Supplementary Note 3: Fibrinogen related proteins

Authors: Tracey Bell

Introduction

Fibrinogen related proteins (FREPS) include a highly conserved fibrinogen like domain (FBG domain) found universally in vertebrates and invertebrates (Wang et al, 2005). In invertebrates the FBG domains are predicted to function in innate immunity by recognizing carbohydrates and their derivatives on the surface of microorganisms. The FBG domain includes approximately 200 amino acids with high sequence similarity to the C terminus of fibrinogen, however some FREPs include a truncated domain (Middah and Wang, 2008). A number of N-terminal domains can be joined to the highly conserved FBG domain allowing for variability and antibodies against mammalian fibrinogen can be used to find proteins with FBG domains in invertebrates (Hanington and Zhang, 2010). FREP genes include angiopoietin, tenascin, scabrous, aslectin, microfibril associated proteins (MAP), ficolin, and tachylectin; all of which were searched for in the *D. citri* genome.

Methods

Fibrinogen related protein genes were collected from NCBI, i5k, Ensembl and FlyBase. BLAT searches were then performed at the i5k workspace to identify homologs in the *D. citri* genome. Potential hits within the *D. citri* genome were explored further using the WebApollo genome browser. NCBI Gnomon annotation gene models were manually annotated within WebApollo using BLAST and RNA-Seq evidence tracks. Completed gene models were analyzed by BlastP to confirm accuracy and completeness of the annotation.
Several models were found when searching the genome for the fibrinogen C-terminal domain. Sequences from completed annotated genes were used to perform BlastP searches against related insects and model organisms. Potential models found in this manner were used in BLASTp searches to verify. This includes gene models named angiopoietin, tenascin, and three fibrillin models.

Pairwise amino acid sequence comparisons were performed between *D citri*, *Acyrthosiphon pisum*, *Drosophila melanogaster*, *C. lectularius*, *Anopheles gambiae*, *A. aegypti*, and *C. quinquefasciatus*. Collected orthologs and annotated *D. citri* FREP sequences were used to generate a multiple sequence alignment using ClustalW in MEGA6. Sequences were chosen to identify close relationships to proteins that include the FBG domain. The *D. citri* proteins were named per the NCBI BLAST results. A neighbor joining phylogenetic tree was produced from the multiple sequence alignment to analyze the evolutionary relationships between the *D. citri* FREPS and well characterized sequences. (Figure 1) The unrooted Neighbor-Joining phylogenetic tree was constructed with the MEGA6 program using the bootstrap method (with 1000

replications) as a test of phylogeny and p-distance method as the substitution model. The analysis includes 38 amino acid sequences with 53 positions in the data set of FREP genes proposed by NCBI.

Results and Discussion

Three FREPS were found within the *D. citri* genome. Several potential gene models had insufficient RNA sequence data and/or domains to facilitate an approved gene model. *Anopheles gambiae* has been shown to have 53 FREPS and *Drosophila melanogaster* to have 20 FREPS (Wang et al, 2005). (Table 1) It is likely the *D. citri* will have a closer amount to *D. melanogaster*, as it has been shown the FREPs have expanded in *A. gambiae* due to recognizing parasites from blood feeding.

One gene model is attributed to the ImmunoDB FREP ortholog search, scabrous protein, which includes the fibrinogen C-terminal domain. Scabrous protein has been shown in Drosophila melanogaster to be involved in the regulation of neurogenesis and (along with Gp150) as an endosomal protein that regulates Notch activity (Li et al, 2003). A search for the protein tenascin revealed one gene model although there are several members in the tenascin family. Tenascins are extracellular matrix proteins, which are associated with organogenetic processes, with neuronal and axon development, and mediate cell adhesion (Wang et al, 2005). A search for the protein angiopoietin revealed one gene model. Angiopoietin is involved with embryonic and postnatal angiogenesis. Other FREPS that were not found in the genome include ficolin, tachylectins and aslectin, which are likely involved in parasite detection within the innate immune system (Wang et al, 2005). Ficolin and tachylectin are involved with N-acetylglucosamine (GlcNAc) binding activity (Middah and Wang, 2008). Aslectin can also bind bacteria and GlcNAc.

