Supplementary Note 1: C-type Lectins

Authors: Chris Cordola

Introduction

C-type Lectins (CTLs) are carbohydrate binding receptors which connect to mostly a wide range of carbohydrate structures (Cambi and Figdor 2003) and proteins as well (Dodd and Drickamer 2001). This gives them the ability to recognize pathogen associated molecular patterns (PAMPs) detecting different pathogens and possibly cancerous cells (Cambi Figdor 2003). They can be involved in transmission of signals after binding to a ligand. Some of them are also known as cell adhesion molecules (CAMs), which enable the transport of immune cells. These are involved in cells traveling to the site of infection through tissue proliferation during the immune response. The C-type Lectins can be either bound to the membrane by a transmembrane region or free-roaming (Weis et al. 1998). The silencing of these genes could lead to the death of an organism, since it will have difficulty preparing an immune response since a C-type Lectin might not be able to detect an organism through carbohydrates usually found on their cell surface, and the absence of CAMs will prevent immune cells from breaking down pathogens as they infect the organism.

Materials and Methods

The C-type Lectin sequences for the following insects were collected and used to search the genome of *Diaphorina citri*: *A. pisum, D. melanogastor, O. fasciatus, A. gambiae, A. aegypti, C. quinquefasciaitus* and *C. lectularius*. These sequences were used in the BLAT search tool in Web Apollo and BLAST from I5K to find gene model predictions. A total of 10 models were found and annotated, correcting anything that was not supported by evidence tracks like RNA sequence data, genomics, and proteomics tracks. Pairwise alignments using BLASTp were performed using the sequences of the completed models to find orthologues in other insects for additional support. Table 1 contains a list of the top results found in closely related insects as well as their ID, Query coverage, and bit score. The listed insects include: *Acyrthosiphon pisum, Cimex lectularius, Anopheles gambiae, Aedes aegypti,* and *Culex quinquefasciatus*.

The sequences from the annotated models were collected in FASTA format. Sequences of the same genes from other insects were found in NCBI and collected as FASTA files. Using a total of 38 sequences a multiple sequence alignment was made with default MUSCLE settings within MEGA 7. The phylogenetic tree was made with sequences from the following organisms: *Apis mellifera, Apis Dorsata, Bactrocera cucurbitae, Ceratitis capitate, Culex quinquefasciatus, Cimex lectularius, Acyrthosiphon pisum, Bombus impatiens, Nicrophorus vespilloides, Stomoxys calcitrans, Athalia rosae, Plutella xylostella, Homo sapiens, Pediculus humanus corporis*, and *Bombyx mori*. The names of most of the genes were simplified to abbreviations, and the species names to the first

letter of the genus name and the whole species name. A neighbor joining tree was then made, based on the previous alignment, with 1000 replicates for bootstrap, Poisson model, and pairwise deletion also within MEGA 7.

Results and Discussion

Ten C-type lectin gene models were located and annotated after using orthologs from closely related insects including *A. pisum, Drosophila melanogastor, Anopheles gambiae, Aedes aegypti, Culex quinquefasciaitus* and *Cimex lectulius*. Most of the models had a C-type Lectin domain that BLAST was able to detect. A multiple alignment was constructed which illustrated sections of conserved sequence. The phylogenetic tree shows that I have indeed found and annotated CTLs, like oxidized low- density lipoprotein receptor, due to their location on the tree in relation to orthologs (Figure 1). Even though the CTLs are conserved from metazoans to humans, there are little to no similarities between CTLs of vertebrates and invertebrates (Dodd and Drickamer 2001). SVEP had not paired with the orthologs found in *H.* sapiens. The lack of accurate pairing possibly indicates that there is a need for more sequences of more closely related organisms to illustrate relations, which was not available. 10 CTL sequences were found and annotated in the *Diaphorina citri* genome while 34 have been found in *Drosophila melanogaster,* 25 in *Anopheles gambiae,* 10 in *Apis mellifera,* and 16 in *Tribolium castaneum* (Tanaka et al. 2008). A comparison to the numbers of CTLs found in those insects show that possibly with better sequence and expression data there might be more CTLs found in the citrus psyllid's genome. Since these sequences have been supported as probable CTLs as more sequence data is available they could be knocked out with RNAi to determine if it could reduce the citrus psyllid population and prevent the spread of citrus greening.

