SPANDIDOS PUBLICATIONS

MOLECULAR MEDICINE REPORTS 16: 4903-4908, 2017

Phylogenetic and specific sequence analysis of four paralogs in insect Aquaporins

WEI XIA, PANWEN ZHAO, ZHONGQUAN YI and YUBAO CUI

Department of Central Laboratory, The Third People's Hospital of Yancheng, Affiliated Yancheng Hospital, School of Medicine, Southeast University, Yancheng, Jiangsu 224000, P.R. China

Received March 17, 2017; Accepted July 27, 2017

DOI: 10.3892/mmr.2017.7148

Abstract. Aquaporins (AQP) are proteins that form channels to facilitate the movement of water across cell membranes in plants, bacteria and animals. Insect AQPs are indispensable for cellular water management under stress, including dehydration and cold. To better understand the biological significance of molecular evolution of gene sequences, followed by structural and functional specialization, the present study used ClustalX2.1, MEGA7.0, Jalview and Mesquite software to build an insect AQP phylogenetic tree and visualize the evolutionary associations among insect AQPs. It was demonstrated that 45 AQPs were classified as four major paralogs with each amino acid sequence containing two conserved NPA (Asp-Pro-Ala) motifs located in the center and C-terminal domains, and other residues conserved within the paralogous groups, however not among them. All these differences in amino acid content may affect the structure, function and classification of the AQPs. The findings provide a basis for further study to understand insect AQPs through sequence comparison, structure and predicted function.

Introduction

Aquaporins (AQPs) are integral membrane proteins belonging to the larger family of major intrinsic proteins (MIP) that facilitate rapid transport of water across cell membranes (1). Some AQPs, the so-called aquaglyceroporins, also transport other small uncharged solutes across the membrane such as glycerol, urea, ammonia and CO_2 (2). AQPs are expressed widely in animals, plants, yeast, insects, amphibians, and bacteria (3). Denker *et al* found that a 28-kD hydrophobic transmembrane protein, known as channel-forming integral membrane protein, or CHIP28, was formed during the isolation and purification of the Rh polypeptide from erythrocyte membranes (4). Through the determination of the protein's activation energy and permeability coefficient, subsequent inhibitor sensitivity studies confirmed that CHIP28 specifically transports water across cell membranes.

AQPs comprise six highly hydrophobic transmembrane α-helices with the amino and the carboxyl termini located on the cytoplasmic surface of the membrane (1,5). The sequence of amino and carboxyl halves show high similarity to each other in what appears to be a tandem repeat. There are also five inter-helical loop regions (A-E) that form the extracellular and cytoplasmic vestibules. Loops B and E are hydrophobic loops that contain the most highly conserved sequence, an NPA (Asp-Pro-Ala) motif (6), which overlaps the center of the lipid bilayer of the membrane forming a 3-D 'hourglass' structure where the water molecules pass through. Different AQPs have variations in peptide sequence, which permits the different size of the pore between AQPs. The resultant pore size affects what molecules are able to flow through the pore, with small pore sizes only allowing small molecules like water to flow through the pore (7).

To date, more than 200 AQPs have been identified in different species (8,9). Mammalian AQPs 0-12 have been identified (10) in a variety of tissues including brain, kidneys, eyes, the gastrointestinal system and pulmonary system (11). Based on their structure, function, cellular location, and permeability, AQPs can be divided into three categories: orthodox AQPs (AQPs 0, 1, 2, 4, 5, 6, and 10) with exclusive water permeability, aquaglyceroporins (AQPs 3, 7, and 9) permeable to water and glycerol, and AQPs with undefined function (AQPs 8, 11, and 12) (12,13). The study of ten AQPs (AQPs 0-9) found in mammals has been relatively thorough and their functions have been identified through research related to human disease. In contrast, AQPs 10-12 have been less studied and their role in human disease remains poorly understood.

Insects have evolved to inhabit a multitude of ecological niches (14). The physiological adaptations diversity have been employed through their evolutionary process enabling insects to exploit a wide range of food sources and survive a multitude of hostile environments successfully (15,16). Among insects, only seven AQPs have been functionally expressed, and four of these were found to either directly or indirectly function in excretion (17). To define the evolutionary relationships among insect AQPs, we undertook a detailed analysis of AQP amino

Correspondence to: Dr Yubao Cui, Department of Central Laboratory, The Third People's Hospital of Yancheng, Affiliated Yancheng Hospital, School of Medicine, Southeast University, 299 Jiefangnan Road, Yancheng, Jiangsu 224000, P.R. China
E-mail: ybcui1975@hotmail.com

Key words: aquaporin, insects, phylogenetic tree

acid sequences to construct a phylogenetic tree to serve as the basis for ongoing evolutionary studies.

