

WHITEFLY VECTOR (*BEMISIA TABACI*) PROTEOME ELUCIDATION: FIRST STEPS TOWARD UNRAVELING THE COMPLEXITY OF WHITEFLY-BEGOMOVIRUS INTERACTIONS

Brown, J.K.¹; Breci, L.²; Saripalli, C.¹; He, R.³; Cicero, J.¹; Nelson, W.M.⁴; Tsprailis, G.⁵; Gang, D.R.³; Soderlund, C.⁴.

¹*Department of Plant Sciences, University of Arizona, Tucson, AZ 85721; jbrown@ag.arizona.edu*

²*Department of Chemistry and Biochemistry, University of Arizona;*

³*Institute of Biological Chemistry, Washington State University;*

⁴*BIO5 Institute, University of Arizona;*

⁵*Center for Toxicology, University of Arizona.*

The whitefly *Bemisia tabaci* (Gennadius) sibling species group is one of most important arthropod vectors of plant viruses. Among these are the widespread and damaging genus, *Begomovirus*, family, Geminiviridae. While much research has been focused on whitefly-vector interactions at the level of transmission behavior, less is known about molecular, cellular, or genomic level interactions. To elucidate proteins (genes) essential to the begomovirus-whitefly transmission pathway we are characterizing the adult whitefly, alimentary canal, and salivary gland proteins at the proteome and/or transcriptome levels for *B. tabaci* biotype B, with a particular focus on those that interact directly and indirectly with virions during whitefly-mediated transmission. A total of 12 mass spectrometry runs were performed on total proteins extracted of the adult (i) whole whitefly (9 MudPIT's) and (ii) whitefly gut (3 MudPIT's), respectively. Protein preparations were digested with trypsin and subjected to LC-LC-MS/MS to determine each peptide spectra using Sequest and X! Tandem software. Scaffold (ver. 2_05_01) was used to validate MS/MS peptides and perform protein identifications, with a minimum of two peptides required for a valid protein call. The resultant proteins were identified using three different 'search databases' (dbs): (i) whitefly ESTs (sequenced, assembled, and annotated in PAVE by our group), referred to as whitefly version 4 ('WFV4') that contains 774,065 translated ESTs), (ii) a set of translated ESTs from ~12 insect species selected from the available insect sequence databases, herein, 'Select Insect' (SI-Ins), and (iii) all proteins of all organisms of the class 'Insecta' (Insecta). The latter two databases were downloaded to the The University of AZ supercomputer (last download 03.11). A total of 597 unique whitefly proteins were identified from the 12 MudPIT runs. Over 500 unique proteins were identified from the 'whole whitefly' preparation (9 MudPITs; 2,145 unique peptides; 35,663 spectra), whereas, whitefly gut preparations alone yielded 411 unique proteins (3 MudPITs; 1484 unique peptides; 14,703 spectra). Using the combined 'SI-Ins' and 'Insecta' dbs as the search databases, 245 whole whitefly and 166 gut protein hits were identified. Protein hits (gi numbers) from the three databases were imported into the Uniprot batch retrieval system to facilitate identification of top hits to organism and insect species, and based on GO Slim, KEGG, and Pfam descriptions. Using the combined 'SI-Ins' and 'Insecta' dbs, pea aphid (Hemiptera: s.o. Homoptera: Aphididae) sequences provided the greatest number of hits, followed by fruit fly (Diptera: Drosophilidae), and then body louse (Phthiraptera: Pediculidae). In contrast, both the 'WFV4' db, and the

batch retrieval from Uniprot had the greatest number of hits to fruit fly, followed by many other species, and then yellow fever mosquito. The top protein hits were dominated by predicted (61) and putative (59) hits in the combined 'SI-Ins-Insecta' dbs, whereas, putative hits (42) dominated for the 'WFV4' db. Overall the extent of shared homology between the whitefly proteins and ESTs/proteins in the three search databases was expected to be highest for other homopteran sequences, namely, the pea aphid, which is the best represented homopteran to date. This pattern was observed for the 'SI-Ins' and 'Insecta' dbs annotations, but not for the 'WFV4' db, an observation that likely reflects the less than optimal number of annotated ESTs in 'WFV4' database, even though it contains ~775,000 ESTS (15,045 contigs and 39,864 singletons) and whitefly transcripts share 100% homology with whitefly proteins. Further only 17% of whitefly transcripts in 'WFV4' had a top hit against a Uniprot accession. These results underscore the need for additional functional genomics projects for plant virus vectors and other homopterans that transmit plant pathogens.