

FIFTH ANNUAL Arthropod Genomics SYMPOSIUM

Center for Genomic Studies on Arthropods Affecting Human, Animal and Plant Health • Kansas State University

June 9-12, 2011 • Kansas City

POSTER ABSTRACTS

*Indicates name of person presenting poster

(Cross-reference list on last page lists first authors alphabetically)

1

Sequencing of the Southern House mosquito (*Culex quinquefasciatus*) establishes a platform for mosquito comparative genomics

*Arensburger, P.; Megy K.; Waterhouse R. M.; Atkinson P. W.; the *Culex* genome consortium (PA, PWA) University of California Riverside, Center for Disease Vector Research, Riverside, CA 92507; (KM) European Bioinformatics Institute, Hinxton, Cambridge CB10 1SD, UK; (RMW) University of Geneva Medical School, 1 rue Michel-Servet, 1211 Geneva, Switzerland.

Culex quinquefasciatus is an important mosquito vector of viruses, such as West Nile virus and St. Louis encephalitis virus as well of nematodes that cause lymphatic filariasis. It is one species within the *Culex pipiens* species complex and enjoys a distribution throughout tropical and temperate climates of the world. Following the whole genome sequencing of *Anopheles gambiae* (vector of malaria) and *Aedes aegypti* (vector of dengue and yellow fever) this is the third mosquito genome to be published and completes the goal of sequencing a representative species from the three major taxonomic groups of vector mosquitoes. These insects, which have been the subject of control efforts, can now be studied using three way comparative genomics.

C. quinquefasciatus contains a large number of protein-coding genes (18,883 genes), 22% more than *Ae. aegypti* and 52% more than *An. gambiae*. This increase in gene number is partially explained by multiple gene family expansions, including expansion in olfactory and gustatory receptors, salivary gland genes, and genes associated with xenobiotic detoxification. These expansions may have contributed to the ability of this species to inhabit a wider geographic area than either of the other two sequenced mosquitoes as well as to its ability to blood feed on a wide range of hosts (humans, birds, and livestock). The three mosquito species also exhibit significant differences in transposable element abundances between genomes (28% of the *C. quinquefasciatus* sequenced genome, 47% of *Ae. aegypti*, 16% of *An. gambiae*), suggesting either increased levels of transposable element activity or reduced level of selection against transposable elements in some lineages.

2

Fungus-Growing Ant Genomics: A comparative approach

*Marsh, Sarah E.; Grubbs, Kirk J.; Caldera, Eric J.; Scott, Jarrod J.; Li, Lewyn; Teiling, Clotilde; Warren, Wes; Elsik, Christine G.; Clifton, Sandra W.; Smith, Christopher R.; Holt, Carson; Yandell, Mark; Smith, Christopher D.; Gadau, Jürgen; Weinstock; George; Gerardo, Nicole M.; Suen, Garret; Currie, Cameron R.

(SEM, KJG, EJC, JJS, GS, CRC) University of Wisconsin-Madison, Department of Bacteriology, Madison, WI 53706; (SEM) University of Wisconsin-Madison, Graduate Training Program in Genetics, Madison, WI 53706; (LL, CT) 454 Life Sciences, Branford, CT 06405; (WW, SWC, GW) Washington University at St. Louis, The Genome Center, St. Louis, MO 63108; (CGE) Georgetown University, Department of Biology, Washington, DC 20057; (CRS) Earlham College, Department of Biology, Richmond, IN 47374; (CH, MY) University of Utah, Department of Human Genetics, Salt Lake City, UT 84111; (CDS) San Francisco State University, Department of Biology, San Francisco, CA 94132; (JG) School of Life Sciences, Arizona State University, Tempe, AZ 85287; (NMG) Emory University, Department of Biology, Atlanta, GA 30322; (GS, CRC) University of Wisconsin-Madison, Great Lakes Bioenergy Research Center, Madison, WI 53706;

The fungus-growing ant-microbe symbiosis originated 50 million years ago. The ants obligately depend on their mutualistic fungal cultivar, which serves as their primary food source. The diversity of fungus-growers includes over 230 species of ants, classified into 13 genera and grouped into four distinct agricultural systems depending on the fungal symbionts they cultivate. Leaf-cutter ants (genera *Atta*, *Acromyrmex*) are the most derived, use fresh plant material, and can reach massive colony

sizes with elaborate caste structures and millions of workers. In contrast, basal attine ants (genera *Apterostigma*, *Mirmycocrypta*, *Mycocepurus*) live in small colonies composed of dozens to a few hundred individuals and simple social structures. We recently sequenced the genome of the leaf-cutter ant *Atta cephalotes* from three haploid males from Panama using a combination of 454 whole-genome shotgun, 8kb and 20kb paired end reads. Assembly of shotgun reads yielded 42,754 contigs with an average length of 6,788 bp and an N50 of 14,240 bp. The addition of 8 and 20 kb insert paired-end sequencing generated 2,835 scaffolds, with an N50 value of 5,154,504. Our estimates suggest a genome size of 290-317 Mb. Our analyses provide evidence that the *Atta cephalotes* genome reflects this ant's nutritional dependence on its fungal mutualist. These include extensive reductions in serine proteases (likely unnecessary because proteolysis is less important when processing nutrients supplied by the fungus), a loss of two genes in the arginine biosynthesis pathway (suggesting that this amino acid is provided by the fungus), and the absence of a hexamerin (which sequesters amino acids during larval development in other insects). We are in the process of sequencing a basal attine ant, *Apterostigma dentigerum*, from Panama. In addition to shotgun sequencing of five haploid males using Roche's titanium chemistry (with an average read length ~400 bp) and paired-end sequencing of a pool of seven haploid males and diploid females, we have early access to Roche's new 1 kb reads. We are examining the benefit of these longer reads on genome assembly. Interestingly, estimates put the *A. dentigerum* genome size at nearly twice that of *Atta cephalotes*. Our *A. Dentigerum* genome will allow us to explore the evolutionary patterns of genome reduction in this tribe of ants, and to identify whether the observed patterns of genomic changes in response to the cultivation of fungi occurred at the early origins of agriculture in these ants. Comparison of these and other genome characteristics across the diversity of fungus-growing ants promises insights into the genome architecture of this ecologically important group.

3

The First Sequenced Genome of a Twisted-Wing Parasite (Insecta: Strepsiptera) and its Phylogenetic Implications

*Niehuis, Oliver; Hartig, Gerrit; Grath, Sonja; Pohl, Hans; Donath, Alexander; Eisenhardt, Carina; Hertel, Jana; Krauss, Veiko; Lehmann, Jörg; Mayer, Christoph; Tafer, Hakim; Meusemann, Karen; Beutel, Rolf G.; Stadler, Peter F.; Bornberg-Bauer, Erich; Misof, Bernhard

(AD, BM, CM, KM, GH, ON) Zoologisches Forschungsmuseum Alexander Koenig, Zentrum für Molekulare Biodiversitätsforschung, 53113 Bonn, Germany; (SG, EBB) Westfälische Wilhelms-Universität Münster, Institut für Evolution und Biodiversität, Evolutionäre Bioinformatik, 48149 Münster, Germany; (CE) Universität Leipzig, Institut für Biologie II, 04103 Leipzig, Germany; (HT, JH, JL, PFS, VK) Universität Leipzig, Institut für Informatik, 04107 Leipzig, Germany; (HP, RGB) Friedrich-Schiller-Universität Jena, Institut für Spezielle Zoologie und Evolutionsbiologie mit Phyletischem Museum, 07743 Jena, Germany

Twisted-wing parasites are endopterygote insects with a highly derived morphology and life history and whose phylogenetic relationship to other endopterygote insects has proven notoriously difficult to resolve ("The Strepsiptera problem"). We sequenced the genome of a newly discovered twisted-wing parasite from the Tunisian Sahara. The new taxon belongs to the family Mengenillidae, whose species are plesiomorphic in many aspects. The aim of the genome sequencing was (1) to clarify the phylogenetic affinities of this intriguing and enigmatic group of insects, (2) to provide the most detailed description of a newly discovered insect species up to this date that includes both comprehensive morphological and molecular data, and (3) to assess problems associated with sequencing a veritable non-model organism. We compiled 5 million reads from a fragmented library, 2 million reads from a 3-Kb mate pair library, and 1 million reads from an EST library using 454-pyrosequencing technology. The genome was sequenced to an estimated $\geq 14\times$ coverage and the genomic sequences assembled into 100,940 scaffolds and contigs (N50 = 8.7 Kb). The large number of comparatively short scaffolds/contigs is possibly a consequence of the genome's unusually high AT content (72.9%), which seems to hamper sequence assembly. Polymorphism, introduced by pooling specimens collected in the field, may represent another confounding parameter. Using a combination of *ab initio* and evidence-based gene prediction, we annotated ~13,000 genes with high confidence. Roughly 4,500 of these proved to be orthologous or inparalogous among endopterygote and paraneopteran insects. After removing ambiguously aligned sites, the compiled dataset of orthologous and inparalogous genes consisted of 1.8 million codon sites. To minimize a confounding impact of inhomogeneous nucleotide or amino acid frequencies among sequences in the phylogenetic analysis, we analyzed RY-recorded second codon positions only. Maximum likelihood estimates overwhelmingly suggest a close phylogenetic relationship of twisted-wing parasites and beetles. Hadamard conjugation indicates no plausible conflict in the phylogenetic signal. We are currently analyzing genomic meta-characters, such as protein domain and gene content, gene order, presence/absence of specific miRNAs and snoRNAs, and the position of introns within genes, to obtain additional, independent phylogenetic estimates.

4

The *Heliconius melpomene* genome project: the complete genomic sequencing of the long-wing passion-flower butterfly.

*Walters, James R.; Heliconius Genome Consortium

Cambridge University, Department of Zoology, Downing St, Cambridge, CB23EJ, United Kingdom

Heliconius butterflies have been a model system for ecological and evolutionary research for over 150 years. They are known as a 'textbook' example of mimicry where polymorphic wing color patterns within species repeatedly converged between species. Ongoing quantitative, population, and developmental genomic investigations of speciation, wing-coloration, mate choice, and other adaptive phenotypes make *Heliconius* butterflies a prominent emerging model genomic system.

Here we report on our strategy and progress in generating a complete genomic sequence for *H. melpomene*. This sequencing effort reflects a collaboration between the broader *Heliconius* genetics research community and the Baylor Human Genome

Sequencing Center. The estimated 295 Mbp genome of *H. melpomene* has been sequenced entirely with 'next generation' technologies, primarily via 454 but with some Illumina for homopolymer correction.

The current assembly includes 270 Mb at ~35x coverage. The assembly includes 4097 scaffolds with N50 = 221 Kb and ~11,000 contigs with N50 = 50.8 Kb. Gene predictions using the MAKER pipeline are ongoing and suggest ~12,000 coding gene loci will be identified. RNA-seq generated from egg, pupa, brain, abdomen, and developing wings support the gene prediction process. Illumina resequencing of 5 different geographic races from Central and South America will elucidate genome-wide patterns of historical gene flow and differentiation.

5

An in depth analysis of the *Phlebotomus papatasi* transcriptome

*Abrudan, Jenica L.

Eck Institute of Global Health, Department of Biological Sciences, University of Notre Dame, IN 46556

Phlebotomine sand flies are important vectors for disease in both the Old and the New World, transmitting a variety of diseases, including leishmaniasis, sand fly fever and bartonellosis. The most devastating of these diseases is leishmaniasis with 350 million people at risk and approximately two million new cases each year. We performed a global gene discovery analysis of two sand fly vectors, *Phlebotomus papatasi* and *Lutzomyia longipalpis*. Expressed Sequence Tags (ESTs) from normalized cDNA libraries from each species were compared. The resulting high quality reads (*Ph. papatasi* - 37,487; *Lu. Longipalpis* - 27,928) were assembled into unique sequences (6,187 contigs and 10,993 singlets for *Ph. papatasi* and 6,049 contigs and 11,101 singlets for *Lu. longipalpis*) using CAP3. Of these sequences 25% for each species had no similarity to proteins available in NR or in UniProt databases, when searched for using BLAST (blastx, e-value: 10^{-5}). Functional annotation was performed using the blast2GO algorithm: 8,837 (50%) *Ph. papatasi* sequences were annotated in one of the 3 GO categories (Biological process, Molecular function or Cellular component). The highest subcategories were catabolic process (Biological Process; 15%), nucleotide binding (Molecular Function; 16%) and protein complex (Cellular Component; 21%). *Lu. longipalpis* exhibited a similar distribution of GO subcategories (catabolic process, 13%; nucleotide binding, 16%; protein complex 19%). An in depth analysis of *Ph. papatasi* genes of interest was performed resulting in identification of 1 novel trypsin, 11 chymotrypsin, 2 aminopeptidases, 5 carboxypeptidases, 4 chitinases, 2 galectins, 4 C-type lectins, genes implicated in digestion and immune response. This project is the initial step in the sand fly genome sequencing effort to generate full genome sequences of both *Ph. papatasi* and *Lu. longipalpis*.

6

A beetle of another color: progress toward lady beetle sequencing

*Allen, Margaret L.

USDA ARS Biological Control of Pests Research Unit, Stoneville, MS 38776

Lady beetles (Coleoptera: Coccinellidae) are a family of insect predators known for their voracious feeding on aphids, scale insects, and other pests. While lady beetles have been popular with growers, gardeners, and biological control scientists, they have had little visibility in the world of genetics. Generalist predators have been essentially ignored from a genetic standpoint. Insectivory is a rather important part of the insect world, and terribly important from an environmental standpoint, and certainly worthy of molecular genetic investigation. The recent completion of the aphid genome and the availability of the complete and well annotated genome of the red flour beetle indicate a sound basis for the initiation of an effort to sequence a generalist predator of aphids so that predator/prey genetic interactions can be examined. Based on this reasoning I propose that a lady beetle, *Coleomegilla maculata* is an overlooked potential insect model organism, and should be added to the list of organisms scheduled for rigorous genetic study and genomic sequencing. Lady beetles belong to the order Coleoptera which is genetically represented by many nucleotide, EST, or protein sequences, led by the complete genome of *Tribolium castaneum*. Therefore a genetic model in the family Coccinellidae would add substantially to the resources available to the systematics, phylogenetics, and evolution communities. Initial molecular genetic sequencing could focus on several broadly representative sets of expressed sequences from the developmental stages of the insect and adult males, virgin and gravid females, and possibly specific body parts or tissues.

7

Comparative Genomic and Functional Genetic Analysis of Vector Mosquito Development

Behura, Susanta K.; Haugen, Morgan; Clemons, Anthony; Flannery, Ellen; Sarro, Joseph; Tessier, Charles R.; Severson, David W.; Duman-Scheel, Molly.

(SKB, AC, EF, JS, DWS, and MDS) University of Notre Dame, Department of Biological Sciences and Eck Institute for Global Health, Notre Dame, IN, 46530; (MH, AC, CRT, DWS, and MDS) Indiana University School of Medicine, Department of Medical and Molecular Genetics, South Bend, IN, 46617.

Although vector mosquitoes transmit a number of the world's deadliest diseases, little is known about their developmental biology. Analysis of mosquito developmental genetics may reveal novel vector control strategies and advance the study of insect evolutionary development. We recently described methods for functional analysis of embryonic genes in *Aedes aegypti*, the yellow and dengue fever vector and an emerging model for vector mosquito development. Although it was anticipated that *Drosophila* developmental genes and their functions would be generally well conserved in mosquitoes, our analyses are revealing striking differences. For example, knockdown of the *frazzled* and *semaphorin-1a* axon guidance genes in *A. aegypti* embryos indicates that genetic regulation of nerve cord development has diverged in dipteran insects. We are also investigating developmental genetics of the *A. aegypti* salivary gland, a tissue that is critical for the spread of infection. Analysis of the expression patterns of a number of *D. melanogaster* salivary gland development gene orthologues in *A.*

aegypti suggests that regulation of the development of this tissue has also diverged between the two species. In light of these results, we are initiating comparative *D. melanogaster* vs. *A. aegypti* developmental transcriptome analyses. Our observations also led us to perform a large-scale comparative analysis of developmental genes in the *D. melanogaster*, *A. aegypti*, *Culex quinquefasciatus*, and *Anopheles gambiae* genomes. While the study was comprehensive, special emphasis centered on genes that 1) are components of developmental signaling pathways, 2) regulate fundamental developmental processes, 3) are critical for the development of tissues of vector importance, 4) function in developmental processes known to have diverged within insects, and 5) encode microRNAs (miRNAs) that regulate developmental transcripts in *Drosophila*. While many fruit fly developmental genes are conserved in the three vector mosquito species, lineage specific duplications and expansions, as well as the loss of several genes known to be critical for *Drosophila* development were found. Notably, gain and loss of several Wnt, FGF, and Notch signaling pathway components was observed. Furthermore, losses and gains of genes that have been implicated in a number of developmental processes, in particular germline, salivary gland, and head development, apoptosis, and segmentation were noted. Sequence analyses suggest that some developmental processes, including salivary gland development, head development, and egg diapause, may be under more stringent selection processes than others. Finally, analysis of predicted miRNA binding sites in fruit fly and mosquito developmental genes suggests that the repertoire of developmental genes targeted by miRNAs is species-specific. The results of these investigations provide insight into the evolution of developmental genes and processes in dipterans and other arthropods, will serve as a resource for those pursuing analysis of mosquito development, and will promote further design and refinement of functional analysis experiments.

8

Helitrons gain and transpose DNA in lepidopteran genomes

*Coates, Brad S.; Hellmich, Richard L; Grant, David M.; Abel, Craig A.
USDA-ARS, Corn Insects & Crop Genetics Research Unit, Ames, IA 50011

Transposable elements (TEs) affect the structure of genomes through their acquisition and transposition of novel DNA sequences. The 134 bp repetitive elements, *Lep1*, are conserved non-autonomous *Helitron* in lepidopteran genomes that have characteristic 5'-CT and 3'-CTAY termini, a 3'-terminal hairpin structure, a 5'- and 3'-subterminal inverted repeat (SIR), and integrations that occur between AT or TT nucleotides. The distribution of 5541 *Lep1*s within the *Bombyx mori* genome show a co-localization with genes and propagate (GTTT)_n microsatellites. *Lep1 Helitrons* have acquired and propagated sequences downstream of their 3'-CTAY termini that are 57 to 344 bp in length and have termini composed of a 3'-CTRR preceded by a 3'-hairpin structure and a region complementary to the 5'-SIR (3'-SIRb). Features of both the ancestral *Lep1 Helitron* and multiple acquired sequences indicate that secondary structures at the 3'-terminus may have a role in rolling circle replication or genome integration mechanisms, and are a prerequisite for novel end creation by *Helitron*-like TEs. The preferential integration of *Lep1 Helitrons* in proximity to gene coding regions results in the creation of genetic novelty that is shown to impact gene structure and function through the introduction of novel exon sequence (exon shuffling), and transcript splice and polyadenylation site mutations. This study provides valuable insight into the mode and implication of *Helitron*-induced mutations, and is the first to predict that secondary structures are formed between ancestral *Helitron* and novel end sequences. These findings are important in the understanding the creation and function of *Helitron* termini.

9

Challenges in sequencing the highly repetitive *Diabrotica virgifera virgifera* genome.

*Coates, Brad S.; Miller, Nicholas J.; Walden, Kimberly; French, B. Wade; Sappington, Thomas W.; Siegfried, Blair D.; Robertson, Hugh M.
(BSC, TWS) USDA-ARS, Corn Insects & Crop Genetics Research Unit, Ames, IA 50011
(NJM, BDS) University of Nebraska, Department of Entomology, Lincoln, NE, 68583
(BWF) USDA-ARS, North Central Agricultural Research Laboratory, Brookings, SD, 57006
(HMR, KW) University of Illinois at Urbana-Champaign, Urbana, IL 61801

The western corn rootworm, *Diabrotica virgifera virgifera*, is destructive to corn plants in North America and Europe where control remains challenging due to evolution of resistance traits that allow survival when exposed to chemical and transgenic toxins. Genome sequencing of an inbred non-diapause strain is soon to be underway, which will enhance the devising of novel control tactics for this insect (NIFA FY2010 Sustainable Bioenergy Grants; HM Robertson et al.). Herein we report an initial characterization of *D. v. virgifera* genome size, and genome composition through paired end sequencing of 1037 BAC inserts and full BAC insert sequencing. End sequencing generated 1.17 Mb of sequence data (~ 0.05-fold genome coverage), and showed that ~ 9.4 and 16.0% of the genome corresponds, respectively, to protein coding regions and transposon-derived protein coding genes. Full insert sequences further suggest that transposable element densities are high within the intergenic and intron regions of the genome, and have contributed to a large genome size of 2.56 Gb. A moderate density single nucleotide polymorphism (SNP)-based genetic linkage map and BAC sequence data are anticipated to be valuable resources for contig assembly and genome scaffolding.