Table 1: A comparison table between *D.citri* and *A. pisum, D. melanogastor, A. gambiae, A. aegypti, C. quinquefasciaitus* and *C. lectulius* derived from the BLAST results. Lists amino acid length of *D.citri* genes, their domains, query coverage, identity, and bit score.

Diaphorina citri gene comparison

Dcitr-	AA	Domains	
oxidized low density lipoprotein receptor 1	140	[1]smart0003	CLECT (CTL or CRD)
oxidized low density lipoprotein receptor 1	134	[1]cd00037	CLECT (CTL or CRD)
oxidized low density lipoprotein receptor 1	93	N/A	N/A
Agglucetin subunit alpha-1	156	[1]smart0003	CLECT (CTL or CRD)
C-type Lectin 8	283	[1]cd00037	CLECT (CTL or CRD)
C-type Lectin 3	188	[2]cd00037, [4]cd00041, [2]pfam00008, [1]cd00054	CLECT (CTL or CRD), CUB, EGF, EGF-CA
C-type Lectin 5	2355	[1]cd00020, [1]cd03590, [5]cd00033	ARM, CLECT DC-SIGN, CCP
E selectin	697	[1]pfam00754, [1]cd03590, [4]smart00032, [3]pfam00084	F5/8type C domain (discoidin; DS), CLECT DC-SIGN, CCP, Sushi
Selectin-like osteoblast-derived protein	794	[1]cd03591, [1]smart00034	CLECT (CTL or CRD), CLECT collectin like
Perlucin	128	[1]cd00037	CLECT (CTL or CRD)
Dcitr-	AA	Domains	
oxidized low density lipoprotein receptor 1	140	[1]smart0003	CLECT (CTL or CRD)
oxidized low density lipoprotein receptor 1	134	[1]cd00037	CLECT (CTL or CRD)
oxidized low density lipoprotein receptor 1	93	N/A	N/A
Agglucetin subunit alpha-1	156	[1]smart0003	CLECT (CTL or CRD)
C-type Lectin 8	283	[1]cd00037	CLECT (CTL or CRD)
C-type Lectin 3	188	[2]cd00037, [4]cd00041, [2]pfam00008, [1]cd00054	CLECT (CTL or CRD), CUB, EGF, EGF-CA
C-type Lectin 5	2355	[1]cd00020, [1]cd03590, [5]cd00033	ARM, CLECT DC-SIGN, CCP
E selectin	697	[1]pfam00754, [1]cd03590, [4]smart00032, [3]pfam00084	F5/8type C domain (discoidin; DS), CLECT DC-SIGN, CCP, Sushi
Selectin-like osteoblast-derived protein	794	[1]cd03591, [1]smart00034	CLECT (CTL or CRD), CLECT collectin like
Perlucin	128	[1]cd00037	CLECT (CTL or CRD)

A. pisum				
Name	Accession	query	ID	Bit Score
PREDICTED: oxidized low-density lipoprotein receptor 1	XP_003245001.1	53%	56%	254
PREDICTED: oxidized low-density lipoprotein receptor 1	XP_003245001.1	75%	56%	342
PREDICTED: oxidized low-density lipoprotein receptor 1	XP_003245001.1	94%	52%	244
NA	NA	NA	NA	NA
PREDICTED: mucin-5AC	XP_001945032.2	91%	66%	663
PREDICTED: uncharacterized protein LOC100159662	XP_001950366.1	98%	77%	811
PREDICTED: uncharacterized protein LOC100167178 isoform X4	XP_016657887.1	61%	43%	1255
PREDICTED: sushi, von Willebrand factor type A, EGF and pentraxi	XP_001948504.2	66%	30%	458
PREDICTED: sushi, von Willebrand factor type A, EGF and pentraxi	XP_001948504.2	91%	27%	464
PREDICTED: perlucin isoform X2	XP_016663197.1	80%	60%	430
A. aegypti				
Name	Accession	query	ID	Bit Score
CTL9	XP_001663957.1	54%	44%	182
CTL9	XP_001663957.1	73%	36%	206
CTL9	XP_001663957.1	93%	46%	180
AAEL006825-PA	XP_001658055.1	66%	45%	252
CTL8	XP_001661274.2	78%	67%	631
CTLGA3	XP_001655089.1	98%	78%	833
CTL5	XP_001661639.2	13%	39%	467
C-Type Lectin (CTL) - selectin like	XP_001653527.2	69%	59%	1507
C-Type Lectin (CTL) - selectin like	XP_001653527.2	95%	43%	1571
C-Type Lectin (CTL) - mannose binding	XP_011493611.1	80%	63%	436