Materials and methods

The data of amino acid sequence. Protein databases at the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov) were searched using 'major intrinsic protein,' 'insect,' and 'aquaporin' as text queries. A total of 251 amino acid sequences were. Of these, 206 sequences were excluded from further analyses because they were short partial sequences or corresponded to putative proteins, or were redundant sequences that contained point mutations with respect to a sequence already included in the analyses. The 45 AQP proteins included in the phylogenetic analyses are listed in Table I.

Redundant sequence data were excluded using a sequence identity cut-off of 100% (http://weizhong-lab.ucsd.edu/cdhit_suite/cgi-bin/). Sequence data were aligned using the default options of ClustalX (18) and the result of multiple alignment was refined by eye. From alignment, the gap results were treated as missing data. Through the phylogenetic analyses, ambiguous alignments with highly variable regions were excluded. NJ (Neighbor-joining) analysis (19) of the amino acid alignment was based on the current distance matrix calculate the matrix Q. Robustness of the NJ tree was tested by bootstrap analyses (as implemented in MEGA7.0) (20) with 1,000 pseudo-replications.

Four major groups produced by the NJ analysis were studied further. Every group was analyzed separately. The amino acid sequences of each group were aligned using ClustalX. A consensus sequence for each group was inferred using JalView (21). Robustness of the phylogenetic results was tested by bootstrap analyses with 1,000 pseudoreplications (as implemented in MEGA7.0) and the quartet puzzling method (with 1,000 puzzling steps).

Estimated the conserved regions of each group for the water channel protein were inferred using the rates variation option in MEGA7.0, which includes Maximum Likelihood analysis and NJ tree.

Results and Discussion

Phylogenetic relationships among insect AQPs. Some insects are able to live in extreme environments with little to no water or periods of drought conditions. On the other hand, some of the insects that feed on blood or plant juices are able to take on too much fluid, which is associated with a particular functional water channel protein (19). An NJ analysis of 446 positions within the AQP protein were analyzed to produce an unrooted phylogenetic tree showing that insect members of the AQP family can be classified into four major groups (Fig. 1).

The first group of insect AQP includes aquaglyceroporin, Aea AQP and AQPcic, which have homology with the AQPs found in *Belgica antarctica*, *T. septentrionalis*, *A. echinatior*, and *Culex quinquefasciatus*. This group is referred as Rpips for *Belgica antarctica* protein. The Rpip group was found to be functionally characterized in the filter chamber of *Cicadella viridis* (22). At least four AQPs in the Rpip group

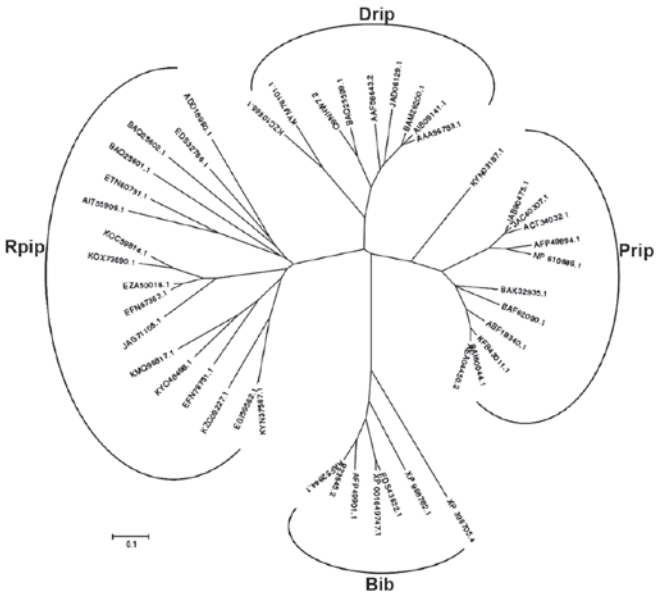

Figure 1. Homologous relationships of the insects AQPs family. Representative unrooted phylogenetic tree of 45 full-length insects AQPs. Tree constructed by MEGA7.0. The NJ was used to estimate the phylogeny of amino acid sequences; fifty percent majority-rule bootstrap NJ was utilized (1,000 replications). Based on the NJ tree, the amino acid sequences were classified into four major groups: Rpips, Drips, Prips, and Bibs. The scale bar indicates the average number of amino acid substitutions per site. The scale bar represents 0.1 amino acid substitutions per site. AQP, aquaporin; NJ, neighbor-joining.