10

RNA-seq identification of trans-regulatory factors that determine *Ostrinia nubilalis* Bt resistance traits

*Coates, Brad S.; Hellmich, Richard L; Siegfried, Blair D.
(BSC, RLH) USDA-ARS, Corn Insects & Crop Genetics Research Unit, Ames, IA, 50011
(BDS) University of Nebraska, Department of Entomology, Lincoln, NE, 68583

Crystalline (Cry) toxins from *Bacillus thuringiensis* (Bt) control insect feeding damage on crop plants via foliar applications or by expression within transgenic plants, but continued Bt use is threatened by the buildup of insect resistance traits. Aminopeptidase N (*apn*) gene family members encode midgut-expressed membrane-bound digestive enzymes in larval

Lepidoptera that also act as receptors for Cry toxin binding. In this study, we used individuals from an *Ostrinia nubilalis* colony that show an ~1400-fold increase in Cry1Ab toxin tolerance to model the inheritance of Bt resistance traits. Comparative real-time RT-PCR indicated that transcripts from the *apn1* locus are suppressed ~146.2±17.3-fold within the midgut of Cry1Ab resistant compared to susceptible larvae. Despite a perfect correlation between *apn1* transcript level and inheritance of Cry1Ab resistance, the *apn1* locus segregated independently of corresponding *apn1* transcription levels and suggests that modification of one or more *trans*-regulatory factors results in suppression of *O. nubilalis apn1* gene transcription. Quantification of midgut-derived transcription by RNA-seq has identified putative differentially-expressed genes that show constitutive linkage with a the segregating Cry1Ab resistance trait. RNA-seq is a method that allowed for high throughput and high read depth transcription data from a non-model organism.

11

De novo transcriptome sequencing in *Anopheles funestus* using Illumina RNA-seq technology.

*Crawford, Jacob E.; Guelbeogo, Wamdaogo M.; Sanou, Antoine; Traoré, Alphonse; Vernick, Kenneth D.; Sagnon, N'Fale; Lazzaro, Brian P.

(JEC, BPL) Department of Entomology, Cornell University, Ithaca, New York, United States of America; (WMG, SA, TA, NS) Centre National de Recherche et de Formation sur le Paludisme, Ouagadougou, Burkina Faso; (KDV) Department of Parasitology and Mycology, Centre National de la Recherche Scientifique Unit URA3012: Hosts, Vectors and Infectious Agents, Institut Pasteur, Paris, France; (KDV) Department of Microbiology, University of Minnesota, Saint Paul, Minnesota, United States of America

Anopheles funestus is one of the primary vectors of human malaria, which causes a million deaths each year in sub-Saharan Africa. Few scientific resources are available to facilitate studies of this mosquito species and relatively little is known about its basic biology and evolution, making development and implementation of novel disease control efforts more difficult. The *An. funestus* genome has not been sequenced, so in order to facilitate genome-scale experimental biology, we have sequenced the adult female transcriptome of *An. funestus* from a newly founded colony in Burkina Faso, West Africa, using the Illumina GAIIx next generation sequencing platform. We assembled short Illumina reads *de novo* using a novel approach involving iterative *de novo* assemblies and 'target-based' contig clustering. We then selected a conservative set of 15,527 contigs through comparisons to four Dipteran transcriptomes as well as multiple functional and conserved protein domain databases. Comparison to the *Anopheles gambiae* immune system identified 339 contigs as putative immune genes, thus identifying a large portion of the immune system that can form the basis for subsequent studies of this important malaria vector. We identified 5,434 1:1 orthologues between *An. funestus* and *An. gambiae* and found that among these 1:1 orthologues, the protein sequence of those with putative immune function were significantly more diverged than the transcriptome as a whole. Short read alignments to the contig set revealed almost 367,000 genetic polymorphisms segregating in the *An. funestus* colony and demonstrated the utility of the assembled transcriptome for use in RNA-seq based measurements of gene expression. We developed a pipeline that makes *de novo* transcriptome sequencing possible in virtually any organism at a very reasonable cost (\$6,300 in sequencing costs in our case). We anticipate that our approach could be used to develop genomic resources in a diversity of systems for which full genome sequence is currently unavailable. Our *An. funestus* contig set and analytical results provide a valuable resource for future studies in this non-model, but epidemiologically critical, vector insect.

12

Deep transcriptome insights into cave beetle eyes

*Friedrich, Markus; Chen, Rui; Daines, Bryce; Bao, Riyue; Caravas, Jason; Rai, Puneet K.

(FM, BR, CJ, RPK) Wayne State University, 5047 Gullen Mall, Detroit, MI 48202; (RC, BD) Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030

Cave species tend to exhibit a consistent suite of dramatic regressive and constructive trait modifications such as eye reduction and touch receptor increase. Because molecular studies of cave adaptation have largely concentrated on vertebrate models, it is not yet possible to probe for universalities in the genetic evolution of cave adaptation. We have begun to study the molecular evolution of the strongly cave adapted small carrion beetle *Ptomaphagus hirtus*. This flightless signature inhabitant of the world's largest known cave system (Mammoth cave, Kentucky) has been considered blind despite the presence of residual lens structures. Readdressing this question by deep sequencing of the *P. hirtus* adult head transcriptome, we discovered the expression all essential insect phototransduction genes and all core members of the circadian clock gene network. However, the absence of select structural photoreceptor and eye pigmentation gene transcripts suggests a cave adaptation-specific pattern of gene loss. We initiated sequencing of the *P. hirtus* genome to obtain a global view of cave adaptation associated gene inventory changes.

13

Functional genomics of the cricket *Gryllus bimaculatus*, a model system for regeneration and evolutionary developmental studies

*Mito, Taro; Nakamura, Taro; Watanabe, Takahito; Bando, Tetsuya; Noji, Sumihare
University of Tokushima, Department of Life Systems, Tokushima 770-8506, Japan

The two-spotted cricket, *Gryllus bimaculatus* De Geer (Orthoptera: Gryllidae), is one of the most populous crickets and inhabits the tropical and subtropical regions of Asia, Africa and Europe. The cricket can be bred easily in a laboratory and has been widely used to study insect physiology and neurobiology. Since RNA interference (RNAi) works effectively in this species, the elucidation of mechanisms of development and nymphal leg regeneration has been expedited through loss-of-function analyses of genes. In addition, a technique for a germline transformation has been established. Furthermore, we recently

succeeded targeted mutagenesis using custom-designed zinc finger nucleases (ZFNs) in this species. These techniques make this cricket species a very useful system for functional genomics.

To explore the cricket genome (genome size ca. 2 Gbp), we are conducting whole genome sequencing and transcriptome analysis of the cricket using an Illumina GALL sequencer. Our ongoing progress on development of functional analysis tools and genome sequencing will be presented.

14

First steps toward whitefly (*Bemisia tabaci*) proteome elucidation using the whitefly *Bemisia tabaci* AZ-B biotype transcriptome database.

Saripalli, C.; Breci, L.; He, R.; Cicero, J.; Nelson, W M.; Tsprailis, G.; Cameron, C T., Kramer, R., Crow, J A., May, G A., Kim, R.; Gang, D R.; Soderlund, C.; *Brown, J K.

(CSAR, JC, JKB) School of Plant Sciences, University of AZ, Tucson, Arizona (AZ) 85721; (LB) Department of Chemistry and Biochemistry, University of Arizona, Tucson, AZ 85721; (RH, DRG) Institute of Biological Chemistry, Washington State University, Pullman, WA 99164; (WMN, CSOD) BIO5 Institute, University of AZ Tucson, AZ 85721; (GT) Center for Toxicology, University of AZ, Tucson, AZ 85721; (CTC, RKRA, JAC, GAM, RKIM) National Center for Genome Resources, Santa Fe, NM 87505

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 (three 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 then used to validate MS/MS peptide 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 (15,045 contigs and 39,864 singletons), (ii) a set of translated ESTs from ~12 insect species selected from the available insect sequence databases, herein, 'Select Insect' (SI), 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+Insecta dbs as the search databases, 245 whole whitefly and 166 gut protein hits were identified. Protein hits (as GI numbers) from the three databases were imported into the UniProt batch retrieval system to facilitate identification of top hits to insect species, which provide GO Slim, KEGG, and Pfam descriptions. Using the combined SI+Insecta dbs, the 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, WFV4 db had the greatest number of hits to fruit fly, followed by many other species (others), and then yellow fever mosquito. The top protein hits were dominated by predicted (61) and putative (59) hits in the combined SI+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 with other homopteran sequences, namely, the pea aphid, presently the best represented genome for plant-feeding homopterans. The expected pattern was observed for SI+Ins and Insecta dbs annotations, but not for WFV4 db annotations. Further, only 17% of whitefly transcripts contained in WFV4 had a top hit to a UniProt accession. These latter results are probably due to the lack of a genome sequence for whitefly (and other closely related plant feeding homopterans), which if available would facilitate annotation of a larger number of the WFV4 db ESTs. Further the breadth of the WFV4 EST db is promising, currently containing ~775,000 ESTs: 15,045 contigs and 39,864 singletons, and housing transcripts that are expected to be 100% homologous with whitefly proteins. Although this undertaking provides the first partial whitefly *B. tabaci* proteome and an extensive EST library for the B biotype of *B. tabaci*, the inability to identify most if not all available spectra determined in the whitefly proteome underscores the need for functional genomics efforts for whitefly and other phyto-homopteran exemplars.

15

Small Non Coding RNAs In The Pea Aphid Genome

*Tagu, D.; Legeai, F.; Gallot, A.; Guernec, G.; Jaubert-Possamai, S.

(DT, FL, AG, SJP) INRA 1099 BiO3P, Rennes, France

(GG) INRA, UR1037 SCRIBE, F-35000 Rennes, France

Small non-coding RNAs (sncRNAs) have been abundantly described as strongly implicated in the post-transcriptional regulation of transcribed units in eukaryote genomes. Silencing of genes, pseudogenes or transposable elements can be operated in the germline or in the soma through microRNA (miRNAs), PIWI associated RNAs (piRNAs) or endogenous short interfering RNAs (siRNAs). All these sncRNAs are thus sculpting the whole functional transcriptome of an organism.

Aphids cause serious damage on crops, partly due to their amazing ability to adapt to environmental variations. This adaptive capacity is largely explained by their phenotypic plasticity that allows the production of distinct phenotypic morphs in response to environmental changes. This is illustrated by the switch of reproductive mode in response to seasonal changes: parthenogenetic females are produced in spring and summer whereas sexual females and males are produced in autumn.

Determination of aphid reproductive mode was shown to occur during embryonic development. In this work, we describe and compare the microRNAome of aphid developing embryos during sexual and asexual embryogeneses using next generation sequencing and combining three algorithms for microRNA identification: GR4500, mirDeep and mirTrap. MicroRNA expression profiles were analyzed all along a developmental kinetic of three points in order to identify microRNAs regulated during the determination of reproductive mode of embryos. For example, among the 615 microRNA candidates identified by mirdeep, 85 were differentially expressed during the developmental kinetic of the determination of aphid reproductive mode. In parallel, transcriptomic analyses were performed on the same biological material and identified 35 genes differentially regulated. Combination of microRNAome and transcriptome analyses will initiate the construction of a regulatory network between miRNA and mRNA that regulate reproductive phenotypic plasticity in the pea aphid.

16

Small RNAs of the soybean aphid, *Aphis glycines* Matsumura

*Vijayendran, Diveena; Liu, Sijun; Bonning, Bryony C.

Iowa State University, Department of Entomology, Ames, IA 50011

Small non-coding RNAs (17-30nt) are key players in many biological processes such as gene regulation, epigenetic modification, development and immunity against invading foreign organisms. We characterized the small RNA (sRNA) of the soybean aphid (*Aphis glycines* Matsumura), an invasive agricultural pest in the U.S. The soybean aphid small RNA was sequenced using the Illumina technology. The dataset was then analyzed for detection of known sRNAs using DSAP and miRanalyzer and potential novel micro RNAs (miRNA) were predicted using miRanalyzer. sRNAs were classified according to class, including miRNA, short-interfering RNA, small nuclear and nucleolar RNA. We identified a total of 116 known miRNAs representing 33% of the dataset. Additionally, over 200 potential novel miRNA have been predicted. The majority (84%) of the detected miRNAs have previously been identified in other arthropods, including 9 miRNAs that are unique to the pea aphid (*Acyrtosiphon pisum* Harris). Other non-coding RNA of bacterial and viral origin was also present in the dataset. A greater understanding of the role of sRNAs in soybean aphid biology could aid in the development of improved aphid management strategies.

17

Identification and Developmental Profiling of Conserved and Novel MicroRNAs in *Manduca sexta*

*Zhang, Xiufeng; Jagadeeswaran, Guru; Zheng, Yun; Sunkar, Ramanjulu; Jiang, Haobo

(XZ, HJ) Oklahoma State University, Department of Entomology and Plant Pathology, Stillwater, OK 74078; (GJ, RS)

Oklahoma State University, Department of Biochemistry and Molecular Biology, Stillwater, OK 74078; (YZ) Fudan University, Institute of Development Biology and Molecular Medicine, Shanghai, China

Due to its large body size and hemolymph volume, the tobacco hornworm *Manduca sexta* has long been used as a model for physiological and biochemical studies of insects. However, there is no information available in this species on post-transcriptional regulation of gene expression by microRNAs. We have constructed four small RNA libraries from embryos, fourth instar larvae, pupae and adults, obtained ~2 million reads of 18-30 nucleotides by Illumina sequencing, and predicted 202 conserved and 48 species-specific microRNAs. By searching the recently sequenced *M. sexta* genome, we identified precursors of 108 conserved microRNAs, 72 of which had mapped reads in at least one of the four libraries. There are twelve microRNA clusters, including both conserved and species-specific ones. Mse-miR-1 may have anti-sense microRNA, but no mapped reads of mse-miR-1-as was found in the libraries. miR-2 family has two loci, each encoding two microRNA variants; mse-miR-3389 has twelve loci and may have derived from repetitive elements. After normalization, we compared numbers of microRNA and microRNA-star reads in these libraries and observed abundance changes during development. Interestingly, mse-miR-281-star, mse-miR-31-star, and mse-miR-965-star are more abundant than mse-miR-281, mse-miR-31, and mse-miR-965, respectively. In summary, these results provided the first set of *M. sexta* microRNA sequences and insights for development-related functions of microRNAs.

18

Transcriptome variations among strains of the vector mosquito, *Aedes aegypti*

*Bonizzoni, Mariangela; Dunn, W. Augustine; Campbell, Corey L.; Olson, Ken E., Marinotti, Osvaldo; James, Anthony A.

(MB) University of California at Irvine, Program in Public Health, Irvine, CA 92697; (MB, WAD, OM, AAJ) University of California at Irvine, Department of Molecular Biology and Biochemistry, Irvine, CA 92697; (WAD) University of California at Irvine, Institute for Genomics and Bioinformatics, Irvine, CA 92697; (CCL) Colorado State University, Department of Biochemistry and Molecular Biology, Ft Collins, CO; (KEO) Colorado State University, Department of Microbiology, Immunology and Pathology, Fort Collins, CO, (AAJ) University of California at Irvine, Department of Microbiology and Molecular Genetics, Irvine, CA 92697.

Aedes aegypti (Diptera, Culicidae) is the principal vector of Dengue viruses (DENVs) and Yellow Fever Virus (YFV), and is responsible for Chikungunya virus (CV) epidemics in countries around the Indian Ocean and in Europe. This mosquito species exhibits great adaptability to diverse ecological conditions, is phenotypically polymorphic and shows variation in vectorial capacity for arboviruses. Indeed, population genetic approaches revealed a varying degree of genetic polymorphism worldwide. Sequencing of the *Ae. aegypti* genome showed a richness in transposable elements (TE), simple and tandem repeats and repetitive sequences that can contribute to genome plasticity, and more than 200 of the identified TE had at least one copy with an intact open reading frame. However, it is not known how much of this genetic polymorphism is functionally important. We used RNA-seq technology to compare the transcriptional profiles of three strains of *Ae. aegypti* (Rexville D-Puerto Rico [Rex-D], Chetumal [CTM], and Liverpool [LVP]), before a blood meal and shortly after blood feeding. Rex-D and CTM are a long-established laboratory and a recently-derived strain from Puerto Rico and the Yucatan Peninsula of Mexico,

respectively. A significant phenotypic distinction between the two strains is their differential susceptibility to Dengue virus type 2 (DENV2). Rex-D shows a slower and less intense DENV2 dissemination than CTM mosquitoes. The LVP strain originated from West Africa in the late 1930s and was used to derive the currently available *Ae. aegypti* genome sequence. We analyzed the dynamics of the *Ae. aegypti* transcriptome following a blood meal to gain insight into the complexity of transcriptional regulation and its effects on phenotype. We hope to link aspects of phenotypic plasticity with vectorial capacity and identify *cis* regulatory elements (CRE) that are responsible for coordinated gene regulation following a blood meal. CREs are essential component of genetic-based vector control strategies whereby mosquito populations are either replaced or suppressed by released genetic modified mosquitoes incapable of transmitting the pathogen or killed in a sex-specific manner, respectively. Significant differences were found in the magnitude and direction of accumulation of specific transcripts emphasizing a surprising amount of variation among the species.

19

Rates, Mechanisms and Functional Implications of Modular Protein Rearrangements during 400 Million Years of Arthropod Evolution

Moore, Andrew D.; Grath, Sonja*; Schüler, Andreas; Huylmans, Ann K.; Bornberg-Bauer, Erich
Westfälische Wilhelms-Universität Münster, Institut für Evolution und Biodiversität, Evolutionäre Bioinformatik, 48149 Münster, Germany

Modularity is a hallmark of molecular evolution. Whether considering gene regulation, the components of metabolic pathways or signaling cascades, the ability to reuse autonomous modules in different molecular contexts can expedite evolutionary innovation. On the protein level, modularity is mediated by protein domains, the functional, structural and evolutionary units of proteins. Modular domain rearrangements, can create vast diversity with seemingly few operations in turn allowing for swift changes to an organisms' functional repertoire. Here, we assess the mechanics and functional effects of modular rearrangements at a to date unprecedented resolution. We reconstruct ancestral proteomes and quantify the events that facilitate modular protein evolution across a well resolved tree of 20 arthropods and 2 outgroups. We illustrate the vast diversity in arrangements, *i.e.* unique combinations of protein domains, apparent in even closely related organisms, estimate arrangement turnover frequency and establish for the first time branch specific rates for fusion, fission, domain gain and terminal loss. Our results show that roughly 17 new arrangements arise per million years and that between 64% - 81% of these can be explained by simple, one-step modular rearrangement events. We find evidence that the frequency of fission and terminal deletion events increase over time, and that modular rearrangements strongly impact all levels of the cellular signaling apparatus and thus have strong adaptive potential. Novel arrangements our model fails to explain contain a significant amount of repeat domains occurring in complex patterns. Furthermore, gained arrangements that lack an explanation are significantly longer than explained gains, suggesting that such arrangements may result from multi-step events. In summary, this analysis provides a detailed quantification and integrated view of the mechanics and functional impact of modular protein evolution in one of the most diverse and successful taxonomic groups on earth. In a second step, we illustrate the means by which the adaptational value of novel domains in protein evolution can be assessed. We compare the domain contents of species across the Arthropod clade and estimate that every lineage has gained and lost thousands of domains since their last common ancestor. However, despite this large turnover, in total only 30 new detectable domain families have emerged. Emerging domains are more likely disordered and spread rapidly within their genomes. Furthermore, we find strong evidence that the emergence of novel domains is foremost associated with environmental adaptation such as abiotic stress response. The here illustrated approach for inferring the evolutionary potential of novel modules is easily applicable to the wealth of newly sequenced genomes, and may provide a valuable complement to the classical site-based analyses of adaptation.

20

COMPARATIVE ANALYSES OF SEVEN ANT GENOMES

*Helmkampf, Martin; on behalf of the Ant Genome Consortium (tAGC)
Arizona State University, School of Life Sciences, Tempe, AZ 85287

Ants are among the most abundant and ecologically successful animals on earth, and display a staggering array of diverse morphologies, ecological adaptations and behaviors. As emerging model species for social organization, phenotypic plasticity and chemical communication, deciphering ant genomes promises to reveal the genetic architecture underlying these hallmarks of their success. With the recent publication of seven ant genomes representing four major ant subfamilies and spanning divergence times between 10 and 100 million years, there is an unprecedented opportunity to study ant genome organization and content on a comparative level, significantly improving the accuracy of evolutionary inference over single-genome analyses. Here we compare the different strategies used in sequencing these species, and report overall similarities in genome size, repeat fraction and gene content across the seven ant species analyzed. Low synteny suggests, however, that substantial re-organization of the genomes has occurred during ant evolution. According to global gene homology analyses, ants possess a relatively small number of paralogous genes compared to non-apocritan insect genomes, but a considerable number of genes without sequence similarity to other arthropods. These genes are significantly enriched in GO terms related to chemical signaling and perception, and might represent unique adaptations to the lifestyle of individual or all ant species. Several gene families related to the social lifestyle are characterized by significant expansions along the ant lineages, as demonstrated by the desaturase and Olfactory Receptor gene families which are involved in the production and reception of chemical signals. This suggests that not only the co-option of conserved pathways, but also the emergence of new genes was essential for the evolution of some social traits in ants.

