Genome-Wide Bioinformatics Analysis of Aquaporin Gene Family in Maize (Zea mays L.)

Amna Bari1, Muhammad Farooq2, Athar Hussain2*, Muhammad Tahir ul Qamar2, Malik Waseem Abbas2, Ghulam Mustafa3, Asad Karim4, Imtiaz Ahmed5 and Tahir Hussain1

1Government College University Faisalabad, Faisalabad, Pakistan
2National Institute for Biotechnology and Genetic Engineering (NIBGE), Faisalabad, Pakistan
3College of Informatics, Huazhong Agricultural University, Wuhan, China
4Jamil-ur-Rahman Center for Genome Research, International Center for Chemical and Biological Sciences, University of Karachi, Pakistan
5School of Chemistry and Environment, Beihang University, Beijing 100191, China

*Corresponding author: Athar Hussain, National Institute for Biotechnology and Genetic Engineering (NIBGE), Jang Road, Faisalabad, Pakistan, Tel: +923469556582; E-mail: atharmutahari@gmail.com

Received date: March 08, 2018; Accepted date: March 28, 2018; Published date: April 06, 2018
Copyright: © 2018 Bari A, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Aquaporins are a super family of major intrinsic proteins, which facilitate the fast and passive movement of water across the cell membrane. This study presented genome-wide identification, characterization and functional prediction of aquaporins in maize using bioinformatics. A total of 41 non-redundant putative aquaporins were identified and were classified into four subfamilies: 18 TIPs, 12 PIPs, 8 NIPs and 3 SIPs. The finding reveals that exon-intron organization were conserved within subfamilies. Several transmembrane domains(TM1-TM6) were predicted by analyzing of conserved domains and motifs, along with various selectivity filters(ar/R). The functional prediction demonstrated ZmAQPs roles in regulation of multiple compounds i.e. water, glycerol, carbohydrates, metal ions and others small solutes. Furthermore, ZmAQPs were the crucial constituent of membranous structure such as plasma membrane and vacuolar membrane etc. These results deliver valuable information to address function of ZmAQPs as well as provide basic data for the improvement of plant growth and development.

Keywords: Bioinformatics; Aquaporins; Conserved motifs; Phylogenetic tree; Gene structure; Trans-membrane

Introduction

Plants are the most adversely affected entities by water scarcity due to exponential depletion of underground water level. The movement of water from underground to plants is mainly carried out through their roots by three parallel pathways i.e. symplastic, trans-cellular or apoplastic [1]. Aquaporins (AQPs) are an ancient channel protein family embedded in membranous structures of plants that transport water and other neutral metabolites across membranes [2]. Most of them are involved in hydraulic conductivity and have potential to increase 10 to 20-fold of water transportation across plasma membrane [3]. This property of AQPs is very important in plants to hold different activities like maintenance and regulation of water [4], cell elongation [5], soil-water relations [6,7], plant cell osmo-regulation [8], seed germination [9] and even in plant reproduction [10,11]. Aquaporins also influence leaf movements and their physiology [12], salt tolerance, fruit ripening [13] and drought resistance in plants [14].

As these proteins cover major portion of membranous structure so named as major intrinsic protein (MIP). The phylogenetic analysis revealed that the MIP encoding genes can be broadly classified into four different sub-families i.e. plasma membrane intrinsic proteins (PIPs), tonoplast intrinsic proteins (TIPs), nodulin-26 intrinsic proteins (NIPs) and small basic intrinsic proteins (SIPs) [3,15-17]. Additionally, three more subfamilies including glycerol intrinsic proteins (GIPs), hybrid intrinsic proteins (HIPs), and unrecognized X intrinsic proteins (XIPs) have also been reported in non-vascular entities by water scarcity due to exponential depletion of underground water level. The movement of water from underground to plants is mainly carried out through their roots by three parallel pathways i.e. symplastic, trans-cellular or apoplastic [1]. Aquaporins (AQPs) are an ancient channel protein family embedded in membranous structures of plants that transport water and other neutral metabolites across membranes [2]. Most of them are involved in hydraulic conductivity and have potential to increase 10 to 20-fold of water transportation across plasma membrane [3]. This property of AQPs is very important in plants to hold different activities like maintenance and regulation of water [4], cell elongation [5], soil-water relations [6,7], plant cell osmo-regulation [8], seed germination [9] and even in plant reproduction [10,11]. Aquaporins also influence leaf movements and their physiology [12], salt tolerance, fruit ripening [13] and drought resistance in plants [14].

As these proteins cover major portion of membranous structure so named as major intrinsic protein (MIP). The phylogenetic analysis revealed that the MIP encoding genes can be broadly classified into four different sub-families i.e. plasma membrane intrinsic proteins (PIPs), tonoplast intrinsic proteins (TIPs), nodulin-26 intrinsic proteins (NIPs) and small basic intrinsic proteins (SIPs) [3,15-17]. Additionally, three more subfamilies including glycerol intrinsic proteins (GIPs), hybrid intrinsic proteins (HIPs), and unrecognized X intrinsic proteins (XIPs) have also been reported in non-vascular...
characterization and evolutionary analysis of aquaporins of Arabidopsis and chickpea. Furthermore, we also predicted many important biological features of maize aquaporins, including protein sequence analysis, identification of trans-membrane domains, conserved motif analysis with respect to phylogenetic tree, gene structure analysis with evolutionary tree, chromosomal distribution, gene ontology includes biological process, molecular function and sub-cellular localization. These bioinformatics results might be helpful for further experimental analysis of aquaporins in maize genome.

