Bioinformatics and Biology Insights

ORIGINAL RESEARCH

The Aquaporin Channel Repertoire of the Tardigrade *Milnesium tardigradum*

Markus A. Grohme[1], Brahim Mali[1], Weronika Wełnicz[1,3], Stephanie Michel[1], Ralph O. Schill[2] and Marcus Frohme[1]

[1]Molecular Biotechnology and Functional Genomics, Technical University of Applied Sciences Wildau, Wildau, Germany. [2]Biological Institute, Zoology, University of Stuttgart, Stuttgart, Germany. [3]Current address: DFG-Center for Regenerative Therapies Dresden - Cluster of Excellence (CRTD), Technische Universität Dresden, Dresden, Germany.
Corresponding author email: mfrohme@th-wildau.de

Abstract: Limno-terrestrial tardigrades are small invertebrates that are subjected to periodic drought of their micro-environment. They have evolved to cope with these unfavorable conditions by anhydrobiosis, an ametabolic state of low cellular water. During drying and rehydration, tardigrades go through drastic changes in cellular water content. By our transcriptome sequencing effort of the limno-terrestrial tardigrade *Milnesium tardigradum* and by a combination of cloning and targeted sequence assembly, we identified transcripts encoding eleven putative aquaporins. Analysis of these sequences proposed 2 classical aquaporins, 8 aquaglyceroporins and a single potentially intracellular unorthodox aquaporin. Using quantitative real-time PCR we analyzed aquaporin transcript expression in the anhydrobiotic context. We have identified additional unorthodox aquaporins in various insect genomes and have identified a novel common conserved structural feature in these proteins. Analysis of the genomic organization of insect aquaporin genes revealed several conserved gene clusters.

Keywords: unorthodox aquaporin, anhydrobiosis, tardigrade

Bioinformatics and Biology Insights 2013:7 153–165

doi: 10.4137/BBI.S11497

This article is available from http://www.la-press.com.

Introduction

Although water is a prerequisite for life, some organisms have developed mechanisms to cope with frequent episodes of water limitation. Tardigrades and other small invertebrates are known for their capability of surviving complete desiccation by undergoing anhydrobiosis.[1–5] This phenomenon of "life without water" can be found in a wide range of taxa, including bacteria, plants and animals.[6–8] Tardigrades, also called "water bears" are micrometazoans living in marine, freshwater and terrestrial environments. The habitat of limno-terrestrial tardigrades may fall dry repeatedly during their normal life cycle. During time of drought, tardigrades assume the shape of compact tuns and enter a phase of dormancy. In this anhydrobiotic state, metabolism and aging stops and if necessary, tuns can last for several years.[9–11] This evolutionary adaptation has allowed tardigrades to colonize a wide range of extreme habitats that are frequently subjected to drastic changes in water availability.

During anhydrobiosis, cellular water content changes drastically and after complete desiccation, only a residual amount of cellular water is present (<10% of absolute water content).[1,11] Members of the genus *Milnesium* represent a well-studied group of limno-terrestrial tardigrades and their tuns show a remarkable resistance to environmental stresses.[12–15] Although tardigrades vary in their sensitivity and response to desiccation, most limno-terrestrial tardigrades appear to have developed mechanisms for survival under desiccating conditions. The speed of dehydration has a great influence on survival, as abrupt dehydration in a low humidity environment correlates with a reduced survival rate.[16] Therefore, anhydrobiosis requires a priming of the organism to be able to cope with its changing surroundings. As the rehydration of desiccated tuns is a quick process, a controlled accelerated water transport into the cells by aquaporins (AQPs) seems like a probable source of rehydration. Because there is no information on AQPs in tardigrades available, we sought to identify corresponding transcripts in *Milnesium tardigradum* (Doyère, 1840) and investigated the potential role of AQPs in the modulation of water content and osmolytes during anhydrobiosis.

AQPs are small integral plasma membrane proteins in the major intrinsic protein (MIP) family. They selectively allow water or other small uncharged molecules to permeate across membranes.[17,18] As biological membranes exhibit very low water permeability, aquaporins are important for water transport across membranes and cellular homeostasis. Since the publication of the first aquaporin,[19] these channel molecules have been identified in a wide range of tissues in mammals, invertebrates, plants as well as in microorganisms.[20–22] AQPs share a common structure comprising 6 transmembrane domains (TMD1–TMD6) that are connected by 5 loops (A–E) and both amino- and carboxyl termini are located in the cytoplasm.[23] Functionally, they are present as homo-tetrameric assemblies in the cell membrane.[24] Transmembrane domains 1–3 and 4–6 seem to be homologous and are likely to have originated from a duplication event in this family (MIPs).[25,26] In addition to water-specific aquaporins that only allow the passage of water, aquaglyceroporins have also been identified, which are permeated by larger solutes such as glycerol or urea.[27] Some AQPs have even been shown to allow passage of unconventional substrates, eg, arsenite or nitric oxide.[28] In addition to AQPs residing in the outer membrane, intracellular AQPs have been reported.[29–31] In addition to their main transport function, they also play a role in cell migration.[32,33]

Two structural features determine AQP selectivity and are responsible for the exclusion of larger solutes and charged solutes such as protons. The "asparagine-proline-alanine" (NPA) motifs, embedded in the plasma membrane in the center of the pore forming a selectivity filter resembling an hourglass, are one of the highly conserved signature sequences.[34,35] There are rare deviations of asparagine in the NPA motif and, although it does not change water or solute permeability, this residue is very important for cation exclusion such as Na^+.[36] A second selectivity filter is the aromatic/arginine region (ar/R) at the extracellular pore mouth, representing the narrowest part of the pore.[37,38] The diameter of the ar/R region determines if the pore is water specific or if it allows larger solutes to pass. These filters act alone or in conjunction to determine the permeation properties of the channel.[39]

Materials and Methods

Animal culture

M. tardigradum was originally collected in Tübingen, Germany and has been available as a well-established laboratory culture for a decade. The animals

were cultured on Petri dishes (ø 9.4 cm) with a layer of agarose (3%) (peqGOLD Universal Agarose, PEQLAB, Erlangen Germany) covered with a thin layer of Volvic® water (Danone Waters, Wiesbaden, Germany) at 20 °C. The animals were fed bdelloid rotifers, *Philodina citrina* (Ehrenberg, 1832), which had been raised on the green algae *Chlorogonium elongatum* (Dangeard, 1888). Animals were starved for 2 days to avoid contamination from undigested food. After repeated washing with Volvic® water, 200 animals were transferred into 1.5 mL tubes. Residual water was removed using a micropipette. The open tubes were then exposed to 85% relative humidity (RH) in a small chamber containing a saturated solution of KCl (Roth, Karlsuhe, Germany). After 24 hours (h), the samples were dried at 35% RH for a further 48 h in a chamber containing a saturated solution of $MgCl_2$ (Roth). For rehydration experiments the dried tuns were transferred to a Petri dish and were submerged in Volvic® water. Rehydration was monitored using a stereo-microscope and animals failing to rehydrate after 1 h or showing abnormal morphology were discarded. Subsequently 3 different stages were used in this study: active (live animals used as control), inactive (62 h of dehydration) and rehydrated stage (1 h of rehydration).

