

Identification of three aquaporin subgroups from *Blomia tropicalis* by transcriptomics

YING ZHOU^{1*}, LING LI^{2*}, JUN QIAN², HAOYUAN JIA³ and YUBAO CUI³

Departments of ¹Pediatric Laboratory and ²Pediatric Respiratory Medicine, Wuxi Children's Hospital; ³Department of Clinical Laboratory, Wuxi People's Hospital Affiliated to Nanjing Medical University, Wuxi, Jiangsu 214023, P.R. China

Received April 18, 2018; Accepted July 25, 2018

DOI: 10.3892/ijmm.2018.3877

Abstract. Aquaporins (AQPs), or water channel proteins, are highly conserved across species. These transmembrane proteins promote water and solute transport across cell membranes. No AQP-related proteins have been identified in mites to date. The present study used transcriptomics (RNA-sequencing) to identify potential AQPs in the mite species *Blomia tropicalis*. Molecular cloning techniques were then used to obtain the full-length gene sequences encoding these AQP family members, and bioinformatics analyses were used to categorize them based on similarity to AQPs in other species. This approach led to the identification of 5 putative AQP-coding sequences, known as BlotAQP1-5 (GenBank accession numbers: KX655540, KX655541, KX655542, KX655543 and KX655544, respectively), which were indexed into all three subgroups, i.e., AQPs, aquaglyceroporins and superAQPs. To the best of our knowledge, these represent the first known AQPs in any mite species. Further studies are required to investigate their functional roles in water transport and their potential as drug targets.

Introduction

Aquaporins (AQPs), a large family of integral membrane proteins termed the major intrinsic proteins (MIPs), have been identified in nearly all organisms. Indeed, these water channel proteins are conserved in unicellular (archaea, bacteria, yeasts and protozoa) and multicellular (plants, animals and humans) organisms. However, the numbers of AQP isoforms vary significantly between species. For example, two isoforms were detected in *Escherichia coli*, one

being an AQP (AqpZ) and the other an AQP-like sequence (glycerol facilitator GlpF) (1). The yeast *Saccharomyces cerevisiae* possesses two AQPs (ScAqp1 and ScAqp2) and two aquaglyceroporins (YFL054Cp and ScFps1) (2). Since the completion of the human genome project, 13 mammalian AQPs (AQP0-12) with varying degrees of homology have been defined; these are expressed in a number of epithelial and endothelial cells involved in fluid transport, as well as in skin, fat, brain, liver and urinary bladder cells that may not be involved in fluid transport (3). Based on the primary sequences and a phylogenetic framework (rather than function), human AQPs are classified into three major paralogous groups: i) Orthodox or classical AQPs (AQP0, 1, 2, 4, 5, 6 and 8), considered to be specific water channels; ii) aquaglyceroporins (AQP3, 7, 9 and 10), permeable to glycerol and perhaps urea and other small solutes in addition to water; and iii) superAQPs (S-AQPs) (AQP11 and 12), also known as unorthodox S-AQPs or subcellular AQPs, a third subfamily present in animals, but not in plants, fungi or bacteria, and with uncertain permeability (4). AQPs are implicated in numerous physiological processes and in the pathophysiology of a wide range of clinical disorders, including mediating allergic responses (5), encouraging the developing of novel approaches for the treatment of human diseases based on AQP function or dysfunction (6,7).

Two conserved regions of amino acids (NPA motifs) are considered markers for a common origin of water channel proteins (8-10). These highly conserved regions can be used to design degenerate polymerase chain reaction (PCR) primers to enable amplification of an MIP gene fragment. Indeed, this was performed in the insect *Rhodnius prolixus*, in which the full-length cDNA encoding a protein termed *R. prolixus*-MIP was isolated from malpighian tubules and amplified using reverse-transcription PCR (RT-PCR) with degenerate primers to highly conserved regions of the members of the AQP family (11). Similarly, an open reading frame was obtained from a λ gt-11 cDNA library constructed from an adult buffalo fly by screening with degenerate PCR primers designed from highly conserved regions of amino acids found in all members of the AQP gene family (12). Thus, certain information on AQPs in insects has been gained by cloning. Studies of AQPs allow precise biophysical measurements of their functions and exploration of high affinity inhibitors for control of these insects.

Correspondence to: Dr Yubao Cui, Department of Clinical Laboratory, Wuxi People's Hospital Affiliated to Nanjing Medical University, 299 Qingyang Road, Wuxi, Jiangsu 214023, P.R. China
E-mail: ybcui1975@hotmail.com

*Contributed equally

Key words: aquaporin, *Blomia tropicalis*, transcriptomics, RNA-sequencing

Molecular cloning with degenerate PCR primers, however, cannot fully resolve all AQP-coding genes within a given organism due to sequence diversity. By contrast, transcriptomic approaches such as RNA-sequencing (RNA-seq) can reveal the full complement of messenger RNA molecules, and *de novo* assembly of RNA-seq data enables the study of genes without the requirement for a genome sequence (13). The use of transcriptomics approaches opens novel possibilities for advancing our understanding of insect genomes (14). In our recent study, RNA-seq, combined with tandem mass spectrometry, was used to identify and study potential allergens from *Tyroglyphus putrescentiae* (15), a mite species with importance for human allergic disease.

Another domestic mite species, *Blomia tropicalis*, is an important source of indoor allergens associated worldwide with asthma, as well as rhinoconjunctivitis and atopic eczema. *B. tropicalis* mites coexist with *Dermatophagoides pteronyssinus* mites in tropical and subtropical regions; although *D. pteronyssinus* contributes to more allergic disease worldwide, *B. tropicalis* allergens are probably more important clinically than *D. pteronyssinus* allergens in these regions (16,17). To the best of our knowledge, to date, no AQP-related proteins have been identified in mite species. Thus, in the present study, RNA-seq was conducted in *B. tropicalis*, its AQPs were annotated by BLASTX and five AQPs were identified by molecular cloning techniques. To the best of our knowledge, this is the first report of AQPs in any house dust or storage mite species.

Materials and methods

Mite culture. Mites were collected from dust samples from Haikou, Hainan, located in the South China Sea. Each mite isolated under a stereomicroscope was cultivated in a small flask (5 ml) with culture medium in the bottom and filter paper covering the top. After 2 months, the mites were removed, mounted on slides and viewed under a stereomicroscope. Any mites identified as *B. tropicalis* were placed in a large flask (200 ml) for continuous culture. Flasks were placed in an artificial climate incubator (catalog no. RXZ-280D; Jiangnan Instrument Factory, Ningbo, China) at 25±1°C and 70±5% relative humidity. The culture medium was comprised of fish meal and yeast powder.

