Evolutionary conservation of candidate osmoregulation genes in plant phloem sap-feeding insects

X. Jing*, T. A. White†, J. Luan*, C. Jiao‡, Z. Fei‡ and A. E. Douglas*§

*Department of Entomology, Cornell University, Ithaca, NY, USA; †Lancaster Environment Centre, Lancaster University, Lancaster, UK; ‡Boyce Thompson Institute for Plant Research, Cornell University, Ithaca, NY, USA; and §Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, USA

Abstract

The high osmotic pressure generated by sugars in plant phloem sap is reduced in phloem-feeding aphids by sugar transformations and facilitated water flux in the gut. The genes mediating these osmoregulatory functions have been identified and validated empirically in the pea aphid Acyrthosiphon pisum: sucrase 1 (SUC1), a sucrase in glycoside hydrolase family 13 (GH13), and aquaporin 1 (AQP1), a member of the Drosophila integral protein (DRIP) family of aquaporins. Here, we describe molecular analysis of GH13 and AQP genes in phloem-feeding representatives of the four phloem-feeding groups: aphids (Myzus persicae), coccids (Planococcus citri), psyllids (Diaphorina citri, Bactericera cockerelli), and whiteflies (Bemisia tabaci MEAM1 and MED). A single candidate GH13-SUC gene and DRIP-AQP gene were identified in the genome/transcriptome of most insects tested by the criteria of sequence motif and gene expression in the gut. Exceptionally, the psyllid Ba. cockerelli transcriptome included a gut-expressed Pyrocoelia rufa integral protein (PRIP)-AQP, but has no DRIP-AQP transcripts, suggesting that PRIP-AQP is recruited for osmoregulatory function in this insect. This study indicates that phylogenetically related SUC and AQP genes may generally mediate osmoregulatory functions in these diverse phloem-feeding insects, and provides candidate genes for empirical validation and development as targets for osmotic disruption of pest species.

Keywords: α-glucosidase, aphid, aquaporin, osmoregulation, psyllid, sucrase, whitefly.

Introduction

The capacity to utilize plant phloem sap as the sole food source throughout the life cycle has evolved multiple times in insects of the order Hemiptera but, apparently, in no other animals (Dolling, 1991; Douglas, 2003). There is widespread recognition that a key physiological barrier to phloem feeding is the osmotic challenge posed by the very high sugar content of phloem sap and that critical features of hemipteran insects enabling phloem utilization are localized to the gut (Fisher et al., 1984; Dolling, 1991; Cohen, 2013). In particular, efficient processing of the phloem sap is facilitated by anatomical structures (eg filter chamber) that mediate the rapid transfer of water between the proximal and distal regions of the gut and by the absence of a distensible crop (Goodchild, 1966). In the pea aphid, Acyrthosiphon pisum, these morphological traits are paralleled by two key molecular functions, mediated by osmoregulatory genes (Huang et al., 2015): a sucrase-transglucosidase, coded by the gene sucrase 1 (SUC1), which transforms excess sucrose to long-chain oligosaccharides in the distal midgut (Price et al., 2007), and an aquaporin, coded by the gene aquaporin 1 (AQP1), which facilitates the osmotic movement of water from the distal gut to the proximal midgut (stomach) and promotes equilibration of the osmotic pressure between the gut lumen and the haemolymph (Shakesby et al., 2009). SUC1 is a member of α-glucosidase hydrolase family 13 (Price et al., 2007; see also www.cazypedia.org) and AQP1 is a member of the Drosophila integral protein (DRIP) subfamily of insect water-specific aquaporins (Kaufmann et al., 2005; Campbell et al., 2008; Goto et al., 2011;
Wallace *et al.*, 2012; Finn *et al.*, 2015). The genes underpinning sugar transformations in the gut of other hemipterans have not been investigated, although biochemical data are available for sugar-modifying enzymes in the whitefly *Bemisia tabaci* (Salvucci *et al.*, 1997; Salvucci, 2003). Water-specific aquaporin genes with predicted osmoregulatory function in the gut have been identified in *B. tabaci* MEAM1 (Mathew *et al.*, 2011), the potato/tomato psyllid *Bactericera cockerelli* (Ibanez *et al.*, 2014), the leafhopper *Cicadella viridis* (Le Caherec *et al.*, 1996) and the western tarnished plant bug *Lygus hesperus* (Fabrick *et al.*, 2014).

