Research article

Major intrinsic proteins repertoire of *Morus notabilis* and their expression profiles in different species

CrossMark

Vinay Kumar Baranwal, Paramjit Khurana[*]

Department of Plant Molecular Biology, University of Delhi South Campus, Benito Juarez Road, New Delhi, 110021, India

ABSTRACT

Leaf moisture content in *Morus* is a significant trait regulating the yield of silk production. Studies have shown that fresh leaves or leaves with high water content are preferably eaten by silk worm. Water and certain other molecules transport in plants is known to be regulated by aquaporins or Major Intrinsic Proteins (MIPs). Members of the MIP gene family have also been implicated in plant development and stress responsiveness. To understand how members of MIP gene family are regulated and evolved, we carried out an extensive analysis of the gene family. We identified a total of 36 non redundant MIPs in *Morus notabilis* genome, belonging to five subfamilies −PIPs, TIPs, NIPs, XIPs and SIPs) have been identified. We performed a Gene ontology (GO) term enrichment analysis and looked at distribution of *cis* elements in their 2K upstream regulatory region to reveal their putative roles in various stresses and developmental aspects. Expression analysis in developmental stages revealed their tissue preferential expression pattern in diverse vegetative and reproductive tissues. Comparison of expression profiles in the leaves of three species including *Morus notabilis*, *Morus serrata* and *Morus laevigata* led to identification of differential expression in these species. In all, this study elaborates a basic insight into the structure, function and evolutionary analysis of MIP gene family in *Morus* which is hitherto unavailable. Our analysis will provide a ready reference to the mulberry research community involved in the *Morus* improvement program.

© 2016 Elsevier Masson SAS. All rights reserved.

1. Introduction

Water is one of the most important constituent in cells and the basis for major chemical and physiological processes of the plant. Sophisticated mechanisms have evolved to maintain water homeostasis in every organism. In higher plants, Major Intrinsic Proteins (MIPs) form important component of symplastic and transcellular water transport mechanism (Reddy et al., 2015). MIPs are channel proteins facilitating transport of water and other solutes across membranes (Chaumont et al., 2001). Besides transporting water, MIPs are also involved in transportation of small molecules including glycerol, urea, ammonia etc. and in regulating cell differentiation and growth. They show a tissue specific expression profile and are known to express in various organs including root, leaves, pollen and pistil. Accumulation of MIP proteins and their enhanced expression has been established in hormone treatment and stress. The gene family has been widely studied in monocots like *Sorghum bicolor* (Reddy et al., 2015), Maize (Chaumont et al., 2001), Rice (Sakurai et al., 2005) and dicots including *Arabidopsis* (Johanson et al., 2001), Brassica (Diehn et al., 2015), *Populus trichocarpa* (Gupta and Sankararamakrishnan, 2009), Cotton (Park et al., 2010), Tomato (Reuscher et al., 2014), Soybean (Zhang et al., 2013) Potato (Venkatesh et al., 2013) and Sweet Orange (de Paula Santos Martins et al., 2015). The functionality of several aquaporins has been studied in great detail in many model systems. They regulate stomatal aperture gating, an important mitigatory mechanism of drought tolerance, in rice (Vinnakota et al., 2015). Wheat transgenics over-expressing SbPIP11from *Salicornia bigelovii* have shown enhanced accumulation of the osmolyte proline and increased soluble sugar biosynthesis (Yu et al., 2015). In *Gossypium*, MIPs are regulated as early responsive proteins when salinity stress treatment was given to the seedlings (Li et al., 2015). NIP5; 1 channel, a major intrinsic protein has been shown to regulate the selective uptake of urea (Yang et al., 2015) in *Arabidopsis*. NOD-26 like major intrinsic protein SspNIP2 in sugarcane has been shown to be up-regulated after 30 min of cold stress treatment (Park et al., 2015). Further, SsNIP2 harboring

transgenic has been shown to perform better with respect to control when challenged with salinity stress. Expression of specific MIPs has also been observed in reproductive tissues. Previously, our lab has shown the differential responses of various transporters in ten genotypes of Mulberry (Das et al., 2013). In *Morus*, TIP1, TIP2 and PIP1 have been shown to have differential expression in various development stages including leaf, seedling, root and flower with higher accumulation in leaf tissue (Lal et al., 2009). In response to various abiotic stresses including aerial drying, cold, salinity, ABA and mercury treatment, expression of the genes varied differentially. TIP1 has accumulates preferentially in response to salinity, mercury and ABA stress (Lal et al., 2009). Further, this analysis was extended and qPCR based estimation was made for four other PIPs and a TIP in 30 days old seedling of same cultivar (Checker et al., 2012). Two PIPs i.e. PIP2B and PIP2; 8 have shown relatively higher up-regulation in leaf of thirty day old seedling with respect to root. Two PIP I.e.e PIP1C and PIP1; 5 along with TIP2; 1 have shown relatively more up-regulation in root tissue with respect to leaf. All these five MIPs studies have shown up-regulation with respect to control. Water has important role in anther dehiscence and hydration of pollen. PIP1 and PIP2 of *Nicotiana tabacum* which encode for active water channels have been shown to express during pollen developmental stages (Bots et al., 2005). Similarly, in wild potato *Solanum chacoense,* PIP2 expression was found to be positively regulated in pistil and anther developmental stages (O'Brien et al., 2002). These analyses led us to conclude that this gene family is associated with a broad range of activity ranging from stress response to vegetative and reproductive growth.

