Title

Genome-wide identification and characterization of aquaporin gene family in common bean (Phaseolus vulgaris L.)

Journal

Molecular Genetics and Genomics, 290(5)

ISSN

1617-4615

Authors

Ariani, Andrea
Gepts, Paul

Publication Date

2015-10-01

DOI

10.1007/s00438-015-1038-2

Peer reviewed

Genome-wide identification and characterization of aquaporin gene family in common bean (*Phaseolus vulgaris* L.)

Andrea Ariani, Paul Gepts

Department of Plant Sciences/MS1, University of California, 1 Shields Avenue, Davis, CA 95616-8780, USA

Corresponding author:
Andrea Ariani
email: aariani@ucdavis.edu
Tel: +1-530-220-3208
Fax: +1-530-752-4361

Abstract

Plant ~~A~~aquaporins are a large and diverse family of water channel proteins that are essential for several physiological processes in living organisms. ~~in all living organisms~~. ~~In plants, this family has the highest diversity of protein isoforms and substrate specificities compared to animals and bacteria.~~ Numerous studies have linked plant aquaporins with a plethora of processes, such as nutrient acquisition, CO_2 transport, plant growth and development, and response to abiotic stresses. However, little is known~~s~~ about this protein family in common bean. Here, we present a genome-wide identification of the aquaporin gene family in common bean (*Phaseolus vulgaris* L.), a legume crop essential for human nutrition. We identified 41 full-length coding aquaporin sequences in the common bean genome, divided by phylogenetic analysis into five sub-families (PIPs, TIPs, NIPs, SIPs and XIPs). Residues determining substrate specificity of aquaporins (i.e., NPA motifs and ar/R selectivity filter) seem conserved between common bean and other plant species, allowing inference of substrate specificity for these proteins. Thanks to the availability of RNA-sequencing datasets, expression levels in different organs and in leaves of wild and domesticated bean accessions were evaluated. Three aquaporins (PvTIP1;1, PvPIP2;4 and PvPIP1;2) have the overall highest mean expressions, with PvTIP1;1 having the highest expression among all aquaporins. We performed an EST database mining to identify drought responsive aquaporins in common bean. This analysis showed a significant increase in expression for PvTIP1;1 in drought stress conditions compared to well-watered environments. The pivotal role suggested for PvTIP1;1 in regulating water homeostasis and drought stress response in the common bean, should be verified by further field experimentation under drought stress.

Keywords: Aquaporin, drought response, gene expression, gene structure, water homeostasis

Introduction

Aquaporins are water channel proteins that facilitate diffusion of water and small molecules across biological membranes (Gomes et al. 2009). In plants, aquaporins showed surprisingly diverse and numerous isoforms per species, with more than 30 members identified in different plant genomes (Chaumont et al. 2001; Johanson et al. 2001; Reuscher et al. 2013). Analysis of the *Arabidopsis* genome, the first plant with a completely sequenced genome (AGI 2000), identified four sub-families of aquaporins, namely plasma membrane intrinsic proteins (PIPs), with two groups (PIPs1 and PIPs2); tonoplastic intrinsic proteins (TIPs); nodulin 26-like intrinsic proteins (NIPs) and small intrinsic proteins (SIPs) (Johanson et al. 2001; Gomes et al. 2009). The sequencing of other plant genomes also identified an additional sub-family absent in *Arabidopsis*, named X-intrinsic proteins (XIPs) (Gupta and Sankararamakrishnan 2009; Bienert et al. 2011; Reuscher et al. 2013).

Aquaporins assemble in biological membranes as homo- or hetero-tetramers in which each monomer forms a water channel (Chaumont et al. 2005). Different experiments have shown that tetramerization between members of the PIP1 and PIP2 sub-families modifies both water channel activity and protein sub-cellular localization (Fetter et al. 2004; Zelazny et al. 2007, 2009). The main features of aquaporin protein monomers are highly conserved in all living organisms, consisting of six transmembrane domains with a central aqueous pore characterized by two NPA motifs (Asp-Pro-Ala). These motifs are critical for the transport function of these proteins (Murata et al. 2000). Other important residues in aquaporin sequences are the ones forming the aromatic/arginine selectivity filter (ar/R). This region is characterized by four residues, one in helix 2 (H2), one in helix 5 (H5), and two in loop E (LE1 and LE2). The ar/R selectivity filter acts as a size-exclusion barrier and regulates the transport specificity of these proteins (Hub and de Groot 2008; Mitani-Ueno et al. 2011).

Aquaporins are involved in almost every physiological process in plants, such as cell differentiation and elongation, plant transpiration, and regulation of plant hydraulics (Maurel 2007; Maurel et al. 2008; Li et al. 2014). Experimental evidence points to variations in expression or in protein accumulation in response to environmental stresses, hormone treatments, and the ability to transport small molecules like ammonia, urea, boron, silicon and CO_2 (Maurel et al. 2008; Li et al. 2014). Several experiments identified aquaporins as involved in response to drought stress in common beans varieties, and marked differences in gene expression were observed in drought-resistant versus susceptible genotypes (Aroca et al. 2006; Montalvo-Hernández et al. 2008; Recchia et al. 2013). In particular, a tolerant genotype showed increased expression of a tonoplast aquaporin in drought-stressed roots, with re-localization of its mRNA in phloem tissues (Montalvo-Hernández et al. 2008).

Because it is a major source of protein, complex carbohydrates, and micronutrients, the common bean (*P. vulgaris* L.) is one of the most important crops for human consumption worldwide (Broughton et al. 2003). Wild common bean is organized into two genetically different gene pools (Mesoamerican and Andean) (Koenig and Gepts 1989; Kwak and Gepts 2009; Mamidi et al. 2013), with a broad environmental distribution in Mexico and Central to South America (Cortés et al. 2013). These two gene pools were independently domesticated in what is now Mexico and South America (Gepts et al. 1986; Bitocchi et al. 2013) about 8,000 years ago. The release of a reference *P. vulgaris* genome sequence (Schmutz et al. 2014), together with the development and application of Next Generation Sequencing technologies in plant species (Bräutigam and Gowik 2010) and synteny among plants, such as soybean and *Arabidopsi*s, enabled a genome-wide identification of actual or putative loci involved in the domestication process of this species (Repinski et al. 2012; Kwak et al. 2012; Bellucci et al. 2014; Schmutz et al. 2014).

The availability of a reference genome sequence also allows the identification and characterization of important plant gene families, thus facilitating genomic, evolutionary, and molecular studies in different species. From the current genome sequence of *P. vulgaris*, we identified the full set of aquaporin-coding sequences in the bean genome and characterized them at phylogenetic and structural levels. We also identified single nucleotide polymorphism (SNP) markers from the BARCBEAN6K_3 array (BeanCAP; www.beancap.org) located near the identified aquaporins. We then analyzed the expression of these genes in different plant organs and in different wild and domesticated bean accessions using publicly available expression datasets (see Material and Methods). Furthermore, we integrated the previously characterized aquaporins in a proposed nomenclature. We then analyzed the expression of aquaporin genes in response to drought stress using ESTs resources available at NCBI. Overall, this first genome-wide identification and analysis of common bean aquaporins provides a foundation for identifying and further functionally characterizing important aquaporins in common bean. Eventually, this information can lead to molecular breeding improvement of drought-related traits in this species.

Materials and Methods

Identification of aquaporin genes and phylogenetic analysis

All the sequence-related information (gene sequences, reference genome, protein sequences, annotation and features) were retrieved from Phytozome (http://phytozome.jgi.doe.gov/pz/portal.html) (Goodstein et al. 2012). The multiple sequence alignments and phylogenetic analyses were performed using the Seaview toolkit (Guoy et al. 2010). For multiple sequence alignments, the MUSCLE algorithm, with default parameters, was used. Pairwise sequence alignments were performed on protein sequence with the needle algorithm of the EMBOSS suite using default parameters (Rice et al. 2000). *P. vulgaris* aquaporins were identified using a Hidden Markov Models (HMMs) probabilistic method implemented in HMMER3 (http://hmmer.janelia.org/). In brief, the complete set of aquaporin protein sequences of *Arabidopsis*, identified by Johanson et al. (2001), were aligned with the MUSCLE algorithm for building a HMM profile. This profile was queried to the full proteome of *P. vulgaris*. The resulting putative aquaporins were aligned with MUSCLE and manually checked for the presence of the two motifs forming the aqueous pore typical of aquaporins. Proteins without this two motifs were discarded from the analysis.