Isolation of total RNA. Total RNA was isolated from mites using an RNAiso Plus kit (catalog no. 9108Q; Takara Biotechnology Co., Ltd., Dalian, China) according to the manufacturer's protocols. Briefly, isolated cultured mites were rapidly frozen with liquid nitrogen, and then 2 ml RNAiso Plus was added. Mites were then homogenized in a PowerGen 125 Tissue Homogenizer (Thermo Fisher Scientific, Inc., Waltham, MA, USA) starting at 5,000 rpm and gradually increasing to ~20,000 rpm for 30–60 sec at room temperature. Samples were transferred to an Eppendorf tube for RNA isolation. To evaluate the quality of the total RNA, the concentrations, 28S/18S and RNA Integrity Number (RIN) were detected on an Agilent 2100 bioanalyzer with Agilent RNA 6000 nano Reagents Port 1 (Agilent Technologies GmbH, Waldbronn, Germany). Finally, total RNA was dissolved in RNase-free water and stored at -80°C.

Transcriptome library construction, sequencing, *de novo* assembly and annotation. Previously described methods were used to sequence and assemble the transcriptome of *B. tropicalis* (15). Briefly, to create a pooled cDNA library for large-scale sequencing, a SMART™ cDNA Library Construction kit (catalog no. 634901; Clontech Laboratories, Inc., Mountainview, CA, USA) was used. mRNA was isolated using magnetic beads with Oligo(dT). Fragmented mRNA was used as a template for cDNA production. Elution buffer (10 mM Tris ·Cl, pH 8.5) was used for end repair and single nucleotide adenine addition, then fragments were connected with adapters. Suitable fragments were selected as templates for PCR amplification.

The library was sequenced using Illumina HiSeq™ 2000 (Illumina, Inc., San Diego, CA, USA). Image data were transformed by base calling into sequence data and stored in Fastq format. Raw data were cleaned by removing reads with adapters, unknown nucleotides >5% and low-quality reads (base quality ≤10) >20%.

Trinity program v2.2.0 (13) was used to align contigs and generate unigenes. Unigenes were aligned by BLASTX (E-value ≤1.0×10⁻⁵; <https://blast.ncbi.nlm.nih.gov/Blast.cgi>) to databases in the priority order non-redundant (NR), Nucleotide (NT), Swiss-Prot, Kyoto Encyclopedia of Genes and Genomes (KEGG), Clusters of Orthologous Groups (COG) and Gene Ontology (GO) database. Unigenes aligned to a high priority database were not aligned to databases of lower priority. The process ended when all alignments were finished. Proteins with the highest ranks in BLAST results were selected to decide the coding region sequences of unigenes, and then the coding region sequences (CDS) were translated into amino sequences with the standard codon table. Unigenes that could not be aligned to any database were scanned by the program ESTScan (18) producing nucleotide sequence (5'-3') direction and amino sequences of the predicted coding region.

Simple sequence repeats (SSRs) were identified with the program MicroSATellite using unigene as a reference. Single-nucleotide polymorphisms (SNPs) were identified using the Short Oligonucleotide Analysis Package (<http://soap.genomics.org.cn>).

Identification and cDNA cloning of AQPs from *Blomia tropicalis*. During functional annotation by BLASTX in transcriptomic analysis, AQP sequences were selected and clustered to create an Excel table. Using total RNA of *B. tropicalis* as a template, these sequences were then tested by RT-PCR with the Takara PrimeScript™ RT-PCR kit (catalog no. bRR014A) and the Takara Tks Gflex DNA Polymerase kit (catalog no. R060A) (both Takara Biotechnology Co., Ltd.). The RT-PCR products were analyzed by agarose electrophoresis (1.0%), visualized with ImageMaster® VDS (Pharmacia Biotech Inc., San Francisco, CA, USA), recovered with a Takara MiniBEST Agarose Gel DNA Extraction kit Ver.4.0 (catalog no. 9762; Takara Biotechnology Co., Ltd.), and submitted to direct sequencing. Following alignment by BLASTp (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), candidate sequences were obtained. Finally, the full cDNA encoding these AQPs were directly synthesized by RT-PCR from the total RNA of *B. tropicalis* with the primers listed in Table I and using the temperature protocols mentioned in the RT-PCR

Table I. Identification of AQP_s from *Blomia tropicalis*.

AQP _s	GenBank accession number	cDNA length, bp	Deduced no. of amino acids	Deduced molecular weight, Da	Homolog similarity ^a (%)	Forward primers in PCR	Reverse primers in PCR
BlotAQP1	KX655540	843	282	30518.8	ANC28171 (48)	AATGGGTCGCGGATCCATG GTCAATTGGTCAATTA AAATGC	TCAGTGGTGGTGGTGGTGC TCGAGTGGCGCCGCTAGTTCT GGGTGACTCGTACA
BlotAQP2	KX655541	918	305	33164.6	KRX48204 (41)	AATGGGTCGCGGATCCATG AATAATATTGGAAAGAAAT	T:CAAGTGGTGGTGGTGGTGTG CTCGAGTGGCGCCGCTCAATAA CGTGAATGTGGTG
BlotAQP3	KX655542	771	256	28598.7	KFM59126 (51)	AATGGGTCGCGGATCCATG TTACCGCACGATATCATCC CGTACA	TCAGTGGTGGTGGTGGTGGTGC TCGAGTGGCGCCGCTTATTCCA ATTTTCCCTGGGC
BlotAQP4	KX655543	1,377	458	49839.3	XP_001866597 (43)	AATGGGTCGCGGATCCATG ATGGCAACACCAAAATG GGTC	TCAGTGGTGGTGGTGGTGGTGC TCGAGTGGCGCCGCTTAGAATT TGGGATTCACACCA
BlotAQP5	KX655544	825	274	30087.6	KRX92759 (35)	AATGGGTCGCGGATCCATG AAACGTGCTGCAAAAGAG	TCAGTGGTGGTGGTGGTGGTGC TCGAGTGGCGCCGCTCATTGTA CTAATATATTTT

^aBLASTp GenBank accession numbers and similarity are listed. AQP, aquaporin.

section. Subsequent to agarose gel electrophoresis, the recovered PCR products were inserted into expression plasmid pET-28a(+) vector (kit catalog no. N72770; Novagen; Merck KGaA, Darmstadt, Germany) with In-Fusion® HD Cloning kit (catalog no. 639633; Clontech Laboratories, Inc.) for sequence determination.

RT-PCR. RT was performed using the total RNA isolated from mites with Takara PrimeScript™ RT-PCR kit (catalog no. RR014A) in the PCR Thermal Cycler Dice (catalog no. TP600) (both Takara Biotechnology Co., Ltd.). The reaction mix contained 1 µl total RNA, 1 µl Random 6 (20 mM), 1 µl dNTPs (10 mmol/l each) and 5.5 µl RNase-free dH₂O. The reaction mix was placed at 65°C for 5 min, and then on an ice-bath for 2 min. Additionally, 2 µl 5X PrimeScript RT Buffer, 0.25 µl RNase inhibitor (40 U/µl), 0.5 µl PrimeScript RTase and 10 µl of the final reaction mixture were incubated at 30°C for 10 min, 45°C for 30 min and 95°C for 5 min. Next, the RT product was used as the template for PCR in the same thermal cycler with PrimeSTAR HS DNA Polymerase (catalog no. DR010A; Takara Biotechnology Co., Ltd.). The total reaction mixture contained 2 µl RT products, 25 µl 2X Gflex PCR Buffer (Mg²⁺, dNTP plus), 1 µl Tks Gflex DNA Polymerase (1.25 U/µl), 1 µl forward primer (10 µM), 1 µl reverse primer (10 µM) and 20 µl dH₂O. PCR conditions comprised an initial incubation for 1 min at 94°C, followed by 30 cycles of 10 sec at 98°C, 15 sec at 55°C and 1 min at 68°C. Subsequent to a final incubation for 5 min at 68°C, 5 µl of the PCR product was analyzed by agarose electrophoresis (1.0%) and visualized with ImageMaster® VDS (Pharmacia Biotech Inc.). The PCR products were recovered by Takara MiniBEST Agarose Gel DNA Extraction kit Ver.4.0 (catalog no. 9762; Takara Biotechnology Co., Ltd.).

