Insect Biochemistry and Molecular Biology 45 (2014) 125–140

Contents lists available at ScienceDirect

Insect Biochemistry and Molecular Biology

journal homepage: www.elsevier.com/locate/ibmb

Molecular and functional characterization of multiple aquaporin water channel proteins from the western tarnished plant bug, *Lygus hesperus*

Jeffrey A. Fabrick [a,*], Jinxin Pei [b], J. Joe Hull [a], Andrea J. Yool [b]

[a] USDA-ARS, U.S. Arid Land Agricultural Research Center, 21881 North Cardon Lane, Maricopa, AZ 85138, USA
[b] University of Adelaide, School of Medical Sciences, Frome Rd., Medical School South, Adelaide, SA 5005, Australia

ARTICLE INFO

Article history:
Received 17 September 2013
Received in revised form
27 November 2013
Accepted 1 December 2013

Keywords:
Aquaporin
Major intrinsic protein family
Lygus hesperus
Western tarnished plant bug
Hemiptera
Superaquaporin
S-aquaporin

ABSTRACT

Aquaporins (AQPs) are integral membrane channel proteins that facilitate the bidirectional transfer of water or other small solutes across biological membranes involved in numerous essential physiological processes. In arthropods, AQPs belong to several subfamilies, which contribute to osmoregulation, respiration, cryoprotection, anhydrobiosis, and excretion. We cloned and characterized five novel AQPs from the western tarnished plant bug, *Lygus hesperus*, a polyphagous insect pest of food and fiber crops throughout western North America. The *L. hesperus* AQPs (LhAQP1-5) belong to different phylogenetic subfamilies, have unique transcription profiles and cellular localizations, and all transport water (but not glycerol) when heterologously expressed in *Xenopus laevis* oocytes. Our results demonstrate that multiple AQPs with possible compensatory functions are produced in *L. hesperus* that likely play important roles in maintaining water homeostasis in this important insect pest.

Published by Elsevier Ltd.

1. Introduction

Aquaporins (AQPs) belong to the ancient class of major intrinsic proteins (MIPs), which are integral membrane channel proteins found in all kingdoms of life. These proteins facilitate the bidirectional transfer of water or sometimes other small neutral solutes across biological membranes involved in numerous essential physiological processes (Hachez and Chaumont, 2010; Gomes et al., 2009; King et al., 2004). AQPs are best known for their diverse roles in water transport related to cell water balance regulation [reviewed in Agre et al. (1998); Carbrey and Agre, 2009], but they also transport a wide range of solutes, including urea and glycerol, hydrogen peroxide, dissolved gasses (CO_2, NO, NH_3), and certain metalloids (Cohen, 2013; Hachez and Chaumont, 2010). Whereas most solutes permeate through the monomeric channel, AQPs usually assemble as tetramers forming a fifth central pore that in some cases functions to conduct ions (Yool and Weinstein, 2002; Yool, 2007).

The functional diversity of AQPs is facilitated by key modifications to their structure. In general, members of the MIP superfamily share a number of structural features, including six helical transmembrane (TM) domains connected by five alternating extracellular/intracellular loops with intercellular amino and carboxyl termini, two canonical Asp-Pro-Ala (NPA) motifs, and the aromatic/arginine (ar/R) selectivity filter (Zardoya, 2005). Together, the canonical amino- and carboxyl-terminal halves form two tandem repeats, each containing a single NPA motif, forming the 'aquaporin' or 'hourglass' fold (Jung et al., 1994; Murata et al., 2000) and ultimately shaping the selective TM channel. 'Classical' or 'orthodox' AQPs allow water diffusion through an extremely narrow and electrostatically-selective pore that precludes the permeability of large, hydrophobic, or incorrectly charged solutes (Murata et al., 2000). In another functional class of AQPs known as the aquaglyceroporins, the composition of the ar/R constriction site and the corresponding larger, more hydrophobic three-dimensional shape of the pore enables permeation of additional solutes, such as

Abbreviations: AQP, aquaporin; MIPs, major intrinsic proteins; NPA motif, Asparagine-Proline-Alanine motif; ar/R, aromatic arginine; DRIPs, *Drosophila* intrinsic proteins; BiB, Big Brain protein; RNAi, RNA interference; PRIPs, *Pyrocoelia rufa* integral proteins; LhAQP, *Lygus hesperus* aquaporin; RACE, rapid amplification of cDNA ends; cDNA, complementary DNA; CDS, coding sequence; ORF, open reading frame; RT-PCR, reverse transcriptase polymerase chain reaction; Tni cells, cultured *Trichoplusia ni* cells; EGFP, enhanced green fluorescent protein; cRNA, complementary RNA; TM, transmembrane; RPIPs, *Rhodnius prolixus* integral proteins; DVIPs, *Dermacentor variabilis* integral proteins; LHIPs, *Lygus hesperus* integral proteins; ER, endoplasmic reticulum; EST, expressed sequence tag; SIP, short basic intrinsic proteins.

* Corresponding author. Tel.: +1 520 316 6335; fax: +1 520 316 6330.
E-mail addresses: jeff.fabrick@ars.usda.edu, jeff.fabrick@ars.usda.gov (J. A. Fabrick).

0965-1748/$ – see front matter Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.ibmb.2013.12.002

polyols (Fu et al., 2000; Hub and de Groot, 2008). Members of a recently identified subfamily known as 'superaquaporins' or "S-aquaporins" exhibit low sequence similarity with other AQPs, altered NPA signature motifs, and anomalous subcellular localization (Ishibashi, 2006, 2009; Nozaki et al., 2008; Calvanese et al., 2013).

Whereas vertebrate AQPs have been studied extensively [reviewed in King et al. (2004); Verkman 2008; and Ishibashi et al., 2009], far less attention has been given to invertebrate AQPs. In arthropods, AQPs are involved in regulating the movement of high volume liquid diets, osmoregulation, respiration, cryoprotection and anhydrobiosis [reviewed in Campbell et al. (2008); Spring et al., 2009; and Cohen 2013]. Genome sequencing indicates that insects may contain up to 5-8 genes encoding putative aquaporin-like transcripts (Campbell et al., 2008; Drake et al., 2010), but the complete repertoire from any single species is yet to be fully characterized. Multiple functional AQPs have been found in several insects, including three from *Aedes aegypti* (Duchesne et al., 2003; Drake et al., 2010) and *Bombyx mori* (Kataoka et al., 2009a; Azuma et al., 2012), and two from the oriental fruit moth, *Grapholita molesta* (Kataoka et al., 2009b), the pea aphid, *Acyrthosiphon pisum* (Shakesby et al., 2009; Wallace et al., 2012), the sleeping chironomid, *Polypedilum vanderplanki* (Kikawada et al., 2008), and *Drosophila melanogaster* (DRIP and BiB) (Kaufmann et al., 2005; Yanochko and Yool, 2002). For *A. aegypti*, studies have shown that a water-specific DRIP (AeaAQP) is involved in water movement in Malpighian tubules and tracheoles (Duchesne et al., 2003). RNA interference (RNAi) knockdown further revealed a role in diuresis in Malpighian tubules for three of the six putative AQPs (Drake et al., 2010). Functional AQPs from *B. mori* include, AQP-Bom1 (a water-specific DRIP), AQP-Bom2 (an aquaglyceroporin), and AQP-Bom3 (a water-specific PRIP), all of which play important roles in water reabsorption and recycling within the cryptonephric rectal complex (Kataoka et al., 2009a; Azuma et al., 2012). *G. molesta* has a functional water-specific DRIP (AQP-Gra1) and an aquaglyceroporin (AQP-Gra2) (Kataoka et al., 2009b). In *A. pisum*, ApAQP1 was shown to be a water-specific AQP (Shakesby et al., 2009) whereas ApAQP2 exhibited aquaglyceroporin activities, transporting water and glycerol in *Xenopus* oocytes (Wallace et al., 2012). Two *P. vanderplanki* AQPs (PvAQP1 and PvAQP2) facilitated water permeation, but not transport of glycerol in *Xenopus* oocytes (Kikawada et al., 2008). Among the *D. melanogaster* AQPs characterized are a water-selective AQP called *Drosophila* integral protein or DRIP, which is involved in excretion and osmoregulation (Kaufmann et al., 2005), and the novel AQP known as Big Brain (DmBIB), which does not permeate water, but functions as a cation channel (Yanochko and Yool, 2002). Other arthropods having at least one functionally characterized AQP include: *Cicadella viridis* (La Caherec et al., 1997); *Rhodnius prolixus* (Echevarria et al., 2001); *Anopheles gambiae* (Liu et al., 2011); *Blattella germanica* (Herraiz et al., 2011); *Eurosta solidaginis* (Philip et al., 2011); *Bemisia tabaci* (Mathew et al., 2011); *Belgica antarctica* (Goto et al., 2011); *Anomala cuprea* (Nagae et al., 2013); and the ticks, *Dermacentor variabilis* (Holmes et al., 2008), *Rhipicephalus sanguineus* (Ball et al., 2009), and *Ixodes ricinus* (Campbell et al., 2010).

Numerous species within the hemipteran family Miridae are important agricultural pests causing crop damage from feeding with specialized piercing-sucking mouthparts (Wheeler, 2001). One such mirid bug is *Lygus hesperus* Knight (the western tarnished plant bug), a highly polyphagous pest causing economic losses in numerous cropping systems in western North America (Wheeler, 2001). Lygus bugs are "cell rupture" feeders (previously termed "lacerate-and-flush") that first puncture host tissue and inject saliva containing digestive enzymes (Miles, 1987; Strong and Kruitwagen, 1968; Shackel et al., 2005; Backus et al., 2007). The pre-digested fluid is then pumped through the food canal for absorption of nutrients primarily within the midgut epithelium of the alimentary tract (Wheeler, 2001; Habibi et al., 2008). Some hemipterans use specialized biochemical processes or gut morphology (e.g. filter chambers) to abate osmotic stress and remove excess dietary fluid (Douglas, 2006; Gullan and Cranston, 2005; Lehane and Billingsley, 1996; Hubert et al., 1989; Mathew et al., 2011). Lygus bugs lack the specialized filter chamber or gastric cecae of sap feeding hemipterans, and instead have a simple, cimicomorph type alimentary tract typical of other mesophyll feeders (Goodchild, 1966; Habibi et al., 2008).

Here, we characterize five AQPs (LhAQP1-5) from *L. hesperus*. Based on sequence conservation and phylogenetics, bioinformatic predictions, cellular and tissue localization, temporal expression, and functional analysis, the LhAQPs belong to different subfamilies but all function as water-specific channel proteins. Localization and transcription profiles for LhAQPs provide clues about the putative roles these functionally redundant AQPs may play in maintaining *L. hesperus* water homeostasis. The combination of standard molecular cloning techniques and the mining of transcriptome data for AQPs provides an unprecedented opportunity to comprehensively identify and characterize AQPs within *L. hesperus*.

2. Materials and methods

2.1. Insects

A laboratory colony of *Lygus hesperus* was maintained at the USDA-ARS U.S. Arid Land Agricultural Research Center in Maricopa, AZ, USA. The colony was fed green beans and artificial diet at 25 °C under 20% humidity and with an L14:D10 photoperiod (Debolt, 1982; Patana, 1982).

2.2. cDNA isolation and 5′-RACE of LhAQPs

Total RNA was extracted from 100 mg of *L. hesperus* adults and 2nd–3rd instar nymphs using TRIzol® reagent (Invitrogen-Life Technologies, Carlsbad, CA). cDNA was produced using random hexamer primers and SuperScript III First-strand Synthesis System (Invitrogen-Life Technologies) according to manufacturer's recommendations. PCR primers 1LhAQP5-8LhAQP3 (Table 1) were designed using Primer3Plus (Untergasser et al., 2007) from four putative *Lygus lineolaris* AQPs obtained from a *L. lineolaris* transcriptome sequencing project (O.P. Perera, unpublished). Four partial LhAQPs (named LhAQP1-4) were PCR amplified using ExTaq DNA polymerase premix (Takara-Clontech, Palo Alto, CA) and primer pairs (Table 1). Products were electrophoresed on a 1% agarose gel and visualized using SYBR Safe (Invitrogen-Life Technologies). Bands were gel-purified using Montage DNA Gel Extraction Kit (EMD Millipore, Billerica, MA) and ligated into pCR2.1-TOPO using TOPO TA Cloning Kit (Invitrogen-Life Technologies). Plasmid DNA was propagated in OneShot TOP10 chemically competent *Escherichia coli* and purified using QIAprep Spin Miniprep Kit (Qiagen, Valencia, CA). Inserts were sequenced with T7 and M13 Reverse vector primers by the Arizona State University DNA Core Lab (Tempe, AZ).

The 5′ and 3′ ends of LhAQP1-4s were identified by rapid amplification of cDNA ends (RACE) using the SMARTer RACE cDNA Amplification Kit (Clontech). At least two sense and two antisense primers were designed for each partial LhAQP consensus sequence (Table 1). The primers were used with Universal Primer A (Clontech) in fully nested PCR to amplify 5′ and 3′ ends (Table 1). PCR products were sub-cloned into pCR2.1-TOPO (Invitrogen-Life Technologies) and sequenced as indicated above.

Table 1
Nucleotide primers used to obtain partial and full-length *Lygus hesperus* aquaporins (LhAQP1-5), rapid amplification of cDNA ends (5′- and 3′-RACE), reverse transcriptase-PCR (RT-PCR) amplification of cDNA, and for sub-cloning into pIB expression vector.