21

Genome Size Variation within and among Aphididae and other Arthropod Families.

*Johnston, J. Spencer; Medina, Raul F.; Tamborindeguy, Cecilia; Hanrahan, Shawn J.
Department of Entomology, Texas A&M University, College Station, TX 77843

Genome size is relatively constant within each Arthropod species, but varies by orders of magnitude between species. Neither a genotype nor a phenotype, genome size correlates with both via cell size and metabolic rate, and the relative number of genes, introns and mobile genetic elements. The constraints limiting intraspecific variation in genome size are poorly understood as are the evolutionary forces resulting in interspecific variation. The available data is limiting. There are 830 insect species currently on the Animal Genome Size Database. Here we report genome size on an additional 150 species from 15 orders and 54 families. Included are new and corrected values for economically important species of aphids and some quite surprising new first-reported values for *Phylloxera*. The reported genomes range 50 fold in size, from 154 Mb to 7752 Mb. Some general trends can be seen, Hymenoptera tend to have small genomes (here, 167 Mb to 543 Mb), while Orthoptera have large genomes (here, 1675 Mb to 7752 Mb). Holometabolous insects tend to have small genomes (154 Mb to 1500 Mb), while ametabolous insects tend to have large genomes (2000 Mb or greater). Yet closely related species can differ by an order of magnitude and added taxa can be exceptions to the above trends. The new values reported for Hemiptera here, extend the genome size range in this order (213.9 Mb to 7002 Mb) to approximately that observed across arthropods. With the availability of high throughput sequencing, genome size has taken on an important new dimension. The expected coverage, the quality of the final assembly, and the expected genomic architecture are all functions of genome size. Lacking genome size data, or relying upon unreplicated and weakly documented literature values can lead to uneconomical and unproductive choices for sequencing efforts. Genome size is an important metric that needs to be extended to comprise quality data from many more species, including species of economic, medical and phylogenetic importance.

22

Exploring Genomic Divergence in Geographically Diverse *Anopheles gambiae* M and S Populations using a SNP Genotyping Array

*Reidenbach, Kyanne R.; Neafsey, Dan E; Costantini, Carlo; Sagnon, N'fale; Lobo, Neil; Muskavitch, Marc A. T.; Besansky, Nora J.

(KRR, NJB, NL) University of Notre Dame, Eck Institute of Global Health, Department of Biological Sciences, Notre Dame, IN 46556; (DEN, MATM) Broad Institute, Cambridge, Massachusetts, MA 02142; (CC) Institut de Recherche pour le Développement, Unité de Recherche R016, Montpellier, France; (CC) Organisation de Coordination pour la Lutte contre les Endémies en Afrique Centrale, Yaoundé, Cameroon; (SN) Centre National de Recherche et Formation sur le Paludisme, Ouagadougou, Burkina Faso

Due to its anthropophilic nature and ecological versatility, *Anopheles gambiae* is one of the most deadly and widespread vectors of human malaria in sub-Saharan Africa. Incipient speciation within this mosquito has resulted in two molecular forms, M and S, which differ in larval habitat preference. A previously developed *An. gambiae* SNP genotyping array has been used to examine patterns of genomic variation in M and S populations from Mali. We extend this analysis to eight populations of *An. gambiae* M and S as well as two populations of sibling species, *An. arabiensis*, from four additional localities across Africa. DNA samples were pooled together and hybridized to the SNP genotyping array. Subsequent pairwise and multi-population analyses were carried out to assess patterns of intraform, interform, and interspecific divergence and population structure. Populations from western and central Africa, where M and S are sympatric, show a similar pattern of heterogeneity in areas surrounding all three centromeres, where elevated divergence is expected, as well as outside of these regions. The shared pattern of M-S divergence across diverse parts of Africa may reflect recent common ancestry of M and S form populations and/or similar local selection pressures acting across Africa. These data constitute the first high-resolution comparisons of genomic patterns of divergence among wild populations spanning most of Africa.

23

Genome Biology of Disease-Vector Mosquitoes.

Waterhouse, Robert M.

Department of Genetic Medicine and Development, University of Geneva Medical School, 1 rue Michel-Servet, 1211 Geneva, Switzerland; Swiss Institute of Bioinformatics, 1 rue Michel-Servet, 1211 Geneva, Switzerland.

A major human health concern of the new millennium centres on pathogens transmitted by blood-feeding mosquitoes, including malaria parasites, dengue and west nile viruses, and filarial worms. Sequencing the genome of the southern house mosquito, *Culex quinquefasciatus*, achieved the important objective of obtaining a reference genome from each of the three major genera of disease-vector mosquitoes. With the sequenced genomes of the malaria mosquito, *Anopheles gambiae*, and the dengue mosquito, *Aedes aegypti*, this triad of representative species offers new opportunities to explore the genetic and genomic diversity underlying mosquito biology.

Anophelinae and *Culicinae* mosquitoes diverged some 145-200 million years ago, while *Aedes* and *Culex* lineages are separated by about 52-54 million years. Differential transposable element activity in each lineage has contributed greatly to differences in genome sizes: from the 278Mbp *Anopheles* genome to the ~2-fold larger *Culex* genome, and the ~5-fold larger *Aedes* genome. The intermediate-size *Culex* genome encodes the largest gene repertoire with almost 19,000 protein-coding genes: ~22% more than *Aedes* and ~52% more than *Anopheles*. Expansions of C-type lectins, fibrinogen-related proteins, and serine protease inhibitors account for much of the increased number of *Culex* immune-related genes. The majority of mosquito genes have orthologues across Insecta, most of which are found as single-copy genes, many with functional characteristics studied in the fruit fly, *Drosophila melanogaster*. Genomic shuffling has eroded ancestral mosquito gene arrangements,

particularly in the *Aedes* genome, but genes found in conserved synteny blocks suggest possible functional constraints that have maintained their local gene neighbourhoods. Examining evolutionary rates of mosquito orthologues indicates a slower accumulation of amino acid substitutions in *Aedes* despite an apparently elevated rate of nucleotide substitutions among corresponding codons, suggesting that different constraints drive contrasting evolutionary patterns in protein versus DNA sequences. Studying the genome biology of these three major disease-vector mosquitoes highlights possible links between their genetic and genomic diversity and their eco-ethological traits, and sets the stage for the analysis of additional species from these three genera, responsible for the vast majority of human morbidity and mortality attributable to insect-transmitted pathogens.

Of ~500 Anophelines, only about two dozen are responsible for human malaria transmission. Their vectorial capacities, and other traits such as insecticide resistance and chemosensory abilities, derive from a genetic basis. The sequencing of multiple additional Anopheline genomes will build an extensive genomic data resource facilitating examination of the evolution of genetic determinants of vectorial capacity. Comprehensive, phylogenetically informed comparative genome analysis of multiple mosquito species with variable vectorial capacities will advance our understanding of the biology underlying pathogen transmission, thereby contributing to the development of novel methods for disease control.

ResearcherID: <http://www.researcherid.com/rid/A-1858-2010>

Acknowledgements: *Culex* genomics (P. Arensburg, K. Megy, *et al*) and pathogenomics (L. Bartholomay, M. Muskavitch, *et al*) working groups (<http://cegg.unige.ch/Culex>); *Anopheles* Genomes Cluster Committee (Nora Besansky *et al*); and members of the Computational Evolutionary Genomics Group (CEGG, University of Geneva) and the Laboratory of Immunogenomics (Imperial College London).

24

Molecular Evolution of Insect Chemosensory Gene Families

*Almeida, Francisca C.; Sanchez-Gracia, Alejandro; Campos, José L.; Rozas, Julio (FCA, ASC, JR) Universitat de Barcelona, Departament de Gènetica, Barcelona, Spain; (JLC) University of Edinburgh, Institute of Evolutionary Biology, Edinburgh, UK

Insect chemosensation is mediated by multigene families. Three of these families encode for receptors - Gustatory Receptors (GR, with 73 genes in *D. melanogaster*), Odorant Receptors (OR, 62 genes), and Ionotropic Receptors (IR, 66 genes) and two encode for binding proteins - Odorant Binding Proteins (OBP, 52 genes) and Chemosensory Proteins (CSP, 4 genes). Previous studies have shown that these families evolve according to the birth-and-death model and evidence of reduced evolutionary constraints in subfamilies with duplicated genes. These studies, however, have used different approaches to estimate gene gain and loss rates, which are also strongly dependent on how ortholog groups are determined. Here we readdress the evolution of the *Drosophila* chemosensory genes in a comparative framework by using the same ortholog group criteria and estimate methods across 3 gene families (OR, IR, and OBP) in 12 *Drosophila* species. First, for each gene family, we delimited ortholog groups using a combination of similarity clustering (using blast E-values) and phylogenetic (gene trees) methods. We then obtained estimates of birth and death rates using three different methods: 1) manual reconstruction of gene gain and loss by comparing gene trees (per orthologous group) and species trees, 2) using the program CAFE, which assumes equal birth and death rates, 3) and using the program BadiRate, which estimates birth and death rates independently by both parsimony and likelihood methods. The comparison among approaches pointed to a slight tendency of underestimation by the automated methods, but only when pseudogenes are accounted for in the manual estimates. On the other side, the direct use of gene number per species without a previous determination of ortholog groups with automated methods strongly bias the results of rate estimates. The most dynamic family was the OR, followed by the IR (specially of the divergent group), and OBP. Independent estimates allowed us to detect differences between birth and death rates, suggesting expansions of the OR and OBP repertoires with about twice as high birth rates as compared to death rates. For the IR family, our estimates suggest a slight contraction of the gene repertoire. Using BadiRate, we investigated whether rates of gene duplication and loss were affected by diet specialization and endemism. An elevated death rate was observed in all chemosensory families in specialist species, but not endemic ones. This difference was much higher when *D. sechellia*, a strict specialist was analyzed alone, pointing to increased gene loss as the primary cause for the reduction of the chemosensory genes in this species. To investigate the evolutionary forces involved in the maintenance of duplicated chemosensory genes, we used the branch model in the program PAML to examine whether branches following duplications had different evolutionary constraints as compared to those of other branches of the gene trees. In 31 of 33 ortholog groups analyzed d_N/d_S (the ratio between the non-synonymous substitution rate and the synonymous rate) was higher in branches following duplications. In 22 cases, branch differences in d_N/d_S were statistically significant, all of which had higher d_N/d_S in post-duplication branches. This suggests faster evolutionary rates in duplicated genes either by relaxation of selective constraints or by positive selection for amino acid change. Although most d_N/d_S of duplicated genes was below 1 (pointing to a predominance of purifying selection), our tests cannot rule out that positive selection may have had a role in the evolution of these genes.

25

Conserved insect sequence similar to the BTB domain of *NACC1* oncogene.

*Brown, Thomas M.

Genectar Com LLC, Whitefish, MT 59937. www.Genectar.com

Human *NACC1* oncogene encodes a protein which is over expressed in certain carcinomas. It is also important in stem cell self renewal. The BTB domain in the N-terminal 110 amino acids is recognized among those insects, including *Aedes aegypti* and *Tribolium castaneum*, represented in BLAST with arthropod genomes of NCBI. Searching public databases, candidate *nacc1*-like sequences were found in other insects and assembled from short contigs. These newly assembled insect sequences will be presented as aligned to *NACC1* and to inferred insect *nacc1* sequences.

26

Genomic organization of the glutathione S-transferase family in insects

Clardy K.A.; *Friedman R.

University of South Carolina, Columbia, SC 29208

Cytosolic glutathione S-transferases (GSTs) are a large and diverse gene family in insects. They are classified into six subclasses where Sigma, Omega, Zeta, and Theta have representatives across Metazoa. Of the remaining two subclasses, the Epsilon subclass has an early origin along the endopterogote lineage while the Delta subclass has an earlier origin in Insecta. Several unclassified GSTs were assigned to a subclass by phylogenetic and genomic evidence. Moreover, the recently duplicated GSTs frequently cluster by genomic position. The pattern of these gene expansions is largely explained by the number of protein-coding genes in the genome.

27

Seeking novel *cis*-regulatory modules with cross-species application for driving transgenes in the midguts of mosquitoes through comparative, midgut-specific transcriptomic analysis of *Anopheles gambiae*, *Aedes aegypti* and *Culex quinquefasciatus* mosquitoes.

*Dunn, W. Augustine; Marinotti, Osvaldo; James, Anthony A.

(WAD, OM, AAJ) Department of Molecular Biology and Biochemistry, University of California, Irvine, California, USA; (WAD) Institute for Genomics and Bioinformatics, University of California, Irvine, California, USA; (AAJ) Department of Microbiology and Molecular Genetics, University of California, California, Irvine, USA

Empirical definition of active *cis*-regulatory elements (CRE) through the classical “promoter bashing” method is difficult in mosquitoes due to the time and effort that is currently required to produce transgenic mosquito strains. Bioinformatic methods combined with existing biological knowledge and quality mRNA abundance data should allow the inference of biologically active CRE combinations, *cis*-regulatory modules (CRM), without the need to build a multitude of transgenic mosquito strains.

The ecdysone (20E) response cascade is conserved throughout insects and has been shown to drive changes in mRNA abundance following the ingestion of a quality bloodmeal by female mosquitoes. This supports the hypothesis that it may be possible to deduce a common regulatory module able to drive transgene expression across multiple mosquito species. 20E has multiple early-response factors that have been described previously including the ecdysone receptor (EcR), its binding partner ultraspiracle (USP), and the 20E inducible gene E74. Multiple isoforms exist for these response factors and work by other laboratories have shown that levels of these isoforms vary in a time- and tissue-specific fashion in response to pulses of 20E following a bloodmeal. This allows one hormone to simultaneously regulate diverse cellular responses.

Tissue-specific, time-course RNA-seq data with high temporal resolution (2 hours) from each species will be used to compare 20E early-response factor isoform mRNA expression levels across evolutionarily distant species. Additionally the expression data will allow for a combined comparative-genomics and expression-profile based CRE/CRM discovery strategy to search for putative CRMs able to drive early-midgut transgenes in multiple mosquito species.

28

Comparative Analysis of Aphid Salivary Transcriptomes

*Marimuthu, M.; *Oh, J.; Cilia, M.; Reeck, G; Reese, J.; Smith, C.M.; White F.F.

(JOH, FFW) Department of Plant Pathology, Kansas State University Manhattan, KS 66506; (MM, JR, CMS) Department of Entomology, Kansas State University Manhattan, KS 66506; (MC) Functional and Comparative Proteomics Center, Robert W. Holley Center for Agriculture and Health, Ithaca, NY 14850; (GR) Department of Biochemistry, Kansas State University Manhattan, KS 66506

Aphid species exhibit host specialization and success of aphids to utilize host plant nutrients is highly dependent on salivary gland secretions. Though plant responses to aphid feeding are well-studied, the components of aphid salivary fluids are not well characterized. We hypothesized that salivary proteins vary between different aphids species as adaptations to the respective host species. We have selected three aphid species for transcriptional analysis of their respective salivary gland, including the reference pea aphid (*Acyrtosiphon pisum*), the bird cherry oat aphid, (*Rhopalosiphum padi*) and the greenbug (*Schizaphis graminum*). RNA was extracted from salivary glands of species and sequenced by high through-put 454 sequencing. Comparison of the transcriptomes and additional proteomic evidence for secreted salivary gland protein profiles will be presented and discussed.

29

From transcriptome to function: pro-apoptotic cell death genes from the Caribbean fruit fly, *Anastrepha suspensa*

*Schetelig, Marc F.; Nirmala, Xavier; Handler, Alfred M.

(MFS, XN, AMH) USDA/ARS, Center for Medical, Agricultural and Veterinary Entomology, 1700 SW 23rd Drive, Gainesville, FL 32608, USA; (XN) Department of Entomology and Nematology, University of Florida, Gainesville, FL 32611, USA.

Pro-apoptotic proteins from the *reaper* (*rpr*), *hid*, *grim* (RHG) family are primary regulators of programmed cell death in *Drosophila* due to their antagonistic effect on inhibitor of apoptosis (IAP) proteins, thereby releasing IAP inhibition of caspases that effect apoptosis. Despite the importance of apoptosis to eukaryotic development and the use of cell death genes for controlling insect pest populations, homologs to the *Drosophila* pro-apoptotic genes have been rarely identified. This has been due, primarily, to their high level of divergence throughout eukaryotes with most identifications requiring customized algorithms

to detect common IAP-binding motifs (IBM). Indeed, in 24 non-Drosophilid insect genomes *hid* and *rpr* cell death genes could not be identified by simple BLAST searches. However, using conserved domains from RHG genes identified in the 12 *Drosophila* species genome project, a degenerate PCR approach allowed the first identification of *reaper* and *hid* orthologs from the embryonic transcriptome of a tephritid fruit fly, *Anastrepha suspensa*. The newly isolated *As-hid* is the first identified non-drosophilid homolog of *hid*, and *As-rpr* is the second non-drosophilid *rpr* homolog. Functional verification of both genes as cell death effectors was demonstrated by cell death assays in *A. suspensa* (AsE01) and *D. melanogaster* (S2) embryonic cell cultures. In common with their *Drosophila* cognates, *As-hid* and *As-rpr* negatively regulated the *Drosophila* inhibitor of apoptosis (DIAP1) gene to promote apoptosis. By comparing the newly isolated genes to the *Drosophila hid*-variant *hid^{Ala5}*, which has been used as an efficient lethal effector in a tetracycline-suppressible lethality system in both *Drosophila* and the Mediterranean fruit fly, we could not only show that *As-hid* is more effective than *Dm-hid^{Ala5}*, but that *As-hid* and *As-rpr* act together co-operatively and, notably, are more effective than either gene alone. This suggests that both genes could be used to design improved transgenic lethality systems, making transgenic strains for insect population control more effective and with less ecological risk due to lethal gene redundancy.

30

Phylogeny and Classification of Lepidopteran Lipases

*Tan-Kristanto, Ariadne; Heckel, David G.

Max Planck Institute of Chemical Ecology, Entomology Department, Jena, D-07745, Germany

In comparison to their mammalian counterparts, insect lipases and their genes have been little studied. However, the genome sequencing of several insect species has facilitated comparisons between enzymes of various insect orders.

Early analysis indicated that lipases of one insect order often lack homologues from a separate order. Here, the focus is narrowed to lipases of lepidopterans. This was possible not only via the thorough sequencing of the silkworm (*Bombyx mori*) genome, but also through sequencing and/or EST projects involving other species. These include the cotton bollworm (*Helicoverpa armigera*), the fall armyworm (*Spodoptera frugiperda*) and the European corn borer (*Ostrinia nubilalis*).

A number of putative phospholipase and triacylglycerol lipase families were found, two of which appear to be largely expressed in the larval gut. One family was then further divided into subfamilies.

Families/subfamilies of apparently inactive lipases have previously been observed. However, examination of the sequences revealed other family-specific variances likely to affect the structure and/or function of the lipases. One example is the movement of a cysteine residue from its usual position at one end of the putative lid to the lid's centre. Another is the absence of a cysteine pair that is conserved in all other lepidopteran phospholipases.

It was considered desirable to classify lepidopteran lipases in a similar manner to members of the cytochrome P450 superfamily. An attempt to do so through amino acid identity was unsuccessful, but apparent structural variances allowed members of certain families and subfamilies to be classified. More thorough categorization may be achievable through the sequence motifs surrounding the catalytic triad. The emergence of additional lepidopteran lipase sequences will assist in refining and correcting the current classification systems.

31

“ArthropodaCyc”: a BioCyc database powered by CycADS to study and compare the metabolism of arthropods.

Baa-Puyoulet, Patrice; Vellozo, Augusto F.; Huerta-Cepas, Jaime; Febvay, Gérard; Gabaldon, Toni; Sagot, Marie-France; Charles, Hubert and Colella, Stefano.

(PBP, GF, HC, SC) UMR203 BF21, Biologie Fonctionnelle Insectes et Interactions, INRA, INSA-Lyon, Université de Lyon, Villeurbanne, France; (AFV, MFS) Université Lyon 1, CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, Villeurbanne, France; (PBP, AFV, GF, MFS, HC, SC) BAMBOO, INRIA Rhône-Alpes, France; (JHC, TG) Center for Genomic Regulation, Barcelona, Spain.