Preparation of vector DNA. The vector pET28a(+) (Novagen; Merck KGaA) was digested with *Bam*HI and *Not*I. Digestion mixture (50 µl) contained 10 µl pET28a(+) (100 ng/µl), 2.5 µl 10X K Buffer, 1.5 µl *Bam*HI (15 U/µl), 1.5 µl *Not*I (15 U/µl), 5 µl 0.1% bovine serum albumin (Sigma-Aldrich; Merck KGaA) and 29.5 µl dH₂O. Digestion was performed at 37°C for 3 h, and then samples were separated by 1% agarose gel electrophoresis. The PCR products were recovered by Takara MiniBEST Agarose Gel DNA Extraction kit Ver.4.0.

Cloning and DNA sequencing. The cDNA fragments coding for AQP were sub-cloned into the expression vector pET28a(+) to create recombinant plasmids using the DNA Ligation kit (catalog no. D6023; Takara Biotechnology Co., Ltd.). *E. coli* JM109 competent cells (catalog no. D9052; Takara Biotechnology Co., Ltd.) were transformed with recombinant plasmids, and positive clones were selected by blue/white screening. Clones were sequenced on an ABI PRISM™ 377XL DNA Sequencer (Applied Biosystems; Thermo Fisher Scientific, Inc.).

Bioinformatics. Amino acid sequences were translated from the nucleotide sequencing results by the ExPASy Translate tool (<http://www.expasy.org/>) and aligned with Clustal Omega (<http://www.ebi.ac.uk/>). A tree was constructed using amino acid sequences with the neighbor-joining method using MEGA 6.0 (<https://www.megasoftware.net/mega6/>). Homology modeling was used to construct a tertiary structure of these

AQPs from *B. tropicalis* one by one. BLASTp search with default parameters was performed against the Protein Data Bank (<https://www.rcsb.org/>) to find suitable templates (high score, low E-value and maximum sequence identity). MODELLER v9.16 (19) was used to predict the tertiary structure of the aforementioned AQPs. Estimating the quality of structural models is a vital step in protein structure construction. PROCHECK (20) checked the stereochemical quality of the constructed structure. ERRAT (21) analyzed statistics of non-bonded interactions between different atom types. VERIFY_3D (22) determined the compatibility of an atomic model (3-dimensional) with its the amino acid sequence (1-dimensional) and compared the results with favorable structures. Superimposition of query and template structure, and visualization of generated models was performed using UCSF Chimera 1.10.2 (23).

Results

Sequencing, assembly and annotation of *B. tropicalis* transcriptome. Total RNA was obtained at a concentration of 508 ng/µl and a RIN value of 7.2, which suggested that the total RNA was of high quality. A cDNA library was constructed from the RNA template, and sequencing runs were performed. Trinity v2.2.0 was used for the *de novo* assembly of 57,492,140 raw reads. Using default assembly parameters, raw reads were screened and trimmed of 5' and 3' adapters, and low quality reads were removed. A total of 55,113,524 total clean reads were obtained with 4,960,217,160 bp. The Q20 value, describing the proportion of nucleotides with a quality value of >20 in reads, was 97.96%, and the GC percentage, describing the proportion of guanine and cytosine nucleotide among total nucleotides, was 40.55%. Overall, this assembly obtained 68,658 unigenes (transcripts) with a total length of 62,406,675 bp and a mean length of 909 bp. A total of 36,645 unigenes were annotated with 33,545 in the non-redundant (NR) database, 12,475 in the nucleotide (NT) database, 30,057 in Swiss-Prot, 26,686 in KEGG, 15,995 in COG and 18,426 in the Gene Ontology (GO) database. In total, 45,632 CDS were generated, including 34,195 CDS by BLASTX protein database searches and 11,437 predicted as aforementioned. The NR database queries revealed *B. tropicalis* sequences to closely match sequences of *Ixodes dammini* (11.7%), *Galendromus occidentalis* (7.5%), *Daphnia pulex* (3.6%), *Crassostrea gigas* (2.6%), *Pediculus humanus corporis* (2.7%) and *Capitella spp.* (2.5%).

Functional annotation performed by searching against the COG database resulted in a classification of 25 categories. The cluster for 'cytoskeleton' constituted the major group (n=5,161; 16.68%), followed by 'nuclear structure' (n=2,593; 8.38%), 'extracellular structures' (n=2,424; 7.83%) and 'defense mechanisms' (n=2,306; 7.45%); only 5 unigenes were allocated to 'RNA processing and modification' (Fig. 1A).

In the GO annotation, the 18,426 unigenes were allocated to one or more GO terms based on sequence similarity. The three main categories of GO annotations were biological process (n=6,016; 38.1%), cellular component (n=4,793; 30.36%) and molecular function (n=4,979; 31.54%) (Fig. 1B). In the category of biological process, the term 'cellular process' was in the highest proportion of annotations, followed by 'metabolic process', 'single-organism process', 'biological regulation' and 'regulation of biological process'. For cellular component,