Primer	Primer DNA sequence[a]	Direction	LhAQP	Application
AcF	5′-ATGTGCGACGAAGAAGTTG-3′	Sense	–	RT-PCR
AcR1	5′-GTCACGGCCAGCCAAATC-3′	Antisense	–	RT-PCR
1LhTub5	5′-CACCAAGTCGTTCATGTTGG-3′	Sense	–	RT-PCR
2LhTub3	5′-GTCACCACTTGCCTCAGGTT-3′	Antisense	–	RT-PCR
1LhAQP5	5′-CAAACTCTGAGCCTTCCCACA-3′	Sense	3	Partial cloning[b]
2LhAQP3	5′-GTCGAGTCAACACGCTGGT-3′	Antisense	3	Partial cloning
3LhAQP5	5′-GTCGAGTCAACACGCTGGT-3′	Sense	4	Partial cloning
4LhAQP3	5′-CTTTTCCAAGCCGCTTCAT-3′	Antisense	4	Partial cloning
5LhAQP5	5′-TTGGGTACGGAGGATGTGG-3′	Sense	1	Partial cloning
6LhAQP3	5′-TTACAAATCGTAAGAACTGTCGTC-3′	Antisense	1	Partial cloning
7LhAQP5	5′-GGAAGTGTTTGGAGCAGAGC-3′	Sense	2	Partial cloning
8LhAQP3	5′-GCAGGTGGTGTAGAAAAGTGG-3′	Antisense	2	Partial cloning
18LhAQP5	5′-TGCCTCTGACTGGATCAGGA-3′	Sense	1	3′-RACE[c]
19LhAQP5	5′-GTTGGACAGTGGGACAACCA-3′	Sense	1	3′-RACE
20LhAQP3	5′-GCCTGACCCTCAGTGAGTCC-3′	Antisense	1	5′-RACE
21LhAQP3	5′-CATCTGAACGACCGAAGCAA-3′	Antisense	1	5′-RACE
22LhAQP5	5′-TGGCCCTCATCTTTCTGGTT-3′	Sense	2	3′-RACE
23LhAQP5	5′-CCCCGTCAGAAGTCTTGGTC-3′	Sense	2	3′-RACE
24LhAQP3	5′-TGATTGCTCCCATGATTTGC-3′	Antisense	2	5′-RACE
25LhAQP3	5′-TGCCCCAGTACGTAGAATGC-3′	Antisense	2	5′-RACE
34LhAQP3	5′-CTGGCGAGCAGCACATCGCGGGTCA-3′	Antisense	2	5′-RACE
35LhAQP3	5′-CACCAACGCCCATGATTGCTCCCATGA-3′	Antisense	2	5′-RACE
36LhAQP5	5′-TGCTGCTCTGCATCGGCTGCGCTTC-3′	Sense	2	3′-RACE
37LhAQP5	5′-CCGCGATGTGCTGCTCGCCAGGACT-3′	Sense	2	3′-RACE
26LhAQP3	5′-AGCGCCGAAATTTGGGGCAGTCCAG-3′	Antisense	3	5′-RACE
27LhAQP3	5′-CGAGCGACATGTGCCCCATGACCAG-3′	Antisense	3	5′-RACE
28LhAQP5	5′-GGTCATGGGGCACATGTCGCTCGTC-3′	Sense	3	3′-RACE
29LhAQP5	5′-GACCGCCGGGTTTCTGCACCACTCA-3′	Sense	3	3′-RACE
30LhAQP3	5′-GCGGGATTCATGCTGGAACCCGTGGA-3′	Antisense	4	5′-RACE
31LhAQP3	5′-CGGCACCTCCGATGGCGCCAAGACA-3′	Antisense	4	5′-RACE
32LhAQP5	5′-AAGGCGGTGGTGGCCGAAGCACTCG-3′	Sense	4	3′-RACE
33LhAQP5	5′-TGGCGCCATCGGAGGTGCCGCTATG-3′	Sense	4	3′-RACE
38LhAQP5	5′-ATGATGGGCGACCACAGTTCTATC-3′	Sense	1	Cloning CDS[d]
39LhAQP3	5′-TTACAAATCGTAAGAACTGTCGTCG-3′	Antisense	1	Cloning CDS
40LhAQP5	5′-ATGGTGACCGACTCGGGCGGGA-3′	Sense	2A	Cloning CDS
41LhAQP3	5′-CTACGATTGGCCTCCCCTGAGGAA-3′	Antisense	2A	Cloning CDS
42LhAQP5	5′-ATGGCTACTCCACTCGTTTTGAACAGGGTG-3′	Sense	2B	Cloning CDS
43LhAQP5	5′-ATGGGAGTCACGACCGGATCGG-3′	Sense	2C	Cloning CDS
44LhAQP5	5′-ATGAACAGTGGGAACAGCGTTTTTCTG-3′	Sense	2D	Cloning CDS
86LhAQP5	5′-GGATCCATGACGATGAGCGAAT-3′	Sense	2E	Cloning CDS & pIB
45LhAQP5	5′-ATGCCTCCGAGATCAAAAGCACG-3′	Sense	3	Cloning CDS
46LhAQP3	5′-TTATATGGTGGCGTCGACTTCTCTTGT-3′	Antisense	3	Cloning CDS
47LhAQP5	5′-ATGCCAGACAAGAGGAGTCGAGTCAA-3′	Sense	4	Cloning CDS
48LhAQP3	5′-TCATGCGTTGCTGTTTTCATTTCCT-3′	Antisense	4	Cloning CDS
79LhAQP5	5′-ATGGGGGGCATCTTCGG-3′	Sense	5	Cloning CDS
80LhAQP3	5′-TTAGAGATCTTTTTCTTTCCCCAGGA-3′	Antisense	5	Cloning CDS
61LhAQP5	5′-ATTCGATGCTGTTGGTAGGG-3′	Sense	1	RT-PCR[e]
62LhAQP3	5′-CTGATCCAGTCAGAGGCACA-3′	Antisense	1	RT-PCR
63LhAQP5	5′-GGAAGTGTTTGGAGCAGAGC-3′	Sense	2	RT-PCR
64LhAQP3	5′-TGCGCCAGTGTATTTACCAA-3′	Antisense	2	RT-PCR
65LhAQP5	5′-GGTGCTTAGTTTCGCTTTGG-3′	Sense	3	RT-PCR
66LhAQP3	5′-CGGTAGATCACTGCTGCAAA-3′	Antisense	3	RT-PCR
67LhAQP5	5′-CGGTAGATCACTGCTGCAAA-3′	Sense	4	RT-PCR
68LhAQP5	5′-CCCAGTGGTTATCCCATTTG-3′	Antisense	4	RT-PCR
77LhAQP5	5′-CGACTTACGCCCTCTTCTTG-3′	Sense	5	RT-PCR
78LhAQP3	5′-GGAACCGATCCAGTAGACGA-3′	Antisense	5	RT-PCR
81LhAQP5	5′-GGATCCATGGTGACCGACTC-3′	Sense	2A	pIB
82LhAQP3	5′-AGGCCTCGATTGGCCTCCCCTG-3′	Antisense	2A	pIB
83LhAQP5	5′-GGATCCATGGCTACTCCACTC-3′	Sense	2B	pIB
84LhAQP5	5′-GGATCCATGGGAGTCACGACCGGAT-3′	Sense	2C	pIB
85LhAQP5	5′-GGATCCATGAACAGTGGGAAC-3′	Sense	2D	pIB
86LhAQP5	5′-GGATCCATGACGATGAGCGAAT-3′	Sense	2E	pIB
73LhAQP5	5′-GGATCCATGCCTCCGAGATCAAAAGC-3′	Sense	3	pIB
74LhAQP3	5′-AGGCCTTATGGTGGCGTCGACTTCTCTTGT-3′	Antisense	3	pIB
75LhAQP5	5′-GGATCCATGCCAGACAAGAGGAGTCGA-3′	Sense	4A	pIB
76LhAQP3	5′-AGGCCTTGCGTTGCTGTTTTCATTTCCTT-3′	Antisense	4A	pIB
89LhAQP5	5′-GGAGAGAGCTCCACTGAGGTACGCATCAAGGTATTA-3′	Sense	4A	pIB
90LhAQP3	5′-TAATACCTTGATGCGTACCTCAGTGGAGCTCTCTCC-3′	Antisense	4A	pIB
87LhAQP5	5′-GGATCCATGGGGGGCATCTT-3′	Sense	5	pIB
88LhAQP3	5′-AGGCCTGAGATCTTTTTCTTTCC-3′	Antisense	5	pIB
LhAQP1 Met3 F	5′-ATGCCACGTTCTGAAGAAG-3′	Sense	1	pIB
EGFP-LhAQP1 OE F	5′-gacgagctgtacaagATGATGGGCGACCAC-3′	Sense	1	Domain swap/pIB

(*continued on next page*)

Table 1 (*continued*)

Primer	Primer DNA sequence[a]	Direction	LhAQP	Application
EGFP-LhAQP1 OE R	5′-GTGGTCGCCCATCATcttgtacagctcgtc-3′	Antisense	1	Domain swap/pIB
BtAQP1N-LhAQP1 F	5′-*TAGTTGCTGAGTTCC*TGGGAAC-3′	Sense	1	Domain swap/pIB
BtAQP1N-LhAQP1 R	5′-GTTCCCAG*GAACTCAGCAACTA*-3′	Antisense	1	Domain swap/pIB
LhAQP1N-BtAQP1 F	5′-CGTCGCTGAGTT*CGTAGGGAC*-3′	Sense	1	Domain swap/pIB
LhAQP1N-BtAQP1 R	5′-*GTCCCTACG*AACTCAGCGACG-3′	Antisense	1	Domain swap/pIB
LhAQP1-EGFP F	5′-GTTCTTACGATTTGatggtgagcaagggcgag-3′	Sense	1	Domain swap/pIB
LhAQP1-EGFP R	5′-ctcgcccttgctcaccatCAAATCGTAAGAAC-3′	Antisense	1	Domain swap/pIB
*Stu*I-EGFP F	5′-AGGCCTATGGTGAGCAAG-3′	Sense	–	Domain swap/pIB
EGFP R stop	5′-TTACTTGTACAGCTCGTCCATG-3′	Antisense	–	Domain swap/pIB
20BtAQP3	5′-*TCAGAAATCATAAGAGCTCTCATC*-3	Antisense	–	Domain swap/pIB

[a] Underlined sequence corresponds to restriction enzyme sites (G*GATCC = *Bam*HI, GGTAC*C = *Kpn*I, GAGCT*C = *Sac*I, AGG*CCT = *Stu*I). Sequence in lower case indicates regions within primers corresponding to EGFP used in overlap extension PCR to produce LhAQP1 chimera with EGFP at amino- or carboxyl-terminus. Italicized sequence corresponds to primer regions from BtAQP1 (ABW96354.1) used in overlap extension PCR to produce LhAQP1–BtAQP1 amino-terminal domain swaps.

[b] Partial LhAQPs 1–4 were PCR amplified using primer pairs 1LhAQP5 + 2LhAQP3 (LhAQP1), 3LhAQP5 + 4LhAQP3 (LhAQP2), 5LhAQP5 + 6LhAQP3 (LhAQP3), and 7LhAQP5 + 8LhAQP3 (LhAQP4).

[c] For 5′ and 3′ RACE at least two sense and two antisense primers were used to amplify ends of LhAQP1-4 (18LhAQP5 and 19LhAQP5 for 3′-RACE of LhAQP1; 20LhAQP3 and 21LhAQP3 for 5′-RACE of LhAQP1; 22LhAQP5, 23LhAQP5, 36LhAQP5, and 37LhAQP5 for 3′-RACE of LhAQP2; 24LhAQP3, 25LhAQP3, 34LhAQP3, and 35LhAQP3 for 5′-RACE of LhAQP2; 26LhAQP3 and 27LhAQP3 for 5′-RACE of LhAQP3; 28LhAQP5 and 29LhAQP5 for 3′-RACE of LhAQP3; 30LhAQP3 and 31LhAQP3 for 5′-RACE of LhAQP4; 32LhAQP5 and 33LhAQP5 for 3′-RACE of LhAQP4).

[d] Full-length coding sequences for LhAQPs were confirmed by PCR amplification from adult *Lygus hesperus* cDNA using primer pairs 38LhAQP5 + 39LhAQP3 (LhAQP1), 40LhAQP5 + 41LhAQP3 (LhAQP2A), 42LhAQP5 + 41LhAQP3 (LhAQP2B), 43LhAQP5 + 41LhAQP3 (LhAQP2C), 44LhAQP5 + 41LhAQP3 (LhAQP2D), 86LhAQP5 + 41LhAQP3 (LhAQP2E), 45LhAQP5 + 46LhAQP3 (LhAQP3), 47LhAQP5 + 48LhAQP3 (LhAQP4), and 79LhAQP3 + 80LhAQP (LhAQP5).

[e] For reverse transcriptase-PCR (RT-PCR), LhAQP-specific primers were used to amplify products from within each ORF using primer pairs 61LhAQP5 + 62LhAQP3 (LhAQP1), 63LhAQP5 + 64LhAQP3 (LhAQP2), 65LhAQP5 + 66LhAQP3 (LhAQP3), 67LhAQP5 + 68LhAQP3 (LhAQP4), and 77LhAQP5 + 78LhAQP3 (LhAQP5).

Concurrent with cloning the four putative partial cDNAs (LhAQP1-4), we sequenced and *de novo* assembled a Roche 454 transcriptome from normalized cDNA libraries prepared from 0 to 5 day old post-eclosion *L. hesperus* adults (Hull et al., 2013). In addition to confirming the presence of the LhAQP1-4 transcripts (Table 2), the *L. hesperus* 454 transcriptome also revealed an additional putative full-length aquaporin-like transcript, designated as LhAQP5.

Full-length LhAQP coding sequences (CDS) corresponding to open reading frames (ORFs) were confirmed by PCR amplification from *L. hesperus* adult cDNA using primer pairs (Table 1). All products corresponding to full-length LhAQP CDS were PCR amplified using ExTaq DNA polymerase (Takara-Clontech), subcloned into pCR2.1-TOPO (Invitrogen-Life Technologies), and sequenced as indicated above.

Table 2
Aquaporin protein transcripts identified by cDNA cloning and from *de novo*-assembled *Lygus hesperus* and *Lygus lineolaris* transcriptomes.

LhAQP	LhAQP NCBI accession number	Number of *L. hesperus* AQP isotigs per isogroup[a]	CDS (bp)[b]	ORF (aa)[c]	Number of putative *L. lineolaris* AQP contigs[d]	CDS (bp)[e]	ORF (aa)[f]
1	KF048092	2	864	287	5	864	287
2A	KF048093	–[g]	789	262	–	–	–
2B	KF048094	–	813	270	–	–	–
2C	KF048095	–	885	294	–	–	–
2D	KF048096	–	846	281	–	–	–
2E	KF048097	9	768	255	2	768	255
3	KF048098	10	804	267	2	804	267
4A	KF048099	1	837	278	–	–	–
4B	KF048100	–	798	265	2	798	265
5	KF048101	5	798	265	–	–	–

[a] Total number of contigs comprising LhAQPs obtained from *de novo*-assembled adult *Lygus hesperus* 454 transcriptome (Hull et al., 2013).

[b] Number of nucleotides comprising *L. hesperus* AQP coding sequences (CDS) obtained from RACE and PCR cloning.

[c] Number of residues from translation of *L. hesperus* AQP CDS using ExPASy Translate tool (http://web.expasy.org/translate/).

[d] Total number of putative AQP contigs obtained from *de novo*-assembled *Lygus lineolaris* transcriptome sequencing libraries (personal communication from O.P. Perera).

[e] Number of nucleotides comprising *L. lineolaris* CDS for AQP transcripts obtained by aligning contigs and generating consensus sequences.

[f] Number of residues from translation of *L. lineolaris* AQP CDS using ExPASy Translate tool (http://web.expasy.org/translate/).

[g] "–" indicates that for cloned AQP cDNAs no contigs/isotigs were found in *de novo*-assembled *Lygus hesperus* or *Lygus lineolaris* transcriptomes.

2.3. Bioinformatics and phylogeny

Sequences were analyzed in Vector NTI (Life Technologies) and comparison against the non-redundant public sequence database was made using BLAST search programs (Altschul et al., 1990). Sequence analysis tools of the ExPASy Molecular Biology Server of Swiss Institute of Bioinformatics, including Translate and Compute pI/MW were used to analyze the deduced LhAQP protein sequences. Predictions of intra/extracellular domains and transmembrane helices were made using TMPred (Hofmann and Stoffel, 1993), TMHMM 2.0 (Krogh et al., 2001), Phobius (Käll et al., 2004), RHYTHM (Rose et al., 2009), TOPCONS (Bernsel et al., 2009), HMMTOP (Tusnady and Simon, 2001), and TopPredII (Claros and von Heijne, 1994). Protein subcellular localization was predicted using WolF PSORT (Horton et al., 2007). Multiple sequence alignments were performed using CLUSTALW (Larkin et al., 2007). Phylogenetic analysis using predicted full-length AQP protein sequences was performed with the unweighted pair group method with arithmetic mean (UPGMA, Sneath and Sokal, 1973), neighbor-joining (Saitou and Nei, 1987), and minimum evolution and maximum parsimony (Saitou and Nei, 1986) methods. Trees were constructed with 10,000 bootstrap replicates using MEGA version 5 (Tamura et al., 2011). Only the UPGMA tree is shown, as all trees were similar. Glycosylation predictions were made using post-translational modification servers found at the Center for Biological Sequence Analysis (http://www.cbs.dtu.dk/services/).

2.4. RNA extraction and semi-quantitative reverse transcription PCR

Lygus representing different developmental stages (egg, 1st instar nymph, 2nd–3rd instar nymph, 4th–5th instar nymph, and adult) were collected, weighed, and frozen in liquid nitrogen. Total tissue weight for each stage ranged from 60 to 112 mg. For tissue distribution RT-PCR, heads, legs, accessory glands, testes, ovaries,

midguts, and Malpighian tubules were dissected from adults and stored in RNAlater (Ambion-Life Technologies) at −80 °C. Tissue samples were homogenized using Kontes pestles (Fisher Scientific, Pittsburg, PA) and total RNA was extracted using TRIzol® reagent (Invitrogen-Life Technologies). RNA concentration was determined using a NanoDrop ND1000 spectrophotometer (NanoDrop Technologies/Thermo Scientific, Wilmington, DE). Total RNA was treated with DNA-free™ DNase I (Ambion-Life Technologies) to remove genomic DNA contamination, if any, carried over during total RNA extraction. Integrity of the total RNA was confirmed on the Agilent 2100 Bioanalyzer using the RNA Nano 6000 LabChip kit (Agilent Technologies, Santa Clara, CA) according to manufacturer's instructions.

