Journal of Insect Science OPEN ACCESS

RESEARCH

Evolutionary Diversification of Insect Innexins

Austin L. Hughes[1]

Department of Biological Sciences, University of South Carolina, Columbia, SC 29208, USA
[1]Corresponding author, e-mail: austin@biol.sc.edu

Subject Editor: Yoonseong Park

J. Insect Sci. 14(221): 2014; DOI: 10.1093/jisesa/ieu083

ABSTRACT. Phylogenetic analysis of insect innexins supported the hypothesis that six major clades of insect innexins arose by gene duplication prior to the origin of the endopterygote insects. Within one of the six clades (the Zpg Clade), two independent gene duplication events were inferred to have occurred in the lineage of *Drosophila*, after the most recent common ancestor of the dipteran families Culicidae and Drosophilidae. The relationships among this clades were poorly resolved, except for a sister relationship between ShakB and Ogre. Gene expression data from FlyAtlas supported the hypothesis that the latter gene duplication events gave rise to functional differentiation, with Zpg showing a high level of expression in ovary, and Inx5 and Inx6 showing a high level of expression in testis. Because unduplicated members of this clade in *Bombyx mori* and *Anopheles gambiae* showed high levels of expression in both ovary and tests, the expression patterns of the *Drosophila* members of this clade provide evidence of subdivision of an ancestral gene function after gene duplication.

Key Words: gap-junction channel, gene duplication, innexin, multigene family

Intercellular communication is a fundamental need of multicellular organisms (Panchin 2005). In animals, an important pathway of cell–cell communications is provided by gap junctions, constituted by the junction of two hexameric protein arrays or hemichannels (Bauer et al. 2005, Phelan 2005). Each of the two communicating cells produces one hemichannel, and their alignment creates a channel allowing the passage of ions and small molecules (Phelan and Starich 2001). In invertebrates, the hexamers are composed of proteins belonging to a family known as innexins, which have apparent distant homologs (pannexins) in vertebrates but are distinct from the widely studied vertebrate connexin family of gap-junction proteins (Yen and Saier 2007).

Genomic studies have revealed multiple innexin family genes in insect species. In the best-studied insect model, *Drosophila melanogaster*, there are seven genes (Bauer et al. 2005): 1) *ogre* ('optic ganglion reduced'), also known as *inx1*, 2) *inx2*, 3) *inx3*, 4) *zpg* ('zero population growth'), also known as *inx4*, 5) *inx5*, 6) *inx6*, and 7) *inx7*, and *shakB* ('shaker B'). Each of these genes encodes a protein with four transmembrane domains (Bauer et al. 2005). Certain hemichannels may be heteromeric, while others are heteromeric. When two homomeric semichannels of different types form a channel together, that channel is known as heterotypic, whereas a channel formed by two homomeric semichannels of the same type is called homotypic (Phelan and Starich 2001). In *Drosophila*, the Ogre and Inx3 proteins both form heteromeric channels with Inx2, while Zpg can form heterotypic channels with Inx2 (Phelan 2005). By contrast, ShakB forms homotypic channels, as can Inx2 (Phelan 2005). Finally, there is evidence that at least in some animal species, innexins or their homologs can form non-junctional channels ('hemichannels' or innexons) of poorly understood function (Bao et al. 2007, Scemes et al. 2009).

A number of researchers have presented phylogenetic analyses of selected innexins (Phelan and Starich 2001, Phelan 2005, Hong et al. 2009). However, no study has attempted to use phylogenetic methods to estimate the time of gene duplications within the innexin family relative to major events of cladogenesis within the insects. Here, I take advantage of the information available from insect genome projects to reconstruct the evolutionary history of insect innexins. In addition, making use of data from gene expression atlases of model species, I examine patterns of functional differentiation of duplicate innexin genes.