Materials and Methods

Database searching and identification of ZmAQP genes

Aquaporin protein sequences from Arabidopsis thaliana (35 sequences) [24] and Cicer aritinum (40 sequences) [31] were used as a query in different database search engines including NCBI-Blastp, Phytosome-Blast and MaizDB-Blast. Furthermore, the position specific iteration (PSI) was also used to make it more specific [32]. The proteins having query similarity more than 75% identified through blastp were selected. After removal of redundant sequences, the identified aquaporins were validated using CD-Search program [33] for presence of MIPs domain. Finally, the data related to these selected ZmAQPs was collected including amino acid sequences, cDNA and genomic DNA sequences from MaizDB [34].

Multiple sequence alignment and phylogenetic tree analysis

The full length amino acid sequences of 41 identified ZmAQP genes were aligned using MUSCLE program [35] and the aligned sequences used for phylogenetic tree construction through maximum likelihood method with 1000 bootstrap value in PhyML3.0 [36]. The phylogenetic tree was visualized in MEGA 6.0 [37]. In order to authenticate the evolutionary tree another program, Mr Bayes v3.2.6 [38] was also used. Sampling over the space of all possible reversible substitution models and prior for the amino acid model to mixed. The Phylogenetic tree was classified into four subfamilies NIPs, TIPs, PIPs and SIPs, based on known nomenclature of AQPs that were used as query in initial BLAST search. Thirty-one ZmAQPs were already annotated with AQPs names by Chaumont [3] and the 10 ZmAQPs genes were named according to their cumulative phylogenetic tree of ZmAQPs with Arabidopsis [24] and chickpea AQPs [31].

Protein characterization and identification of conserved domain, conserved motifs, trans membrane domain and ar/R selectivity filter of ZmAQPs

In order to determine the characteristics of ZmAQPs, the amino acid length (a.a), molecular weight (kDa) and iso-electric point (pI) were predicted through the ExPaSy server (http://web.expasy.org). Conserved domains in ZmAQPs identifies using conserved domain database, NCBI Batch CD search [33]. Transmembrane domains were predicted using TMHMM Server v.2.0 (http://www.cbs.dtu.dk/services/TMHMM/). The NPA motifs, ar/R selectivity filters (H2, H5, LE1, LE2) and Foger’s positions (P1–P5) were identified by careful comparison of multiple sequence alignment of ZmAQPs with AQP protein structure as identified in literature [39-42]. Beside NPA motifs other conserved motifs also predicted using MEME motif discovery server 4 [43] with parameters as described by Hussain [44]. The cellular localization of ZmAQPs were predicted using two other algorithms; Plant-mPLOC (http://www.csbio.sjtu.edu.cn) and WoLF PSORT (http://www.genscript.com/wolf-psort.html).

Gene structure analysis, chromosomal distribution and Gene Ontology (GO) of ZmAQPs

The genomic data including DNA and cDNA sequences, chromosome number, start base pair, end base pair, number of exons as well as introns were retrieved from MaizeGDB (http://www.maizegdb.org) for gene structure analysis and chromosomal distribution. Exon-intron organization of ZmAQPs were analyzed by using Gene Display Server 2 (http://gds.cbi.pku.edu.cn/) [45] and the physical location graph was manually created using MS excel sheet. Gene Ontology of ZmAQP proteins was predicted by Blast2GO program [46] using amino acid sequences with default parameters and using different databases like Swiss-Prot protein, NCBI non-redundant protein (nr), Gene ontology (GO), Kyoto Encyclopedia of Genes(KEGG) protein family and Cluster of Orthologs Groups (COGs).

Results

Identification and prediction of Zea mays Aquaporin's gene family

The database searches resulted in 46 putative AQP genes showing strong matches in maize (Zea may L.) genome. After removal of duplication in gene locus and validation of MIP superfamily domain in amino acid sequences, 41 putative ZmAQPs were proposed in maize genome. A literature survey of AQPs superfamily along with their subfamilies in other plant species were also included for a comparative study, given in Table 1. Generic names (ZmAQPs: ZmPIPs, ZmNIPs, ZmTIPs and ZmSIPs) were assigned for Zca mays aquaporins. The individual gene data and their predicted characteristics are listed in Table 2, including gene name, NCBI accession, chromosome number (Chr#), amino acid residues (aa), Trans-membrane domain (TMD), isoelectric point (Ip), molecular weight (MWkDa), base start pair, end base pair and sub-cellular localization (C.L) (Table 2). Our results demonstrated that all ZmAQPs were unevenly distributed on all 10 chromosomes of maize genome. The amino acid sequences length ranged from 116 bp (ZmTIP4:3a) to 311 bp (ZmTIP4:2a) and lp ranged 5.30 pH (ZmTIP2:3) to 9.79pH (ZmTIP3:3). Similarly, molecular weight (MW) ranged 24.988kDa (ZmTIP2:2) to 31.760 (ZmNIP2:1) (Table 2).