For phylogenetic analysis and nomenclature, *Arabidopsis* and *Populus trichocarpa* (Johanson et al. 2001; Lopez et al. 2012) aquaporin protein sequences, with their proposed nomenclature, were aligned with the *P. vulgaris* aquaporins using MUSCLE. A Bio Neighbor-Joining (BioNJ) tree, based on observed distance with 1000 bootstrap replicates, was built.

Tandem duplicated genes were identified as those with adjacent chromosomal localization, as determined also by the ID of the loci.

Gene structure, conserved residues, and *in silico* prediction of transmembrane domains and sub-cellular localization

Gene structure for each aquaporin was illustrated with the Gene Structure Display Server (http://gsds.cbi.pku.edu.cn/). Conserved residues of the NPA domain and ar/R selectivity filter (Murata et al. 2000; Hub and de Groot 2008; Mitani-Ueno et al. 2011) were identified by multiple sequence alignment with MUSCLE of the different sub-families of *P. vulgaris* aquaporins together with the tomato aquaporins identified by Reuscher et al. (2013). Prediction of transmembrane domains was performed using TMHMM server v.2.0, using default parameters (http://www.cbs.dtu.dk/services/TMHMM/) (Krogh et al. 2001). Prediction of sub-cellular localization was performed using both WoLF PSORT (Horton et al. 2007) and Plant-mPLoc (Chou

and Shen 2010) algorithms with default parameters.

Expression analysis

For expression analysis in different organs, the FPKM (fragments per kilobase of transcript per million mapped reads) values for each aquaporin were extracted from the common bean genes.fpkm_tracking file, available in Phytozome. From this expression table, the FPKM values < 1 were excluded and treated as a lack of expression of the corresponding gene (FPKM=0). For analysis of aquaporin expression in different wild and domesticated *P. vulgaris* accessions, raw RNA-sequencing reads from Bellucci et al. (2014) (SRP028116) were downloaded from Short Read Archive (SRA) in NCBI using the R package (www.r-project.org) SRAdb (Zhu et al. 2013). Reads were quality-filtered with Sickle (https://github.com/najoshi/sickle) using default parameters, and only reads without any uncertain called bases (Ns) were retained for subsequent alignment. Reads were aligned on the *P. vulgaris* genome sequence using TopHat (Trapnell et al. 2009) with a maximum of five mismatches, a maximum gap length of 3 and a maximum of one mismatch in the anchor region of a spliced alignment. Transcripts expression, represented as FPKM, were estimated using cufflinks (Trapnell et al. 2010) using upper-quartile normalization (-N) and multi-read correction (-u) parameters.; For subsequent analyses, only the FPKM value of each gene model, without considering splicing isoforms, was taken into account. For co-expression analysis, the Pearson correlation coefficient (r) based on FPKM was calculated.

To analyze gene expression of aquaporins in response to drought stress, all the *P. vulgaris* ESTs were downloaded from NCBI. We then divided all the ESTs in drought-related (i.e., libraries prepared from plants under drought stress) and background ESTs (all ESTs), and aligned them with BLASTN to *P. vulgaris* coding DNA sequences (CDSs). BLAST searches were performed with an e value cut-off of $1 \times e^{-10}$ and only the best hit for each EST was considered for subsequent analysis. Only aquaporins identified by both datasets were taken into account for differential expression analysis. Expression was estimated by counting the sequences aligning to a specific CDS. Differential expression was evaluated with a χ^2 test, based on a 2x2 contingency table of alignment counts. For each χ^2 test, p values were corrected with 10^3 Monte Carlo simulations of p values. Expression between the two datasets were normalized by dividing the number of ESTs aligning to a specific aquaporin by all the aligned ESTs in the dataset. Differential expression was expressed as the ratio between the two normalized values.

Integration of previously characterized aquaporins and BARCBean6K_3 SNP markers

Protein sequences of the common bean aquaporins previously characterized by Aroca et al. (2006),

and Montalvo-Hernández et al. (2008) were downloaded from the NCBI nucleotide database (http://www.ncbi.nlm.nih.gov/nuccore) using the respective GenBank ID (U97023, AY995196, AY995195, DQ087217, and DQ087218). The sequences were blasted against the *P. vulgaris* proteome using the BLASTP program, with default parameters, in the Phytozome platform (Goodstein et al. 2012). Since the genomic positions of the BARCBean6K_3 SNP loci are known (Bello et al. 2014), the BEDTools suite (Quinlan and Hall 2010) was used for calculating the distance between common bean aquaporins and BARCBean6K_3 SNP markers. To identify drought-related SNP markers in the BARCBean6K_3 chip, we chose the markers nearest todrought-QTLs identified by Mukeshimana et al. (2014).

Results

Aquaporin identification and analysis

Analysis of the available *P. vulgaris* genome identified 42 putative aquaporin-encoding genes located on all but one chromosome, in a non-uniform manner. Chromosome 3 (Chr 03) contained one aquaporin (Phvul.003G040100). In contrast, Chromosomes 1 and 11 contained six aquaporins each. Chromosome 6 had two members of the NIP sub-family exclusively. Manual checking of multiple protein alignment of the putative identified aquaporins showed that Phvul.003G040100 ~~lacked the first~~has a complete deletion of the first NPA domain (Supplementary File S1). Analysis of RNA-Sequencing read coverage using the jbrowse plugin on Phytozome showed a low log-coverage (< 1) for this gene (Phytozome jbrowse RNA-Sequencing log-coverage link in Supplementary File S1). Moreover, analysis of expression with the Phytomine plug-in on Phytozome showed a FPKM of 0 in every organ analyzed (Phytomine FPKM expression link in Supplementary File S1). Thus, t~~T~~his protein was removed for further analysis because it is probably a ~~mis~~miss~~s~~-annotation or a pseudo-gene sequence. With the removal of this sequence, the total number of aquaporins in the *P. vulgaris* genome is 41, with no~~out~~ aquaporin located on Chr 03 (Fig. 1). The number of aquaporins identified in common bean is higher than that identified in *Arabidopsis* and maize, but lower compared to species like *P. trichocarpa* and tomato. The list of the identified aquaporins, together with basic statistics on primary protein sequences, is reported in Table 1. Protein length ranged from 235 (Phvul.001G108800) to 324 residues (Phvul.008G041100), with a median of 272 amino acids per protein. Prediction of transmembrane domains showed six transmembrane domains (TMDs) for the majority of aquaporins (32 of 41, 78%), except for six aquaporins having five predicted TMDs and three having seven predicted TMDs (Table 1). Multiple sequence alignment of the 6th helix (H6) region does not show any gap in this region for protein predicted to have 5 TMDs, except ~~for PvPIP1~~for PvPIP1;4, ~~that~~ which shows a shorter C-term (Fig. 2). However, ~~–~~PvPIP1;4 is expressed in almost every organ and genotype analyzed, except in ~~G~~green mature pods and in the leaves of genotype PI325677 (Supplementary File S2). The two different sub-cellular localization prediction tools identified the same localization for 23 aquaporins (56%), locating them mainly in the plasma membrane and tonoplast. For the other 18 aquaporins, Plant-mPLoc predicted ~~mainly~~ a localization mainly in the plasma membrane or tonoplast; in contrast, WoLF PSORT predicted several different sub-cellular localizations, including chloroplast, cytoplasm, and mitochondria (Table 1).

Phylogenetic and tandem duplication analysis of aquaporin family

A phylogenetic analysis combining common bean aquaporins with those of *Arabidopsis* and *P. trichocarpa* clearly grouped this protein family into five sub-families consisting of 12 PIPs, 13 TIPs, 10 NIPs, 4 SIPs, and 2 XIPs (Fig. 3~~2~~, Table 1). The TIPs sub-family was the biggest one with 13 members divided into five groups (TIP1 to TIP5), six members in the TIP1 group, three in TIP2, two in TIP3 and one each in groups TIP4 and TIP5. The PIP sub-family was divided into PIP1, with 5 members, and PIP2, with 7 members. NIPs were divided into NIP1, with 3 members, NIP6, with 2 members, and NIP2, 4, 5 and 7, with one member each. SIPs were divided into SIP1, with 3 members, and SIP2, with one member. Only one group of XIPs was identified, with 2 members.

