were characterized. Analysis of the highly conserved dual NPA motifs, ar/R selectivity filters, and P1-P5 Froger's positions for each member of the MaMIP are summarized in Table 1 and described in the following section. Prediction of the subcellular localization of MaMIPs showed that all are localized to the plasma membrane (Table 2). It is expected that TIPs would be localized in the vacuolar membrane. Prediction of the MaTIP subcellular localization identifies plasma membrane as well as cytosolic localization in parallel with the vacuolar targets. Most banana MaNIP members were localized to plasma membrane except MaNIP1-2 and MaNIP2-2, which were predicted to localize in vacuolar and chloroplast membranes. MaSIPs are also predicted to localize in the plasma and chloroplast membranes.

3.3.1. MaPIP subfamily: transporters of water, boron, CO_2, H_2O_2, and urea

MaPIP subfamily, which is divided into MaPIP1 and MaPIP2 subgroups, has the highest sequence similarity within MaMIPs from 77-98%. The MaPIP1s have relatively longer and shorter N- and C-terminal, in comparison to the MaPIP2 subgroup. MaPIP1 peptide sequence length ranges

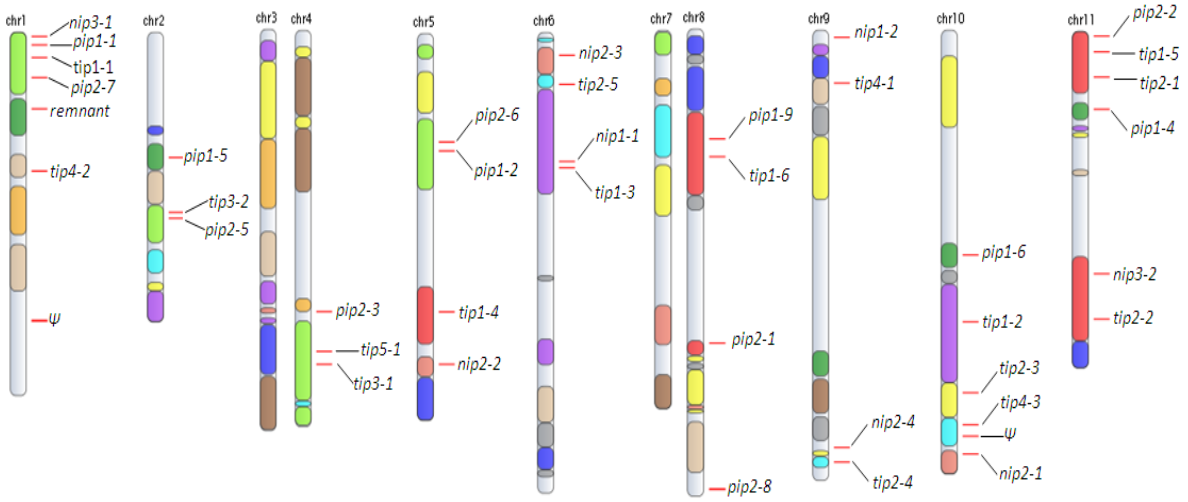

Figure 2. Annotation of MaMIP genes localized on the chromosomes of *M. acuminata* generated by tblastn using MaTIP1-1 as query in the banana genome server (Ψ=Pseudogene).

from 285 to 287 amino acids, while this range varies for the MaPIP2 from 281 to 288 amino acids (Table 2). The MaPIP subfamily has a MW of 29-30 KDa with a basic pI except for MaPIP2-1 and MaPIP2-4, which has a pI of 6.95 (Table 2). The dual NPA motif is conserved in all MaMIPs with a uniform configuration of F-H-T-R residue according to the ar/R selectivity filter, which is representative for the MIP water transporters (Table 1 and Figure 3). Except the P1 position that shows variability, strict conservation of P2-P5 residues (S-A-F-W) was observed in all MaPIPs. According to the selectivity related sequences for non-aqua substrates described in the materials and methods, all MaPIPs except MaPIP1-1 are potentially urea transporters (Table 1). MaPIP1-2 to MaPIP1-4, MaPIP1-6 to MaPIP1-9, MaPIP2-1, MaPIP2-5 to MaPIP2-7, and MaPIP2-10 are CO2 transporters. All MaPIP1 subgroup except MaPIP1-1 and MaPIP1-5 are potential boron transporters. MaPIP1-2, MaPIP1-3, MaPIP1-5 to MaPIP1-9, and MaPIP2-1 to MaPIP2-4 are predicted as H2O2 transporters.