<table>
<thead>
<tr>
<th>Species</th>
<th>Total AQPs</th>
<th>PIP</th>
<th>TIP</th>
<th>NIP</th>
<th>SIP</th>
<th>XIP</th>
<th>Mono/dicot</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Oryza sativa)</td>
<td>33</td>
<td>11</td>
<td>10</td>
<td>10</td>
<td>2</td>
<td>0</td>
<td>nonocot</td>
<td>[48]</td>
</tr>
<tr>
<td>Moso bamboo (Phyllostachys chysedulits)</td>
<td>26</td>
<td>10</td>
<td>6</td>
<td>8</td>
<td>2</td>
<td>0</td>
<td>nonocot</td>
<td>[47]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene name</th>
<th>NCBI accession</th>
<th>Chr#</th>
<th>AA</th>
<th>TMDd</th>
<th>Ip (pH)</th>
<th>MW (kDa)</th>
<th>Start bp</th>
<th>End bp</th>
<th>C.La</th>
<th>C.Lb</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZmNIP1;1</td>
<td>NP_001105721</td>
<td>5</td>
<td>282</td>
<td>6</td>
<td>8.58</td>
<td>29.55</td>
<td>148343777</td>
<td>148345925</td>
<td>C.M</td>
<td>plas: 7</td>
</tr>
<tr>
<td>ZmNIP1;2</td>
<td>NP_001151947</td>
<td>6</td>
<td>284</td>
<td>5</td>
<td>7.67</td>
<td>29.36</td>
<td>131635199</td>
<td>131636926</td>
<td>C.M</td>
<td>plas: 8</td>
</tr>
<tr>
<td>ZmNIP2;1</td>
<td>NP_001105637</td>
<td>5</td>
<td>295</td>
<td>6</td>
<td>6.79</td>
<td>31.76</td>
<td>206972623</td>
<td>206976131</td>
<td>C.M</td>
<td>plas: 6</td>
</tr>
<tr>
<td>ZmNIP2;1a</td>
<td>NP_001131324</td>
<td>4</td>
<td>303</td>
<td>6</td>
<td>6.39</td>
<td>32.33</td>
<td>182093856</td>
<td>182096521</td>
<td>C.M</td>
<td>plas: 5</td>
</tr>
<tr>
<td>ZmNIP2;2</td>
<td>NP_001105020</td>
<td>6</td>
<td>294</td>
<td>6</td>
<td>7.71</td>
<td>31.39</td>
<td>113136429</td>
<td>113140203</td>
<td>C.M</td>
<td>plas: 6</td>
</tr>
<tr>
<td>ZmNIP2;3</td>
<td>NP_001105517</td>
<td>9</td>
<td>301</td>
<td>6</td>
<td>7.04</td>
<td>31.65</td>
<td>4504397</td>
<td>4507849</td>
<td>C.M</td>
<td>E.R.: 5</td>
</tr>
<tr>
<td>ZmNIP3;1</td>
<td>XP_008660914</td>
<td>9</td>
<td>284</td>
<td>6</td>
<td>6.07</td>
<td>30.14</td>
<td>77428241</td>
<td>77429525</td>
<td>C.M</td>
<td>cyto: 5</td>
</tr>
<tr>
<td>ZmNIP7;1</td>
<td>NP_001150784</td>
<td>4</td>
<td>288</td>
<td>6</td>
<td>9.44</td>
<td>29.26</td>
<td>34365878</td>
<td>34367308</td>
<td>C.M</td>
<td>vacu: 9</td>
</tr>
<tr>
<td>ZmPIP1;1</td>
<td>NP_001105466</td>
<td>2</td>
<td>288</td>
<td>6</td>
<td>9</td>
<td>30.67</td>
<td>18757697</td>
<td>18759478</td>
<td>C.M</td>
<td>plas: 7</td>
</tr>
<tr>
<td>ZmPIP1;2</td>
<td>NP_001104934</td>
<td>5</td>
<td>289</td>
<td>6</td>
<td>9</td>
<td>30.79</td>
<td>193359281</td>
<td>193362537</td>
<td>C.M</td>
<td>plas: 7</td>
</tr>
<tr>
<td>ZmPIP1;3</td>
<td>NP_001105022</td>
<td>4</td>
<td>292</td>
<td>6</td>
<td>8.83</td>
<td>30.99</td>
<td>153654785</td>
<td>153658151</td>
<td>C.M</td>
<td>plas: 7</td>
</tr>
<tr>
<td>ZmPIP1;5</td>
<td>NP_001105131</td>
<td>4</td>
<td>288</td>
<td>6</td>
<td>8.3</td>
<td>30.72</td>
<td>170004670</td>
<td>170006094</td>
<td>C.M</td>
<td>plas: 8</td>
</tr>
<tr>
<td>ZmPIP1;6</td>
<td>NP_001105023</td>
<td>9</td>
<td>296</td>
<td>6</td>
<td>6.7</td>
<td>31.02</td>
<td>6467916</td>
<td>6469385</td>
<td>C.M</td>
<td>cyto: 4</td>
</tr>
<tr>
<td>ZmPIP2;1</td>
<td>NP_001105024</td>
<td>7</td>
<td>290</td>
<td>6</td>
<td>7.69</td>
<td>30.21</td>
<td>41435404</td>
<td>41438423</td>
<td>C.M</td>
<td>plas: 8</td>
</tr>
<tr>
<td>ZmPIP2;2</td>
<td>NP_001105638</td>
<td>2</td>
<td>292</td>
<td>6</td>
<td>8.29</td>
<td>30.25</td>
<td>169186930</td>
<td>169190239</td>
<td>C.M</td>
<td>plas: 6</td>
</tr>
<tr>
<td>ZmPIP2;3</td>
<td>NP_001105025</td>
<td>4</td>
<td>289</td>
<td>6</td>
<td>6.95</td>
<td>30.42</td>
<td>143891383</td>
<td>143893926</td>
<td>C.M</td>
<td>plas: 1</td>
</tr>
</tbody>
</table>