Methods

Phylogenetic Analyses. Phylogenetic analyses were based on the 79 selected innexin protein sequences from 14 insect species representing two orders of exopterygotes (insects with incomplete metamorphosis) from the infraclass Paraneoptera; and four orders of the infraclass Endopterygota (endopterygotes or insects with complete metamorphosis) were downloaded from the NCBI website (Table 1; Fig. 1). From the genus *Drosophila*, *D. melanogaster* was chosen along with *Drosophila grimshawi*, which represents the clade of *Drosophila* species with sequenced genomes that are phylogenetically most distant from *D. melanogaster* (Drosophila 12 Genomes Consortium 2007). Sequences were aligned by the CLUSTAL algorithm in MEGA 5.05 (Tamura et al. 2011); and any site at which the alignment postulated a gap in any of a set of aligned sequences was excluded from analyses involving that set of sequences. Phylogenetic trees rooted with four sequences from the nematode *Ascaris sum* (Fig. 1). Phylogenetic trees were reconstructed by two methods: 1) maximum likelihood (ML), based on the JTT+G+I+F model and 2) minimum evolution (ME) based on the JTT+G distance.

The model for the ML analysis was chosen in MEGA 5.05 using the Bayes Information Criterion (Tamura et al. 2011). The gamma parameter (measuring rate variation among sites) used in the ME analysis (1.9926) was estimated by the ML analysis. The reliability of branching patterns in ML trees was tested by bootstrapping; 1,000 bootstrap pseudosamples were used. Significance of internal branches in the ME tree was tested by the interior branch test, with the standard error of branch lengths estimated by bootstrapping (Nei and Kumar 2000). The ML method was used to reconstruct ancestral sequences (most probable ancestors) at major nodes within the phylogenetic tree in MEGA 5.05.

Gene Expression Data. Gene expression data were downloaded from the Gene Expression Omnibus (GEO) platform at the National Institute for Biotechnology Information (http://www.ncbi.nlm.nih.gov/geo/). For *D. melanogaster*, gene expression data across nine tissue types (larval fat body and Malpighian tubules; and adult hindgut, midgut, accessory gland, brain, crop, ovary, and tests) were derived from the FlyAtlas database (GEO accession GSE7763; Chintapalli et al. 2007). Each tissue was represented by four biological samples,

© The Author 2014. Published by Oxford University Press on behalf of the Entomological Society of America.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Table 1. Insect species and major innexin groups[1] from which sequences are analyzed

Infraclass	Order	Family	Species	Innexin group (number of sequences)						
				Inx2	Inx3	Inx7	Ogre	ShakB	Zpg/Inx5/Inx6	Other
Paraneoptera	Hemiptera	Aphididae	*Acyrthosiphon pisum*	1	1	2	1	1	–	2
	Phthiraptera	Pediculidae	*Pediculus humanus*	1	1	1	1	1	1	–
Endopterygota	Coleoptera	Tenebrionidae	*Tribolium castaneum*	1	3	1	1	1	1	–
	Lepidoptera	Bombycidae	*Bombyx mori*	1	–	1	–	–	1	–
	Diptera	Culicidae	*Aedes aegypti*	1	1	1	1	1	–	–
			Anopheles gambiae	1	1	1	1	2	1	–
			Culex quinquefasciatus	1	–	1	–	1	–	–
		Drosophilidae	*Drosophila melanogaster*	1	1	1	1	1	3	–
			Drosophila grimshawi	1	1	1	1	1	3	–
	Hymenoptera	Pteromalidae	*Nasonia vitripennis*	–	1	1	–	1	–	–
		Apidae	*Apis mellifera*	1	1	1	1	1	–	–
			Bombus impatiens	1	1	1	1	1	–	–
			Camponotus floridanus	1	1	1	1	1	–	–
			Harpegnathos saltator	1	1	1	1	–	1	–

[1]*Groups are named as in D. melanogaster.*

and expression scores were normalized by global scaling. Expression of selected innexin family members in *Anopheles gambiae* and *Bombyx mori* were obtained from, respectively, GEO datasets GSE21689 and GSE1757. The *A. gambiae* data provided expression data for adult Malpighian tubules, midgut, head, salivary gland, testis, and ovary (four to eight replicates per tissue). The *B. mori* data provided expression data for eight tissues (fat body, Malpighian tubules, midgut, head, integument, hemocyte, anterior/median silk gland, and posterior silk gland) from fifth instar larvae and from adult testis and ovary (4–12 replicates per tissue).