RACE-PCR and cloning of *Milnesium tardigradum* aquaporin cDNAs

For RNA extraction the tardigrades were sonicated on ice for 1 minute (min) (duty cycle 0.5 seconds (s)) using a microsonicator (Probe 73, Sonopuls; Bandelin, Berlin, Germany). Total RNA extraction was performed by following the instructions of QIAGEN Rneasy® Mini Kit (Qiagen, Hilden, Germany).

To obtain the 5′- and 3′-ends of AQP-5 and 6 transcripts, rapid amplification of cDNA ends (RACE-PCR) was performed using the SMART RACE cDNA Amplification Kit (Clontech, Saint-Germain-en-Laye, France). cDNA was generated from 100 ng total RNA according to the manufacturer's protocol. 5′ and 3′ cDNAs were then amplified by PCR (30 cycles 30 s at 94 °C, 2 min at 64 °C, 2 min at 72 °C). Primer sequences used for RACE-PCR are available upon request. RACE PCR products were cloned into the pGEM-T vector (Promega, Mannheim, Germany) and sequenced on a capillary sequencer at StarSEQ (Mannheim, Germany).

Quantitative real-time PCR

Quantitative real-time PCR (qPCR) analysis was performed on a LightCycler 480 real-time PCR machine (Roche Applied Science, Mannheim, Germany). Our preliminary experiments indicated that the *M. tardigradum* orthologue of the *rho GDP dissociation inhibitor* (rhoGDI) transcript has stable expression under current experimental conditions. Thus, rhoGDI was used as a reference transcript for the normalization of qPCR results. Primers were designed using the PerlPrimer software (v1.1.17)[55] using standard parameters and were obtained as HPLC purified oligonucleotides (Thermo Fisher Scientific, Ulm, Germany). A volume of 20 µL PCR reaction mixture contained 1 µL cDNA, 1 µL of each AQP primer (10 µM), 7 µL nuclease-free water and 10 µL 480 SYBR Green I Master (Roche Applied Science). All reactions were performed in duplicate in 96-well plate format. The thermal cycling conditions were as follows: 1 cycle (10 min at 95 °C) followed by 45 cycles (10 s at 95 °C, 10 s at 59 °C and 15 s at 72 °C) followed by a melting curve analysis to check primer specificity. The expression of the eleven aquaporins was analyzed by using the $2^{-\Delta\Delta Ct}$-method.[56] Primers and qPCR data is available in file S3.

Sequence clustering and assembly

To collect additional sequence information covering the identified putative aquaporin transcripts we iteratively assembled our 454 short sequence reads together with Sanger ESTs using a combination of GenSeed (v1.0.22) and TGICL (v2.1).[57,58] GenSeed performs a BLAST search with a "seed" sequence to identify matching reads in a FASTA database and incrementally tries to expand the sequence alternating between assembly and BLAST search. Both programs use CAP3 to assemble matching reads into larger contigs.[59] The input for GenSeed consisted of aquaporin fragments identified by BLAST searches.[60,61] The GenSeed parameters were: genseed. pl -g no -t 50 -j 95 -x complete -b "-F F -b 1000 -e 1e-06 -a 3" -a "-o 25 -g 1 -s 450". As GenSeed does not use quality values associated with the nucleotide sequences, the final contig sequence might not represent the optimal assembly. In a second step we therefore clustered and assembled against GenSeed generated maximum length contigs as references using TGICL. TGICL, in contrast to GenSeed, also

takes base quality information into account, but lacks the iterative assembly functionality. This 2-phased approach led to longer and more accurate contigs in our hands. The TGICL parameters were: Clustering step: tgicl -F <input_reads.fasta> -q <input_reads.qual> -d aquaporin_reference.fasta -l 20; assembly parameters: tgicl -F <input_reads.fasta> -q <input_reads.qual> -a <clusters> -O "-o 20 -g 1 -s 450". For the assembly we reduced the overlap cutoff to 20 bases and reduced the gap penalty to 1 to accommodate for the shorter nature and error model of 454 reads.[62,63] The resulting contigs were inspected using the tablet assembly viewer for obvious sequencing and/or assembly errors and spoiler read IDs were removed from the cluster file and the clusters were re-assembled.[64]

Database search for undescribed aquaporins

We used a combination of different BLAST search strategies against ecdysozoan genome and EST databases to identify known and new AQP homologues. We found DELTA-BLAST, which is a modified BLAST utilizing domain information to identify distant homologues to be especially sensitive for detecting AQP sequences.[65] It readily identified already known, undescribed and even erroneously annotated AQP sequences in the analyzed genomes as we found new unorthodox aquaporins for *B. mori* and *A. mellifera*. Database searches using DELTA BLAST were performed against NCBI non-redundant protein database (nr) using MtAqp-11 against sub-databases of *D. melanogaster*, *A. aegypti*, *B. mori*, *A. mellifera*, *P. humanus corporis* and *C. elegans*. The search parameters for the substitution matrix were [BLOSUM45, word size 2] and for the gap parameters [existence 16, extension 1]. For *A. mellifera* the predicted protein for XP_001122456.2 was fused with exons of an upstream gene attaching it to its C-terminus. This was evidenced by several undetermined bases in the genome sequence ("n") closely upstream of a start codon, which likely represents the start of the true open reading frame (File S6). For *B. mori* EST/cDNA was assembled, revealing a putative unorthodox AQP.

Phylogenetic analysis

Multiple sequence alignments (MSA) for the highly diverged AQPs were generated using the PROMALS3D web-server.[66] PROMALS3D has been specifically developed for aligning distantly related proteins by taking structural information into account. We then converted the amino acid alignment into its respective codon alignment using PAL2NAL (v14).[67] The codon alignments were loaded into MEGA (v. 5.05).[68] The evolutionary history was inferred by using the maximum likelihood method using the generalized time reversible substitution model. Initial tree(s) for the heuristic search were obtained automatically as follows. When the number of common sites was <100, or <1/4 the total number of sites, the maximum parsimony method was used; otherwise the BIONJ method with a MCL distance matrix was used. A discrete gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.4637)). Bootstrapping was performed using 500 replicates. The analysis involved 59 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 3018 positions in the final dataset. The final tree was edited using inkscape (http://inkscape.org/).

Structure analyses

M. tardigradum AQP predicted protein sequences were submitted to the Iterative implementation of the Threading ASSEmbly Refinement (I-TASSER) server for 3D structure prediction.[42] The I-TASSER predictions for *M. tardigradum* AQPs as well for unorthodox AQPs can be found in file S4. Structures were rendered using pyMOL.[69]

Sequence accession numbers

The *M. tardigradum* aquaporin cDNA nucleotide sequences reported in this study have been submitted to GenBank under the accession numbers JN378733–JN378746. Further sequences used in this study were: *Apis mellifera* (GenBank: XP_624531.2, XP_394391.1, XP_396705.3, XP_624194.2, XP_001121899.2, XP_001121043.2, XP_001122456.2), *Drosophila melanogaster* (GenBank: NP_725051.2, NP_725052.1, NP_610686.1, NP_476837.1, NP_611810.1, NP_611812.1, NP_726347.1, NP_611813.1), *Polypedilum vanderplanki* (GenBank: BAF62090.1, BAF62091.1), *Belgica antarctica* (GenBank: BAK32935.1), *Caenorhabditis elegans* (GenBank: NP_505691.2, NP_001021552.2, P_001022480.1,