**Table 1:** Genome-wide studies of aquaporins. Literature survey on genome wide studies of aquaporins in plant species.

**ZmPIP2;4**  NP_001105026  5  288  6  6.5  30.32  195239919  195242604  C.M  plas:11

**ZmPIP2;5**  NP_001105616  2  285  6  7.7  29.83  28493234  28495453  C.M  plas: 9

**ZmPIP2;6**  NP_001105027  7  288  7  8.38  30.19  41553964  41555424  C.M  plas: 7

**ZmPIP2;7**  XP_008670063  2  286  6  8.38  29.82  168928036  168929344  C.M  plas: 8

**ZmTIP1;1**  NP_001105029  8  254  7  5.87  25.35  156113368  156114463  vacu  vacu: 6

**ZmTIP1;2**  NP_001104896  1  250  6  6.02  25.82  10068811  10070364  vacu  cyto: 4

**ZmTIP2;1**  NP_001104907  10  248  6  6.16  25.13  139149979  139151283  vacu  vacu: 7

**ZmTIP2;2**  XP_008669054  2  248  6  5.87  24.98  20452237  20453860  vacu  plas: 8

**ZmTIP2;3**  NP_001105030  4  249  6  5.3  24.86  152458474  152459746  vacu  cyto: 5

**ZmTIP2;4**  NP_001105031  5  250  6  5.59  25.04  191963086  191964272  vacu  cyto: 5

**ZmTIP3;1**  NP_001105032  5  262  6  8.12  27.2  25531209  25532312  Vacu  mito: 9

**ZmTIP3;2**  NP_001105045  1  266  6  8.37  27.39  228464363  228465624  Vacu  mito: 9

**ZmTIP3;3**  NP_001146930  9  267  5  9.79  27.28  154814247  154815523  vacu  cyto: 5

**ZmTIP4;1**  XP_008648279  6  274  6  8.86  28.8  133537380  133539724  vacu  chlo:11

**ZmTIP4;1a**  NP_001105036  10  260  6  7.74  26.49  139151482  139151482  vacu  chlo:12

**ZmSIP1;1**  NP_001105028  8  243  6  9.35  25.7  108815373  108819980  C.M, vacu  cyto: 5

**ZmSIP1;2**  AF326499  1  249  6  9.86  26.6  52645832  52648286  C.M  chlo, nucl

**Table 2:** Bioinformatics analysis of ZmAQPs. Bioinformatics predicted features of ZmAQPs, including gene name, accession, chromosome number, number of amino acids, trans-membrane domain, iso-electric point molecular weight, start and end base pair and sub-cellular location.

**Phylogenetic analysis**

Phylogenetic analysis revealed 41 ZmAQPs, which were divided into 4 subfamilies: ZmTIPs, ZmNIPs, ZmPIPs and ZmSIPs. ZmTIPs contained 18 ZmAQPs genes forming five subgroups: ZmTIP1 (two genes), ZmTIP2 (four genes), ZmTIP3 (four genes), ZmTIP4 (seven genes), and ZmTIP5 (one gene). Eight genes ZmNIPs were found and these genes were divided into four subgroups i.e. ZmNIP1 (2 genes), ZmNIP2 (4 genes), ZmNIP3 (1 genes) and ZmNIP7 (1 gene). Twelve ZmPIPs were found and were divided into two subgroups, ZmPIP1 (5 genes) and ZmPIP2 (7 genes). Similarly, ZmSIPs possessed only 3 genes: ZmSIP1(two genes) and ZmSIP2 (one gene). All ZmAQPs distributed into 14 sisters branches and 13 single genes (Figure 1). In order to understand the evolutionary relationship of AQPs orthologs between monocots and dicots plants, a phylogenetic analysis was performed among maize (41 AQPs genes), Arabidopsis (35 AQPs genes) and chickpea (40 AQPs genes) using their amino acid sequences. Result depicted that all AQPs were distributed into 38 sister groups including 13 sisters of ZmAQPs-ZmAQPs, 11 of AtAQPs-AtAQPs, 9 of CaAQPs-CaAQPs, 5 of CaAQPs-AtAQPs and 36 single genes (Figure 2). Interestingly there in no any sister branch of either ZmAQPs-AtAQP or ZmAQPs-CaAQPs. Moreover most of the maize aquaporin gene formed separated branches in comparison to chickpea and Arabidopsis. For example, in NIP subfamily all ZmNIPs formed two distinct branches. Similarly, in PIPs, ZmPIP1 and ZmPIP2 showed separate branches except ZmPIP1;6. Within TIPs, ZmTIP2 formed three separated branches e.g all ZmTIP3, ZmTIP2 and ZmTIP4.
Figure 1: Phylogenetic tree of *Zea mays* Aquaporins. Phylogenetic tree was constructed with bootstrap 1000 replicates, using maximum likelihood algorithms through PhyML 3, and visualized by MEGA6 software. Numeric values on node represent evolutionary divergence. The outermost blue circular lines represent subfamilies like ZmTIP, ZmNIP, ZmPIP and ZmSIP while the inner circular lines show groups within subfamilies.
Figure 2: Cumulative phylogenetic tree of plant species. An unrooted phylogenetic tree of Arabidopsis, chickpea and maize’s AQPs was constructed using maximum likelihood algorithm through PhyML at with 1000 bootstrap value. Numeric values on node represent evolutionary divergence. Accessions of species indicated with color filled symbol. At: Arabidopsis thaliana; Car: Cicer aritinum and Zm: Zea mays.

