

Supplementary Note 2: Galactoside-Binding Lectins

Authors: Krystal Villalobos

Introduction

Galactoside-binding lectins (galectins) bind β -galactose with their structurally similar carbohydrate-recognition domains (CRDs) (Cummings and Liu 2009). CRDs can function alone or in clusters creating a β -sandwich structure without Ca^{2+} binding sites (Leffler et al., 2004; Mitchell et al., 2001; Wang et al., 2012).

In the invertebrate *Clavelina picta*, it has been shown that galectin may be an effector molecule in the innate immune system (Vasta et al., 1999), whereas it can be similar in insects. *Anopheles* mosquito galectins were upregulated in the gut and salivary glands after infection of both malaria and bacteria (Dimopoulos et al., 1996, 1998, 1988). It was suggested that galectins may function as a pattern recognition receptor (PRR) by binding saccharides (β -galactose) on the pathogen surface to trigger the innate immune response or agglutinate bacteria in the midgut following blood-feeding (Dimopoulos et al., 1998).

Methods

Insecta galectins orthologs were collected from Immunodb and were used to blat against the *Diaphorina citri*, Asian citrus psyllid (ACP), genome. Identification and annotation of galectins was performed in WebApollo where RNA sequence data (adult, nymph, and egg) supported the manually curated gene models. BLASTp was used to confirm accuracy of the annotated galectins and determine the number of CRDs each galectin contains.

Galectin orthologs were collected from a limited NCBI search to include different insect orders. The conserved domains database in NCBI was used to identify and retrieve the CRD amino acid sequences of each collected galectin orthologs. ACP CRDs were globally pairwise aligned with *Acyrtosiphon pisum*, *Cimex lectularius*, *Diuraphis noxia*, and *Drosophila melanogaster* using NCBI Needleman-Wunsch tool. Multiple sequence alignments of ACP and orthologous insect galectins CRDs was performed in MEGA6 using ClustalW (Tamura et al 2003). The phylogenetic analysis of the CRDs included, constructing a midpoint rooted neighbor-joining tree in MEGA6 using p-distance to show evolutionary distance and 1,000 bootstrap replicates.

Results and Discussion

A total of three galectins and one partial galectin were identified and manually curated within the ACP genome (Table 1). *Acyrtosiphon pisum*, ACP closest relative, only has

two galectin paralogs (Gerardo et al., 2010) (Table 3), which are isoforms (NCBI accession: XP_008186241 and XP_001943769). The naming of the ACP galectins are based on NCBI Gnomon. Galectin-5 and 32 kDa beta-galactoside-binding lectin were identified in the same scaffold (WebApollo accession: gi|645507215|ref|NW_007377695.1|). Blastp showed that galectin-4, 32 kDa beta-galactoside-binding lectin, galectin-5, and galectin-4 partial has two CRDs, two CRDs, one CRD, and one CRD, respectively. ACP galectins containing two CRDs, like that of mammalian galectins, may mediate cell-cell communication or trigger signal cascades, as also suggested in *Drosophila melanogaster* (Pace et al., 2002; Liu, 2000; Hadari et al., 2000; Pace et al., 2000; Baum et al., 1995).

In the pairwise alignment analysis (Table 2), the ACP CRDs have an identity of 17% or greater compared to other related organisms. *D. citri* gal-4 CRD1, 32 kDa gal CRD2, and gal-5 CRD1 were analyzed with related insects CRDs from the orders Hymenoptera, and Coleoptera (Figure 1). The tree shows that the ACP CDRs have evolved from each other but still grouped with other hemipterans, further supporting the manually curated models. Coleoptera and Hymenoptera CRDs have also grouped with themselves independently.

Table 1: All the galectins found in the *D. citri* genome. Table shows all galectins with names based on NCBI Gnomon with BLASTp matches, and corresponding protein length, number of CRDs, and domains.