The overall goal of our research is to investigate the evolutionary relationship amongst gut osmoregulation genes in phloem-feeding hemipterans, focusing on representatives of the four sternorrhynchan superfamilies (Aleyrodoidea, Aphidoidea, Coccoidea and Psylloidea) for which genomic or transcriptomic data are available. The common ancestor of the Sternorrhyncha was probably a phloem-feeder (Dolling, 1991), raising the possibility that homologous genes mediate the osmoregulatory function in the different insect groups. Alternatively, as for amino acid transporters (Duncan *et al.*, 2014), the patterns of osmoregulation gene diversification and tissue distribution of expression may vary amongst lineages. Resolution of this issue is important to understand the evolutionary history of the phloem-feeding insects. Gene identification is also of practical value because these genes have potential as targets for novel strategies to control phloem-feeding insects (Douglas, 2003).

Recently, genomic and large transcriptomic data sets have become available for representative species of the four superfamilies of Sternorrhyncha (Table 1). In this study, we combined phylogenetic reconstruction of osmoregulation genes with analysis of functionally important sequence motifs and gene expression data in order to investigate whether the osmoregulation genes in the different sternorrhynchan groups are homologous or genetically distinct.

### Results

**AQP genes**

Examination of the genomes/transcriptomes of seven phloem-feeding sternorrhynchan insects (Table 1) revealed three to eight sequences with the AQP domain (PF00230) in each species (Table 2; Fig. 1). The primary focus of this study was the water-specific AQPs expressed in the insect gut, which are members of the subclade DRIP (Kaufmann *et al.*, 2005) and *Pyrocoelia rufa* integral protein (PRIP; Lee *et al.*, 2001). DRIPs and PRIPs have the sequence motifs common to empirically determined water-specific channels, comprising two asparagine-proline-alanine (NPA) boxes that define the channel and phenylalanine-histidine-alanine/serine/threonine-arginine (F-H-A/S/T-R) for the aromatic/arginine selectivity filter, together with six predicted motifs.
transmembrane domains characteristic of AQP proteins (Heymann & Engel, 2000; Agre, 2006; Fig. 1; Table S1; Fig. S1). The empirically validated water-specific gut aquaporin AQP1 of the aphid Ac. pisum and whitefly Be. tabaci are DRIPs (Shakesby et al., 2009; Mathew et al., 2011). The single homologous DRIP sequences in the mealybug Planococcus citri and psyllid Diaphorina citri possess the sequence motifs of water channels and are predicted to function as the osmoregulatory gut aquaporin. Both species of Be. tabaci have a second DRIP aquaporin gene, which we annotate as AQP2, but analysis of published RNAseq data of adult Be. tabaci MEAM1 indicated that the expression of AQP1 in the gut is 31-fold greater than for AQP2 (fragments per kilobase of exon per million fragments mapped (FPKM) values of 373 and 12, respectively) and only AQP1, and not AQP2, has enriched expression in the gut relative to the whole body (Table 3A). This result suggests that the AQP2 protein does not play a significant osmoregulatory role in adult whitefly.

The water-specific channels of the PRIP subclade were represented by single genes in the psyllids (AQP2) and whiteflies (AQP3), but no gene in the aphids or mealybugs (Fig. 1). The transcript of AQP2 of Ba. cockerelli, but not AQP3 of the whitefly Be. tabaci MEAM1, was significantly enriched in the gut, relative to the whole body (Ibanez et al., 2014; Table 3A). Taken with the apparent absence of AQP1 from the Ba. cockerelli transcriptome, these data raise the possibility that the gut osmoregulatory function is mediated by a PRIP in psyllids, either exclusively or, for Di. citri, in conjunction with a DRIP-AQP.