Samples were diluted and 1 μg of DNase-treated total RNA was used to prepare cDNA using random decamer primers according to the protocol described in RetroScript cDNA Synthesis Kit (Ambion-Life Technologies). For reverse transcriptase-PCR (RT-PCR), LhAQP-specific primers were used to amplify 503–547-bp products from within each ORF (Table 1). Nucleotide primer sequences for 63LhAQP5 and 64LhAQP3 are conserved in LhAQP2A-E and therefore, end-point RT-PCR products represent transcriptional profiles for LhAQP2 as a group. Likewise, 67LhAQP5 and 68LhAQP3 primer sequences are found in both LhAQP4A and LhAQP4B. The appropriate number of PCR cycles was determined empirically and included 25, 28, and 30 cycles. Actin (AcF + AcR1) and β-tubulin (1LhTub5 + 2LhTub3) were used as amplification controls for stages and tissues, respectively (Hull et al., 2013). Products were electrophoresed on 0.8% agarose gels and visualized using SYBR Safe (Invitrogen-Life Technologies).

2.5. Expression in Tni cell culture and fluorescence microscopy

To observe cellular localization, recombinant LhAQPs were expressed using the pIB/V5-His TOPO TA (Invitrogen-Life Technologies) insect cell expression vector in cultured *Trichoplusia ni* (Tni) cells (Allele Biotechnology, San Diego, CA). Recombinant proteins were expressed as chimeras with the enhanced green fluorescent protein (EGFP) produced in frame with LhAQPs either at their amino- or carboxyl termini.

A pIB/*Stu*I-EGFP reporter cassette was modified from pIB/V5-His TOPO TA (Life Technologies) and used to express chimeric LhAQP2-5-EGFP recombinant proteins in cultured Tni cells. Full-length LhAQP2A-E, LhAQP3, LhAQP4A-B, and LhAQP5 coding sequences were first cloned into pCR2.1-TOPO (Invitrogen-Life Technologies) using PCR with KOD DNA polymerase (EMD Millipore) and primers that introduced *Bam*HI and *Stu*I restriction sites on the 5′ and 3′ ends of LhAQPs, respectively [81LhAQP5 + 82LhAQP3 (LhAQP2A), 83LhAQP5 + 82LhAQP3 (LhAQP2B), 84LhAQP5 + 82LhAQP3 (LhAQP2C), 85LhAQP5 + 82LhAQP3 (LhAQP2D), 86LhAQP5 + 82LhAQP3 (LhAQP2E), 73LhAQP5 + 74LhAQP3 (LhAQP3), 75LhAQP5 + 76LhAQP3 (LhAQP4A and 4B), and 87LhAQP5 + 88LhAQP3 (LhAQP5)]. LhAQP/pCR2.1-TOPO plasmids and pIB/*Stu*I-EGFP were digested with *Bam*HI-HF and *Stu*I, gel-purified using an EZNA gel extraction kit (Omega Bio-Tek, Norcross, GA), and ligated using Mighty Mix DNA ligation kit (Takara-Clontech). Because LhAQP4B is preferentially PCR amplified over LhAQP4A, additional overlap extension PCR steps were needed to make LhAQP4A/pCR2.1-TOPO with the appropriate restriction sites. The first PCR amplified the 5′ end of LhAQP4A with a 5′ *BamH*I site but missing 36 bp from LhAQP4B using 75LhAQP5 + 90LhAQP3. The second PCR amplified the 3′end of LhAQP4A including the unique 36 bp and a 3′ *Sac*I site using 89LhAQP5 + 76LhAQP3. The final step involved PCR amplification using the two previous reactions as templates with 75LhAQP5 + 76LhAQP3. The resulting PCR product was cloned into pCR2.1-TOPO (Invitrogen-Life Technologies), digested with *Bam*HI and *Stu*I, and sub-cloned into pIB/*Stu*I-EGFP as described above. High-fidelity KOD DNA polymerase (EMD Millipore) was used for all PCR amplifications. Orientation and correctness of the pCR2.1-TOPO and pIB/*Stu*I-EGFP inserts were confirmed by sequencing at the ASU DNA Core Lab. All restriction enzymes were from New England Biolabs (Ipswich, MA).

Because LhAQP1 has endogenous *Bam*HI and *Stu*I restriction sites, we used overlap extension PCR to make the EGFP-LhAQP1 pIB expression vector variants. To make LhAQP1-EGFP/pIB (chimeric LhAQP1 with EGFP at carboxyl-terminus), the first PCR reactions (38LhAQP5 + LhAQP1-EGFP R and LhAQP1-EGFP F + EGFP R stop) amplified the coding sequences for LhAQP1 and EGFP, respectively. The final PCR amplification used the two previous reactions as templates with the LhAQP1 gene-specific sense primer 38LhAQP5 and the EGFP-specific antisense primer, EGFP R stop. Similarly, for EGFP-LhAQP1/pIB (chimeric LhAQP1 with EGFP at amino-terminus), the first PCR reactions (*Stu*I-EGFP F + EGFP-LhAQP1 OE R and EGFP-LhAQP1 OE F + 39LhAQP3) amplified the coding sequences for EGFP and LhAQP1, which were then used as templates for final PCR using the sense primer *Stu*I-EGFP F and the LhAQP1-specific antisense primer, 39LhAQP3. Final products were sub-cloned into pIB/V5-His-TOPO (Invitrogen-Life Technologies) and DNA was sequenced at ASU DNA Core Lab.

An expression vector with a truncated LhAQP1 beginning at base position 28 (corresponding to the third methionine in translated LhAQP1, see Supplemental Fig. 1) and a carboxyl-terminal EGFP was prepared by overlap extension PCR. The first PCRs used LhAQP1 Met3 F + LhAQP1-EGFP R and LhAQP1-EGFP F + EGFP R stop to amplify the CDS corresponding to Met3-truncated LhAQP1 and EGFP, respectively. Products of these reactions were used as templates for final amplification using LhAQP1 Met3 F + EGFP R stop and were sub-cloned in pIB/V5-His-TOPO (Invitrogen-Life Technologies) and sequenced as above.

To compare the cellular localization of LhAQP1 with that of *B. tabaci* aquaporin 1 (BtAQP1) we used EGFP-BtAQP1/pIB plasmid DNA from Mathew et al. (2011) and prepared amino-terminal domain swaps between LhAQP1 and BtAQP1. To make EGFP-BtAQP1N-LhAQP1/pIB (chimeric LhAQP1 tagged at amino-terminus EGFP with 117 bp from 5′ of BtAQP1 replacing 207 bp from LhAQP1 5′ end), *Stu*I-EGFP F + BtAQP1N-LhAQP1 R and BtAQP1N-LhAQP1 F + 39LhAQP3 products were used as templates for overlap extension PCR with *Stu*I-EGFP F + 39LhAQP3. For EGFP-LhAQP1N-BtAQP1/pIB (chimeric BtAQP1 tagged at the amino-terminus with EGFP and with 207 bp from the 5′ end of LhAQP1 replacing the first 117 bp of BtAQP1), PCR products from *Stu*I-EGFP F + LhAQP1N-BtAQP1 R and LhAQP1-BtAQP1 F + 20BtAQP3 were templates for overlap extension PCR with *Stu*I-EGFP F + 20BtAQP3. The resulting products from final overlap extension PCR were sub-cloned and sequenced as above.

Transfections and fluorescent imaging were performed according to protocols largely based on those of Hull et al. (2012) and Mathew et al. (2011). Tni cells maintained as adherent cultures in serum-free insect culture media (Allele Biotechnology) at 28 °C were transfected in 35 mm #1.5 glass bottom dishes (In Vitro Scientific, Sunnyvale, CA) with 2 μg plasmid DNA using Insect GeneJuice transfection reagent (Novagen/EMD Biosciences, Madison, WI) for 5–12 h. Two days post transfection, media was replaced with IPL1-41 insect media (Invitrogen-Life Technologies) and cells were imaged using an Olympus FSX-100 fluorescence microscope with FSX-BSW imaging software (Olympus, Center Valley, PA). Images were processed using IrfanView (http://www.irfanview.com/) with gamma correction set at 1.5.

2.6. cRNA synthesis, Xenopus oocyte expression and permeability assays

LhAQP ORFs were sub-cloned into a modified *Xenopus laevis* β-globin plasmid expression vector (Anthony et al., 2000), linearized with *Sac*II, and transcribed *in vitro* (T3 mMessage mMachine; Ambion-Life Technologies). *In vitro*-transcribed complementary RNA (cRNA) corresponded to the full-length ORFs (residues 1–287 for LhAQP1, residues 1–262 for LhAQP2A, 1–267 for LhAQP3, 1–278 for LhAQP4A, 1–265 for LhAQP5). cRNA (20 ng μL^{-1}) was suspended in sterile water prior to injection.

Unfertilized *X. laevis* oocytes were defolliculated with collagenase (type 1A, 1.5 mg mL^{-1}; Sigma, St. Louis, MO) and trypsin inhibitor (15 mg mL^{-1}) in modified OR-2 saline [125 mM NaCl, 3.25 mM KCl, 2.5 mM $MgCl_2$, and 10 mM 4-(2-Hydroxyethyl) piperazine-1-ethanesulfonic acid (HEPES); pH 7.6] at 18 °C for 1.5 h, washed in modified OR2 saline solution, and maintained in modified ND96 saline (96 mM NaCl, 2 mM KCl, 5 mM $MgCl_2$, 0.6 mM $CaCl_2$, 5 mM HEPES, pH 7.6) supplemented with 100 μg/ml penicillin and 100 U/ml streptomycin, and 10% (V/V) heat-inactivated horse serum. Oocytes were injected with 50 nL of water containing 1 ng of LhAQP cRNA and were incubated for 2 or more days at 18 °C to allow protein expression.

For quantitative swelling assays, oocytes incubated in isotonic modified ND96 saline (without serum or antibiotics) for 1–2 h at room temperature were tested for water permeability by swelling in 50% hypotonic saline (isotonic diluted with an equal volume of water). Glycerol permeability was assayed by measuring swelling rates of oocytes in saline in which NaCl was partially substituted by glycerol [31 mM NaCl, 130 mM glycerol, 2 mM KCl, 5 mM $MgCl_2$ and 5 mM HEPES]. Swelling rates were quantified by relative increases in oocyte cross-sectional area imaged by video microscopy (charge-coupled device camera; Cohu, San Diego, CA) at 0.5 frames per second for 45 s using NIH ImageJ software. Rates were measured as described previously (Anthony et al., 2000; Boassa and Yool, 2003) using Prism (GraphPad Software Inc., San Diego, CA). Data for oocyte volume (V), standardized to the initial volume (V_0) as a function of time, were fit by linear regression to determine the swelling rate from the slope value $(V/V_0)(s^{-1} * 10^5)$. The University of Adelaide Animal Ethics committee approved all surgical procedures and the procedures followed protocols approved under the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes.

3. Results

3.1. Cloning, topology and homology of LhAQPs

RT-PCR and primer pairs 1LhAQP5 + 2LhAQP3, 3LhAQP5 + 4LhAQP3, 5LhAQP5 + 6LhAQP3, and 7LhAQP5 + 8LhAQP3 (Table 1) amplified four partial LhAQP fragments of 638, 759, 717, and 643 bp, respectively, from *L. hesperus* 2nd–3rd instar nymph and adult cDNA. The cloned fragments were sequenced and BLAST analysis revealed high similarity with the aquaporin family (data not shown). Using the sequences from the partial clones, we identified the 5′ and 3′ ends by RACE and full-length LhAQP1-4 cDNAs were PCR-amplified, cloned, and sequenced (Supplemental Fig. 1 and Table 2). Coincidental with the process of cloning the full-length LhAQP cDNAs, we completed a *de novo* assembled Roche 454 transcriptome of *L. hesperus* adults (Hull et al., 2013) that independently validated LhAQP1-4 transcripts and revealed an additional putative full-length aquaporin-like transcript, designated LhAQP5.

The 864-bp full-length coding sequence (CDS) of LhAQP1 (KF048092) corresponds to an open reading frame (ORF) encoding 287 amino acids (Supplemental Fig. 1 and Table 2). The predicted isoelectric point (pI) and molecular weight of LhAQP1 is 5.6 and 29.7 kDa, respectively. While no N- or C-glycosylation sites are predicted for LhAQP1, one O-glycosylation site is predicted at Thr19. However, because LhAQP1 has no predicted signal peptide, it is possible the mature protein is not exposed to glycosylation machinery. Like other members of the aquaporin superfamily (Murata et al., 2000), LhAQP1 has two Asn-Pro-Ala (NPA) motifs and residues predicted to comprise the ar/R constriction (Phe100, His224, Ser233, and Arg239) (Supplemental Figs. 1, Fig. 1, and Supplemental Table 1). Seven different TM/topology prediction programs revealed that LhAQP1 likely has six TM-spanning regions with intracellular amino- and carboxyl-termini (Table 3 and Supplemental Table 2), fitting the pattern known for other members of the aquaporin superfamily.

Nested 5′-RACE PCR using primers 24LhAQP3, 25LhAQP3, 34LhAQP3, and 35LhAQP3 revealed sequence variation resulting in different start sites for LhAQP2 (Supplemental Fig. 1, Fig. 2, and Table 2), including LhAQP2A (KF048093, 789-bp CDS encoding 262 amino acids), LhAQP2B (KF048094, 813-bp CDS encoding 270 amino acids), LhAQP2C (KF048095, 885-bp CDS encoding 294 amino acids), LhAQP2D (KF048096, 846-bp CDS encoding 281 amino acids), and LhAQP2E (KF048097, 768-bp CDS encoding 255 amino acids). While all variant LhAQP2s contain the 768-bp CDS that comprise LhAQP2E, RT-PCR and cloning verified that transcripts for all five full-length variant forms are present in *L. hesperus*. Predicted pIs for LhAQP2A-E are 5.1, 5.2, 4.9, 5.6, and 5.4, respectively. Molecular weights predicted by Compute pI/MW ExPASy tool for LhAQP2A-E are 27.9, 28.7, 31.2, 30.0, and 28.8 kDa, respectively. The same two putative N-glycosylation sites (NGSK and NLTG) were predicted in all LhAQP2 translated sequences. Several amino-terminal threonines (Thr3 for LhAQP2A-B, Thr4-5 for LhAQP2C, and Thr2 for LhAQP2E) are predicted to be O-glycosylated. No C-glycosylation sites were predicted for LhAQP2A-E. As for LhAQP1, all LhAQP2s are predicted to lack signal peptide, glycosylation may not occur on those predicted motifs. In contrast with many aquaporins that contain two NPA motifs, the LhAQP2s have one conserved amino-terminal NPA motif and a second motif comprised of Asn-Pro-Val (NPV). The four residues predicted to comprise the ar/R selectivity filter (Phe, Ser, Ala, and Arg) are conserved in all LhAQP2 translated sequences (Supplemental Fig. 1; Fig. 1, and Supplemental Table 1). Analysis of transmembrane topology indicates that LhAQP2A-E likely have six TM helices with intracellular amino and carboxyl termini, although Phobius predicts only five TM helices for all LhAQP2 translated proteins and RHYTHM and TOPCONS predict seven TM helices for LhAQP2C and LhAQP2D, respectively (Table 3 and Supplemental Table 2).

LhAQP3 (KF048098) has an 804-bp full-length CDS containing ORF encoding 267 amino acids (Supplemental Fig. 1 and Table 2). The predicted pI and molecular weight of LhAQP3 is 5.9 and 28.8 kDa, respectively. While no C-mannosylation sites are predicted, two N-glycosylation motifs (20-NLSM-23 and 93-NVTW-96) and three O-glycosylation sites (Thr141, Thr142, and Thr266) are predicted for LhAQP3. LhAQP3 has an Asn-Pro-Cys (NPC) motif at residues 89–91 and an NPA motif at residues 207–209 (Supplemental Figs. 1, Fig. 1, and Supplemental Table 1) with conserved ar/R constriction residues Phe69, Ala195, Gly204, and Arg210 (Supplemental Fig. 1 and Supplemental Table 1). Predicted TM topology for LhAQP3 is consistent with known patterns of AQP-folding, and all seven analyses predicted six TM-spanning regions and intracellular amino and carboxyl termini (Table 3 and Supplemental Table 2).