The multiple sequence alignment and phylogenetic tree position the scabrous model in a separate clade alongside the other hemipteran scabrous proteins. The angiopoietin and tenascin models were also in the proper clades. (Figure 1)

Table 1: The numbers of named FREP gene copies in *Diaphorina citri*, *Acyrthosiphon pisum*, *Drosophila melanogaster*, *Cimex lectularius*, *Anopheles gambiae*, *Aedes aegypti*, and *Culex quinquefasciatus*. An extensive search was performed through NCBI search, ENSEMBL, and www.supfam.org to obtain final figures for numbers of FREP genes in other hemipterans.

Fibrinogen Related Proteins	Scabrous	MAP	Ficolin	Angiopoietin	Tenascin	Other
D. citri [a]	1	0	0	1	1	0
A. pisum [a]	1	1	0	1	1	0

C. lectularius [a]	1	1	0	1	0	N/A
D. melanogaster [a] [c]	1	1	0	0	2	16
A. gambiae [b] [c]	1	0	2	1	0	55
A. aegypti [b] [c]	1	1	0	0	0	33
C. quinquefasciatus [a] [c]	1	1	3	3	2	83

[a] NCBI search, [b] ENSEMBL, [c] www.supfam.org

Table 2: FREP sequence table. The protein sequences used for the phylogenetic analysis with their corresponding BLAST match results for specific species, with the bit scores followed by the coverage %, ID%, and accession numbers.

| *D. citri Gene Named*	*D. citri AA Length*	*A. pisum BLAST Match Bit Score*	*BLAST Query Percent*	*BLAST ID Percent*	*A. pisum Acc #*	*D. melanogaster BLAST Match Bit Score*	*BLAST Query Percent*	*BLAST ID Percent*	*D. melanogaster Acc #*
Scabrous	897	434	86	83	XP_001951011.3	419	87	78	NP_001188463.1
Angiopoietin	567	337	99	32	XP_001944698.3	337	95	33	AAA28880.1
Tenascin	977	313	72	31	XP_008178803.2	57.8	9	33	NP_609691.6
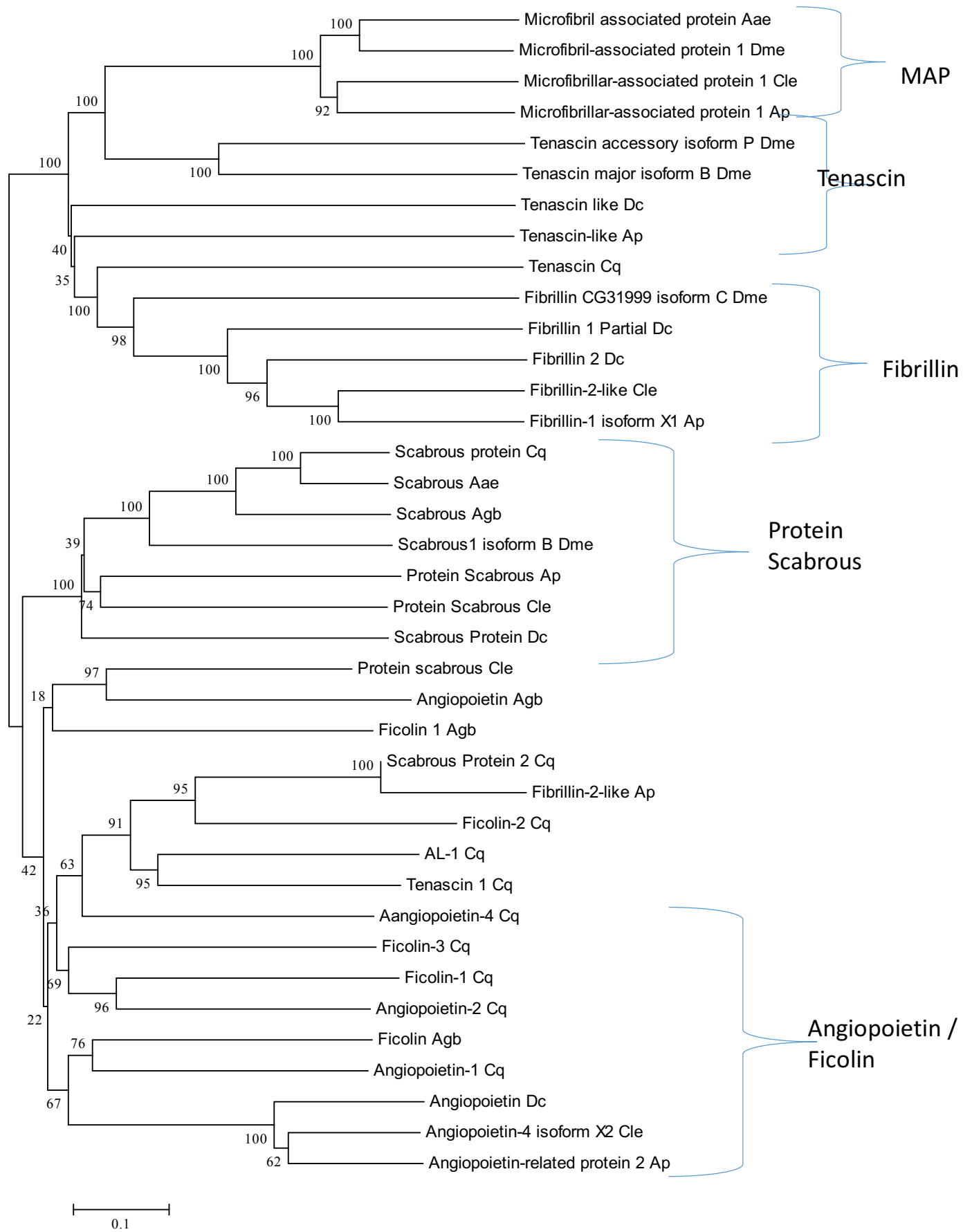
Figure 1: Phylogenetic analysis of Hemipteran FREP family protein sequences

