Genome-wide identification, characterization, and expression profile of aquaporin gene family in flax (Linum usitatissimum)

S. M. Shivraj1,*, Rupesh K. Deshmukh2,*, Rhiitu Rai3, Richard Bélanger4, Pawan K. Agrawal3 & Prasanta K. Dash1

Membrane intrinsic proteins (MIPs) form transmembrane channels and facilitate transport of myriad substrates across the cell membrane in many organisms. Majority of plant MIPs have water transporting ability and are commonly referred as aquaporins (AQPs). In the present study, we identified aquaporin coding genes in flax by genome-wide analysis, their structure, function and expression pattern by pan-genome exploration. Cross-genera phylogenetic analysis with known aquaporins from rice, arabidopsis, and poplar showed five subgroups of flax aquaporins representing 16 plasma membrane intrinsic proteins (PIPs), 17 tonoplast intrinsic proteins (TIPs), 13 NOD26-like intrinsic proteins (NIPs), 2 small basic intrinsic proteins (SIPs), and 3 uncharacterized intrinsic proteins (XIPs). Amongst aquaporins, PIPs contained hydrophilic aromatic arginine (ar/R) selective filter but TIP, NIP, SIP and XIP subfamilies mostly contained hydrophobic ar/R selective filter. Analysis of RNA-seq and microarray data revealed high expression of PIPs in multiple tissues, low expression of NIPs, and seed specific expression of TIP3 in flax. Exploration of aquaporin homologs in three closely related Linum species bienne, grandiflorum and leonii revealed presence of 49, 39 and 19 AQPs, respectively. The genome-wide identification of aquaporins, first in flax, provides insight to elucidate their physiological and developmental roles in flax.

Water absorption from soil through roots and its translocation to different parts is of paramount importance for innate physiological processes in plants. Within the plant system water movement takes place through apoplastic, symplastic and transcellular pathways1. Amongst the three defined pathways, transcellular movement of water in plants is facilitated by small integral membrane proteins (21–34 kD) called aquaporins (AQPs)2,3. These aquaporins belong to major intrinsic protein (MIP) superfamily with members spanning in animals4,5, plants as well as in microbes6. However, plants, unlike animals, harbor highly abundant and divergent AQPs7. Based on phylogenetic analysis, plant AQPs are grouped into five subfamilies: (i)plasma membrane intrinsic proteins (PIPs), (ii) tonoplast intrinsic proteins (TIPs), (iii) nodulin26-like intrinsic proteins (NIPs), (iv) small basic intrinsic proteins (SIPs), and (v) uncharacterized X intrinsic proteins (XIPs)8–11. All AQP subfamilies are identified in different plant species including primitive land plant physcomitrella (Physcomitrella patens) except the XIPs that are absent in Brassicaceae and monocots12–15. In addition, two unique classes of AQPs, GlpF-like intrinsic protein (GIPs) and hybrid intrinsic proteins (HIPs) have been identified in physcomitrella13.

Primarily, plant AQPs are involved in water transport, though some of them are implicated in transport of small solutes such as urea, ammonia, glycerol, silicic acid, boric acid, CO2 and H2O23,16–18. AQPs from different plant species have also been reported to be involved in abiotic stress tolerance. Additionally, the role of AQPs in key developmental processes such as seed germination, stomatal movement, cell elongation, and reproductive growth including male fertility has also been established19–23. Understanding the multiple roles of AQPs have been facilitated by studying the interaction of AQPs with cognate solvents at molecular level. High resolution, three-dimensional structures of AQPs from different organisms including plants revealed their hourglass structure24. The structure of the protein is formed by six transmembrane (TM) α helices (helix 1 to helix 6), and five loops (loop A to E) that penetrates into the lipid bi-layer to make route for water transport. While, loops B and

1ICAR-NRC on Plant Biotechnology, PUSA, New Delhi, India. 2Département de Phytophysique, University Laval, Quebec, QC, Canada. 3National Agricultural Science Fund (NASF), PUSA, ICAR, New Delhi, India. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to P.K.D. (email: pdas@ncrbp.org)
E contain highly conserved NPA (Asn-Pro-Ala) motifs in half-helices to form one of the two constrictions in the channel; the aromatic arginine (ar/R) region formed by each residue from helix 2 and helix 5, and two residues from loop E (LE1 and LE2) represent the other constriction. These two constrictions mostly determine the substrate permeability. The NPA motifs are known to have role in plasma-membrane targeting for the AQPs and also shown to be involved in substrate specificity. Recent study in plants highlighted the importance of the precise spacing between two NPA motifs in determining substrate specificity. Change in amino acid residue at ar/R selectivity filters (SF), NPA motifs, and the spacing between NPA motifs are found to affect substrate specificity of AQPs in plants. Similarly, few conserved amino acids known as Froger’s residues are found to be associated with substrate specificity particularly in microbial AQPs. Experimentally validated structural and functional information of AQPs has led to identification of their orthologs in different plant species such as arabidopsis (A. thaliana), rice (O. sativa), poplar (P. trichocarpa), soybean (G. max) and tomato (L. esculentum).

Flax (Linum usitatissimum) is an important dual purpose industrial crop cultivated for high quality fibre (linen) and seed oil. The flax stem is the main source of cellulose-rich bast fibre, used by textile industries for production of linen clothes. Its seed (linseed) oil is beneficial for human health owing to the presence of high amount of omega-3 fatty acids. Also, the oil of linseed is used in preparation of many industrial solvents. Considering the economic importance of flax, its genome was sequenced and subsequently there is an accumulation of genomic information in flax. Water transporting aquaporin family members are implicated to play an important role in seed development and fiber formation in different crop plants. In this study, genome sequence of flax was analyzed to identify the aquaporin encoding genes. This is the first report that identified 51 AQPs in flax through genome-wide analysis. Hitherto, flax aquaporins were classified based on the phylogenetic analysis with known AQPs, which showed five groups representing different sub-families of AQPs. Their distribution patterns, genetic architecture, structural properties and expression pattern were investigated to identify candidate AQPs with pivotal role in physiology and development in flax.

Results

Genome-wide distribution of Flax AQPs. Tabulation of blast output, BLASTp, employing AQP genes from ten different plant species as query (with high scoring pairs >100bit score) led to identification of 51 putative AQP genes (see Supplementary Table S1) in flax. While, domains of these putative proteins were analysed using conserved domain database (CDD) tool hosted in NCBI (see Supplementary Table S2), functional annotations of these sequences using protein database (PANTHER) confirmed function of these candidate sequences as aquaporins (see Supplementary Table S3). Protein domain analysis revealed the presence of six signature transmembrane domains (see Supplementary Table S4) in 35 out of the 51 identified AQPs. These AQPs were found to be distributed among 43 scaffolds (Table 1). Out of the 43 scaffolds, 35 contained only one AQP while eight scaffolds (scaffold 28, scaffold 59, scaffold 76, scaffold 123, scaffold 156, scaffold 280, scaffold 605 and scaffold 612) were found to harbor two AQPs each (Table 1).

Phylogenetic distribution of AQPs in flax. Phylogenetic tree of flax AQPs along with the known AQPs from A. thaliana, O. sativa and P. trichocarpa showed five distinct clusters representing different class of AQPs (Fig. 1). The flax AQPs were named according to their grouping with known AQPs and were classified into 16 LuPIPs, 17 LuTIPs, 13 LuNIPs, 2 LuSIPs and 3 LuXIPs. Within the groups formed by flax AQPs, two major subgroups were found in LuPIPs (LuPIP1, LuPIP2). Among them, LuPIP1 represented five members while LuPIP2 comprised of 11 members. Similarly, the LuNIPs formed two subgroups of LuNIP1 and LuNIP3 comprising of six and seven members respectively. Surprisingly, the NIP2 subfamily was found to be missing in flax genome. LuTIPs formed five subgroups (Fig.1) with LuTIP1 having seven members, LuTIP2 and LuTIP3 having four members each and, LuTIP4 and LuTIP5 having one member each. LuXIPs formed two subgroups, LuXIP1 comprised of two members and LuXIP2 with one member. LuSIPs did not form any subgroup except two isoforms SIP1-1 and SIP1-2. In our study, although homologs of all TIP, PIP and XIP family members have been identified, the homologs of NIP2 and SIP2 were not found.