3.3.2. MaTIP subfamily: transporters of water, H_2O_2, urea, and NH_3

The sequence similarity within MaTIP subfamily members, which are clustered from MaTIP1 to MaTIP5 ranges from 52-95%. MaTIP peptide sequence length ranges from 243 to 259 amino acids with a MW of 24-26 KDa. The pI in almost all MaTIPs is acidic except for MaTIP3-1

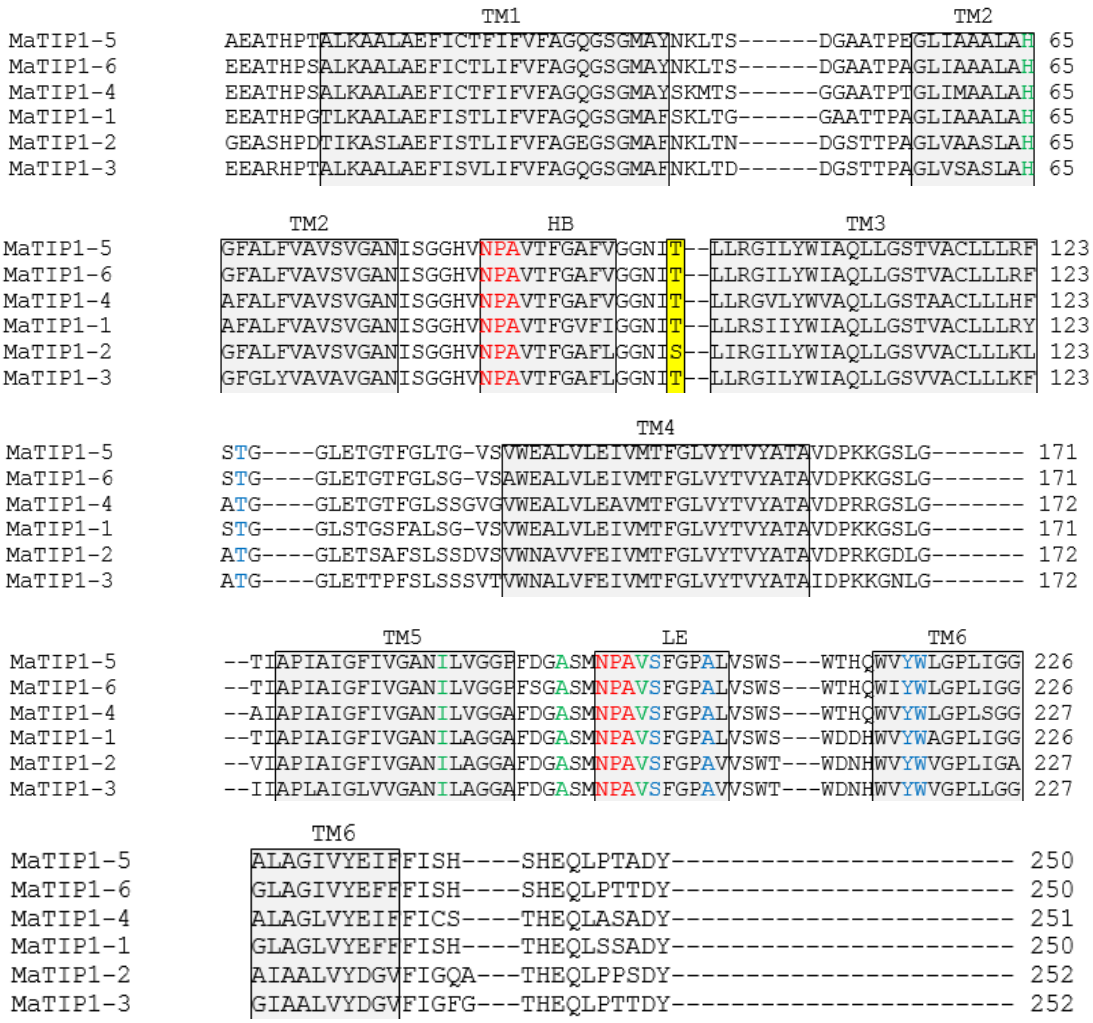

Figure 3. Part of the multiple sequence alignment of MaMIPs. Trans-membrane helices (TM1-TM6) and the two short helices forming the two NPAs (HB and HE) (shaded), NPA motifs (red) P1-P5 residues (blue), ar/R selectivity filter residues (green).

Table 1. Characteristics of conserved residues in primary sequences which determine substrate selectivity in banana MIPs.