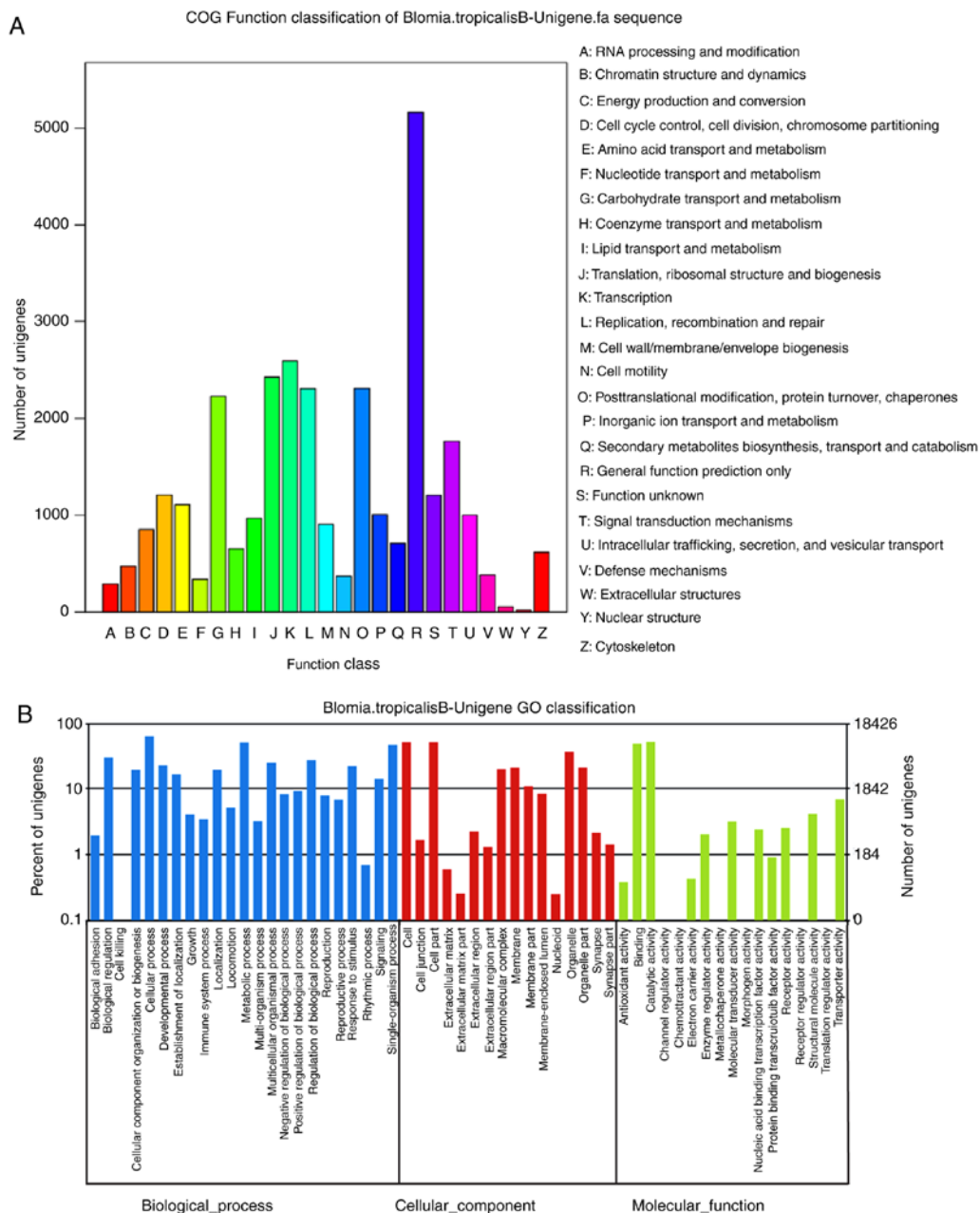


Figure 1. Functional annotation of *Blomia tropicalis* transcriptome. (A) COG function classification of unigenes in *B. tropicalis* transcriptome analysis. COGs of proteins were delineated by comparing protein sequences encoded in complete genomes, representing major phylogenetic lineages. Each COG consists of individual proteins or groups of paralogs from at least 3 lineages and thus corresponds to an ancient conserved domain. (B) GO enrichment analysis of unigenes in *B. tropicalis* transcriptome analysis. COGs, Clusters of Orthologous Groups; GO, Gene ontology.

genes involved in 'cell' and 'cell part' were most represented. In the category of molecular function, the most frequent GO term was 'catalytic activity', followed by 'binding'.

In total, CDS of 34,195 unigenes were generated by BLASTX protein database searches. The ratio of gap lengths to the sizes of unigene CDS were analyzed. The majority of the unigene CDS ($n=24,513$; 71.69%) had $<1,000$ bp, and 8,307 had $\geq 1,000$ bp. The size frequency distributions of these unigene CDS and proteins are depicted in Fig. 2A and B, respectively. A total of 11,437 CDS of unigenes were scanned that could not be aligned to any database by ESTScan. Of these, the majority of the unigene CDS assigned by ESTScan ($n=10,676$; 93.34%) were shorter than 1,000 bp (Fig. 2C); this was also the case for protein sequences obtained from ESTScan (Fig. 2D).

Discovery of microsatellites. SSRs, or DNA microsatellites, can be used for genomic mapping, DNA fingerprinting and marker-assisted selection. To investigate the SSR profile in unigenes of *B. tropicalis*, in the 68,658 investigated unigene sequences, 13,298 SSRs were detected, with 3,896 (29.29%) sequences containing more than one SSR. Six types of SSRs were found, among which the tri-nucleotide repeats motif represented the largest group ($n=10,394$), followed by the di-nucleotide repeats ($n=7,653$), quad-nucleotide repeats ($n=541$), mono-nucleotide repeats ($n=336$), penta-nucleotide repeats ($n=182$) and hexa-nucleotide repeats ($n=87$).

Finally, based on the alignment of short reads to unigenes with the corresponding sequencing quality scores, a total of

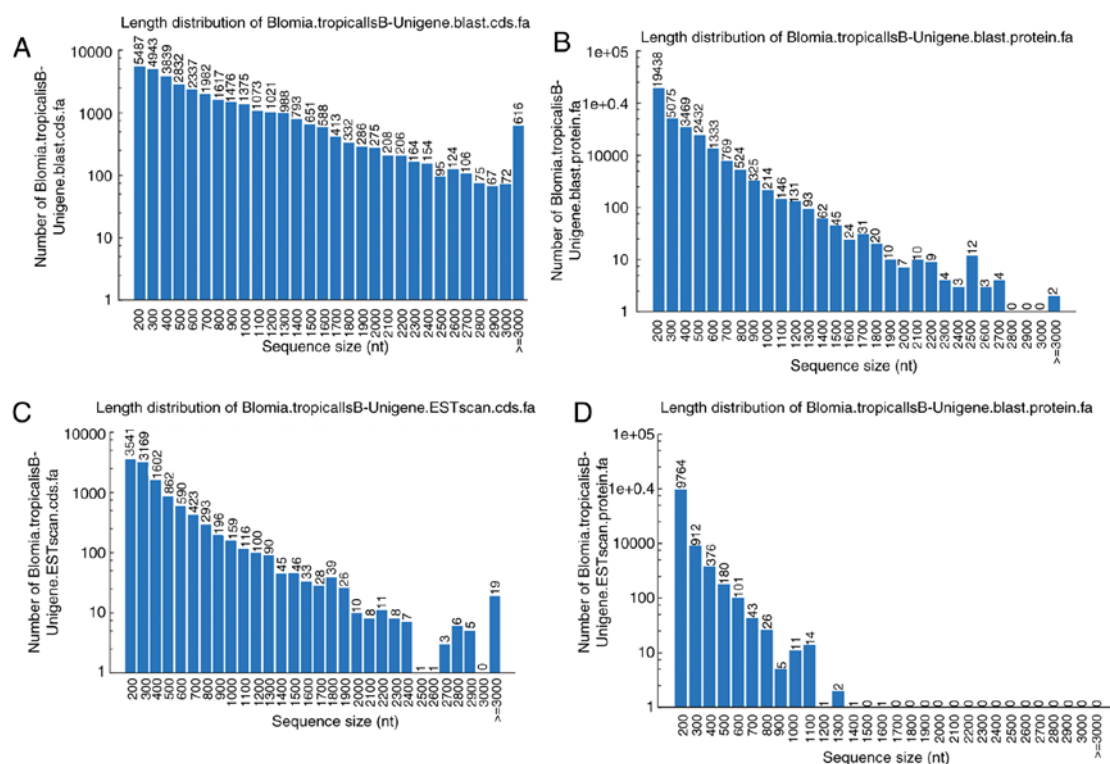


Figure 2. Size frequency distributions of unigene CDS in *B. tropicalis* transcriptome analysis. (A) The length distribution of unigenes CDS by BLAST. (B) Length distribution of unigene protein by BLAST. (C) Length distribution of unigene CDS by ESTscan. (D) Length distribution of unigene protein by ESTscan. CDS, coding region sequences.