Two full-length LhAQP4 isoforms were identified by RT-PCR and cloning, with LhAQP4A (KF048099) having CDS of 837 bp encoding 278 amino acids and LhAQP4B (KF048100) having CDS of 798 bp

```
LhAQP4A         ---------------------------------------------MPDKRSRVNTLVGM 14
LhAQP4B         ---------------------------------------------MPDKRSRVNTLVGM 14
LhAQP1          --------MMGDHSSIDMPRSEEGVQTEKTPLLHKGSDPVSKLSKKKKNHADGSKSFLGT 52
LhAQP2B         ------------------------MATPLVLNRVTDSGGMTMSEFTNNDATIKVEEGKDT 36
LhAQP2C         MGVTTGSDVSQELTLSGAEIRVPEMATPLVLNRVTDSGGMTMSEFTNNDATIKVEEGKDT 60
LhAQP2E         ------------------------------------MTMSEFTNNDATIKVEEGKDT 21
LhAQP2A         -------------------------------MVTDSGGMTMSEFTNNDATIKVEEGKDT 28
LhAQP2D         ------------MNSGNSVFLASHDRIGRQPRVTDSGGMTMSEFTNNDATIKVEEGKDT 47
LhAQP3          ---------------------------------------MPPRSKARYTSLLRDDGFTNL 21
LhAQP5          ----------------------------------MGGIFGLIVSSAYIGLTCLIAWWAR 25

LhAQP4A         NELARAGDLG---KAVVAEALGTLFITYFGIMSCIA---LVPGNLVQISLCFGFVVMVSV 68
LhAQP4B         NELARAGDLG---KAVVAEALGTLFITYFGIMSCIA---LVPGNLVQISLCFGFVVMVSV 68
LhAQP1          EDVEKFDAVG---RALVAEFLGTMLLVVVGCGACVGSD-VAQPTTLSIALAFGFIIASVV 108
LhAQP2B         DSNGSKTDVGRYLEVFGAELVGTGVLLCIGCASCIAGDDDKPITDFHSALTFGFTVSTII 96
LhAQP2C         DSNGSKTDVGRYLEVFGAELVGTGVLLCIGCASCIAGDDDKPITDFHSALTFGFTVSTII 120
LhAQP2E         DSNGSKTDVGRYLEVFGAELVGTGVLLCIGCASCIAGDDDKPITDFHSALTFGFTVSAII 81
LhAQP2A         DSNGSKTDVGRYLEVFGAELVGTGVLLCIGCASCIAGDDDKPITDFHSALTFGFTVSTII 88
LhAQP2D         DSNGSKTDVGRYLEVFGAELVGTGVLLCIGCASCIAGDDDKPITDFHSALTFGFTVSTII 107
LhAQP3          SMIKVRRLLG----IGLAEMFGVALFLATGCGNLVSTISNSEPSHINTVLSFALGISSAI 77
LhAQP5          QLADRLQIES----NFKKLLLLEGIATWELCATCFELIIVADNYGVSTYALFLFVLT--I 79
                .       .          .   .         .          .     * : :   :

LhAQP4A         QALGHVSGGNLNPAVTCGLLITGRITIIRAALYIAAQCLGAIGGAAMAKFFTPE--DKVG 126
LhAQP4B         QALGHVSGGNLNPAVTCGLLITGRITIIRAALYIAAQCLGAIGGAAMAKFFTPE--DKVG 126
LhAQP1          QMIGHISGAHINPAVTIGLLACGRIGLIMSILYIPFQLAGAVAGAALLQYLVPE--LKKD 166
LhAQP2B         CIFGHISAAHLNPAVTVAFYVLGHINVPMMLVYFIAQIMGAIMGVGVLVLLTPEG-WIKS 155
LhAQP2C         CIFGHISAAHLNPAVTVAFYVLGHINVPMMLVYFIAQIMGAIMGVGVLVLLTPEG-WIKS 179
LhAQP2E         CIFGHISAAHLNPAVTVAFYVLGHINVPMMLVYFIAQIMGAIMGVGVLVLLTPEG-WIKS 140
LhAQP2A         CIFGHISAAHLNPAVTVAFYVLGHINVPMMLVYFIAQIMGAIMGVGVLVLLTPEG-WIKS 147
LhAQP2D         CIFGHISAAHLNPAVTVAFYVLGHINVPMMLVYFIAQIMGAIMGVGVLVLLTPEG-WIKS 166
LhAQP3          IIFAPISGCMLNPCLNVTWLVMGHMSLVKFVYYTVFQVIGAYLGVAFIAAVTPDIPEGPP 137
LhAQP5          WWSRNWGDATACPYTHFEDVAEGKKLIGQALVMILAQLAGGLLTYGYSQFLWSLE--VSA 137
                      .     *         *:  :         *  *.    .    . .

LhAQP4A         DLGATLLNPSITPGQGVAIEFMLGFVLVFVIFGVIDPNKPDAKIVAPLAIGLTV-AVGHL 185
LhAQP4B         DLGATLLNPSITPGQGVAIEFMLGFVLVFVIFGVIDPNKPDAKIVAPLAIGLTV-AVGHL 185
LhAQP1          AIGVVSVGAGLTEGQAFAVEAVITAVLLLVVCAVTDPNRSDLANSAPVAIGLAI-ACSHI 225
LhAQP2B         DMCVTQPHDNLTGYQAAGIEFVATMALIFLVCGLSDPRCAARQDSVPLKFAALL-IALSL 214
LhAQP2C         DMCVTQPHDNLTGYQAAGIEFVATMALIFLVCGLSDPRCAARQDSVPLKFAALL-IALSL 238
LhAQP2E         DMCVTRPHDNLTGYQAAGIEFVATMALIFLVCGLSDPRCAARQDSVPLKFAALL-IALSL 199
LhAQP2A         DMCVTQPHDNLTGYQAAGIEFVATMALIFLVCGLSDPRCAARQDSVPLKFAALL-IALSL 206
LhAQP2D         DMCVTQPHDNLTGYQAAGIEFVATMALIFLVCGLSDPRCAARQDSVPLKFAALL-IALSL 225
LhAQP3          GFCTTHPNPDVTTSQAFAVEFFLGFFLSLCLCHVLDKRASQQHGLVFAKFAVLV-VALAL 196
LhAQP5          NHRGRAYEACTADLQVPALIGAVIEGAATCVCRLASRFIAELNPAFGSELDAFIGTSLVV 197
                          :  *  .:           :  : .   .         :   :     :

LhAQP4A         SCIDSTGSSMNPARSLGSVAMIN---KWDNHWVYWVGPCMGGIVAAILYQFVLSAPPQEY 242
LhAQP4B         SCIDSTGSSMNPARSLGSVAMIN---KWDNHWVYWVGPCMGGIVAAILYQFVLSAPPQEY 242
LhAQP1          FAVPLTGSGMNPARSFGPAAMVG---QWDNQWIYWVAPLVGAACAGAVYRIIFRAPKDD- 281
LhAQP2B         AVGKYTGASLNPVRSLGPAIWTG---NWNQHWVYWVSPLSAGIVTPLFYTTCFLRGGQS- 270
LhAQP2C         AVGKYTGASLNPVRSLGPAIWTG---NWNQHWVYWVSPLSAGIVTPLFYTTCFLRGGQS- 294
LhAQP2E         AVGKYTGASLNPVRSLGPAIWTG---NWNQHWVYWVSPLSAGIVTPLFYTTCFLRGGQS- 255
LhAQP2A         AVGKYTGASLNPVRSLGPAIWTG---NWNQHWVYWVSPLSAGIVTPLFYTTCFLRGGQS- 262
LhAQP2D         AVGKYTGASLNPVRSLGPAIWTG---NWNQHWVYWVSPLSAGIVTPLFYTTCFLRGGQS- 281
LhAQP3          PLGKYEGGSANPARSLGPALLSG---DWNDQWLYWTAPNFGAAFAAVIYRAFFDAPIFED 253
LhAQP5          AAFNYSGGYFNPALATSLKLGCDGHTAEQHFFVYWIGSTVGAVASVFLFKNPEVRAKVFS 257
                      *.  **. : .     .     :. ::** ..  ..  :  .:

LhAQP4A         SQVPGESSTEVRIKVLPDESSSEMKRLGKGNENSNA 278
LhAQP4B         SQVPG-------------ESSTEMKRLGKGNENSNA 265
LhAQP1          ------------------DSSYDL------------ 287
LhAQP2B         ------------------------------------
LhAQP2C         ------------------------------------
LhAQP2E         ------------------------------------
LhAQP2A         ------------------------------------
LhAQP2D         ------------------------------------
LhAQP3          VDSDGKTR----------EVDATI------------ 267
LhAQP5          LLGKEKDL---------------------------- 265
```

Fig. 1. Amino acid sequence alignment of LhAQPs. The deduced amino acid sequences of the LhAQPs were aligned using CLUSTALW. NPA motifs are shown with grey highlight and residues that correspond to the ar/R constriction site (Phe56, His180, Cys189, and Arg195 from *Homo sapiens* AQP1, EAL24446.1) are shown in bold. A predicted mercury-sensitive cysteine (Cys86) from LhAQP3 is shown with white text on black highlight. Identical residues in all sequences are shown as "*" and partial identity or residues with conserved side chains are shown as ":" and ".".

Table 3
Prediction of intra/extracellular and transmembrane topology for *Lygus hesperus* aquaporins 1–5 (LhAQP1-5).

	No. residues	Number of predicted transmembrane helices by topology prediction tools[a]							Consensus position[b]	
		TMPred	TMHMM	Phobius	RHYTHM	TOPCONS	HMMTOP	TopPredII	N-terminus	C-terminus
LhAQP1	287	6	6	6	6	6	6	6	Inside	Inside
LhAQP2A	262	6	6	5	6	6	6	6	Inside	Inside
LhAQP2B	270	6	6	5	6	6	6	6	Inside	Inside
LhAQP2C	294	6	6	5	7	6	6	6	Inside	Inside
LhAQP2D	281	6	6	5	6	7	6	6	Inside	Inside
LhAQP2E	255	6	6	5	6	6	6	6	Inside	Inside
LhAQP3	267	6	6	6	6	6	6	6	Inside	Inside
LhAQP4A	278	6	6	6	6	6	6	6	Inside	Inside
LhAQP4B	265	6	5	6	6	6	6	6	Inside	Inside
LhAQP5	265	6	5	5	6	7	5	6	Outside	Inside

[a] Predictions of intra/extracellular domains and transmembrane helices were made using TMPred (Hofmann and Stoffel, 1993), TMHMM 2.0 (Krogh et al., 2001), Phobius (Käll et al., 2004), RHYTHM (Rose et al., 2009), TOPCONS (Bernsel et al., 2009), HMMTOP (Tusnady and Simon, 2001), and TopPredII (Claros and von Heijne, 1994).
[b] The amino- and carboxyl-terminal ends of LhAQPs were predicted to be either intracellular or extracellular based on consensus positions from the seven different prediction software programs.

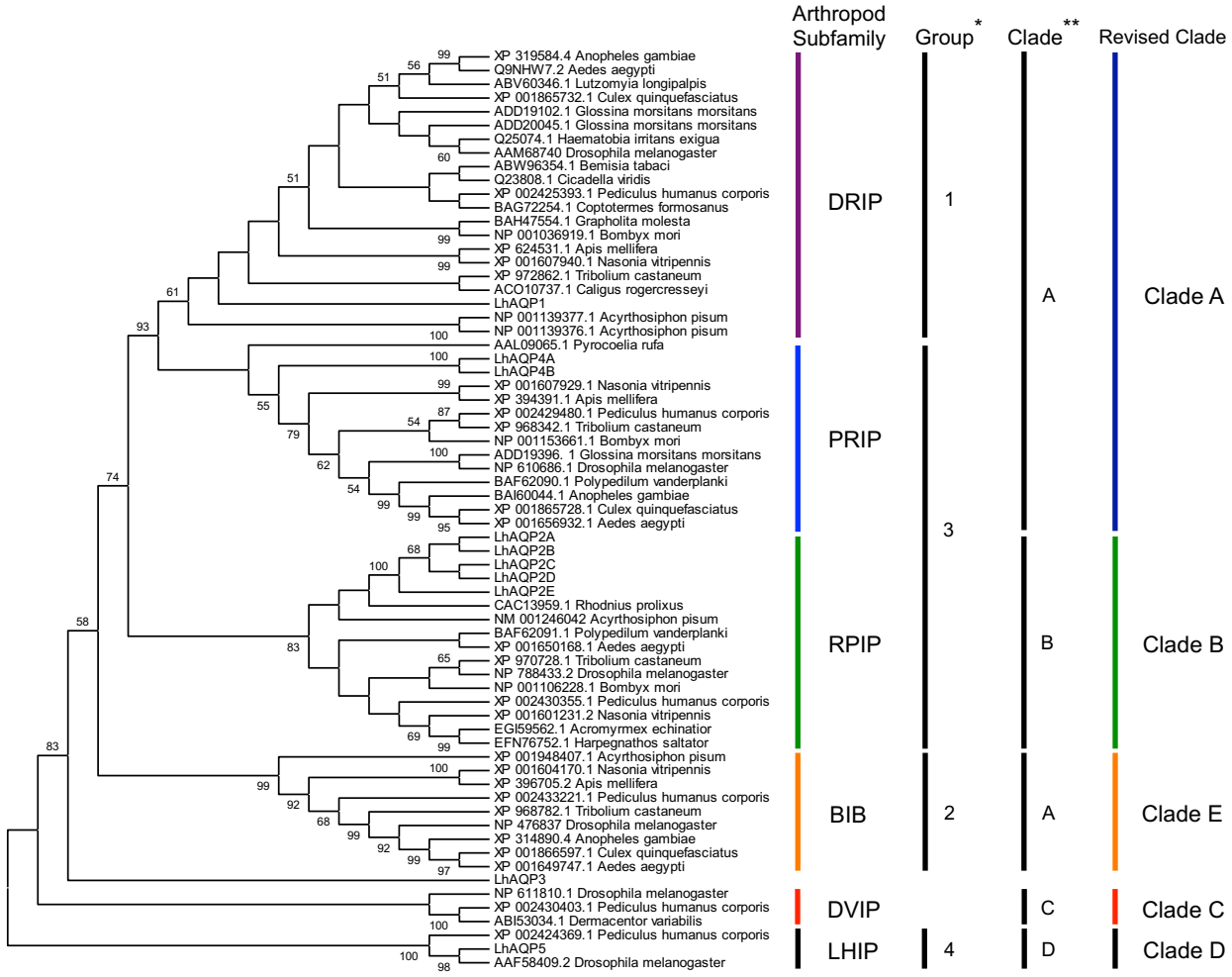

Fig. 2. Phylogenetic analysis and classification of LhAQPs and other arthropod aquaporins. Translated LhAQP sequences were compared with other arthropod aquaporins and the evolutionary history was inferred using the UPGMA method (Sneath and Sokal, 1973) in MEGA5 (Tamura et al., 2011). The bootstrap consensus tree inferred from 10,000 replicates (Felsenstein, 1985) and branches corresponding to partitions reproduced in less than 50% of bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (10,000 replicates) are shown next to the branches. The evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) and are in units of the number of amino acid substitutions per site. The analysis involved 67 amino acid sequences. All positions containing gaps and missing data were eliminated. A total of 147 positions were in the final dataset. Taxonomic name and accession number for each sequence is given. Representatives from six major arthropod subfamilies (DRIPs, PRIPs, RPIPs, BiBs, DVIPs, and LHIPs) are shown. The proposed subfamily names follow the nomenclature proposed by Campbell et al. (2008) and includes *Drosophila* intrinsic proteins DRIPs, *Pyrocoelia rufa* integral proteins PRIPs, *Rhodnius prolixus* integral proteins RPIPs, Big Brain BIBs, *Dermacentor variabilis* integral proteins DVIPs, and *Lygus hesperus* integral proteins LHIPs. The group and clade (as indicated by superscript "*" and "**") correspond to the four groups (1–4) [by Kambara et al. (2009) and Goto et al. (2011)] or clades (A–D) [by Wallace et al. (2012)], respectively. Revised clades A-E are indicated.

encoding 265 amino acids. The predicted pI and molecular weight of LhAQP4A is 6.3 and 29.1 kDa, respectively, and 6.8 and 27.6 kDa for LhAQP4B. The CDS for LhAQP4A and LhAQP4B are identical except for 39 bp (positions 757–796) that are missing from LhAQP4B. This difference likely results from alternative splicing of an exon encoding 13 amino acids. The LhAQP4B transcript is either more abundant in *L. hesperus* and/or preferentially cloned over LhAQP4A, as 11 of the cloned 47LhAQP5 + 48LhAQP3 PCR products corresponded to LhAQP4B versus none for LhAQP4A. LhAQP4A is, however, present in *L. hesperus*, as transcripts were found by RACE, overlap extension PCR, and 454 transcriptome sequencing. Because LhAQP4A and LhAQP4B only differ after residue 252, both have two conserved NPA motifs at positions 80–82 and 196–198 and ar/R residues Phe60, His184, Ser193 and Arg199 (Supplemental Figs. 1, Fig. 1, and Supplemental Table 1). TM topology analysis indicated the presence of six TM-helices for LhAQP4A and 5-6 helices for LhAQP4B, with only TMHMM showing five and a weak probability score for a sixth putative TM-spanning region corresponding to residues 170–190 in LhAQP4B (Table 3 and Supplemental Table 2).