Silicon accumulation in flax plants. The NIP2s having signature sequence of Gly-Ser-Gly-Arg (GSGR) at ar/R selectivity filter are well characterized as a silicon transporters. On the basis of presence or absence of NIP2s a previous study had characterized 25 plant species as poor or rich silicon accumulators. Results of phylogenetic analysis and subsequent identification of conserved motifs and ar/R selectivity filter revealed loss of Si transporting signature (GSGR) containing NIP2 members in flax. We validated the effect of loss of NIP2s in flax by measuring the silicon accumulation capability in flax shoot (Fig. 2). Quantification of silicon in one-month old flax plant revealed 0.24% silicon on dry weight basis in the leaf tissues. Low accumulation of Si in flax corroborated during the silicon accumulation capability in flax shoot (Fig. 2). Further, comparison with previously reported data grouped flax into poor silicon accumulator category along with Arabidopsis thaliana, Brassica rapa and Solanum tuberosum all of which have less than 1% silicon accumulation when grown with continuous supply of sufficient silicon (1.7 mM). All these poor silicon accumulator plants are devoid of NIP2s whereas the high silicon accumulator plants like soybean, Brachypodium and rice possess NIP2s and are found to be accumulating over 2.4% silicon on dry weight basis (Fig. 2).

Gene structure, organization and evolution of flax AQPs. Flax AQPs showed less variation in transcript length (ranging from 498 bp to 945 bp) compared to variation in gene length (ranging from 706 bp to 6670 bp). Exon intron structure analysis depicted presence of varied number of introns among the AQPs contributing to the variation in gene length (Fig. 3). Our study revealed that the number of introns in flax AQPs ranged from zero (LuSIP1-1; LuSIP1-2) to four (LuNIP1-2, LuNIP1-5, LuNIP3-1, LuNIP3-2) introns. While, both SIP
Table 1. Description and distribution of aquaporins identified from flax genome. Fifty one aquaporins belonging to five different classes such as PIPs(1–16), TIPs(17–33), NIPs(34–46), SIPs(47–49), and XIPs(49–51) along with their gene identifiers, gene length, and transcript length are identified in 43 scaffolds in flax genome.
homologs were found to be intronless; least number of introns were observed in TIPs (1-2) followed by PIPs (1-3) and NIPs (1-4). The exon-intron features observed in flax AQPs were similar to gene structure of AQPs observed in other crop plants. Of the 17 TIPs, maximum number of homologs (12) contained two introns while five homologs harbored single intron each. Similarly, of the sixteen PIPs identified in flax genome ten homologs contained three introns, while rest six PIPs contained less than three introns. Among the NIPs, maximum numbers of NIPs (9) harbored either three or four introns. Conserved intron-exon organization of AQPs in the four subfamilies suggest diversification of AQPs occurred early before the evolution of flax. The identified AQPs from flax showed amino acids ranging from 165 (LuTIP5-1) to 314 (LuXIP1-2) and the molecular weight of proteins ranged from 16.80 kD (LuTIP5-1) to 32.65 kD (LuPIP1-1).

Figure 1. Phylogenetic analysis of flax aquaporins (AQPs) with rice. *Arabidopsis* and *Populus*. 16 PIPs, 17 TIPs, 13 NIPs, 2 SIPs and 3 XIPs were identified in flax genome and all the 51 AQPs grouped into five different classes such as PIPs, TIPs, NIPs, SIPs, and XIPs. The genes from rice, *Arabidopsis*, *Solanum*, *Lotus*, and *Populus* are indicated with the prefixes Os, At, Sl, Lj, and Pt respectively.

Figure 2. Estimation of silicon content in flax. Si content in flax was compared with Si accumulation in other reported plant species such as rice (4%) and Brachypodium (3.6%) and are classified as high Si accumulators and possess NIP2 while crops like tomato, potato (0.2–0.5%) and flax (0.24%) are classified as poor accumulators and are devoid of NIP2s.
The distinct pattern of intron-exon organization structure observed amongst flax AQPs correlated well with their phylogenetic distribution (Fig. 3). Most of the phylogenetically related AQPs shared similar gene organization suggesting possible gene duplication event. Unlike previous report of presence of 2-3 introns in SIPs, both the SIP genes in flax were devoid of introns. Similarly, two groups of SIPs are reported in soybean, rice, arabidopsis and chickpea while SIP2 is missing in flax. It suggests, SIPs are evolutionary more vulnerable in flax. Additionally, distribution of selectivity filters also resembled well with the phylogenetic distribution in flax AQPs. The NIP groups posses relatively less conserved selectivity filter compared to groups in the other AQP subfamilies suggesting less selection pressure on NIPs.

Characterization of NPA motif, transmembrane domains and sub-cellular localization of flax AQPs. The Flax AQPs displayed difference in Asn-Pro-Ala (NPA) motifs and residues at ar/R selectivity filters and Froger’s positions (Table 2). Most of the AQPs contained dual NPA motifs except LuPIP2-9 and LuTIP1-6 which showed Asparagine to Glycine and Asparagine to Proline substitution respectively, in the first NPA motif. The GPA (LuPIP2-9) and PPA (LuTIP1-6) motifs observed in flax were not reported in any other plant species.
| Loci                      | NPA (LB) | NPA (LE) | ar/R selectivity filters | Froger's Residue | P1 | P2 | P3 | P4 | P5 |
|---------------------------|----------|----------|--------------------------|-----------------|----|----|----|----|----|
| **Plasma membrane intrinsic proteins (PIPs)** |          |          |                          |                 |    |    |    |    |    |
| LuPIP1-1                  | NPA      | NPA      | F H T R Q S A F W        |                 |    |    |    |    |    |
| LuPIP1-2                  | NPA      | NPA      | F H T R E S A F W        |                 |    |    |    |    |    |
| LuPIP1-3                  | NPA      | NPA      | F H T R E S A F W        |                 |    |    |    |    |    |
| LuPIP1-4                  | NPA      | NPA      | F H T R E S A F W        |                 |    |    |    |    |    |
| LuPIP1-5                  | NPA      | NPA      | F H T R E S A F W        |                 |    |    |    |    |    |
| LuPIP2-1                  | NPA      | NPA      | F H T R M S A F W        |                 |    |    |    |    |    |
| LuPIP2-2                  | NPA      | NPA      | F H T R Q S A F W        |                 |    |    |    |    |    |
| LuPIP2-3                  | NPA      | NPA      | — H T R Q S A F W        |                 |    |    |    |    |    |
| LuPIP2-4                  | — NP A   | — NP A   | — H T R Q S A F W        |                 |    |    |    |    |    |
| LuPIP2-5                  | — NP A   | — NP A   | — H T R S A F W          |                 |    |    |    |    |    |
| LuPIP2-6                  | NPA      | NPA      | F H T R Q S A F W        |                 |    |    |    |    |    |
| LuPIP2-7                  | NPA      | NPA      | F H T R Q S A F W        |                 |    |    |    |    |    |
| LuPIP2-8                  | NPA      | NPA      | — H T R Q S A F W        |                 |    |    |    |    |    |
| LuPIP2-9                  | GPA      | NPA      | F H T R Q S A F W        |                 |    |    |    |    |    |
| LuPIP2-10                 | NPA      | NPA      | F H T R Q S A F W        |                 |    |    |    |    |    |
| LuPIP2-11                 | NPA      | NPA      | F H T R Q S A F W        |                 |    |    |    |    |    |
| **Tonoplast intrinsic proteins (TIPs)** |          |          |                          |                 |    |    |    |    |    |
| LuTIP1-1                  | NPA      | NPA      | H I A V T S A Y W        |                 |    |    |    |    |    |
| LuTIP1-2                  | NPA      | NPA      | H I A V T S A Y W        |                 |    |    |    |    |    |
| LuTIP1-3                  | NPA      | NPA      | H I A V T S A Y W        |                 |    |    |    |    |    |
| LuTIP1-4                  | NPA      | NPA      | H I A V T S A Y W        |                 |    |    |    |    |    |
| LuTIP1-5                  | NPA      | NPA      | H I A V T S A Y W        |                 |    |    |    |    |    |
| LuTIP1-6                  | PPA      | NPA      | H I A V T S A Y W        |                 |    |    |    |    |    |
| LuTIP1-7                  | NPA      | NPA      | H I A L T S A Y W        |                 |    |    |    |    |    |
| LuTIP2-1                  | NPA      | NPA      | H I G R T S A Y W        |                 |    |    |    |    |    |
| LuTIP2-2                  | NPA      | NPA      | H I G R T S A Y W        |                 |    |    |    |    |    |
| LuTIP2-3                  | NPA      | NPA      | H I G R T S A Y W        |                 |    |    |    |    |    |
| LuTIP2-4                  | NPA      | NPA      | H I G R T S A Y W        |                 |    |    |    |    |    |
| LuTIP2-5                  | NPA      | NPA      | H I A R T A S Y W        |                 |    |    |    |    |    |
| LuTIP2-6                  | NPA      | NPA      | H I A R T A S Y W        |                 |    |    |    |    |    |
| LuTIP2-7                  | NPA      | NPA      | — V G C T A A Y W        |                 |    |    |    |    |    |
| **Nodulin-26 like intrinsic proteins (NIPs)** |          |          |                          |                 |    |    |    |    |    |
| LuNIP1-1                  | NPA      | NPA      | W V A R F S A Y I        |                 |    |    |    |    |    |
| LuNIP1-2                  | NPA      | NPA      | W V A R F S A Y M        |                 |    |    |    |    |    |
| LuNIP1-3                  | NPA      | NPA      | W V A R F S A Y V        |                 |    |    |    |    |    |
| LuNIP1-4                  | NPA      | NPA      | W V A R F S A Y V        |                 |    |    |    |    |    |
| LuNIP1-5                  | NPA      | NPA      | W V A R F S A Y I        |                 |    |    |    |    |    |
| LuNIP1-6                  | — NP A   | — NP A   | — V A R F S A Y I        |                 |    |    |    |    |    |
| LuNIP3-1                  | NPA      | NPA      | — V G R Y S A Y I        |                 |    |    |    |    |    |
| LuNIP3-2                  | NPA      | NPA      | A V G R Y S A Y I        |                 |    |    |    |    |    |
| LuNIP3-3                  | NPA      | NPA      | A R F A A Y I            |                 |    |    |    |    |    |
| LuNIP3-4                  | NPA      | NPA      | S I A R F A A Y M        |                 |    |    |    |    |    |
| LuNIP3-5                  | NPA      | NPA      | S I A R F T A Y L        |                 |    |    |    |    |    |
| LuNIP3-6                  | NPA      | NPA      | T I A R F T A Y L        |                 |    |    |    |    |    |
| LuNIP3-7                  | NPA      | NPA      | — I A R F T A Y I        |                 |    |    |    |    |    |
| **Small basic intrinsic proteins (SIPs)** |          |          |                          |                 |    |    |    |    |    |
| LuSIP1-1                  | NPT      | NPA      | A I P N R A A Y W        |                 |    |    |    |    |    |
| LuSIP1-2                  | NPT      | NPA      | A I P N R A A Y W        |                 |    |    |    |    |    |
| **Uncharacterized X intrinsic proteins (XIPs)** |          |          |                          |                 |    |    |    |    |    |
| LuXIP1-1                  | NP V     | NPA      | I L V R V C A F W        |                 |    |    |    |    |    |
| LuXIP1-2                  | NP I     | NPA      | V V A R V C A F W        |                 |    |    |    |    |    |
| LuXIP2-1                  | NP V     | NPA      | I V V R M C A F W        |                 |    |    |    |    |    |