The genome sequence for several arthropods is available and more genomes will be sequenced in the near future. The availability of multiple species genome sequences open the way to comparative studies to better understand different aspects of arthropods' biology. Such comparative studies rely heavily on the quality of genome annotation. In particular, in order to use systems biology approaches to study metabolism, genomic data have to be collected from various sources and updated regularly. Furthermore, metabolic databases can be also very powerful for researchers looking for combined and well-organized information about the metabolism of a newly sequenced organism.

During the genome annotation for the pea aphid (*Acyrtosiphon pisum*) we developed CycADS¹ (Cyc Annotation Database System), an automated annotation management system that allows the seamless integration of the latest sequence information into metabolic networks reconstruction. Data from GenBank and/or organism specific databases and obtained using different annotation tools (such as KAAS, PRIAM, Blast2GO, PhylomeDB) are collected into the CycADS database and later extracted to generate a complete input file to build and/or update BioCyc databases using the 'Pathway tools' software. We have used CycADS to generate Cyc metabolism reconstructions for several arthropods for which the genome sequence was available and we collected them in the “ArthropodaCyc” database². At present “ArthropodaCyc” includes metabolic reconstructions for 11 arthropods: *Acyrtosiphon pisum*, *Aedes aegypti*, *Anopheles gambiae*, *Apis mellifera*, *Culex quinquefasciatus*, *Daphnia pulex*, *Ixodes scapularis*, *Nasonia vitripennis*, *Pediculus humanus corporis*, *Tribolium castaneum* and *Drosophila melanogaster* (both the CycADS version and the FlyCyc database manually curated by the FlyBase team are available). The “ArthropodaCyc” database allows researchers to browse their model organism metabolism and to perform

comparative analyses using the BioCyc comparative tools. Thanks to the CycADS software, we included, in each database, information on annotation source and links to genomics databases (including AphidBase, BeetleBase, VectorBase, Hymenoptera Genome Database, FlyBase and wFleaBase). Future plans include adding other sequenced genomes. We are also open to collaborations with arthropod genome sequencing projects to help the annotation of metabolism.

¹ Vellozo A, Véron A, Baa-Puyoulet P, Huerta-Cepas J, Cottret L, Febvay G, Calevro F, Rahbé Y, Douglas AE, Gabaldon T, Sagot M-F, Charles H, Colella S. CycADS: an annotation database system to ease the development and update of BioCyc databases. Database, doi: 10.1093/database/bar008, 2011. [<http://cycadsys.org/>]

² <http://arthropodacyc.cycadsys.org>

32

Evolutionary Analysis As The Basis For Interpreting, Comparing, And Presenting Genomes: The GATOR And PHRINGE System

*Boore, Jeffrey L.; Fuerstenberg, Susan I.

(JLB) Genome Project Solutions, Inc., 1024 Promenade St., Hercules, CA 94547 USA; University of California Berkeley, 3060 Valley Life Sciences Building, Berkeley, CA 94720 USA; (SIF) Genome Project Solutions, Inc., 1024 Promenade St., Hercules, CA 94547 USA

We anticipate a great acceleration in whole genome sequencing over the next few years. Current tools for interpreting, comparing, and presenting these data cannot handle the expected pace, lack integration, require extensive IT support and computational expertise, and do too little to facilitate biological discovery. In particular, the standard “browser” format is anachronistic, with the genome assembly, rather than the biological information, being the organizing principle. It requires great manual effort to identify any particular gene or biochemical pathway. Fortunately, two new developments are enabling a better approach. First, next generation sequencing technology allows very deep sequence coverage of the set of expressed genes. For example, 200-fold mean sequence coverage can be obtained on a typical transcriptome for under \$4,000. This means that genes can be modeled with much greater accuracy, so even the early stages of analysis can focus on biological discovery instead of manual gene curation. Second, we have developed an effective tool (“PHRINGE”, for Phylogenetic Resources for Interpreting Genomes) for assigning orthologous and paralogous relationships among genes based on phylogenetic analysis of complete gene sets. In the absence of biochemical characterization, the best inference of gene function is by inferring that orthologous genes retain the same function. This is incorporated into the “GATOR” (Genome Analysis Tools and Online Resources) system under development, a “gene-centric”, user-friendly, streamlined approach to genome interpretation, comparison, and presentation. The entry point is the gene catalog itself, sortable by many categories, including domain content, intracellular location, SNP content, biochemical pathway, protein characteristics, number of members in any gene family, and many others. Users can view evolutionary trees, gene colinearity maps, and links to protein structures for all genes in multiple sequenced genomes.

33

Will Your Favorite Genome Be Sequenced?

*Fuerstenberg, Susan I.; Boore, Jeffrey L.

(SIF) Genome Project Solutions, Inc., 1024 Promenade St., Hercules, CA 94547 USA; (JLB) Genome Project Solutions, Inc., 1024 Promenade St., Hercules, CA 94547 USA; University of California Berkeley, 3060 Valley Life Sciences Building, Berkeley, CA 94720 USA

The past decade has seen a fantastic acceleration in our ability to sequence complete genomes. We anticipate that this exponential increase will continue for at least the next several years. Will your favorite genome be sequenced? We review the several technologies that are state-of-the-art for genome sequencing, describe their capabilities, provide an interpretation of the pros and cons of each and argue for the most appropriate mix of techniques for whole genome sequencing. We describe the current potential for funding such a large project and the options for follow-on analyses that will maximize biological discovery and the utility of the work to the broader scientific community.

34

VectorBase: biocuration of the genomes of invertebrate vectors of human pathogens

*Lawson, D.; Besansky, N.J.; Birney, E.; Christophides, G.K.; Collins, F.H.; Emrich, S.J.; Gelbart, W.M.; Kafatos, F.; Kersey, P.J.; Louis, C.; Madey, G.; Werner-Washburne, M.; VectorBase.

(DL,EB,PJK) European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SD, UK; (NJB, FHC, SJE, GM) Center for Global Health and Infectious Diseases, Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46656-0369, USA; (GKC,FK) Department of Cell and Molecular Biology, Imperial College London, South Kensington Campus, London SW7 2AZ, UK; (WMG) The Biological Laboratories, 16 Divinity Avenue, Harvard University, Cambridge MA 02138, USA; (CL) Institute of Molecular Biology and Biotechnology, FORTH, Vassilika Vouton, PO BOX 1385, Heraklion, Crete, Greece; (MWW) Department of Biology, University of New Mexico, Albuquerque, NM 87131, USA.

VectorBase (<http://www.vectorbase.org>) is an NIH-NIAID funded Bioinformatics Resource Center (BRC) relating to the curation of the genomes of invertebrate vectors of human pathogens. VectorBase works with the Microbial Sequencing Centers (MSCs) on the annotation of new sequencing targets (recent examples include *Culex quinquefasciatus* and *Ixodes scapularis*) as well as other sequencers working on vector genomes (such as *Pediculus humanus* and *Rhodnius prolixus*).

VectorBase is involved with all phases of genome annotation from gene prediction, functional analysis of the predicted protein-coding genes and comparative analyses of the predicted proteomes. We work with the sequencers in the first-pass annotation for release and publication and then assume responsibility for the long-term curation of the genome annotations including

submission to GenBank and the public sequence repositories. Maintaining extensive cross-references to other databases such as GenBank/EMBL/DDBJ and UniProt is an integral part of VectorBase activities. VectorBase integrates other data types (such as gene expression studies, proteomics, RNA-Seq transcriptomics and population genomics) within the context of the genome.

In the past year VectorBase has been trialling pre-sites allowing early access to the genome assembly and computed similarities prior to the release of the reference gene set. Pre-sites have been released for the M & S molecular forms of *Anopheles gambiae*, *Rhodnius prolixus* and *Glossina morsitans* (scheduled for a June release).

35

Ensembl Genomes: extending Ensembl across arthropod taxonomy

*Lawson, D.; Ensembl Genomes.

European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SD

Ensembl Genomes (<http://www.ensemblgenomes.org>) is a portal offering integrated access to genome scale data from non-vertebrates species of scientific interest. Developed using the Ensembl genome annotation and visualization platform Ensembl Genomes consists of five sub-portals (for bacteria, protists, fungi, plants and invertebrate metazoa). Ensembl Genomes provides a common set of user interfaces (which include a genome browser, FTP, BLAST search, a query optimised data warehouse, programmatic access and a Perl API) for each species.

Data types incorporated include annotation of protein-coding and non-coding genes, cross-references to external resources, high-throughput experimental data (e.g. data from large scale gene expression studies and population genomics studies visualized in their genomic context). Comparative analysis, both within metazoa and across the wider taxonomy has been performed to generate gene trees and sequence alignments.

Ensembl Metazoa supports databases that have been built in close collaboration with the scientific community, which we consider as essential for maintaining the accuracy and usefulness of the resource. We work closely with the VectorBase project (arthropod vectors of human pathogens including *Anopheles gambiae*) and FlyBase project (12 Drosophilid genomes including *Drosophila melanogaster*) to reflect the latest annotations for these genomes. Recent releases have included the *Acyrtosiphon pisum*, *Apis mellifera* and *Daphnia pulex* genomes as well as key non-arthropod genomes.

35A

The Hymenoptera Genome Database

*Munoz-Torres, Monica C., Reese, Justin T., Childers, Christopher P., Bennett, Anna K., Sundaram, Jaideep P., Hagen, Darren E. Vile, Donald C., Elisk, Christine G.

Georgetown University, Department of Biology, Washington, DC 20057. USA.

One hundred and fifteen thousand described species in the order Hymenoptera account for approximately 10% of the species diversity on Earth, and they directly affect human health and agriculture acting as pollinators, bioindicators, parasitoids and agents of biological control. The Hymenoptera Genome Database (HGD) is an informatics resource supporting genomics of insect species in this group. Our relational database implements open-source software and components that provide access to curated data contributed by an extensive, active research community. HGD contains genomic data of honey bee *Apis mellifera*, the parasitoid wasp *Nasonia vitripennis* and a portal to the recently released genomes of six species of ants. Bringing these species together across 200 MY in the phylogeny of Hymenoptera allows to leverage genetic data, genome sequences and gene expression data, as well as the biological knowledge of related model organisms. The availability of resources across the order greatly facilitates comparative genomics and will contribute to advance our understanding of processes such as foraging, reproductive division of labor due to unique genetic caste determination systems, the evolution of social behavior, and other aspects of the biology of agriculturally important Hymenoptera species. HGD has orchestrated research contributions from an extensive community of nearly 80 institutions from 14 countries, constituting what is perhaps one of the largest dispersed manual annotation efforts reported. Community annotation efforts are made possible thanks to a remote connection to a Chado database by Apollo Genome Annotation client software. Curated data at HGD includes predicted and annotated gene sets supported with evidence tracks such as ESTs/cDNAs, small RNA sequences and GC composition domains. Data at HGD can be queried using genome browsers and / or BLAST/PSI-BLAST servers, and it may also be downloaded to perform local searches. We encourage the public to access and contribute data to HGD at <http://HymenopteraGenome.org>.

36

Arthropod genome support at NCBI

*Murphy, Terence; Church, Deanna; Kitts, Paul; Maglott, Donna; Tatusova, Tatiana; Pruitt, Kim; and the NCBI Genome Annotation Team

NCBI/NLM/NIH/DHHS, 45 Center Drive, MSC 6510 Building 45, Bethesda, MD 20892-6510

Recent years have seen a tremendous surge in arthropod genomic information, with 39 sequenced genomes in the public databases and dozens more anticipated in the next few years. The National Center for Biotechnology Information (NCBI) provides genomic resources to support arthropod research at multiple levels including finding information about genomes, genes, sequences, expression, variation, and publications. These resources are extensively cross-linked and facilitate navigation across a broad range of biological information.

The great taxonomic diversity of arthropod genomes provides considerable challenges for genome annotation. The NCBI has a well-developed gene annotation pipeline that combines evidence-based prediction with *ab initio* modeling to predict protein-coding genes with varying degrees of alignment support. We have also developed methods to predict likely protein-coding loci that contain insertions or deletions in the assembly sequence, which may occur at higher frequency in future assemblies based on pyrosequencing (aka 454) technology. The NCBI pipeline has been used to help annotate a variety of taxonomically diverse species including 17 arthropods, and has recently been retooled for higher throughput in anticipation of greater demands from future sequencing projects. Details of recent work on the *Apis mellifera* (honey bee) and *Acyrtosiphon pisum* (pea aphid) genome annotations using the latest assemblies and short read RNAseq evidence will be presented.

NCBI provides a variety of resources to help arthropod researchers efficiently utilize the wealth of available data. The Reference Sequences (RefSeq) collection integrates transcript, protein, and genomic sequences for over 12,000 species, and is searchable by query or BLAST. Gene models for some organisms, such as *Drosophila melanogaster* and *Anopheles gambiae*, are based on annotation data provided by model organism databases such as FlyBase and VectorBase, whereas gene models for other organisms are based on data submitted to GenBank and models calculated by NCBI's gene prediction pipeline. NCBI's Gene database provides gene-oriented information including RefSeq and GenBank accession numbers, map information, nomenclature, publications, and external links including links to model organism databases for each locus. It also supports GeneRIF (Gene References into Function), which allows users to contribute functional annotation in Gene.

Additional NCBI resources include a genome browser (Map Viewer); organism-specific BLAST pages; pre-computed BLASTp results (Blink); computed homologous clusters (HomoloGene); clusters of related transcripts (UniGene); and the BioProjects database. Further information is available at the NCBI web site (<http://www.ncbi.nlm.nih.gov/>), or through organism-specific Genome Resources web pages (<http://www.ncbi.nlm.nih.gov/Genomes/>). The NCBI is interested in making these resources highly useful to the arthropod research community by including additional data, providing links to other resources, and more. In this regard, they have designated an "Arthropod Genome Champion" to act as a contact within NCBI. Please feel free to contact us at murphyte@ncbi.nlm.nih.gov or info@ncbi.nlm.nih.gov.

37

A community-centric bioinformatic infrastructure for new genome consortia

*Papanicolaou, Alexie; Brockhouse, Charles; Campbell, M; East, Peter D; French-Constant, Richard H; Gordon, Karl HJ; Karatolos, Nikos; McMillan, W. Owen; Ott, Michael; Terenius, Olle; the Lepidoptera RNAi International Working Group; Edwards, Owain R.

(AP, PC, PE, GK,MO) CSIRO Ecosystem Sciences, Black Mountain Laboratories, Clunies Ross St., Canberra, ACT 2601, Australia; (CB) Department of Biology, Creighton University, 2500 California Plaza, Omaha, NE 68178-0103, USA; (RFC) School of Biological Sciences, Centre for Ecology and Conservation, University of Exeter, Penryn, TR10 9EZ, UK; (NK) Rothamsted Research, Harpenden, Hertfordshire, AL5 2JQ, UK; (WOM) Smithsonian Tropical Research Institute, Panamá, República de Panamá; (TO) Department of Ecology, Swedish University of Agricultural Sciences, Box 7044, 750 07 Uppsala, Sweden; (ORE) CSIRO Ecosystem Sciences, CSIRO Centre for Environment and Life Sciences, Underwood Ave, Floreat, WA 6014, Australia

Due to evolving molecular and informatics technologies, modern sequencing projects have radically different characteristics to what most biologists were accustomed to during the capillary-based sequencing era. As evidenced by the recently published genomes of multiple social insects, the current paradigm is one of smaller consortia building -omic resources, for one or multiple species, on relatively limited budgets, without the extensive support previously available through the white paper process. Nonetheless, Next Generation Sequencing (NGS) has removed the sequencing bottleneck, making biology extraordinarily data-rich, but has produced a bioinformatics one. As we are currently core members of multiple insect genome consortia, we understand that a lack of centralized genome databases is a serious obstacle to delivering high quality genome projects. In the past, these were provided by the white paper process. The deployment of an analysis infrastructure is an immediate need for each project, but in order to be deployable, such solutions must adhere to community standards and not be project-specific. Further, they must serve the wider community and not be limited to whole genome projects. We present, thus, novel, robust and species-neutral bioinformatic solutions for analysis and dissemination of -omic data. First, we build standardized reference transcriptomes from Sanger, 454 and RNA-seq. We accomplish this using *est2assembly* (<http://gmod.org/est2assembly>) and the Trinity RNA-Seq software developed at the Broad (<http://trinityrnaseq.sourceforge.net>) which we have HPC-optimized for a high-throughput environment. Subsequently, we build online communities using the GMOD Drupal Bioinformatic Server Framework (http://gmod.org/gmod_dbsf). Finally, we store all the data in a Chado database warehouse and visualize them using Genes4all (<http://drupal.org/project/genes4all>). As proof of concept, we deployed the InsectaCentral database (<http://insectacentral.org>), a resource to support the entire insect genomics community. Our resources can support even nascent genome project as they begin with annotated *de-novo* transcriptome assemblies and provide a rich resource for mining and analyzing data. Our software is freely available, actively developed and allows one to rapidly deploy new community-specific 'Centrals'. As a result, we are able to support the communities of *Helicoverpa* sp., *Heliothis* sp., *Heliconius* sp., *Simulium* sp. and the wider whitefly community (*Bemisia* sp. and *Trialeurodes* sp.) as well as our new genome sequencing effort for the *Myrmecia* species group. We focus on interactive community-centric resources and have already driven synthesis studies from entire taxa. One such example is our cataloging of RNAi wet-lab experiments from 43 laboratories and subsequently publishing a comprehensive review in Lepidoptera. Current developments focus on developing interactive population genetic tools with the Atlas of Living Australia (<http://www.ala.org.au>).

38

The *Drosophila* Genetic Reference Panel: A Whole Genome Association Resource for Quantitative Traits.

*Richards, Stephen; Zhu, Dianhui; Han, Yi; Ayroles, Julien; Carbone, Mary Anna; Stone, Eric; Mackay, Trudy; Gibbs, Richard A.

(SR, DZ, YH, RAG) Molecular Human Genetics, Baylor Col Medicine, Houston, TX; (JA, MAC, ES, TM) Department of Genetics, North Carolina State University, Raleigh, NC.

Many of the traits most interesting to geneticists, biologists and entomologists are quantitative in nature, and involve many gene alleles with modest effect on the trait in question. Often these traits are more ephemeral than classic Mendelian traits – for example host plant selection between similar hosts Vs. segmentation defects identified in EMS segmentation screens. These quantitative traits require specific new genetic tools. The *Drosophila* Genetic Reference Panel (DGRP) is a community resource of 192 inbred *Drosophila melanogaster* lines for GWAS analysis of quantitative traits.

Freeze 1 of the DGRP comprises 162 sequenced and phenotyped inbred lines. The majority of the inbred lines have been sequenced to >12 X genome coverage on the Illumina platform, 40 lines have been sequenced to a minimum of 12X coverage using both 454 and Illumina sequencing platforms. The remaining lines were sequenced using the Illumina platform, and additional lines are in progress to bring freeze 2 up to the goal of 192 lines. Alignments have been generated using BWA, and polymorphisms identified using a newly developed method, accounting for the exact population structure of this data. We find approximately 500,000 SNPs, and 50,000 indels per inbred line, relative to the reference sequence. Extensive quality control genotyping ensured both sequence and strain sample integrity, matching and true homozygosity for all of the lines. Contaminated and heterozygous lines have been replaced to ensure no confounding effects on association studies.

Test whole genome association experiments on the released Freeze 1, (162 lines) lines have identified between 20-100 polymorphism associated with complex traits with p- values less than 10⁻⁵ Most of these genes are novel, and many have pleiotropic effects on multiple traits. Thus, we believe this tool will provide a new understanding of *Drosophila* genetics, complementary to the large body of work based on artificially induced mutational screens.

As a community resource we are working to facilitate the reference panels use in as many setting as possible. The fly stocks are currently available from the Bloomington stock center. We are also preparing a website for genome wide associations allowing input of phenotypes measured on the lines, and outputting putative polymorphisms and associated p-values. We hope this web tool will allow any *Drosophila* geneticist with the ability to work on flies to utilize the reference panel – even to the level of high school students counting leg bristles.

Finally, we discuss the possibility of bringing this technique to other arthropods, (for example, natural variation in pest resistance, behavior) and the requirements and costs needed for such an experiment.

39

OrthoDB: A Hierarchical Catalog of Orthologs across Arthropoda.

Zdobnov, Evgeny M.

Department of Genetic Medicine and Development, University of Geneva Medical School, 1 rue Michel-Servet, 1211 Geneva, Switzerland; Swiss Institute of Bioinformatics, 1 rue Michel-Servet, 1211 Geneva, Switzerland; Imperial College London, South Kensington Campus, London, SW7 2AZ, United Kingdom.