Name	NPA		ar/R filter				Froger's Position					Putative non-aqua substrate
	LB	LE	TM2	TM5	LE1	LE2	P1	P2	P3	P4	P5	
MaPIP1-1	NPA	NPA	F	H	T	R	K	S	A	F	W	
MaPIP1-2	NPA	NPA	F	H	T	R	Q	S	A	F	W	Boron, CO_2, H_2O_2, Urea
MaPIP1-3	NPA	NPA	F	H	T	R	Q	S	A	F	W	Boron, CO_2, H_2O_2, Urea
MaPIP1-4	NPA	NPA	F	H	T	R	R	S	A	F	W	Boron, CO_2, Urea
MaPIP1-5	NPA	NPA	F	H	T	R	Q	S	A	F	W	H_2O_2, Urea
MaPIP1-6	NPA	NPA	F	H	T	R	Q	S	A	F	W	Boron, CO_2, H_2O_2, Urea
MaPIP1-7	NPA	NPA	F	H	T	R	Q	S	A	F	W	Boron, CO_2, H_2O_2, Urea
MaPIP1-8	NPA	NPA	F	H	T	R	Q	S	A	F	W	Boron, CO_2, H_2O_2, Urea
MaPIP1-9	NPA	NPA	F	H	T	R	Q	S	A	F	W	Boron, CO_2, H_2O_2, Urea
MaPIP2-1	NPA	NPA	F	H	T	R	Q	S	A	F	W	Boron, CO_2, H_2O_2, Urea
MaPIP2-2	NPA	NPA	F	H	T	R	Q	S	A	F	W	H_2O_2, Urea
MaPIP2-3	NPA	NPA	F	H	T	R	Q	S	A	F	W	H_2O_2, Urea
MaPIP2-4	NPA	NPA	F	H	T	R	Q	S	A	F	W	H_2O_2, Urea
MaPIP2-5	NPA	NPA	F	H	T	R	M	S	A	F	W	CO_2, Urea
MaPIP2-6	NPA	NPA	F	H	T	R	M	S	A	F	W	CO_2, Urea
MaPIP2-7	NPA	NPA	F	H	T	R	M	S	A	F	W	CO_2, Urea
MaPIP2-8	NPA	NPA	F	H	T	R	M	S	A	F	W	Urea
MaPIP2-9	NPA	NPA	F	H	T	R	M	S	A	F	W	Urea
MaPIP2-10	NPA	NPA	F	H	T	R	M	S	A	F	W	CO_2, Urea
MaTIP1-1	NPA	NPA	H	I	A	V	T	S	A	Y	W	H_2O_2, Urea
MaTIP1-2	NPA	NPA	H	I	A	V	T	S	A	Y	W	H_2O_2, Urea
MaTIP1-3	NPA	NPA	H	I	A	V	T	S	A	Y	W	H_2O_2, Urea
MaTIP1-4	NPA	NPA	H	I	A	V	T	S	A	Y	W	H_2O_2, Urea
MaTIP1-5	NPA	NPA	H	I	A	V	T	S	A	Y	W	H_2O_2, Urea
MaTIP1-6	NPA	NPA	H	I	A	V	T	S	A	Y	W	H_2O_2, Urea
MaTIP2-1	NPA	NPA	H	I	G	R	T	S	A	Y	W	Ammonia, H_2O_2, Urea
MaTIP2-2	NPA	NPA	H	I	G	R	T	S	A	Y	W	Ammonia, H_2O_2, Urea
MaTIP2-3	NPA	NPA	H	I	G	R	T	S	A	Y	W	Ammonia, H_2O_2, Urea
MaTIP2-4	NPA	NPA	H	I	G	R	T	S	A	Y	W	Ammonia, H_2O_2 , Urea
MaTIP2-5	NPA	NPA	H	I	G	R	T	S	A	Y	F	Urea
MaTIP3-1	NPA	NPA	H	I	A	R	T	A	A	Y	W	Boron, H_2O_2, Urea
MaTIP3-2	NPA	NPA	H	I	A	R	T	A	A	Y	W	H_2O_2, Urea
MaTIP4-1	NPA	NPA	Q	T	A	R	A	S	A	Y	W	GlpF like
MaTIP4-2	NPA	NPA	Q	S	A	R	A	S	A	Y	W	GlpF like
MaTIP4-3	NPA	NPA	H	I	A	R	T	S	A	Y	W	H_2O_2, Urea
MaTIP5-1	NPA	NPA	Q	V	G	R	S	S	A	Y	W	GlpF like
MaNIP1-1	NPA	NPA	W	V	A	R	F	T	A	Y	I	
MaNIP1-2	NPA	NPA	W	V	A	R	F	S	A	Y	V	Urea
MaNIP2-1	NPA	NPA	G	S	G	R	L	T	A	Y	F	Silicon, Urea
MaNIP2-2	NPA	NPA	G	S	G	R	L	T	A	Y	L	Urea
MaNIP2-3	NPA	NPA	G	S	G	R	L	T	A	Y	F	Silicon, Urea
MaNIP2-4	NPA	NPA	G	S	G	R	L	T	A	Y	F	Silicon, Urea

Table 1. Continued.