Table 2. Conserved domains, selectivity filter and amino acid residues of AQPs in flax genome.
Such changes are expected to alter the substrate specificity of the aquaporins in flax. In the NIP sub-family, the first NPA domain was found to be conserved in all the members, while the second NPA motif showed Alanine to Valine substitution in five members of LuNIP sub-family (LuNIP3-3 to LuNIP3-7). In SIP and XIP sub-family, the first NPA motif showed substitution, while the second NPA motif was found conserved. Similarly, the first NPA motif of SIP sub-family harbored threonine in place of alanine; while isolateducine (LuXIP1-2) or valine (LuXIP1-1, LuXIP2-1) substituted alanine in XIP sub-family in the same motif.

All the PIP sub-family members showed a conserved ar/R filter residues with Phenylalanine in H2, Histidine at H5, Threonine at LE1 and Arginine at LE2 (see Supplementary Fig. S1). In TIP sub-family, the H2 position of ar/R filter comprised of Histidine, H5 position comprised of Isoleucine except for LuTIP5-1, which contains Valine residue (Fig. 4). LE1 and LE2 positions were found to be specific for each subgroup of LuTIPs. LuTIP1 subgroup was mostly characterized by Alanine (LE1) and Valine (LE2) except LuTIP1-7 that contained Leucine (LE2). LuTIP2 subgroup was characterized by Glycine (LE1) and Arginine (LE2). LuTIP3 and LuTIP4 sub-groups were characterized by Alanine (LE1) and Arginine (LE2). In NIP sub-family, the NIP1s were characterized by Tryptophan (H2), Valine (H5), Alanine (LE1) and Arginine (LE5) whereas the NIP3s were comprised of Alanine/Serine/Threonine (H2), Valine/Isoleucine (H5), Valine/Arginine (LE1), Arginine (LE2) (see Supplementary Fig. S2).

To ascertain expression of flax AQPs at different cellular/organellar levels, their sub-cellular localizations were predicted (Supplementary Table S5). Majority of flax PIPs were predicted to localize in the plasma membrane. Out of seventeen TIP subfamily members, only two were predicted to localize in plasma membrane, while nine TIPs were targeted to cytoplasm and five were targeted to vacuoles. Majority of NIPs were predicted to be associated with plasma membrane. While SIPs localized into the vacuoles, XIPs localized into the plasma membrane or cytoplasm.

**Identification of Aquaporins in different Linum species.** RNA-seq data for three different *Linum* species such as *L. bienne*, *L. grandiflorum* and *L. leonii* were downloaded from short read archive (SRA) at NCBI and analysed with an aim to identify orthologs of aquaporins (see Supplementary Fig. S3, Table 3). The de novo assembly of RNA reads showed N50 values of 520 bp, 797 bp and 1254 bp for *L. bienne*, *L. grandiflorum* and *L. leonii*, respectively (see Supplementary Table S6). In comparison to *L. usitatissimum*, forty-nine aquaporins were observed in *L. bienne* while thirty-nine aquaporins were found in *L. grandiflorum* and nineteen AQPs were observed in *L. leonii*. Amongst forty-nine aquaporins found in *L. bienne*, twenty-nine were PIPs, twelve were TIPs, five were NIPs, two were SIPs, and one was XIP. Expansion of PIP specific family members was observed in *L. bienne* as compared to sixteen PIPs in *L. usitatissimum*. Comparable number of PIP family members with fifteen and eleven PIPs was also observed in *L. grandiflorum*, and *L. leonii* respectively. The phylogenetic analysis of AQPs identified from different *Linum* species showed grouping in accordance to their sequence homology (see Supplementary Fig. S3B).