78,391 SNPs, both transition and transversion, were identified for *B. tropicalis*. In the total 53,761 transition variants, there were 27,032 'A-G' and 26,729 'C-T'. In the total 24,630 transversion variants, there were 5,981 'A-C', 9,954 'A-T', 3,113 'C-G' and 5,582 'G-T'.

Gene cloning and nucleotide sequencing of AQP from *B. tropicalis*. According to functional annotation in transcriptome analysis, the unigenes coding for AQPs were selected and clustered to create an Excel table. Based on these unigenes sequences, the primers were designed, synthesized and used in RT-PCR for amplification of the sequences from the total RNA of *B. tropicalis*. By agarose electrophoresis analysis, five genes were obtained, which were then recovered and inserted into the vector pET28a(+) to create recombinant plasmids. The recombinant plasmids were transformed into competent cells for subsequent selection of positive clones. Sequencing of the positive clones produced five full-length genes of 843, 918, 771, 1,377 and 825 bp. The five sequences were registered in GenBank under accession numbers KX655540, KX655541, KX655542, KX655543 and KX655544, and termed BlotAQP1, BlotAQP2, BlotAQP3, BlotAQP4 and BlotAQP5, respectively.

Phylogenetic tree, amino acid sequences alignment and homolog modeling. Fig. 3 shows the molecular evolution tree constructed with the aforementioned five amino acid sequences and human AQPs (AQP0-12), from which it can be observed that BlotAQP3 was clustered with hAQP11 and hAQP12, BlotAQP4 was clustered with hAQP1, hAQP2, hAQP4, hAQP5, hAQP6 and hAQP8, and BlotAQP1, BlotAQP2 and BlotAQP5 were clustered with hAQP3, hAQP7 hAQP9 and

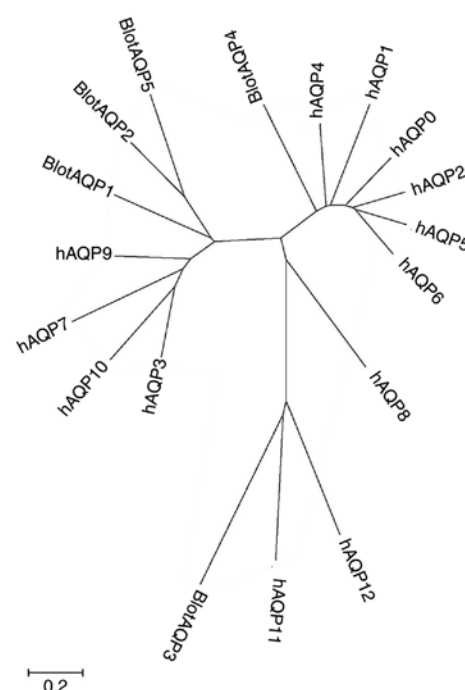


Figure 3. Phylogenetic relationship of AQPs from *Blomia tropicalis* and *Homo sapiens*. Tree constructed using amino acid sequences with neighbor-joining method using MEGA (6.0). Scale bar represents an estimate of the number of amino acid substitutions per site. Accession and database sequences identifiers are as follows: hAQP1 (no. AB451275), hAQP2 (no. AH007817), hAQP3 (no. BT007199), hAQP4 (no. BC022286), hAQP5 (no. AH006636), hAQP6 (no. NM_001652), hAQP7 (no. BC119672), hAQP8 (no. AF067797), hAQP9 (no. AB008775), AQP10 (no. BC069607), AQP11 (no. BC040443), AQP12 (no. AB040748), BlotAQP1 (no. KX655540), BlotAQP2 (no. KX655541), BlotAQP3 (no. KX655542), BlotAQP4 (no. KX655543) and BlotAQP5 (no. KX655544). AQP, aquaporin.

Table II. Evaluation parameters for tertiary structures of AQP.

AQPs	Template (.pdb)	Identity, %	PROCHECK	ERRAT	VERIFY 3D, %
BlotAQP1	3C02	68.5	88.9% core 8.9% allowed 0.4% generously allowed 1.7% disallowed	83.209	69.29
BlotAQP2	3C02	63.8	91.5% core 7.7% allowed 0.8% generously allowed 0.0% disallowed	68.889	70.33
BlotAQP3	3C02	56.7	85.5% core 11.0% allowed 3.1% generously allowed 0.4% disallowed	61.134	65.55
BlotAQP4	2D57	42.3	91.4% core 7.6% allowed 0.8% generously allowed 0.3% disallowed	55.016	67.26
BlotAQP5	3C02	65.0	88.3% core 9.6% allowed 1.7% generously allowed 0.4% disallowed	63.910	66.35

AQP, aquaporin.