The OrthoDB resource presents a catalog of eukaryotic orthologs delineated at each radiation of the species phylogeny in an explicitly hierarchical manner, and currently covers 25 arthropod species from the arachnid deer tick to the dipteran fruit flies. New database features include functional annotations in the form of Gene Ontology and InterPro attributes and characterised phenotypes from *Drosophila melanogaster*, as well as quantification of evolutionary divergence and relations among orthologous groups. The interface features extended phyletic profile querying and enhanced text-based searches. OrthoDB orthology data have facilitated comprehensive comparative gene repertoire analyses as part of several recent arthropod genome projects including the water flea, *Daphnia pulex*, the human body louse, *Pediculus humanus*, the *Culex* mosquito, *Culex quinquefasciatus*, and the parasitoid wasp, *Nasonia vitripennis*. The ever increasing sampling of sequenced arthropod genomes brings a clearer account of the majority of gene genealogies that will facilitate informed hypotheses of gene function in newly sequenced genomes, as well as providing essential data for uncovering and quantifying long-term trends of gene evolution. OrthoDB is referenced with link-outs from a number of resources including UniProt and FlyBase, and is freely accessible from www.orthodb.org.

References: Kriventseva *et al.*, 2008; Waterhouse *et al.*, 2011.

Acknowledgements: Dr Evgenia V. Kriventseva and members of the Computational Evolutionary Genomics Group (<http://cegg.unige.ch>, CEGG, University of Geneva).

40

Patterning the dorsal-ventral axis in the jewel wasp *Nasonia vitripennis*.

*Buchta, Thomas; Özüak, Orhan; Roth, Siegfried; Lynch; Jeremy A.

University of Cologne, Institute of Developmental Biology, Cologne, Germany

Patterning the dorsal-ventral (D-V) axis relies on complex gene regulatory networks, which are best understood in the fruit fly *Drosophila melanogaster*. To gain further insights into the evolution of developmental patterning mechanisms we compare the D-V system of *Drosophila* to that of the flour beetle *Tribolium castaneum* and the jewel wasp *Nasonia vitripennis*. In comparison to the beetle, which has a more ancestral mode of embryogenesis, *Nasonia* embryogenesis represents an independently derived *Drosophila*-like mode. While patterning the D-V axis of *Drosophila* relies on a strict, hierarchical Toll/NF- κ B signaling, self-regulatory circuits are responsible in *Tribolium*. In both organisms, the BMP signaling is restricted to the dorsal half of the embryo. To understand how the Toll/NF- κ B - and BMP signaling pathways contribute to the establishment of the D-V axis in *Nasonia*, a detailed understanding of the normal development of the embryo is required. Here we show expression patterns of representative dorsal-ventral marker genes during embryogenesis and compare them to *Drosophila* and *Tribolium*.

41

Ectopic expression of hunchback during germband growth of *Tribolium* results in ectopic gap gene expression

*Distler, Jutta; Cerny, Alexander; Klingler, Martin

(DJ, MK) Friedrich Alexander Universität Erlangen, Department of Developmental Biology, 91052 Erlangen; (AC) Universität Hohenheim, Department of Biosensorics, 70593 Stuttgart

The knock-down of gap genes like *hunchback* (*hb*) and *giant* (*gt*) by parental RNAi in *Tribolium* results in segmentation and homeotic defects. Both genes are expressed in several domains during embryogenesis, i.e. in the blastoderm as well as during subsequent growth stages. Parental RNAi cannot distinguish between separate functions in early and late stages. To overcome this problem, *hb* dsRNA was injected into germ rudiment embryos, which resulted in homeotic defects but did not identify a segmentation function for the posterior domain of *Tc'hb*.

As an alternative method to interfere with gap gene function in germband stages, we generated transgenic strains based on the *Tribolium* heatshock promoter (Schinko and Bucher, unpublished) which allow to induce a gap gene expression at any timepoint of development. While overexpression of *Tc'gt*, which is normally active in the anterior abdomen of the embryo, results in anterior abdominal defects, *Tc'hb* overexpression in the early germband often results in larvae displaying additional abdominal segments. Given that we do not induce spatially specific expression in these experiments, we interpret this as evidence for a segmentation clock mechanism in short germ segmentation while a stripe-specific function of the posterior *hb* domain seems less likely.

42

The function of the serosa in *Tribolium castaneum*.

*Jacobs, Chris C.G.; Rezende, Gustavo L.; van der Zee, Maurijn
(CJ, MvdZ) Leiden University, Department of Biology, Leiden, 2333 BE, the Netherlands;
(GLR) *Laboratório de Fisiologia e Controle de Artrópodes Vetores, Instituto Oswaldo Cruz, FIOCRUZ Rio de Janeiro, Brazil*

Early in development the eggs of insects are divided into embryonic and extraembryonic primordia. The embryonic primordium forms the embryo proper, whereas the extraembryonic primordium forms transient epithelia: the amnion and the serosa. These wrap the germ band embryo. The serosa is proposed to protect against desiccation and might have facilitated the spectacular radiation of insects on land. However these extraembryonic tissues are reduced to a dorsal vestige, the amnioserosa, in the intensely studied *Drosophila* and other higher flies. Furthermore, in *Tribolium castaneum*, it has been shown that when removing the serosa with *Tc-zen1* RNAi, the eggs can hatch, and normally sized larvae arise from these eggs. This questions the function of the serosa.

We have removed the serosa in *T.castaneum* with *Tc-zen1* RNA interference and tested the role of the serosa in desiccation resistance. We found a dramatic increased mortality of serosa-less eggs at low humidity, compared to the control or wild-type eggs, confirming a role for the serosa in desiccation resistance. Surprisingly, there was a strong increase of the mortality in serosa-less eggs at high humidity. The mortality in wild-type and control eggs only increased in completely saturated air. Interestingly, we observed swelling of eggs at high humidity and presumed that water enters the egg because of osmotic pressure. Currently we are investigating the role of the serosal cuticle in resisting this osmotic pressure.

Summarizing, we found a clear relation between the serosa and the ability to withstand extreme low humidities. The observation of the swelling of the egg suggests that deformations could cause higher mortality at high humidities. In conclusion, the serosa protects the insect egg against terrestrial challenges: desiccation and the presence of fresh water.

43

Insights into the evolution of germline determination from the genome of the wasp *Nasonia vitripennis*

*Lynch, Jeremy A.; Özüak, Orhan; Khila, Abderrahman; Abouheif, Ehab; Desplan, Claude; Roth, Siegfried
(JAL, OÖ, SR) Institute for Developmental Biology, University of Cologne, 50674 Cologne, Germany;
(AK, EA) Department of Biology, McGill University, avenue Dr Penfield, Montreal, Quebec H3A 1B1, Canada;
(CD) Center for Developmental Genetics, Department of Biology, New York University, New York, NY 10003 USA

In *Drosophila*, the gene *oskar* (*osk*) is both necessary and sufficient for the localized assembly of the germ plasm at the posterior pole of the oocyte and embryo. Syncytial nuclei that come into contact with this germ plasm during blastoderm formation become specified as the primordial germ cells. This mode of maternal provision of germ plasm is not ancestral among insects. Rather, most basally branching hemimetabolous species specify the germline later in embryogenesis, through zygotic induction mechanisms. *Drosophila*-like modes of maternal provision are found only among some lineages of the Holometabola, but it has not been clear whether there was a single, or multiple derivations of this strategy. In addition, *osk* orthologs had so far only been found in dipteran (flies and mosquitos) genomes, and is absent from the genomes of other sequenced holometabolans (*Apis*, *Bombyx*, *Tribolium*). Interestingly, these species that lack *osk* all also lack identifiable germ plasm in the early embryo, and do not form pole cells. We sought to understand the origin of maternal provision of germ plasm and pole cells in insects by examining this process in the wasp *Nasonia*, a relatively close relative of *Apis*. We found that an *osk* ortholog is present in the *Nasonia* genome, and has conserved position within a regulatory network for the localized production of maternal germ plasm. This result, along with supporting data from ants and beetles with maternal germ plasm, indicates that there was a single origin of maternal provisioning of germ plasm, and that there multiple independent losses of this strategy, within the Holometabola. In addition, the *Nasonia* sequence appears to be evolving more slowly than its fly counterpart, allowing the detection of clues as to the origin of this novel factor, which may have included horizontal transfer, and gene fusion events.

44

Life history differences between two social bees affect expression of genes involved in senescence: *Apis* vs *Bombus*

*Margotta, Joseph W.; Mancinelli, Georgina E.; Benito, Azucena A.; Ammons, Andrew; Roberts, Stephen P.; Elekonich, Michelle M.
(JWM, GEM, AAB, MME) School of Life Sciences, University of Nevada, Las Vegas, Las Vegas, Nevada 89154; (AA) Department of Biology and Environmental Science, Goshen College, Goshen, Indiana 46526; (SPR) Department of Biology, Central Michigan University, Mount Pleasant, Michigan 48858

Understanding how the relationship between life history characteristics and the environment are mediated by physiological mechanisms provides insight into how behavior can move an organism more quickly or slowly along its life history pathway. Differences in life history characteristics, specifically seasonality, between *Apis mellifera* and *Bombus impatiens* lead to drastically different lifespan potentials. Because of its ability to overwinter, the perennial honey bee has the potential for a 3-fold increase in lifespan compared to the annual bumble bee. Here, we utilized a naturally occurring behavior that limits lifespan (flight) to compare the effect of different life history trajectories on expression of genes involved in senescence. We used whole-genome microarrays to identify tissue specific genomic changes associated with senescence in honey bees and bumble bees, which were collected during their metabolically intense flight behavior. In both the honey bee and bumble bee, we found that flight muscle had increased numbers of differentially regulated transcripts compared to brain. This result suggests that flight muscle undergoes pronounced genomic changes during flight likely because it is the most metabolically

active tissue in the bee. Specific genes involved in mitigating damage from oxidative stress and preventing the formation of reactive oxygen species were more pronounced in honey bee brains and flight muscle compared to brains and flight muscle of the bumble bee. More pronounced mechanisms that mitigate oxidative stress may have co-evolved with the longer lifespan potential of the honey bee.

45

The role of the BMP and Toll/NF- κ B Pathway in patterning the dorsal-ventral axis of the jewel wasp *Nasonia vitripennis*.

*Özüak, Orhan; Buchta, Thomas; Roth, Siegfried; Lynch, Jeremy A.
University of Cologne, Institute of Developmental Biology, Cologne, Germany

Bone Morphogenetic Proteins (BMPs) play a major role in establishing the dorsal-ventral (D-V) axis of most bilaterian animals. In the fruit fly *Drosophila melanogaster* the BMP pathway patterns only the dorsal half of the embryo and acts downstream of the Toll/NF- κ B signaling pathway. Outside the insects the Toll/NF- κ B -Pathway is primarily used for innate immunity and lacks a function for D-V patterning indicating that the employment of Toll/NF- κ B signaling in D-V axis formation is an evolutionary novelty of insects. Recent studies in our lab revealed an important role of the Toll/NF- κ B pathway in patterning the D-V axis of the short germ flour beetle *Tribolium castaneum*. To address the question when the transition from a mainly BMP dependent patterning system, to one dominated by Toll/NF- κ B signaling, occurred within the insects we analyze both pathways in the parasitic jewel wasp *Nasonia vitripennis*, a representative of the most basal branch of the Holometabolous insects. In addition, *Nasonia* has a *Drosophila* like, however independently derived, long germ mode of embryogenesis. These characteristics make *Nasonia* an ideal model system with which to understand the evolution of D-V patterning mechanisms.

46

Developmental patterning gene inventory of an emerging model chelicerate, *Tetranychus urticae*

*Pace, Ryan M.; Rombauts, Stephane; Van de Peer, Yves; Grbic, Mike; Nagy, Lisa
(RMP, LN) University of Arizona, Department of Molecular and Cellular Biology, Tucson, AZ 85719 USA; (SR, YVP) Department of Plant Systems Biology, VIB, 9052 Ghent, Belgium; (SR, YVP) Department of Plant Biotechnology and Genetics, Ghent University, 9052 Ghent, Belgium; (MG) University of Western Ontario, Department of Biology, London, Ontario, N6A 5B7, Canada

The spider mite *Tetranychus urticae* (Arthropoda, Chelicerata) is an agricultural pest known for its broad host range and rapid evolution of pesticide resistance. Occupying a basal position within the arthropod phylogeny, the chelicerates are differentiated from other arthropods by features such as chelicerae, four-pairs of walking legs, and a body plan consisting of two regions, the prosoma and opisthosoma. Along with other related mites, *Tetranychus* is further distinguished by an extremely reduced body plan through the apparent fusion of the opisthosoma and the prosoma. The current paradigm is that the formation of the arthropod body plan is established through multiple gene regulatory networks that are part of a conserved metazoan developmental "toolkit". Using a comparative genomics approach, we asked to what extent this toolkit is preserved in emerging genome datasets like *Tetranychus*. As expected, we find the majority of the developmental toolkit in the *Tetranychus* genome. We find a novel organization and inventory of Hox genes. We also find differential *Tetranychus*-specific gene gains and losses, with preferential losses of ligands and duplications of receptors in developmental signaling pathways. Preliminary analysis of RNA-seq data suggests that certain duplicated receptors may have evolved functions outside of canonical developmental pathways. Experiments are in progress to assay the expression and function of the *Tetranychus* Hox genes.

47

Identification of maternal effect genes in *Tribolium castaneum* using a tiled whole genome microarray

*Preuss, Kevin M.; Colbourne, John K.; Wade, Michael J.
(KMP, MJW) Indiana University, Department of Biology, Bloomington, IN 47405; (JKC) Indiana University, Center for Genomics and Bioinformatics, Bloomington, IN 47405

Maternal effect genes are known to play critical roles in the development of offspring as well as influence evolutionary dynamics. The study of maternal effect genes in emerging model systems has been largely limited to the homologs of maternal effect genes known from more established model systems, making the discovery of novel maternal effect genes unlikely. We identified maternal effect genes in an unbiased manner in an emerging model system, *Tribolium castaneum*, using total RNA from fertilized and unfertilized eggs on a tiled whole genome microarray. Maternal genes were identified as those genes present in unfertilized egg samples. Comparison of RNA levels in fertilized and unfertilized eggs also suggests the rapid degradation upon fertilization of some maternal effect genes. These findings demonstrate the utility of tiled whole genome microarrays in identifying maternal effect genes in an unbiased manner at the genomic level.

48

Mating Induces Gene Expression Changes in the Parasitic Wasp *Nasonia vitripennis*

*Watt, Rebekah; Trivedi, Urmi; Park, Jennifer; Blaxter, Mark; Shuker, David M.
(RW, UT, MB) Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, U.K; (JP, DMS) School of Biology, University of St Andrews, St Andrews, Fife, U.K.

Mating is known to induce a suite of changes in the behaviour and physiology of female insects. For many species this includes the loss of receptivity towards further mating attempts from males. How and why female receptivity changes after mating is important for understanding mating system evolution, and the opportunity for sexual selection and sexual conflict over mating. Here we consider the genomic basis of post-mating changes in the parasitoid wasp *Nasonia vitripennis*. This

species is typically monandrous, with female receptivity sharply declining following mating. This behavioural change is believed to be driven by male courtship behaviour as opposed to a mechanism associated with seminal protein transfer during copulation as in *Drosophila melanogaster*. Previous studies have considered the transcriptional changes which occur post-mating in species such as *D. melanogaster* and *Apis mellifera* but it is not yet known whether similar pathways control loss of receptivity and other changes in reproductive physiology across insects. We used a Digital Gene Expression (DGE) technique to generate estimates of gene expression in virgin and recently-mated females, comparing two time points. Head and body (considering thorax and abdomen) tissues were analysed separately. We found significant changes in gene expression associated with mating across both heads and bodies, and that these patterns changed between 30 minutes and four hours post-mating. In particular, head samples demonstrated increasing levels of differential expression by four hours. For body samples we saw a dip in gene expression after 30 minutes followed by resumption of virgin-like expression levels by four-hours post-mating. These broad patterns of expression reflect those found in *D. melanogaster*. Amongst a number of more specific changes, there were increases in gene expression among the GO classes associated with signalling and transcriptional regulation in the head samples, although unlike other studies we found less evidence for changes in immune-related genes. In general, our results support the general patterns of changes in gene expression following mating seen in other insects, albeit with different gene classes underlying those changes.

49

Functional Analysis of Members of the ABC Transporter Gene Superfamily in *Tribolium castaneum*

*Broehan, Gunnar; Muthukrishnan, Subbaratnam; Lorenzen, Marcé D.; Merzendorfer, Hans (GB, HM) University of Osnabrück, Department of Biology/Chemistry, 49069 Osnabrück, Germany; (SM) Department of Biochemistry, Kansas State University, Manhattan, KS 66506, USA; (MDL) Department of Entomology, North Carolina State University, Raleigh, NC 27695, USA

The ATP binding cassette (ABC) transporter superfamily contains membrane bound proteins involved in the translocation of substrates across intra- and extracellular membranes as well as soluble proteins involved in DNA repair and protein synthesis. We screened the *Tribolium castaneum* genome for genes encoding ABC transporters (*TcABCs*) using a homology based approach and identified 72 non-redundant *TcABCs* from 8 distinct subfamilies (*TcABC-A* to *-H*). We analyzed the expression pattern of selected members of each subfamily by RT-PCR. While most of the *TcABCs* showed constitutive expression throughout all tested developmental stages, some showed induced expression in particular stages. Nearly all *TcABCs* were expressed in the midgut and the carcass with the exception of *TcABC-A9B*, which was expressed exclusively in the midgut. To analyze *TcABCs* with conspicuous expression patterns in more detail, we performed systemic RNAi experiments injecting dsRNA to various *TcABCs* including *TcABC-G8A*, *TcABC-G9B* (*TcWhite*), *TcABC-C9A* (*TcSUR-1*), *TcABC-C5H* (homologous to vertebrate *CFTR*) and *TcABC-F2A*. While RNAi mediated knockdown of *TcSUR-1* and *TcABC-C5H* expression did not affect growth and development, knockdown of *TcWhite*, responsible for the transport of eye pigment precursors, led to the expected white-eyed phenotype (Lorenzen & Beeman, unpublished), and knockdown of *TcABC-F2A* resulted in molting defects with impaired development of adults eyes. Injection of dsRNA for *TcABC-G8A*, like *TcWhite* a member of subfamily ABCG and a homologue of the *Drosophila* Early gene at 23 (*TcE23*), produced a distinct phenotype. When the dsRNA was injected into late larvae, animals were not able to shed their old cuticle during larval to pupal molt. This severe molting defect resembles a phenotype observed when injecting dsRNA for *TcCTLP-5C*, a gene encoding a chymotrypsin-like peptidase from the molting fluid, and for *TcETH*, the gene encoding the ecdysis triggering hormone. A polymorphic phenotype could be observed when dsRNA for *TcE23* was injected into prepupae. The animals were either not able to complete adult molt or if they succeeded, their posterior abdominal segments became elongated and protruded under the elytra. Further analysis of *TcE23* will allow us to draw conclusions on its suggested role in ecdysteroid transport.

50

Chymotrypsin-like Peptidases from *Tribolium castaneum*: Functional Analysis by Systemic RNAi and Site-directed Mutagenesis of Recombinant Proteins

Broehan, Gunnar; Albaum, Daniel; Muthukrishnan, Subbaratnam; Merzendorfer, Hans* (GB, DA, HM) University of Osnabrück, Department of Biology/Chemistry, 49069 Osnabrück, Germany; (SM) Department of Biochemistry, Kansas State University, Manhattan, KS 66506, USA

The genome of the red flour beetle, *Tribolium castaneum*, comprises 14 genes encoding peptidases with S1 specificity pocket residues typically found in chymotrypsin-like peptidases (*TcCTLPs*). We analyzed the expression patterns of seven *TcCTLP* genes at various developmental stages by RT-PCR. While some *TcCTLP* genes were exclusively expressed in feeding stages, others were also detected in non-feeding embryonic and pupal stages. All tested *TcCTLP* genes were expressed predominantly in the midgut. However, *TcCTLP-5C* and *TcCTLP-6C* also showed considerable expression in the carcass and hence may act as molting fluid enzymes. In line with this hypothesis, crude extracts from larval exuviae reacted with antibodies to *TcCTLP-6C* and *TcCTLP-5C*, and injection of dsRNA for *TcCTLP-5C* and *TcCTLP-6C* into larvae resulted in severe molting defects suggesting that they have specific functions in degrading the old cuticle. To further investigate *TcCTLP-5C* function, we produced a recombinant version of this enzyme in insect Hi5 cells using a baculoviral expression system. We mutated the activation peptide of *TcCTLP-5C*, and replaced its conserved trypsin cleavage site by a Factor Xa cleavage site. After purification and complete Factor Xa activation, the recombinant enzyme exhibited a significant chymotrypsin-like activity using *N*-succinyl-Ala-Ala-Pro-Phe-p-nitroanilide as a substrate ($K_m=0.72 \text{ mmol l}^{-1}$, $K_{cat}=0.15 \text{ s}^{-1}$, $K_{cat}/K_m=208.3 \text{ l mol}^{-1} \text{ s}^{-1}$). In SDS-PAGE with and without β -mercaptoethanol activated *TcCTLP-5C* migrated with a Mr of about 35 and 25 kDa, respectively, the latter corresponded with the theoretical Mr of 24.6 kDa. *TcCTLP-5C* contains a highly conserved *N*-glycosylation site. To test for *N*-glycosylation, we performed western blots and lectin stainings. We found that *TcCTLP-5C* is specifically decorated by the *Galanthus nivalis* agglutinin, which exhibits an exclusive mannose-binding specificity. In addition, glycosidase F treatment resulted in a considerable band shift suggesting that *TcCTLP-5C* is indeed glycosylated. Finally, we mutated *TcCTLP-5C* by

site-directed mutagenesis in various ways to examine the effects of single amino acid substitutions within the *N*-glycosylation site, active center and the substrate specificity pocket. The obtained data will provide more detailed insights into the structure-function relationship of insect CTLPs and their role during molt.