MaNIP3-1	NPS	NPV	A	I	G	R	F	T	A	Y	L	Boron
MaNIP3-2	NPS	NPV	A	I	G	R	F	T	A	Y	M	
MaSIP1-1	NPT	NPA	I	V	P	N	M	A	A	Y	W	
MaSIP2-1	NPL	NPA	V	N	G	S	F	A	A	Y	W	
MaSIP2-2												

and MaTIP3-2, which have a basic and natural pI, respectively (Table 2). All MaTIP transporters have the two canonical NPA motifs. The ar/R selectivity filter varies among different subgroups of banana MIPs. The residue at TM2 in all of the MaTIPs is H except in MaTIP4-1, MaTIP4-2, and MaTIP5-1, where Q is found. In the 5th TM, all MaTIPs have I which is replaced by T and S in MaTIP4-1 and MaTIP4-2, respectively. The residues at LE1 and LE2 were found to be specific for each subgroup in MaTIPs. While MaTIP1, MaTIP3, and MaTIP4 have an A in LE1, MaTIP2 and MaTIP5 subgroups show G at the similar positions. The residue at LE2 in all MaTIPs is R except for MaTIP1 subgroup, which has a V residue. The P1 in MaTIPs is a conserved T, except for MaTIP4 and MaTIP5 subgroups. P2 was found to be S in all MaTIPs, but in MaTIP3 subgroup A is replaced. The Froger's position A-Y-W at P3-P5 is found to be strictly conserved in MaTIPs except for MaTIP2-5, where F has occupied the P5 position. Prediction of functional characters of MaTIPs shows that MaTIP2 subgroups except MaTIP2-5 are NH3 transporters (Table 1). All MaTIPs are predicted to be the transporter of urea except for MaTIP4-1, MaTIP4-2 and MaTIP5-1. Except for MaTIP2-5, MaTIP4-1, MaTIP4-2, and MaTIP5-1, H_2O_2 is the non-aqua substrate for all the rest of MaTIPs.

3.3.3. MaNIP subfamily: transporters of Si, urea, and boron

62-92% sequence similarity is observed in the MaNIP subfamily, which is divided into three subgroups in banana. The polypeptide length of this subfamily is quite variable from 263-301 amino acids. Their MW ranges from 27-31 KDa and have a basic pI (Table 2). The two NPA motifs in MaNIPs are variable. In MaNIP3 subgroup, the A at NPA is replaced by S and V at the first and second NPA motifs. Residues that form the ar/R constrictions are also variable, though within different subgroups a high degree of conservation is observed. The residue at TE2 is a totally conserved R in all MaNIPs. The TM2, TM5, and TE1 are W-V-A, G-S-G, and A-I-G in MaNIP1, MaNIP2, and MaNIP3, respectively. Except for P3 (A) and P4 (Y), which are conserved across all the MaNIPs, residues at P1 and P5 are quite variable (Table 1). While P1 (F/L) is conserved in each MaNIP subgroup, P5 is any of the I/V/F/L/M. P2 is T in all MaNIPs, except for MaNIP1-2, where S is found instead. Evaluation of MaNIP functions shows that NIP2-1, NIP2-3, and NIP2-4 are potentially Si transporters. All the MaNIP2 subgroups have been predicted as urea transporters, while MaNIP3-1 is a boron transporter.

3.3.4. MaSIP subfamily

Banana has relatively fewer SIPs compared to the other MaMIPs. The three MaSIPs are clustered into two subfamilies with a 45-89% similarity. With a polypeptide length of 238-261 amino acids, MW of MaSIPs is in the range of 25-29 KDa, and the pI is about 9 (Table 2). While the second NPA is perfectly conserved, MaSIP subfamily shows a less conserved first NPA motif, where A is replaced by T and L at MaSIP1 and MaSIP2 subgroups, respectively (Table 1). The residues at TM2 and TM5 are completely variable in MaSIPs, but the LE1 and LE2 of ar/R selectivity filter is conserved in each MaSIP subfamily. Interestingly the P2-P5 position is conserved in all MaSIPs with the A-A-Y-W. The P1 of Froger's position is M in MaSIP1, where is replaced by F in MaSIP2 subfamily. So far no data regarding the physiological role of SIPs is available.