**AQP expression profiling in flax.** In order to provide transcriptional evidence, homology search was performed against flax specific ESTs at the NCBI database (http://blast.ncbi.nlm.nih.gov/) that revealed existence of...
| Aquaporin Family | *L. usitatissimum* | *Linum bienne* | *Linum grandiflorum* | *Linum leonii* |
|-----------------|------------------|---------------|----------------------|---------------|
| NIP             |                  |               |                      |               |
| LaNIP1-1        | —                | —             | LgNIP1-1             | —             |
| LaNIP1-2        | LbNIP1-2a        | LgNIP1-2a     | LINIP1-2             | —             |
| LbNIP1-2b       | LgNIP1-2b        | —             | —                    | —             |
| LaNIP3-1        | —                | —             | —                    | —             |
| LaNIP3-2        | LbNIP3-2a        | LgNIP3-2      | —                    | —             |
| LbNIP3-2b       | LgNIP3-2b        | —             | —                    | —             |
| LaNIP3-3        | LbNIP3-3         | —             | —                    | LINIP3-3      |
| LaNIP3-4        | —                | LgNIP3-4      | —                    | —             |
| LaNIP3-5        | —                | LgNIP3-5      | —                    | LINIP3-5      |
| LaNIP3-6        | —                | LgNIP3-6      | —                    | —             |
| LaNIP3-7        | —                | —             | —                    | —             |
| LaNIP3-1        | —                | —             | —                    | —             |
| LaNIP3-1        | —                | —             | —                    | —             |
| LaNIP1-5        | —                | —             | —                    | —             |
| LaNIP1-6        | —                | —             | —                    | —             |
| PIP             |                  |               |                      |               |
| LaPIP1-1        | LbPIP1-1         | LgPIP1-1      | —                    | —             |
| LaPIP1-2        | LbPIP1-2a        | LgPIP1-2a     | LPIP1-2              | —             |
| LbPIP1-2b       | LgPIP1-2b        | —             | —                    | —             |
| LaPIP1-3        | LbPIP1-3a        | LgPIP1-3      | LPIP1-3              | —             |
| LbPIP1-3b       | LgPIP1-3b        | —             | —                    | —             |
| LaPIP1-4        | LbPIP1-4a        | LgPIP1-4      | LPIP1-4              | —             |
| LbPIP1-4b       | LgPIP1-4b        | —             | —                    | —             |
| LaPIP1-5        | LbPIP1-5a        | —             | —                    | —             |
| LbPIP1-5b       | LgPIP1-5b        | —             | —                    | —             |
| LaPIP2-1        | —                | —             | LPIP2-1              | —             |
| LaPIP2-2        | LbPIP2-2a        | —             | —                    | —             |
| LbPIP2-2b       | LgPIP2-2b        | —             | —                    | —             |
| LaPIP2-3        | LbPIP2-3         | LgPIP2-3     | LPIP2-3              | —             |
| LaPIP2-4        | LbPIP2-4a        | LgPIP2-4     | —                    | —             |
| LbPIP2-4b       | LgPIP2-4b        | —             | —                    | —             |
| LaPIP2-5        | LbPIP2-5a        | LgPIP2-5     | —                    | LPIP2-5      |
| LbPIP2-5b       | LgPIP2-5b        | —             | —                    | —             |
| LaPIP2-6        | LbPIP2-6a        | LgPIP2-6     | LPIP2-6              | —             |
| LbPIP2-6b       | LgPIP2-6b        | —             | —                    | LPIP2-6b     |
| LaPIP2-7        | LbPIP2-7a        | LgPIP2-7a    | —                    | —             |
| LbPIP2-7b       | LgPIP2-7b        | —             | —                    | —             |
| LbPIP2-7c       | —                | —             | —                    | —             |
| LaPIP2-8        | LbPIP2-8         | —             | —                    | —             |
| LbPIP2-9a       | LgPIP2-9a        | —             | —                    | LPIP2-9      |
| LbPIP2-9b       | LgPIP2-9b        | —             | —                    | —             |
| LbPIP2-9c       | —                | —             | —                    | —             |
| LbPIP2-9d       | —                | —             | —                    | —             |
| LaPIP2-10       | LbPIP2-10        | LgPIP2-10    | —                    | —             |
| LbPIP2-11       | LbPIP2-11a       | LgPIP2-11    | LPIP2-11a            | —             |
| LbPIP2-11b      | —                | —             | LPIP2-11b            | —             |

Continued
Table 3. List of aquaporins identified in different *Linum* species. In comparison to 51 AQPs in *L. usitatissimum*, three related species of *Linum* such as *L. bienne* harbors 49 AQPs, *L. grandiflorum* possesses 39 AQPs and *L. leonii* possesses 19 AQPs.

| Aquaporin Family | L. usitatissimum | Linum bienne | Linum grandiflorum | Linum leonii |
|------------------|-----------------|-------------|-------------------|-------------|
| TIP              | LuTIP1-1        | LbTIP1-1    | LgTIP1-1          | —           |
|                  | LuTIP1-2        | LbTIP1-2a   | LgTIP1-2a         | LITIP1-2    |
|                  |                 | LbTIP1-2b   | LgTIP1-2b         | —           |
|                  | LuTIP1-3        | LbTIP1-3    | LgTIP1-3          | —           |
|                  | LuTIP1-4        | LbTIP1-4    | LgTIP1-4          | —           |
|                  | LuTIP1-5        | LbTIP1-5    | LgTIP1-5          | LITIP1-5    |
|                  | LuTIP1-6        | LbTIP1-6    | —                 | —           |
|                  | LuTIP1-7        | —           | LgTIP1-7          | —           |
|                  | LbTIP1-2        | —           | —                 | —           |
|                  | LbTIP1-3        | LgTIP2-2    | LITIP2-2          | —           |
|                  | LbTIP1-4        | LgTIP2-3    | —                 | —           |
|                  | LbTIP1-5        | LgTIP2-4    | —                 | —           |
|                  | LbTIP1-6        | —           | —                 | —           |
|                  | LbTIP1-2        | —           | —                 | —           |
|                  | LbTIP1-3        | —           | —                 | —           |
|                  | LbTIP1-4        | —           | —                 | —           |
|                  | LbTIP1-5        | —           | —                 | —           |
|                  | LbTIP2-1        | —           | —                 | —           |
|                  | LuSIP1-1        | LbSIP1-1    | LgSIP1-1          | LLSIP1-1    |
|                  | LuSIP1-2        | LbSIP1-2    | —                 | —           |
|                  | LuXIP1-1        | LbXIP1-1    | LgXIP1-1          | LLEXIP1-1   |
|                  | LuXIP1-2        | —           | —                 | —           |
|                  | LuXIP2-1        | —           | LgXIP2-1          | —           |
| Number of MIPs   | 51              | 49          | 39                | 19          |

Figure 5. Analysis of flax aquaporins expression using microarray data. Majority of PIPs showed higher level of expression compared to other MIPs across multiple tissues in flax. The different tissues included for expression analysis are root, leaf, stem inner at vegetative stage (SIV), stem inner at green capsule stage (SIGC), stem outer at vegetative stage (SOV), stem outer at green capsule stage (SOGC), developing seed 10 DAF (S1), 20 DAF (S2), 40 DAF (S3).
ESTs for 31 out of the 51 identified AQPs (Table 1). The highest numbers of ESTs were found for LuTIP3-3 (91), LuTIP3-4 (80) and LuPIP2-1 (52), whereas, least number of ESTs (0-1) were observed for the NIP family members among the different groups of AQPs.

Analysis of a priori reported microarray data revealed the expression of thirty three AQPs out of 51 AQPs represented on the array (Fig. 5). Among different AQPs, majority of the TIPs showed low expression whereas many PIP family members showed higher expression across nine different tissues of flax such as root, leaf, stem (stem inner at vegetative stage, stem inner at green capsule stage, stem outer at vegetative stage, stem outer at green capsule stage) and developing seed at 10 days after flowering (DAF), 20DAF and 40DAF. The expression was calculated in terms of fold change of AQPs in different tissues of flax in comparison to root revealed similar results (see Supplementary Fig. S4). LuTIP3-1, LuTIP3-2, LuTIP3-3 and LuTIP3-4 showed higher accumulation in seeds as compared to other tissues. Majority of PIPs showed higher expression in leaf, root, stem, and initial stage of embryo development (10 DAF) while low expression was observed during later stages of embryo development (20 DAF and 40 DAF). While, PIP1s (PIP1-3 to PIP1-5) and PIP2s (PIP2-1, PIP2-2, PIP2-4) showed higher expression in both outer and inner stem, few PIP2s (PIP2-5, PIP2-11) showed differential expression in outer as well as inner stem during vegetative and green capsule stage in flax.

The RNA-seq data was congruent with expression signatures observed in microarray data-set. The majority of PIPs showed higher level of expression across thirteen different tissues (Fig. 6). All NIP homologs were well represented in the RNA-seq dataset and majority of them showed lower accumulation in different tissues compared to other family members. Among TIPs, TIP3 members (LuTIP3-1, LuTIP3-2 and LuTIP3-3) showed gradual increase in expression from globular stage of embryo to mature embryo stage during seed development in flax. The expression was also calculated in terms of fold change of AQPs in different tissues in comparison to root (see Supplementary Fig. S4). Similarly, in other plant species such as soybean, rice, Arabidopsis and Medicago, expression of TIP3s specific to seed tissue was observed (Fig. 7). The pattern of TIP3s expression gradually increasing from early stage of seed development to maturation was commonly observed in all species analyzed (Fig. 7).