Trans-membrane domains (TMDs) and sub-cellular localization of ZmAQPs

The Trans-membrane domain (TMDs) prediction demonstrated that majority of ZmAQPs consist of six TMDs (34 of 41, 83%), two ZmAQPs (ZmPIP2;6 and ZmTIP1;1) showed seven TMDs, three ZmAQPs (ZmTIP3;3, ZmNIP1;2 and ZmSIP1;1) carried five TMDs and only one ZmAQP (ZmTIP4;2) has two TMDs. The sub-cellular localization prediction revealed that 21 aquaporins were localized in cell membrane, whereas, 18 of them demonstrated multiple locations including vacuole membrane, plasma membrane, mitochondrial membrane, rough endoplasmic reticulum and chloroplast membrane (Table 2, Figure S1).

Sequence alignment and identification of NPA motifs and ar/R selectivity filter residues and Froger’s positions

Multiple sequence alignment of ZmAQPs exhibited that most of the amino acid residues within domains were conserved among all 41 ZmAQPs. In NPA motifs, ar/R selectivity filter and Froger’s positions
some residues were family specific. A complete full length multiple sequence alignment of all ZmAQPs amino acid sequences has been shown in (Figure S2), indicating NPA motifs, ar/R filter, foreigner position and trans-membrane domain. The 1st NPA motif pattern (Asp-Pro-Ala) was conserved in all subfamilies except ZmSIPs which showed replacement of alanine (A) with threonine (T), while ZmTIP4p2 depicted deletion of aforementioned motif. Similarly, the 2nd NPA motif was also conserved in all 41 ZmAPQs excluding ZmSIP4p2 i.e. A(alanine) replaced with T(tyrosine) (Table 3). In case of ar/R selectivity filter regions amino acid residues were conserved intra-subfamilies but it varied inter-subfamilies like in H2 helix, histidine (H) remained intact in all of the ZmTIPs except ZmTIP4p3 and ZmTIP5p1(histidine was replaced with glutamine). In ZmPIP subfamily, histidine (H) was substituted with phenylalanine (F), in case of ZmNIPs; histidine was changed with either glycine (G) or tryptophan (W). Similarly, ZmSIPs possessed leucine (L) in place of histidine. The H5 helix exhibited bit different amino acids placements e.g. ZmTIP subfamily has isoleucine/valine/serine residue, in ZmNIPs; alanine/serine/valine, and in ZmSIPs; valine/isoleucine. However, in all ZmP1Ps; histidine(H) was conserved in H5 helix. In case of loops, (G) in all ZmAQPs, while in LE2 most of the ZmAQPs found arginine (R) except ZmSIPs (R replace with N) and two gene of ZmTIPs i.e

<table>
<thead>
<tr>
<th>Gene name</th>
<th>NPA motifs</th>
<th>ArR selectivity filter</th>
<th>Froger’s positions</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZmTIP2p2</td>
<td>NPA</td>
<td>H I G R T S A Y W</td>
<td></td>
</tr>
<tr>
<td>ZmTIP2p1</td>
<td>NPA</td>
<td>H I G R T S A Y W</td>
<td></td>
</tr>
<tr>
<td>ZmTIP2p4</td>
<td>NPA</td>
<td>H I G R T S A Y W</td>
<td></td>
</tr>
<tr>
<td>ZmTIP2p3</td>
<td>NPA</td>
<td>H I G R T S A Y W</td>
<td></td>
</tr>
<tr>
<td>ZmTIP1p1</td>
<td>NPA</td>
<td>H I G V T S A Y W</td>
<td></td>
</tr>
<tr>
<td>ZmTIP1p2</td>
<td>NPA</td>
<td>H I G V T S A Y W</td>
<td></td>
</tr>
<tr>
<td>ZmTIP3p1</td>
<td>NPA</td>
<td>H V G R T V A Y W</td>
<td></td>
</tr>
<tr>
<td>ZmTIP3p2</td>
<td>NPA</td>
<td>H V G R T V A Y W</td>
<td></td>
</tr>
<tr>
<td>ZmTIP3p3</td>
<td>NPA</td>
<td>H V G R T A A Y W</td>
<td></td>
</tr>
<tr>
<td>ZmTIP3p4</td>
<td>NPA</td>
<td>H I G R S A A Y W</td>
<td></td>
</tr>
<tr>
<td>ZmTIP4p2</td>
<td>NPA</td>
<td>H S G R S A A Y W</td>
<td></td>
</tr>
<tr>
<td>ZmTIP4p2</td>
<td>NPT</td>
<td>- S G R - S A Y W</td>
<td></td>
</tr>
<tr>
<td>ZmTIP4p4</td>
<td>NPA</td>
<td>H S G R S A Y W</td>
<td></td>
</tr>
<tr>
<td>ZmTIP4p1</td>
<td>NPA</td>
<td>H S G R S A Y W</td>
<td></td>
</tr>
<tr>
<td>ZmTIP4p1</td>
<td>NPA</td>
<td>H S G R S A Y W</td>
<td></td>
</tr>
<tr>
<td>ZmTIP4p4</td>
<td>NPA</td>
<td>H V G R A S A Y W</td>
<td></td>
</tr>
<tr>
<td>ZmTIP4p3</td>
<td>NPA</td>
<td>Q S G R T S A Y W</td>
<td></td>
</tr>
<tr>
<td>ZmTIP5p1</td>
<td>NPA</td>
<td>Q V G R : S A Y W</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: ar/R selectivity regions. Specificity determining conserved residues including NPA, ar/R filter region and P1-P5 in Zea mays aquaporins.