The 798-bp CDS of LhAQP5 (KF048101) encodes 265 amino acids (Supplemental Fig. 1 and Table 2) with a predicted molecular weight of 28.9 kDa and a pI of 5.6. No C-, N- or O-glycosylation sites are predicted for LhAQP5. While the size of the translated LhAQP5 and presence of two NPA motifs is consistent with other insect AQPs (Campbell et al., 2008), the positioning and sequence conservation of the motifs differ from those seen in previously characterized insect AQPs. Based on alignment with the other LhAQPs, the residues comprising what would be the LhAQP5 amino-terminal NPA motif are instead Cys-Pro-Tyr (Supplemental Figs. 1, Fig. 1, and Supplemental Table 1). An NPA motif sequence is located 89 residues downstream at position 180–182, however, the spatial positioning of this motif is not consistent with the classical two-NPA motif "hour-glass" aquaporin folding pattern (Jung et al., 1994). A second NPA motif, conserved in the carboxyl terminal domain in other aquaporins, is located at position 208–210. Furthermore, other than Phe73, the predicted ar/R constriction residues are not conserved with orthodox AQPs (Val196, Gly205, and Leu211) (Supplemental Figs. 1, Fig. 1, and Supplemental Table 1). TM topology predictions for LhAQP5 varied, but the consensus topology from TmPred, TMHMM, RHYTHM, HMMTOP, and TopPredII predicted six TM-spanning regions with intracellular amino and carboxyl termini (Table 3 and Supplemental Table 2). Phobius predicted five TM helices and TOPCONS predicted seven TM regions (Table 3 and Supplemental Table 2).

We used CLUSTALW and MEGA 5.1 UPGMA, Maximum Likelihood, Maximum Parsimony, and Neighbor-Joining phylogenetic analysis to compare LhAQPs with other arthropod aquaporins (Fig. 2). Invertebrate aquaporins were previously classified by Campbell et al. (2008) and Mathew et al. (2011) into at least three subfamilies (*Drosophila* intrinsic proteins - DRIPs, Big Brain proteins - BIBs, and *Pyrocoelia rufa* integral proteins - PRIPs). More recently they were grouped into four major clusters [groups 1–4 by Kambara et al. (2009) and Goto et al. (2011)] or clades A–D (Wallace et al., 2012). Sequence alignment and phylogenetic analysis of full-length LhAQP protein sequences with 57 other invertebrate aquaporin sequences retrieved from GenBank showed the existence of at least six distinct families of arthropod AQPs, including DRIPs, PRIPs, *R. prolixus* integral proteins – RPIPs, BIBs, *Dermacentor variabilis* integral proteins – DVIPs, and *Lygus hesperus* integral proteins – LHIPs (Fig. 2). While analysis by Wallace et al. (2012) includes BIBs, DRIPs, and PRIPs within a single clade (e.g., A), analysis here and by others (Kambara et al., 2009; Drake et al., 2010; Goto et al., 2011; Herraiz et al., 2011) indicates that BIBs are sufficiently diverged from DRIPs and PRIPs to warrant classification into a separate clade, which we term clade E using the nomenclature of Wallace et al. (2012) (Fig. 2). Here, we classify six subfamilies based on the nomenclature of Campbell et al. (2008) that are contained within five clades (A–E). Whereas, clades B, C, and D are identical to those from Wallace et al. (2012), the clade A described here contains both the DRIP and PRIP subfamilies and does not contain the BIBs (clade E). Furthermore, of the four AQP groups from Kambara et al. (2009) and Goto et al. (2011), group 1 corresponds to subfamily DRIP, group 2 corresponds to BIB, group 3 contains both PRIPs and RPIPs, and group 4 corresponds to the new subfamily LHIP (Fig. 2).

Based on phylogenetic analysis, the LhAQP proteins segregated into different subfamilies comprising DRIPs (LhAQP1), PRIPs (LhAQP4A and LhAQP4B), RPIPs (LhAQP2A-E), and LHIPs (LhAQP5). LhAQP3 is unique and appears to fall outside currently classified aquaporin-like sequences (Campbell et al., 2008; Kambara et al., 2009; Goto et al., 2011) and may represent a novel subfamily/clade of AQPs. Furthermore, while a BIB-like protein has been found in the hemipteran *A. pisum*, no lygus BIB-like sequence has yet been identified.

LhAQP1 aligned within the DRIP subfamily of clade A and overall shares 42% identity with members of that subfamily. LhAQP1 is most homologous (46% identity) to the *B. mori* (NP_001036919.1), *A. aegypti* (Q9NHW7.2), and *Coptotermes formosanus* (BAG72254.1) DRIP-like aquaporins. LhAQP4A and LhAQP4B aligned within the PRIP subfamily, which also belongs to clade A (Fig. 2). LhAQP4B shares 41% identity across the PRIP subfamily and is most closely related with aquaporins from *Tribolium castaneum* (XP_968342.1) (45% identity), *Pediculus humanus corporis* (XP_002429480.1) (45% identity), *Culex quinquefasciatus* (XP_001865728.1) (44% identity), and *B. mori* (NP_001153661.1) (43% identity).

The LhAQP2s aligned within the RPIP aquaporin subfamily or clade B (Fig. 2). LhAQP2E shared an average of 36% identity with other RPIPs, with the most closely related sequences corresponding to aquaporins from *R. prolixus* (CAC13959.1) and *A. pisum* (NM_001246042.2) having 42 and 35% identity, respectively.

LhAQP5 aligns within subfamily LHIP (clade D) and includes other putative aquaporin sequences that lack functional characterization (Fig. 2). LhAQP5 shares 41 and 52% identity with *P. humanus corporis* (XP_002424369.1) and *D. melanogaster* (AAF58409.2), respectively. BLAST analysis (data not shown) reveals that these LHIP subfamily members share significant sequence similarity with vertebrate S-aquaporins 11 and 12, which, although not well characterized, have highly deviated NPA motifs and are thought to function in intracellular trafficking of water (Ishibashi, 2006, 2009).

3.2. Temporal expression and tissue localization

Transcriptional expression patterns of the LhAQPs were investigated using end-point RT-PCR over the life cycle of *L. hesperus*, including eggs, nymphs (1st, 2nd–3rd, 4th–5th instars), and adults. LhAQP1 transcripts are present in all development stages and most highly abundant in eggs (Fig. 3A). LhAQP2 and LhAQP5 transcripts have similar profiles through development, with little or no transcripts evident in eggs, but constitutive expression in all nymphal instars and adults. LhAQP3 and LhAQP4 transcripts are constitutively present in all developmental stages.

End-point RT-PCR using RNA extracted from different adult tissues show that LhAQP transcripts are differentially expressed (Fig. 3B). All five LhAQPs were found in midguts and transcripts for all but LhAQP5 were present in Malpighian tubules. LhAQP1 transcripts were the only LhAQP abundant in ovaries. LhAQP2 and LhAQP3 transcripts were found primarily in midguts and Malpighian tubules. LhAQP4 transcripts were most abundant in male testes and Malpighian tubules. Surprisingly, in addition to midguts,

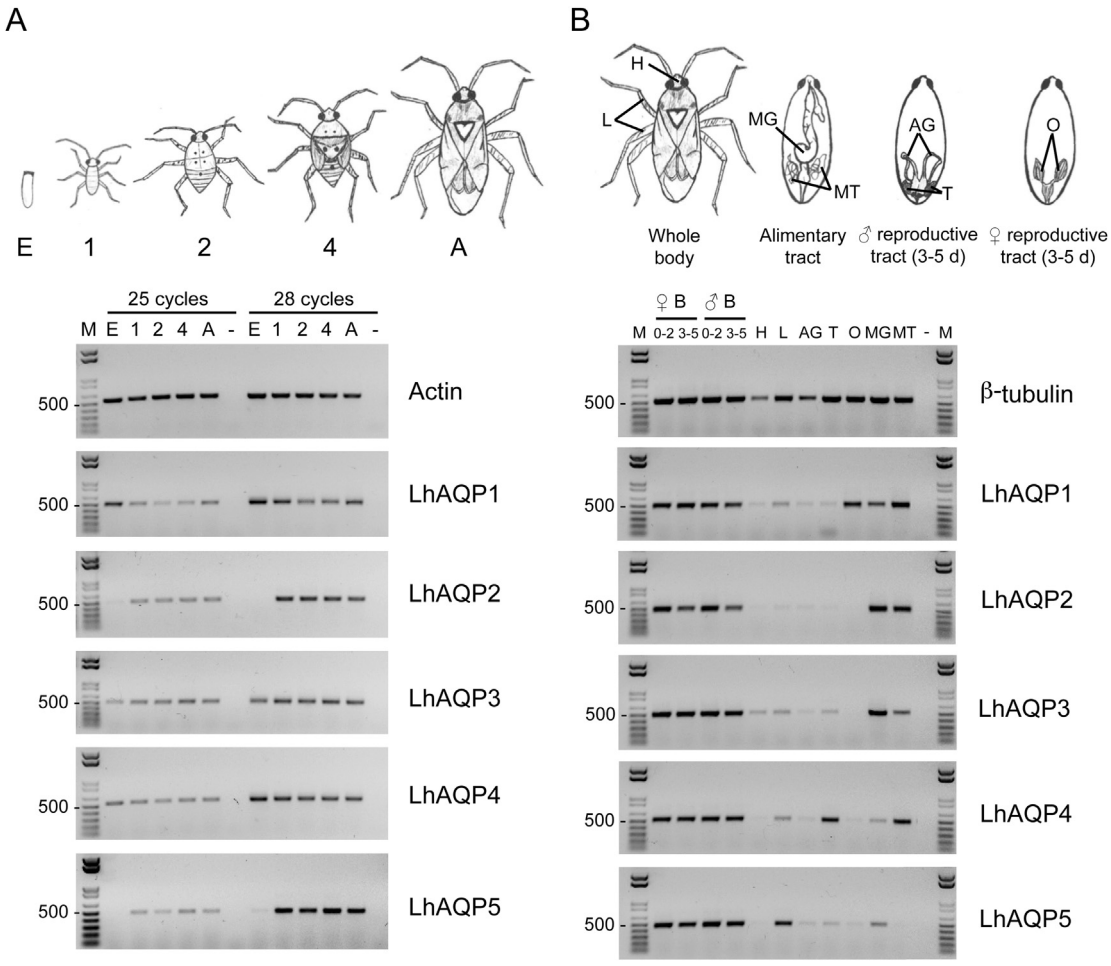

Fig. 3. Transcriptional profiles of LhAQPs through development and in different tissues. (A) Transcriptional expression of all five LhAQPs in *L. hesperus* developmental stages. cDNA was prepared from 1.0 μg of DNase-treated total RNA from eggs (E), 1st instar nymphs (1), 2nd–3rd instar nymphs (2), 4th–5th instar nymphs (4), and adults (A). The appropriate annealing temperatures and number of RT-PCR cycles was determined empirically with 25 and 28 cycles shown. Actin was amplified as an internal reference gene and "–" designates a no template control. (B) Transcriptional expression profile of all five LhAQPs from dissected *L. hesperus* tissues. cDNA made from female and male whole bodies (0–2 d or 3–5 d old adults), heads (H), legs (L), accessory glands (AG), testis (T), ovaries (O), midguts (MG), and Malpighian tubules (MT) were analyzed by RT-PCR (results with 28 cycles is shown). β-tubulin was used as an amplification control for tissues. Products were electrophoresed on 0.8% agarose gels and visualized using SYBR Safe (Invitrogen). Lane M corresponds to 1 kb Plus DNA Ladder (Invitrogen).

LhAQP5 transcripts were present in leg tissue. Faint bands were also visible in leg tissue for LhAQP1, LhAQP3, and LhAQP4.

3.3. Cellular localization

All arthropod aquaporins characterized to date involve intracellular protein synthesis and subsequent translocation to plasma membranes, where they exhibit their transport function (Campbell et al., 2008; Cohen, 2013). WOLF PSORT (Horton et al., 2007) predictions of subcellular localization of each translated sequence were consistent with presumed functionality; all LhAQPs were predicted to be associated with the plasma membrane (data not shown). However, LhAQP2A, LhAQP2E, LhAQP3, LhAQP4A, LhAQP4B, and LhAQP5 either had potential endoplasmic reticulum (ER) retention signals, or in the case of LhAQP2A, LhAQP2E, and LhAQP3, had PSORTII k-NN scores for ER equal to or above those for the plasma membrane (data not shown).

To examine cellular localization of each LhAQP, chimeric recombinant LhAQP proteins tagged with enhanced green fluorescent protein (EGFP) were produced in cultured Tni cells. First, we made LhAQP/pIB constructs with carboxyl-terminal-tagged EGFP chimeric proteins (Fig. 4A, products #1–11). Examination of transfected cells by fluorescence microscopy showed that LhAQP2A-EGFP, LhAQP2B-EGFP, LhAQP2C-EGFP, LhAQP2D-EGFP, LhAQP2E-EGFP, LhAQP3-EGFP, and LhAQP4B-EGFP recombinant chimeras were produced and transported primarily to the plasma membrane in Tni cells (Fig. 4B). LhAQP1-EGFP, LhAQP4A-EGFP, and LhAQP5-EGFP were not detectable at the cell surface, but rather were primarily intracellular with distribution patterns consistent with either localization in the Golgi apparatus (Fig. 4B, panel c and panel j for LhAQP1-EGFP and LhAQP4A-EGFP, respectively) or the ER (Fig. 4B, panel l for LhAQP5-EGFP).