<i>D. citri</i> galectins	BLASTP match (organism, bit score)	Protein Length	Number of CRDs	Domains
Galectin 4	Galectin 4 (<i>Zootermopsis nevadensis</i> , 101)	286	2	Dimerization interface, sugar binding pocket, dimerization swap strand, Putative alternate dimerization interface, Gal-bind_lectin (CRD), Gal-bind_lectin (CRD), two Gal-bind_lectin, two GLECT, two GLECT superfamily
Galectin 4, partial	Galectin 9 (<i>Habropoda laboriosa</i> , 52.4)	142	1	Putative alternate dimerization interface, sugar binding pocket, Gal-bind_lectin (CRD), GLECT, Gal-bind_lectin, GLECT superfamily
Galectin 5	Galectin 8 (<i>Zootermopsis nevadensis</i> , 116)	218	1	Dimerization swap strand, sugar binding pocket, dimerization interface, putative alternate dimerization interface, Gal-bind_lectin (CRD), Gal-bind_lectin, GLECT, GLECT superfamily
		494	2	Two sugar binding pocket, dimerization interface, dimerization swap strand, two

32 kDa beta-galactoside-binding lectin	Galectin 8 (<i>Zootermopsis nevadensis</i> , 184)			putative alternate dimerization interface, two Gal-bind_lectin (CRD), two Gal-bind_lectin, two GLECT, two GLECT superfamily
---	--	--	--	---

Table 2: Global pairwise alignment of *D. citri* CRDs with related organisms displaying percent identity and (Needleman-Wunsch score).

	Galectin-4		32 kDa galectin		Galectin-5
	CRD1	CRD2	CRD1	CRD2	CRD1
<i>Diaphorina citri</i>					
<i>Acyrtosiphon pisum</i> (XP_008186241.1)	33% (146)	22% (-27)	28% (104)	20% (7)	28% (104)
<i>Cimex lectularius</i> (XP_014259365.1)	17% (12)	19% (-32)	31% (147)	18% (-29)	31% (147)
<i>Diuraphis noxia</i> (XP_015371421.1)	32% (139)	22% (-28)	27% (91)	19% (2)	27% (91)
<i>Drosophila melanogaster</i> (NP_001245813.1)	23% (52)	23% (-24)	31% (155)	20% (25)	31% (155)

Table 3: Number of galectins in *D. citri* and related organisms.

Organism	# of galectins
<i>Diaphorina citri</i>	3
<i>Acyrtosiphon pisum</i>	1
<i>Cimex lectularius</i>	1
<i>D. melanogaster</i>	1
<i>Aedes aegypti</i>	12
<i>Anopheles gambiae</i>	15
<i>Culex quinquefasciatus</i>	5

*Numbers based on Ensembl and NCBI, does not include isoforms/partial.