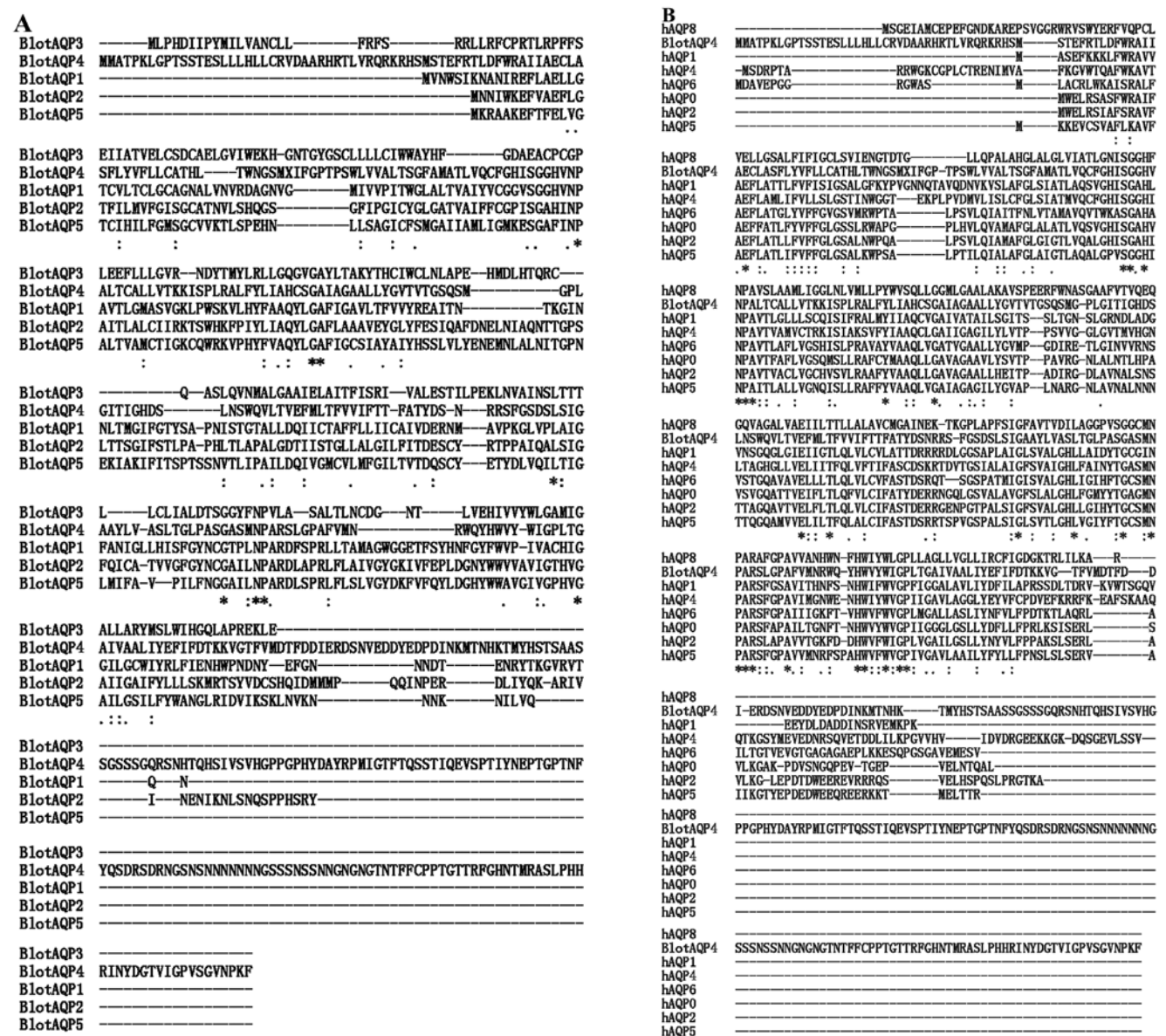
Figure 4. AQP amino acid sequence alignments. Amino acid sequence alignments (A) among AQPs of *Blomia tropicalis*; (B) among BlotAQP4 and hAQP1, hAQP2, hAQP4, hAQP5, hAQP6 and hAQP8. Alignments performed using Clustal omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>). '*' indicates that residues or nucleotides in that column are identical in all sequences in the alignment; ':' indicates conserved substitutions; and '.' indicates semi-conserved substitutions. AQP, aquaporin.



Figure 4. Continued. AQP amino acid sequence alignments. (C) among BlotAQP1, BlotAQP2, BlotAQP5 and hAQP3, hAQP7, hAQP9 and hAQP10; (D) among BlotAQP3, hAQP11 and hAQP12. Alignments performed using Clustal omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>). ‘*’ indicates that residues or nucleotides in that column are identical in all sequences in the alignment; ‘:’ indicates conserved substitutions; and ‘.’ indicates semi-conserved substitutions. AQP, aquaporin.

hAQP10. By Clustal Omega, the five amino acid sequences from *Blomia tropicalis* were aligned with the human orthodox AQPs (AQP0, 1, 2, 4, 5, 6 and 8), aquaglyceroporins and S-AQPs (AQP11 and 12). Alignments are shown in Fig. 4A-D.

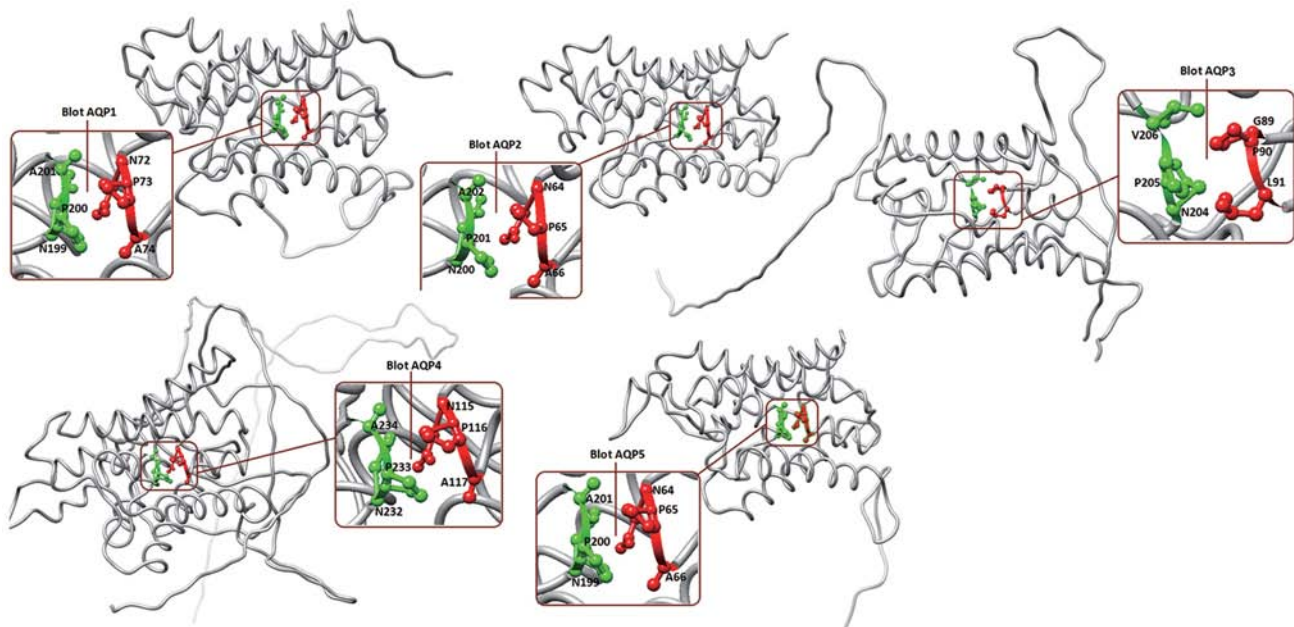


Figure 5. Locations of two conserved NPA motifs on the tertiary structures of *Blomia tropicalis* aquaporins.

By BLASTp, the sequence similarity between BlotAQP4 and *Rattus norvegicus* AQP is 42.3%, whereas the sequence similarities between BlotAQP1, BlotAQP3, BlotAQP5, BlotAQP2 and *Plasmodium falciparum* AQP (PDB ID: 3C02) are 68.5, 56.7 and 63.8%, respectively (Table II). Homolog models of the AQPs identified from *B. tropicalis* were built based on the crystal structures of *Rattus norvegicus* AQP4 (PDB ID: 2D57) for BlotAQP4, and *Plasmodium falciparum* AQP (PDB ID: 3C02) for BlotAQP1, BlotAQP3, BlotAQP5 and BlotAQP2, respectively, by MODELLER v9.16. The predicted crystal structures were then evaluated by PROCHECK in Table II. For each AQP from *B. tropicalis*, 20 homology models were built and the one with the best discrete optimized protein energy score was chosen. The conserved motifs are marked on Fig. 5: these are N72-P73-A74 and N199-P200-A201 for BlotAQP1, N64-P65-A66 and N200-P201-A202 for BlotAQP2, G89-P90-L91 and N204-P205-V206 for BlotAQP3, N115-P116-A117 and N232-P233-A234 for BlotAQP4, and N64-P65-A66 and N199-P200-A201 for BlotAQP5. From the sequences alignment and homolog modeling, these mite AQPs can be divided into three clades: BlotAQP4 clustered with hAQP1, hAQP2, hAQP4, hAQP5, hAQP6 and hAQP8 (AQPs); BlotAQP1, BlotAQP2 and BlotAQP5 clustered with hAQP3, hAQP7, hAQP9 and hAQP10 (aquaglyceroporins); and BlotAQP3 clustered with hAQP11 and hAQP12 (S-AQPs).