\(\alpha\)-glycohydrolases

The only experimentally validated sucrose hydrolase gene with osmoregulatory function is the sucrase-transglucosidase SUC1 \([\text{ACYPI000002}:\text{a member of}\ \alpha\text{-glucoside hydrolase family 13 (GH13, Pfam: PF00128)}]\) of the pea aphid, Ac. pisum (Price et al., 2007). The number of genes coding the PF00128 domain varies widely amongst the phloem-feeding insect species:

### Osmoregulation genes in phloem-feeding insects

© 2016 The Authors. *Insect Molecular Biology* published by John Wiley & Sons Ltd on behalf of The Royal Entomological Society, 25, 251–258
species (Table 2, Table S2), from just seven genes in the psyllid Ba. cockerelli to 33 and 42 genes recovered from Be. tabaci MEAM1 and MED, respectively.

The Bayesian analysis of GH13 proteins assigned the full complement of Drosophila melanogaster and Anopheles gambiae genes to a well-supported subgroup (Fig. 2), indicating that (in contrast to the aquaporin genes) much of the evolutionary diversification in GH13 probably occurred subsequent to the Diptera/Hemiptera common ancestor.

The Ac. pism SUC1 was assigned to a well-supported clade that comprises two groups, one of aphid proteins and the other of mealybug proteins, indicative of independent diversification in the two insect groups. The MpgGH13-1 and MpgGH13-2 of Myzus persicae cluster with Ac. pism SUC1, and are strong candidates for sucrase-transglucosidase because they have both of the required traits: gene expression is enriched in the gut (Price et al., 2007), as confirmed by analysis of the gut RNAseq data set used in this study (Table 3B), and the cognate protein bears both the catalytic site residues (Table S2) and a predicted signal peptide (Fig. 2; Table S2) required for extracellular localization and function in sugar transformations in the gut lumen. PcgGH13-1 of the mealybug Pl. citri is the sole protein in this clade with a predicted signal peptide, but its localization has not been investigated. The sister group of this clade of aphid and mealybug genes includes a single protein in the psyllids (DcGH13-1 and BcGH13-1), five proteins in Be. tabaci MEAM1 and six proteins in Be. tabaci MED (Fig. 2). The Ba. cockerelli protein BcGH13-1 is a strong candidate for osmoregulatory function because it has enriched expression in the gut (Ibanez et al., 2014; Tzin et al., 2015), and possesses the catalytic residues and predicted signal peptide (Fig. 2; Table S2). Similarly, the orthologous protein GH13-1 in Di. citri has a predicted signal peptide (D-score = 0.77; Table S2), suggesting that the GH13-1s in the two psyllid species are functionally equivalent. All the GH13 genes in Be. tabaci under consideration have the catalytic residues (Table S2). In Be. tabaci/MEAM1, only one gene (GH13-1) has elevated transcript abundance in the RNAseq data set for the gut relative to the whole body (Table 3) and a predicted signal peptide (Fig. 2, Table S2), consistent with a role in sugar transformations in the gut.

**Discussion**

In this study, we built on empirically validated osmoregulatory gene identifications in Ac. pism to identify candidate AQP and SUC genes with osmoregulatory function, using the following inclusive criteria: that the AQP gene has the sequence motifs of a water channel; the SUC gene is a member of glucohydrolase family-13 with the predicted catalytic residues and codes a predicted signal peptide (required for localization to the gut lumen); and, where data were available, the gene is expressed in the gut. Our analysis correctly identified the empirically validated Be. tabaci AQP gene (Fig. 1; Mathew et al., 2011) and yielded single candidate genes in most insects, with phylogenetic patterns that are generally indicative of strong genetic conservation of osmoregulatory function within the phloem-feeding Sternorrhyncha.