Differential expression of genes at apical and basal region of flax stem was delineated from another set of RNA-seq transcriptome profiling. Comparison of expression pattern of AQPs in apical and basal region of flax stem identified eighteen AQPs with two fold transcript enrichment in apical region (AR) compared to basal region (BR) (see Supplementary Table S7). The differentially expressing 18 AQPs comprised of seven PIPs, five TIPs, five NIPs and one XIP. Among different sub-family of MIPs, members of PIP subfamily, PIP2-4, PIP2-5 and PIP2-7 showed higher expression levels with 12, 11 and 5.3 fold respectively in AR compared to BR.

Analysis of tertiary protein structure of flax AQPs. Homology based tertiary (3D) protein structure of all 51 flax AQPs predicted to form hourglass like structure with six transmembrane domains (see Supplementary Fig. S5). Pore structure and three dimensional geometry of a channel of TIP3 (TIP3–1, TIP3–2, TIP3–3, TIP3–4) family members obtained with “PoreWalker” software identified a pore that longitudinally runs from the extracellular to intracellular opening of the protein. The pore morphology clearly showed conservation
of pore size and two constrains that are known to act as selectivity barrier in the pore (Fig. 8). Though, PoreWalker analysis does not provide information about solute interaction, the data of pore morphology obtained with it helps to predict the solute permeability. Conservation of pore size and similar constrain in all the four TIP3s indicates its plausible role in water transport. Similarly, pore structure and three dimensional geometry of PIP2s (PIP2-4, PIP2-5 and PIP2-7) family members of flax obtained with “PoreWalker” software showed conservation for pore size and constrains in the pore (see Supplementary Fig. S6).

Figure 7. Expression profiling of TIP3 genes across different tissues and developmental stages in multiple crops, (A) Soybean, (B) Rice, (C) Arabidopsis, and (D) Medicago. Higher levels of TIP3 transcripts are found in seeds compared to different tissues and developmental stages across all the four plant species analyzed.
Discussion

Aquaporins are the key membrane transport proteins involved in transport of water and substrate in the plant system. Plants have, relatively, a high number of aquaporins that are evolved into specific subfamilies and groups comprising constitutively expressed, tissue-specific, temporal or environmental factors and stress induced AQPs. Recent studies also revealed role of AQPs in abiotic stress tolerance in arabidopsis and barley plants3,50,51. Recent spurt of decoding whole genome sequence of crops has led to the identification of varied gene families including AQPs in many plant species. The available draft genome sequence of flax provided us an opportunity to analyze the AQP gene family in flax vis-à-vis other taxa including rice, poplar and Arabidopsis. We identified 51 putative AQPs in flax genome, which is more than the number of AQPs identified in rice (34) and Arabidopsis (35).

Additionally, AQPs search performed using RNA-seq data in the three Linum species identified similar number of aquaporins in L. bienne having comparable chromosome number (n = 15) as flax. The L. grandiflorum (n = 8) and L. leonii (n = 9) showed less number of aquaporins besides having much longer contigs than the L. bienne with de novo assembly. The variation in the AQPs across the Linum species could be due to the differences in the genome size. However, there is a possibility of identification of more number of AQPs in these species with the sequencing of more whole genomes. In particular, the NIPs with low tissue specific expression can be a serendipitous discovery. The number of AQPs observed in flax was found to be similar to poplar (55) which is also a member of the order malpighiales. Whole-genome duplications and inter-specific hybridizations have resulted in expansion of gene copy number in plants. Thus, the presence of more number of AQPs in flax compared to Arabidopsis and rice is attributed to recent whole-genome duplication event that occurred about 5–9 MYA in flax lineage, after it’s divergence from poplar and castor38. Since, chromosome-scale assembly is not available in flax, the assembled sequence in the form of scaffolds was helpful to locate tandem duplications. However, analysis of genomic distribution of AQPs revealed the absence of tandem duplications among the flax AQPs.

As observed in most of the plant genomes, flax AQPs grouped into five sub-families (PIPs, TIPs, NIPs, SIPs, XIPs) except monocots and Brassicaceae which harbor four AQP (PIPs, TIPs, NIPs, SIPs) sub-families13,14. Number of flax AQPs in different sub-families was also similar to that of Populus. However, the number of XIPs and SIPs varied in both genomes compared to other sub-families. Six members each of SIPs and XIPs were reported in Populus, while in flax two SIPs and three XIPs were observed. Specifically, members of SIP2s and NIP2s were not observed in flax genome while NIP2s were also absent in Arabidopsis genome. The exon-intron distribution in members of flax AQP subfamilies were found to be similar to the gene structure of AQPs observed in other crop plants35,47. Similar gene structure indicates conserved function of AQPs in flax. It is well documented that the intron number is correlated with the gene expression, gene duplication, and diversification12 of genes in plants. Usually, the substrate specificity of the AQPs is determined by the hydrophobicity and size of the amino acids forming the pores24,26,27. These positions in flax AQPs are based on the previously known AQPs from other plant.

Figure 8. Pore morphology and dimensions of flax TIP3s. Protein tertiary structure showing pore morphology of LuTIP3 (A) family members. Cross section of the proteins showing pore is depicted for each family member along with the graph showing pore dimensions obtained from PoreWalker software (B).
species. Two highly conserved NPA motifs in loops B and E along with four amino acid residues forming aromatic/arginine filter determine the transport of substrate molecule. All PIP family members from flax contained more hydrophilic ar/R selectivity filter (F/H/T/R) a hallmark of water transporting aquaporins compared to other families. Similar ar/R selectivity filter was also observed in PIP family of aquaporins from other plant species such as A. thaliana, B. rapa, G. max, P. vulgaris, and R. communis.19,35,36,53–55. The water transporting AQPs from humans also contains a similar ar/R selectivity filter with large hydrophilic amino acid residues (F/H/C/R). PIPs are known to play a central role in transport of water regulating root and leaf hydraulics.5 In addition to water transport, PIPs are known to facilitate diffusion of CO₂ in mesophyll tissue of A. thaliana and N. tabacum affecting photosynthesis.56,57. Our expression analysis also showed abundant expression of PIPs in flax root, stem as well as leaves suggesting possible role of PIPs in water transport and CO₂ diffusion in flax.

Among LuTIPs, LuTIP1s were found to have residues (H/I/A/V, H/I/A/L) forming more hydrophilic ar/R filter compared to LuTIP2s and LuTIP3s which contained ar/R filter with H/I/G/R and H/I/A/R residues respectively. The residues present in ar/R selectivity filter in LuTIP subfamily were similar to TIPs from other plant species. TIPs are located mainly in vacuolar membrane and act as functional water transporters. A number of experiments have shown the ability of TIPs to facilitate transport of small solutes such as NH₄⁺, H₂O₂, and urea.58–60. Conserved structural motifs such as NPA motif and ar/R filter in TIPs involved in transport of water as well as substrate were also observed in flax.

Among the NIPs, NIP1s were found to be more hydrophilic (WVAR) compared to NIP3s (AVGR, SIAR, TIAR). Interestingly, in the present investigation, members of NIP2 sub-group were not observed in flax genome. In plant kingdom different species accumulate wide range of silicon.61. The ability of plants to absorb Si is attributed to the presence of NIP2s containing GSGR selectivity filter with a precise distance of 108 amino acids distance between the NPA domains62,63. Plant species considered as high accumulators of Si are known to accumulate up to 10% of Si on dry weight basis.64. Low expression of NIPs in plants without precise distance between NPA domains accumulate around 0.2% or less of silicon. Thus, less accumulation of Si in flax leaves is possibly due to absence of NIP2 members. NIPs exhibit low level of expression compared to PIPs and TIPs and their expression is confined to specialized cells and organs.65,66. Low expression of NIPs as observed in microarray and RNA-seq data of flax was also supported by least number of ESTs found in the flax EST database.

The variations in ar/R selective filter for XIP family members specifically at H5 position have been reported in different studies.13,34. In one study, Serine/Threonine residue was reported at H5 position in plants;34 while in other study Valine/Isoleucine was reported at H5 position in plants.34 The ar/R selectivity filter in XIPs from different plants is more hydrophobic in nature, while Valine/Isoleucine at H5 position increases its hydrophobicity. In flax, Valine/Leucine imparting more hydrophobicity occupies H5 position. This hydrophobic nature of XIPs facilitates transport of bulky and hydrophobic molecules such as glycerol, urea, and boracic acid in plants.67.