are obvious in the genome of *Aedes aegypti*, the yellow fever mosquito, but only one has been functionally characterized. The result of AQP expressed in *Xenopus* oocytes showed that AQPcic had a higher water permeability than human AQP1 and was reversibly inhibited by Hg^{2+}. Furthermore, the water permeability of Aea AQP is even greater degree than AQPcic and is reversibly inhibited by $HgCl_2$ when expressed in *Xenopus* oocytes (23).

The second group of insect AQPs include the characterized *D. melanogaster* protein known as Drosophila integral protein, or Drip, and is therefore referred to as the Drip group (24). *D. melanogaster* Drip (Dm Drip) is most similar to the mammalian AQP4 isoform, which has been functionally expressed in *Xenopus* oocytes and acts as a water-specific AQP (25). As well as having representatives from each insect genome, this family of AQPs also contains AeaAQP from the *Dufourea novaeangliae* and AngAQP from *Atta colombica*. Currently, the Drip group appears to be restrained to insects, although it is conceivable that these highly unusual Drip-like AQPs may be identified in other invertebrates in the future. The expression of Drip mRNA in Xenopus oocytes and yeast secretory vesicles shows that the DRIP protein channel has a very high permeability to water (26), implicating its role in the regulation of water permeability. So far, only a small part of the water channel protein function has been verified.

The third group of insect AQPs includes Aea AQP and Ang AQP and contains a representative from the insect genome research. Most of the AQP proteins in this group have a high degree of water-selectivity. This group was referred to as the Prips for *P. rufa* integral protein (27). No other homologous proteins have yet been characterized. This is the least

Table I. The AQP family of proteins phylogenetically analyzed in this study.

Protein	Species	Accession no.
Rpip		
Aquaglyceroporin	*Anopheles gambiae*	AIT55906.1
Rpip2 aquaporin	*Belgica antarctica*	BAQ25601.1
Rpip1 AQP transcript variant A	*Belgica antarctica*	BAQ25602.1
Aquaporin -like protein	*Lasius niger*	KMQ96817.1
Aquaporin AQPcic	*Trachymyrmex Sseptentrionalis*	KYN37567.1
Aquaporin AQPcic	*Harpegnathos saltator*	EFN76751.1
Aquaporin AQPAe.a	*Trachymyrmex zeteki*	KYQ48498.1
Aquaporin AQPAe.a	*Fopius arisanus*	JAG71105.1
Aquaporin AQPAe.a	*Camponotus floridanus*	EFN67363.1
Aquaporin AQPAe.a	*Melipona quadrifasciata*	KOX73690.1
Aquaporin AQPAe.a	*Cerapachys biroi*	EZA50016.1
Aquaporin AQPAe.a	*Habropoda laboriosa*	KOC59814.1
Aquaporin AQPAe.a	*Acromyrmex echinatior*	EGI59562.1
Aquaporin	*Culex quinquefasciatus*	EDS32766.1
Aquaporin	*Glossina morsitans morsitans*	ADD18960.1
Aquaporin	*Anopheles darlingi*	ETN60731.1
Aquaporin	*Dufourea novaeangliae*	KZC09227.1
Drip		
Aquaporin	*Musca domestica*	AIB09141.1
Aquaporin	*Phormia regina*	BAM26200.1
Aquaporin	*Bactrocera cucurbitae*	JAD06129.1
Aquaporin AQPAe.a	*Dufourea novaeangliae*	KZC10666.1
Aquaporin AQPAn.G	*Atta colombica*	KYM76101.1
Drip1 aquaporin	*Belgica antarctica*	BAQ25599.1
Drip, isoform B	*Drosophila melanogaster*	AAF58643.2
RecName: Full=Aquaporin AQPAe.a	*Aedes aegypti*	Q9NHW7.2
Water channel	*Haematobia irritans exigua*	AAA96783.1
Prip		
Aquaporin water channel isoform A	*Anopheles gambiae*	AEA04450.2
Aquaporin AQPAn.G	*Cyphomyrmex costatus*	KYN03187.1
Aquaporin AQPAe.a	*Ceratitis capitata*	JAB90475.1
Aquaporin AQPAe.a	*Bactrocera dorsalis*	JAC40307.1
Aquaporin 2a	*Glossina morsitans morsitans*	AFP49894.1
Aquaporin 2	*Aedes aegypti*	ABF18340.1
Aquaporin 1	*Anopheles gambiae*	BAI60044.1
Aquaporin 1	*Anopheles sinensis*	KFB43011.1
Aquaporin	*Eurosta solidaginis*	ACT34032.1
Aquaporin	*Polypedilum vanderplanki*	BAF62090.1
Aquaporin water channel isoform B	*Drosophila melanogaster*	NP_610686.1
Aquaporin	*Belgica antarctica*	BAK32935.1
Bib		
Neurogenic protein big brain	*Apis mellifera*	XP_396705.4
Neurogenic protein big brain	*Musca domestica*	P23645.2
Neurogenic protein big brain	*Tribolium castaneum*	XP_968782.1
AAEL004741-PA	*Aedes aegypti*	XP_001649747.1
Big brain, isoform A	*Drosophila melanogaster*	AAF52844.1
Big brain	*Glossina morsitans morsitans*	AFP49901.1
Aquaporin transporter	*Culex quinquefasciatus*	EDS43622.1