Our findings raise three general issues. The first relates to all the data: that the identification of genes with enriched expression in the gut was conducted on specific life stages of insects reared in the laboratory, and the possibility cannot be excluded that other genes are expressed in the gut at different life stages (eg different morphs of the aphids, pre-adult stages of the whiteflies) or under different environmental conditions, especially in the field. The second issue is the evidence that a PRIP-AQP may have an osmoregulatory function in psyllids, in contrast to the apparently exclusive assignment of osmotic function to DRIP-AQPs in the other sternorrhynchan insects investigated (Shakesby et al., 2009; Mathew et al., 2011). Consistent with the role of this gene in osmoregulation, its expression is enriched in the insect gut relative to the whole body (Ibanez et al., 2014; Tzin et al., 2015), and orally delivered RNA interference (RNAi) against this gene causes elevated haemolymph osmotic pressure and insect mortality (Tzin et al., 2015), indicating that the osmotic function of the RNAi-treated insects was perturbed. Other data,
however, give cause for caution: that the apparent absence of a DRIP-AQP gene is based on transcriptome data, and definitive evidence must await the availability of the sequenced genome for *Ba. cockerelli*; and that the uniform expression pattern in the gut (Ibanez et al., 2014) is contrary to its predicted role in promoting water flux between specific proximal and distal regions of the gut (Goodchild, 1966; Douglas, 2003). The third issue relates to the predicted SUC genes. We cannot exclude the involvement of glucohydrolases of other families (not investigated in this study) in the sucrase and transglucosidase functions because, although GH13 apparently predominates in insect digestive enzymes (Terra et al., 1996) including the pea aphid SUC1 (Price et al., 2007), members of different carbohydrase families could contribute to the transformations of ingested sugars in other phloem-feeding insects. For example, the mammalian digestive sucrase activity is mediated by a GH31 isomaltase-sucrase that, unlike all known insect digestive sucrases, comprises two domains with distinct carbohydrate catalytic activities (Sim et al., 2010).

One striking result of this study was the duplication of the DRIP-AQP in *Be. tabaci* and inclusion of the candidate *Be. tabaci* sucrase-hydrolysing gene within a gene expansion of five/six (MEAM1/MED) GH13 genes, compared with the single candidate genes of the psyllids. Genome-scale data for multiple whitefly species will be required to establish the chronological and phylogenetic timing of these gene diversification events, building on the current estimated divergence time of 0.5–29 million years ago for *Be. tabaci* MEAM1 and MED (Boykin et al., 2013). These results raise multiple questions, including whether the duplicated genes differ in pattern of expression (as discussed above) and functional traits, such as substrate specificity and kinetics of enzymatic or transport flux, and how these functional consequences may influence the osmoregulatory capacity of the insect. Two features of the biology of whiteflies, including *Be. tabaci*, may be relevant.

---

Figure 2. Phylogenetic relationships of α-glucosidases resolved by Bayesian inference. The sequences within the dotted box are the hemipteran sequences allied with the experimentally validated digestive sucrase of *A. pismum* (grey shaded). * sequences with predicted signal peptide [discrimination score (SIGNALP 4.1) provided in Table S2]. Posterior probabilities are given on the branches.
First, these insects probably experience very severe hazard of dehydration because the high osmotic pressure of ingested phloem sap is compounded by enhanced risks of water loss linked to its low latitude distribution and considerably smaller size than the other insects investigated in this study. The possibility that these traits have selected for functional flexibility in osmoregulation through gene amplification and diversification can be tested by functional analysis of the different genes in *Be. tabaci*, together with detailed phylogenetic analysis of these genes across the whitefly radiation. Possibly related to this issue, indications of unusual sugar transformations in *Be. tabaci* come from the discovery that trehalulose [O-α-D-glucopyranosyl-(1, l)-D-fructose] is a major honeydew sugar in this species (Byrne & Miller, 1990; Hendrix et al., 1992), but little is known about the genetic and enzymatic basis of this capability and its significance for osmoregulation and water conservation in whiteflies (Salvucci et al., 1997).

In conclusion, this study has defined the candidate osmoregulatory genes in representatives of the phloem-feeding Sternorrhyncha. Our results suggest that the genetic basis of osmoregulation is conserved across the different taxa, in contrast to some other molecular functions for which phylogenetically distant genes are recruited to equivalent biological functions in different sternorrhynchan groups. Furthermore, our study focused principally on major pest species. Linked to the importance of osmoregulation for the fecundity and survival of phloem-feeding insects (Karley et al., 2005; Tzin et al., 2015), these genes have potential as molecular targets for novel strategies to control phloem-feeding insects.