Analysis of microarray and RNA-seq data in toto revealed, both were congruent, higher accumulation of TIP3 specific transcripts in developing seeds of flax. TIP3s are generally involved in cellular osmoregulation and maturation of the vacuolar apparatus to support optimal water uptake and growth of the embryo during seed development and germination. Increasing level of TIP3 expression from early stages to seed maturation also suggests its role in seed desiccation process. Similar observations have been noted during seed maturation and initial phase of seed germination in Arabidopsis and seed specific expression in soybean.68,69. Differential expression analysis between apical and basal region of flax also identified many aquaporins in fiber development. Contrary to role of TIP3 in seed development, PIP2s are envisaged to control fibre length in cotton by mediation of turgor pressure in developing fibres. Among aquaporin encoding genes, PIP2s showed higher differential expression in flax.

Recently, AQPs are envisaged to control fibre length in cotton.43 Rapid elongation of cotton fibres is accomplished by the coordinated action of turgor potential across the tonoplast that pushes against and loosens the cell wall of fibre initials. During fibre elongation, enhanced turgor potential is generated by accumulation of sugar, malate and K⁺ besides influx of water by AQPs.44–46. Panoply of genes involved in osmoregulation and cell expansion during fibre elongation in cotton has been identified. Phylogenetically, AQPs of cotton involved in fibre development belong to five subfamilies (PIP, TIP, NIP, SIP, and unrecognized intrinsic proteins XIP) of which PIP2s are up-regulated during fibre development. PIP2s inflict rapid influx of water by resulting high turgor pressure that accentuates longitudinal and polar expansion of cotton fibre primordia. Precise role of PIP2s was further proved by developing RNAi transgenic plants targeting PIP2 that exhibited “Short-fibre phenotype” with >20% reduction in fibre production in cotton.47. It is reported that in the short fibre mutant of cotton (“Ligon lintless”)71–73 most of the AQPs such as PIP (seven genes), TIP (four genes) and NIP (two genes) are massively down-regulated (p < 0.0001) compared to the wild type G. hirsutum. Nonetheless, equivalent information is meager to ascertain the role of AQPs in bast fibre development in flax. Thus, the role of aquaporins in flax fibre development needs further investigation.

Conclusions

To the best of our knowledge, this is the first comprehensive genome-wide analysis of the AQP gene family in flax. The sequence comparison, phylogenetic analysis and expression analysis of AQPs in flax revealed presence of flax AQPs clustering into five sub-families. The global expression profiles of 51 AQP genes through microarray and RNA-seq data analysis revealed TIPs exhibit higher expression in flax. The RNA-seq data precisely pointed out low expression of NIPs in multiple tissues compared to other AQPs. A gradual increase in TIP3 expression was observed from globular stage till seed development in flax envisaging a pivotal physiological role of TIP in seed development. Further, absence of NIP2 AQP in flax was observed and was commensurate with low accumulation of silicon in flax. Besides water and substrate transport, AQPs are
reported to be intrinsically involved in fiber development in cotton. Coincidentally, both cotton and flax fibres are cellulosic in nature. Thus, a cardinal role of AQP s in flax fibre development is envisaged. In flax, particularly PIP family members (PIP1-3, -4, -5; PIP2-1, PIP2-2, -4, -5, and -11) showed higher level of expression in stem during vegetative stage followed by green capsule stage indicating possible role of PIPS in bast fibre development. Targeted identification of the AQPs specifically involved in water equilibrium via-a-vis fibre elongation will delineate the molecular mechanism of fibre development in flax. The AQPs identified in the present study provide wealth of information to be used for further characterization of aquaporins to understand their physiological role in this industrial cash crop.

Methods

Genome-wide identification and distribution of aquaporins in flax. The *Linum usitatissimum* v1.0 annotated scaffold assembly of flax genome was retrieved from phytozome database (https://phytozome.jgi.doe.gov/pz/portal.html). A local database of the protein sequences of flax genes was created in BioEdit ver. 7.2.574. The 466 AQP genes from 10 different plant species were employed as query to identify putative orthologs of AQP genes in flax in local database using BLASTp. An e-value of 10^-5 was kept as initial cut-off to identify high scoring pairs (HSPs). The blast output was tabulated and the HSPs showing >100 bit score were selected. Finally, redundant hits were removed to select unique sequences for further analysis.

Structural characterizations of flax aquaporins. The AQP sequences retrieved from phytozome database were employed to identify respective genes from flax genome retrieved from GigaDB database (http://gigadb.org/) using local blast in BioEdit ver. 7.2.574. The details of length and location of AQPs were obtained from phytozome database. The gene structure of flax AQPs was analyzed using GSDS ver. 2.07.

Identification of functional motif and transmembrane domains. The NPA motifs were identified in protein sequences using conserved domain database at NCBI (CDD). Transmembrane domains in the genes were identified using TMHMM and SOSUI software tools (http://www.cbs.dtu.dk, http://harrier.nagahama-i-bio.ac.jp). The transmembrane domains were further analysed manually to detect altered and/or missing domains.

Phylogenetic analysis of flax AQPs. The AQP sequences were aligned using CLUSTALW alignment function in MEGA6. The phylogentic tree was constructed by using maximum likelihood method and the stability of the branch node was measured by performing 1000 bootstrap. The subfamilies PIP, SIP, TIP, NIP and XIPs were classified in accordance to the nomenclature used for arabidopsis, rice and poplar. A phylogenetic tree was constructed using arabidopsis, rice, poplar and flax AQP sequences.

Tertiary protein structure prediction. The protein sequences of *LuTIP3* (TIP3-1, TIP3-2, TIP3-3, TIP3-4) and *LuPIP2* (PIP2-4, PIP2-5, PIP2-7) were submitted to the Phyre2 protein-modeling server (www.sbg.bio.ic.ac.uk/*phyre2). The results obtained in the form of PDB files were uploaded to PoreWalker server to predict tertiary protein structure via-a-vis pore size (http://www.ebi.ac.uk/thornton-srv/software/PoreWalker/) in identified TIPS and PIPS.

Identification of major intrinsic protein coding orthologs in different *Linum* species. Raw RNA sequencing reads SRR957663, SRR957662, SRR957228 from *Linum bienne*, *Linum grandiflorum* and *L. leonii* respectively, were downloaded from Short Read Archive (SRA) in NCBI. The raw reads were examined for the adaptor sequences. The reads were used for de novo assembly using CLC Genomics Workbench (version 9.0; CLC bio, Aarhus, Denmark). Parameters used for the de novo assembly were: word size 20, automatic bubble size 20, and minimum contig length 200. The N50 contig value was determined by sorting all contigs in decreasing order of size and the contigs were added until the total size reached at least half of the total size of all assembled contigs. To map reads back to contigs, options as mismatch cost 2, insertion cost 3, deletion cost 3, length fraction 0.5, similarity fraction 0.8, and color space error cost 3 were provided. A local database of the assembled sequences of flax transcripts was created in BioEdit ver. 7.2.574. The 51 AQP genes from flax were employed as query to identify putative orthologs of AQP genes in three different *Linum* species in local database using BLASTn. An e-value of 10^-5 was kept as initial cut-off to identify high scoring pairs (HSPs). The blast output was tabulated and the HSPs showing >100 bit score were selected. Finally, redundant hits were removed to select unique sequences. Further, unique sequences having length of >290 bp were considered as bona fide AQP orthologs.

Expression profiling of flax aquaporins. To identify the transcriptional evidence for the putative flax AQPs, their transcript sequence were used as query to search flax specific ESTs at the NCBI database (dated: Mar 2016; 2,86,856 sequences; http://blast.ncbi.nlm.nih.gov/) using BLASTn. The ESTs showing >99% identity were selected and the redundant hits were removed before determining number of EST hits for each AQP transcript. The microarray data by Fenart, et al. was downloaded from NCBI GEO database (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE21868). RMA normalized gene-level signal intensity (log2) values for the unigenes exhibiting similarity to AQPs were used from all replicates and averaged further. Similarly, the normalized RNA-Seq dataset generated by Kumar, et al. available at http://linum.ca/downloads/RNAseq was also used to analyze the expression of AQPs. A heat map for expression of AQPs was constructed with these values using TIGR Multi Experiment Viewer (MeV, http://www.tm4.org/mev.html). Hierarchical clustering with average linkage method was performed to cluster the genes.