AQP, aquaporin.

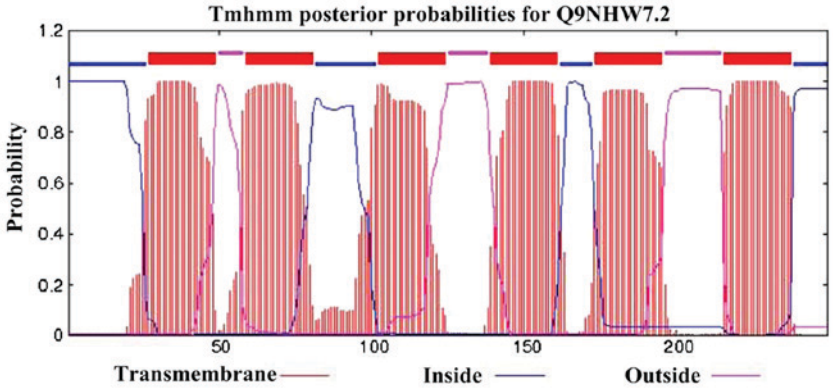

Figure 2. The secondary structure diagram of *Aedes aegypti* AQP (Q9NHW7.2). AQP, aquaporin.

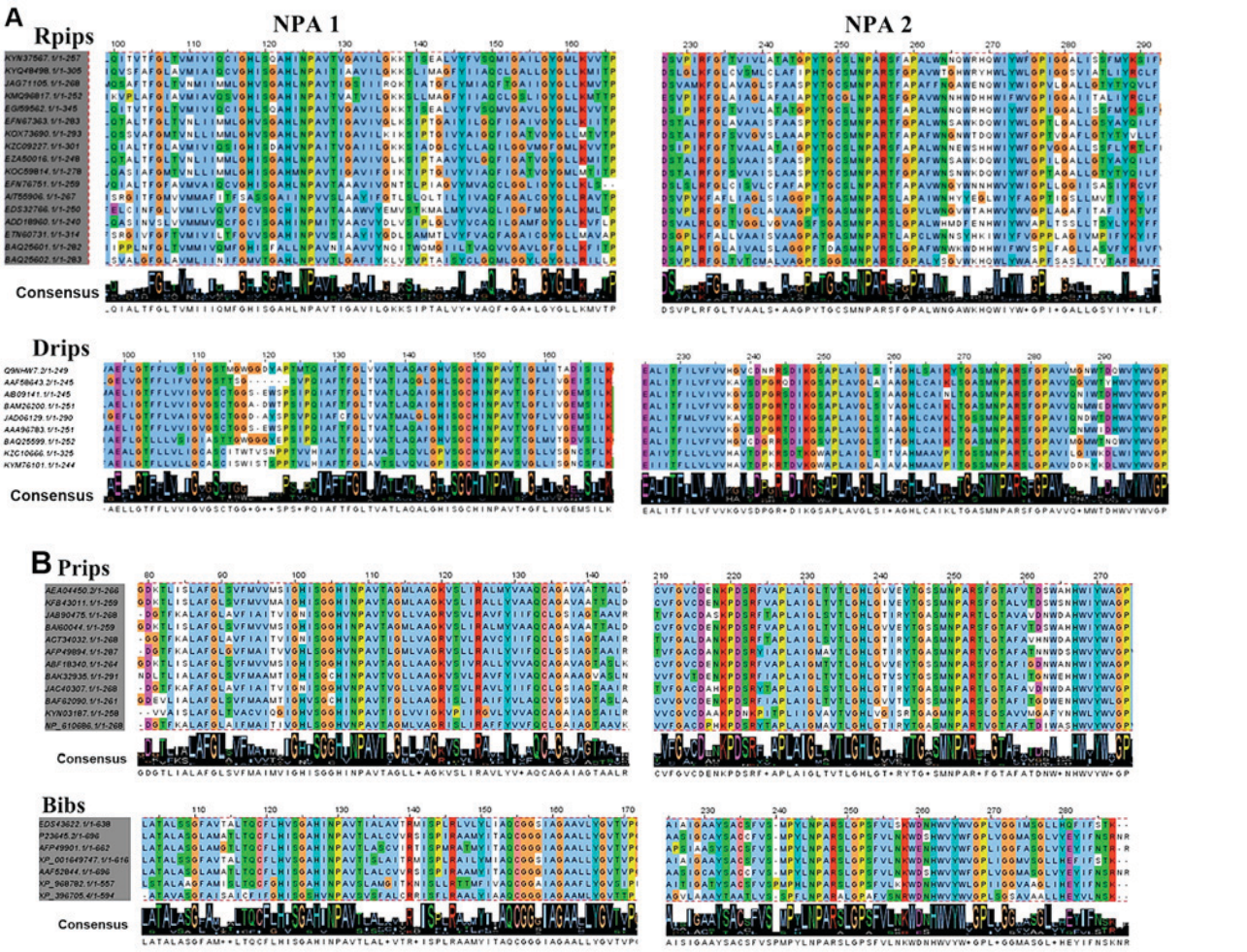

Figure 3. Conserved residues for each paralogous group were aligned using ClustalW. The two NPA motif alignments between each group were assessed by JalView 2.1. The two NPA motif alignments sequences of (A) Rpips and Drips and (B) Prips and Bibs groups.

characterized family of the insect AQPs and deserves further investigation (23).

The fourth group of insect AQPs contains the *D. melanogaster* AQP Big Brain (Dm Bib) (28), as well as Bib-like homologues from each of the other insect species for which complete genome data are available. As the function of the Dm Bib genome has been characterized, this insect AQP-like family was referred to as Bibs appropriately. When expressed in *Xenopus* oocytes, Dm Bib is the only characterized Bib to be found to date and functions as a non-selective cation channel; its permeability is apparently regulated by tyrosine-kinase activity (29). In addition, compared to all other known AQPs, Dm Bib has an extended C-terminal tail with a high number of putative phosphorylation sites. An all Bib homologues alignment reveals four stretches of the tail with high conservation across all insect species examined. Within these regions are five fully conserved and two semi-conserved tyrosine residues. Dm Bib plays an important role in determining *Drosophila* neural

fate. Bibs sequences found across other insects play a similar functional role, and one of the conserved or semi-conserved tyrosine residues confers functional regulation in the tail (24).