A pair of putatively tandem duplicated genes was detected for two PIP2 aquaporins, four NIP1 (two pair of duplicated genes), and the two members of the XIP sub-family. A pairwise protein sequence analysis showed family-specific percentage of identity among the putatively duplicated pairs. In particular, PvPIP2;1 and PvPIP2;2 showed ~98% identity, the two pairs of NIPs (PvNIP1;1 and PvNIP1;2, PvNIP1;3 and PvNIP1;4) ~71% and the two XIPs less than 60% (Supplementary File S3~~1~~).

Analysis of exon-intron structure

The complete transcript annotation of *P. vulgaris* genome enables the analysis and comparison of the structural features of aquaporins in different families. Within each aquaporin sub-family, the number, order and length of exons were almost conserved; in contrast, the introns and UTRs showed variable lengths and patterns (Fig. 4~~3~~). The majority of the members of the PIP sub-family were characterized by four exons of similar length, except PvPIP1;4, which showed a short 4th exon without a 3'UTR, and PvPIP2;6, which had only three exons. Most of the members of the TIP sub-family showed three exons, except PvTIP1;1 and PvT~~P~~IP1;6, with two, and PvTIP1;5, with four exons. In addition, PvTIP1;2, PvTIP1;5 and PvTIP5;1 did not have UTRs at both 5' and 3' ends. The genes in the NIP sub-family showed mainly five exons, except for PvNIP2;1 PvNIP5;1 and PvNIP6;1, which exhibited four. In this family, PvNIP4;1 did not have any UTRs and PvNIP7;1 showed a longer 5' UTR compared to other aquaporins in the same family. SIP genes showed three exons, except PvSIP1;3 that had only one exon and no introns. The two members of the XIP sub-family showed completely different features, with PvXIP1;1 having two exons, and PvXIP1;2 having one exon without any UTRs.

Analysis of conserved and substrate-specific residues in aquaporin proteins

To understand the possible physiological role and substrate specificity of common bean aquaporins, we identified and analyzed the NPA motifs and ar/R selectivity filter sequences (Table 2). In the PIP

and TIP sub-families, both NPA domain are conserved and showed the typical Asp-Pro-Ala residues of aquaporins. In the NIP sub-family, the first NPA showed the same sequence as in PIPs and TIPs, except for PvNIP5;1, where the alanine is replaced by a serine residue. The second NPA motif showed an alanine to valine substitution in four NIPs (PvNIP1;3, PvNIP1;5, PvNIP6;1 and PvNIP6;2). SIP aquaporins showed a second NPA motif completely conserved with the other sub-families, instead all of the first NPA motifs show the replacement of alanine by threonine (PvSIP1;1 and PvSIP1;2), serine (PvSIP1;3) or leucine (PvSIP2;1). The XIP sub-family showed a second, completely conserved NPA, while the alanine is changed to isoleucine (PvXIP1;1) or leucine (PvXIP1;2) in the first NPA.

The ar/R positions showed an increased family-specific sequence compared to the two NPA motifs. In all PIPs, this selectivity filter showed the conserved residues, observed in other species, phenylalanine in H2, histidine in H5, threonine in LE1, and arginine in LE2. In the TIP sub-family, the ar/R is formed by histidine/asparagine/serine in H2; isoleucine/valine in H5, alanine/glycine in LE1 and valine/alanine/arginine/leucine/cysteine in LE2. In the NIP sub-family, this selection filter is constituted by tryptophan/alanine/threonine/asparagine in H2, valine/serine/isoleucine in H5, alanine/glycine/serine in LE1 and arginine in LE5. The SIP sub-family showed phenylalanine/alanine/tyrosine in H2, threonine/lysine in H5, proline/glycine in LE1, and aparagine/phenylalanine/serine in LE2. The XIPs sub-family is quite homogeneous in these sites showing valine in H2 and LE1, methionine in H5, and asparagine in LE2.

Expression analysis

Gene expression data in different organs with RNA-sequencing are available from Phytozome (see Materials and Methods) and was downloaded for characterizing the expression in different organs of the identified aquaporins. A heat map showing aquaporin expression is shown in Fig. 5~~4~~a. Thirty-six aquaporins (~88%) were expressed in at least one organ analyzed, and 17 of them (~41%) were expressed in all organs, including 10 PIPs, 4 TIPs, 2 SIPs, and 1 NIP. The number of aquaporins expressed in different organs range from 25 to 33, with the highest number in young pods (33, ~80%) and flower buds (30, ~73%), and the lowest in green mature pods and leaves (25 each, ~61%). Among all the *P. vulgaris* aquaporins, PvTIP1;1 showed the highest expression level in all the organs analyzed; in contrast, five aquaporins (PvTIP1;2, PvTIP3;2, PvTIP5;1, PvNIP1;1 and PvXIP1;2) were not expressed in any of these organs. The aquaporins with the highest mean expression (> 90% percentile) across all organs were PvTIP1;1 (mean FPKM=2303.06), PvPIP1;2 (mean FPKM= 672.22), PvPIP2;4 (mean FPKM=589.01) and PvPIP2;7 (mean FPKM=448.68).

Analysis of expression in different genotypes showed a similar expression pattern observed within

different tissues (Supplementary file S4~~2~~), with PvTIP1;1 having the highest expression of all aquaporins in almost all genotypes. The aquaporins with the highest mean expression (> 90% percentile) across all genotypes were PvTIP1;1 (mean FPKM=2291.62), PvPIP2;4 (mean FPKM=1363.05), PvPIP1;2 (mean FPKM=1244.26) and PvPIP1;1 (mean FPKM=781.36).

Since hetero-~~tetrametization~~ tetramerization between aquaporin isoforms, especially between members of PIP1 and PIP2 sub-families, have been previously observed, we performed a co-expression analysis among the different genotypes for identifying possible interacting proteins. This analysis showed that the majority of common bean aquaporins did not have any correlation at the expression level, except for a few members of PIPs sub-family (Fig. 5~~4~~b). Among these co-expressed proteins, we found three pairs (PvPIP1;4 and PvPIP2;7, PvPIP1;4 and PvPIP2;3, PvPIP1;3 and PvPIP2;7) with a Pearson correlation coefficient > 0.9 (Supplementary file S5~~3~~).

Expression analysis in response to drought stress was performed using EST databases available in NCBI. ESTs for 28 aquaporins (~68%) were identified in at least one of the datasets, while in the drought-related dataset 254 ESTs identify only 12 aquaporins (~30%). The aquaporins with the highest expression (based on the number of ESTs aligning to them) were PvTIP1;1, PvPIP1;2, PvPIP2;4 and PvPIP2;5. Differential expression analysis based on χ^2 test showed a significant differential expression ($p < 0.01$) for PvTIP1;1, with a relative increase of ~1.6 fold in comparison to the control dataset (Supplementary file S6~~4~~).

Integration of previously characterized aquaporins and BARCBean6K_3 SNPs markers

Protein sequences of the aquaporins previously characterized by Aroca et al. (2006) and Montalvo-Hernández et al. (2008), were assigned to the current gene nomenclature based on the best blast alignment. PvPIP1;2 is the best blast hit for U97023 (97.9% identity), PvPIP1;3 for AY995196 (99.7% identity), PvPIP2;5 for AY995195 (98.6% identity), and PvTIP2;3 for both DQ087217 (100% identity) and DQ087218 (100% identity).

Nine (22%) of the common bean aquaporins are the nearest annotated locus to eight different SNP markers, with PvPIP2;1 being the closest locus to both ss715641384 and ss715650422 (Supplementary file S7~~5~~). In particular, three SNP markers are located inside gene sequences of three different aquaporins: ss715641100 is located in an intron sequence of PvSIP1;2; in contrast, ss715645362 and ss715639327 are located in exon sequences of PvTIP2;1 and PvSIP1;3, respectively. The other aquaporins are located between 1500bp and 13,000bp from the closest SNP marker (Supplementary file S8~~5~~). Recently, Mukeshimana et al. (2014) identified several Quantitative Trait Loci (QTLs) associated with drought stress in common bean using the BARCBean6K_3 genotyping platform. None of the common bean aquaporins was the closest

annotated locus to ~~an~~ identified drought QTLs. The distance between the QTLs and the closest aquaporins ranged from 0.4 to 14 Mb (Supplemental file S6).