4. Discussion

4.1. Characterization of the MaMIPs

Substrate transport selectivity of MIPs varies in different plant species. While Froger's position discriminates AQPs from GlpFs (7), the

Table 2. Physicochemical characteristics of MaMIPs. Polypeptide length (PL), Isoelectric point (pI), Molecular weight (MW), Protein subcellular localization (PSL).

Protein name	PL (aa)	pI	MW (KDa)	PSL	Protein name	PL (aa)	pI	MW (KDa)	PSL
MaPIP1-1	285	9.10	30.4033	plasma	MaTIP1-6	250	5.79	25.8439	vacuole
MaPIP1-2	285	8.96	30.4764	plasma	MaTIP2-1	249	5.37	24.8489	cytosol
MaPIP1-3	287	8.81	30.7797	plasma	MaTIP2-2	249	5.67	24.8560	cytosol
MaPIP1-4	287	8.83	30.7607	plasma	MaTIP2-3	243	5.33	24.4926	vacuole
MaPIP1-5	287	9.00	30.9089	plasma	MaTIP2-4	245	5.03	24.7296	plasma
MaPIP1-6	287	8.99	30.6286	plasma	MaTIP2-5	245	5.58	24.8258	plasma
MaPIP1-7	286	9.10	30.6567	plasma	MaTIP3-1	254	9.00	26.6713	cytosol
MaPIP1-8	287	9.16	30.6747	plasma	MaTIP3-2	254	7.07	26.4329	cytosol
MaPIP1-9	287	9.14	30.6717	plasma	MaTIP4-1	255	6.04	26.6515	cyto/plas
MaPIP2-1	287	6.95	30.1659	plasma	MaTIP4-2	254	5.54	26.3899	cyto/plas
MaPIP2-2	287	7.66	30.4863	plasma	MaTIP4-3	249	6.17	25.4390	vacuole
MaPIP2-3	288	8.27	30.6225	plasma	MaTIP5-1	259	8.60	26.3583	vacuole
MaPIP2-4	286	6.95	30.1808	plasma	MaNIP1-1	288	8.61	30.5754	plasma
MaPIP2-5	281	8.71	29.9098	plasma	MaNIP1-2	263	8.64	27.9739	vacuole
MaPIP2-6	282	8.71	30.0481	plasma	MaNIP2-1	293	7.72	30.7356	plasma
MaPIP2-7	282	8.71	29.8958	plasma	MaNIP2-2	293	7.75	30.7297	chloroplast
MaPIP2-8	282	9.02	30.0141	plasma	MaNIP2-3	292	8.98	30.9948	plasma
MaPIP2-9	281	9.42	30.3155	plasma	MaNIP2-4	301	8.57	31.9318	plasma
MaPIP2-10	281	9.16	29.6687	plasma	MaNIP3-1	296	8.63	31.1542	plasma
MaTIP1-1	250	5.55	25.6736	vacuole	MaNIP3-2	264	8.44	27.4558	plasma
MaTIP1-2	252	4.65	25.7316	plasma	MaSIP1-1	238	9.94	25.2168	plasma
MaTIP1-3	252	5.36	26.0322	vacuole	MaSIP2-1	261	9.57	25.9566	chloroplast
MaTIP1-4	251	5.56	25.6877	vacuole	MaSIP2-2	238	9.11	29.2994	plasma
MaTIP1-5	250	5.34	25.8269	vacuole					

ar/R tetrad selectivity filter, as the narrowest part of the pore, acts as the size exclusion barrier (4). The ar/R selectivity filter not only affects the pore size, but also pore hydrophobicity. These two filters, as well as the dual NPA motifs were used to functionally characterize the MaMIPs.