Results

Phylogenetic Analysis. Figure 1 shows the ME tree of insect innexins, rooted with nematode sequences. There were six major clusters in the phylogenetic tree, each supported by a significant internal branch (Fig. 1). Five of these clusters are here designated according to the name of the *D. melanogaster* gene belonging to each cluster: 1) Inx2, 2) Ogre, 3) ShakB, 4) Inx7, and 5) Inx3 (Fig. 1). The sixth cluster included Drosophila Zpg, Inx5, and Inx6 (Fig. 1). In the following, the latter cluster is designated as the 'Zpg Clade.' The relationships between the six major clusters of insect innexins were not well resolved except for the fact that there was a strong support for a sister relationship between the Ogre and ShakB clades (Fig. 1). This pattern was supported by a significant internal branch in the ME tree (Fig. 1). The ML tree showed a similar overall topology to that seen in the ME tree, with the same six clusters (Supp Fig. S1), although the bootstrap values in the ME tree were lower than the confidence levels of the interior branch tests in the ME tree.

Each of the six major clusters included sequences from three or four exopterygote orders as well as either of both of the endopterygote species *Acyrthosiphon pisum* and *Pediculus humanus* (Fig. 1). Thus, the phylogenetic analysis supported the hypothesis that each of the six clusters arose by a gene duplication even that occurred prior to the most recent common ancestor (MRCA) of exopterygotes and endopterygotes. However, the phylogenetic tree supported the hypothesis that the gene duplication events giving rise to *Drosophila* Zpg, Inx5, and Inx6 occurred more recently. Each of these proteins from *D. melanogaster* clustered with an apparent ortholog from *D. grimshawi*, indicating that the gene duplication events giving rise to Zpg, Inx5, and Inx6 occurred prior to the MRCA of these two *Drosophila* species (Fig. 1).

Sequences from mosquitoes (Culicidae) fell outside the cluster of Drosophila Zpg, Inx5, and Inx6; and this topology was supported by a significant internal branch (Fig. 1). The latter topology supports the hypothesis that the gene duplication events giving rise to Zpg, Inx5, and Inx6 occurred after the MRCA of Culicidae and Drosophilidae. A zsignificant internal branch supported a sister relationship between Inx5 and Inx6, with Zpg falling outside (Fig. 1). This topology supported the hypothesis that an initial gene duplication event separated the *zpg* gene from the gene ancestral to the *inx5* and *inx6* genes, followed by the duplication that gave rise to separate *inx5* and *inx6* genes.

Conserved Amino Acid Residues. Reconstructed ancestral residues were used to estimate the percentage of residues in the ancestor of each clade that were conserved in all members of the clade included in the analysis (Fig. 2A). The highest percentage of conserved residues was seen in ShakB (57.7%; Fig. 2A). The percentage conserved was significantly lower in each of the other clades (Fig. 2A). Considering only the residues that resulted from new replacements (apomorphies) in the ancestor of each clade, again ShakB showed the highest proportion conserved (30.8%; Fig. 2B). Thus, the ShakB clade was characterized by unique conserved residues to a greater extent than the other five clades.

Gene Expression Patterns. One-way analysis of variance (ANOVA) was used to test for difference in mean expression score across nine tissues of *D. melanogaster* innexins (Figs. 3 and 4 and Supp Fig S2). In all cases except ShakB, there were significant differences in expression across tissues ($P < 0.01$; Bonferroni-corrected). Inx5 and Inx6 showed a pattern of generally low expression scores in all tissues except testis (Fig. 3). By contrast, Zpg showed a pattern of low expression in all tissues except ovary (Fig. 4A). Inx2 showed a more complicated expression pattern than Zpg, but Inx2 also showed high expression in ovary but low expression in testis (Fig. 4B).

Similar analyses were applied to gene expression data for the members of the Zpg Clade from *A. gambiae* (Fig. 5A) and *B. mori* (Fig. 5B). In both cases, there was a significant difference among tissues (one-way ANOVA; $P < 0.001$). In both cases, the expression levels in ovary and testis were substantially greater than those of other tissues, although the mean expression score for ovary was highest in *A. gambiae* and that for testis was highest in *B. mori* (Fig. 5).