Discussion

Aquaporins are an abundant and highly divergent protein family in plants. Recent efforts in plant genome sequencing projects have enabled the identification and characterization of the complete set of aquaporins in several plant species (Chaumont et al. 2001; Johanson et al. 2001; Gupta and Sankararamakrishnan 2009; Reuscher et al. 2013; Tao et al. 2014). These analyses could be extremely helpful in the identification of physiologically important aquaporins for a specific species. In addition, *in silico* analyses of aquaporin protein family could guide other researchers in hypothesis generation and testing, facilitating in depth functional characterization of this protein family.

The recent release of *P. vulgaris* genome (Schmutz et al. 2014) allowed the identification and characterization of the entire aquaporin family in this study. We identified 41 full-length aquaporin-coding sequences in the common bean genome, with 12 genes belonging to PIPs sub-family, 13 to TIPs, 10 to NIPs, four to SIPs, and two to XIPs. The number of aquaporins identified in the common bean is higher than those identified in species like *Arabidopsis* and maize, but lower compared to species like *P. trichocarpa* and tomato. A note of caution should be introduced here. The distribution and count of aquaporin sequences was based on the current reference sequence of common bean (Schmutz et al. 2014), which originated in the Andean landrace G19833. Based on the divergence between the Andean and Mesoamerican gene pools, which includes partial reproductive isolation (Koinange and Gepts 1992), a similar analysis should be conducted with Mesoamerican reference sequences, once these become available (http://phasibeam.crg.eu/wiki/Main_Page ; http://www.beangenomics.ca/).

Analysis of chromosomal distribution of aquaporins showed the existence of tandem duplications within aquaporin families. Gene duplication is an important mechanism for increasing genetic variability and creating novel genes (Magadum et al. 2013). In plants, this mechanism is a major force in the increase of gene family size and in functional expansion of genomes (Severing et al. 2009; Magadum et al. 2013). Duplicated genes could undergo pseudogenization, gene conservation and sub- or neo-functionalization (Lynch and Conery 2000; Flagel and Wendel 2009).

The high similarity at both protein and gene structure levels between PvPIP2;1 and PvPIP2;2 suggests a recent tandem duplication and, possibly, a redundant function of these two genes in common bean. This redundant function could also be inferred from similar expression patterns in different organs. In the NIP sub-family, two pairs of putatively tandem duplicated genes were identified: PvNIP1;1 - PvNIP1;2 and PvNIP1;3 - PvNIP1;4. These pairs of genes showed a similar gene structure, but displayed just ~70% similarity at the protein level and divergent expression patterns. In particular, PvNIP1;1 could have undergone pseudogenization since it was not expressed

in any organ. PvNIP1;3 is only slightly expressed in 19-day-old roots, while PvNIP1;4 is highly expressed in both 10- and 19-day-old roots. The most striking difference among thisese pair of putatively duplicated genes is in the two members of the XIP sub-family. Even though they are adjacent on Chromosome 11 they showed a low level of protein sequence similarity, different expression patterns, and a completely different gene structure. PvXIP1;1 is expressed only in flowers, while PvXIP1;2 is not expressed in any organ.

An essential feature of aquaporins is the presence of two NPA motifs that form the water pore, essential for selective transport of water transport (Murata et al. 2000). In addition, the ar/R selectivity filter is essential in determining transport capabilityspecificity of aquaporins. Point mutations or other sequence variations in these residues confer different substrate specificities to aquaporins (Beitz et al. 2006; Hub and de Groot, 2008; Mitani-Ueno et al. 2013).

All *P. vulgaris* PIPs aquaporins showed the characteristic double NPA motif in their protein sequences. In addition, the ar/R selectivity filter is highly conserved and typical of water-transporting aquaporins (F/H/T/R). The same ar/R selectivity filter sequence is shared by the PIP sub-family in different plant species such as *Z. mays*, *A. thaliana*, *P. trichocarpa*, *S. lycopersicum*, *G. max* and *B. rapa* (Chaumont et al. 2001; Johanson et al. 2001; Gupta and Sankararamakrishnan 2009; Reuscher et al. 2013; Zhang et al. 2013; Tao et al. 2014). SIn plants,everal researchers hypothesize PIP aquaporins have a central role in roots water absorption in rootsfor plant PIP aquaporins, suggesting thatand this proteins could also actively regulate root and leaf hydraulics (Maurel et al. 2008; Li et al. 2014). In addition to water transport, different studies have showedn that PIPs aquaporins from *N. tabacum*, *A. thaliana* and barley could facilitate the diffusion of CO_2 in mesophyll and could directly affect photosynthesis (Flexas et al. 2006; Heckwolf et al. 2003; Mori et al. 2014). The conservation of important transport residues (i.e., NPA motifs and ar/R residues) in common bean PIPs could suggest a similar role of these proteins in regulating water absorption, plant hydraulics and/or CO_2 diffusion. Expression analysis showed that among PIPs, PvPIP1;2 and PvPIP2;4 have the highest expression across all organs and in leaves of different genotypes. Thus, we suggest that these two aquaporins could have a central physiological role in regulating water movement and homeostasis in common bean.

Members of the PIP1 and PIP2 sub-families interact *in vivo*. This interaction determines an additive increase in water transport and the re-localization of PIP1 aquaporins in plasma membrane (Fetter et al. 2004; Zelazny et al. 2007, 2009). Co-expression network analysis is widely used in plant and other organisms for identifying and annotating novel gene functions (Usadel et al. 2009), but it showed to be less useful for identifying interacting proteins (Xulvi-Brunet and Li 2010; Giorgi et al. 2013). Despite this, our co-expression analysis showed that pairs of PIP1 and PIP2 members have a

high Pearson correlation coefficient (> 0.9), pointing to a possible interaction between these proteins pair.

All *P. vulgaris* TIP aquaporins showed the canonical double NPA motif and a group specific ar/R selectivity filter that were conserved across different species. TIP aquaporins are located mainly in vacuolar membranes and regulate ~~cellular oo~~smotic~~sis~~ potential and water flow across this essential plant sub-cellular compartment (Maurel et al. 2008, Li et al. 2014). The expression levels of TIPs were also associated with cell growth, cell expansion, and response to gibberellic acid (Ludevid et al. 1992; Phillips and Huttly 1994). Several experiments have shown the ability of TIPs to also transport a wide variety~~ies~~ of small solutes, such as NH_4^+, H_2O_2, and urea, besides water (Liu et al. 2003, Holm et al. 2005; Loque et al. 2005; Bienert et al. 2007). The conservation of ar/R residues and NPA motifs between TIPs of common bean and other species could suggests a conserved function for these proteins. Expression analysis showed the highest mean expression in all organs for PvTIP1;1, suggesting its possible central role in regulating water homeostasis and cellular osmosis in common bean. Since the homologs of PvTIP1;1 in *Arabidopsis* and *Panax ginseng* were related to plant growth (Phillips and Huttly 1994), or increased growth rate and biomass production when expressed in transgenic plants (Peng et al. 2007), we suggest that PvTIP1;1 could be involved in regulation of plant growth in common bean as well.

The NIPs aquaporin sub-family was named by its first described member, the *G. max* nodulin 26 protein, identified in the symbiosome membranes of the nitrogen-assimilating root nodules (Wallace et al. 2006). This aquaporin sub-family has a wide range of substrate specificity such as water, glycerol, urea, boric acid, silicon, and lactic acid (Dean et al. 1999; Wallace and Roberts 2005; Ma et al. 2006; Takano et al. 2006; Choi and Roberts 2007). This specificity is based mainly on sequence differences in the ar/R selectivity filter that are used also for dividing this sub-family in three distinct subgroups (Wallace and Roberts 2005). The ar/R selectivity filter of *P. vulgaris* NIP1 group is typical of the subgroup I of plant NIP aquaporins. These proteins are characterized by the residues W/V/A/R in the ar/R filter, low water permeability, and the capability of transporting uncharged solutes like glycerol and formamide (Wallace and Roberts 2005). Due to these structural similarities, we suggest that common bean NIP1 group could also have similar transport specificity as previously demonstrated for homologous aquaporins from other species.