**Experimental procedures**

**Transcriptomic analysis of *M. persicae***

The green peach aphid *M. persicae* Sulzer clone SB10/1, derived from a single parthenogenetic female from a long-term laboratory colony originally obtained from S. Gray (United States Department of Agriculture (USDA) Plant Soil and Nutrition Laboratory, Ithaca, NY, USA), was maintained on pre-flowering tobacco (*Nicotinia tabaci* cv. Xanthi) at 24/20°C [light/dark (L/D)] and 14L: 10D photoperiod. Total RNA was extracted from five whole bodies and 60 dissected guts of adult, apterous aphids using the RNeasy Mini kit (Qiagen, Valencia, CA, USA) following the manufacturer’s instructions. Following removal of contaminating genomic DNA using the DNA-Free kit (Ambion/Thermo Fisher, Rochester, NY, USA), the purity of the extracted RNA was checked with Bioanalyzer (Agilent, Santa Clara, CA, USA) 2100 ([RNA Integrity Number (RIN) ≥ 8.5, customized to insect material], and quantified by the Quant-iT™ PicoGreen® Kit (Thermo Fisher, Rochester, NY, USA). A multiplex library was built with 1 μg RNA from each extraction following the TruSeq RNA Samples Prep v2 LS Protocol (Illumina, San Diego, CA, USA) and 100-bp single-end reads were sequenced by an Illumina Hi-Seq2000 Platform. The reads that passed quality control of the Illumina pipeline v. 1.8 were used for subsequent analysis.

**Transcriptome assembly and expression profiling**

Raw reads for *Be. tabaci* MED and MEAM1 [National Center for Biotechnology Information (NCBI) SRR39231, SRR835756, SRR835757 and SRR059302]. Ba. cockerelli [unpublished data provided by C. Tamborindegey, Texas A&M University (Nachappa et al., 2012)], *Pl. citri* (SRP021919; Husnik et al., 2013) and *M. persicae* gut and whole body (this study; NCBI Short Read Archive: accession no. SRP061935) were assembled by the TRINITY package (trinityrnaseq_r20131110) with default parameters (Haas et al., 2013). To remove redundant contigs, pipelines were used to align the reads to reference indexes of the *Ac. pisum* genome (version 2), and an annotation file provided coordinates of gene transcript boundaries (Trapnell et al., 2012; Kim et al., 2013). Expression differences between gut and whole body in *Ac. pisum, M. persicae* and *Be. tabaci* MEAM1 were calculated by CUFFDIFF (Trapnell et al., 2012). To determine expression differences between the gut and whole body of *Be. tabaci* MEAM1, raw reads [NCBI SRR835757 (Ye et al., 2014) and SRR1523522 (Luan et al., 2015)] were processed using TRIMMOMATIC (Bolger et al., 2014). The cleaned reads were *de novo* assembled using TRINITY (Grabherr et al., 2011) with minimum k-mer coverage set to 2. The assembled TRINITY contigs were further clustered with iAssembler (Zheng et al., 2011) to remove redundant transcripts. The cleaned reads were then mapped to the final assembled contigs using BOWTIE using the single end mapping mode and allowing up to two mismatches. Following mapping, raw counts and FPKM values of each contig were derived. The significance of differential expression was determined using NOISEQ (Tarazona et al., 2011).

**Identification of osmoregulation genes**

Candidate genes coding aquaporins and sucrases were searched in the translated proteomes [longest open reading frame (ORF) for transcriptomes] of 10 insect species (Table 1). The sequences with an aquaporin domain (PF00230) or α-amylase (family 13 of glycosyl hydrolases) domain (PF00128) were identified separately using HMMER v. 3.0 (Eddy, 2011). To obtain an exhaustive inventory of genes, amino acid sequences translated in all six reading frames from transcriptomic assemblies by ORFPREDICTOR (Min et al., 2005) were also searched using HMMER and the nonlongest ORF sequences with the corresponding domain were also included.