**Conserved motif within ZmAQPs**

Besides NPA motifs and hexa-helical domain, there were other motifs too which might play important role in regulation of ZmAQPs. MEME motif discovery resulted that most of the motifs remained conserved in all subfamilies, but some showed different pattern i.e. few were deleted out and few were unique and family specific. Motifs 1-5 (1: NPA, 2: NPA, 3: Phosphoserine, 4: Amidation-Site, 5: Casein kinase II) were conserved among all four subfamilies except ZmSIPs which had only 2nd (NPA) and 4th motif. Some motifs were family specific like motif 6th, 7th, and 8th (6: Novel motif, 7: N-myristoylation site, 8: Novel , 9: N-myristoylation site) were present only in ZmPIPs while its subgroups ZmPIP1 and ZmPIP2 exhibited some differences, ZmPIP1 has 11th (PK_Phospho site) motif, which was deleted in ZmPIP2 protein. Motif 10th (: N-myristoylation site) and motif 12th (N-myristoylation site) were present only in ZmTIPs and ZmSIPs, while motif 13th (Novel motif ) was present only in ZmNIPs. Motif 14th (Phosphothreonine kinase) was only present in ZmTIPs subfamily members except the ZmTIPS3, ZmTIPS4,
ZmTIP4;4, ZmTIP4;3 and ZmTIP5;1. Motif 15th (N-myristoylation site) was present in ZmNIPs and ZmPIPs subfamily members. However, few genes were found as outlier to aforementioned pattern (Figure 3, supplementary file S1).


Gene ontology (GO)
Gene Ontology analysis revealed a critical role of ZmAQPs in distinct biological, cellular and molecular processes. The biological process involved transportation of ions, glycerol and other small solute (Figure 4a, Figure S3). The molecular function revealed that ZmAQPs have crucial role in substrate-specific transmembrane transporter activity, active and passive transmembrane transporter activity, carbohydrate and organic hydroxyl compound transporter activity, receptor activity, metal binding activity, heterocyclic compound binding activity and structural molecular activity (Figure 4a, Figure S4). In summary, significant activity of ZmAQPs was observed in transportation of compounds such as glycerol, ion, water and alcohol across plasma membrane. These activities may be due to substrate-specific channel forming activity (Figure 4a). Sub-cellular localization exhibited that all ZmAQPs were incorporated into membrane of various cellular components like plasma membrane, vacuolar membrane, membrane bounded organelles and other intrinsic component of membrane. ZmAQPs were also integrated into cell periphery, intracellular organelle, plasmodesmeta and cell-cell junction (Figure 4b, Figure S5).
**Figure 4a:** Gene Ontology. Gene ontology of ZmAQPs proteins Biological process of 41ZmAQPs.

**Figure 4b:** Gene Ontology. Gene ontology of ZmAQPs proteins Cellular components of 41ZmAQPs.
Gene structure analysis of ZmAQPs gene family / Genomic organization of ZmMIPs

The comprehensive transcriptomic data of Z. mays made it possible to analyze the gene structural components of aquaporins within Z. mays genome. The gene displays server (GDS) resulted that all four families showed differences in number of exon and introns. Among ZmAQPs families, ZmNIPs has highest number of exons (five exons) followed by ZmPIPs (four exons), ZmTIPs (three exons) and ZmSIPs (three exons). For more depth study, in ZmTIPs; nine genes contain two exons and remaining eight genes carried three exons while ZmTIP4;2 showed no intron. In ZmNIPs all genes carried five exons except ZmNIP1;1 and ZmNIP1;2, which showed loss of 2nd exon. In ZmPIPs; five genes had four exons, three genes (ZmPIP2;5, ZmPIP2;6 and ZmPIP2;7) has three exons and only two genes (ZmPIP1;5 and ZmPIP1;5) has two exons. However, all ZmSIPs genes carried three exons (Figure 5).

Figure 5: Gene structure analysis. Exon-intron structures of Zea mays AQP genes with phylogenetic tree. The graphic representation of the gene models is displayed using GSDS. Phylogenetic tree with sub families shown on left side of gene model graph. Blue boxes indicate upstream/downstream region, yellow color boxes for exons and straight lines represent introns part of gene.