4.1.1. Characterization of the MaPIPs

MaPIP1s like other plant species show longer N terminal than MaPIP2, which might have a role in the permeability difference. Physical interaction between AQPs can change their subcellular localization by forming homo and hetero tetramers (29). Although plasma localization of MaPIPs is in agreement with the traditional view on PIPs, but *in silico* prediction should be validated with experiments (30). Similar to water transporting AQPs, the ar/R filter of banana, rice and maize PIPs is the conserved tetrad F-H-T-R. MaPIP1s transport small molecules like boron, CO_2, H_2O_2, and urea. MaPIP2s are mainly transporters of water molecule, but also permeable to H_2O_2, and urea. We have predicted that most members of the MaPIP1 subfamily and four members of the MaPIP2 subfamily, including MaPIP2-1, MaPIP2-2, MaPIP2-3 and MaPIP2-4 as H_2O_2 transporters. MaPIP2-1 and predominant members of the MaPIP1 were predicted as boron transporters. According to our prediction, a predominant number of PIP1 and some members of the PIP2 subfamily are the only CO_2 transporters in banana. None of the other subfamilies are permeable to CO_2. Similar observa-

tions have been reported for other species as well (31). Over-expression of MaPIP1-2 and MaPIP2-6 in banana plants had a pivotal role in drought and salinity tolerance (11, 32). These transgenic plants also exhibited higher photosynthetic efficiency. Photosynthetic efficiency is a sensitive factor of overall plant health. This is in line with our prediction of MaPIP1-2 and MaPIP2-6 as CO_2 transporters (Table 1). Urea is the only solute to which all the members of MaPIP1 and MaPIP2 subfamilies are permeable.

4.1.2. Characterization of the MaTIPs

Like other species, MaTIPs are classified into five distinct classes (33). The ar/R selectivity filter in MaTIPs differs in each subgroup and has the same profile as other monocots (34). The MaTIP1 subfamily mainly transports H_2O_2 and urea, while acquisition of NH_3 transport is observed in the MaTIP2 subfamily. Experimental evidence on the transport of all three molecules exists for other species as well (35). The H residue at the TM2 of the ar/R selectivity filter in MaTIP4-1, MaTIP4-2, and MaTIP5-1 is replaced by Q. At the TM5 of the ar/R filter of MaTIP4-1, MaTIP4-2, the nonpolar Ileu is substituted by polar residues (ser/thr). Similar observation is reported for other monocot species, such as rice and maize. Substitution of small polar residues results in a wider hydrophilic constriction. The increase in constriction diameter facilitates the transport of hydrophilic solutes like glycerol, and likens this group of MaTIPs to the GlpFs (36). MaTIP1, MaTIP2, and MaTIP3 subfamilies are potentially able to transport H_2O_2. The moderating H_2O_2 effect after pathogen attack might be the origin of vast distribution of H_2O_2 transporters in different MaTIP subfamilies. We have predicted MaTIP3-1 as a boron transporter. All members of MaTIP1, MaTIP2, MaTIP3, and MaTIP5 subfamilies are potentially able to transport urea. Based on our prediction, MaTIP2 subfamily is the only NH_3 transporter (Table 1). Similar observation has been clarified in other plants (37).

4.1.3. Characterization of the MaNIPs

MaNIPs are classified into three subclasses. The MaNIP1-1 and MaNIP1-2 with the signature motif W-V-A-R at the ar/R selectivity filter shows the characteristics of GlpFs. The hydrophobic W-V-A forms van der Waals interaction with the substrate (38). The ar/R selectivity filter of the MaNIP2 subgroup is larger than other MaNIPs, due to the first three small residues in the tetrad of ar/R filter, which is G-S-G at TM2, TM5, and LE1, respectively. Homology modeling of MIPs in rice and maize has been conducted. Results showed that this region is about 0.1 nm larger compared to GlpFs and 0.22 nm larger than water specific plant PIPs (36). Our prediction on the silicon permeability of MaNIP2-1, MaNIP2-3, and MaNIP2-4 is in agreement with the change in pore size and the experimental findings in rice and maize (39). The ar/R selectivity filters in MaNIP3-1 and MaNIP3-2, such as ZmNIP3-1 and OSNIP3-1 are designated as A-I-G-R. It has been reported that NIPs, such as AtNIP6-1, which has A residue at the TM2 position, are water impermeable glyceroporins able to transport larger uncharged solutes, including formamide, glycerol, and urea (38). Despite the small size of water molecule and the large pore diameter in the ar/R selectivity filter of these channels, NIPs have limitations in water transport. It has been demonstrated that a wider pore makes the organization of water molecules difficult (40). None of the MaNIPs in our prediction were H_2O_2 transporters. All MaNIP2 subfamily are urea transporters. We predicted MaNIP3-1 as a boron transporter.