Discussion

Phylogenetic analysis of insect innexins strongly supported the hypothesis that six major clades of insect innexins arose by gene duplication prior to the origin of the endopterygote insects. The two exopterygote species used in these analyses belong to Paraneoptera, whose MRCA with the endopterygotes occurred in the Carboniferous, over 300 million years ago (Mya) (Grimaldi and Engel 2005). Because members of all six clades were found in Paraneoptera as well as in endopterygote orders, it can be concluded that the six major clades of

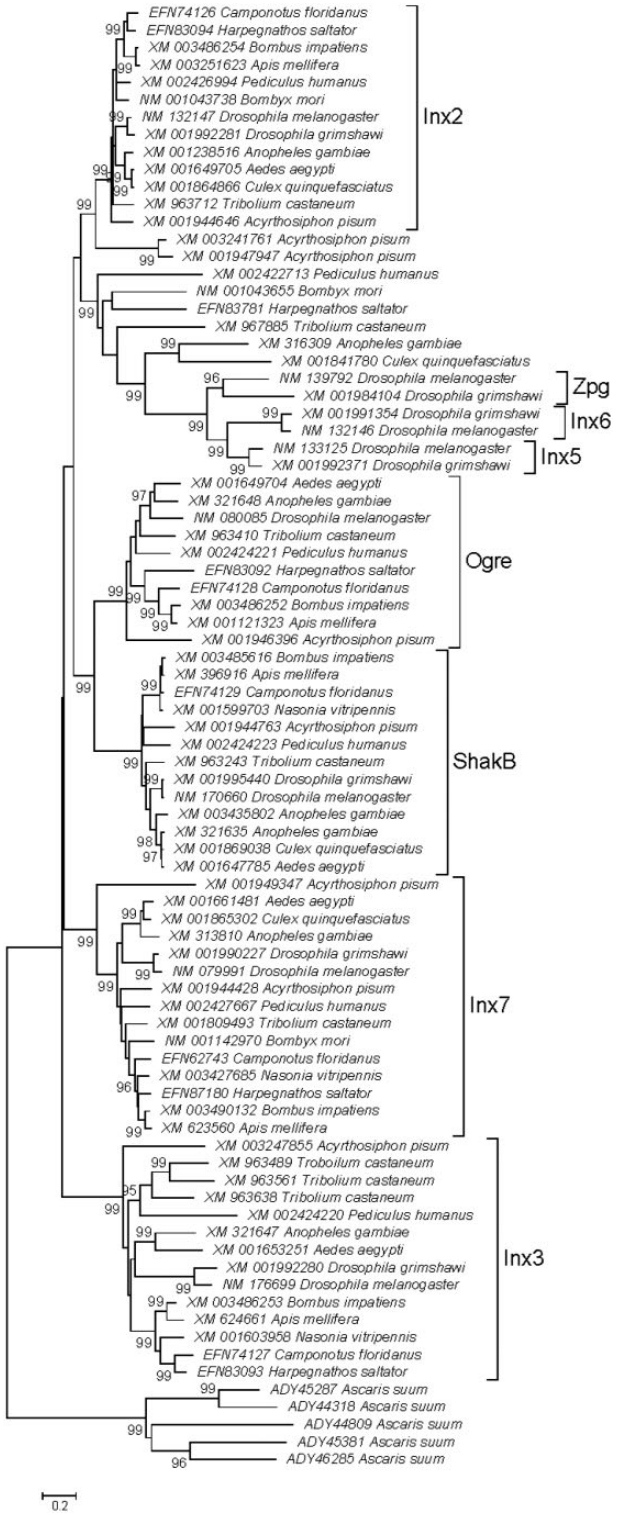

Fig. 1. ME tree of insect innexins, rooted with sequences from the nematode *A. sum*. The tree was based on the JTT+G distance at 267 aligned amino acid positions. Numbers on the branches are the confidence levels of the interior branch test; only values ≥95% are shown.

