Supplementary Note 5: Thioester containing Proteins

Authors: Gabe DeAvila and Kayla Shore

Introduction

Thioester containing Proteins (TEPs) are members of an ancient protein family that includes vertebrate C complement and alpha-2-macroglobulin proteins (2). Insect TEPs seem to play a similar role to their vertebrate homologs, binding to invaders such as parasites or microbes, marking them for degradation (1). These proteins may be upregulated by the JAK/STAT pathway during innate immune response (1). In some TEPs, the interaction between invader and TEP occurs via a conserved thioester motif (GCGEQ), specifically TEP 1-4, but other TEPs lack this domain (2). TEPs also include a long C-terminal cysteine signature (2). Previous studies indicate that insect TEPs fall into three clades, with one clade being specific to mosquitos (3). Potential TEPs in *Diaphorina citri* were identified and investigated in comparison to sequences from other insects, including mosquito species, *Drosophila* species, and the pea aphid.

Methods

Investigation into potential TEPs in *Diaphorina citri* was initiated by first collecting known TEP sequences from other insects, including *Anopheles gambiae, Aedes aegypti, Culex quinquefasciatus, Drosophila melanogaster, Tribolium castaneum* and *Acrythosiphon pisum* (4) (7). Using the gathered sequences, individual searches were conducted for each of the 46 orthologs provided using the protein blat and blast functions in i5k and NCBI, respectively. Two predicted proteins were implicated as those most closely related to the known TEPs of related organisms. The corresponding sequences were then extracted from NCBI, and searched within the online annotation program, WebApollo. In WebApollo, the NCBI predicted sequence, the RNA seq data, and the Maker predicted genes were used as reference for manual annotation. Using the NCBI predicted sequences as a base, exons were edited based on the evidence provided by the RNAseq data. After ensuring that the NCBI model corresponded as precisely as possible to the RNAseq data, a multiple sequence alignment was created through MUSCLE to determine whether the constructed *D. citri* model was similar to the known sequences from the previously listed organisms. MEGA 7 was used to perform a local alignment on all of these sequences, from which a proposed phylogenetic relationship was created, using a neighbor-joining phylogeny, as illustrated in Figure 1.

Results and Discussion

Based on the current gene assembly, only two TEPs were able to be identified within the psyllid, *Diaphorina citri.* This number of TEPs is comparable to that of *Acrythosiphon pisum* and *Nasonia vitripennis,* indicating that these are mostly likely the only two TEP genes in the psyllid's genome, although isoforms may be present.

Organism	**Number of TEPs**
Drosophila melanogaster	6
Anopheles gambiae	13
Tribolium castaneum	4
Apis mellifera	4
Nasonia vitripennis	3
Acyrthosiphon pisum	2
Diaphorina citri	2

Table 1 : TEP in *D. citri* and related organisms (5)(6)

The first gene TEP3 (Thioester_containing_protein 3) is 28,982 base pairs in length. It includes 24 exons, which are typically short and evenly spread between long introns. It was found to contain the distinct GCGEQ motif towards the C terminal.The second gene TEP6 (Thioester_containing_protein 6) is 30,776 base pairs in length. Within this gene there are 19 exons in three distinct clusters. The GCGEQ motif was not found within this protein, giving more evidence to its identity as an ortholog of TEP6. RNAseq data suggests that this model is one of many possible isoforms. Due to the length and the number of exons, determining the number of isoforms present was not possible at this time, and requires further investigation.

		A. pisum	***D. melanogaster***	***T. castaneum***
Thioester containing protein 3	Bit Score (name)	1400 (CD109 antigen)	704 (macroglobulin complement-related)	981 (CD 109 antigen)
	Query (Identity)	95% (65%)	95% (38%)	96% (46%)
	E value	0	0	0
Thioester containing protein 6	Bit Score (name)	1000 (CD109 antigen)	805 (macroglobulin complement-related)	834 (CD109 antigen)
	Query (Identity)	96% (58%)	93% (67%)	94% (69%)
	E value	0	0	0