Molecular features of insect AQP proteins. All AQP proteins show relatively conserved motifs and overall molecular structure along the multiple alignment. The first motif contains six membrane-spanning segments with five connecting loops connecting the inside and outside of the channel (for example, accession Q9NHW7.2 of *Aedes aegypti* AQP is shown in Fig. 2). Two motifs are NPA (Asp-Pro-Ala) signature motifs (30). These two NPA motifs are the most commonly conserved feature among AQPs and likely play an important role in the structure and function of water channel proteins. A SG-H-NPAVT-G and TG-MNPAR-G motif that is highly conserved in the Rpips clade is replaced by SG-H-NPAV-IS and P-LNPAR-GP motifs and in the Drips clade, the SGCH-NPAV-G and SM-NPAR-GP motifs in the Prips clade, and the V-NPAV-Q and S-NPAR-P motifs in the Bibs clade. Other consensus residues among AQPs are presented in Fig. 3. Most of them are localized in the six transmembrane domains and are likely involved in the selectivity of the channel. Because the full-length sequences of insect AQPs are about 300 bp, special constraints prevent us from depicting the full multiple alignment results in the text. Instead, we chose the most important conserved region, which is the NPA area, and displayed the two conserved region sequence comparison results of insect AQP. We can also identify the homology and evolutionary relationship between them. The different specific conserved sequence has already been responsible for the functional properties (19). Despite many advances in understanding mammalian AQPs since Peter Agre's initial discoveries, not much has been done to understand the evolutionary relationships among insect AQPs. We undertook a detailed analysis of AQP amino acid sequences to construct a phylogenetic tree that could serve as the basis for ongoing evolutionary studies. The evolutionary relationships presented here among AQPs assist with the correct identification of AQP family members moving toward further functional investigation. The insect AQP phylogenetic tree will allow for the potential manipulation of insect evolution, such as introducing water resistance and drought-resistant properties. During the development of mammalian follicles, the formation and maturation of follicles, the transport of follicular fluid, the formation of antral cavities, and the atresia of follicles are closely related to the regulation and expression of AQPs.

In summary, the evolutionary relationships presented here among AQPs assist with the correct identification of AQP family members toward the goal of further functional investigation (31). The insect AQP phylogenetic tree presented here will allow for the potential manipulation of insect evolution, such as introducing water resistance and drought-resistant properties (32). In addition, this study provides a basis for further research on the structure and function of insect sequences (33).

References

1. Preston GM, Jung JS, Guggino WB and Agre P: Membrane topology of aquaporin CHIP. Analysis of functional epitope-scanning mutants by vectorial proteolysis. J Biol Chem 269: 1668-1673, 1994.
2. Kong Y and Ma J: Dynamic mechanisms of the membrane water channel aquaporin-1 (AQP1). Proc Natl Acad Sci USA 98: 14345-14359, 2001.
3. Kwon TH, Nielsen J, Møller HB, Fenton RA, Nielsen S and Frøkiaer J: Aquaporins in the kidney. Handb Exp Pharmacol 190: 95-132, 2009.