NP_499821.2, NP_508515.2, NP_505512.3, NP_001024758.1, NP_505727.1, NP_502044.1, NP_496105.1, NP_495510.1, NP_495973.1, NP_495972.1), *Aedes aegypti* (GenBank: XP_001656932.1,XP_001650171.1,XP_001650170.1, XP_001650169.1,XP_001650168.1,XP_001649747.1, XP_001648319.1 XP_001648046.1), *Bombyx mori* (GenBank: NP_001036919.1, NP_001106228.1, NP_001153661.1, AK383046.1, DC438195.1, AK383046.1, FS881627.1, FY739398.1, BP183079.1, BP184052.1, HX273263.1, HX271531.1, SilkDB:[70] BGIBMGA008236, BGIBMGA007338, BGIBMGA007339), *Pediculus humanus corporis* (GenBank: EEB20483.1,EEB12655.1,EEB16742.1,EEB15451.1, EEB17617.1, EEB17665.1), and *Homo sapiens* (GenBank: NP_766627.1, NP_945349.1).

Results

Identification and in silico analysis of *Milnesium tardigradum* aquaporins

We identified eleven unique AQP transcripts (denoted MtAqp-1 through MtAqp-11) by screening the available *M. tardigradum* expressed sequence tags (EST)[40] generated by Sanger sequencing and also the assembly and analysis of normalized 454 Titanium sequencing data (unpublished data). The majority of full-length AQP transcripts could be retrieved from the sequence data alone. The full-length sequences of incomplete AQP transcripts were either obtained by 5′- and/or 3′ RACE-PCR amplifications or targeted seed driven assembly approaches.

The open reading frames of the transcripts ranged from 270 amino acids (aa) to 409 aa (Table 1). Relative to the channel core, MtAqp-1, -4 and -10 showed elongated amino-termini, whereas MtAqp-2 and -8 contained longer carboxy-terminal domains.

Upon analyzing the pore region residues of the MtAqps, we found that 7 of them contained canonical NPA motifs (Table 1, Fig. 1). Deviations from the NPA motifs were found in MtAqp-6, which contains a common alanine to valine change in the third position of the NPA box in loop B and a methionine in the third position of the loop E box.[21] MtAqp-2 and MtAqp-8 also show a common exchange of alanine for a serine in the second box in loop E. Typically, for aquaporin-like proteins (or unorthodox aquaporins), MtAqp-11 contains a highly derived NPA box in loop B with the first residue changed into aspartic acid and the third into a serine (DPS). Also, in loop E the third residue is changed to a threonine (NPT). The asparagine of the NPA boxes is therefore conserved in all identified sequences, except MtAqp-11.

The majority of the identified AQPs (except MtAqp-5, -6, -11) show signature residues of aquaglyceroporins, possessing an aspartic acid (D) directly after the second NPA box.[41]

As there are several crystal structures for various aquaporins available we attempted modeling the protein structures of the identified *M. tardigradum* proteins using experimentally determined structures. This was performed on the I-TASSER server.[42] Threading templates consisted exclusively of structures from MIPs and we retrieved good modeling of the well-conserved protein cores, ie, from helix 1 to helix 6. The lowest estimated template modeling score (TM-Score) was 0.46 ± 0.15 for MtAqp-1 and the highest of 0.74 ± 0.11 for both MtAqp-5 and MtAqp-9 (Table 2).

Models predicted for MtAqp-1, -4 and -10 had lower estimated C- and TM-scores due to their longer unstructured N- and C-termini, whereas the "core" of channel forming structures was well modeled. In agreement with the corresponding signature residues, the most similar structures in the Protein Data Bank (PDB)[43] represented aquaglyceroporin-type channels, except for MtAqp-5, -6 and -11, where the structures belonged to water-specific aquaporins (Table 3). Although MtAqp-11 has highly deviated NPA boxes the estimated accuracy of the top model was reasonably high (0.57 ± 0.14), implicating minor deviations from the common channel structure.

Expression dynamics of *Milnesium tardigradum* aquaporins during anhydrobiosis

To retrieve expression values for individual stages and to determine if the expression of the AQP transcripts identified in *M. tardigradum* is dependent on the desiccation/rehydration process, quantitative real-time PCR analysis was performed at 3 different stages: active animals, inactive tuns and during rehydration of the tuns. The qPCR transcript quantification showed that the expression of certain AQPs is, to a small extent, affected by desiccation (Fig. 2). Relative levels of transcripts from MtAqp-8 and MtAqp-9 were slightly up-regulated in the inac-

Table 1. Accession numbers and features of identified aquaporin proteins.

Name	Accession ID	Length	1st NPA box	2nd NPA box
MtAqp-1	AEP14555.1	333	GAGVSGGII**NPA**LTFAVALLG	FSYSAGAAM**NPA**RDLSPRLWS
MtAqp-2	AEP14556.2	323	AGGVTGAFL**NPA**IAVAFSVLG	LAYNAGAAL**NPS**RDLAPRLFT
MtAqp-3	AEP14557.1	326	SGGISGAHL**NPA**VTTTMLVMG	FSYNAGAAL**NPA**RDLAPRLFL
MtAqp-4	AEP14558.2	349	AGGISGGLL**NPA**ITFTIAFLG	FSYSAGAAM**NPA**RDISPRLWT
MtAqp-5	AEP14559.1	277	ICGVSGGHI**NPA**VSLGFLVTR	AIPLTGTSM**NPA**RSLGPAVLI
MtAqp-6	AEP14560.2	285	CFKSISAHF**NPV**ITIAALLLR	ARQATGGFL**NPM**RAFSLALFE
MtAqp-7	AEP14561.1	315	AGGVSGGHL**NPA**VTLAFAVIN	FGFNCGYPI**NPA**RDFGPRLFT
MtAqp-8	AEP14562.1	342	AGGVSGAFL**NPA**VALAFAVLG	LSYNAGAAM**NPS**RDLAPRFFS
MtAqp-9	AEP14563.1	281	GGAMSGAII**NPA**LTLAVALLG	FSYSTGAAM**NPA**RDFSPRLWS
MtAqp-10	AEP14564.1	409	AGGISGGHL**NPA**VSLAFAALG	YGFNCGYPI**NPA**RDMGPRLFT
MtAqp-11	AEP14565.2	270	TFTFQDGTC**DPS**ECYEKFCKR	GLFVSGGYF**NPT**LSFAMEYGC

Note: Signature NPA boxes are shown in bold.

tive stage and MtAqp-1 was slightly down-regulated. Although the fold expression changes were rather low, distinct expression patterns could be observed. The strongest expression shifts could be identified in inactive animals compared to active animals. The transcript expression during rehydration was, according to its physiological state, an intermediate stage between levels of the inactive and active stages.