51

***Tribolium castaneum* Knickkopf is a chitin-binding protein that protects procuticular chitin from chitinases**

*Chaudhari, Sujata S; Arakane, Yasuyuki; Specht, Charles A; Moussian, Bernard; Boyle, Daniel L; Park, Yoonseong; Kramer, Karl J; Beeman, Richard W; Muthukrishnan, Subbaratnam (SC, KJK, SM) Department of Biochemistry, Kansas State University, 141 Chalmers Hall, Manhattan, KS 66506, USA; (YA) Division of Plant Biotechnology, Chonnam National University, Korea; (CS) Department of Medicine, University of Massachusetts, Worcester, USA; (BM) Interfaculty Institute for Cell Biology, Animal Genetics, University of Tübingen, Tübingen, Germany; (DB) Department of Biology, Kansas State University, 116 Ackert Hall, Manhattan, KS 66506, USA; (YP) Department of Entomology, Kansas State University, Manhattan, KS, USA; (KJK, RWB) USDA, Agricultural Research Service, Center for Grain and Animal Health Research, 1515 College Avenue, Manhattan, KS 66502, USA

The rigid, chitinous cuticle or exoskeleton is the primary protective barrier that protects insects against infection, injury and desiccation. In order to accommodate saltatory growth, insects periodically replace their old cuticles with new ones, a process that involves significant chitinolytic activity. Chitinases degrade the old exoskeleton and recycle some of the resulting products for new cuticle synthesis. Curiously, chitinases are found not only in those portions of old cuticle targeted for degradation and recycling, but also in the newly developing procuticle. This raises a question of fundamental importance: What protects the new cuticular chitin from degradation by these chitinases? In the present study we demonstrate that *Tribolium castaneum knickkopf* (*TcKnk*) encodes a chitin binding protein that selectively co-localizes to the new cuticle and protects chitin from the activity of molting fluid chitinases. Down-regulation of *TcKnk* transcripts results in loss of chitin and severe molting defects that are lethal at all developmental stages. The conservation of *Knk* in all insect, crustacean, echinoderm and nematode species examined suggests that such a critical role for this protein in exoskeletal chitin accumulation and maintenance may be conserved in all chitinous invertebrates.

52

Sand fly Peritrophins: Domain Architecture, Expression Profile, and Functional Analysis

*Coutinho-Abreu, Iliano V.; Sharma, Narinder K.; Robles-Murgia, Maricela; Ramalho-Ortigao, Marcelo Kansas State University, Department of Entomology, Manhattan KS, 66506, USA.

Sand fly peritrophic matrix (PM) plays an important role in blood meal digestion and also functions as a barrier against *Leishmania* infection. Although the molecular characterization of sand fly peritrophins, the main peritrophic matrix proteins, can have significant impact in the development of strategies to counteract the spread of leishmaniasis, little is known about sand fly PM molecular constituents or their roles in sand fly vector competence. Recently, our group and others have identified six peritrophin transcripts in midgut cDNA libraries of the sand flies *Phlebotomus papatasi* and *Lutzomyia longipalpis*. An additional sequence has been detected in whole body *L. longipalpis* cDNA libraries. Using peritrophin sequences previously published, we have carried out a deeper search in a *L. longipalpis* sequence database and found seven new peritrophin sequences. Altogether, these 14 sand fly peritrophins exhibit up to four chitin binding domains (CBDs), and some of those also display a mucin-like sequence. Interestingly, the *P. papatasi* and *L. longipalpis* ortholog peritrophins PpPer3 and LIPer3 exhibit a conserved putative CBD devoid of the functional aromatic amino acids, and their mucin-like counterparts display sequence and length polymorphisms. Whereas LIPer3 exhibits a mucin-like sequence twice as long as PpPer3, the former displays 10% less predicted sites of *O*-linked glycosylation. Moreover, LIPer3 mucin-like sequence is enriched in serine residues (30.3%) while threonine amino acids predominate (36.4%) as *O*-linked glycosylation sites in the related PpPer3 mucin-like sequence. We also have detected a potential new CBD domain bearing 8 cysteines in the recently identified *L. longipalpis* LuloPer7 peritrophin. The mRNA expression profiles revealed that only one *P. papatasi* (*PpPer1*) and two *L. longipalpis* (*LuloPer1* and *LuloPer2*) peritrophins display midgut-specific expression in adult females. All the other peritrophins were expressed in many different adult tissues, such as ovary, Malpighian tubule, hindgut, carcass, and/or fat body, pointing to other roles besides PM scaffolding. Nine sand fly peritrophins also were expressed in the developmental stages. We have biochemically characterized the *P. papatasi* PpPer1 and PpPer2 peritrophins, showing that their recombinant counterparts (expressed in CHO-S cells) bind to chitin *in vitro* and exhibit similar retention times as the native ones by HPLC. Moreover, expression profiles of two *P. papatasi* peritrophins (PpPer1 and PpPer3) were differentially modulated in the midgut upon *Leishmania major* infection, suggesting their participation in sand fly vector competence. Indeed, knocking down *PpPer1* expression by double-strand RNA injections reduced *PpPer1* mRNA and protein levels by 45% and led to increased *Le. major* loads in *P. papatasi* midguts at 48h (39%) and 96h (22%) post-infection. Therefore, PpPer1 is a harmful molecule for *Leishmania*, and the presence of PpPer1 in *P. papatasi* PM strengthens its role as a barrier against *Le. major* infection.

53

Distinctive Amino Acid Composition Profiles in Salivary Proteins of the Tick *Ixodes scapularis*

*Hughes, A.L., Friedman, R.

Department of Biological Sciences, University of South Carolina, Columbia SC 29208

Analysis of the amino acid composition of proteins of the tick *Ixodes scapularis* showed that salivary proteins in general have relatively high frequencies of small residues and polar residues and relatively low frequencies of the non-polar residues leucine and valine. Computer prediction of linear B-cell epitopes provided no evidence that salivary proteins as a whole have evolved characteristics minimizing their antigenicity to the vertebrate host. However, certain salivary proteins such as ZNF706 poly-lysine basic tail salivary protein, showed unusually high frequencies of polar residues, consisted with the hypothesis of

adaptation to minimize antigenicity. By contrast, *I. scapularis* proline- and glycine-rich protein and salivary mucins had high predicted antigenicity, and may represent suitable candidates for salivary antigen components of an anti-tick vaccine.

54

Peritrophic Matrix Proteins of *Tribolium castaneum*

Jasrapuria, Sinu; *Williams, Clay; Arakane, Yasuyuki; Beeman, Richard W.; Kramer, Karl J.; Muthukrishnan, Subbaratnam (JS, CW, YA, KJK, SM) Department of Biochemistry, Kansas State University, Manhattan, KS 66506; (RWB, KJK) Center for Grain and Animal Health Research, ARS-USDA, Manhattan, KS 66502; (YA) Division of Plant Biotechnology, Chonnam National University, Gwangju 500-757, Korea

This study focuses on Peritrophic Matrix Proteins (PMPs), especially their domain organization, expression, and RNAi induced phenotypes, in *Tribolium castaneum*. The peritrophic matrix lines the midgut epithelium of insects. It is made up of chitin and structural proteins, and functions to compartmentalize digestion and prevent mechanical damage and microorganism attack. PMPs represent one family of proteins with chitin binding domains related to the peritrophin A domain. Unlike other proteins with chitin binding domains, PMPs are expressed only in the midgut, and only during feeding stages. Furthermore, *Tribolium castaneum* PMPs have a variable number of chitin binding domains, with as many as 14 chitin binding domains in one protein. Also, each PMP contains a cleavable signal peptide. Because PMPs have cleavable signal peptides and multiple chitin binding domains, they may interact with chitin in an extracellular matrix by crosslinking chitin fibrils. RNA interference was used to determine loss of function phenotypes for each PMP gene. RNAi of PMP3C showed fat body depletion, reduced egg lay, and high mortality in adults as well as a lethal shrinking phenotype in larvae. RNAi of PMP5B showed reduced hatching of the offspring of dsRNA-injected adults, and a molting defect in the larval stage.

This work was supported by the NSF grant IOS-1022227

55

Transient Receptor Potential (TRP) Channels in the Red Flour Beetle, *Tribolium castaneum*

*Kim, Hong Geun; Margolies, David C.; Park, Yoonseong
Kansas State University, Department of Entomology, Manhattan, KS 66506

Transient Receptor Potential (TRP) channels are a family of six-transmembrane polypeptide that forms a tetrameric channel that is permeable to cations. Various sensory functions, including photo-, olfactory-, and mechano-reception, are exerted through the TRPs. Among the TRPs, four members – TRPA, TRPC, TRPM and TRPV - are the subfamilies known for including thermo-TRPs, which are activated by specific ranges of temperature and involved in thermosensory information processing. We have identified 15 genes encoding TRP channel genes in the genome sequence of *Tribolium castaneum* Herbst. (Coleoptera: Tenebrionidae). Phylogenetic analysis with 13 *Drosophila melanogaster* TRPs, 15 *T. castaneum* TRPs and six mammalian thermo-TRPs suggested that seven subfamilies named TRPA, TRPV, TRPC, TRPN, TRPM, TRPP and TRPML, are ancestrally conserved. Among 15 *T. castaneum* TRPs, three members of the TRPA subfamily, previously described as the thermosensory TRP channels in *D. melanogaster*, were identified as *pyrexia* (TC09731), *painless* (TC007561), and *trpA1* (TC002449). We examine the three selected thermo-TRP channels by using the RNA interference to assess the effect on behavioral phenotype.

56

Catecholamine Biosynthetic Enzymes, Tyrosine Hydroxylase And Dopa Decarboxylase In *Ixodes Scapularis* Tick.

Koči, J.; Šimo, L.; Park, Y.
Kansas State University, Entomology Department, Manhattan, KS 66506

Catecholamines, such as dopamine (DA), norepinephrine (NE) and epinephrine (E) are essential molecules in the physiology of diverse animals. DA is the major inducer of tick salivation and its biosynthetic pathway involves eukaryotic enzymes: (i) tyrosine hydroxylase (TH), catalyzing the first step in catecholamine biosynthesis by hydroxylation of L-tyrosine to produce the L-dihydroxyphenylalanine (L-DOPA) and (ii) L-DOPA decarboxylase (DDC), which catalyses the decarboxylation of L-DOPA to DA. L-DOPA can be further processed to tyramine and octopamine in insects and DA processed to NE and E in vertebrates. Evidence of DA in tick salivary glands led us to the study of the biosynthetic enzymes TH and DDC. Taking advantage of the genome sequence available in *Ixodes scapularis*, we have identified at least two TH genes and a DDC gene. Spatial and temporal expression results obtained by in situ hybridization and qRT-PCR support the existence of the catecholamine biosynthetic enzymes in the tick salivary glands. This study lays a foundation to investigate the mechanisms involved in the regulation of catecholamine production in the salivary glands of tick.

57

Mode of Action of the Insect Growth Regulator Diflubenzuron: A Comprehensive Study in *Tribolium castaneum* Utilizing Genomic Tiling Array and Immunolocalization Approaches

*Kumari, Meera; Merzendorfer, Hans; Kim, Hee S.; Brown, Susan J., Specht, Charles A.; Beeman, Richard W.; Kramer, Karl J.; Muthukrishnan, Subbaratnam
(HM, MK, KJK, SM) Department of Biochemistry, Kansas State University, Manhattan, KS 66506; (HM) University of Osnabrück, Department of Biology/Chemistry, 49069 Osnabrück, Germany; (CAS) Department of Medicine, University of Massachusetts, Worcester, MA 01605, USA; (RWB, KJK) Center for Grain and Animal Health Research, ARS-USDA, Manhattan, KS 66502; HK, SJB, Division of Biology, Kansas State University

Several benzoyl-phenyl-urea-derived insecticides such as diflubenzuron (DFB, Dimilin®) are widely used to control various insect pests. Although this class of compounds disrupts molting and affects chitin content, their precise mode of action is not known. To evaluate the mechanism underlying the insect growth regulating effects, we are conducting a comprehensive study utilizing genomics, and immunolocalization tools in the model insect species and stored product pest, *Tribolium castaneum* (red flour beetle). The results showed that DFB affects larval-larval and larval-pupal molting. In addition, the ensuing adults had walking defects and died prematurely after adult eclosion.

The chitin content of whole larvae or isolated larval midguts was reduced by ~50% after DFB treatment as revealed by the Morgan-Elson method. This result was consistent with confocal analysis of elytra of control and DFB-treated pharate adults using a chitin-binding probe. In contrast, chitosan content was elevated significantly after DFB treatment. Results of RT-PCR and immunolocalization studies of several genes involved in chitin metabolism will be presented. Adult females fed a DFB-containing diet laid normal a number of eggs with reduced chitin content, but the embryos failed to hatch

In summary, DFB treatment results in an increase chitosan, a reduction in chitin and mislocalization of specific proteins of chitin metabolism and cuticle formation. We also observed defects in the mobility (leg movement) of adults that eclosed following DFB-treatment of larvae. Hence, our results support the hypothesis that DFB affects the expression of multiple genes involved in cuticle chitin metabolism, which are critical for normal insect growth and development.

This work was supported by the NSF grant IOS-1022227.

58

Novel Functions of Two Acetylcholinesterase Genes in *Tribolium castaneum* Revealed by RNA Interference

Lu, Yanhui; Park, Yoonseong; Gao, Xiwu; Zhang, Xin; Yao, Jianxiu; Pang, Yuan-Ping; *Zhu, Kun Yan (YL, YP, XZ, JY, KYZ) Department of Entomology, Kansas State University, Manhattan, KS 66506; (YL, XG) Department of Entomology, China Agricultural University, Beijing, China; (YPP) Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN 55905

We investigated biological functions of each of two acetylcholinesterase genes (*Ace1* and *Ace2*) in the red flour beetle (*Tribolium castaneum*), a notorious stored grain pest worldwide, by RNA interference (RNAi). Silencing *TcAce1* gene in 20-day-old larvae resulted in significantly increased susceptibilities of the larvae when exposed to carbaryl, carbofuran, dichlorvos or malathion 3 days later and 100% mortality within two weeks after adult eclosion. In contrast, silencing *TcAce2* under the same conditions showed no or limited effect on the susceptibility of the larvae to these insecticides and did not lead to the adult mortality. However, late larvae injected with *TcAce2* dsRNA showed slow development, reduced female egg-laying and reduced egg hatchability as compared with those injected with *TcAce1* dsRNA. These results suggest that AChE1 encoded by *Ace1* is a key enzyme for cholinergic functions and is the target of organophosphate and carbamate insecticides, whereas AChE2 encoded by *Ace2* appears to play an important role in insect growth, female reproduction, and embryo development.

59

RNAi-based functional analysis of *yellow-e* in *Tribolium castaneum*

*Noh, Mi Young; Kanost, Michael R; Muthukrishnan, Subbaratnam; Beeman, Richard W; Kramer, Karl J; Arakane, Yasuyuki (MYN, YA) Chonnam National University, Division of Plant Biotechnology, Gwangju 500-757, Korea; (MRK, SM, KJK) Kansas State University, Department of Biochemistry, Manhattan, Kansas 66506, USA; (KJK, RWB) USDA-ARS Center for Grain and Animal Health Research, Manhattan, Kansas 66502, USA

Pigmentation/melanization is an important physiological event in insect development and is involved in cuticle tanning, wound healing and encapsulation as a defensive response as well as in egg chorion hardening. “Yellow” protein (dopachrome conversion enzyme, DCE) is involved in the melanin biosynthetic pathway and significantly accelerates pigmentation and melanization reactions. Recent studies have suggested that *yellow* is a rapidly evolving gene family generating functionally diverse paralogs. However, the exact physiological functions of several *yellow* genes are still not understood. *Tribolium castaneum* (*Tc*) is an excellent experimental insect to determine the roles of individual *yellow* genes because it is possible to carry out RNA interference (RNAi) experiments by injecting gene-specific double-stranded RNAs (dsRNAs) into insects at all developmental stages.

Injection of dsRNAs for *Tcyellow-e* (*TcY-e*) into last instar larvae had no effect on larval-pupal molting, and the resulting pupae developed normally. The pupal cuticle including the setae, gin traps and urogomphi showed normal tanning. Adult cuticle could also be viewed through the translucent pupal cuticle. Adult cuticle tanning including the head, mandibles and legs was initiated on schedule (pupal day 4-5), indicating that *TcY-e* is not required for pupal or adult cuticle pigmentation in *T. castaneum*. The subsequent pupal-adult molt, however, was adversely affected. Although pupal cuticle apolysis and slippage were evident, some of the adults (~20%) were unable to shed their exuvium and died entrapped in their pupal cuticle. In addition, the resulting adults rapidly became completely desiccated. Failure of the pupal-adult molt and desiccation-induced mortality were prevented by maintaining dsRNA *TcY-e*-treated insects at 100% relative humidity (rh). When the humidity-rescued adults were subsequently moved from 100% rh to 50% rh, they quickly desiccated and died (wild-type insects thrive throughout development at 50% rh). These results suggest that *TcY-e* has a “critical waterproofing” function for cuticle rather than cuticle pigmentation.

This work was partially supported by Arthropod Genomics Seed Grant NOBO2-83015.

60

A conserved cuticular protein sequence suggests important functions and serves as an indicator of contamination

*Prince, Michael R.; He, Ningjia; Willis, Judith H.

Department of Cellular Biology, University of Georgia, Athens, GA 30602

The most recent family of cuticular proteins to be named is the CPAP3 family (Jasrapuria et al 2010 IBMB 40:214). Members have three putative chitin binding domains (ChtBD2) with conserved spacing between cysteine residues. They are found throughout the arthropods, including chelicerates. The first member of the family recognized was *DmelGasp* (CG10287) named because it was expressed in embryonic tracheae (Barry et al. 1999 IBMB 29:319). When Behr and Hoch (2005 FEBS Lett 579:6827) discovered that there were 10 similar proteins in the *D. melanogaster* genome, they named the family obstrucsters and *Gasp* became *obstC*. With the latest classification its name has been changed again to *DmelCPAP3-C*. Sequence conservation among orthologs includes almost invariant residues within and between domains. The conservation of sequence across more than 600 million years suggests that this protein must serve important functions.

In situ hybridization analyses with the *Anopheles* CPAP-C ortholog (AGAP003308) revealed that its mRNA is associated with more than just the tracheal system, with provocative hybridization found in the vibration-sensing Johnston's organ at the base of the antennae.

EST sequences reported from two plants (Sitka spruce, *Picea sitchensis*, and Key Lime, [Citrus aurantiifolia](#)) and the Atlantic halibut ([Hippoglossus hippoglossus](#)) were clear orthologs of CPAP3-C. Further analysis revealed that these sequences came from specific arthropods that could be expected to be eating the plants and serving as a food source for the halibut.

Thus highly conserved sequences not only serve the species that bear their genes but can reveal contamination of sequence libraries by predators or prey.

Acknowledgement: We thank Marjorie Hoy for allowing us to search the transcriptome of *Metaseiulus occidentalis* for cuticular protein sequences.

61

A new model beetle "superworm" *Zophobas atratus*.