Chromosomal location of ZmAQP gene

The chromosomal locations of 41 ZmAQPs were graphically presented in Figure 6. The ZmAQPs distribution in genome showed that 4th and 5th chromosome has highest number of ZmAQPs genes (7 genes), followed by chr2 (6 genes), chr6 (4 genes) and chr9 (4 genes). Chromosome 1, 7 and 8 carried three ZmAQPs genes, while chromosome 3 and 10 has only two. Furthermore, chromosomal location identification disclosed that most of the ZmAQPs genes were
found in cluster form and located either on top or on bottom of the chromosome. Interestingly, most of the genes belonging to same family were located on same chromosome e.g. ZmPIP2;2, ZmPIP2;7 and ZmPIP2;5 formed a cluster, within a ~500kbp segment on chromosome 2 while ZmPIP1;5 and ZmPIP1;3 were clustered on chromosome 4. Similarly, ZmNIP1;2 and ZmNIP2;2 were clustered on chromosome 6 (Figure 6).

Figure 6: Chromosomal distribution. Chromosomal location of Zea mays AQPs. The chromosome numbers are shown on the right side of each chromosome and correspond to the approximate location of each AQP gene.

**Discussion**

Robust advancements in the computational biology helped to sequence large number of plant genomes, which certainly improved the identification and characterization of physiologically vital gene families in plants. The computational analysis of aquaporin proteins will definitely help in hypothesis generation and subsequent experimental validations and ultimately lead towards genetically engineered improved crop. Researchers have always been concerned to study different genes and their expression in order to understand their role in the context of a particular plant. Here in this study, AQPs family was selected due to its significant input to growth and development of a plant [47]. The prime function performed by AQPs is the regulation of water and some solutes across cell membrane [47,48]. Considering AQPs importance in plant growth and development, its genome wide analysis has been conducted in many plant species including monocots, such as rice, wheat, barley etc, as well as in dicots i.e. potato, tomato, cabbage, carrot, celery and Arabidopsis (see Table 1). Current study on ZmAQPs has its own worth because to the best of our knowledge few studies conducted on genome wide identification, characterization and functional prediction of aquaporins in maize [49]. However, Chaumont [3] carried a study on ZmAQPs based on EST/cDNA sequence and identified 31 AQPs, while current study is based on whole genome identification of aquaporins gene and proposed 41 AQPs gene. The difference in number of AQPs between the two studies might be due to varying number of expressed transcripts at a given time. The limitation in Chaumont study [3] was addressed in this study by exploiting the whole genome sequence of maize. In contrast to Chaumont study, we also performed a genome-wide comparative study of ZmAQPs with Arabidopsis and chickpea that gave the evolutionary insight among the monocots and dicots. Chaumont Study was restricted to expressed sequence tags and its phylogenetic tree, while in this study, we provided many other biological parameters like ZmAQPs protein sequences analysis, conserved motif prediction, gene ontology, gene structure analysis and chromosomal and sub-cellular localization prediction in details. The genome wide study of aquaporins in maize suggested that the total number of AQP genes in maize were higher than in Arabidopsis (35 AQPs), sweet orange (34 AQPs), rice (34 AQPs) and physic nut (32 AQPs), but lesser than other few species like banana (47 AQPs), tomato (47 AQPs), cottonwood (55 AQPs), soybean (66 AQPs), and chinese cabbage (53 AQPs) (see Table 1 for the comparison chart). These changes in number of gene in
different species may be due to size of their genome or due to evolutionary process for adaptation in natural environment [50].

The phylogenetic tree demonstrated that all ZmAQPs were classified into four sub-families viz ZmTIPs, ZmPIPs, ZmNIPs and ZmSIPs (Figure 1), which is in agreement with the Chaumont [3] and other studies listed in Table 1. Additionally, we also indicated groups and sub-groups of subfamilies; for instance, ZmTIPs was divided into five subgroups (TIP1, TIP2, TIP3, TIP4 and TIP5). Such findings were similar to other monocots AQPs i.e river, banana, sorghum, and barley (Table 1). As TIPs proteins involve in the transport of various small solutes like NH₄⁺, H₂O₂, and urea [51-53], so the information about TIPs in maize genome may lead to improvement of transportation mechanism. The ZmPIPs were divided into two subgroups; ZmPIP1 and ZmPIP2 which are seeming to be conserved in all other plant species (listed in Table 1). Experimental analysis of PIPs conformed its role in water absorption inside roots as well as turgor pressure in leaf [54-56]. Furthermore, PIPs also facilitate CO₂ diffusion in mesophyll that enhances photosynthesis process [57,58]. So the study of PIPs individual genes in maize will help in understanding of C4 photosynthesis mechanism that are crucial in engineering of C4 crops. Similarly ZmNIPs carried H, I, G, R or H, I, A, V residues in the ar/R selectivity filter region, and T, S, A, Y, W or T/S, A, Y, W amino acid residues at P1-P5 positions are reported to transport urea [63,64]. Chaumont study [3] only presented the NPA motifs, while current study included the ar/R selectivity filter and Frogger’s positions that are most important for selection of molecule across biological membrane [66,67]. All ZmAQPs showed two representative NPA motifs as reported by Chaumont [3]. The ZmPIPs contain highly conserved ar/R selectivity filter region (F-H-G-R) and same pattern was also identified in other PIPs, like in Arabidopsis, tomato, wheat, barley and poplar (Table 1).