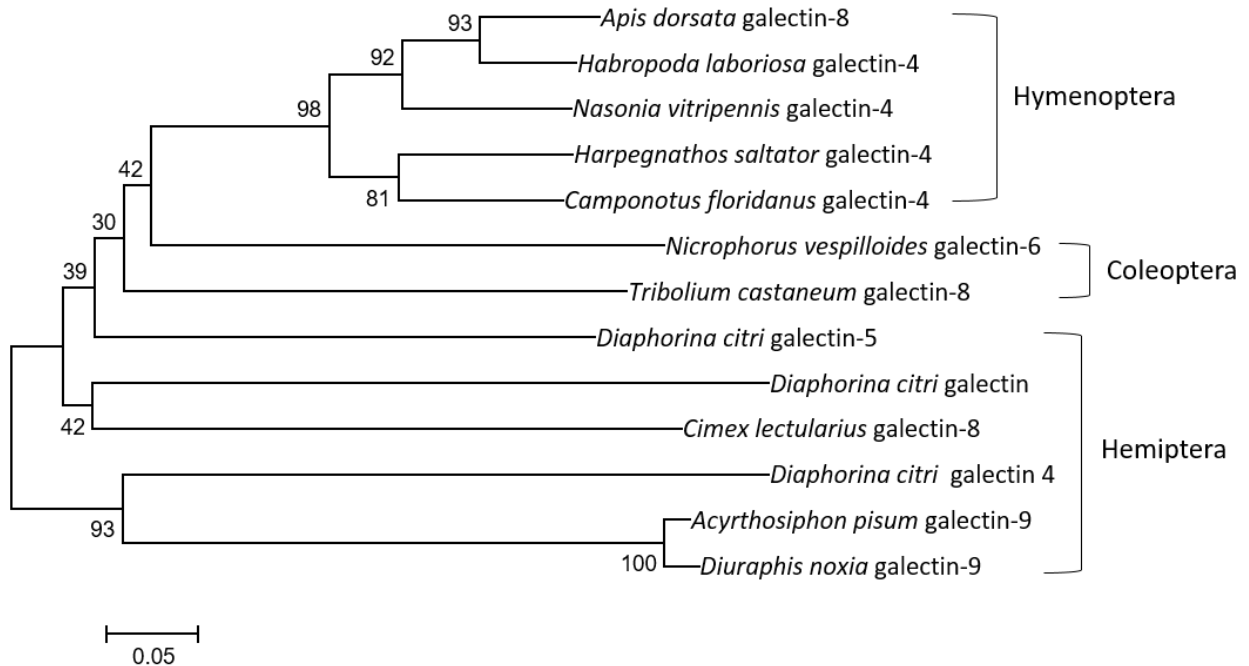


Figure 1: Neighbor-joining tree of 13 galectin CRDs was constructed in MEGA6 using 1000 bootstrap replications. The p-distance method was used to show evolutionary distance. A total of 60 amino acid positions were analyzed for the tree. NCBI accessions: *A. dorsata* [XP_006607868.1], *H. laboriosa* [XP_017794332.1], *N. vitripennis* [XP_008211704.1], *H. saltator* [EFN82261.1], *C. floridanus* [EFN73449.1], *N. vespilloides* [XP_017773770.1], *T. castaneum* [XP_015834566.1], *C. lectularius* [XP_014259365.1], *A. pisum* [XP_008186241.1], *D. noxia* [XP_015371421.1].

References

1. Cummings RD, Liu FT. 2009. Galectins. *Essentials of Glycobiology* 2nd edition.
2. Vasta GR, Quesenberry M, Ahmed H, O'Leary N. 1999. C-type lectins and galectins mediate innate and adaptive immune functions: their roles in the complement activation pathway. *Developmental & Comparative Immunology* [Internet]. 23(4-5):401-420. PMID:10426431
3. Dimopoulos G, Richman A, della Torre A, Kafatos FC, Louis C, Proc Natl Acad Sci USA 93, 13066–71 (1996).
4. Dimopoulos G, Seeley D, Wolf A, Kafatos FC, Embo J 17, 6115– 23. (1998).
5. Dimopoulos G, Muller HM, Levashina EA, Kafatos FC, Curr Opin Immunol 13, 79– 88 (1998).
6. Gerardo NM, et al. 2010 Feb. Immunity and other defenses in pea aphids, *Acyrtosiphon pisum*. *Genome Biology* [Internet]. [cited 2016 May 17];1-17. PMID: 20178569
7. Leffler, H., Carlsson, S., Hedlund, M., Qian, Y., Poirier, F., 2004. Introduction to galectins. *Glycoconj. J.* 19, 433e440.

8. Mitchell, D.A., Fadden, A.J., Drickamer, K., 2001. A novel mechanism of carbohydrate recognition by the C-type lectins DC-SIGN and DC-SIGNR. Subunit organization and binding to multivalent ligands. *J. Biol. Chem.* 276, 28939e28945
9. Wang, L., Wang, L., Yang, J., Zhang, H., Huang, M., Kong, P., Zhou, Z., Song, L., 2012. A multi-CRD C-type lectin with broad recognition spectrum and cellular adhesion from *Argopecten irradians*. *Dev. Comp. Immunol.* 36, 591e601.
10. Tamura K., Stecher G., Peterson D., FilipSKI A., and Kumar S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*30: 2725-2729.
11. Pace KE, Lebestky T, Hummel T, Arnoux P, Kwan K, Baum LG. 2002. Characterization of a Novel *Drosophila melanogaster* Galectin. *THE JOURNAL OF BIOLOGICAL CHEMISTRY.* 227(13):13091-13098
12. Liu, F.-T. (2000) *Clin. Immunol.* 97, 79–88
13. Hadari, Y. R., Arbel-Goren, R., Levy, Y., Amsterdam, A., Alon, R., Zakut, R., and Zick, Y. (2000) *J. Cell Sci.* 113, 2385–2397
14. Pace, K. E., Hahn, H. P., Pang, M., Nguyen, J. T., and Baum, L. G. (2000)*J. Immunol.* 165, 2331–2334
15. Baum, L. G., Pang, M., Perillo, N. L., Wu, T., Delegeane, A., Uittenbogaart, C. H., Fukuda, M., and Seilhammer, J. (1995) *J. Exp. Med.* 181, 877–887