Although global silk production has shown a positive growth trend, it is has not been able to cope with the growing demand. *Morus* is the basis of silk industry as *Bombyx mori* which is the source of 90% silk is monophagous on it. The worms preferentially feed on Morus leaves. It has been established that there is a direct relation between leaf quality and yield. Wild species like *Morus notabilis* (MN) (He et al., 2013), *M. laevigata* (ML) and *M. serrata* (MS) (Tikader and Kamble, 2008) harbor an array of genes which could be used in mulberry improvement programs. A comparative account of various gene families in these three genotypes have been made showing differential response in various conditions (Baranwal et al., 2016; Baranwal and Khurana, 2016; Saeed et al., 2016a) including development and stresses. MN is a wild species and native to Sichuan province on China. Many important genes related to biotic and abiotic stress have been isolated and characterized from MN recently (Liu et al., 2015; Wei et al., 2014). ML and MS grow widely in India with varied distribution pattern. ML is distributed throughout the country while MS is restricted to the foothills of the Himalayas (Tikader and Kamble, 2008). ML shows greater phenotype diversity and comparatively better adaptability than MS. Both ML and MS are resistant to powdery mildew caused by Erysiphaceae member *Phyllactinia corylea* (Babu et al., 2002). Apart from being used in silk industry, ML in India is widely used for timber and forage (Chatterjee et al., 2004). Various accessions of mulberry including ML and other such species have been analyzed for their utility in breeding programs (Chattopadhyay et al., 2010). Availability of genome has created an opportunity to analyze the gene family, its orientation, evolution and also facilitate comparative transcriptomic studies in related species. Based on its reported roles in stress and reproductive developments, we selected the MIP gene family for a detailed analysis in a genome-wide context that could give detailed insights of its structure, function and evolutionary dynamics. A comparison of expression profiles of these genes was also made in three wild species of *Morus*. This work would immensely benefit those working on Mulberry across the world to select genes and characterize them further for their role and creating high yielding varieties.

2. Material and methods

2.1. Identification of MIPs and phylogenetic analysis

PfamA.hmm file provided by Pfam database (ftp://ftp.ebi.ac.uk/pub/databases/Pfam/current_release/) was used to fetch hmm profile of Major Intrinsic Protein with accession number PF00230.15 using hmmfetch program of hmm suit version 3.12b2. This hmm profile was used to query the entire predicted proteome of MN (Li et al., 2014) with default parameters using hmmsearch program of the same suit. Along with that, protein sequences harboring these domains from *Arabidopsis* and rice from TAIR (https://www.arabidopsis.org/tools/bulk/sequences/index.jsp) and RGAP (http://rice.plantbiology.msu.edu/) databases respectively were downloaded and used to query the *Morus* proteome using blastp program. Members thus identified were queried for the presence of MIP domain using NCBI Conserved Domain Database (NCBI-CDD) (http://www.ncbi.nlm.nih.gov/cdd/); (Marchler-Bauer et al., 2011). Domains from those members which showed the presence of this domain were fetched and incorporated into existing hmm profile. The new hmm profile thus generated was used again to search the proteome and this exercise was repeated till no new members could be identified. These identified members were searched for the presence of transmembrane helices using five tools including TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) (Krogh et al., 2001), TMPred Server (http://embnet.vital-it.ch/software/TMPRED_form.html), HMMTop (http://www.enzim.hu/hmmtop/) (Tusnády and Simon, 2001), DAS-TMfilter server (http://mendel.imp.ac.at/sat/DAS/DAS.html) (Cserzö et al., 2002) and CCTOP (http://cctop.enzim.ttk.mta.hu/) (Dobson et al., 2015) which employ different algorithms. Those MIPs which showed ≤5 such helices were removed from the list.

For phylogenetic analysis, complete protein sequences of identified *Morus,* rice *Arabidopsis, Physcomitrella patens* and *Populus trichocarpa* were aligned with default parameters in Clustal X version 2.1. Based on the alignment, a bootstrapped maximum likelihood tree was prepared in MEGA version 7.0.18 with 1000 replications. Bootstrap replication is an established statistical method to find out nonparametric errors where taxa are held constant and the amino acids are resampled randomly with replacement (Hedges, 1992). Further to understand the evolution of the MIP members in *Morus,* a bootstrapped NJ tree based on the alignment of only *Morus* members were created following the same parameters defined above.

For estimation of non-synonymous to synonymous mutations ratio (dN/dS), PAML (http://abacus.gene.ucl.ac.uk/software/paml.html) based calculation was done using a dNdS-calculator (Shaw et al., 2012). To make the comparison, best matches for *Morus* MIPs in BLASTn search were compared to MIP proteins of *Populus trichocarpa* and *Physcomitrella patens* and pair-wise comparisons were made.

2.2. Gene structure and putative promoters and cis regulatory elements

To analyze the gene structure, sequences of predicted genes and mRNAs from the MorusDB (Li et al., 2014) were fetched and used in Gene Structure Display Server (GSDS; http://gsds.cbi.pku.edu.cn/) (Hu et al., 2015) from Center for Bioinformatics, Peking University to display the intron exon junctions. Junctions were further confirmed by the predictions made at MorusDB. Molecular mass and Isoelectric focusing point was calculated using the protein sequences in custom written perl scripts utilizing bioperl modules (Stajich et al., 2002). Conserved motifs in protein sequences were identified using MEME suit (http://meme-suite.org/) (Bailey et al.,