Where multiple sequences that included identical regions were obtained, only the longest isoform was retained for downstream analysis (Tables S1 and S2). For transcriptome data, the divergence times between identified transcripts for each function in each species were estimated by the pairwise rate of synonymous substitutions (dS) and the transcripts with dS < 0.25 were clustered and the longest was used to represent the corresponding locus (Duncan et al., 2014). The candidate transcripts of *Pl. citri* were checked by alignment to the draft genome for this species (provided by John McCutcheon, University of Montana), to assist in selecting the longest ORF in
each transcript cluster (the quality of the genome data was inadequate to support independent genome-based gene identification). All resulting sequences were further analysed by the NCBI conserved domain search tools to confirm the functional domains for aquaporin and sucrase, (Marchler-Bauer et al., 2011). Candidate signal peptides were identified by SIGNALP v. 4.1 (cbs.dtu.dk/services/SignalP).

Phylogenetic analysis

Gene phylogenies were estimated using the sequences identified above (Tables S1 and S2). Full-length protein sequences were aligned in CLUSTALX 2.1 (Larkin et al., 2007) with default parameters and the resulting alignments were trimmed in TRIMAL v. 1.2 (Capella-Gutiérrez et al. 2009) using a gap threshold of 0.25. PROTTest 2.4 (Abascal et al., 2005) was used to determine the best model of protein evolution by the Akaike information criterion. The best model was LG + I + G for the aquaporin sequences and WAG + G for the α-glucohydrolase sequences. Bayesian phylogenies were reconstructed in MrBayes v. 3.2.3b MPI (parallel) version (Ronquist et al., 2012) using two runs with four chains per run. Analyses were allowed to run until the standard deviation of split frequencies between runs dropped below 0.05. Convergence was confirmed in TRACER v. 1.6 (beast.bio.ed.ac.uk/Tracer), assuming a burn-in of 10% of generations. As convergence was supported, the first 10% of generations were discarded and phylogenies sampled in the remaining generations were used to estimate a 50% majority-rule consensus tree.

Acknowledgements

We thank Dr Cecilia Tamborinde guy (Texas A&M University) for providing the raw reads for the Ba. cockerell i tran scriptome, Dr John McCutcheon for proving access to the draft genome of Pl. citri and Dr Stewart Gray (USDA-ARS) for the M. persicae culture. This publication is an output of a research project funded by National Institute of Food and Agriculture (NIFA) grant NYW-2011-04650, and the Natural Resources Institute, University of Greenwich from a grant provided by the Bill & Melinda Gates foundation grant Agreement OPP1058938.

References

Abascal, F., Zardoya, R. and Posada, D. (2005) ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21: 2104–2105.
Agre, P. (2006) The aquaporin water channels. Proc Am Thorac Soc 3: 5–13.
Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114–2120.
Boykin, L.M., Bell, C.D., Evans, G., Small, I. and De Barro, P. (2013) Is agriculture driving the diversification of the Bemisia tabaci species complex (Hemiptera: Sternorrhyncha: Aleyrodidae)? Dating, diversification and biogeographic evidence revealed. BMC Evol Biol 13: 228.