C.lectularius				
Name	**Accession**	**query**	**ID**	**Bit Score**
PREDICTED: oxidized low-density lipoprotein receptor 1-like	XP_014254818.1	54%	50%	223
PREDICTED: oxidized low-density lipoprotein receptor 1-like	XP_014254818.1	70%	63%	333
PREDICTED: oxidized low-density lipoprotein receptor 1-like	XP_014254818.1	96%	50%	211
NA	NA	NA	NA	NA
PREDICTED: uncharacterized protein LOC106673215 isoform	XP_014260729.1	91%	71%	679
PREDICTED: uncharacterized protein LOC106666810	XP_014249760.1	98%	81%	865
PREDICTED: uncharacterized protein LOC106663479 isoform	XP_014243827.1	84%	36%	1811
PREDICTED: CUB and sushi domain-containing protein 3	XP_014245415.1	85%	58%	1759
PREDICTED: CUB and sushi domain-containing protein 3	XP_014245415.1	95%	42%	1626
PREDICTED: perlucin isoform X2	XP_014243366.1	81%	58%	397

A. gambiae				
Name	**Accession**	**query**	**ID**	**Bit Score**
CTL9	XP_312307.1	53%	43%	170
CTL9	XP_312307.1	73%	33%	204
CTL9	XP_312307.1	59%	62%	191
CTL7	XP_309125.4	89%	45%	366
CTL8	XP_313387.4	91%	64%	618
CTLGA3	XP_319371.2	98%	77%	822
CTL5	XP_310650.4	70%	38%	2059
C-Type Lectin (CTL) - selectin like	XP_309111.3	78%	59%	1529
C-Type Lectin (CTL) - selectin like	XP_309111.3	96%	42%	1562
C-Type Lectin (CTL) - mannose binding	XP_315346.4	80%	63%	420

D.melanogaster				
Name	**Accession**	**query**	**ID**	**Bit Score**
C-type Lectin	NP_001138057.1	53%	43%	174
C-type Lectin	NP_001138057.1	66%	39%	235
C-type Lectin	NP_650444.1	64%	61%	187
C-type Lectin	NP_572269.1	89%	45%	355
narrow, isoform B	NP_001188963.1	90%	64%	613
C-type lectin 27kD, isoform A	NP_608858.3	98%	74%	798
cubilin ortholog	NP_727348.2	28%	32%	213
furrowed, isoform A	NP_511136.2	80%	56%	1494
furrowed, isoform A	NP_511136.2	95%	41%	1521
triforce, isoform A	NP_728586.1	81%	68%	445