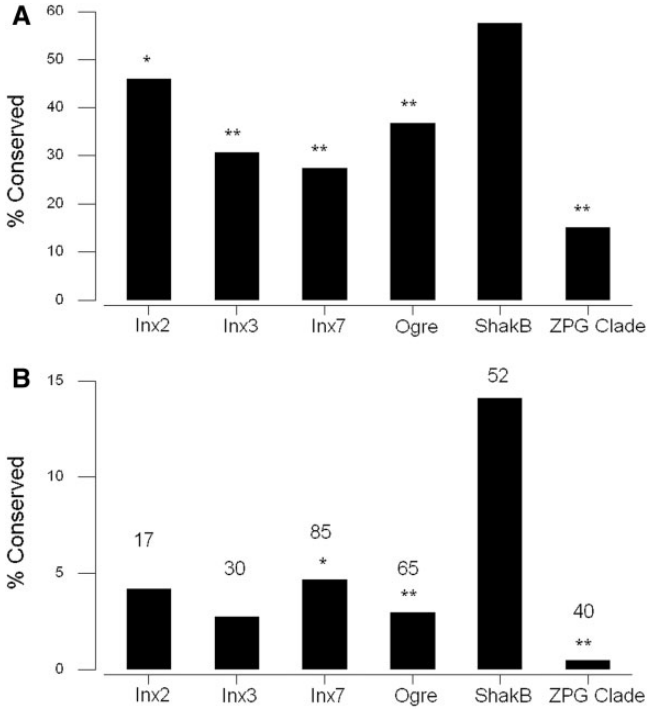

Fig. 2. (A) The percentage of residues inferred to have been present in the common ancestor of each of six clades of insect innexins that were conserved in all clade members analyzed (out of 267 total aligned residues in each case). Fisher's exact test of the equality of the proportion conserved with the ShakB clade: $*P < 0.05$; $**P < 0.01$ (Bonferroni-corrected). (B) The percentage of apomorphic residues inferred to have arisen in the common ancestor of each clade that were conserved in all clade members analyzed (total numbers of apomorphic residues in the ancestor are shown above each bar). Fisher's exact test of the equality of the proportion conserved with the ShakB clade: $*P < 0.05$; $**P < 0.01$ (Bonferroni-corrected).

insect innexins have been separated for at least 300 Mya. The relationships among the sic clades were not well resolved, except for evidence of a sister relationship between ShakB and Ogre.

Within one of the six clades (the Zpg Clade), two independent gene duplication events were inferred to have occurred in the lineage of *Drosophila*, after the MRCA of dipteran families Culicidae and Drosophilidae, which probably occurred in the early Jurassic around 250 Mya (Grimaldi and Engel 2005). These duplications gave rise to *Drosophila zpg*, *inx5*, and *inx6*, which lack orthologs in Culicidae or other available insect genomes. The presence of *zpg*, *inx5*, and *inx6* orthologs in both *D. melanogaster* and *D. grimshawi* indicates that these duplications occurred prior to the MRCA of these two species, which has been estimated to have occurred over 60 Mya (Tamura et al. 2004). More information about the timing of these duplications will be provided by more fully sequenced genomes of Diptera.

Gene expression data from the FlyAtlas database (Chintapalli et al. 2007) indicated marked differences in expression among tissues in the case of every *D. melanogaster* innexin except ShakB. The ShakB protein also differed from other insect innexins in its high level of amino acid sequence conservation, with much higher proportions of ancestral residues and of unique ancestral residues being conserved in the ShakB clade than in other clades. A high level of sequence conservation in the case of a very broadly expressed protein is consistent with data suggesting that broadly expressed proteins tend to be highly conserved (Zhang and Li 2004). In spite of the sister relationship between ShakB and Ogre, there was no evidence of a similar pattern of gene expression in *Drosophila* (Supp Fig S2).

The Zpg Clade of Drosophila provided a striking contrast in patterns of gene expression, with Zpg expressed at a high level in ovary and Inx6 and Inx7 expressed at a high level in testis. Because the unduplicated members of this clade from *B. mori* (Lepidoptera) and *A. gambiae* (Diptera: Culicidae) were expressed at high levels in both ovary and