Discussion

Mites belong to the taxon Acari, which also includes other acarines such as ticks. Ball *et al* (24) constructed a phylogenetic tree of tick AQPs that branched into two clear groups: One is tick AQP1, which was found in *R. sanguineus*, *R. appendiculatus*, *A. variegatum* and *I. scapularis*; the other is tick AQP2, found only in *I. scapularis* in addition to the original *D. variabilis* source, and considerably different from the tick AQPs.

The present results identified five AQP-like proteins by direct PCR, which is in accordance with the identification of at least four AQPs in the genome of the yellow fever mosquito, *Aedes aegypti* (25). To the best of our knowledge, the present study is the first to suggest the existence of at least five AQP family members in mites. Further, the predicted proteins were divided into three clades (i.e., paralogs), as has been demonstrated for human AQPs. Therefore, it may be concluded that all three subgroups, i.e., AQPs, aquaglyceroporins and S-AQPs, are present in mites.

All AQPs share a relatively conserved overall molecular structure, containing six transmembrane domains with five connecting loops (A-E), and the amino and carboxyl termini are located in the cytoplasm. Two 'NPA' motifs are embedded into the loop connecting the second and third transmembrane domains (loop B), and in the loop connecting the fifth and sixth transmembrane domains (loop E), respectively (9,16,26). In the present study, following alignment of the five sequences from *Blomia tropicalis*, two conserved NPA motifs were identified, except in BlotAQP3, which had GPL and NPV motifs. Furthermore, five other fully conserved residues were found in these mite AQPs, including a conserved 'GA' ~26 residues downstream from the first NPA motif, an 'L' ~23 residues and a 'G' ~4 residues upstream from the second NPA motif, and a 'G' ~38 residues downstream from the second NPA motif.

There are several residues that characterize each of the three clades (i.e., paralogs) and are likely responsible for their specific functional properties. In the first clade of AQPs, BlotAQP4 clustered with hAQP1, hAQP2, hAQP4, hAQP5, hAQP6 and hAQP8. AQP0, AQP1, AQP2, AQP4 and AQP5 have been demonstrated to be permeated by water (27); AQP6 is also permeated by anions such as nitrate (28), and AQP8 perhaps by water and urea (29). By sequence alignment of BlotAQP4 and hAQP1, hAQP2, hAQP4, hAQP5, hAQP6 and hAQP8, three fairly conserved motifs are observed. An AEF box that is located ~60 residues upstream from the first NPA

motif is replaced by AEC in BlotAQP4. The highly conserved 'E', 'T', 'IG', 'L' and 'G' residues located ~50, ~46, ~20, ~11 and ~4 residues, respectively, upstream from the second NPA motif were present in all of these AQPs from mites and humans. A 'HW-W-GPL' motif was also found ~16 residues downstream of the second NPA motif. This sequence alignment supports the hypothesis that the NPA box could be included within a larger, less-stringent consensus 'SG-H-NPAVT' motif, whereas a second NPA box could be part of a 'G-NPAR-GP' motif (30).

In the second clade of AQPs, according to the molecular evolution tree constructed by amino acid sequences coding for AQPs from *B. tropicalis* and humans, BlotAQP1, BlotAQP2 and BlotAQP5 clustered with hAQP3, hAQP7, hAQP9 and hAQP10. Human AQP3, 7, 9 and 10 are the aquaglyceroporins, which transport small neutral solutes such as glycerol and urea, as well as water. Aquaglyceroporins also facilitate the diffusion of charged and non-charged molecules of the metalloids arsenic and antimony, and serve a crucial role in metalloid homeostasis (31). The sequence alignment in the present study satisfied the two highly conserved NPA motifs in BlotAQP1, BlotAQP2 and BlotAQP5. Furthermore, the second NPA box could be included within a larger common consensus 'N-G-NPSRD-PRL' motif. A highly conserved 'AQYLGA' motif (~22 residues downstream from the first NPA motif) was observed in these three proteins. There are certain other residues that characterize these paralogs and are likely to be responsible for their specific functional properties. Although the AEF box only appears in BlotAQP2 and hAQP10, the 'E' is highly conserved.

In the last clade of AQPs, BlotAQP3 was clustered with hAQP11 and hAQP12, 'S-AQPs', with two deviated NPA motifs. The first NPA motif was 'CPC' in BlotAQP3, 'NPC' in hAQP11 and 'NTP' in hAQP12. The second NPA motif could be included within a larger common consensus 'FNPALA' motif. Unexpectedly, a conserved motif 'VYWL', ~20 residues downstream from the second NPA motif, was found in all three of these sequences. The substrate specificities for hAQP11 and hAQP12 remain to be investigated, but water transport activity has been shown for hAQP11 (32), while another study failed to detect any water or glycerol transport activities (33).

Notably, from the *B. tropicalis* transcriptome, eight sequences were predicted that appeared to be water channel proteins, but only five gene fragments were obtained by RT-PCR using the total RNA of *B. tropicalis*. These five full-length cDNA were not expressed in *E. coli* BL21 (DE3) T1R cells. The current study presents these five genes in full length, and use bioinformatics to classify them into three paralogous groups, i.e. AQPs, aquaglyceroporins and S-AQPs. The signature sequences summarized in the present study will be useful for the investigation of mite AQPs in the future. AQP proteins identified in other insects are implicated in excretion (34), and excretions are a major source of allergens from mites (35). Thus, these proteins may represent key mediators of allergic disease in humans. Furthermore, AQPs have been noted as potential drug targets (36); in the case of mite species, the identification of AQPs presents the potential to design or identify drugs that, by targeting AQPs, may enable control of the population.

Acknowledgements

The authors would like to thank Ms. Xiaowei Zhang, Ms. Ying Han and Ms. Liwei Lin (Takara Biotechnology Co., Ltd.) for assisting with the molecular techniques, and Mr. Zhiyuan Huang (LC-Bio Co. Ltd., Hangzhou, China) for assistance with the transcriptomic analyses.

Funding

This study was supported by the National Natural Sciences Foundation of China (grant no. NSFC31272369), the 333 project of Jiangsu Province in 2017, Medical Innovation Team of Jiangsu Province (grant no. CXTDB 2017016) and the Major Program of Wuxi health and Family Planning Commission (grant nos. z201606 and Z201701).