OsNIP2;1 is a rice silicon transporter characterized by a double NPA motif and a G/S/G/R ar/R selectivity filter (Ma et al. 2006). This aquaporin is also able to transport arsenite and boric acid when expressed in *Xenopus* oocytes. Mutation in the H5 position of the ar/R selectivity filter resulted in a loss of transport activity for these molecules (Mitani-Ueno et al. 2011). On the other hand, a G to A mutation in the H2 position of the ar/R selectivity filter does not affect the transport

capability of this protein (Mitani-Ueno et al. 2011). Since PvNIP2;1 showed sequence similarities with OsNIP2;1, with a G to A mutation in the H2 position of the ar/R selectivity filter, we suggest that this protein could be involved in both silicon or boric acid homeostasis in common bean. AtNIP5;1 is an *Arabidopsis* boric acid transporter characterized by a NPS/NPV aqueous pore and an A/I/G/R selectivity filter (Takano et al. 2006). Mutations in the ar/R selectivity filter regions alter the transport capability of this protein. Since AtNIP5;1 and its common bean homolog, PvNIP5;1, share the same NPS/NPV aqueous pore and ar/R selectivity filter, we suggest that PvNIP5;1 could be involved in boron transport in common bean.

Several experiments showed differential expression of plant aquaporins in response to hormones and to environmental stresses, such as drought, cold, and salinity (Li et al. 2014). This differential expression could be plant-, aquaporin-, or stress-specific. Differential expression of aquaporins in response to different environmental stresses has been comprehensively evaluated in several plant species such as *Arabidopsis*, maize, poplar and *B. rapa* in the past years (Alexandersson et al. 2005; Zhu et al. 2005; Cohen et al. 2013; Tao et al. 2014). Regarding common bean, few aquaporins have been characterized to date. Aroca et al. (2006) cloned and characterized three *P. vulgaris* PIP aquaporins (GenBank ID: U97023, AY995196 and AY995195), whose best BLAST hits were, according to our nomenclature, PvPIP1;2 (U97023), PvPIP1;3 (AY995196) and PvPIP2;5 (AY995195). Under drought stress, only PvPIP1;2 was up-regulated in leaves, while all the three PIPs were up-regulated in roots under similar stress (Aroca et al. 2006).

Transcriptome analysis in common bean under drought stress using EST approaches showed contrasting expression patterns between tolerant and susceptible genotypes (Montalvo-Hernández et al. 2008; Recchia et al. 2013). Susceptible genotypes showed a marked down-regulation of aquaporins expression in response to drought, while resistant showed a stable or higher expression of these genes. In particular, under drought stress, a putative TIP aquaporin was highly expressed in root phloem tissues of a drought-resistant genotype compared to the susceptible one (Montalvo-Hernández et al. 2008). The best BLAST hit of this aquaporin, identified and cloned in a drought-tolerant (GenBank ID: DQ087217) and a susceptible (GenBank ID: DQ087218) genotype, is PvTIP2;3. Interestingly, three of the previously characterized drought-responsive aquaporins were identified as putative molecular targets of selection during common bean domestication. In particular, PvPIP2;5 was identified as putatively selected during the Mesoamerican gene pool domestication, whereas PvPIP1;3 and PvTIP2;3 were putatively selected during the Andean gene pool domestication (Schmutz et al. 2014).

Two different classes of transcription factors involved in drought-response were previously characterized in wild and domesticated common beans. Two *Asr* genes, involved in an ABA-

dependent drought-response (Cortés et al. 2012a), and two *Dreb* genes, involved in an ABA-independent drought response (Cortés et al. 2012b). Both studies showed a higher genetic diversity and signature of selection in wild common bean in comparison to domesticated ones. Even though additional studies are needed, we suggest that PvPIP1;3, PvPIP2;5, and PvTIP2;3 could show similar patterns of diversity and selection, which could be a potential source of drought tolerance traits in cultivated common bean.

The EST-database-mining analysis showed a marginal but not significant increase of PvTIP2;3 expression in response to drought stress. The only aquaporin that showed a significant increase of expression was PvTIP1;1. This result suggests a possible role for PvTIP1;1 in response to drought stress in common bean. One of the mechanisms for avoiding drought stress in common bean is an increased root depth (Beebe et al. 2013). The expression pattern of AtTIP1;1 promoter is correlated with elongating cells in *Arabidopsis* roots (Ludevid et al. 1992). In addition, over-expression of *P. ginseng PgTIP1* in transgenic *Arabidopsis* plants determined an increased tolerance to drought stress in comparison to wild-type plants (Peng et al. 2007). Thus, we propose a putative role for PvTIP1;1 aquaporin in regulating drought response in common bean.

Despite the possible importance of aquaporin in regulating drought stress response in common bean (Aroca et al. 2006; Montalvo-Hernández et al. 2008; Recchia et al. 2013), none of the drought-QTLs identified by Mukeshimana et al. (2014) are located near aquaporin loci. The response variables analyzed in the above-mentioned study were mainly related to flowering time, maturity, and seed productivity, suggesting that aquaporins could have only a marginal role in these developmental processes in common bean. Early flowering and maturity are known escape mechanisms in drought tolerance, but do not represent biochemical or physiological tolerance mechanisms to which aquaporins might contribute. Nonetheless, we ~~could~~can ~~–~~not ~~completely~~ exclude that aquaporins could be important in regulating drought stress in other bean cultivars.

In summary, the recent release of the *P. vulgaris* genome sequence enabled us to comprehensively identify and analyze the aquaporin protein family in common bean at the structural and sequence levels. We identified 41 full-length aquaporins in the common bean genome, divided in PIPs, TIPs, NIPs, SIPs, and XIPs, based on a phylogenetic analysis. Analysis of conserved structural residues showed a sub-family-specific differentiation that is conserved across different plant species. Through these residues, we inferred putative functions and substrate-specificities of common bean aquaporins, based on the functional characterization of homologous proteins in other species. Analysis of expression datasets of common bean, in the forms of RNA-sequencing and EST databases, suggests that a few aquaporins regulate water homeostasis and drought stress response in *P. vulgaris*. In particular, PvTIP1;1 has the highest FPKM expression in the RNA-Sequencing

datasets analyzed.~~,~~ ~~and~~It is the only aquaporin showing significant differential expression based on EST database mining. ~~In particular, PvTIP1;1 seems to have a pivotal role~~These results ~~could~~ suggest an important role for PvTIP1;1 in regulating water homeostasis, plant hydraulic and drought response in this species. Nonethless, f~~F~~urther experiments~~,~~ aiming at~~focusing on~~ the functional characterization of the identified aquaporins, are needed for a better understanding of the molecular and physiological role of these genes in common bean.

Acknowledgments

This project was supported by Agriculture and Food Research Initiative (AFRI) Competitive Grant no. 2013-67013-21224 from the USDA National Institute of Food and Agriculture.

Compliance with ethical standards

This article does not contain any studies with human participants or animals performed by any of the authors.

References

Alexandersson E, Fraysse L, Sjövall-Larsen S, Gustavsson S, Fellert M, Karlsson M, Johanson U, Kjellbom P (2005) Whole gene family expression and drought stress regulation of aquaporins. Plant Mol Biol 59:469-484.

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408: 796-815.

Aroca R, Ferrante A, Vernieri P, Chrispeels MJ (2006) Drought, abscisic acid and transpiration rate effects on the regulation of PIP gene expression and abundance in *Phaseolus vulgaris* plants. Ann Bot 98: 1301-1310.

Beebe SE, Rao IM, Blair MW, Acosta-Gallegos JA (2013) Phenotyping common beans for adaptation to drought. Front Physiol 4: 35.

Beitz E, Wu B, Holm LM, Schultz JE, Zeuthen T (2006) Point mutations in the aromatic arginine region in aquaporin 1 allow passage of urea, glycerol, ammonia and protons. Proc Natl Acad Sci USA 103: 269-274.