Our localization results for LhAQP1-EGFP were unexpected, as other insect DRIP-like aquaporins are associated with the plasma membrane (Le Caherec et al., 1997; Duchesne et al., 2003; Liu et al., 2011; Mathew et al., 2011; Azuma et al., 2012) where they facilitate transport of cellular and extracellular water. A Met3-LhAQP1-EGFP/pIB construct (Fig. 4A, product #11) was made to test whether a truncated LhAQP1 produced using the methionine located at base 27 as the start codon would influence Tni cellular localization. Like LhAQP1-EGFP, Met3-LhAQP1-EGFP was intracellular (Fig. 4C, panel a) and exhibited Golgi-like localization (Kawar and Jarvis, 2001). Furthermore, placement of EGFP at the amino-terminus of LhAQP1 (EGFP-LhAQP1, Fig. 4A, product #12) did not alter the translocation

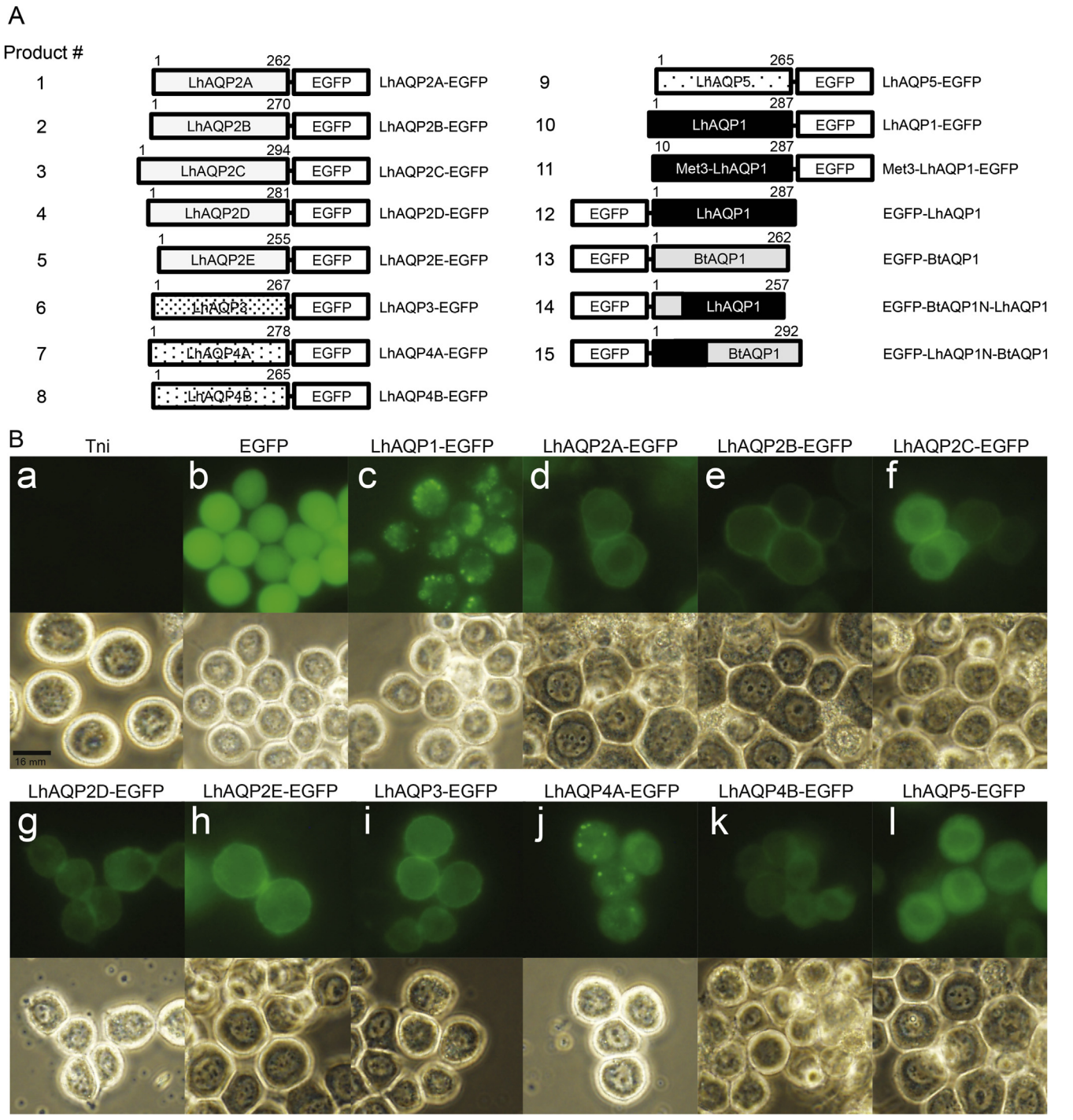

Fig. 4. Cellular localization of LhAQPs heterologously expressed in cultured Tni cells. (A) Schematic diagram of LhAQPs produced as EGFP chimeras in cultured Tni cells. Recombinant LhAQPs were expressed as chimeras with the enhanced green fluorescent protein (EGFP) using the pIB/V5-His TOPO TA (Invitrogen-Life Technologies) insect cell expression vector in Tni cells. Whereas LhAQP2-5 were all produced with EGFP at their carboxyl termini (products #1–9), recombinant LhAQP1 was produced with EGFP either at the carboxyl terminus (Product #10) or at the amino terminus (product #12). Product #11 corresponds to a truncated LhAQP1 beginning at base position 28 (corresponding to third methionine in translated LhAQP1) with EGFP at the amino terminus. Domain swapping was performed between LhAQP1 and *B. tabaci* AQP1 (BtAQP1) from Mathew et al. (2011) with product #13 corresponding to the full-length BtAQP1 chimera with EGFP at the amino terminus. Product #14 corresponds to LhAQP1 with 69 residues from the amino terminus replaced by the first 39 residues from BtAQP1. Product #15 corresponds to a chimeric protein having the first 39 residues of BtAQP1 replaced with 69 residues from the amino terminus of LhAQP1 produced in frame with the final 223 residues from BtAQP1. (B) Expression of recombinant carboxyl-terminal EGFP-labeled LhAQPs in Tni cells. Cells were transfected with 2 μg pIB plasmid DNA corresponding to LhAQP1-5-EGFP (panels c-l) using Insect GeneJuice transfection reagent (Novagen/EMD Biosciences) for 5–12 h with microscopic imaging (top image detecting fluorescence and bottom image is phase contrast) performed two days post transfection. Panel (a) shows untransfected Tni cells and panel (b) shows cells transfected with EGFP/pIB. Panels (c-l) show cells transfected with LhAQP-EGFP/pIB plasmid DNA. All proteins were produced as chimeras with carboxyl-terminal EGFP. (C) Domain swapping of LhAQP1 and BtAQP1 amino termini and expression in Tni cells. Panel (a) corresponds to Met3-LhAQP1 (10 residues truncated from LhAQP1). Panels (b) and (c) show localization of amino-terminal tagged LhAQP1 and BtAQP1, respectively. Panel (d) shows expression of LhAQP1 with the amino terminus of BtAQP1 is at the cell surface, whereas panel (e) shows expression of BtAQP1 with the amino terminus of LhAQP1 is subcellular. All cells were imaged using an Olympus FSX-100 fluorescence microscope with FSX-BSW imaging software (Olympus, Center Valley, PA) and processed using IrfanView with gamma correction set at 1.5.

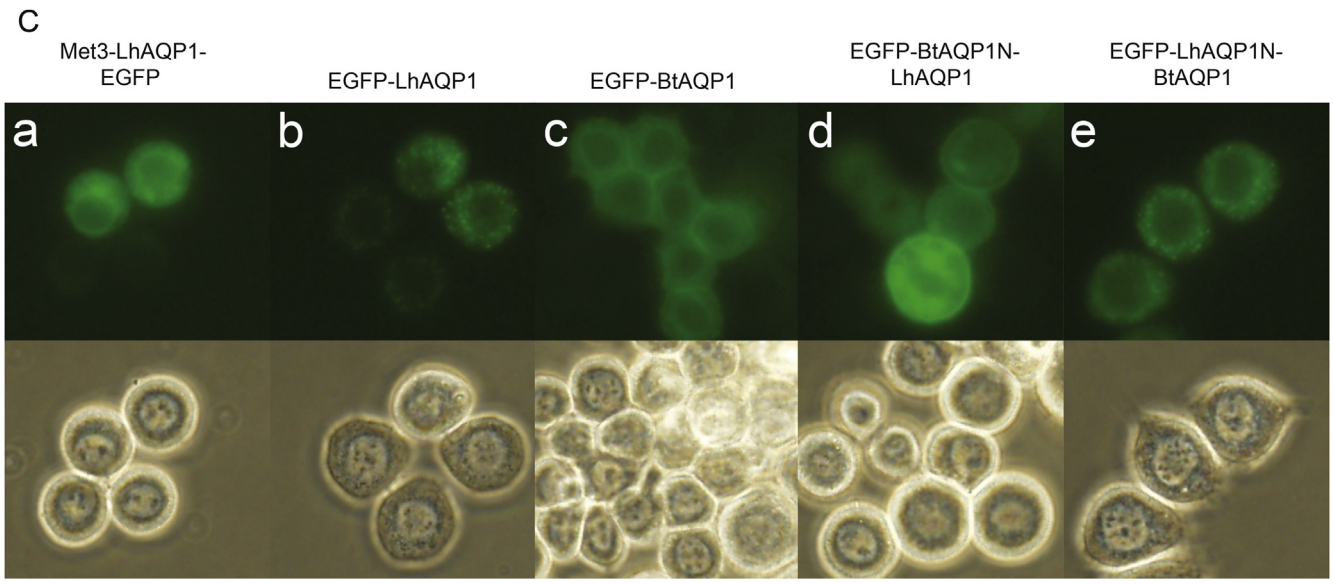

Fig. 4. (*continued*).

of LhAQP1, indicating the position of EGFP was not responsible for the observed intracellular localization (Fig. 4C, panel b).

To test the influence of the amino terminus of LhAQP1 on cellular localization, we produced EGFP chimeras that switched the first 69 amino acid residues from LhAQP1 with the first 39 residues from the *B. tabaci* aquaporin 1 (BtAQP1) and vice versa. Whereas the addition of the amino terminus of BtAQP1 to LhAQP1 (EGFP-BtAQP1N-LhAQP1) rescued localization to the cellular plasma membrane in Tni (Fig. 4C, panel d), the addition of the LhAQP1 amino terminus to BtAQP1 (EGFP-LhAQP1N-BtAQP1) resulted in impaired translocation, with intracellular fluorescence primarily limited to small punctae reminiscent of the Golgi apparatus (Fig. 4C, panel e). These results indicate that the first 69 residues at the amino terminus of LhAQP1 are responsible for intracellular localization of LhAQP1 in Tni cells.

3.4. Oocyte functional analysis

The functions of LhAQPs in water and glycerol transport were tested using the *Xenopus* oocyte expression system, measuring volume changes indicative of net influx of water or glycerol in hypotonic and isotonic solutions, respectively (Fig. 5A). We observed significantly higher swelling rates ($p < 0.01$) in LhAQP1, 2, 3, 4, and 5-injected oocytes in hypotonic saline (mean ± SEM 14.4 ± 2.5 for LhAQP1, 14.2 ± 1.2 for LhAQP2A, 27.1 ± 2.6 for LhAQP3, 19.0 ± 1.8 for LhAQP4A, and 13.4 ± 0.6 for LhAQP5 $V/V_0 \times 10^5 \times sec^{-1}$) compared to water-injected control oocytes (5.1 ± 0.5 $V/V_0 \times 10^5 \times sec^{-1}$) (Fig. 5B). Furthermore, none of the LhAQP-injected oocytes showed appreciable transport of glycerol, based on the absence of swelling. Instead, observed negative slopes (Fig. 5A) indicated a slight shrinkage in the glycerol-substituted saline (−3.1 ± 1.3 for LhAQP1, −5.6 ± 2.6 for LhAQP2A, −3.1 ± 1.3 for LhAQP3, −5.8 ± 1.2 for LhAQP4A, and −6.4 ± 1.1 for LhAQP5; $V/V_0 \times 10^5 \times sec^{-1}$), indicating relative aquaporin impermeability to glycerol.

4. Discussion

Aquaporins are integral membrane proteins belonging to a large family of water channel proteins or MIPs that facilitate transport water and/or other small solutes across cellular membranes (Benga, 2009; Kruse et al., 2006). Because water and solute transport are universal requirements for living cells, these proteins are found in most organisms. The total number of invertebrate AQPs identified in insects is rapidly increasing, with genome sequencing implicating at least 5–8 genes per organism encoding putative AQPs (Campbell et al., 2008; Drake et al., 2010). High-throughput next generation transcriptomic sequencing and *de novo* assembly provides an opportunity to obtain comprehensive "snapshots" of active gene transcripts in organisms without prior genomic support (Hull et al., 2013; Grabherr et al., 2011). Here, we gleaned the transcriptomes of two mirid plant bugs, *L. hesperus* (Hull et al., 2013) and *L. lineolaris* (O.P. Perera, unpublished), to identify expressed sequence tag (EST) sequences encoding putative AQPs.

Arthropods have multiple functional AQPs, and based on other animal systems these proteins likely play diverse physiological roles (Campbell et al., 2008; Gomes et al., 2009). Unfortunately, for most invertebrates comprehensive molecular and functional characterization of AQPs trails their initial identification. For example, *D. melanogaster* has eight putative AQPs, but only two have been assigned functions (Kaufmann et al., 2005; Yanochko and Yool, 2002). *A. aegypti* has six putative AQPs and RNAi knockdown was used to show that three have apparent roles in diuresis in Malpighian tubules (Drake et al., 2010). However, the roles of those transcripts in mosquitoes not exhibiting an obvious phenotype are yet to be defined, and biochemical characterization of all AeAQPs is not complete. This report attempts to identify and characterize a complement of AQPs from the hemipteran pest, *L. hesperus*.

Five full-length LhAQPs that likely represent the primary AQP transcripts from *L. hesperus* were cloned and characterized. We used phylogenetic alignments of LhAQPs with other arthropod representatives to classify at least six subfamilies of arthropod AQPs, including three (DRIPs, PRIPs, and BIBs) previously described by Campbell et al. (2008). Three new subfamilies [named using nomenclature described by Campbell et al. (2008)] include the RPIPs (named for *R. prolixus* integral protein), DVIPs (named for *D. variabilis* integral protein), and LHIPs (named for *L. hesperus* integral protein). Within these phylogenetic subfamilies, three functional classes of AQPs have been characterized for arthropods, including water-specific AQPs, aquaglyceroporins, and monovalent cation channels (BIBs) (Campbell et al., 2008; Echevarria et al.,

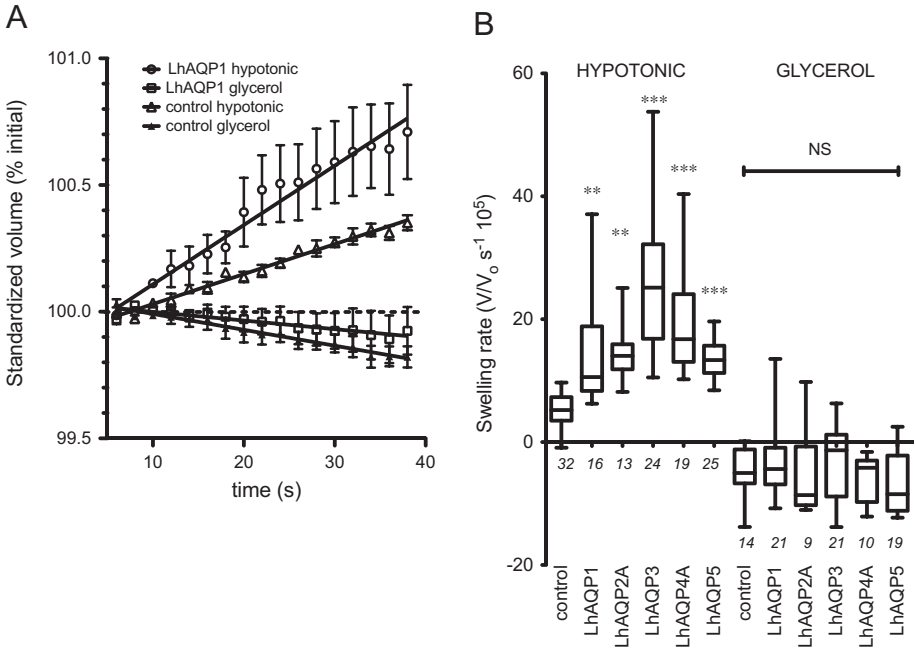

Fig. 5. Osmotic water and glycerol permeability of *Xenopus* oocytes expressing LhAQPs. (A) Representative examples of mean swelling rates (±SEM) of control and LhAQP1-expressing oocytes as a function of time after introduction of oocytes into 50% hypotonic saline or isotonic glycerol-substituted saline at time zero. Volumes of the oocytes at each time point (V) are standardized to the initial area at time zero (Vo). (B) Compiled swelling rates (slope values of linear regression fits of V/Vo versus time) showed significant osmotic water permeability of LhAQP-expressing oocytes (LhAQP1-5) compared with control oocytes (ANOVA and post-hoc Bonferroni tests, with ** indicating $p < 0.01$ and ***$p < 0.001$). No significant glycerol permeability was found using saline in which 130 mM glycerol replaced 65 mM NaCl. Box plots show 25th and 75th quartiles (box), median values (solid horizontal line within box), and error bars span the full range of data points. Sample sizes (n) are shown near the x-axis.