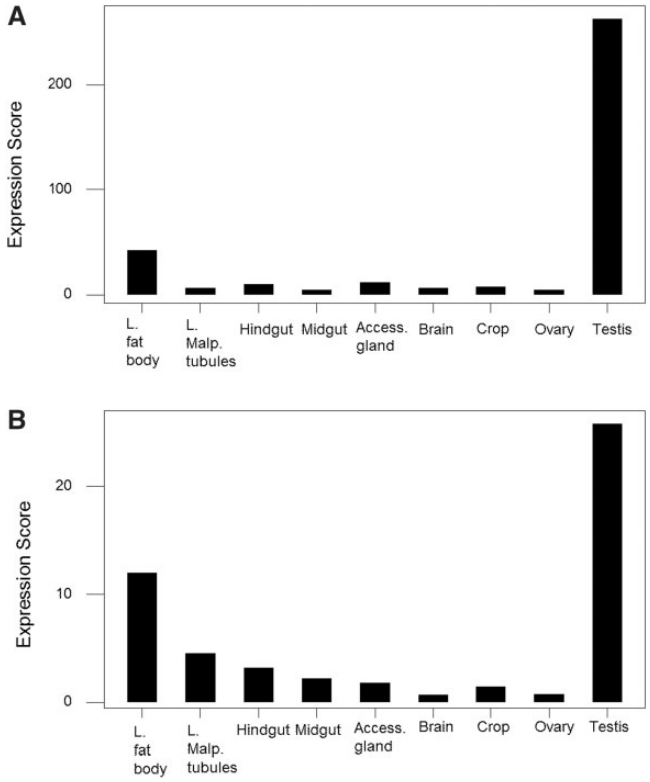

Fig. 3. Mean gene expression scores in nine tissues of (A) ShakB and (B) Inx2. There was no significant difference among tissues in the case of ShakB (one-way ANOVA; $P > 0.05$, Bonferroni-corrected), but there was a significant difference in the case of Inx2 (one-way ANOVA; $P < 0.01$, Bonferroni-corrected).

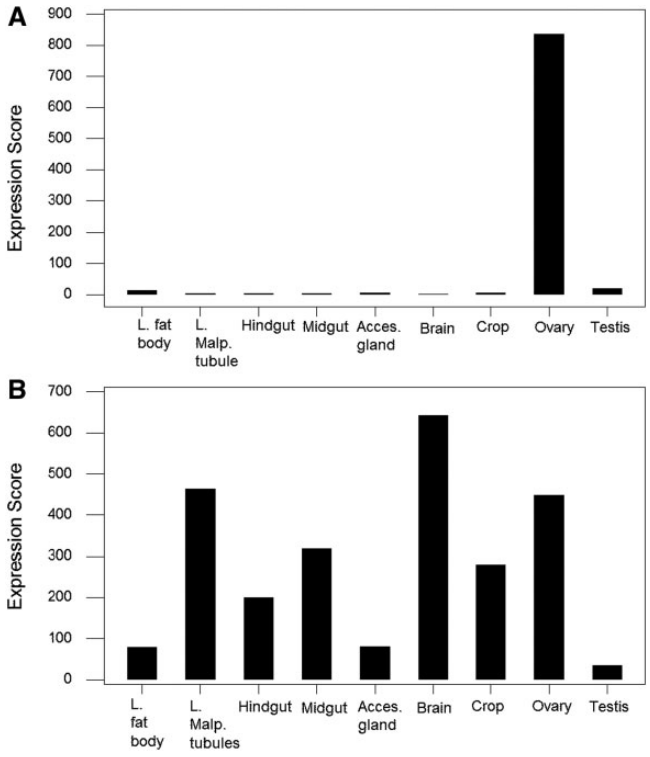

Fig. 4. Mean gene expression scores in nine tissues of (A) Inx5 and (B) Inx6. There was a significant difference among tissues in each case (one-way ANOVA; $P < 0.01$ in each case, Bonferroni-corrected).

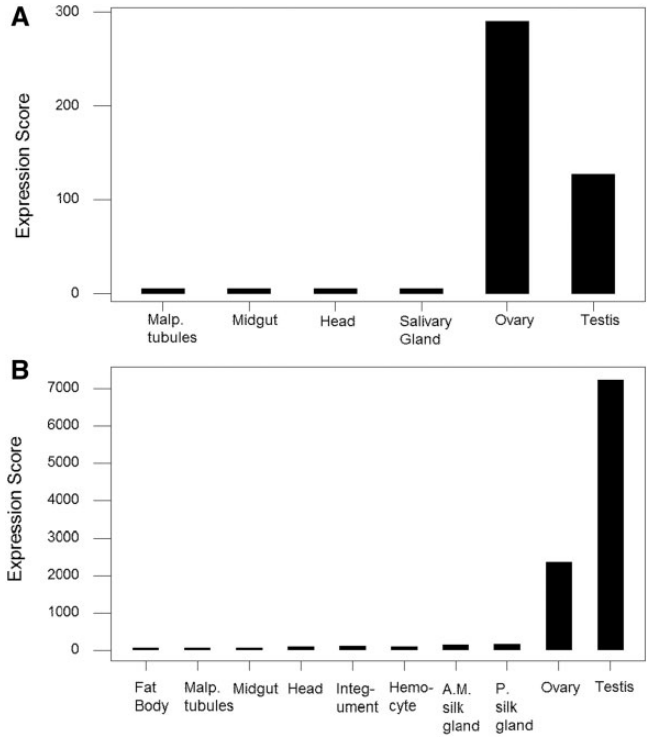

Fig. 5. Mean gene expression scores of Zpg Clade members from (A) *A. gambiae* (six tissues) and (B) *B. mori* (10 tissues). There was a significant difference among tissues in each case (one-way ANOVA; $P < 0.01$).

testis, it seems likely that this broader pattern of expression in gonads of both sexes was the ancestral pattern prior to gene duplication in the *Drosophila* lineage. Thus, the patterns of expression of the *Drosophila* members of this clade provide evidence of subdivision of an ancestral gene function after gene duplication (Hughes 1994, Lynch and Force 2000).

