Supplementary Note 12: Cactus proteins

Authors: Teresa Shippy and Yen-Ta Chen

Introduction

Cactus is an I kappa B factor that acts in the Toll pathway in insects (Geisler et al, 1992). In *Drosophila*, Cactus binds the NF-kappa B transcription factors Dorsal and Dif, keeping them in the cytoplasm. When the Toll pathway is activated by the presence of an infectious agent (usually Gram-positive bacteria or fungi), Cactus is phosphorylated and then degraded. This releases Dorsal or Dif to translocate to the nucleus to activate target genes such as antimicrobial peptides (see Valanne et al. 2011)

Most insects whose genomes have been sequenced have a single ortholog of Cactus (see Viljakainen 2015 and Table 1), although the honeybee *Apis mellifera* has three (Evans et al 2006) and the Argentine ant *Linepithsma humile* has two (Smith et al 2011).

Organism	Order	# of cactus genes
Tribolium castaneum	Coleoptera	1
Anopheles gambiae	Diptera	1
Drosophila melanogaster	Diptera	1
Acyrthosiphon pisum	Hemiptera	1
Nilaparvata lugens	Hemiptera	1
Apis mellifera	Hymenoptera	3
Nasonia vitripennis	Hymenoptera	1
Linepithsma humile	Hymenoptera	2
Bombyx mori	Lepidoptera	1

Table 1. Gene counts of *cactus* orthologs in various insects

Predicted Protein	Top BLASTp hit	Bitscore	E value	Percent Identity
Cactus (MCOT19824.0.CT)	PREDICTED: NF-kappa-B inhibitor cactus [Bemisia tabaci]	212	6e-63	39%

Table 2. BLAST analysis of the predicted Cactus protein.

Methods

Acyrthosiphon pisum Cactus was used to query the predicted *D. citri* protein sets (Diacit_International_psyllid_consortium_proteins_v1 and

Diacit_RefSeq_proteins_Release_100) at i5k@NAL. The best hit was BLASTed against Insecta with NCBI BLAST to verify its identity. The locus encoding this protein was identified and manually annotated in Web Apollo. We also performed a BLAST search of the *D. citri* MCOT set to compare our gene model to *de novo* assembled transcripts. Multiple alignments were performed in Muscle (http://www.ebi.ac.uk/Tools/msa/muscle/). We used MEGA7 to construct a phylogenetic tree via the neighbor-joining method. Sequences for multiple alignment and phylogenetic analysis were obtained from NCBI, FlyBase, ImmunoDB and the Bordenstein Lab (NSF DEB-1046149).

Results

Using *Acyrthosiphon pisum* Cactus as a BLAST query, we identified a single *D. citri cactus* gene. An existing NCBI gene model was manually edited to include additional exons encoding motifs shared with other Cactus orthologs. However, comparing the predicted protein to orthologous proteins reveals that these motifs do not occur in the expected order, suggesting a possible problem with the genome assembly in this region. We then used the predicted Cactus protein to BLAST the MCOT predicted protein set. We identified a single MCOT protein (MCOT19824.0.CT), which perfectly matched several stretches of the predicted Cactus protein. Again, these regions were not in the same order. The region corresponding to the C-terminal portion of the MCOT protein was located in the middle of the assembly-predicted protein. Since the MCOT protein is a *de novo*-assembled transcript from Trinity and the order of its sequence matches that found in Cactus genes from other insects, we conclude that the genome is incorrectly assembled in this region causing the *cactus* exons to be in the wrong order. Therefore, we used the MCOT Cactus protein for the remainder of our analysis.

Using the MCOT Dcitri_Cactus protein to BLAST against insect proteins retrieves Cactus orthologs from many insects. The top hit is to the Cactus ortholog of another hemipteran, *Bemisia tabaci*, the silverleaf whitefly (Table 2). In a phylogenetic tree, MCOT Dcitri_Cactus also clusters with another hemipteran Cactus ortholog (Fig. 1)

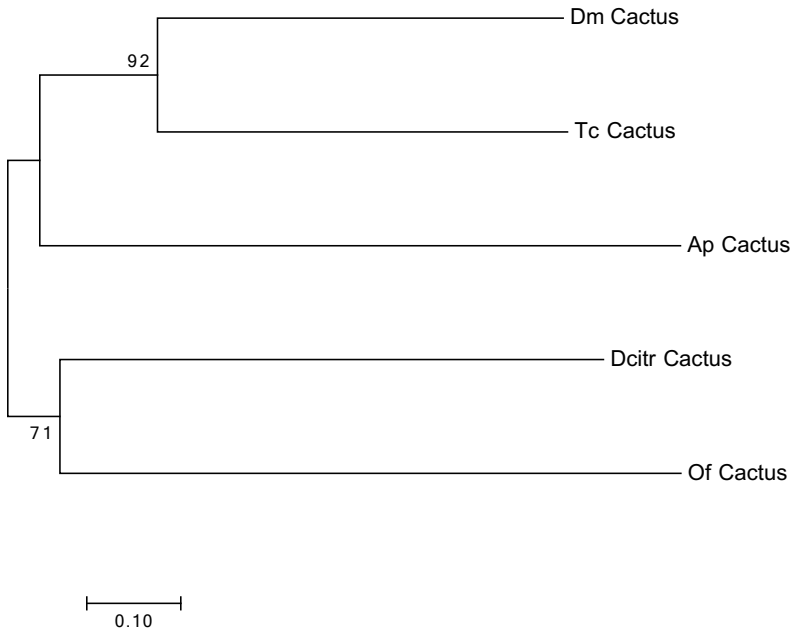

Figure 1. A neighbor-joining tree of Cactus proteins from *Drosophila melanogaster* (Dm), *Tribolium castaneum* (Tc), *Acyrthosiphon pisum* (Ap), *Diaphorina citri* (Dcitr) and *Oncopeltus fasciatus* (Of).

References

1. Evans, J., Aronstein, K., Chen, Y., Hetru, C., Imler, J.-L., Jiang, H., Kanost, M., Thompson, G., Zou, Z., and Hultmark, D. (2006). Immune pathways and defence mechanisms in honey bees Apis mellifera. Insect Mol Biol *15*, 645–656.

2. Geisler, R., Bergmann, A., Hiromi, Y., and Nüsslein-Volhard, C. (1992). cactus, a gene involved in dorsoventral pattern formation of Drosophila, is related to the IκB gene family of vertebrates. Cell *71*, 613–621.

3. Smith, C., Zimin, A., Holt, C., Abouheif, E., Benton, R., Cash, E., Croset, V., Currie, C., Elhaik, E., Elsik, C., et al. (2011). Draft genome of the globally widespread and invasive Argentine ant (Linepithema humile). Proc Natl Acad Sci *108*, 5673–5678.

4. Valanne, S., Wang, J.-H., and Rämet, M. (2011). The Drosophila Toll Signaling Pathway. J Immunol *186*, 649–656.

5. Viljakainen, L. (2015). Evolutionary genetics of insect innate immunity. Briefings Funct Genom *14*, 407–412.