2001; Duchesne et al., 2003; Le Caherec et al., 1996; Shakesby et al., 2009; Mathew et al., 2011; Kataoka et al., 2009a; Wallace et al., 2012; Yanochko and Yool, 2002). An important distinction is that the functional classes do not necessarily correspond with phylogenetic classifications. For example, the RPIP subfamily includes some AQPs that function to transport water and glycerol/urea and hence are aquaglyceroporins (Fig. 2). But, not all members of the RPIP subfamily transport glycerol and are therefore not functional aquaglyceroporins (includes PvAQP2 in Kikawada et al., 2008; the LhAQP4 isoforms, Fig. 5) [see Supplemental Table 1 for phylogenetic and functional classification of representative invertebrate AQPs].

Phylogenetic analysis assigned the LhAQP proteins to different subfamilies, including DRIPs (LhAQP1), PRIPs (LhAQP4A and LhAQP4B), RPIPs (LhAQP2A-E), and LHIPs (LhAQP5). The arthropod DRIP subfamily includes a number of water-specific AQPs generally promoting fluid homeostasis by moving water through tissues to facilitate excretion and/or maintenance of osmotic potential (Campbell et al., 2008; Mathew et al., 2011). Based on conserved AQP motifs and sequence similarity (Fig. 2), predicted topology (Table 3 and Supplemental Table 2), and formation of water-specific pores in *Xenopus* oocytes (Fig. 5), LhAQP1 is a water-specific AQP from the DRIP subfamily. Two LhAQP4 isoforms belong to the arthropod PRIP subfamily which includes AQPs with diverse physiological functions, such as water recycling in the cryptonephric rectal complex (Azuma et al., 2012), hydration and dehydration responses to environmental stress (Kikawada et al., 2008), and water and nitrogen homeostasis during oogenesis (Herraiz et al., 2011). The LhAQP2 isoforms belong to the RPIP arthropod subfamily, which contains a number of AQPs with putative multifunctional transport channel functions, including the aquaglyceroporins. Members of the RPIP subfamily are important for osmoregulation and *Buchnera* symbiosis in aphids (Wallace et al., 2012), water homeostasis in fat body (Kikawada et al., 2008), urine formation (Echevarria et al., 2001), and water homeostasis/urea excretion in midgut and Malpighian tubules (Kataoka et al., 2009a). We classify LhAQP5 into a new subfamily of arthropod AQPs called the LHIPs. These proteins share similarity with 'superaquaporins' (Morishita et al., 2004) or "S-aquaporins" (Ishibashi, 2006), which are quite diverged from other families of AQPs. These divergences include altered NPA motifs and anomalous subcellular localization (Ishibashi, 2006, 2009; Nozaki et al., 2008; Calvanese et al., 2013). Although LhAQP3 transports water (and not glycerol) in *Xenopus* oocytes, it falls outside currently classified aquaporin-like sequences and may represent an additional unclassified subfamily of water-transporting AQPs (Fig. 2). Contigs encoding a BiB-like protein were not found in our search of the *L. hesperus* or *L. lineolaris* transcriptomes.

Two canonical NPA motifs along with residues that comprise the ar/R constriction site (Sui et al., 2001; Beitz et al., 2006) are critical for determining the selectivity for solutes that transverse the AQP channel. Of the LhAQPs, only LhAQP1 and LhAQP4 have two conserved NPA motifs (Fig. 1 and Supplemental Table 1). Both LhAQP3 and LhAQP5 have substitutions within the first NPA motif (Asn-Pro-Cys for LhAQP3; Cys-Pro-Tyr for LhAQP5), and for LhAQP2 the second NPA motif is Asn-Pro-Val (Fig. 1 and Supplemental Table 1). In human AQPs, four residues (Phe56, His180, Cys189, and Arg195 from AQP1 and Phe77, His201, Ala210, and Arg216 from AQP4) form the narrowest restriction point within the AQP water channel that is only slightly larger than size of a single water molecule (Sui et al., 2001; Beitz et al., 2006; Ho et al., 2009). The corresponding ar/R constriction residues for LhAQPs include Phe/His/Ser/Arg for LhAQP1, Phe/Ser/Ala/Arg for LhAQP2, Phe/Ala/Gly/Arg for LhAQP3, Phe/His/Ser/Arg for LhAQP4, and Phe/Val/Gly/Leu for LhAQP5 (Fig. 1 and Supplemental Table 1). Wallace et al. (2012) reported that insect MIPs (including some aquaglyceroporins) within the RPIP subfamily (clade B from Wallace et al., 2012) have small neutral residues at position H5 that contribute to a wider, more hydrophilic ar/R constriction, which is required for passage of

larger solutes. However, we found that none of the LhAQPs facilitate the transport of glycerol in *Xenopus* oocytes, including LhAQP2, LhAQP3, or LhAQP5, which contain similar ar/R residues to those found in other insect aquaglyceroporins. Therefore, insect AQPs may have diverse functional roles that are not necessarily predicted by sequence conservation. Residues yet to be defined could additionally influence permeation pathways in the AQP channel and participate in defining solute permeabilities.

Levels of LhAQP transcripts differed little during *L. hesperus* development with all five apparently constitutively present in nymphs and adults (Fig. 3A). LhAQP1, 3, and 4 transcripts were detected in eggs (Fig. 3A), indicating a possible role in embryonic development. It is not known whether these LhAQP transcripts were embryonic or maternally derived. LhAQP transcripts are abundant in various adult *L. hesperus* tissues (Fig. 3B), with all LhAQPs present in Malpighian tubules and LhAQP1-4 present in the midgut. These results are consistent with reports of other AQPs associated with fluid absorption and excretion in arthropods (Spring et al., 2009; Le Caherec et al., 1997; Echevarria et al., 2001; Kaufmann et al., 2005; Kataoka et al., 2009a; Liu et al., 2011; Mathew et al., 2011; Azuma et al., 2012). LhAQP1 is the only transcript found in ovaries, and LhAQP4 is abundant in male testes (Fig. 3B). The presence of LhAQPs in reproductive tissues is not unprecedented, as 12 of 13 known mammalian AQPs are expressed in female and male reproductive systems [reviewed in Zhang et al., (2012)]. Vertebrate AQPs have numerous known functions in males, including spermatogenesis, seminal fluid production, and fluid transport in testis and numerous ducts and tubules (Zhang et al., 2012). Vertebrate females have AQPs that function in tubal fluid transport, water homeostasis of the uterus, cervix, placenta, and oviduct, maintenance of amniotic fluid, gamete transport and fertilization, early embryo development, and implantation (Zhang et al., 2012). Arthropod AQPs may also function in reproduction, including those found in ovaries of the tick *D. variabilis* (Holmes et al., 2008), the cockroach *B. germanica* (Herraiz et al., 2011), and the mosquitoes *A. aegypti* (Drake et al., 2010) and *A. gambiae* (Liu et al., 2011). Interestingly, transcripts from all five LhAQPs were detected in leg tissue, which could indicate a possible role in muscle respiration. Indeed, water transport is important in flight muscles during periods of high-energy demand (Wigglesworth, 1965; Wigglesworth and Lee, 1982) and a water-specific AQP identified in *A. aegypti* Malpighian tubules has been shown to be involved in respiration by transporting water through tracheolar cells (Duchesne et al., 2003).

Heterologous expression of LhAQP-EGFP chimeras in Tni cells showed that LhAQP2 (isoforms A-E), LhAQP3, and LhAQP4B are produced and transported to the cell surface (Fig. 4B). Carboxyl-terminal EGFP-tagged LhAQP1, LhAQP4A, and LhAQP5 were likewise produced in Tni cells but did not localize at the plasma membrane. They instead localized primarily to cytosolic punctae (LhAQP1 and LhAQP4A) reminiscent of the Golgi apparatus (Kawar and Jarvis, 2001) or exhibited a diffuse cytosolic distribution with a prominent perinuclear ring (LhAQP5) similar to that described (Thomas et al., 1998) for endoplasmic reticulum localization (Figs. 4B, C). We also performed domain-swapping experiments between LhAQP1 and the *B. tabaci* AQP1 (BtAQP1) previously shown to translocate to the cell surface in cultured insect cells (Mathew et al., 2011). Here, the replacement of the amino-terminus of BtAQP1 with that of LhAQP1 prevented the chimera from reaching the cell surface (Fig. 4C). Furthermore, when the amino terminus of LhAQP1 was replaced with that of BtAQP1, localization to the cellular plasma membrane in Tni was rescued (Fig. 4C). These results indicate that some LhAQPs may either have an intracellular localization or that *L. hesperus* and Tni cells differ such that some proteins are not correctly processed by the Tni intracellular protein trafficking machinery. The latter is likely for LhAQP1. LhAQP5 however, might have subcellular function, as it resembles other superaquaporins thought to function as intracellular water channels that regulate free and bound water movement across subcellular organelle membranes (Ishikawa et al., 2005; Nozaki et al., 2008; Ishibashi, 2006).

S-aquaporins or superaquaporins represent a subfamily of AQPs characterized as having relatively poor sequence similarity with other AQPs and atypical NPA motifs (Morishita et al., 2004; Ishibashi, 2006). Here, we report the first functional S-aquaporin (LhAQP5) from an insect. Both plants and animals have S-aquaporins with atypical amino-terminal NPA motifs (HsAQP11, HsAQP12, SIP1.1, SIP1.2, and SIP2.1) (Supplemental Table 1). These AQPs localize on the membrane of intracellular organelles (Ishikawa et al., 2005; Morishita et al., 2005; Itoh et al., 2005) and are thought to participate in the transport of intracellular water through cell organelles, regulation of organelle volume, regulation of intracellular waste products via secretory granules, and maintenance of cellular activities through the distribution of free water (Nozaki et al., 2008). Several S-aquaporins have high lysine content at their carboxyl termini that may act as ER-localization signals (Cosson and Letourneur, 1994), including LhAQP5 with residues 261-KEKD-264 that resemble carboxyl terminal ER retention motifs. Furthermore, Guan et al. (2010) indicates that NPA motifs are important for targeting orthodox AQPs to the plasma membrane and the atypical NPA motifs of the S-aquaporins may aid in their subcellular localization.

Water homeostasis is a critical function for all life. Animals (including insects) have developed both behavioral and physiological adaptations to facilitate osmoregulation and balance internal water equilibrium with water lost to the environment. Terrestrial insects in particular are under continuous osmotic stress and must balance water losses to evaporation through the cuticle, respiratory transpiration, and excretion, and gain water by direct ingestion of fluids, generation of metabolic water, and/or absorption of atmospheric water vapor (Cohen, 2013). Disrupting water balance by targeting the osmoregulatory system including water channel proteins in arthropod pests remains a promising avenue for developing selective and environmentally compatible pest control strategies (Cohen, 2013). However, there are no commercially effective and selective methods yet available that target either water channels or the regulatory pathways controlling water homeostasis in arthropod pests. Furthermore, arthropods have evolved mechanisms by which critical functions (such as embryogenesis, immunity, apoptosis, membrane transport, digestion, metamorphosis, neural-motor function, to name a few) may have redundant or compensatory routes of activity (Masel and Siegal, 2009; Moczek et al., 2011). Indeed, insects produce multiple AQPs with overlapping tissue expression patterns (see Fig. 3B; Drake et al., 2010) and knockdown of individual AQP transcripts by RNAi does not necessarily lead to negative phenotypes (Shakesby et al., 2009; Drake et al., 2010; Herraiz et al., 2011), which is indicative of possible functional redundancy. Yet, the AQPs and the mechanisms by which water homeostasis is maintained remain promising areas of research that may lead to new controls for pests of medical and agricultural importance (Cohen, 2013). Further studies are needed to develop comprehensive models of arthropod AQP function and novel strategies to disrupt osmoregulation.

Acknowledgements

We thank Lolita Mathew, Lynn Forlow-Jech, Sarah Bjorklund, and Eric Hoffmann for technical support, O.P. Perera for providing partial *Lygus lineolaris* aquaporin transcriptome sequences (unpublished) that helped with initial cloning of LhAQPs, and Dale Spurgeon for helpful comments. This is a cooperative investigation

between USDA-ARS and the University of Adelaide. The research described in this manuscript was partially supported by funds from Cotton Inc. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.ibmb.2013.12.002