Bello MH, Moghaddam SM, Massoudi M, McClean PE, Cregan PB, Miklas PN (2014) Application of in silico bulked segregant analysis for rapid development of markers linked to Bean common mosaic virus resistance in common bean. BMC Genomics 15: 903.

Bellucci E, Bitocchi E, Ferrarini A, Benazzo A, Biagetti E, Klie S, Minio A, Rau D, Rodriguez M, Panziera A, Venturini L, Attene G, Albertini E, Jackson SA, Nanni L, Fernie AR, Nikoloski Z, Bertorelle G, Delledonne M, Papa R, (2014) Decreased nucleotide and expression diversity and modified coexpression patterns characterize domestication in the common bean. Plant Cell 26: 1901-1912.

Bienert GP, Møller AL, Kristiansen KA, Schulz A, Møller IM, Schjoerring JK, Jahn TP (2007) Specific aquaporins facilitate the diffusion of hydrogen peroxide across membranes. J Biol Chem 282: 1183-1192.

Bienert GP, Bienert MD, Jahn TP, Boutry M, Chaumont F (2011) Solanaceae XIPs are plasma

membrane aquaporins that facilitate the transport of many uncharged substrates. Plant J 66: 306-317.

Bitocchi E, Bellucci E, Giardini A, Rau D, Rodriguez M, Biagetti E, Santilocchi R, Spagnoletti Zeuli P, Gioia T, Logozzo G, Attene G, Nanni L, Papa R (2013) Molecular analysis of the parallel domestication of common bean (*Phaseolus vulgaris*) in Mesoamerica and the Andes. New Phytol 197: 300-313.

Bräutigam A, Gowik U (2010) What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. Plant Biol 12: 831-841.

Broughton WJ, Hernández G, Blair M, Beebe S, Gepts P, Vanderleyden J (2003) Beans (*Phaseolus* spp.) - model food legumes. Plant Soil 252: 55-128.

Chaumont F, Barrieu F, Wojcik E, Chrispeels MJ, Jung R (2001) Aquaporins constitute a large and highly divergent protein family in maize. Plant Physiol 152: 1206-1215.

Chaumont F, Moshelion M, Daniels MJ (2005) Regulation of plant aquaporin activity. Biol Cell 97:746-764.

Choi WG, Roberts DM (2007) *Arabidopsis* NIP2;1:a major intrinsic protein transporter of lactic acid induced by anoxic stress. J Biol Chem 282: 24209-24218.

Chou KC, Shen HB (2010) Plant-mPLoc: a top-down strategy to augment the power for predicting plant protein subcellular localization. PLoS ONE 5: e11335.

Cohen D, Bogeat-Triboulot MB, Vialet-Chabrand S, Merret R, Courty PE, Moretti S, Bizet F, Guillot A, Hummel I (2013) Developmental and environmental regulation of aquaporin gene expression across *Populus* species: divergence or redundancy? PLoS One 8:e55506.

Cortés AJ, Chavarro MC, Madriñán S, This D, Blair MW (2012a) Molecular ecology and selection in the drought-related Asr gene polymorphisms in wild and cultivated common bean (*Phaseolus vulgaris* L.). BMC Genet 13: 58.

Cortés AJ, Monserrate FA, Ramirez-Villegas J, Madriñán S, Blair MV (2013) Drought tolerance in wild plant population: the case of common beans (*Phaseolus vulgaris* L.). PLoS One 8: e62898.

Cortés AJ1, This D, Chavarro C, Madriñán S, Blair MW (2012b) Nucleotide diversity patterns at the drought-related DREB2 encoding genes in wild and cultivated common bean (*Phaseolus vulgaris* L.). Theor Appl Genet 125: 1069-1085.

Dean RM, Rivers RL, Zeidel ML, Roberts DM (1999) Purification and functional reconstitution of soybean nodulin 26. An aquaporin with water and glycerol transport properties. Biochemistry 38: 347-353.

Fetter K, Van Wilder V, Moshelion M, Chaumonf F (2004) Interactions between plasma membrane aquaporins modulate their water channel activity. Plant Cell 16: 215-228.

Flexas J, Ribas-Carbó M, Hanson DT, Bota J, Otto B, Cife J, McDowell N, Medrano H, Kaldenhoff R (2006) Tobacco aquaporin NtAQP1 is involved in mesophyll conductance to CO_2 *in vivo*. Plant J 48: 427-439.

Gepts P, Osborn T, Rashka K, Bliss F (1986) Phaseolin-protein variability in wild form and landraces of the common bean (*Phaseolus vulgaris*): evidence for multiple centers of domestication. Econ Bot 40: 451-468.

Giorgi FM, Del Fabbro C, Licausi F (2013) Comparative study of RNA-seq- and microarray-derived coexpression networks in *Arabidopsis thaliana*. Bioinformatics 15: 717-724.

Gomes D, Agasse A, Thiébaud P, Delrot S, Gerós H, Chaumont F (2009) Aquaporins are multifunctional water and solute transporters highly divergent in living organism. Biochem Biophys Acta 1788: 1213-1228.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acid Res 40: D1178-1186.

Guoy M, Guindon S, Gascuel O (2010) SeaView Version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 27: 221-224.
Gupta AB, Sankararamakrishnan R (2009) Genome wide analysis of major intrinsic proteins in the tree plant *Populus trichocarpa*: characterization of XIP subfamily of aquaporins from evolutionary perspective. BMC Plant Biol 9: 134
Heckwolf M, Pater D, Hanson DT, Kaldenhoff R (2011) The *Arabidopsis thaliana* aquaporin AtPIP1:2 is a physiologically relevant CO_2 transport facilitator. Plant J 67: 734-737.
Holm LM, Jahn TP, Møller AL, Schjoerring JK, Ferri D, Klaerke DA, Zeuthen T (2005) NH_3 and NH_4^+ permeability in aquaporin-expressin *Xenopus* oocytes. Pflugers Arch 450: 415-428.
Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K (2007) WoLF PSORT: protein localization predictor. Nucleic Acid Res 35: W585-587.
Hub JS, de Groot BL (2008) Mechanisms of selectivity in aquaporins and aquagliceroporins, Proc Natl Acad Sci USA 105: 1198-1203.
Johanson U, Karlsson M, Johansson I, Gustavsson S, Sjövall S, Fraysse L, Weig AR, Kjellbom P (2001) The complete set of genes encoding major intrinsic proteins in *Arabidopsis* provides a framework for a new nomenclature for major intrinsic protein in plants. Plant Physiol 126: 1358-1369.
Koenig R, Gepts P (1989) Allozyme diversity in wild *Phaseolus vulgaris*: further evidence for two major centers of diversity. Theor Appl Genet 78: 809-817
Krogh A, Larsson B, von Heijne G, Sonnhammer LL (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J Mol Biol 305: 567-580.
Kwak M, Gepts P (2009) Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). Theor Appl Genet 118: 979-992
Kwak M, Toro O, Debouck D, Gepts P (2012) Multiple origins of the determinate growth habit in domesticated common bean (*Phaseolus vulgaris* L.). Ann Bot 110: 1573-1580
Li G, Santoni V, Maurel C (2014) Plant aquaporins: role in plant physiology. Biochim Biophys Acta 1840: 1574-1582.
Liu L, Ludewig U, Gasset B, Frommer WB von Wiren N (2003) Urea transport by nitrogen-regulated tonoplastic intrinsic proteins in *Arabidopsis*. Plant Physiol 133: 1220-1228.
Lopez D, Bronner G, Brunel N, Auguin D, Bourgerie S, Brignolas F, Carpin S, Tournaire-Roux C, Maurel C, Fumanal B, Martin F, Sakr S, Label P, Julien JL, Gousset-Dupont A, Venisse JS (2012) Insight into *Populus* XIP aquaporins: evolutionary expansion, protein functionality and environmental regulation. J Exp Bot 63: 2217-2230.
Loque D, Ludewig U, Yuan L, von Wiren N (2005) Tonoplastic intrinsic proteins AtTIP2;1 and AtTIP2,3 facilitate NH_3 transport into the vacuole. Plant Physiol 137: 671-680.
Ludevid D, Höfte H, Himelblau E, Chrispeels MJ (1992) The expression pattern of the tonoplastic intrinsic protein γ-TIP in *Arabidopsis thaliana* is correlated with cell enlargement. Plant Physiol 100: 1633-1639.
Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290: 1151-1155.
Ma JG, Tamai K, Yamaji N, Mitani N, Konishi S, Katsuhara M, Ishiguro M, Murata Y, Yano M (2006) A silicon transporter in rice. Nature 440: 688-691.
Magadum S, Banerjee U, Murugan P, Gangapur D, Ravikesavan R (2013) Gene duplication as a major force in evolution. J Genet 92: 155-161.
Mamidi S, Rossi M, Moghaddam SM, Annam D, Lee R, Papa R, McClean PE (2013) Demographic factors shaped diversity in the two gene pools of wild common bean *Phaseolus vulgaris* L. Heredity 110: 267-276.
Maurel C (2007) Plant aquaporins: novel function and regulations properties. FEBS Lett 581: 2227-2236.