Table 2: Blast matches of ACP TEP proteins to sequenced arthropods

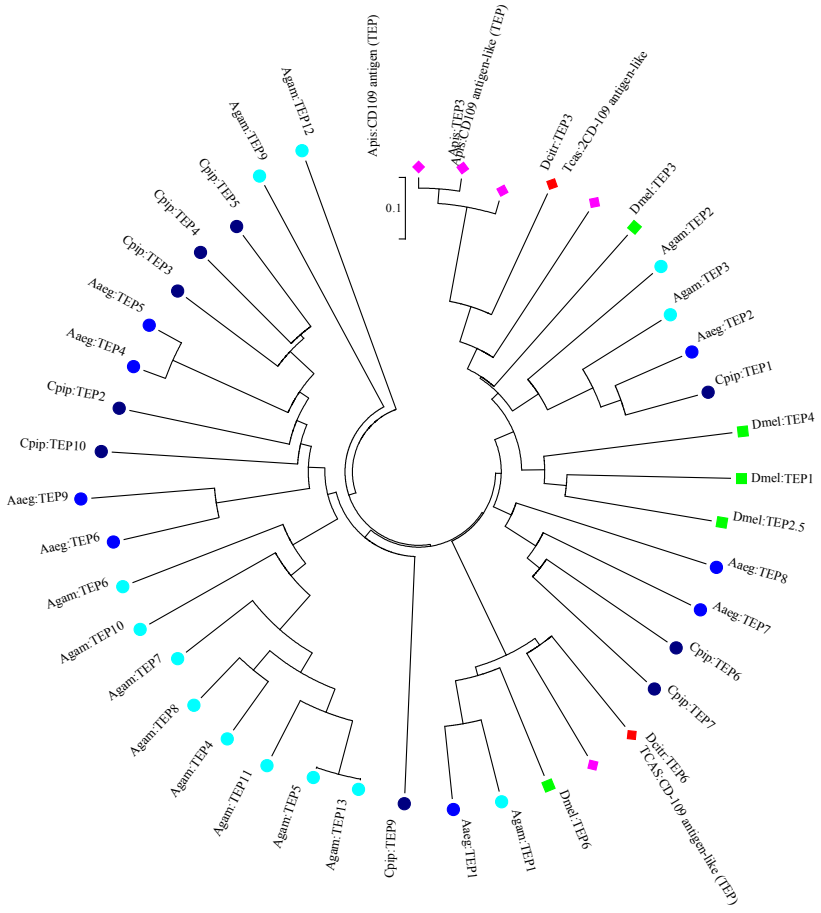

Figure 1: Proposed phylogenetic relationship between Diaphorina citri (Dcitr), Acrythosiphon pisum (Apis), Tribolium castaneum (Tcas), Anopheles gambiae (Agam), Aedes aegypti (Aaeg), Culex quinquefasciatus (Cpip) and Drosophila melanogaster (Dmel)

References

1. Agaisse, H. and Perrimon, N. “The roles of JAK/STAT signaling in Drosophila immune responses.” *Immunological Reviews* (2004), 198: 72–82. doi:10.1111/j.0105-2896.2004.0133.x

2. Blandlin, Stephanie et al “Thioester-containing proteins and insect immunity.” *Molecular Immunology Vol 40* (2004) 903-908

3. Bou Aoun, Richard et al. “Analysis of Thioester-Containing Proteins during the Innate Immune Response of *Drosophila Melanogaster*.” *Journal of Innate Immunity* 3.1 (2010): 52–64. *PMC*. Web. 28 Sept. 2016.

4. Gerardo et al. “Immunity and other defenses in pea aphids” *Genome Biology* (2010)

5. Lagueux, Marie et al. “Constitutive Expression of a Complement-like Protein in Toll and JAK Gain-of-Function Mutants of *Drosophila*.” *Proceedings of the National Academy of Sciences of the United States of America* 97.21 (2000): 11427–11432. Print.

6. Viljakainen, Lumi “Evolutionary genetics of insect innate immunity” *Briefings in Functional Genomics* (2015)

7. Zou, Zhen et al. “Comparative genomic analysis of the Tribolium immune system” *Genome Biology* (2007)