Simultaneously, differential transcript expression data of apical region (AR) compared to basal region (BR) in flax (measured as normalized FPKM) was retrieved from NCBI GEO (GSE80718; http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE80718).
Si level of expression in any given tissue. Arabidopsis, and Medicago. Absolute expression values were obtained from The Bio-Analytical Resource for Plant Biology (http://bbc.botany.utoronto.ca/) and converted into relative expression by comparing with the maximum level of expression in any given tissue.

**Silicon quantification in flax leaf and comparison of Si content in different plant species.** Flax plants were grown in four replications with continuous supplementation of 1.7 mM Si in the form of potassium silicate as a regular irrigation. Leaf samples of 30 days-old plants were harvested and then dried at 65°C for 24 h in hot air drier. The Si content in the dried powder was measured by HCL-HF extraction method followed by colorimetric analysis. Phylogenetic tree of different plant species was developed based on the NCBI taxonomy using the PhytoT tool (http://phylot.biobyte.de/). Categorization of functional NIP2s with GSGR was based on attributes described in Deshmukh et al. 2015.

**References**

1. Steudle, E. Water uptake by plant roots: an integration of views. *Plant and Soil* **226**, 45–56 (2000).
2. Kaldenhoff, R., Kolling, A. & Richter, G. A novel blue light-and abscisic acid-inducible gene of Arabidopsis thaliana encoding an intrinsic membrane protein. *Plant molecular biology* **23**, 1187–1198 (1993).
3. Maurel, C., Verducq, L., Luu, D. T. & Santoni, V. Plant aquaporines: membrane channels with multiple integrated functions. *Annual review of plant biology* **59**, 595–624 (2008).
4. Kato, I. et al. Expression and localization of aquaporin-4 in sensory ganglia. *Biochemical and biophysical research communications* **451**, 562–567 (2014).
5. Kato, I. et al. A general anaesthetic propofol inhibits aquaporin-4 in the presence of Zn2+. *Biochemical Journal* **454**, 275–282 (2013).
6. Gomes, D. et al. Aquaporinins are multifunctional water and solute transporters highly divergent in living organisms. *Biochimica et biophysica acta* **1758**, 1213–1228 (2009).
7. Chaumont, F., Moshelion, M. & Daniell, M. J. Regulation of plant aquaporin activity. *Biology of the cell* **97**, 749–764 (2005).
8. Chaumont, F., Barrieu, F., Wojcik, E., Chrispeels, M. J. & Jung, R. Aquaporins constitute a large and highly divergent protein family in maize. *Plant physiology* **125**, 1206–1215 (2001).
9. Quigley, F., Rosenberg, J. M., Shachar-Hill, Y. & Bohnert, H. J. From genome to function: the Arabidopsis aquaporins. *Genome Biol.* **3**, research0001.0001–research0001.0017 (2002).
10. Johansson, U. et al. 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* **126**, 1358–1369 (2001).
11. Kaldenhoff, R. & Fischer, M. Functional aquaporin diversity in plants. *Biochimica et biophysica acta* **1758**, 1134–1141 (2006).
12. Borslap, A. C. Early diversification of plant aquaporins. *Trends in plant science* **7**, 529–530 (2002).
13. Danielson, J. A. & Johanson, U. Unexpected complexity of the aquaporin gene family in the moss Physcomitrella patens. *BMC Plant Biol.* **8**, 45 (2008).
14. Chaumont, F. & Tyerman, S. D. Aquaporins: highly regulated channels controlling plant water relations. *Plant physiology* **164**, 1600–1618 (2014).
15. Deshmukh, R., Bélanger, R. R. & Hartley, S. M. The role of aquaporins in silicon transport in plants. *Function and morphology of the xylem sheath bases.* *Physiologia Plantarum* **130**, 560–571 (2007).
16. Törnroth-Horsefield, S. et al. Structural mechanism of plant aquaporin gating. *Nature* **439**, 688–694 (2006).
17. Lee, J. K. et al. Structural basis for conductance by the archaeal aquaporin AqpM at 1.6 Å. *PNAS* **102**, 18932–18937 (2005).
18. Hu, L., Cui, D., Neill, S. & Cai, W. OsEXPA4 and OsRWC3 are involved in asymmetric growth during gravitropic bending of rice leaf sheath bases. *Physiologia Plantarum* **130**, 560–571 (2007).
19. Eriksson, U. K. Plant physiology. *Intrinsic membrane protein.* *Biochemical Journal* **325**, 1–9 (1997).
20. Eriksson, U. K. Subangstrom resolution X-ray structure details aquaporin-water interactions. *Science* **290**, 481–486 (2000).
21. Eriksson, U. K. et al. Subangstrom resolution X-ray structure details aquaporin-water interactions. *Science* **340**, 1346–1349 (2013).
22. Guan, X. G. et al. Structural basis for conductance by the archaeal aquaporin AqpM at 1.6 Å. *PNAS* **102**, 18932–18937 (2005).
23. Hu, L., Cui, D., Neill, S. & Cai, W. OsEXPA4 and OsRWC3 are involved in asymmetric growth during gravitropic bending of rice leaf sheath bases. *Physiologia Plantarum* **130**, 560–571 (2007).
24. Törnroth-Horsefield, S. et al. Structural mechanism of plant aquaporin gating. *Nature* **439**, 688–694 (2006).
25. Lee, J. K. et al. Structural basis for conductance by the archaeal aquaporin AqpM at 1.6 Å. *PNAS* **102**, 18932–18937 (2005).
26. Sui, H., Han, B.-G., Lee, J. K., Walian, P. & Jap, B. K. Structural basis of water-specific transport through the AQP1 water channel. *Nature* **414**, 872–878 (2001).
27. Fu, D. et al. Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* **290**, 481–486 (2000).
28. Eriksson, U. K. et al. Subangstrom resolution X-ray structure details aquaporin-water interactions. *Science* **340**, 1346–1349 (2013).
29. Guan, X. G. et al. NPA motifs play a key role in plasma membrane targeting of aquaporin-4. *IUBMB life* **62**, 222–226 (2010).
30. Deshmukh, R., R. et al. Precise spacing between the NPA domains of aquaporins is essential for silicon permeability in plants. *Journal of experimental botany* **59**, 1307–1316 (2008).
31. Froger, A., Thomas, D., Delamarche, C. & Tallur, B. Prediction of functional residues in water channels and related proteins. *Protein Science* **7**, 1458–1468 (1998).
32. Sakurai, J., Ishikawa, F., Yamaguchi, T., Uemura, M. & Maeshima, M. Identification of 33 rice aquaporin genes and analysis of their expression and function. *Plant cell biology* **46**, 1568–1577 (2005).
33. Gupta, A. B. & Sankaramakrishna, 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* **9**, 134 (2009).
34. Deshmukh, R. K. et al. Identification and functional characterization of silicon transporters in soybean using comparative genomics of major intrinsic proteins in Arabidopsis and rice. *Plant molecular biology* **83**, 303–315 (2013).
36. Zhang da, Y. et al. Genome-wide sequence characterization and expression analysis of major intrinsic proteins in soybean (Glycine max L.). *Plos one* **8**, e56312 (2013).

37. Reuscher, S. et al. The sugar transporter inventory of tomato: genome-wide identification and expression analysis. *Plant and Cell Physiology* **55**, 1123–1141 (2014).

38. Wang, Z. et al. The genome of flax (Linum usitatissimum) assembled de novo from short shotgun sequence reads. *The Plant Journal* **72**, 461–473 (2012).

39. Dash, P. K. et al. Genome-wide analysis of drought induced gene expression changes in flax (Linum usitatissimum). *GM crops & food* **5**, 106–119 (2014).

40. Dash, P., Gupta, P. & Rai, R. Hydroponic method of halophoric response elicitation in flax (Linum usitatissimum) for precise down-stream gene expression studies. *International Journal of Tropical Agriculture* **33**, 1079–1085 (2015).

41. Dash, P. & Dash, P. K. Precise method of in situ drought stress induction in flax (Linum usitatissimum) for RNA isolation towards down-stream analysis. *Annual Agriculture Research* **36**, 10–17 (2015).