Byrne, D.N. and Miller, W.B. (1990) Carbohydrate and amino acid composition of phloem sap and honeydew produced by Bemisia tabaci. J Insect Physiol 36: 433–439.
Campbell, E.M., Ball, A., Hopper, S. and Bowman, A.S. (2008) Invertebrate aquaporins: a review. J Comp Physiol B 1: 935–955.
Capella-Gutiérrez, S., Silla-Martínez, J.M. and Gabaldón, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25: 1972–1973.
Cohen, E. (2013) Water homeostasis and osmoregulation as targets in the control of insect pests. In Target Receptors in the Control of Insect Pests (Cohen, E., ed.), pp. 1–61. Academic Press Ltd and Elsevier Science Ltd, London.
Dolling, W.R. (1991) The Hemiptera. Oxford University Press, Oxford, UK.
Douglas, A.E. (2003) The nutritional ecology of aphids. Adv Insect Physiol 31: 73–140.
Duncan, R.P., Husnik, F., Van Leuven, J.T., Gilbert, D.G., Dávalos, L.M., McCutcheon, J.P. et al. (2014) Dynamic recruitment of amino acid transporters to the insect/symbiont interface. Mol Ecol 23: 1606–1623.
Eddy, S.R. (2011) Accelerated profile HMM searches. PLoS Comput Biol 7: e1002195.
Fabrick, J.A., Pei, J., Hull, J.J. and Yool, A.J. (2014) Molecular and functional characterization of multiple aquaporin water channel proteins from the western tarnished plant bug, Lygus hesperus. Insect Biochem Mol Biol 45: 125–140.
Finn, R.N., Chauvigne, F., Stavang, J.A., Belles, X. and Cerda, J. (2015) Insect glycerol transporters evolved by functional co-option and gene replacement. Nat Comm 6: 7514.
Fisher, D.B., Wright, J.P. and Mittler, T.E. (1984) Osmoregulation by the aphid Myzus persicae: a physiological role for honeydew oligosaccharides. J Insect Physiol 30: 387–393.
Goodchild, A.J.P. (1966) Evolution of the alimentary canal in the Hemiptera. Biol Rev 41: 97–140.
Goto, S.G., Philip, B.N., Teets, N.M., Kawarasaki, Y., Lee, R.E and 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. et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnol 29: 644–652.
Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J. et al. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protocols 8: 1494–1512.
Hendrix, D.L., Wei, Y. and Leggett, J.E. (1992) Homopteran honeydew sugar composition is determined by both the insect and plant species. Comp Biochem Physiol 101B: 23–27.
Heymann, J.B. and Engel, A. (2000) Structural clues in the sequences of the aquaporins. J Mol Biol 295: 1039–1053.
Huang, J.-H., Jing, X. and Douglas, A.E. (2015) The multi-tasking gut epithelium of insects. Insect Biochem Mol Biol 67: 15–20.
Husnik, F., Nikoh, N., Koga, R., Ross, L., Duncan, R.P., Fujie, M. et al. (2013) Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. Cell 153: 1567–1578.
Ibanez, F., Hancock, J. and Tamborinde guy, C. (2014) Identification and expression analysis of aquaporins in the potato psyllid, Bactericera cockerell i. PLoS ONE 9: e111745.
Jing, X., White, T.A. and Douglas, A.E. (2015) The molecular correlates of organ loss: the case of insect Malpighian tubules. *Biol Lett* 11: 20150154.

Karley, A.J., Ashford, D.A., Minto, L.M., Pritchard, J. and Douglas, A.E. (2005) The significance of gut sucrase activity for osmoregulation in the pea aphid, *Acyrthosiphon pismum*. *J Insect Physiol* 51: 1313–1319.

Kaufmann, N., Mathai, J.C., Hill, W.G., Dow, J.A.T., Zeidel, M.L. and Brodsky, J.L. (2005) Developmental expression and biochemical characterization of a *Drosophila melanogaster* aquaporin. *Am J Physiol* 289: C397–C407.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14: R36.

Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10: R25.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H. et al (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.

Le Cacherec, F., Deschamps, S., Delamarche, C., Pellerin, I., Bonnet, G., Guilam, M.T. et al (1996) Molecular cloning and characterization of an insect aquaporin functional comparison with aquaporin 1. *Eur J Biochem* 41: 707–715.

Lee, K.S., Kim, S.R., Lee, S.M., Lee, K.R., Sohn, H.D. and Jin, B.R. (2001) Molecular cloning and expression of a cDNA encoding the aquaporin homologue from the firefly, *Pyrococella rufa*. *Korean J Entomol* 31: 269–279.

Luan, J.B., Chen, W.B., Hasegawa, D.K., Simmons, A.M., Wintermantel, W.M., Ling, K.S. et al (2015) Metabolic coevolution in the bacterial symbiosis of whiteflies and related plant sap-feeding insects. *Genome Biol Evol* 7: 2635–2647.

Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C. et al (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* 39: D225–D229.

Mathew, L.G., Campbell, E.M., Yool, A.J. and 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.

Min, X.J., Butler, G., Storms, R. and Tsang, A. (2005) OrfPredicator: predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Res* 33: W677–W680.

Nachappa, P., Levy, J. and Tambarindey, C. (2012) Transcriptome analyses of *Bactericera cockerelli* adults in response to “Candidatus Liberibacter solanacearum” infection. *Mol Genet Genomics* 287: 803–817.