It is known that Drosophila Zpg can form heterotypic channels with Inx2 (Phelan 2005), and it is of interest that, in spite of a complex pattern of tissue expression, Inx2 shared with Zpg a high level of expression in ovary (Fig. 2B). Because no other *Drosophila* innexin showed the same high level of expression in the testis as Inx5 and Inx6, it is possible that these two proteins together form heterotypic channels in the testis. On the other hand, it is known that Inx6 forms heterotypic channels with Inx7 in the Drosophila brain and that these channels are important for memory (Wu et al. 2011).

A more detailed knowledge of expression patterns and the formation of heteromeric and heterotypic channels in different tissues will shed further light on the functional differentiation of the insect innexins.

References Cited

Bao, L., S. Samuels, S. Locovei., E. R. Macagno, K. J. Muller, and G. Dahl. 2007. Innexins form two types of channels. FEBS Lett. 581: 5703–5708.

Bauer, R., B. Löer, K. Ostrowski, J. Martini, A. Weimbs, H. Lechner, and M. Hoch. 2005. Intercellular communication: the *Drosophila* innexin multiprotein family of gap junction proteins. Chem. Biol. 12: 515–526.

Chintapalli, V. R., J. Wang, J. A. Dow. 2007. Using FlyAtlas to identify better Drosophila models of human disease. Nat. Genet. 39: 715–720.

Drosophila 12 Genomes Consortium. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. Nature 450: 203–218.

Grimaldi, D., and M. S. Engel. 2005. Evolution of the insects. Cambridge University Press, Cambridge, United Kingdom.

Hong, S. M., S. K. Noh, K. A. Kim, H. Mitsunobu, H. Mon, J. M. Lee, Y. Kawaguchi, and T. Kusakabe. 2009. Molecular characterization, localization, and distribution of innexins in the silkworm, *Bombyx mori*. Mol. Biotechnol. 43: 52–58.

Hughes, A. L. 1994. The evolution of functionally novel proteins after gene duplication. Proc. R. Soc. Lond. B 256: 119–124.

Lynch, M., and A. Force. 2000. The probability of duplicate-gene preservation by subfunctionalization. Genetics 154: 459–473.

Nei, M. and S. Kumar. 2000. Molecular evolution and phylogenetics. Oxford University Press, New York.

Panchin, Y. V. 2005. Evolution of gap junction proteins: the pannexin alternative. J. Exp. Biol. 208: 1415–1419.

Phelan, P. 2005. Innexins: members of an evolutionarily conserved family of gap-junction proteins. Biochim. Biophys. Acta 1711: 225–245.

Phelan, P., and T. A. Starich. 2001. Innexins get into the gap. BioEssays 23: 388–396.

Scemes, E., D. C. Spray, and P. Meda. 2009. Connexin, pannexins, innexins: novel roles of 'hemi-channels.' Pflugers Archiv – Eur. J. Physiol. 457: 1207–1226.

Tamura, K., S. Subramanian, and S. Kumar. 2004. Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. Mol. Biol. Evol. 21: 36–44.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28: 2731–2739.

Wu, C.-L., M.-F. Shih, J. S. Lai, H.-T. Yang, G. C. Turner, L. Chen, and A.-S. Chiang. 2011. Heterotypic gap junctions between two neurons in the *Drosophila* brain are critical for memory. Curr. Biol. 21: 848–854.

Yen, M. R., and M. H. Saier. 2007. Gap junction proteins of animals: the innexin/pannexin superfamily. Prog. Biophys. Mol. Biol. 94: 5–14.

Zhang, L., and W.-H. Li. 2004. Mammalian housekeeping genes evolve more slowly than tissue-specific genes. Mol. Biol. Evol. 21: 236–239.

Received 30 April 2013; accepted 11 July 2014.