The unrooted Neighbor-Joining phylogenetic tree was constructed with the MEGA6 program using the bootstrap method (with 1000 replications) as a test of phylogeny and p-distance method as the substitution model. The analysis includes 38 amino acid sequences with 53 positions in the data set of FREP genes proposed by NCBI. All bootstrap values are shown near the nodes. The *D. citri* proteins are numbered in the order found. The *A. pisum* proteins were named according to their EST numbers. The *D. Melanoaster, A gambiae,* and proteins were renamed from their accession number name. Abbreviations: *Diaphorina citri* Dc; *Acyrthosiphon pisum* Ap; *Anopheles gambiae* Agb; *Drosophila melanogaster* Dme; *Cimex lectularius* Cle, *Aedes aegypti* Aae, and *Culex quinquefasciatus* Cq.

References

1. Identification and characterization of the fibrinogen-like domain of fibrinogen-related proteins in the mosquito, Anopheles gambiae, and the fruitfly, Drosophila melanogaster, genomes. Xinguo Wang, Qin Zhao, and Bruce M Christensen. BMC Genomics20056:114 Wang et al; licensee BioMed Central Ltd. 2005. DOI: 10.1186/1471-2164-6-114
2. Scabrous and Gp150 are endosomal proteins that regulate Notch activity. Yanxia Li, Michael Fetchko, Zhi-Chun Lai, Nicholas E. Baker. Development 2003 130: 2819-2827; doi: 10.1242/dev.00495
3. The Scabrous protein can act as an extracellular antagonist of Notch signaling in the Drosophila wing. Lee, E-Chiang et al. Current Biology , Volume 10 , Issue 15 , 931 – 934. DOI: http://dx.doi.org/10.1016/S0960-9822(00)00622-9
4. Hanington P, C, Zhang S, -M. The Primary Role of Fibrinogen-Related Proteins in Invertebrates Is Defense, Not Coagulation. J Innate Immun 2011;3:17-27 DOI: 10.1159/000321882
5. Evolution and potential function of fibrinogen-like domains across twelve Drosophila species. Sumit Middha and Xinguo Wang. BMC Genomics20089:260 DOI: 10.1186/1471-2164-9-260