42. Gattolin, S., Sorieul, M. & Frigerio, L. Mapping of tonoplast intrinsic proteins in maturing and germinating Arabidopsis seeds reveals dual localization of embryonic TIPs to the tonoplast and plasma membrane. *Molecular Plant* **4**, 180–189 (2011).

43. Li, D. D. et al. Cotton plasma membrane intrinsic protein 2s (PIP2s) selectively interact to regulate their water channel activities and are required for fibre development. *The New phytologist* **199**, 695–707 (2013).

44. Utsubo, S., Shibasaki, M., Maekawa, M. & Katsuhara, M. Control of the Water Transport Activity of Barley HvTIP1; 1 Specifically Expressed in Seeds. *Plant and Cell Physiology* **56**, 1831–1840 (2015).

45. Hodson, M., White, P. J. J., Mead, A. & Broadway, M. Phylogenetic variation in the silicon composition of plants. *Annals of Botany* **96**, 1027–1046 (2005).

46. Vulavala, V. K. et al. Silicon fertilization of potato: expression of putative transporters and tuber skin quality. *Plantas** **243**, 217–229 (2016).

47. Deokar, A. A. & Tar’an, B. Genome-Wide Analysis of the Aquaporin Gene Family in chickpea (Cicer arietinum L.). *Front. in Plant Sci* **7** (2016).

48. Fenart, S. et al. Development and validation of a flax (Linum usitatissimum L.) gene expression oligo microarray. *BMC genomics* **11**, 592 (2010).

49. Zhang, N. & Deyholos, M. K. RNASeq Analysis of the Shoot Apex of Flax (Linum usitatissimum) to Identify Phloem Fiber Specification Genes. *Front. in Plant Sci* **7** (2016).

50. Alexandersson, E. et al. Transcriptional regulation of aquaporins in accessions of Arabidopsis in response to drought stress. *The Plant Journal* **61**, 650–660 (2010).

51. Horie, T. et al. Mechanisms of water transport mediated by PIP aquaporins and their regulation via phosphorylation events under salinity stress in barley roots. *Plant and Cell Physiology* **52**, 663–675 (2011).

52. Deshmukh, R. K., Sonah, H. & Singh, N. K. Intron gain, a dominant evolutionary process supporting high levels of gene expression in rice. *Journal of Plant Biochemistry and Biotechnology* **25**, 142–146 (2016).

53. Diehn, T. A., Pomerreneg, B., Bernhardt, N., Hartmann, A. & Bienert, G. P. Genome-wide identification of aquaporin encoding genes in Brassica oleracea and their phylogenetic sequence comparison to Brassica crops and Arabidopsis. *Front. Plant Sci* **6**, 166 (2015).

54. Zou, Z. et al. Gene Structures, Evolution, Classification and Expression Profiles of the Aquaporin Gene Family in Castor Bean (Ricinus communis L.). *Plos one* **10**, e0141022 (2015).

55. Ariani, A. & Gepts, P. Genome-wide identification and characterization of aquaporin gene family in common bean (Phaseolus vulgaris L.). *Mol. Genet. Genomics* **290**, 1771–1785 (2015).

56. Flexas, J. et al. Tobacco aquaporin MAQIP1 is involved in mesophyll conductance to CO_{2} in vivo. *The Plant Journal* **48**, 427–439 (2006).

57. Heckwolf, M., Pater, D., Hanson, D. T. & Kaldenhoff, R. The Arabidopsis thaliana aquaporin AtPIP1; 1 is a physiologically relevant CO_{2} transport facilitator. *The Plant Journal* **67**, 795–804 (2011).

58. Bienert, G. P. et al. Specific aquaporins facilitate the diffusion of hydrogen peroxide across membranes. *Journal of Biological Chemistry* **282**, 1183–1192 (2007).

59. Holm, L. M. et al. NH_{4}^{+} permeability in aquaporin-expressing Xenopus oocytes. *Pflugers Archiv* **450**, 415–428 (2005).

60. Liu, L. H., Ludewig, U., Gassert, B., Frommer, W. B. & von Wirén, N. Urea transport by nitrogen-regulated tonoplast intrinsic proteins in Arabidopsis. *Plant physiology* **133**, 1220–1228 (2003).

61. Jones, L. & Handreck, K. Silica in soils, plants, and animals. *Advances inagonomy* **19**, 107–149 (1967).

62. Ma, J. F. et al. A silicon transporter in rice. *Nature* **440**, 688–691 (2006).

63. Miti, N., Chiba, Y., Yamaji, N. & Ma, J. F. Identification and characterization of maize and barley Lsi2-like silicon efflux transporters reveals a distinct silicon uptake pattern from that in rice. *The Plant Cell* **21**, 2133–2142 (2009).

64. Epstein, E. The anomaly of silicon in plant biology. *PNAS* **91**, 11–17 (1994).

65. Alexandersson, E. et al. Whole gene family expression and drought stress regulation of aquaporins. *Plant molecular biology* **59**, 469–484 (2005).

66. Choi, W. G. & Roberts, D. M. Arabidopsis NIP2; 1, a major intrinsic protein transporter of lactic acid induced by anoxic stress. *Journal of Biological Chemistry* **282**, 24209–24218 (2007).

67. Bienert, G. P., Bienert, M. D., Jahn, T. P., Boutry, M. & Chaumont, F. Solanaceae XIPs are plasma membrane aquaporins that facilitate the transport of many uncharged substrates. *The Plant Journal* **66**, 306–317 (2011).

68. Martini, C., Bhatt, K. & Baumann, K. Shaping in plant cells. *Current opinion in plant biology* **4**, 540–549 (2001).

69. Ruan, Y.-L., Xu, S.-M., White, R. & Furhank, R. T. Genotypic and developmental evidence for the role of plasmodesmatal regulation in cotton fiber elongation mediated by callose turnover. *Plant physiology* **136**, 4104–4113 (2004).

70. Park, W., Scheffler, B. E., Bauer, P. J. & Campbell, B. T. Identification of the family of aquaporin genes and their expression in upland cotton (Gossypium hirsutum L.). *BMC plant biology* **10**, 1 (2010).

71. Narbuth, E. & Kohel, R. Inheritance and linkage analysis of a new fiber mutant in cotton. *Journal of Heredity* **81**, 131–133 (1990).

72. Kohel, R. Linkage Tests in Upland Cotton. Gossypium hirsutum L. *Ill. Crop Science* **12**, 66–69 (1972).

73. Naojima, K., Thysen, G. N. & Fang, D. D. RNA-seq analysis of short fiber mutants Ligon-lintless-1 (Li 1) and –2 (Li 2) revealed important role of aquaporins in cotton (Gossypium hirsutum L.) fiber elongation. *BMC plant biology* **15**, 1 (2015).

74. Hall, T. A. In Nucleic Acids Symposium Series; 95–98 (1999).

75. Hu, B. et al. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* **31**, 1296–1297 (2015).

76. Kumar, S., Nell, M., Dudley, J. & Tamura, K. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in bioinformatics* **9**, 299–306 (2008).

77. Kumar, S., Jordan, M. C., Datla, R. & Cloutier, S. The LuWD40-1 gene encoding WD repeat protein regulates growth and pollen viability in flax (Linum usitatissimum L.). *Plos one* **8**, e69124 (2013).

78. Taber, H. G., Shogren, D. & Lu, G. Extraction of silicon from plant tissue with dilute HCl and HF and measurement by modified inductive coupled argon plasma procedures. *Communications in soil science and plant analysis* **33**, 1661–1670 (2002).
Acknowledgements
The authors are grateful to ICAR-NPTC, ICAR-NASF for financial assistance to P.K.D., R.R., S.M.S. and Dr. Debabrata Sarkar, ICAR-CRIJAF, West Bengal for valuable suggestions.

Author Contributions
P.K.D., R.R. and P.K.A. designed the study. S.M.S., R.K.D. and P.K.D. analyzed the data. S.M.S. and P.K.D. wrote the manuscript with input from R.R., P.K.A. and R.B. All authors read and approved the manuscript.

Additional Information
Supplementary information accompanies this paper at http://www.nature.com/srep

Competing Interests: The authors declare no competing financial interests.

How to cite this article: Shivaraj, S. M. et al. Genome-wide identification, characterization, and expression profile of aquaporin gene family in flax (Linum usitatissimum). Sci. Rep. 7, 46137; doi: 10.1038/srep46137 (2017).

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2017