*Ruang-Rit, Krissana; Park, Yoonseong

Kansas State University, Department of Entomology, Manhattan, KS 66506

We propose to use a large superworm, *Zophobas atratus*, Tenebrionidae as a model system in the physiological study, while the red flour beetle, *Tribolium castaneum*, has been considered to be an excellent genetic model organism. The superworm provides an interesting biology of intraspecific competition by modulating the developmental program in the crowded condition. Delay in metamorphosis to the pupal stage occurs to avoid becoming the prey of cannibalism and likely involves an interesting endocrinology. In order to examine whether RNA interference (RNAi) is applicable to the *Z. atratus*, which is a critical requirement for the species being used as a model system, we performed RNAi targeting *laccase2*, which is the essential enzyme for sclerotization of the cuticle. We cloned a highly conserved region of the *laccase2* gene from the *Z. atratus* by using a degenerate PCR for 1,178 bp messenger RNA. The postembryonic RNAi of the *laccase2* resulted in strong inhibition of cuticle tanning. We find that the RNAi occurs in both larvae and pupae in a dose-dependent manner at the range of doses between 400 ng to 30 ug/individual.

62

Development of transgenic silkworms for functional genomics in *Bombyx mori* as a lepidopteran model insect.

*Sezutsu, H.; Tsubota, T.; Uchino, K.; Kobayashi, I.; Osanai-Futahashi, M.; Suzuki, T.; Tatematsu, K.I.; Iizuka, T.; Yonemura, N.; Tamura, T.

National Institute of Agrobiological Sciences, Transgenic Silkworm Research Unit, Tsukuba, Ibaraki, 305-8634, Japan

The assembly of 432 Mb genome of the silkworm *Bombyx mori* was published in 2008 and has become a reference in the genomic analysis of the very diverse order of Lepidoptera. In the post-genomic studies, various genetic tools such as binary expression systems, enhancer traps, gene traps, insertional mutagenesis, gene targeting, and RNAi are important for analysis of the genome. We have been developing the genetic tools using transgenic silkworms for gene functional analysis and new uses. We established and improved the methods to generate the transgenic silkworm using *piggyBac* and *Minos* transposons. We also developed the binary expression systems in the silkworm using a GAL4/UAS or Tet-On/Off system. Then we constructed the *piggyBac*-based GAL4 enhancer trap and gene trap systems by remobilizing transposon vectors to various chromosomal locations. Such systems can be used for targeted transgene expression in a stage- and organ/tissue-specific manner, and insertion of the transgene may disrupt the endogenous gene, creating a new mutant allele. We have constructed "Bombyx Trap DataBase", which contains information on EGFP-reporter expression and the insertion sites of mutators in the enhancer trap and gene trap lines, and is integrated into the silkworm genome database "KAIKObase". Recently, we are developing the methods for gene targeting using ZFN and so on. We are also trying to improve dsRNAi method in silkworm using a dsRNA transporter gene (*sid-1*) of *Caenorhabditis elegans*. These genetic tools will be crucial for the analyses of the gene functions in the silkworm as a model of lepidopteran insect.

63

Myoinhibitory Peptide and SIFamide Receptors in Neuropeptide Signaling Pathways Controlling the Tick Salivary Glands

* Šimo, Ladislav; Koči, Juraj; Park, Yoonseong

Kansas State University, Entomology Department, Manhattan, KS 66506

Recent discovery of SIFamide and myoinhibitory peptide (MIP) peptidergic innervation in the *Ixodes scapularis* salivary glands directed our attention to representative receptors of these two classes of neuropeptides. SIFamide and MIP immunoreactive axon terminals in the basal cells of the salivary glands acini II and III indicate an important role of these two neuropeptides in tick salivary gland controls. Using the homology blast search, we have identified one gene encoding a SIFamide receptor (SIFa-R) and two tandemly repeated genes encoding separate MIP receptors (MIP-R1 and MIP-R2). Currently, there is no evidence for multiple MIP receptors in any other arthropods. The SIFa-R and MIP-R1 were expressed in a CHO cells system, and showed high sensitivity in the intracellular calcium mobilization for their representative ligands. Tissue specific PCR showed the presence of MIP-R1 and SIFa-R in the central nervous system, as well as in the salivary glands. Immunohistochemistry and real time qPCR approaches were used to investigate the spatial and temporal dynamics of MIP1- and SIFa-R and their ligands in the central nervous system and salivary glands during tick feeding. The molecular and pharmacological characterization and expression patterns of the MIP and SIFamide receptors in this study provide the foundation for understanding the molecular mechanisms of the neuropeptidergic signaling pathways controlling tick salivary glands.

64

Genomic Analysis of Carboxyl/cholinesterase Genes in the Silkworm *Bombyx mori*

*Tsubota, T.; Sezutsu, H.; Shiotsuki, T.

National Institute of Agrobiological Sciences, Tsukuba, Ibaraki, 305-8634, Japan

Carboxyl/cholinesterase (CCE) genes have pivotal roles in dietary detoxification, pheromone or hormone degradation and neurodevelopment. The recent completion of genome projects in various insect species has led to the identification of multiple CCEs with unknown functions. Here, we analyzed the phylogeny, expression and genomic distribution of 69 putative CCEs in the silkworm, *Bombyx mori* (Lepidoptera: Bombycidae).

A phylogenetic tree of CCEs in *B. mori* and other lepidopteran species was constructed. The expression pattern of each *B. mori* CCE was also investigated by a search of an expressed sequence tag (EST) database, and the relationship between phylogeny and expression was analyzed. A large number of *B. mori* CCEs were identified from a midgut EST library. CCEs expressed in the midgut formed a cluster in the phylogenetic tree that included not only *B. mori* genes but also those of other lepidopteran species. The silkworm, and possibly also other lepidopteran species, has a large number of CCEs, and this might be a consequence of the large cluster of midgut CCEs. Investigation of intron-exon organization in *B. mori* CCEs revealed that their positions and splicing site phases were strongly conserved. Several *B. mori* CCEs, including juvenile hormone esterase, not only showed clustering in the phylogenetic tree but were also closely located on silkworm chromosomes. We investigated the phylogeny and microsynteny of neuropeptides in detail, among many CCEs. Interestingly, we found the evolution of this gene appeared not to be conserved between *B. mori* and other insect orders. Our genomic analysis has provided novel information on the CCEs of the silkworm, which will be of value to understanding the biology, physiology and evolution of insect CCEs.

65

Ultrastructural immunolocalization of cuticular proteins in the cuticle of *Anopheles gambiae*

*Vannini, Laura; Balakrishna Pillai, Agiesh Kumar; Willis, Judith H.

Department of Cellular Biology, University of Georgia, Athens, GA 30602

Anopheles gambiae devotes about 2% of its protein coding genes to structural cuticular proteins (CPs) that have been classified into a dozen distinct families. Sequence domains, homology models and some experimental work revealed that some families contribute to the cuticle by binding chitin, others appear to function in different ways. While considerable data are accumulating on the spatial distribution of individual CPs across the insect body, there is almost no information on the localization of well-defined CPs within the cuticle itself. We have begun to remedy this deficiency by using secondary antibodies labeled with colloidal gold to detect antibodies raised against two CPs, AgamCPF3 and AgamCPLCG3. We selected CPF3 for initial analysis and used a 15 aa peptide as the immunizing antigen. This primary antibody was applied to ultrathin sections of legs of pharate and newly emerged adults where mRNAs for CPF3 are abundant. CPF3 was selected because of profoundly different mRNA levels of transcripts in M and S incipient species of *An. gambiae* (Cassone et al. 2008 Mol Ecol 17:2491), the failure of recombinant CPF3 to bind chitin (Togawa et al. 2007 IBMB 37:675), and a homology model showing that the *Drosophila* pheromone 7,11-HD would fit its binding pocket (Papandreou et al. 2010 JIP 56:1420). This information led to the suggestion that CPF3 might be localized in the epicuticle where it could present a contact pheromone (Papandreou et al. 2010). Initial results are not consistent with this suggestion. We also have an antibody against a 14 aa peptide from CPLCG3. Its transcripts are more abundant in insecticide resistant than in insecticide sensitive *An. gambiae* (Awolola et al. 2009 Trans R Soc Trop Med Hyg 103:1139). CPLCG3 expression peaks after adult emergence, but *in situ* hybridization revealed that both CPF3 and CPLCG3 have a similar anatomical distribution of mRNA. Thus we hope that a comparison of their location in the cuticle might provide further clues as to their functions.

66

Analysis of expressed sequence tags from gut tissues of the corn planthopper, *Peregrinus maidis*, exposed to *Maize mosaic rhabdovirus*

Alviar, Karen B.; Rotenberg, Dorith; Hogenhout, Saskia A.; Aritua, Valente and Whitfield, Anna E. (KBA, DR, VA, AEW) Kansas State University, Department of Plant Pathology, Manhattan, KS, 66506; (SAH) The John Innes Centre, Department of Disease and Stress Biology, Norwich Research Park, Norwich, United Kingdom
*Correspondence: Anna E. Whitfield, Kansas State University, Department of Plant Pathology, Manhattan, KS, 66506, USA. Tel.: +1-785-532-3364; fax: +1-785-532-5692; e-mail: aewtospo@ksu.edu

The corn planthopper, *Peregrinus maidis*, causes direct feeding damage to plants and transmits *Maize mosaic rhabdovirus* (MMV) in a persistent-propagative manner. MMV must cross several insect tissue layers for successful transmission to occur, and the gut serves as an important barrier for rhabdovirus transmission. Characterization of insect molecules that interact directly with viral proteins to mediate internalization and movement of a virus in host tissues or indirectly through various host response pathways is essential for defining the relationship between vector and the virus harbored and transmitted by the vector. As a primary approach to addressing this need, we sequenced the gut transcriptome of MMV-infected *P. maidis*. The sequencing effort generated 15,892 unique ESTs (1,860 contigs and 14,032 singletons). The sequences were analyzed using Blast2GO and functional roles were assigned to 49% of the contigs and 35% of the singletons. Examination of the annotated ESTs revealed that a subset shared significant sequence similarity to proteins that may play roles in the infection cycle of MMV (e.g. receptors identified for other members of the family *Rhabdoviridae*) and recognition and response to MMV infection (innate immunity and the virus silencing pathways). ESTs from all six of the MMV genes were found in the insect gut transcriptome and their relative abundance reflected the expression levels documented for other rhabdoviruses. These tools will provide the necessary information to identify differentially-expressed genes in MMV-infected *P. maidis* guts and for comparative studies of arthropod interactions with rhabdoviruses.

67

Functional Serpin expression and analysis in a malaria vector *Anopheles gambiae*

*Gulley, Melissa; Graves, Tinea; Michel, Kristin
Kansas State University, Division of Biology, Manhattan, Ks 66506

Serpins are a group of serine protease inhibitors that play a crucial role in the regulation of innate immune responses in arthropods. Serpins are proteins that have several structurally important amino acid sequences within a so-called reactive center loop (RCL) that are indicative of their ability to act as inhibitors. These areas include an anchor region, hinge region and a P1 cleavage site that show remarked conservation amongst arthropods. Arginine or Lysine at this P1 cleavage site suggests inhibition of trypsin or a trypsin-like proteases, which has been shown in insects such as *Manduca*. In the mosquito *Anopheles gambiae*, the most prominent vector of malaria in Africa, 18 serpin genes encoding 22 distinct proteins have been annotated to date. Of these, SRPN2 and SRPN6 have been shown to be active inhibitors and involved in parasite clearance. This study aims to identify other inhibitory serpins and their roles during immune responses. By determining which serpins are inhibitory and involved in the immune response of *An. gambiae* a more complete understanding of the immune system of this mosquito will be obtained. This could potentially provide novel targets for malaria control.

Of the 22 serpin proteins encoded in the *Anopheles gambiae* genome, 15 retain an arginine/lysine at the P1 cleavage site. They are SRPN1, 2, 4A, 4B, 5-9, 10A, 10B, 12, 14, 16, and 19. Of these, only SRPN12, 14 and 19 vary significantly within their anchor and hinge regions. The predicted sequences of SRPN3, 10C, 10D, 11, 13, 17 and 18 do not retain Lys/Arg residue at the P1 site. Of these, SRPN11, 13, 17 and 18 do not show the conservation of the hinge and anchor regions typical for functional protease inhibitors, suggesting that these proteins may alternative functions.

So far, we have cloned the coding region of 22 *An. gambiae* serpin into the pET28a vector system. Recombinant proteins for serpin 1, 9, and 18 have been expressed in *E.coli* and purified using nickel and Q-Sepharose columns. SRPN1 and 18 have been further analyzed by protease activity assays. SRPN1 has shown no inhibition to several commercially available proteases (trypsin, plasmin, kallikrein, proteinase K, subtilisin C, thrombin, and chymotrypsin) *in vitro*.

In contrast, SRPN18 show statistically relevant inhibition of both trypsin and plasmin. These results are interesting as they contradict the predicted functionality of serpins based on conserved regions within the reactive center loop. To determine if SRPN18 plays a role in *An. gambiae* immune responses, we plan to analyze its expression by quantitative RT-PCR during immune challenge. Additionally, injections of long dsRNAs will be used to determine whether reduction in SRPN18 mRNA levels will change the mosquito's ability to support the development of malaria parasites.

68

Transcriptional Changes of Metabolic Genes in *Manduca sexta* under Immune Challenge

*Gunaratna, Ramesh T.; Dillwith, Jack; Jiang, Haobo
Oklahoma State University, Department of Entomology and Plant Pathology, Stillwater, OK 74078

The tobacco hornworm (*Manduca sexta*) is employed as a model organism to study insect physiological processes including innate immunity and intermediary metabolism. Despite numerous works done in this species, a link between metabolic processes and innate immunity remains unclear and most metabolism-related genes are not yet discovered. Our recent quantitative transcriptome analysis in *M. sexta* yielded 19,020 contigs from fat body and hemocytes of naïve and microbe-injected fifth instar larvae, allowing us to explore abundance changes in transcripts of metabolic enzymes and their regulators. BLAST2GO and KEGG pathway analyses unraveled sets of genes involved in energy metabolism, fatty acid synthesis and

elongation, pentose phosphate pathway, and many others. We also identified genes for nucleotide metabolism, genetic information processing, protein synthesis, and amino acid metabolism. Marked elevation in mRNA levels of key enzymes in glycolytic pathway and tricarboxylic acid cycle suggested that the energy demand increased after the immune challenge. The notion was further supported by increases in mRNA levels for lipases and lipophorin receptors. Surprisingly, transcript levels of methionine-rich storage proteins showed the greatest increase. Other significant changes included the up-regulation of transcript levels for translation initiation factors, certain ribosomal proteins, and endoplasmic reticulum biosynthesis and fusion proteins. Also, genes encoding iron sequestering/storage proteins, heat shock proteins, P450s, esterases, and enzymes for juvenile hormone biosynthesis and subsequent degradation were up-regulated at the transcription level. These results are indicative of a strong connection between metabolism and immunity.

69

Further Analysis of Immunity-related cDNA Contigs from Hemocytes and Fat Body of *Manduca sexta* Larvae upon Microbial Infection

Gunaratna, Ramesh T.; Zhang, Shuguang; *Jiang, Haobo

Oklahoma State University, Department of Entomology and Plant Pathology, Stillwater, OK 74078

As a biochemical model, *Manduca sexta* has significantly contributed to our knowledge on insect innate immune system. The RNA-Seq approach based on massively parallel pyrosequencing was implemented in three studies to explore tissue immunotranscriptomes of this species. With the latest and largest one focusing on process- and tissue-specifically regulated genes, we analyzed the same set of sequence data using BLAST2GO, which provided functional perspectives of the larval fat body and hemocyte transcript-omes before and after the immune challenge. We have detected global differences in terms of cellular components, biological processes, and molecular functions of the CF, IF, CH, and IH contigs and read numbers. We further identified 825 immunity-related contigs and compared them with those discovered based on relative abundance changes and BLASTX results. The major overlap of these two lists validated our method developed for gene discovery and transcript profiling in organisms lacking sequenced genomes. On the other hand, we discovered 685 new contigs for proteins involved in various aspects of innate immunity, such as pathogen recognition, signal transduction, and killing of invading organisms. In summary, these studies set the stage for future analysis of the *M. sexta* immunogenome.

70

Annotation and Functional Analysis of Bacteria-Like Genes in the Red Flour Beetle

Klobasa, William A.; Saltzmann, Kurt D.; Beeman, Richard W.; Lorenzen, Marcé D.*

(WAK, MDL) North Carolina State University, Department of Entomology, Raleigh, NC 27695; (KDS) Purdue University, Department of Entomology, West Lafayette, IN 47907; (RWB) USDA-ARS-CGAHR 1515 College Ave. Manhattan, KS 66502.

The opportunity for gene neofunctionalization following a duplication event is of great significance to organismal evolution. The red flour beetle, *Tribolium castaneum*, is an excellent model for the study of gene duplication since it has a sequenced genome and a variety of powerful genomic tools to facilitate functional genetic analysis. Here, we examine a family of related genes in *Tribolium* that encode proteins highly similar to members of the bacterial PD-(D/E)XK nuclease superfamily. Each of these proteins contains an N-terminal AAA-ATPase like domain (pfam09820), while some also contain a bacterial Domain of Unknown Function (DUF) 1703 (pfam08011) near the C-terminus. Previously, these domains have been detected only in prokaryotic proteins, with the exception of one each in an amoebozoan and a filamentous fungus. We have detected members of this gene family for the first time in a higher eukaryote. Sequence similarity and phylogenetic analyses reveal the presence of at least five subfamilies within the *Tribolium* genome, suggesting lateral transfer from bacteria followed by multiple rounds of gene duplication. Our analyses include data for the three largest subfamilies, which we have tested for both expression and function. Full-length transcription units have been identified from each *Tribolium* life stage via 5' and 3' RACE. Functional analysis of these three subfamilies has been conducted at each life stage using RNA interference (RNAi). Results indicate that one gene subfamily has acquired a critical function in *Tribolium* as demonstrated by a lethal phenotype following transcript depletion. To ascertain whether off-target effects of the dsRNAs are responsible for insect lethality, we are designing transgenic constructs to rescue the lethal phenotype in this RNAi assay. These constructs will either have the 3'UTR replaced by that of a different gene, or will contain a fully functional copy of the targeted bacteria-like gene in which the wobble base of each codon has been mutated in a way that preserves the protein sequence, thus rendering the mRNA nonsusceptible to RNAi targeting directed at the endogenous transcript.

71

Gypsy Moth Larval Midgut Transcriptome Characterization and its Response to *Bacillus thuringiensis* Infection

*Sparks, Michael E.; Blackburn, Michael; Gundersen-Rindal, Dawn E.

USDA-ARS Invasive Insects Biocontrol and Behavior Laboratory, Plant Sciences Institute, Beltsville Agricultural Research Center, Beltsville, MD 20705

The gypsy moth, *Lymantria dispar*, is a significant defoliation threat to urban and rural regions of the United States. Efforts to contain and/or eliminate this pest have been largely unsuccessful, and evidence exists that its range is expanding at an accelerating rate. Bioinformatics analysis, coupled with the immense throughput of next-generation sequencing, offers a powerful means to investigate the genomic and transcriptomic characteristics of this insect. Here, we describe the production of an Illumina-based RNA-Seq library for gypsy moth, the delineation and functional annotation of protein coding genes, and the response of the gypsy moth midgut transcriptome to infection by *Bacillus thuringiensis* (Bt). We demonstrate that Bt infection results in substantial modifications to the gypsy moth midgut transcriptome program: 78 of 554 highest-quality gypsy moth genes exhibited at least a five-fold change in expression levels, 18 at least 20-fold and for five, 90-fold or higher.

72

Clip-serine proteinase CLIPB8 supplements the SRPN2/CLIPB9 regulatory unit that controls melanization in the African malaria mosquito, *Anopheles gambiae* s.s.

*Zhang, Xin; An, Chunju; Jones, Kara J.; Michel, Kristin
Kansas State University, Division of Biology, Manhattan, KS 66506

Melanization immune response encapsulates and kills invading pathogens in insects and other arthropods. Melanization is regulated by the activation of prophenoloxidase (PPO), which is controlled by a proteinase cascade and its serpin inhibitors. To date, the molecular composition of this system is only partially understood especially in mosquitoes. Recently, a regulatory unit of melanization in *Anopheles gambiae* was documented comprising an inhibitory serpin-clip-serine proteinase pair: SRPN2-CLIPB9. Partial reversion of SRPN2 phenotypes in melanotic tumor formation and adult survival by SRPN2/CLIPB9 double knockdown suggests other target proteinases of SRPN2 in regulating melanization. Here we report that CLIPB8 is identified as a target proteinase of SRPN2 and supplements the SRPN2/CLIPB9 regulatory unit in controlling melanization in *An. gambiae*. Heterologously expressed SRPN2 forms a complex with activated recombinant proCLIPB8 and directly inhibits CLIPB8 activity *in vitro*. Similar to CLIPB9, double knockdown SRPN2 and CLIPB8 also reversed the pleiotrophic phenotype induced by SRPN2 silencing, both in adult survival and melanotic tumor formation. Different from CLIPB9, CLIPB8 does not cleave *Manduca sexta* PPO *in vitro*. Biochemical analysis showed that CLIPB8 and CLIPB9 cannot activate each other *in vitro*, suggesting a second parallel pathway of CLIPB8 and CLIPB9 in activation of melanization. To test this hypothesis, experiments are being performed testing if CLIPB8 can activate melanization in *M. sexta* plasma *in vitro*. Additionally, reverse genetic analysis is carried out using triple knockdown of SRPN2, CLIPB8, and CLIPB9 to determine the degree of genetic interaction between the two proteases and their inhibitor.