4. Denker BM, Smith BL, Kuhajda FP and Agre P: Identification, purification, and partial characterization of a novel Mr 28,000 integral membrane protein from erythrocytes and renal tubules. J Biol Chem 263: 15634-15642, 1988.
5. Preston GM, Jung JS, Guggino WB and Agre P: The mercury-sensitive residue at cysteine 189 in the CHIP28 water channel. J Biol Chem 268: 17-20, 1993.
6. Preston GM and Agre P: Isolation of the cDNA for erythrocyte integral membrane protein of 28 kilodaltons: Member of an ancient channel family. P Natl Acad Sci USA 88: 11110-11114, 1991.
7. Jung JS, Preston GM, Smith BL, Guggino WB and Agre P: Molecular structure of the water channel through aquaporin CHIP. The hourglass model. J Biol Chem 269: 14648-14654, 1994.
8. Froger A, Tallur B, Thomas D and Delamarche C: Prediction of functional residues in water channels and related proteins. Protein Sci 7: 1458-1468, 1998.
9. Heymann JB and Engel A: Aquaporins: Phylogeny, structure, and physiology of water channels. News Physiol Sci 14: 187-193, 1999.
10. King LS, Kozono D and Agre P: From structure to disease: The evolving tale of aquaporin biology. Nat Rev Mol Cell Biol 5: 687-698, 2004.
11. Takata K, Matsuzaki T, Tajika Y and Ablimit A: Aquaporin water channels in the kidney. Acta Histochemica Et Cytochemica Official J Japan Soc Histochemistry Cytochemistry 38: 199-207, 2005.
12. Mandal G, Orta J, Sharma M and Mukhopadhyay R: Trypanosomatid aquaporins: Roles in physiology and drug response. Diseases 2: 3-23, 2014.
13. Gourbal B, Sonuc NH, Bhattacharjee H, Legare D, Sundar S, Ouellette M, Rosen BP and Mukhopadhyay R: Drug uptake and modulation of drug resistance in *Leishmania* by an aquaglyceroporin. J Biol Chem 279: 31010-31017, 2004.
14. Anyon MJ, Orchard MJ, Buzza DM, Humphries S and Kohonen MM: Effect of particulate contamination on adhesive ability and repellence in two species of ant (Hymenoptera; Formicidae). J Exp Biol 215: 605-616, 2012.
15. Henry M, Gasco L, Piccolo G and Fountoulaki E: Review on the use of insects in the diet of farmed fish: Past and future. Animal Feed Sci Technol 203: 1-22, 2015.
16. Wagner DL and Van Driesche RG: Threats posed to rare or endangered insects by invasions of nonnative species. Annu Rev Entomol 55: 547-568, 2010.
17. Costa MJ, Balasekaran G, Vilas-Boas JP and Barbosa TM: Physiological adaptations to training in competitive swimming: A systematic review. J Hum Kinet 49: 179-194, 2015.
18. Thompson JD, Higgins DG and Gibson TJ: CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673-4680, 1994.
19. Saitou N and Nei M: The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406-425, 1987.
20. Tamura K, Peterson D, Peterson N, Stecher G, Nei M and Kumar S: MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance and maximum parsimony methods. Mol Biol Evol 28: 2731-2739, 2011.
21. Clamp M, Cuff J, Searle SM and Barton GJ: The Jalview Java alignment editor. Bioinformatics 20: 426-427, 2004.
22. Kambara K, Takematsu Y, Azuma M and Kobayashi J: cDNA cloning of aquaporin gene expressed in the digestive tract of the Formosan subterranean termite, *Coptotermes formosanus* Shiraki (Isoptera; Rhinotermitidae). Appl Entomol Zool 44: 315-321, 2009.
23. Le Cahérec F, Deschamps S, Delamarche C, Pellerin I, Bonnec G, Guillam MT, Thomas D, Gouranton J and Hubert JF: Molecular cloning and characterization of an insect aquaporin functional comparison with aquaporin 1. Eur J Biochem 241: 707-715, 1996.
24. Campbell EM, Ball A, Hoppler S and Bowman AS: Invertebrate aquaporins: A review. J Comp Physiol B 178: 935-955, 2008.