From the raw qPCR data, transcript abundances could roughly be grouped into high abundance (MtAqp-3, -4, -5, -10), medium abundance (MtAqp-1, -2, -7) and low abundance transcripts (MtAqp-6, -8, -9, -11). According to the respective expression levels, the assembly of the trancriptome sequencing data recruited a different number of transcript reads for each AQP, ie, the cluster size. When comparing raw qPCR expression with sequencing data, we found good abundance correlations (file S1). For the low and medium abundant transcripts, we found these two data types correlating well, with Spearman's rank correlation coefficients of ~0.8. Even though the 454 dataset was derived from a normalized cDNA pool, the relative abundances of the low- and medium-expressed transcripts were not strongly affected. The cDNA had been normalized using the duplex specific nuclease method prior to sequencing.[44] The only exception was MtAqp-5, which showed increased sequence read coverage compared to qPCR data, probably caused by biased PCR amplification during the normalization procedure.

Phylogenetic analysis of *Milnesium tardigradum* aquaporins

Another aim of our study was to analyze the identified tardigrade AQPs from a phylogenetic perspective. To collect aquaporins from other invertebrate taxa we performed searches using various BLAST programs against nematode (*Caenorhadbditis elegans*) and insect genomes/ESTs (*Drosophila melanogaster*, *Aedes aegypti*, *Bombyx mori*, *Apis mellifera*, *Pediculus humanus corporis*), including AQPs from the chironomids *Polypedilum vanderplanki* and *Belgica antarctica*, that have been studied under aspects of dehydration and freeze tolerance.[45,46]

As the quality and coverage of genome sequence data differed a lot between organisms, some sequences had to be reconstructed manually using both genome and EST data. We were able identify additional putative unorthodox AQPs in the genomes of *B. mori* and *A. mellifera*. These new sequences were incorporated into our analysis (Fig. 3). As AQPs share a very low sequence identity but good protein structure conservation, we chose the structure-aware PROMALS3D program for generating multiple sequence alignments. PROMALS3D was able to retrieve AQP templates for all of the submitted sequences, which were mostly identical to those used by I-TASSER for structure prediction. The multiple sequence alignment was generated using a total of 48 insect and nematode AQP sequences.

Most of the *M. tardigradum* AQPs formed a clade outside of the insect sequences together with nematode sequences. Given the high speed of AQP evolution, there is strong evidence that some members of the tardigrade AQPs are recent duplication events, as they still show a high sequence identity. The unorthodox AQPs formed a separate clade with MtAqp-11 (AEP14565.2), each of the insects contributing a single member and *C. elegans* being present with three

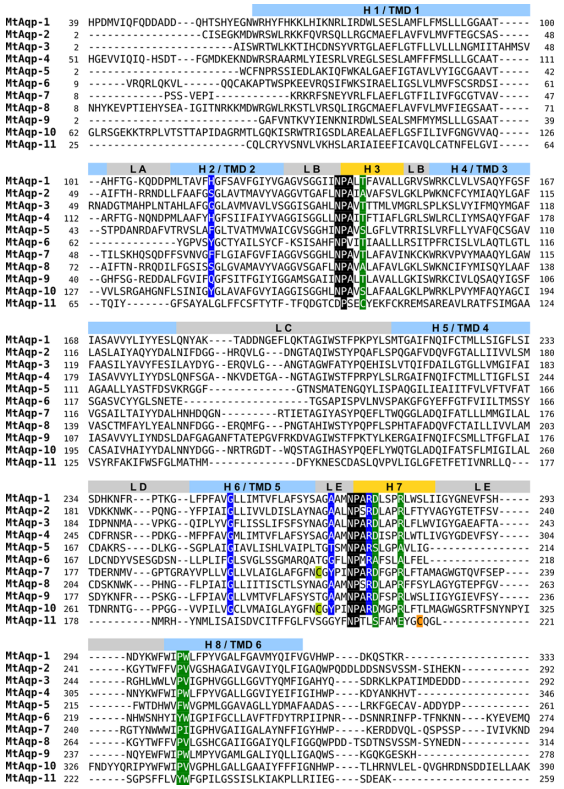

Figure 1. Truncated amino acid sequence alignment of *Milnesium tardigradum* aquaporins.
Notes: Channel transmembrane domains (TMD1–TMD6; light blue) and five inter-helical loops (LA–E; gray) connecting transmembrane domains. Helix 3 (H 3) and helix 7 (H 7) in loop B and E contain conserved NPA motifs (black) that form short α-helices that fold back into the membrane, forming the channel pore. Putative Hg^{2+}-sensitive cysteines are labeled light green. Signature cysteine residue for unorthodox aquaporins in MtAqp-11 is labeled in orange. Residues forming the aromatic/arginine restriction (ar/R) are labeled in dark blue. The conserved residues P1 to P5 according to Froger et al[72] are labeled in dark green.

members. The orthologue of *D. melanogaster big brain* (BIB; NP_476837.1) was also identified in *A. aegypti*, *A. mellifera, P. humanus corporis* and fragmentary evidence could be found in the *B. mori* genome.[47] The long C-terminus of BIB is probably a recent domain acquisition from non-coding DNA in insects, as evidenced by its highly repetitive character. We did not find any BIB orthologue in *M. tardigradum*.

In *D. melanogaster,* the genes coding for NP_611810.1, NP_611811.3, NP_611812.2 and NP_611813.1 as well as NP_725051.2 (DRIP) and NP_725052.1 are present as a cluster in the genome (file S2). The first 4 sequences form a single clade in the tree, whereas the latter 2 form separate clades shared between insects. This suggests that NP_611810-13 were duplicated in *D. melanogaster* alone, whereas the DRIP cluster duplication is older. With MtAqp-5 (AEP14559.2), *M. tardigradum* seems to possess only a single DRIP orthologue and its signature residues also predict a water-specific AQP.[48] A high baseline expression of this AQP and its conservation make it, comparably to DRIP, very likely to have similar important and conserved cellular functions. The second putatively water-specific tardigrade member (MtAqp-6, AEP14560.2) groups with three nematode members, 2 of them shown to be water specific.[49] The remaining tardigrade AQPs group with the expanded *C. elegans* aquaglyceroporin members, which is in agreement with their predicted specificities.

Table 2. Homology modeling results of *M. tardigradum* aquaporins by I-TASSER.

Name	C-score	TM-Score	RMSD	Top threading templates in PDB (% coverage/% identity)
MtAqp-1	–2.12	0.46 ± 0.15	11.4 ± 4.5 Å	1ldf (75/31), 1fx8 (75/31), 3c02 (72/30), 3iyz (67/25)
MtAqp-2	–1.19	0.57 ± 0.15	9.0 ± 4.6 Å	1ldf (77/34), 1fx8 (77/34), 3c02 (75/32), 3iyz (69/21)
MtAqp-3	–1.13	0.57 ± 0.14	8.9 ± 4.6 Å	1ldf (77/34), 1fx8 (77/34), 3c02 (74/33), 3iyz (68/29)
MtAqp-4	–2.27	0.45 ± 0.14	11.9 ± 4.4 Å	1ldf (72/28), 1fx8 (72/28), 3c02 (69/29), 3iyz (63/28)
MtAqp-5	0.17	0.74 ± 0.11	5.7 ± 3.6 Å	2b6p (92/35), 3iyz (80/45), 1z98 (87/38), 3d9s (88/39), 2d57 (81/44), 1j4n (86/39)
MtAqp-6	–0.74	0.62 ± 0.14	7.7 ± 4.3 Å	3d9s (82/24), 3iyz (74/29), 1z98 (82/24), 1j4n (81/25), 2d57 (74/29), 2b6p (87/24)
MtAqp-7	–1.02	0.59 ± 0.14	8.6 ± 4.5 Å	1ldf (78/38), 1fx8 (78/38), 3c02 (77/31), 3iyz (70/30)
MtAqp-8	–1.43	0.54 ± 0.15	9.7 ± 4.6 Å	1ldf (73/36), 1fx8 (73/36), 3c02 (70/31), 3iyz (65/24)
MtAqp-9	0.21	0.74 ± 0.11	5.6 ± 3.5 Å	1ldf (90/29), 1fx8 (90/28), 3c02 (86/27), 3iyz (79/28)
MtAqp-10	–2.51	0.42 ± 0.14	12.9 ± 4.2 Å	1ldf (62/37), 1fx8 (62/37), 3c02 (59/30), 3iyz (54/28)
MtAqp-11	–1.12	0.57 ± 0.14	8.5 ± 4.5 Å	2b6p (85/18), 3iyz (76/16), 1z98 (84/21), 3d9s (86/14), 2zz9 (76/16)