Maurel C, Verdoucq L, Luu DT, Santoni V (2008) Plant aquaporins: membrane channels with multipe integrated functions. Annu Rev Plant Biol 59: 595-624.

Mitani-Ueno N, Yamaji N, Zhao F, Ma JF (2011) The aromatic/arginine selectivity filter of NIP aquaporins play a critical role in substrate selectivity for silicon, boron and arsenic. J Exp Bot 62: 4391-4398.

Montalvo-Hernández L, Piedra-Ibarra E, Gómez-Silva L, Lira-Carmona R, Acosta-Gallegos JA, Vazquez-Medrano J, Xoconostle-Cázares B, Ruíz-Medrano R (2008) Differential accumulation of mRNAs in drought-tolerant and susceptible common bean cultivars in response to water deficit. New Phytol 177: 102-113.

Mori IC, Rhee J, Shibasaka M, Sasano S, Kaneko T, Horie T, Katsuhara M (2014) CO_2 transport by PIP2 aquaporins in barley. Plant Cell Physiol 55: 251-257.

Mukeshimana G, Butare L, Cregan PB, Blair MW, Kelly JD (2014) Quantitative Trait Loci Associated with Drought Tolerance in Common Bean. Crop Sci 54: 923-938.

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

Peng Y, Lin W, Cai W, Arora R (2007) Overexpression of a *Panax ginseng* tonoplast aquaporin alters salt tolerance, drougth tolerance and cold acclimation ability in transgenic *Arabidopsis* plants. Planta 226: 729-740.

Phillips AL, Huttly AK (1994) Cloning of two gibberellin-regulated cDNAs from *Arabidopsis thaliana* by subtractive hybridization: expression of the tonoplast water channel, γ-TIP, is increased by GA3. Plant Mol Biol 24: 603-615.

Quinlan AR1, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841-842.

Recchia GH, Caldas DG, Beraldo AL, da Silva MJ, Tsai SM (2013) Transcriptional analysis of drought-induced genes in the roots of a tolerant genotype in the common bean (*Phaseolus vulgaris* L.). Int J Mol Sci 14: 7155-7179.

Repinski SL, Kwak M, Gepts P (2012) The common bean growth habit gene PvTFL1y is a functional homolog of Arabidopsis TFL1. Theor Appl Genet 124: 1539-1547.

Reuscher S, Akiyama M, Mori C, Aoki K, Shibata D (2013) Genome-wide identification and expression analysis of aquaporins in tomato. PLoS ONE 8: e79052.

Rice P, Longden I, Bleasby A (2000) EMBOSS: The european molecular biology open software suite. Trends Genet 16: 276-277.

Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, Jenkins J, Shu S, Song Q, Chavarro C, Torres-Torres M, Geffroy V, Moghaddam SM, Gao D, Abernathy B, Barry K, Blair M, Brick MA, Chovatia M, Gepts P, Goodstein DM, Gonzales M, Hellsten U, Hyten DL, Jia G, Kelly JD, Kudrna D, Lee R, Richard MM, Miklas PN, Osorno JM, Rodrigues J, Thareau V, Urrea CA, Wang M, Yu Y, Zhang M, Wing RA, Cregan PB, Rokhsar DS, Jackson SA (2014) A reference genome for common bean and genome-wide analysis of dual domestication. Nat Genet 46: 707-713.

Severing EI, Van Dijk ADJ, Stiekema WJ, Van Ham RCHJ (2009) Comparative analysis indicates that alternative splicing in plants has a limited role in functional expansion of the proteome. BMC Genomics 10: 154.

Takano J, Wada M, Ludewig U, Schaaf G, von Wirén N, Fujiwara T (2006) The *Arabidopsis* major intrinsic protein NIP5;1 is essential for efficient boron uptake and plant development under boron limitation. Plant Cell 18: 1498-1509.

Tao P, Zhong X, Li B, Wang W, Yue Z, Lei J, Guo W, Huang X (2014) Genome-wide identification and characterization of aquaporin genes (AQP) in chinese cabbage (*Brassica rapa* spp. p*ekinensis*). Mol Genet Genomics doi:10.1007/s00438-014-0874-9.

Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovery splice junctions with RNA-Seq, Bioinformatics 25: 1105-1111.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28: 511-515.

Usadel B, Obayashi T, Mutwil M, Giorgi FM, Bassel GW, Tanimoto M, Chow A, Steinhauser D, Persson S, Provart NJ (2009) Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. Plant Cell Environ 32: 1633-1651.

Wallace IS, Choi W, Roberts DM (2006) The structure, function and regulation of the Nodulin 26-like intrinsic protein family of plant aquaglyceroporins. Biochim Biophys Acta 1758: 1165-1175.

Wallace IS, Roberts DM (2005) Distinct transport selectivity of two structural subclasses of the nodulin-like intrinsic protein family of plant aquaglyceroporin channels. Biochemistry 44: 16826-16834.

Xulvi-Brunet R, Li H (2010) Co-expression networks: graph properties and topological comparisons. Bioinformatics 26: 205-214.

Zelazny E, Borst JW, Muyalaert M, Batoko H, Hemminga, Chaumont F (2007) FRET imaging in living maize cells reveals that plasma membrane aquaporins interact to regulate their subcellular localization. Proc Natl Acad Sci USA 104: 12359-12364.

Zelazny E, Miecielica U, Borst JW, Hemminga MA, Chaumont F (2009) An N-terminal diacidic motif is required for the trafficking of maize aquaporins ZmPIP2;4 and ZMPIP2;5 to the plasma membrane. Plant J 57: 346-355.

Zhang DY, Ali Z, Wang CB, Xu L, Yi JX, Xu ZL, Liu XQ, He XL, Huang YH, Khan IA, Trethowan RM, Ma HX (2013) Genome-wide sequence characterization and expression analysis of major intrinsic proteins in soybean (*Glycine max* L.). PLoS ONE 8: e56312.

Zhu C, Schraut D, Hartung W, Schaffner AR (2005) Differential responses of maize MIP genes to salt stress and ABA. J Exp Bot 56:2971-2981

Zhu Y, Stephens RM, Meltzer PS, Davis SR (2013) SRAdb: query and use public next-generation sequencing data from within R. BMC Bioinformatics 14: 19.