| C. quinquefasciatus				
Name	**Accession**	**query**	**ID**	**Bit Score**
conserved hypothetical protein	XP_001847975.1	57%	41%	169
conserved hypothetical protein	XP_001847975.1	73%	37%	210
NA	NA	NA	NA	NA
conserved hypothetical protein	XP_001868565.1	68%	45%	261
conserved hypothetical protein	XP_001866032.1	78%	66%	626
conserved hypothetical protein	XP_001847297.1	98%	79%	844
conserved hypothetical protein	XP_001862639.1	16%	43%	760
furrowed	XP_001869209.1	71%	28%	434
furrowed	XP_001869209.1	78%	32%	343
galactose-specific C-type lectin	XP_001842113.1	58%	45%	161
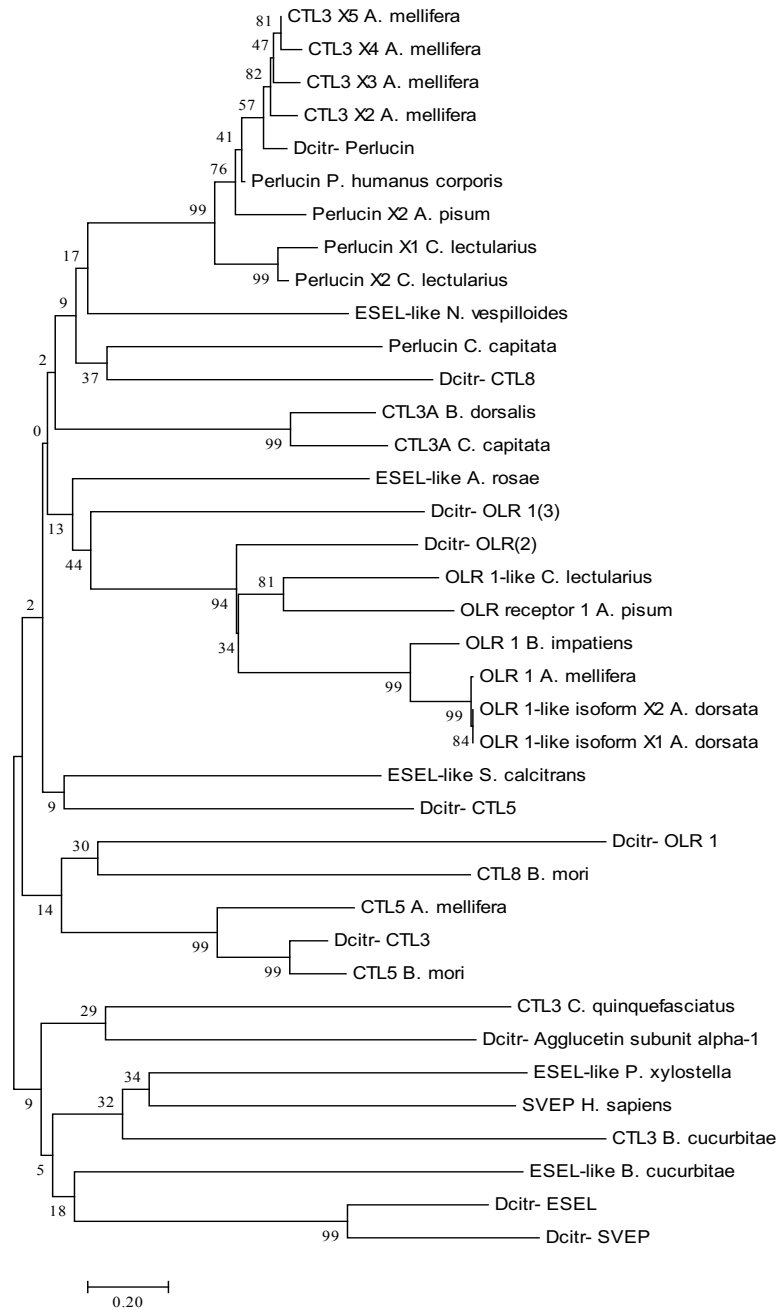
Figure1: Phylogenetic tree consisting of *Diaphorina citri* C-type Lectins and the same genes from other insects. The C-type Lectins found in *D.citri* are the following: oxidized low density lipoprotein receptor (OLR), C-type Lectin 3 (CTL3), C-type Lectin 5 (CTL5), E selectin (ESEL), Perlucin, Agglucetin subunit alpha, and selectin like osteoblast derived protein (SVEP). The species names of the additional organisms are shown next to the gene name. An ortholog from another insect could not be found for Agglucetin. Only a human ortholog could be found for SVEP. High bootstrap values show that some genes like OLR are closely related to those in other insects. OLR is also known as CTL8A.

References

1. Cambi A and Figdor CG. 2003. Dual function of c-type lectin-like receptors in the immune system. Curr Opin Cell Biol. 15:539-546. PMID: 14519388
2. Dodd RB and Drickamer K. 2001. Lectin-like proteins in model organisms: implications for evolutions of carbohydrate- binding activity. Glycobiology 11(5):71R-79R. PMID: 11425795
3. Tanaka H, Ishibashi J, Fujita K, Nakajima Y, Sagisaka A, Tomimoto, K, Suzuki, N, Yoshiyama, M, Kaneko, Y, Iwasaki, T, Sunagawa, T, Yamaji, K, Asaoka, A, Mita, K, Yamakawa, M. 2008. A genome-wide analysis of genes and gene families involved in innate immunity of Bombyx mori. Insect Biochem Mol Biol 38(12): 1087-1110. PMID: 18835443
4. Weis WI, Taylor ME, and Drickamer K. 1998. The c-type lectin superfamily in the immune system. Immunological Reviews 163:19-34. PMID: 9700499