Notes: C-score is a confidence score for estimating the quality of predicted models/template modeling score: TM-Score; A TM-score > 0.5 indicates a model of correct topology and a TM-score < 0.17 means a random similarity. Structures: *E. coli* GlpF mutant (1ldf), *E. coli* GlpF (1fx8), *P. falciparum* PfAQP (3c02), *R. norvegicus* AQP4 (2d57), *R. norvegicus* AQP4 mutant (3iyz/2zz9), *S. oleracea* SoPIP2;1 (1z98), *H. sapiens* AQP5 (3d9s), *R. norvegicus* AQP4 (2d57), *B. taurus* AQP0 (2b6p), *B. taurus* AQP1 (1j4n).
Abbreviation: RMSD, root mean squared deviation.

Figure 3. Maximum Likelihood molecular phylogenetic analysis of selected ecdysozoan aquaporins.
Notes: The evolutionary history was inferred by using the maximum likelihood method using the generalized time reversible substitution model. The tree with the highest log likelihood (–62629.4925) is shown. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 59 nucleotide sequences. All ambiguous positions were removed for each sequence pair. Bootstrap support is given at the respective nodes. There were a total of 3018 positions in the final dataset. AQPs with known permeabilities are marked (● water; ● glycerol).

Unorthodox aquaporins contain a conserved N-terminal α-helical structure

Unorthodox aquaporins have not been studied to a great extent in comparison to the "classical" *H. sapiens* aquaporin-1 or *E. coli* Aqp-Z and GlpF, which have received much attention, not only from structural biologists. By analyzing the PROMALS3D multiple sequence alignment of the unorthodox AQPs, including 2 human members (AQP11 and AQP12), we found a common conserved α-helical structure directly upstream of the first transmembrane domain in MtAqp-11 and other unorthodox members (see File S5). This was in agreement with the I-TASSER secondary structure predictions, which also predicted α-helices at the respective positions in the proteins. The corresponding I-TASSER structure models for this region showed 2 short helices connected by a small loop (Fig. 4).

It is not clear whether this structure actually represents an upstream transmembrane domain or other peripheral membrane structure. In silico consensus transmembrane domain analysis using TOPCONS predicted transmembrane domains at the respective positions of the additional α-helices.[50] Consequently, in contrast to orthodox AQPs, only the C-termini of unorthodox AQPs would be cytosolic.

Table 3. Structures in Protein Data Bank (PDB) most similar to *M. tardigradum* aquaporin models.

Name	PDB hit	TM-score	RMSD	Identity	Coverage
MtAqp-1	1fx8	0.741	0.9 Å	30.7%	75.4%
MtAqp-2	1fx8	0.767	0.65 Å	33.6%	77.4%
MtAqp-3	1ldf	0.758	0.73 Å	34%	76.7%
MtAqp-4	1ldf	0.703	1.07 Å	27.9%	71.9%
MtAqp-5	3d9s	0.839	1.48 Å	38.9%	88.1%
MtAqp-6	2b6p	0.799	2.25 Å	25.2%	87.7%
MtAqp-7	1fx8	0.76	1.21 Å	38.9%	78.4%
MtAqp-8	1ldf	0.724	0.66 Å	35.6%	73.1%
MtAqp-9	1fx8	0.888	0.72 Å	28.9%	90%
MtAqp-10	1fx8	0.614	0.48 Å	36.5%	61.6%
MtAqp-11	2b6p	0.804	1.6 Å	17.9%	84.8%

Notes: Structures: *E. coli* GlpF mutant (1ldf), *E. coli* GlpF (1fx8), *H. sapiens* AQP5 (3d9s), *B. taurus* AQP0 (2b6p). Identified by TM-align.[71]

For *H. sapiens* AQP11 and AQP12, subcellular localization has been shown with typical features of unorthodox aquaporins, containing highly divergent NPA boxes and the conserved cysteine downstream of the second NPA motif.[41,51] In addition to the signature cysteine, which is indispensable for proper function as an amino acid change to serine results in a loss of function phenotype,[52] we have also identified a second highly conserved cysteine (Cys^{151} in MtAqp-11) in the long C loop that is situated directly opposite of the one in loop E (Cys^{218} in MtAqp-11) at the pore opening.

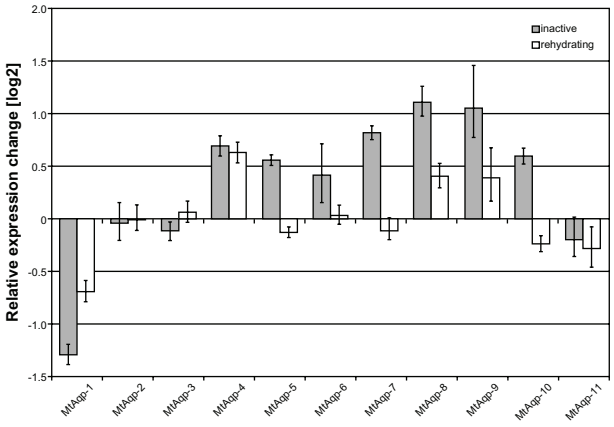

Figure 2. Relative expression changes of aquaporins in *Milnesium tardigradum* in response to dehydration and rehydration measured by quantitative real-time PCR.
Notes: Tardigrades were dehydrated for 62 hours (inactive; gray) and rehydrated at room temperature for 1 h (rehydrating; white). Aquaporin transcript expression differences are given as a $\log_2$ fold change relative to the expression in the active state. Error bars represent standard deviation of duplicate measurements.

As none of the homology modeling templates provided structural clues upstream of the first canonical transmembrane domain, the resulting N-terminus predictions of the unorthodox aquaporins are mostly ab initio models. Nevertheless, even if the predicted tertiary structures for the unorthodox proteins do not represent the native in vivo conformation, the resulting models of the N-termini show distinct similarities.