Additionally, the presence of the S, A, F, W residues at P2-P5 positions in PIPs, has been reported as a signature of CO₂ transporter [68]. Any mutation in these conserved amino acids in ZmPIPs can alter the capacity of the protein for CO₂ diffusion. If the mutation leads increment in CO₂ diffusion, then it will also helpful for establishment of C4 crops. Similarly ZmTIPs carried H, L, G, R or H, I, A, V residues in the ar/R selectivity filter region, and T, S, A, Y, W or T/S, A, Y, W amino acid residues at P1-P5 positions are reported to transport urea and H₂O₂ across membrane [68]. The ZmAQPs ar/R filters in maize were identical to sweet orange and soybean where these genes act as water facilitator [69]. The ZmNIPs sub-group showed G, S, G and R, at ar/R filter that are involved in transportation of water, silicon (Si) and boron (B) as identified in Arabidopsis (AtNIP5 with A, I, G and R in ar/R filter region involved in the transport of arsenic and boron but not silicon) and sweet orange [70]. The SIPs protein’s basic structure and function are still under characterization but in Arabidopsis, AtSIPs are involved in functional water channel [71]. The differences in ar/R selectivity filter residues might alter the selectivity behavior for substrate transport. Thus, point mutations in these patterns either increase or decrease the transportation capacity of aquaporins proteins [70,27,23]. The variation in these conserved segments described the evolutionary diversity among ZmAQPs. Besides NPA and ar/R filter regions, there were some other important motifs, predicted through MEME discovery server (Figure 3) including phosphoserine, amidation site, casein kinase II, N-myristoylation site, PK_Phospho, phosphothreonine and some novel motifs. Such motifs have also essential role in organization and regulation of MIP domain and sometime these motifs can be targeted for regulation and interaction of AQPs protein with other cellular components. The prediction on subcellular localization as plasma membrane, vacuole membrane and mitochondrial membrane for ZmAQPs that were corresponds to the reports in other plant species including soybean, sorghum, rubber tree, sweet orange, common bean, soybean and moso bamboo (see Table 1). The ZmTIPs mainly express in vacuole that may help to control osmotic potential and ZmPIPs integrated within plasma membrane, while ZmNIPs are expressed in various membrane such finding were similar to literature [74,75].

We have presented schematic representation of gene structures of all ZmAQPs with their evolutionary relationship (Figure 5) that were not demonstrated in Chaumont [3]. Our results were in agreement to previous studies in other plant species [17,61,65]. Most of ZmAQPs showed similar gene structure to their orthologs in other plants like in TIPs; ZmTIP1, SbTIP1(sorghum) [60], MaTIP1(banana) [65], PtTIP1 (moss bamboo) [61] has two exons; ZmTIP2 has also covenant with its orthologs in sorghum, banana and moso bamboo except ZmTIP2;1 which demonstrated an additional exon as compare to aforementioned species[17,61,65]. The ZmTIP3 showed similarity to sorghum, while somehow diverges to moso bamboo and banana. Similarly, other subfamilies also demonstrated similar patterns. Thus, most of our findings were similar to previous studies (Table 1). So, we suggest that the loss or gain of exons in ZmAQP genes might have ensued under natural
selection [28,76]. Previous studies have depicted that lost or gain of exons are the common feature of evolutionary process in plants genomes [76,77].

The chromosomal location of genes tells about the expression capacity of a gene [78]. Chaumont study [3] on ZmAQPs did not give any clue about gene location within genome of maize. However, we demonstrated chromosomal locations of ZmAQPs on whole genome sequence. All ZmAQPs genes were distributed on all 10 chromosomes forming groups. The genes belonging to same sub-family mostly form clusters within the window size of 500kb. The clustering of ZmAQPs depicted segmental duplication that is one of a key mechanisms for gene expansion that increase genetic diversity [79]. Moreover, this mechanism may also be responsible for the functional divergence by increasing total members of a given gene subfamily [80].

Conclusion

In this study, we used different bioinformatics tools for the genome-wide identification, analysis and characterization of aquaporins in maize genome. Moreover, we have predicted several physiological structures with their biological functions. The aquaporin genes family widely studied in many important crops demonstrating its structural and functional diversity. The availability of genome sequences made this study possible. Current study identified 41 aquaporins in *Zea mays* L. and these genes assigned nomenclature as well as classified into four subfamilies. Moreover, the structural and functional features of ZmAQPs have been predicted, and a comparative phylogenetic study of ZmAQPs, CaAQPs, and AtAQPs was also conducted, which provide insights about the evolution of AQPs within plant species. The results achieved in this study not only provide valuable information for future functional analysis of ZmAQPs genes but also make a suitable reference to survey the gene family expansion in *Zea mays* and other crops from evolutionary perspective.

Competing Interests

The Authors declare no competing interest.

Acknowledgements

We are thankful to our colleagues and teachers at National Institute for Biotechnology and Genetic Engineering (NIBGE), Faisalabad, Pakistan.

Ethical Approval and Consent to Participate

Not applicable.

References