Availability of data and materials

The datasets used or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

YC conceived and designed the experiments. YZ, LL and HJ performed the experiments. JQ and YC performed the bioinformatics analysis, prepared the figures and wrote the paper. All authors reviewed the results and approved the submitted version of the manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

- Calamita G, Bishai WR, Preston GM, Guggino WB and Agre P: Molecular cloning and characterization of AqpZ, a water channel from *Escherichia coli*. *J Biol Chem* 270: 29063-29066, 1995.
- Soveral G, Prista C, Moura TF and Loureiro-Dias MC: Yeast water channels: An overview of orthodox aquaporins. *Biol Cell* 103: 35-54, 2010.
- Verkman AS: More than just water channels: Unexpected cellular roles of aquaporins. *J Cell Sci* 118: 3225, 2005.
- Ishibashi K, Tanaka Y and Morishita Y: The role of mammalian supraaquaporins inside the cell. *Biochim Biophys Acta* 1840: 1507-1512, 2014.
- Ikezoe K, Oga T, Honda T, Hara-Chikuma M, Ma X, Tsuruyama T, Uno K, Fuchikami J, Tanizawa K, Handa T, *et al*: Aquaporin-3 potentiates allergic airway inflammation in ovalbumin-induced murine asthma. *Sci Rep* 6: 25781, 2016.
- Castle NA: Aquaporins as targets for drug discovery. *Drug Discov Today* 10: 485-493, 2005.
- Ishibashi K, Hara S and Kondo S: Aquaporin water channels in mammals. *Clin Exp Nephrol* 13: 107-117, 2009.
- Gonen T and Walz T: The structure of aquaporins. *Q Rev Biophys* 39: 361-396, 2006.

9. Zhao CX, Shao HB and Chu LY: Aquaporin structure-function relationships: water flow through plant living cells. *Colloids Surf B Biointerfaces* 62: 163-172, 2008.
10. Benga G: Foreword to the special issue on water channel proteins (aquaporins and relatives) in health and disease: 25 years after the discovery of the first water channel protein, later called aquaporin 1. *Mol Aspects Med* 33: 511-513, 2012.
11. Echevarría M, Ramírez-Lorca R, Hernández CS, Gutiérrez A, Méndez-Ferrer S, González E, Toledo-Aral JJ, Ilundáin AA and Whittombury G: Identification of a new water channel (Rp-MIP) in the Malpighian tubules of the insect *Rhodnius prolixus*. *Pflügers Arch* 442: 27-34, 2001.
12. Elvin CM, Bunch R, Liyou NE, Pearson RD, Gough J and Drinkwater RD: Molecular cloning and expression in *Escherichia coli* of an aquaporin-like gene from adult buffalo fly (*Haematobia irritans exigua*). *Insect Mol Biol* 8: 369-380, 2010.
13. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, *et al*: De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8: 1494-1512, 2013.
14. Zhang QL and Yuan ML: Progress in insect transcriptomics based on the next-generation sequencing technique. *Acta Entomologica Sinica* 56: 1489-1508, 2013.
15. Cui Y, Yu L, Teng F, Zhang C, Wang N, Yang L and Zhou Y: Transcriptomic/proteomic identification of allergens in the mite *Tyrophagus putrescentiae*. *Allergy* 71: 1635-1639, 2016.
16. Arlian LG, Vyszynski-Moher DL and Fernandez-Caldas E: Allergenicity of the mite, *Blomia tropicalis*. *J Allergy Clin Immunol* 91: 1042-1050, 1993.
17. Fernández-Caldas E: Allergenicity of *Blomia tropicalis*. *J Investig Allergol Clin Immunol* 7: 402, 1997.
18. Iseli C, Jongeneel CV and Bucher P: ESTScan: A program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proc Int Conf Intell Syst Mol Biol* 99: 138-148, 1998.
19. Eswar N, Eramian D, Webb B, Shen MY and Sali A: Protein structure modeling with MODELLER. *Methods Mol Biol* 426: 145-159, 2008.
20. Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R and Thornton JM: AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 8: 477-486, 1996.
21. Colovos C and Yeates TO: Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Sci* 2: 1511-1519, 1993.
22. Bowie JU, Luthy R and Eisenberg D: A method to identify protein sequences that fold into a known three-dimensional structure. *Science* 253: 164-170, 1991.
23. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC and Ferrin TE: UCSF Chimera-A visualization system for exploratory research and analysis. *J Comput Chem* 25, 2010.
24. Ball A, Campbell EM, Jacob J, Hoppler S and Bowman AS: Identification, functional characterization and expression patterns of a water-specific aquaporin in the brown dog tick, *Rhipicephalus sanguineus*. *Insect Biochem Mol Biol* 39: 105-112, 2009.
25. Duchesne L, Hubert JF, Verbavatz JM, Thomas D and Pietrantoni PV: Mosquito (*Aedes aegypti*) aquaporin, present in tracheolar cells, transports water, not glycerol, and forms orthogonal arrays in *Xenopus* oocyte membranes. *Febs J* 270: 422-429, 2003.
26. Benga G: On the definition, nomenclature and classification of water channel proteins (aquaporins and relatives). *Mol Aspects Med* 33: 514-517, 2012.
27. Agre P, Sasaki S and Chrispeels MJ: Aquaporins: A family of water channel proteins. *Am J Physiol* 265: F461, 1993.
28. Yasui M, Kwon TH, Knepper MA, Nielsen S and Agre P: Aquaporin-6: An intracellular vesicle water channel protein in renal epithelia. *Proc Natl Acad Sci USA* 96: 5808-5813, 1999.
29. Ishibashi K, Kuwahara M, Gu Y, Kageyama Y, Tohsaka A, Suzuki F, Marumo F and Sasaki S: Cloning and functional expression of a new water channel abundantly expressed in the testis permeable to water, glycerol, and urea. *J Biol Chem* 272: 20782-20786, 1997.
30. Zardoya R and Villalba S: A phylogenetic framework for the aquaporin family in eukaryotes. *J Mol Evol* 52: 391-404, 2001.
31. Bienert GP, Schussler MD and Jahn TP: Metalloids: Essential, beneficial or toxic? Major intrinsic proteins sort it out. *Trends Biochem Sci* 33: 20-26, 2008.
32. Yakata K, Hiroaki Y, Ishibashi K, Sohara E, Sasaki S, Mitsuoka K and Fujiyoshi Y: Aquaporin-11 containing a divergent NPA motif has normal water channel activity. *Biochim Biophys Acta* 1768: 688-693, 2007.
33. Gorelick DA, Praetorius J, Tsunenari T, Nielsen S and Agre P: Aquaporin-11: A channel protein lacking apparent transport function expressed in brain. *BMC Biochem* 7: 14, 2006.
34. Spring JH, Robichaux SR and Hamlin JA: The role of aquaporins in excretion in insects. *J Exp Biol* 212: 358-362, 2009.
35. Platts-Mills TA, Vervloet D, Thomas WR, Aalberse RC and Chapman MD: Indoor allergens and asthma: Report of the third international workshop. *J Allergy Clin Immunol* 100: S2-S24, 1997.
36. Frigeri A, Nicchia GP and Svelto M: Aquaporins as targets for drug discovery. *Curr Pharm Des* 13: 2421-2427, 2007.