Discussion

Tardigrades have the ability to survive extreme desiccation to cope with the fluctuating water availability in the environment they inhabit. In this study, we have identified eleven transcripts coding for putative AQPs in *M. tardigradum*. Yet there is still the possibility that our data on aquaporins in *M. tardigradum* is still not exhaustive and that the genome contains an even greater number of yet unidentified AQP genes. This is especially possible given that we have identified uncharacterized putative unorthodox AQPs in other invertebrate genomes in this study. As all analyzed unorthodox AQPs share an additional upstream α-helical domain, we propose this as an additional hallmark feature, in combination with the signature cysteine. To gain further insight into aquaporin expression, the transcript levels of all AQPs were analyzed in *M. tardigradum* under normal (control), inactive and rehydration conditions. The expression levels of the majority of AQPs were almost unaffected in either inactive or rehydrating conditions compared to active animals. Although the inactive stage is not an actual physiological state, it is a very stable end-point measurement of the desiccation process, ie, all expression changes up to complete anhydrobiosis are preserved in the tun form. Peak expression changes could be observed in the inactive stage and differences are less pronounced during rehydration, although they are reminiscent of the inactive stage. MtAqp-2, -3 and -11 expression levels were completely unaffected by desiccation or rehydration. Thus, *M. tardigradums* AQPs might play a minor role during anhydrobiosis; otherwise regulation is mostly happening on a cellular level. Another possibility is that water transport can be efficiently fine-tuned by small changes in expression of the respective proteins or fusion of submembrane vesicles containing channel proteins, as AQPs greatly increase membrane permeability.[53] As Huang and colleagues speculated, the large repertoire of *C. elegans* AQPs could be an adaptive response to changing water abundance and osmolality

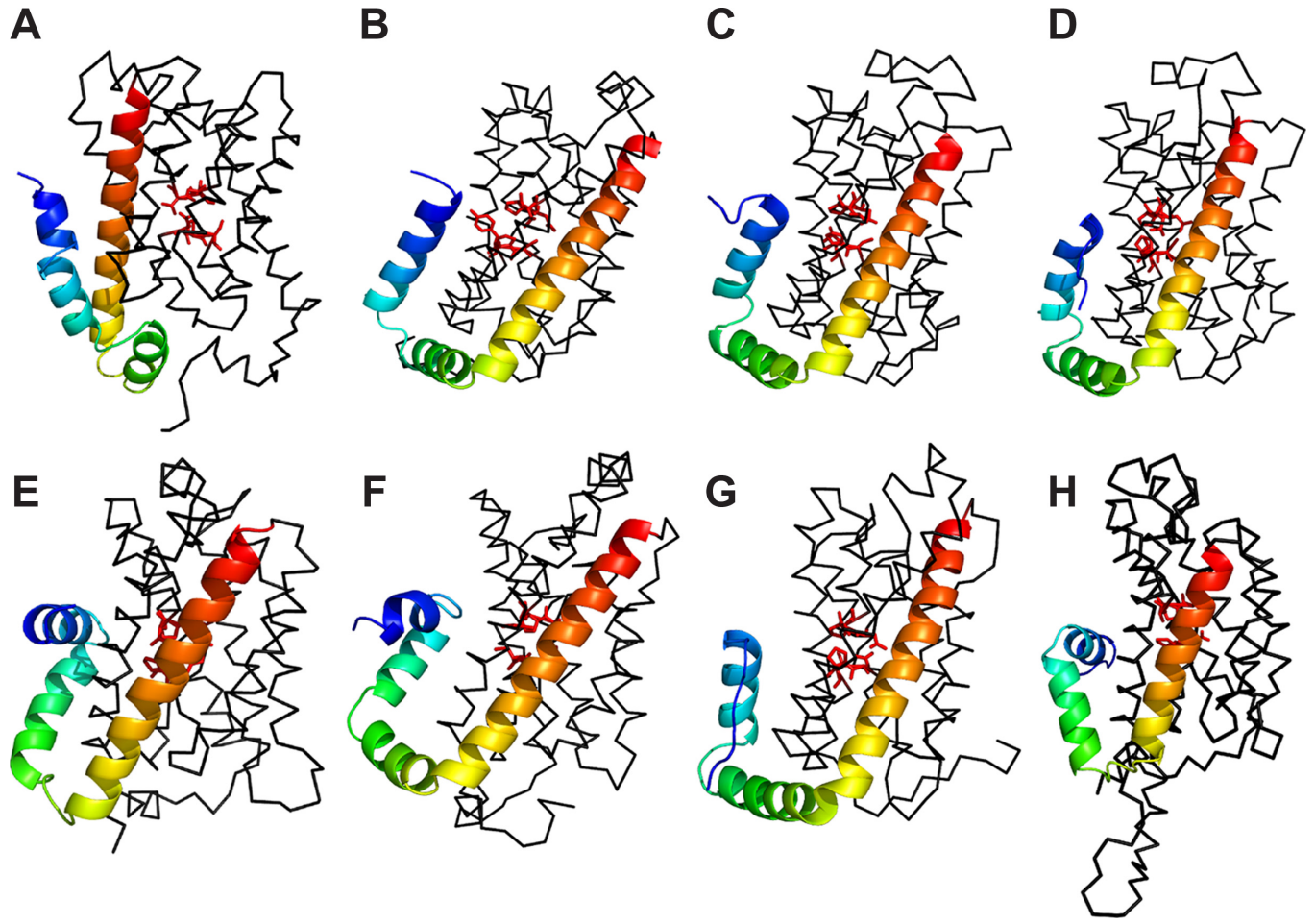

Figure 4. Conserved N-terminal helix in unorthodox aquaporins. I-TASSER structure predictions for protein sequences of (**A**) *M. tardigradum* aquaporin 11 (AEP14565; model2), (**B**) *D. melanogaster* (NP_001163140; model1) (**C**) *A. aegypti* (XP_001648319; model1), (**D**) *C. elegans* (NP_499821; model3), (**E**) *C. elegans* (NP_001021552; model1), (**F**) *C. elegans* (NP_496105; model2), (**G**) *H. sapiens* AQP11 (NP_766627; model1), (**H**) *H. sapiens* AQP12 (NP_945349; model1). Conserved N-terminal structure (dark blue to light green), first canonical transmembrane domain (yellow to dark red). The NPA motifs are given as red stick representations.

of its habitat.[49] Using RNA interference knock-down or deletion mutants of AQPs in *C. elegans,* no strong phenotypic effects could be observed. Unfortunately, the effect of AQP knockdown on *C. elegans* dauer larvae had not been investigated, as they are also capable of anhydrobiosis.[54] Natural osmotic stress in the soil micro-environment, which tardigrades are subjected to with their whole body surface, may require fine-tuning of water and solute homeostasis. This might be the reason for similar aquaglyceroporin member compositions in nematodes and tardigrades. Insects with a greater body size whose main way of water loss is excretion or evaporation seem to have neofunctionalized water specific aquaporins for glycerol transport, as they lack aquaglyceroporin type channels.[41] Unfortunately, genetic manipulation techniques and most cellular staining methods are not established for tardigrades yet, and testing whether our computational predictions of permeability characteristics are accurate will require further functional analyses. We have laid the foundation for these experiments in this study by providing a computational description of the AQP repertoire of *M. tardigradum*. The genomic organization of tardigrade AQPs and inclusion of data from the arthropod sister group of onychophorans, once available, might give a clearer picture on tardigrade AQP classification and evolution.