4.1.4. Characterization of the MaSIPs

MaSIPs are distantly related to other MaMIPs. In MaSIPs, as well as the SIPs in rice and maize, the hydrophobic nature of the ar/R selectivity filter in comparison to PIPs, TIPs, and NIPs might result in a distinct transport activity. Prediction on the permeability of MaPIPs, MaTIPs, and MaNIPs to urea is in agreement with the interpretation of the separation of the subfamilies after acquisition of urea transport activity in a common ancestor (4). The substrate specificity of MaSIPs needs experimental studies.

5. Conclusion

This study characterized 47 AQP genes in banana. Phylogenetic analysis divided MaMIPs

into four subfamilies. Most MaMIPs showed identical ar/R selectivity type filters with their corresponding orthologues in rice and maize. Predictive analysis shows that MaMIPs contain true aquaporins, aquaglyceroporins, and glyceroporins. This finding needs to be validated by *in vitro* experiments. Affecting water permeability and ion transport, different isoforms of MaMIP provide a tight cell osmoregulation under stress. Our analysis of MIP genes from the *Musa* genome deciphers the role of different AQPs during development and stress. This study provides the opportunity to manipulate carbon gain and water use efficiency, which in turn allow banana to cope with environmental stress. It is also a model for how to perform *in silico* data mining of the genomic resources for non-model crops, taking advantage of what is well known about model crops.

Acknowledgment

The author would like to thank Shiraz University of Medical Sciences, Shiraz, Iran.

Conflict of Interest

None declared.

6. References

1. Maurel C. Plant aquaporins: novel functions and regulation properties. *FEBS Lett*. 2007;581:2227-36.
2. Kaldenhoff R, Bertl A, Otto B, Moshelion M, Uehlein N. Characterization of Plant Aquaporins. *Methods Enzymol*. 2007;428:505-31.
3. Maurel C, Santoni V, Luu DT, Wudick MM, Verdoucq L. The cellular dynamics of plant aquaporin expression and functions. *Curr Opin Plant Biol*. 2009;12:690-8.
4. Hove RM, Bhave M. Plant aquaporins with non-aqua functions: deciphering the signature sequences. *Plant Mol Biol*. 2011;75:413-30.
5. Gomes D, Agasse A, Thiebaud P, Delrot S, Geros H, Chaumont F. Aquaporins are multifunctional water and solute transporters highly divergent in living organisms. *Biochim Biophys Acta*. 2009;1788:1213-28.
6. Forrest KL, Bhave M. Major intrinsic proteins (MIPs) in plants: a complex gene family with major impacts on plant phenotype. *Funct Integr Genomics*. 2007;7:263-89.
7. Froger A, Tallur B, Thomas D, Delamarche C. Prediction of functional residues in water channels and related proteins. *Protein Sci*. 1998;7:1458-68.
8. Perrier X, De Langhe E, Donohue M, Lentfer C, Vrydaghs L, Bakry F, *et al*. Multidisciplinary perspectives on banana (*Musa* spp.) domestication. *Proc Natl Acad Sci USA*. 2011;108:11311-8.
9. Kumar KS, Bhowmik D. Traditional and medicinal uses of banana. *J Pharmacogn Phytochem*. 2012;1(3).
10. Fagbemi JF, Ugoji E, Adenipekun T, Adelowotan O. Evaluation of the antimicrobial properties of unripe banana (*Musa sapientum* L.), lemon grass (*Cymbopogon citratus* S.) and turmeric (*Curcuma longa* L.) on pathogens. *African J Biotechnol*. 2009;8:1176-82.
11. Sreedharan S, Shekhawat UK, Ganapathi TR. Transgenic banana plants overexpressing a native plasma membrane aquaporin MusaPIP1;2 display high tolerance levels to different abiotic stresses. *Plant Biotechnol J*. 2013;11:942-52.
12. Shorrocks VM. The occurrence and correction of boron deficiency. *Plant Soil*. 1997;193:121-48.
13. Shapira OR, Israeli Y, Shani U, Schwartz A. Salt stress aggravates boron toxicity symptoms in banana leaves by impairing guttation. *Plant Cell Environ*. 2013;36:275-87.
14. Prasertsak P, Freney J, Saffigna P, Denmead O, Prove B. Fate of urea nitrogen applied to a banana crop in the wet tropics of Queensland. *Nutr Cycl Agroecosys*. 2001;59:65-73.
15. Henriet C, Bodarwé L, Dorel M, Draye X, Delvaux B. Leaf silicon content in banana (*Musa* spp.) reveals the weathering stage of volcanic ash soils in Guadeloupe. *Plant Soil*. 2008;313:71-82.
16. Kaldenhoff R. Mechanisms underlying CO_2 diffusion in leaves. *Curr Opin Plant Biol*. 2012;15:276-81.
17. Quan LJ, Zhang B, Shi WW, Li HY. Hydrogen peroxide in plants: a versatile molecule of the reactive oxygen species network. *J Integr Plant Biol*. 2008;50:2-18.
18. D'Hont A, Denoeud F, Aury JM, Baurens FC, Carreel F, Garsmeur O, *et al*. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature*. 2012 ;488:213-7.
19. Droc G, Lariviere D, Guignon V, Yahiaoui N, This D, Garsmeur O, *et al*. The banana genome hub. *Database* . 2013;2013:bat035.

20. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, *et al.* Pfam: the protein families database. *Nucleic Acids Res*. 2014;42:D222-30.
21. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, *et al*. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* 2015;43:D222-6.
22. Mitchell A, Chang HY, Daugherty L, Fraser M, Hunter S, Lopez R, *et al*. The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res*. 2015;43:D213-21.
23. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* 2001 19;305:567-80.
24. Hofman K. TMbase-A database of membrane spanning protein segments. *Biol Chem Hoppe-Seyler*. 1993;374:166.
25. Wong W-C, Maurer-Stroh S, Eisenhaber F. More than 1,001 problems with protein domain databases: transmembrane regions, signal peptides and the issue of sequence homology. *PLoS Comput Biol.* 2010;6:e1000867.
26. Cserzo M, Eisenhaber F, Eisenhaber B, Simon I. On filtering false positive transmembrane protein predictions. *Protein Eng.* 2002;15:745-52.
27. Nugent T, Jones DT. Transmembrane protein topology prediction using support vector machines. *BMC bioinformatics.* 2009;10:159.
28. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 2013;30:2725-9.
29. Chaumont F, Tyerman SD. Aquaporins: highly regulated channels controlling plant water relations. *Plant Physiol.* 2014;164:1600-18.
30. Chevalier AS, Chaumont F. Trafficking of plant plasma membrane aquaporins: multiple regulation levels and complex sorting signals. *Plant Cell Physiol.* 2015;56:819-29.
31. Hanba YT, Shibasaka M, Hayashi Y, Hayakawa T, Kasamo K, Terashima I, *et al*. Overexpression of the barley aquaporin HvPIP2;1 increases internal CO(2) conductance and CO(2) assimilation in the leaves of transgenic rice plants. *Plant Cell Physiol.* 2004;45:521-9.
32. Sreedharan S, Shekhawat UK, Ganapathi TR. Constitutive and stress-inducible overexpression of a native aquaporin gene (MusaPIP2;6) in transgenic banana plants signals its pivotal role in salt tolerance. *Plant Mol Biol.* 2015;88:41-52.
33. Regon P, Panda P, Kshetrimayum E, Panda SK. Genome-wide comparative analysis of tonoplast intrinsic protein (TIP) genes in plants. *Funct Integr Genomics.* 2014;14(4):617-29.
34. Forrest KL, Bhave M. The PIP and TIP aquaporins in wheat form a large and diverse family with unique gene structures and functionally important features. *Funct Integr Genomics.* 2008;8:115-33.
35. Li GW, Peng YH, Yu X, Zhang MH, Cai WM, Sun WN, et al. Transport functions and expression analysis of vacuolar membrane aquaporins in response to various stresses in rice. *J Plant Physiol.* 2008;165:1879-88.
36. Bansal A, Sankararamakrishnan R. Homology modeling of major intrinsic proteins in rice, maize and *Arabidopsis*: comparative analysis of transmembrane helix association and aromatic/arginine selectivity filters. *BMC Struct Biol.* 2007;7:27.
37. Liu LH, Ludewig U, Gassert B, Frommer WB, von Wiren N. Urea transport by nitrogen-regulated tonoplast intrinsic proteins in *Arabidopsis*. *Plant Physiol.* 2003;133:1220-8.
38. Wallace IS, Roberts DM. Distinct transport selectivity of two structural subclasses of the nodulin-like intrinsic protein family of plant aquaglyceroporin channels. *Biochemistry.* 2005;44:16826-34.
39. Ma JF, Tamai K, Yamaji N, Mitani N, Konishi S, Katsuhara M, *et al*. A silicon transporter in rice. *Nature.* 2006;440:688-91.
40. Tajkhorshid E, Nollert P, Jensen MO, Miercke LJ, O'Connell J, Stroud RM, et al. Control of the selectivity of the aquaporin water channel family by global orientational tuning. *Science*. 2002;296:525-30.