References

Agre, P., Bonhivers, M., Borgnia, M.J., 1998. The aquaporins, blueprints for cellular plumbing systems. J. Biol. Chem. 273, 14659–14662.
Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.
Anthony, T.L., Brooks, H.L., Boassa, D., Leonov, S., Yanochko, G.M., Regan, J.W., Yool, A.J., 2000. Cloned human aquaporin-1 is a cyclic GMP-gated ion channel. Mol. Pharmacol. 57, 576–588.
Azuma, M., Nagae, T., Maruyama, M., Kataoka, N., Miyake, S., 2012. Two water-specific aquaporins at the apical and basal plasma membranes of insect epithelia: molecular basis for water recycling through the cryptonephric rectal complex of lepidopteran larvae. J. Insect Physiol. 58, 523–533.
Backus, E.A., Cline, A.R., Ellerseick, M.R., Serrano, M.S., 2007. *Lygus hesperus* (Hemiptera: Miridae) feeding on cotton: new methods and parameters for analysis of nonsequential electrical penetration graph data. Ann. Entomol. Soc. Am. 100, 296–310.
Ball, A., Campbell, E.M., Jacob, J., Hoppler, S., Bowman, A.S., 2009. 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.
Beitz, E., Wu, B., Holm, L.M., Schultz, J.E., 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.
Benga, G., 2009. Water channel proteins (later called aquaporins) and relatives: past, present, and future. IUBMB Life 61, 112–133.
Bernsel, A., Viklund, H., Hennerdal, A., Elofsson, A., 2009. TOPCONS: consensus prediction of membrane protein topology. Nucleic Acids Res. 37, W465–W468.
Boassa, D., Yool, A.J., 2003. Single amino acids in the carboxyl terminal domain of aquaporin-1 contribute to cGMP-dependent ion channel activation. BMC Physiol. 3, 12.
Calvanese, L., Pellegrini-Calace, M., Oliva, R., 2013. In silico study of human aquaporin AQP11 and AQP12 channels. Protein Sci. 22, 455–466.
Campbell, E.M., Ball, A., Hoppler, S., Bowman, A.S., 2008. Invertebrate aquaporins: a review. J. Comp. Physiol. B. 178, 935–955.
Campbell, E.M., Burdin, M., Hoppler, S., Bowman, A.S., 2010. Role of an aquaporin in the sheep tick *Ixodes ricinus*: assessment as a potential control target. Int. J. Parasitol. 40, 15–23.
Carbrey, J.M., Agre, P., 2009. Discovery of the aquaporins and development of the field. Handb. Exp. Pharmacol., 3–28.
Claros, M.G., von Heijne, G., 1994. TopPred II: an improved software for membrane protein structure predictions. Comput. Appl. Biosci. 10, 685–686.
Cohen, E., 2013. Chapter one – water homeostasis and osmoregulation as targets in the control of insect pests. In: Cohen, E. (Ed.), Target Receptors in the Control of Insect Pests: Part I, Advances in Insect Physiology, vol. 44. Academic Press, pp. 1–61.
Cosson, P., Letourneur, F., 1994. Coatomer interaction with di-lysine endoplasmic reticulum retention motifs. Science 263, 1629–1631.
Debolt, J.W., 1982. Meridic diet for rearing successive generations of *Lygus hesperus*. Ann. Entomol. Soc. Am. 75, 119–122.
Douglas, A.E., 2006. Phloem-sap feeding by animals: problems and solutions. J. Exp. Bot. 57, 747–754.
Drake, L.L., Boudko, D.Y., Marinotti, O., Carpenter, V.K., Dawe, A.L., Hansen, I.A., 2010. The aquaporin gene family of the yellow fever mosquito, *Aedes aegypti*. PLoS One 5, e15578.
Duchesne, L., Hubert, J.F., Verbavatz, J.M., Thomas, D., Pietrantonio, P.V., 2003. Mosquito (*Aedes aegypti*) aquaporin, present in tracheolar cells, transports water, not glycerol, and forms orthogonal arrays in Xenopus oocyte membranes. Eur. J. Biochem. 270, 422–429.
Echevarria, M., Ramirez-Lorca, R., Hernandez, C.S., Gutierriez, A., Mendez-Ferrer, S., Gonzalez, E., Toledo-Aral, J.J., Ilundain, A.A., Whittembury, G., 2001. Identification of a new water channel (Rg-MIP) in the Malpighian tubules of the insect *Rhodnius prolixus*. Pflug. Arch. Eur. J. Phy. 442, 27–34.
Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.
Fu, D., Libson, A., Miercke, L.J., Weitzman, C., Nollert, P., Krucinski, J., Stroud, R.M., 2000. Structure of a glycerol-conducting channel and the basis for its selectivity. Science 290, 481–486.
Gomes, D., Agasse, A., Thiébaud, P., Delrot, S., Gerós, H., Chaumont, F., 2009. Aquaporins are multifunctional water and solute transporters highly divergent in living organisms. BBA-Biomemb. 1788, 1213–1228.
Goodchild, A.J.P., 1966. Evolution of the alimentary canal in the Hemiptera. Biol. Rev. 41, 97–139.
Goto, S.G., Philip, B.N., Teets, N.M., Kawarasaki, Y., Lee Jr., R.E., Denlinger, D.L., 2011. Functional characterization of an aquaporin in the Antarctic midge *Belgica antarctica*. J. Insect Physiol. 57, 1106–1114.
Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotech. 29, 644–652.
Guan, X.G., Su, W.H., Yi, F., Zhang, D., Hao, F., Zhang, H.G., Liu, Y.J., Feng, X.C., Ma, T.H., 2010. NPA motifs play a key role in plasma membrane targeting of Aquaporin-4. IUBMB Life 62, 222–226.
Gullan, P.J., Cranston, P.S., 2005. The Insects: an Outline of Entomology, third ed. Blackwell Publishing Ltd, Massachusetts.
Habibi, J., Coudron, T.A., Backus, E.A., Brandt, S.L., Wagner, R.M., Wright, M.K., Huesing, J.E., 2008. Morphology and histology of the alimentary canal of *Lygus hesperus* (Heteroptera: Cimicomoropha: Miridae). Ann. Entomol. Soc. Am. 101, 159–171.
Hachez, C., Chaumont, F., 2010. Aquaporins: a family of highly regulated multifunctional channels. Adv. Exp. Med. Biol. 679, 1–17.
Herraiz, A., Chauvigne, F., Cerda, J., Bellas, X., Piulachs, M.-D., 2011. Identification and functional characterization of an ovarian aquaporin from the cockroach *Blattella germanica* L. (Dictyoptera, Blattellidae). J. Exp. Biol. 214, 3630–3638.
Ho, J.D., Yeh, R., Sandstrom, A., Chorny, I., Harries, W.E., Robbins, R.A., Miercke, L.J., Stroud, R.M., 2009. Crystal structure of human aquaporin 4 at 1.8 A and its mechanism of conductance. Proc. Natl. Acad. Sci. USA 106, 7437–7442.
Hofmann, K., Stoffel, W., 1993. TMbase – a database of membrane spanning proteins segments. Biol. Chem. Hoppe-Seyler 374, 166.
Holmes, S., Li, D., Ceraul, S., Azad, A., 2008. An aquaporin-like protein from the ovaries and gut of American dog tick (Acari: Ixodidae). J. Med. Entomol. 45, 68–74.
Horton, P., Park, K.-J., Obayahsi, T., Fujita, N., Harada, H., Adams-Collier, C.J., Nakai, K., 2007. WoLF PSORT: protein localization predictor. Nucl. Acids Res. 35, W585–W587.
Hub, J.S., De Groot, B.L., 2008. Mechanism of selectivity in aquaporins and aquaglyceroporins. Proc. Natl. Acad. Sci. USA 105, 1198–1203.
Hubert, J.F., Thomas, D., Cavalier, A., Gouranton, J., 1989. Structural and biochemical observations on specialized membranes of the filter chamber, a water-shunting complex in sap-sucking homopteran insects. Biol. Cell 66, 155–163.
Hull, J.J., Hoffmann, E.J., Perera, O.P., Snodgrass, G.L., 2012. Identification of the western tarnished plant bug (*Lygus hesperus*) olfactory co-receptor Orco: expression profile and confirmation of atypical membrane topology. Arch. Insect Biochem. Physiol. 81, 179–198.
Hull, J.J., Geib, S., Fabrick, J.A., Brent, C.S., 2013. Sequencing and *de novo* assembly of the western tarnished plant bug (*Lygus hesperus*) transcriptome. PLoS ONE 8, e55105.
Ishibashi, K., 2006. Aquaporin subfamily with unusual NPA boxes. Biochim. Biophys. Acta 1758, 989–993.
Ishibashi, K., 2009. New members of mammalian aquaporins: AQP10–AQP12. Handb. Exp. Pharmacol., 251–262.
Ishibashi, K., Hara, S., Kondo, S., 2009. Aquaporin water channels in mammals. Clin. Exp. Nephrol. 13, 107–117.
Ishikawa, F., Suga, S., Uemura, T., Sato, M.H., Maeshima, M., 2005. Novel type aquaporin SIPs are mainly localized to the ER membrane and show cell-specific expression in *Arabidopsis thaliana*. FEBS Lett. 579, 5814–5820.
Itoh, T., Rai, T., Kuwahara, M., Ko, S.B., Uchida, S., Sasaki, S., Ishibashi, K., 2005. Identification of a novel aquaporin, AQP12, expressed in pancreatic acinar cells. Biochem. Biophys. Res. Commun. 330, 832–838.
Jung, J.S., Preston, G.M., Smith, B.L., Guggino, W.B., Agre, P., 1994. Molecular structure of the water channel through aquaporin CHIP. The hourglass model. J. Biol. Chem. 269, 14648–14654.
Käll, L., Krogh, A., Sonnhammer, E.L.L., 2004. A combined transmembrane topology and signal peptide prediction method. J. Mol. Biol. 338, 1027–1036.
Kambara, K., Takematsu, Y., Azuma, M., Kobayashi, J., 2009. 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.
Kataoka, N., Miyake, S., Azuma, M., 2009a. Aquaporin and aquaglyceroporin in silkworms, differently expressed in the hindgut and midgut of *Bombyx mori*. Insect Mol. Biol. 18, 303–314.
Kataoka, N., Miyake, S., Azuma, M., 2009b. Molecular characterization of aquaporin and aquaglyceroporin in the alimentary canal of *Grapholita molesta* (the oriental fruit moth) -comparison with *Bombyx mori* aquaporins. J. Insect Biotech. Sericol. 78, 81–90.
Kaufmann, N., Mathai, J.C., Hill, W.G., Dow, J.A.T., Zeidel, M.L., Brodsky, J.L., 2005. Developmental expression and biophysical characterization of a *Drosophila melanogaster* aquaporin. Am. J. Physiol.-Cell Ph. 289, C397–C407.

Kawar, Z., Jarvis, D.L., 2001. Biosynthesis and subcellular localization of a lepidopteran insect alpha 1,2-mannosidase. Insect Biochem. Mol. Biol. 31, 289–297.
Kikawada, T., Saito, A., Kanamori, Y., Fujita, M., Snigorska, K., Watanabe, M., Okuda, T., 2008. Dehydration-inducible changes in expression of two aquaporins in the sleeping chironomid, *Polypedilum vanderplanki*. BBA-Biomemb. 1778, 514–520.
King, L.S., Kozono, D., Agre, P., 2004. From structure to disease: the evolving tale of aquaporin biology. Nat. Rev. Mol. Cell. Biol. 5, 687–698.
Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305, 567–580.
Kruse, E., Uehlein, N., Kaldenhoff, R., 2006. The aquaporins. Genome Biol. 7, 206.
Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and clustal X version 2.0. Bioinformatics 23, 2947–2948.
Le Caherec, F., Deschamps, S., Delamarche, C., Pellerin, I., Bonnec, G., Guillam, M.T., Thomas, D., Gouranton, J., Hubert, J.F., 1996. Molecular cloning and characterization of an insect aquaporin – functional comparison with aquaporin 1. Eur. J. Biochem. 241, 707–715.
Le Caherec, F., Guillam, M.T., Beuron, F., Cavalier, A., Thomas, D., Gouranton, J., Hubert, J.F., 1997. Aquaporin-related proteins in the filter chamber of homopteran insects. Cell Tiss. Res. 290, 143–151.
Lehane, M.J., Billingsley, P.F., 1996. Biology of the Insect Midgut, first ed. Chapman and Hall, London.
Liu, K., Tsujimoto, H., Cha, S.-J., Agre, P., Rasgon, J.L., 2011. Aquaporin water channel AgAQP1 in the malaria vector mosquito *Anopheles gambiae* during blood feeding and humidity adaptation. Proc. Natl. Acad. Sci. USA 108, 6062–6066.
Masel, J., Siegal, M.L., 2009. Robustness: mechanisms and consequences. Trends Genet. 25, 395–403.
Mathew, L.G., Campbell, E.M., Yool, A.J., Fabrick, J.A., 2011. Identification and characterization of functional aquaporin water channel protein from alimentary tract of whitefly, *Bemisia tabaci*. Insect Biochem. Mol. Biol. 41, 178–190.
Miles, P.W., 1987. Plant-sucking bugs can remove the contents of cells without mechanical damage. Experientia 43, 937–939.
Moczek, A.P., Sultan, S., Foster, S., Ledon-Rettig, C., Dworkin, I., Nijhout, H.F., Abouheif, E., Pfennig, D.W., 2011. The role of developmental plasticity in evolutionary innovation. Proc. Biol. Sci. 278, 2705–2713.
Morishita, Y., Matsuzaki, T., Hara-chikuma, M., Andoo, A., Shimono, M., Matsuki, A., Kobayashi, K., Ikeda, M., Yamamoto, T., Verkman, A., Kusano, E., Ookawara, S., Takata, K., Sasaki, S., Ishibashi, K., 2005. Disruption of aquaporin-11 produces polycystic kidneys following vacuolization of the proximal tubule. Mol. Cell. Biol. 25, 7770–7779.
Morishita, Y., Sakube, Y., Sasaki, S., Ishibashi, K., 2004. Molecular mechanisms and drug development in aquaporin water channel diseases: aquaporin superfamily (superaquaporins): expansion of aquaporins restricted to multicellular organisms. J. Pharmacol. Sci. 96, 276–279.
Murata, K., Mitsuoka, K., Hirai, T., Walz, T., Agre, P., Heymann, J.B., Engel, A., Fujiyoshi, Y., 2000. Structural determinants of water permeation through aquaporin-1. Nature 407, 599–605.
Nagae, T., Miyake, S., Kosaki, S., Azuma, M., 2013. Identification and characterisation of functional aquaporin water channel (*Anomala cuprea* DRIP) in a coleopteran insect. J. Exp. Biol.
Nozaki, K., Ishii, D., Ishibashi, K., 2008. Intracellular aquaporins: clues for intracellular water transport? Pflugers Arch. 456, 701–707.
Patana, R., 1982. Disposable diet packet for feeding and oviposition of *Lygus hesperus* (Hemiptera: Miridae). J. Econ. Entomol. 75, 668–669.
Philip, B.N., Kiss, A.J., Lee Jr., R.E., 2011. The protective role of aquaporins in the freeze-tolerant insect *Eurosta solidaginis*: functional characterization and tissue abundance of EsAQP1. J. Exp. Biol. 214, 848–857.
Rose, A., Lorenzen, S., Goede, A., Gruening, B., Hildebrand, P.W., 2009. RHYTHM–a server to predict the orientation of transmembrane helices in channels and membrane-coils. Nucleic Acids Res. 37, W575–W580.
Saitou, N., Nei, M., 1986. Strategy for resolving the branching order of humans, chimpanzees and gorillas by using DNA-sequence data. Am. J. Phys. Anthropol. 69, 260.
Saitou, N., Nei, M., 1987. The neighbor-joining method – a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.
Shackel, K.A., de la Paz Celorio-Mancera, M., Ahmadi, H., Greve, L.C., Teuber, L.R., Backus, E.A., Labavitch, J.M., 2005. Micro-injection of lygus salivary gland proteins to simulate feeding damage in alfalfa and cotton flowers. Arch. Insect Biochem. Physiol. 58, 69–83.
Shakesby, A.J., Wallace, I.S., Isaacs, H.V., Pritchard, J., Roberts, D.M., Douglas, A.E., 2009. A water-specific aquaporin involved in aphid osmoregulation. Insect Biochem. Mol. Biol. 39, 1–10.
Sneath, P.H.A., Sokal, R.R., 1973. Numerical Taxonomy. W.H. Freeman and Company, San Francisco, pp. 230–234.
Spring, J.H., Robichaux, S.R., Hamlin, J.A., 2009. The role of aquaporins in excretion in insects. J. Exp. Biol. 212, 358–362.
Strong, F.E., Kruitwagen, E.C., 1968. Polygalacturonase in the salivary apparatus of *Lygus hesperus* (Hemiptera). J. Insect Physiol. 14, 1113–1119.
Sui, H., Han, B.G., Lee, J.K., Walian, P., Jap, B.K., 2001. Structural basis of water-specific transport through the AQP1 water channel. Nature 414, 872–878.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony method. Mol. Biol. Evol. 28, 2731–2739.
Thomas, C.J., Brown, H.L., Hawes, C.R., Lee, B.Y., Min, M.K., King, L.A., Possee, R.D., 1998. Localization of a baculovirus-induced chitinase in the insect cell endoplasmic reticulum. J. Virol. 72, 10207–10212.
Tusnady, G.E., Simon, I., 2001. The HMMTOP transmembrane topology prediction server. Bioinformatics 17, 849–850.
Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., Leunissen, J.A.M., 2007. Primer3Plus, an enhanced web interface to primer3. Nucleic Acids Res. 35, W71–W74.
Verkman, A.S., 2008. Mammalian aquaporins: diverse physiological roles and potential clinical significance. Expert Rev. Mol. Med. 10, e13.
Wallace, I.S., Shakesby, A.J., Hwang, J.H., Choi, W.G., Martinkova, N., Douglas, A.E., Roberts, D.M., 2012. *Acyrthosiphon pisum* AQP2: a multifunctional insect aquaglyceroporin. Biochim. Biophys. Acta 1818, 627–635.
Wheeler, A.G., 2001. Biology of the Plant Bugs (Hemiptera: Miridae): Pests, Predators, Opportunists. Cornell University Press, Ithaca, NY.
Wigglesworth, V.B., 1965. Respiration. The Principles of Insect Physiology. Methuen Ltd, London, UK, pp. 317–368.
Wigglesworth, V.B., Lee, W.M., 1982. The supply of oxygen to the flight muscles of insects: a theory of tracheole physiology. Tissue Cell 14, 501–518.
Yanochko, G.M., Yool, A.J., 2002. Regulated cationic channel function in *Xenopus* oocytes expressing *Drosophila* big brain. J. Neurosci. 22, 2530–2540.
Yool, A.J., 2007. Aquaporins: multiple roles in the central nervous system. Neuroscientist 13, 470–485.
Yool, A.J., Weinstein, A.M., 2002. New roles for old holes: Ion channel function in aquaporin-1. News Physiol. Sci. 17, 68–72.
Zardoya, R., 2005. Phylogeny and evolution of the major intrinsic protein family. Biol. Cell. 97, 397–414.
Zhang, D., Tan, Y.-J., Qu, F., Sheng, J.-Z., Huang, H.-F., 2012. Functions of water channels in male and female reproductive systems. Mol. Aspects Med. 33, 676–690.
Zuckerkandl, E., Pauling, L., 1965. Evolutionary divergence and convergence in proteins. In: Bryson, V., Vogel, H.J. (Eds.), Evolving Genes and Proteins. Academic Press, New York, pp. 97–166.