Conclusion

This is the first study to characterize aquaporin channels in a member of the tardigrade phylum. By mining the *M. tardigradum* transcriptome for aquaporin sequences, we identified eleven aquaporin transcripts belonging to the three major classes: canonical aquaporins, aquaglyceroporins and unorthodox

aquaporins. Although unorthodox aquaporins show some divergence to typical aquaporins and aquaglyceroporins, they share well-conserved signatures themselves. The tardigrade aquaporin repertoire is more similar to the distribution found in nematodes than insects, with an abundance of putative aquaglyceroporins. During anhydrobiosis, aquaporin transcript expression is not strongly affected in tardigrades.

Author Contributions

Conceived and designed the experiments: MAG, BM. Analyzed the data: MAG, BM, WW, SM, ROS. Wrote the first draft of the manuscript: BM. Contributed to the writing of the manuscript: MAG, BM, ROS, MF. Agree with manuscript results and conclusions: MAG, BM, WW, SM, ROS, MF. Jointly developed the structure and arguments for the paper: MAG, BM, MF. Made critical revisions and approved final version: ROS, MF. All authors reviewed and approved of the final manuscript.

Funding

This work was supported by the German Federal Ministry of Education and Research, BMBF, as part of FUNCRYPTA (FKY 0313838).

Competing Interests

Author(s) disclose no potential conflicts of interest.

Disclosures and Ethics

As a requirement of publication the authors have provided signed confirmation of their compliance with ethical and legal obligations including but not limited to compliance with ICMJE authorship and competing interests guidelines, that the article is neither under consideration for publication nor published elsewhere, of their compliance with legal and ethical guidelines concerning human and animal research participants (if applicable), and that permission has been obtained for reproduction of any copyrighted material. This article was subject to blind, independent, expert peer review. The reviewers reported no competing interests.

References

1. Jönsson KI. The evolution of life histories in holo-anhydrobiotic animals: a first approach. *Integr Comp Biol*. 2005;45(5):764–70.
2. Watanabe M. Anhydrobiosis in invertebrates. *Appl Entomol Zool*. 2006; 41(1):15–31.
3. Møbjerg N, Halberg K a, Jørgensen A, et al. Survival in extreme environments—on the current knowledge of adaptations in tardigrades. *Acta Physiol (Oxf)*. 2011;202(3):409–20.
4. Wełnicz W, Grohme MA, Kaczmarek Ł, Schill RO, Frohme M. Anhydrobiosis in tardigrades—the last decade. *J Insect Physiol*. 2011;57(5):577–83.
5. Guidetti R, Altiero T, Rebecchi L. On dormancy strategies in tardigrades. *J Insect Physiol*. 2011;57(5):567–76.
6. Keilin D. The problem of anabiosis or latent life: history and current concept. *Proc R Soc Lond B Biol Sci*. 1959;150(939):149–91.
7. Clegg JS. Cryptobiosis—a peculiar state of biological organization. *Comp Biochem Physiol B Biochem Mol Biol*. 2001;128(4):613–24.
8. Alpert P. Constraints of tolerance: why are desiccation-tolerant organisms so small or rare? *J Exp Biol*. 2006;209(Pt 9):1575–84.
9. Guidetti R, Jönsson KI. Long-term anhydrobiotic survival in semi-terrestrial micrometazoans. *J Zool (Lond)*. 2002;257(2):181–7.
10. Hengherr S, Brümmer F, Schill RO. Anhydrobiosis in tardigrades and its effects on longevity traits. *J Zool (Lond)*. 2008;275(3):216–20.
11. Bertolani R, Guidetti R, Jönsson KI, Altiero T, Boschini D, Rebecchi L. Experiences with dormancy in tardigrades. *J Limnol*. 2004;63(Suppl 1):16–25.
12. Horikawa DD, Sakashita T, Katagiri C, et al. Radiation tolerance in the tardigrade Milnesium tardigradum. *Int J Radiat Biol*. 2006;82(12):843–8.
13. Jönsson KI, Rabbow E, Schill RO, Harms-Ringdahl M, Rettberg P. Tardigrades survive exposure to space in low Earth orbit. *Curr Biol*. 2008; 18(17):R729–31.
14. Hengherr S, Worland MR, Reuner A, Brümmer F, Schill RO. Freeze tolerance, supercooling points and ice formation: comparative studies on the subzero temperature survival of limno-terrestrial tardigrades. *J Exp Biol*. 2009;212(Pt 6):802–7.
15. Hengherr S, Worland MR, Reuner A, Brümmer F, Schill RO. High-temperature tolerance in anhydrobiotic tardigrades is limited by glass transition. *Physiol Biochem Zool*. 2009;82(6):749–55.
16. Wright JC. Desiccation tolerance and water-retentive mechanisms in tardigrades. *J Exp Biol*. 1989;142:267–92.
17. Zeuthen T. How water molecules pass through aquaporins. *Trends Biochem Sci*. 2001;26(2):77–9.
18. King LS, Kozono D, Agre P. From structure to disease: the evolving tale of aquaporin biology. *Nat Rev Mol Cell Biol*. 2004;5(9):687–98.
19. Preston GM, Carroll TP, Guggino WB, Agre P. Appearance of water channels in Xenopus oocytes expressing red cell CHIP28 protein. *Science*. 1992; 256(5055):385–7.
20. Zardoya R, Villalba S. A phylogenetic framework for the aquaporin family in eukaryotes. *J Mol Evol*. 2001;52(5):391–404.
21. Zardoya R. Phylogeny and evolution of the major intrinsic protein family. *Biol Cell*. 2005;97(6):397–414.
22. Campbell EM, Ball A, Hoppler S, Bowman AS. Invertebrate aquaporins: a review. *J Comp Physiol B*. 2008;178(8):935–55.
23. Walz T, Hirai T, Murata K, et al. The three-dimensional structure of aquaporin-1. *Nature*. 1997;387(6633):624–7.
24. Verbavatz JM, Brown D, Sabolić I, et al. Tetrameric assembly of CHIP28 water channels in liposomes and cell membranes: a freeze-fracture study. *J Cell Biol*. 1993;123(3):605–18.
25. Wistow GJ, Pisano MM, Chepelinsky AB. Tandem sequence repeats in transmembrane channel proteins. *Trends Biochem Sci*. 1991;16(5): 170–1.
26. Pao GM, Wu LF, Johnson KD, et al. Evolution of the MIP family of integral membrane transport proteins. *Mol Microbiol*. 1991;5(1):33–7.
27. Maurel C, Reizer J, Schroeder JI, Chrispeels MJ, Saier MH. Functional characterization of the Escherichia coli glycerol facilitator, GlpF, in Xenopus oocytes. *J Biol Chem*. 1994;269(16):11869–72.
28. Wu B, Beitz E. Aquaporins with selectivity for unconventional permeants. *Cell Mol Life Sci*. 2007;64(18):2413–21.
29. Yasui M, Hazama A, Kwon TH, Nielsen S, Guggino WB, Agre P. Rapid gating and anion permeability of an intracellular aquaporin. *Nature*. 1999; 402(6758):184–7.
30. Morishita Y, Matsuzaki T, Hara-chikuma M, et al. Disruption of aquaporin-11 produces polycystic kidneys following vacuolization of the proximal tubule. *Mol Cell Biol*. 2005;25(17):7770–9.