Table 1 Nomenclature and analysis of primary structure of common bean aquaporins

	Gene Name	Locus	AA	TMD	WoLF PSORT[a]	Plant-mPLoc[a]
PIP	PvPIP1;1	Phvul.001G241200	289	6	Plas	Plas
	PvPIP1;2	Phvul.008G226000	289	5	Plas	Plas
	PvPIP1;3	Phvul.002G290400	287	6	Plas	Plas
	PvPIP1;4	Phvul.008G106400	249	5	Plas	Plas
	PvPIP1;5	Phvul.010G004500	287	6	Plas	Plas
	PvPIP2;1	Phvul.004G082600	285	6	Plas	Plas
	PvPIP2;2	Phvul.004G082700	285	6	Plas	Plas
	PvPIP2;3	Phvul.007G094600	287	6	Plas	Plas
	PvPIP2;4	Phvul.005G135300	287	6	Plas	Plas
	PvPIP2;5	Phvul.011G079300	285	6	Plas	Plas
	PvPIP2;6	Phvul.001G177000	286	6	Plas	Plas
	PvPIP2;7	Phvul.009G118900	278	6	Plas	Plas
TIP	PvTIP1;1	Phvul.001G181100	250	7	Cyto	Vacu
	PvTIP1;2	Phvul.007G010800	252	6	Cyto/Plas	Vacu
	PvTIP1;3	Phvul.008G014900	252	6	Cyto	Vacu
	PvTIP1;4	Phvul.011G067200	252	6	Vacu	Vacu
	PvTIP1;5	Phvul.005G141100	238	6	Vacu	Vacu
	PvTIP1;6	Phvul.007G151600	253	6	Mito	Vacu
	PvTIP2;1	Phvul.005G170300	248	6	Plas	Vacu
	PvTIP2;2	Phvul.010G147100	248	7	Vacu	Vacu
	PvTIP2;3	Phvul.002G057500	249	6	Plas	Vacu
	PvTIP3;1	Phvul.004G136000	256	6	Mito	Vacu
	PvTIP3;2	Phvul.001G080300	254	6	Vacu	Vacu
	PvTIP4;1	Phvul.009G108800	246	6	Vacu	Vacu
	PvTIP5;1	Phvul.011G015200	252	7	Chlo	Plas/Vacu
NIP	PvNIP1;1	Phvul.006G170900	260	6	Plas	Plas
	PvNIP1;2	Phvul.006G171000	273	6	Plas	Plas
	PvNIP1;3	Phvul.002G242200	272	5	Vacu	Plas
	PvNIP1;4	Phvul.002G242300	272	6	Vacu	Plas
	PvNIP2;1	Phvul.008G041100	324	6	Mito	Plas
	PvNIP4;1	Phvul.002G093700	261	6	Vacu	Plas/Vacu
	PvNIP5;1	Phvul.007G084000	299	5	Plas	Plas
	PvNIP6;1	Phvul.005G182400	303	6	Plas	Plas
	PvNIP6;2	Phvul.010G160200	301	6	Plas	Plas
	PvNIP7;1	Phvul.007G178300	297	6	Plas	Plas
SIP	PvSIP1;1	Phvul.001G097000	248	6	Vacu	Plas
	PvSIP1;2	Phvul.004G077700	247	6	Vacu	Plas/Vacu
	PvSIP1;3	Phvul.011G102700	239	5	Plas	Plas/Vacu
	PvSIP2;1	Phvul.001G108800	235	5	Vacu	Plas
XIP	PvXIP1;1	Phvul.011G025700	308	6	Cyto	Plas
	PvXIP1;2	Phvul.011G025800	280	6	Chlo	Plas

[a] Best possible sub-cellular localization of common bean aquaporins based on two different algorithms (Chlo: chloroplast; Cyto: cytosol; Mito: mitochondria; Plas: plasma membrane; Vacu: vacuolar membrane).

Table 2 NPA motifs and ar/R selectivity filter of common bean aquaporins

Sub-family	Name	NPA motifs		ar/R Selectivity filter			
		I	II	H2	H5	LE1	LE2
PIP	PvPIP1;1	NPA	NPA	F	H	T	R
	PvPIP1;2	NPA	NPA	F	H	T	R
	PvPIP1;3	NPA	NPA	F	H	T	R
	PvPIP1;4	NPA	NPA	F	H	T	R
	PvPIP1;5	NPA	NPA	F	H	T	R
	PvPIP2;1	NPA	NPA	F	H	T	R
	PvPIP2;2	NPA	NPA	F	H	T	R
	PvPIP2;3	NPA	NPA	F	H	T	R
	PvPIP2;4	NPA	NPA	F	H	T	R
	PvPIP2;5	NPA	NPA	F	H	T	R
	PvPIP2;6	NPA	NPA	F	H	T	R
	PvPIP2;7	NPA	NPA	F	H	T	R
TIP	PvTIP1;1	NPA	NPA	H	I	A	V
	PvTIP1;2	NPA	NPA	H	I	A	V
	PvTIP1;3	NPA	NPA	H	I	A	V
	PvTIP1;4	NPA	NPA	H	I	A	V
	PvTIP1;5	NPA	NPA	H	I	A	V
	PvTIP1;6	NPA	NPA	N	I	G	A
	PvTIP2;1	NPA	NPA	H	I	G	R
	PvTIP2;2	NPA	NPA	H	I	G	R
	PvTIP2;3	NPA	NPA	H	I	G	R
	PvTIP3;1	NPA	NPA	H	I	A	L
	PvTIP3;2	NPA	NPA	H	I	A	R
	PvTIP4;1	NPA	NPA	H	I	A	R
	PvTIP5;1	NPA	NPA	S	V	G	C
NIP	PvNIP1;1	NPA	NPA	W	V	A	R
	PvNIP1;2	NPA	NPA	W	V	A	R
	PvNIP1;3	NPA	NPV	W	V	A	R
	PvNIP1;4	NPA	NPA	W	V	A	R
	PvNIP2;1	NPA	NPA	A	S	G	R
	PvNIP4;1	NPA	NPA	W	V	A	R
	PvNIP5;1	NPS	NPV	A	I	G	R
	PvNIP6;1	NPA	NPV	T	I	G	R
	PvNIP6;2	NPA	NPV	N	I	S	R
	PvNIP7;1	NPA	NPA	A	V	G	R
SIP	PvSIP1;1	NPT	NPA	F	T	P	N
	PvSIP1;2	NPT	NPA	F	T	P	F
	PvSIP1;3	NPS	NPA	A	T	P	N
	PvSIP2;1	NPL	NPA	Y	K	G	S
XIP	PvXIP1;1	NPI	NPA	V	M	V	R
	PvXIP1;2	SPV	NPA	V	M	V	R

Figure Legend

Fig. 1 Chromosomal localization of the 41 full-length coding aquaporin sequences in the common bean genome. Position determined in bp based on t*he Phaseolus vulga*ris reference sequence.

Fig. 2 Multiple protein alignment of the *P. vulgaris* aquaporins in the region surrounding the 6th helix (H6). The gray rectangle and the black arrow highlight the position and orientation of H6. The number in the upper left side indicates the starting position of the multiple alignmen.

Fig. 32 Bio Neighbor-Joining (BioNJ) phylogenetic tree (1000 bootstraps), based on protein alignment of aquaporins from *Arabidopsis*, *P. trichocarpa* and *P. vulgaris.* Numbers on branches represent bootstrap values. The abbreviation of the species are as follow: At: *Arabidopsis*; Poptr: *P. trichocarpa*; Pv: *P. vulgaris*.

Fig. 43 Gene structure of the 41 full-length coding aquaporin sequences in common bean. UTRs are represented by gray rectangles, CDSs by black round-corner rectangles, and introns by black lines.

Fig. 54 Expression and co-expression profiles of the 41 aquaporins in common bean. a) Expression heatmap in different organs. The expression values of each aquaporin are $\log_2$ transformed. 10d and 19d represent the expression in the corresponding organs at 10 or 19 days after germination. b) Co-expression heatmap based on Pearson's correlation coefficient (r). The value is calculated using the FPKM values in leaves of different domesticated and wild common bean genotypes.

Supplementary material

Supplementary File S1 Multiple sequence alignment of the first NPA region in the putative aquaporins of *P. vulgaris*. Phytozome jbrowse RNA-Sequencing log-coverage link and Phytomine FPKM expression link refers to link of RNA-Sequencing coverage and FPKM expression of Phvul.003G040100.

Supplementary File S2 FPKM expression of PvPIP1;4 in the different organs and genotypes analyzed.

Supplementary File S1 Pairwise protein alignment of the putative tandemly duplicated aquaporins.

Supplementary File S2 Expression heatmap of the 41 aquaporins in different domesticated and wild common bean genotypes. The expression values of each aquaporins are $\log_2$ transformed.

Supplementary File S3 Putative interacting aquaporins with a Pearson's correlation coefficient (r) > 0.9.

Supplementary File S4 Aquaporin normalized expression and expression-fold change under drought stress and in background ESTs. P value, based on 2 x 2 χ^2 tests, is shown.

Supplementary File S5 Summary of BARCBean6K_3 SNPs that have an aquaporin gene locus as the nearest annotated genomic feature. Distance between the markers and the aquaporins, as well as genomic positions of both, are shown.

Supplementary File S6 Distance and genomic positions of the drought QTLs identified by Mukeshimana et al. (2014) and the nearest annotated aquaporins.