Price, D.R.G., Karley, A.J., Ashford, D.A., Isaacs, H.V., Pownall, M.E., Wilkinson, H.S. et al (2007) Molecular characterisation of a candidate gut sucrase in the pea aphid, *Acyrthosiphon pismum*. *Insect Biochem Mol Biol* 37: 307–317.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S. et al (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61: 539–542.

Salvucci, M.E. (2003) Distinct sucrase isomerases catalyze trehalulose synthesis in whiteflies, *Bemisia argentifolii*, and *Erwinia rhapsontici*. *Comp Biochem Physiol B 135*: 385–395.

Salvucci, M.E., Wolfe, G.R. and Hendrix, D.L. (1997) Effect of sucrase concentration on carbohydrate metabolism in *Bemisia argentifolii*: biochemical mechanism and physiological role for trehalulose synthesis in the silverleaf whitefly. *J Insect Physiol* 43: 457–464.

Shakesby, A.J., Wallace, I.S., Isaacs, H.V., Pritchard, J., Roberts, D.M. and Douglas, A.E. (2009) A water-specific aquaporin involved in aphid osmoregulation. *Insect Biochem Mol Biol* 39: 1–10.

Sim, L., Willemsma, C., Mohan, S., Nair, H.Y., Pinto, B.M. and Rose, D.R. (2012) Structural basis for substrate selectivity in human maltase-glucosamylase and sucrase-isomaltase N-terminal domains. *J Biol Chem* 285: 17763–17770.

Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A. and Conesa, A. (2011) Differential expression in RNA-seq: a matter of depth. *Genome Res* 21: 2213–2223.

Terra, W.R., Ferreira, C., Jordan, B.P., and Dillon, R.J. (1996) Digestive enzymes. In *The Biology of the Insect Midgut* (Lehane, M.J. and Billingsly, P.F., eds), pp. 153–194. Chapman and Hall, London.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R. et al (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 7: 562–578.

Tzin, V., Yang, X., Jing, X., Zhang, K., Jander, G. and Douglas, A.E. (2015) RNA interference against gut osmoregulatory genes in phloem-feeding insects. *J Insect Physiol* 79: 105–112.

Wallace, I.S., Shakesby, A.J., Hwang, J.H., Choi, W.G., Martinakova, N., Douglas, A.E. et al (2012) *Acyrthosiphon pismum AQP2*: a multifunctional insect aquaglyceroporin. *Biochim Biophys Acta – Biomembranes* 1818: 627–635.

Ye, X.D., Su, Y.L., Zhao, Q.Y., Xia, W.Q., Liu, S.S. and Wang, X.W. (2014) Transcriptomic analyses reveal the adaptive features and biological differences of guts from two invasive whitefly species. *BMC Genomics* 15: 370.

Zheng, Y., Zhao, L.J., Gao, J.P. and Fei, Z.J. (2011) Assembler: a package for de novo assembly of Roche-454/Sanger transcriptome sequences. *BMC Bioinformatics* 12: 453.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

**Figure S1.** Structure of aquaporin (AQP) genes. Transmembrane domains (TMs, ‘Curved rectangle’) were predicted using TMHMM and TM_PRED [www.ch.embnet.org/software/TMpred_form; Hofmann and Stoffel (1993) TMBase, a database of membrane spanning protein segments, *Biol Chem* 347: 166. Rost B, Sander C, Houthof K (1993) TMbase, a database of membrane spanning protein segments. *Biochem J* 295: 713–724. The position of aspartagine-proline-alanine (NPA) boxes (‘star’) is highlighted. 1, 2 These Drosophila integral protein (DIRP) genes have been verified empirically as water channels.

**Table S1.** Aquaporins identified in each species and representative sequences used in the phylogenetic analysis.

**Table S2.** β-glucosidases identified in each species and representative sequences used in the phylogenetic analysis. The transcriptome sequences were BLASTed against the National Center for Biotechnology Information (NCBI) online database. NCBI accession numbers are provided for cases in which the transcriptome and NCBI sequences are identical.