31. Itoh T, Rai T, Kuwahara M, et al. Identification of a novel aquaporin, AQP12, expressed in pancreatic acinar cells. *Biochem Biophys Res Commun*. 2005; 330(3):832–8.
32. Verkman AS. More than just water channels: unexpected cellular roles of aquaporins. *J Cell Sci*. 2005;118(Pt 15):3225–32.
33. Papadopoulos MC, Saadoun S, Verkman AS. Aquaporins and cell migration. *Pflugers Arch*. 2008;456(4):693–700.
34. Jung JS, Preston GM, Smith BL, Guggino WB, Agre P. Molecular structure of the water channel through aquaporin CHIP. The hourglass model. *J Biol Chem*. 1994;269(20):14648–54.
35. Murata K, Mitsuoka K, Hirai T, et al. Structural determinants of water permeation through aquaporin-1. *Nature*. 2000;407(6804):599–605.
36. Wree D, Wu B, Zeuthen T, Beitz E. Requirement for asparagine in the aquaporin NPA sequence signature motifs for cation exclusion. *FEBS J*. 2011; 278(5):740–8.
37. Wang Y, Schulten K, Tajkhorshid E. What makes an aquaporin a glycerol channel? A comparative study of AqpZ and GlpF. *Structure*. 2005;13(8): 1107–18.
38. Beitz E, Wu B, Holm LM, Schultz JE, Zeuthen T. Point mutations in the aromatic/arginine region in aquaporin 1 allow passage of urea, glycerol, ammonia, and protons. *Proc Natl Acad Sci U S A*. 2006;103(2):269–74.
39. Wu B, Steinbronn C, Alsterfjord M, Zeuthen T, Beitz E. Concerted action of two cation filters in the aquaporin water channel. *EMBO J*. 2009;28(15): 2188–94.
40. Mali B, Grohme MA, Förster F, et al. Transcriptome survey of the anhydrobiotic tardigrade Milnesium tardigradum in comparison with Hypsibius dujardini and Richtersius coronifer. *BMC Genomics*. 2010;11:168.
41. Ishibashi K, Kondo S, Hara S, Morishita Y. The evolutionary aspects of aquaporin family. *Am J Physiol Regul Integr Comp Physiol*. 2011;300(3): R566–76.
42. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*. 2008;9:40.
43. Berman HM, Westbrook J, Feng Z, et al. The protein data bank. *Nucleic Acids Res*. 2000;28(1):235–42.
44. Zhulidov PA, Bogdanova EA, Shcheglov AS, et al. Simple cDNA normalization using kamchatka crab duplex-specific nuclease. *Nucleic Acids Res*. 2004;32(3):e37.
45. Kikawada T, Saito A, Kanamori Y, et al. Dehydration-inducible changes in expression of two aquaporins in the sleeping chironomid, Polypedilum vanderplanki. *Biochim Biophys Acta*. 2008;1778(2):514–20.
46. Goto SG, Philip BN, Teets NM, Kawarasaki Y, Lee RE, Denlinger DL. Functional characterization of an aquaporin in the Antarctic midge Belgica antarctica. *J Insect Physiol*. 2011;57(8):1106–14.
47. Rao Y, Jan LY, Jan YN. Similarity of the product of the Drosophila neurogenic gene big brain to transmembrane channel proteins. *Nature*. 1990;345(6271):163–7.
48. Kaufmann N, Mathai JC, Hill WG, Dow JAT, Zeidel ML, Brodsky JL. Developmental expression and biophysical characterization of a Drosophila melanogaster aquaporin. *Am J Physiol Cell Physiol*. 2005;289(2): C397–407.
49. Huang CG, Lamitina T, Agre P, Strange K. Functional analysis of the aquaporin gene family in Caenorhabditis elegans. *Am J Physiol Cell Physiol*. 2007;292(5):C1867–73.
50. Bernsel A, Viklund H, Hennerdal A, Elofsson A. TOPCONS: consensus prediction of membrane protein topology. *Nucleic Acids Res*. 2009;37: W465–8.
51. Ishibashi K, Koike S, Kondo S, Hara S, Tanaka Y. The role of a group III AQP, AQP11 in intracellular organelle homeostasis. *J Med Invest*. 2009; 56(Suppl):312–7.
52. Tchekneva EE, Khuchua Z, Davis LS, et al. Single amino acid substitution in aquaporin 11 causes renal failure. *J Am Soc Nephrol*. 2008;19(10):1955–64.
53. Zeidel ML, Ambudkar SV, Smith BL, Agre P. Reconstitution of functional water channels in liposomes containing purified red cell CHIP28 protein. *Biochemistry*. 1992;31(33):7436–40.
54. Erkut C, Penkov S, Khesbak H, et al. Trehalose renders the dauer larva of Caenorhabditis elegans resistant to extreme desiccation. *Curr Biol*. 2011; 21(15):1331–6.
55. Marshall OJ. PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR. *Bioinformatics*. 2004;20(15): 2471–2.
56. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*. 2001;25(4):402–8.
57. Sobreira TJP, Gruber A. Sequence-specific reconstruction from fragmentary databases using seed sequences: implementation and validation on SAGE, proteome and generic sequencing data. *Bioinformatics*. 2008;24(15): 1676–80.
58. Pertea G, Huang X, Liang F, et al. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics*. 2003;19(5):651–2.
59. Huang X, Madan A. CAP3: A DNA sequence assembly program. *Genome Res*. 1999;9(9):868–77.
60. Gish W, States DJ. Identification of protein coding regions by database similarity search. *Nat Genet*. 1993;3(3):266–72.
61. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25(17):3389–402.
62. Margulies M, Egholm M, Altman WE, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 2005;437(7057):376–80.
63. Gilles A, Meglécz E, Pech N, Ferreira S, Malausa T, Martin J-F. Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. *BMC Genomics*. 2011;12(1):245.
64. Milne I, Bayer M, Cardle L, et al. Tablet—next generation sequence assembly visualization. *Bioinformatics*. 2010;26(3):401–2.
65. Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL. Domain enhanced lookup time accelerated BLAST. *Biol Direct*. 2012;7:12.
66. Pei J, Kim B-H, Grishin NV. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res*. 2008;36(7): 2295–300.
67. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res*. 2006;34:W609–12.
68. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011;28(10):2731–9.
69. The pyMOL Molecular Graphics System, Version~1.5.0.1. 2012; Schrödinger, LLC.
70. Duan J, Li R, Cheng D, et al. SilkDB v2.0: a platform for silkworm (Bombyx mori) genome biology. *Nucleic Acids Res*. 2010;38:D453–6.
71. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res*. 2005;33(7):2302–9.
72. Froger A, Tallur B, Thomas D, Delamarche C. Prediction of functional residues in water channels and related proteins. *Protein Sci*. 1998;7(6): 1458–68.

Supplementary Data

File S1 Correlation between qPCR data and cluster sizes of assembled transcripts.

File S2 Aquaporin clusters in the genomes of *D. melanogaster*, *A. aegypti* and *A. mellifera*.

File S3 Primers used for quantitative real-time PCR and raw crossing point values of qPCR experiments.

File S4 Summary of I-TASSER output for *M. tardigradum* aquaporins and unorthodox aquaporins analyzed in this study, including predicted structures.

File S5 Promals3D multiple sequence alignment for unorthodox aquaporins analyzed in this study.

File S6 *A. mellifera* XP_001122456.2 mRNA and proposed annotation error.