25. Terhzaz S, O'Connell FC, Pollock VP, Kean L, Davies SA, Veenstra JA and Dow JA: Isolation and characterization of a leucokinin-like peptide of *Drosophila* melanogaster. J Exp Biol 202: 3667-3676, 1999.
26. Kaufmann N, Mathai JC, Hill WG, Dow JA, Zeidel ML and Brodsky JL: Developmental expression and biophysical characterization of a *Drosophila* melanogaster aquaporin. Am J Physiol Cell Physiol 289: C397-C407, 2005.
27. Kikawada T, Saito A, Kanamori Y, Fujita M, Snigórska K, Watanabe M and Okuda T: Dehydration-inducible changes in expression of two aquaporins in the sleeping chironomid, *Polypedilum vanderplanki*. Biochim Biophys Acta 1778: 514-520, 2008.
28. Drake KD, Schuette D, Chepelinsky AB, Jacob TJ and Crabbe MJ: pH-Dependent channel activity of heterologously-expressed main intrinsic protein (MIP) from rat lens. FEBS Lett 512: 199-204, 2002.
29. Yanochko GM and Yool AJ: Regulated cationic channel function in *Xenopus* oocytes expressing *Drosophila* big brain. J Neurosci 22: 2530-2540, 2002.
30. Park JH and Saier MH Jr: Phylogenetic characterization of the MIP family of transmembrane channel proteins. J Membrane Biol 153: 171-180, 1996.
31. Pietrantonio PV, Jagge C, Keeley LL and Ross LS: Cloning of an aquaporin-like cDNA and in situ hybridization in adults of the mosquito *Aedes aegypti* (Diptera: Culicidae). Insect Mol Biol 9: 407-418, 2000.
32. Martini SV, Goldenberg RC, Fortes FS, Campos-de-Carvalho AC, Falkenstein D and Morales MM: *Rhodnius prolixus* Malpighian tubule's aquaporin expression is modulated by 5-hydroxytryptamine. Arch Insect Biochem Physiol 57: 133-141, 2004.
33. Philip BN, Yi SX, Elnitsky MA and Lee RE Jr: Aquaporins play a role in desiccation and freeze tolerance in larvae of the goldenrod gall fly, *Eurosta solidaginis*. J Exp Biol 211: 1114-1119, 2008.