73

Using Insect Pest Genomics to Identify New Molecular Targets for Insecticide Discovery and Plant Genetic Engineering

*Bailey, David; Bass, Chris; Caccamo, Mario; Colvin, John; Denholm, Ian; Febrer, Melanie; Gorman, Kevin; Hogenhout, Saskia; Kersey, Paul; Lawson, Dan; Overington, John; Seal, Susan; and Williamson, Martin.
(PK, DL, JO) European Molecular Biology Laboratory–European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SD, UK; (SH) John Innes Centre, Norwich NR4 7UH, UK; (DB, JC, SS) Natural Resources Institute, Chatham, Kent ME4 4TB, UK; (CB, ID, KG, MW) Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK; (MC, MF) The Genome Analysis Centre, Norwich NR4 7UH, UK.

A collaborative partnership has been formed between leading insect pest biology laboratories and genome centres in the UK to develop the genomic tools and technologies to support rational target discovery in insect pests, for use both in insecticide discovery and in plant genetic engineering. We are currently working on a number of hemipteran insect pests, including the whiteflies *Bemisia tabaci* and *Trialeurodes vaporariorum* (Aleyrodidae), the aphids *Acyrtosiphon pisum* and *Myzus persicae* (Aphididae), the brown planthopper *Nilaparvata lugens* (Delphacidae), and the maize leafhopper *Cicadulina mbila* (Cicadellidae). These species are chosen for their global economic importance both as direct pests and as virus-disease vectors on numerous arable and horticultural crops.

Entomology and Plant Genetic Engineering The Natural Resources Institute (NRI), Rothamsted Research (RR), and the John Innes Centre (JIC) have considerable and long-standing expertise in the biology of whiteflies (NRI, RR), leafhoppers (RR, NRI) and aphids (RR, JIC). We have also contributed to major advances in the understanding of insecticide mode of action and insecticide resistance development, and to international initiatives in genome sequencing and annotation, including the International Aphid Genome Consortium. We are also developing RNAi-based strategies for the use of genome information in insect gene function analysis and plant genetic engineering. Together, we are committed to using genomic and transcriptomic information in the development of new pest management strategies.

Genome Sequencing and Analysis The European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI) and the UK's Genome Analysis Centre (TGAC) are centres of excellence for next generation sequencing (NGS) data analysis, sequence annotation and bioinformatics. TGAC hosts a sequence facility with access to a broad range of NGS instruments. In addition to providing annotation for our pest genomics project through the automatic genome annotation and chemical biology databases Ensembl and ChEMBL, EMBL-EBI also supports the parallel Wellcome Trust and NIAID initiatives in mosquito biology, and in vector genomics (VectorBase). Comparative genomics is an important part of our project, in which both new pesticide targets and antitargets need to be defined.

Transcriptomics and Functional Genomics to Identify Novel Targets The advent of next generation sequencing (NGS) has revolutionised target discovery using transcriptomics, allowing a comprehensive analysis of genes governing both insecticide metabolism and resistance development at the expression and chromosomal sequence level. Rapid, side-by-side comparisons between target genes in pests and corresponding non-target genes in beneficial species such as the honeybee, *Apis mellifera* (Hymenoptera: Apidae), backed by RNAi-based genetic screens to validate specific molecular targets, will provide the basis of a robust target discovery pipeline.

Insecticide Resistance Mechanisms Our consortium has characterised numerous resistance mechanisms based on target-site modifications and enhanced detoxification. The latter can arise through transcriptional changes or gene amplification, making the development of rapid ways of accessing genome sequence information a key objective of our work.

Summary Our consortium has a strong interest in the sequencing, analysis and exploitation of new pest genomes. We are currently exploring the formation of international partnerships for the analysis and exploitation of the *B. tabaci* and *C. mbila* genomes.

74

Expressed Sequence Tags Profile in Starved and Blood Fed *Triatoma infestans* Midgut

*Buarque, Diego S.; Martins, Rafael M.; Tanaka-Azevedo, Anita, M.; Braz, Glória R C.; Tanaka, Aparecida S.

(DSB, AST) Universidade Federal de São Paulo, Departamento de Bioquímica, Rua Três de Maio, 100, São Paulo, 04044-020; (RMM) Universidade Federal de São Paulo, Departamento de Microbiologia, Imunologia e Parasitologia, Rua Botucatu, 862, 04023-062; (AMTA) Instituto Butantã, Laboratório de Herpetologia, Avenida Vital Brasil, 1500, São Paulo, 05503-900; (GRCB) Universidade Federal do Rio de Janeiro, Instituto de Química, Departamento de Bioquímica, Av Brigadeiro Trompowsk, s/n, Rio de Janeiro, 21949-900.

The blood-sucking bug *Triatoma infestans* is one of the main vectors of *Trypanosoma cruzi*, a protozoan parasite and etiological agent of Chagas' disease (American Trypanosomiasis) in Latin America. *Trypanosoma cruzi* resides in triatomine insects' intestinal tract, which is considered an immune competent tissue. In addition, midgut is also responsible for the insect's digestion. Thus, the aim of this work was to analyze expression profile at RNA level in starved and blood fed *Triatoma infestans* midgut to identify molecules expressed in response to blood feeding. Total RNA was collected from midgut of adult *T. infestans* starved and 24 h after blood feeding on mice. cDNA libraries from *T. infestans* midgut were constructed and 244 and 551 high quality expressed sequence tags (ESTs) were obtained in starved and blood fed midgut libraries, respectively. The sequences were assembled in contigs (unfed – 137 contigs; fed – 291 contigs) and blasted (e-value: 10^{-3}) against several databases: Conserved Domain Database (CDD), Protein families (Pfam), Gene Ontology Consortium (GO), Orthologous Eukariotic Domains (Kog), Non redundant database (NR), Ribosomal RNA (rRNA), Simple Modular Architecture tool (Smart), Swissprot and Mit-pla. Our results showed that 15% and 23% of ESTs obtained from unfed and blood fed libraries had no similarity to sequences available in databases. Among the sequences with BLAST hits, about 25% of the sequences were related to possible secreted proteins in both libraries. We detected the presence of sequences related to aspartyl proteases; the cysteine proteases inhibitor, cystatin; infestin, an inhibitor of blood coagulation, and immune-related peptides, such as defensins in starved and fed insects. On the other hand, it was possible to observe the expression of some transcripts only in blood fed insects. These sequences are related to lectin; lipocalins, such as fatty acid binding and triabin; peritrophins and the antioxidant enzyme thioredoxin reductase. Our findings suggested that some transcripts can be expressed in response to blood feeding. Further studies of functional characterization at protein level could provide information about molecules present in the intestinal tract of these insects. Thus, cDNA sequencing can be helpful to understand blood feeding process in insect vectors and to identify novel active compounds.

Financial Support: FAPESP, CNPq, INCT-EM.

75

Transcriptional Response to Manipulation of Fungal Symbiosis in Leaf-Cutter Ant, *Atta cephalotes*

*Grubbs, Kirk; Suen, Garret; Teiling, Clotilde; Li, Lewyn; Cavanaugh, Amy; Marsh, Sarah; Moeller, Joseph; Currie, Cameron (KG, GS, AC, SM, JM, CC) Department of Bacteriology, University of Wisconsin-Madison, Madison, Wisconsin; (KG) Cellular and Molecular Pathology Graduate Program, University of Wisconsin-Madison, Madison, Wisconsin; (GS, JM, CC) Department of Energy, Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, Wisconsin; (CT) Roche Diagnostics, Indianapolis, Indiana; (LL) 454 Life Sciences, Branford, Connecticut; (AC) Department of Biology, University of Wisconsin-Rock County, Janesville, Wisconsin; (SM) Department of Genetics, University of Wisconsin-Madison, Madison, Wisconsin

Leaf-cutter ants represent an intricate system of symbiosis, which is comprised of a minimum of five well described, interacting members. The longest recognized and possibly most basic interaction in this system is that of the ants and their fungal cultivar. In this relationship, which is evidenced to be up to 50 million years old, ants forage for leaves which are processed into manure for their fungal gardens. In exchange, the fungal garden produces hyphal swellings known as gongylidia which serve as the main nutritional resource for the ants. In order to better understand this relationship and how it affects the other symbiotic interactions of the system we must ascertain its genetic underpinnings. To this end we seek to study transcriptional responses of leaf-cutter ants to specific manipulations of the ant/fungal cultivar interaction.

The transcriptional response of *Atta cephalotes* with their native cultivar, *A. cephalotes* with cultivar from *Acromyrmex octospinosus* and *A. cephalotes* with a glucose solution was sequenced using 454-titanium technology. The resulting data was analyzed in two veins. First, a basic microbial community profile associated with the ants of each trial was produced by BLAST comparison of each read against the NR database and keeping the top-hit (minimum cutoff $1.0E-05$) as an indicator of genus. Notably, after taking into account differences in total sequencing real estate, the control treatment had a bacterial community that was almost twice as large as the other treatments and the bacterial communities of all treatments were largely composed of isolates that are associated with gut flora. Further support for these results was garnered through the use of the phylogenetic binning program RAlphy.

The second line of analysis was to assess and describe gene expression variation between treatments. Initially, GS Reference Mapper software was used to align reads to the annotated *A. cephalotes* genome and obtain a total expression count for each gene. Expression profiles were then TMM normalized with edgeR to account for differences in total reads between treatments. Treatment profiles were then compared to the control in order to obtain differential expression profiles. Initial KOGG ID assignment to genes considered significantly differentially regulated revealed several basic functional groups that were either over- or under-expressed. Further analysis was pursued using GO terms assigned by InterproScan because of low

assignment percentage of genes from the KOGG analysis. Each of the treatment profiles were found to be under-represented for olfactory receptor activity, sensory perception of smell and odorant binding. The glucose treatment was the only treatment found to be significantly over-represented by any GO terms and those corresponded to methyltransferase activity and translation. We are currently in the process of replicating these experiments to bolster statistical power and further refine the suite of genes to be further investigated as playing major roles in the ant/fungal cultivar symbiosis

76

Recycling and de novo synthesis: the *Buchnera*-aphid collaboration for amino acid metabolism

*Hansen, Allison K.; Moran, Nancy A.

Yale University, Department of Ecology and Evolution, West Haven, CT 06516

Animals cannot synthesize their own essential amino acids (EAA) and usually acquire them from their diet. Aphids, which feed on plant sap, a diet low in EAA, obtain EAA from their obligate bacterial endosymbiont (*Buchnera*), which lives inside specialized aphid cells (bacteriocytes). *Buchnera*'s genome encodes genes for all EAA pathways; however, several genes in EAA pathways are missing. How these amino acids are synthesized is unknown. Also, it is unclear if aphids synthesize NEAA which *Buchnera* is unable to provide. Finally, mechanisms by which the symbiotic partners obtain sufficient total nitrogen are unknown. In this study, we compared pea aphid gene expression between aphid bacteriocytes and other aphid body tissues using RNA sequencing and pathway analysis based on the fully sequenced genomes of both partners. First, 26 amino acid related genes are up-regulated in aphid bacteriocytes. Seven of these correspond to enzyme activities that fill in the gaps of *Buchnera*'s EAA and tyrosine pathways. In addition, genes for five NEAA pathways lost from *Buchnera*, were up-regulated in bacteriocytes. Finally, our results reveal that two genes, glutamine synthase (GS) and glutamate synthase (GltS), which work together in the GOGAT cycle to assimilate ammonia into glutamate, are both up-regulated in bacteriocytes. The GOGAT cycle may be a key mechanism fueling this nitrogen-limited integrated amino acid metabolism by recycling waste ammonia into EAA and NEAA.

77

Identifying phosphine resistance genes in *T. castaneum* by high throughput sequencing and bulk segregant analysis of SNPs

Jagadeesan, Rajeswaran; Ebert, Paul R.; [†]Schluppius, David I.

(RJ, PRE) School of Biological Sciences, University of Queensland, St. Lucia, QLD, Australia 4072; (DIS) Agriscience Queensland, Department of Employment, Economic Development and Innovation, Ecosciences Precinct, GPO Box 46, Brisbane, QLD, Australia 4001, Cooperative Research Centre for National Plant Biosecurity, LPO Box 5012, Bruce, ACT 2617, Australia

Phosphine resistance presents a significant challenge to pest management in stored products. In order to be able to monitor and combat resistance to this important fumigant, we are identifying the genes responsible for resistance in the pest species, *Tribolium castaneum*. *T. castaneum* is an ideal model for phosphine resistance studies, as strains strongly resistant to phosphine have been identified and there is a completed genome available. We have mapped short read sequences from selected (strongly resistant) and unselected progeny (F_4 and F_{19}) of segregating crosses (Susceptible X Resistant) and compared the average homozygosity across all chromosomes and unplaced scaffolds using a rolling window that is defined by the SNP density, rather than the physical distance on the chromosome. In this way we have identified two loci that segregate for resistance in *T. castaneum*. SNPs in these regions have been validated for their linkage to resistance and are being used for fine scale localisation of the resistance genes. We have found this method can quickly scan the entire genome for linkage to a trait, including unplaced scaffolds. We have also found that this method is relatively insensitive to errors in sequencing and SNP calling, chromosomal rearrangements as well as large gaps, assembly errors and fragmentation in the reference genome.

78

Translational Insect Genomics: Pheromone Traps and Molecular Markers Genotype Hessian Flies For Virulence To Resistance Gene *H13* In Wheat

Johnson, Alisha J.; Buntin G. David; Flanders, Kathy L.; Reay-Jones, Francis P. F.; Reisig, Dominic D.; Schemerhorn, Brandon J.; ^{*}Stuart, Jeffrey J.; Shukle, Richard H.

(AJJ, BJS, RHS) USDA-ARS, Purdue University, West Lafayette, IN 47907; (DGB) Department of Entomology, University of Georgia, Griffin, GA 30223; (KLF) Department of Entomology Auburn University, Auburn, AL 36849; (FPFR-J) Department of Entomology Clemson University, Clemson, SC 29631; (DDR) Department of Entomology North Carolina State University, Raleigh, NC 27607; (JJS) Department of Entomology, Purdue University, West Lafayette, IN) 47907

In the southeastern United States, the Hessian fly is the major insect pest of wheat and causes significant yield losses to the region. Hessian fly is primarily controlled through the use of resistant wheat cultivars that carry resistance (*R*) genes. Wheat containing the *R* gene *H13* has been found to provide effective protection against Hessian fly attack in the Southeast. However, successive yearly deployment of resistant wheat lines selects for "virulent" genotypes that are able to overcome resistance. Thus, single *R* genes typically lose their initial effectiveness in 7 to 10 years. In this study, we used a map-based genomics approach to identify the mutations that allow Hessian flies to overcome the resistance conferred by the Hessian fly resistance gene *H13*. We then used pheromone traps to collect Hessian fly males from field populations across North Carolina, South Carolina, Georgia, and Alabama and used PCR to test for the presence of the mutations in those populations. The approach was diagnostic for the frequency of Hessian fly *H13*-virulence and -avirulence in each population. Thus, field populations can be monitored regularly to survey the efficacy of *H13*'s ability to protect wheat and inform both growers and

breeders with regard to the gene's effectiveness in relatively small geographic areas. As the mutations conferring virulence to additional *R* genes are identified, we expect this method will serve as a rapid and economical diagnostic that will replace current procedures, which require collecting and rearing live insects on resistant wheat lines.

79

Behavioral and Proteomic Comparisons of Aphid Responses to Host Plant Resistance in Tomato

Pallipparambil, G.R.; Reese, J.C.; Francis, F.; Jia, L.; Goggin, F.L. *

(GRP, LJ, FLG) University of Arkansas, Department of Entomology, Fayetteville, AR 72701; (JCR) Kansas State University, Department of Entomology, Manhattan, KS 66506; (FF) Gembloux Agricultural University, Functional & Evolutionary Entomology, Gembloux, Belgium.

The *Mi-1.2* gene in tomato (*Solanum lycopersicum*) confers resistance against the potato aphid, *Macrosiphum euphorbiae*, but different potato aphid populations vary widely in their virulence, or their ability to overcome resistance. For example, *Mi*-mediated resistance reduces the population growth of one avirulent potato aphid clone (isolate WU11) by ~95%, but causes only a 15% reduction in the semi-virulent WU12 isolate. To investigate this difference in response between WU11 and WU12 aphids, their feeding behavior on resistant and susceptible tomato cultivars was compared using the direct-current electrical penetration graph (DC EPG) technique. DC-EPG analysis shows that both aphid clones take longer to initiate cell sampling and to establish a confirmed sieve element phase on resistant plants (*Mi-1.2+*) than on susceptible hosts (*Mi-1.2-*), and have shorter ingestion periods on resistant plants. However, the magnitude of these deterrent effects is far less for the semi-virulent clone than for the avirulent aphids. In particular, the WU12 clone is less sensitive to factors that limit sieve element ingestion, showing shorter non-probe duration and rapidly establishing sustained phloem ingestion on resistant plants when compared to the WU11 clone. We conclude that, in addition to previously described factors in the phloem that inhibit ingestion, *Mi*-mediated aphid resistance also involves factors (possibly in the mesophyll and/or epidermis) that delay initiation of phloem salivation, and that act in the intercellular spaces to deter the first cell sampling. Furthermore, the relative effectiveness of these components of resistance differs among insect populations. We also examined the impact of exposure to resistant plants on the protein profiles of the two aphid isolates using two-dimensional difference gel electrophoresis (2D DIGE) coupled with protein identification by MALDI-TOF-MS. Eighty-two protein spots showed significant quantitative differences among the four treatment groups (WU11 and WU12 on resistant and susceptible plants), and of these, 48 could be assigned putative identities. Nearly 25% of the differentially regulated proteins originated from aphid endosymbionts and not the aphid itself. Six were assigned to the primary endosymbiont *Buchnera aphidicola*, and 5 appeared to be derived from a *Rickettsia*-like secondary symbiont. Furthermore, PCR screening indicated that the semivirulent WU12 isolate carried a *Rickettsia*-like secondary symbiont that was absent in the avirulent WU11 isolate. These results indicate that expression patterns in both primary and secondary symbionts are influenced by the aphids' host plant, and suggest that symbionts may contribute to differential adaptation of aphids to host plant resistance.

80

A viral histone H4 alters insect gene expression via an epigenetic mode

Prasad, Surakasi Venkata; Kim, Yonggyun

Department of Bio-Sciences, Andong National University, Andong, Korea

A viral histone H4 is identified from an episomal genome of *Cotesia plutellae* bracovirus (CpBV). In contrast to histone H4s that are highly conserved in their amino acid sequences among eukaryotes, the viral histone H4 (CpBV-H4) possesses an N-terminal extra tail (38 residues). Transient expression of CpBV-H4 to the diamondback moth, *Plutella xylostella*, causes significant physiological alterations. However, a truncated CpBV-H4 prepared by removing the N-terminal tail loses the inhibitory activity. Pull-down experiment of nuclear protein extract of *P. xylostella* larvae using a purified recombinant CpBV-H4 possessing His-tag, the viral histone H4 were associated with host nucleosome components. This study was focused on transcriptional regulation of CpBV-H4 against insect gene expression. Nonparasitized *P. xylostella* larvae were micro-injected with a recombinant expression vector containing CpBV-H4 under a baculoviral immediately early promoter. After a high stringent suppressive subtractive hybridization (SSH) of cDNAs prepared from the treated larvae against those from untreated larvae (nonrecombinant vector injection), the resulting cDNAs were sequenced with 454 pyrosequencing and produced 493 isogroups, 610 isotigs, and 463 singletons. Gene ontology data illustrated that the majority of the genes inhibited by expression of CpBV-H4 are involved in growth, metabolism, transcriptional factors, and immune responses. When the target genes were mapped on the three known genomes of *Drosophila melanogaster*, *Apis mellifera*, and *Tribolium castaneum*, the target genes were scattered on the chromosomes. Interestingly, the target genes appear to be clustered around transcriptional factors that are inhibited by CpBV-H4. These results suggest that a viral histone H4 can regulate host gene expression via an epigenetic mode.

81

Strategies for development of RNAi-mediated insect resistant crops

*Ramaseshadri, Partha; Bolognesi, R; Segers, G.

Monsanto Co. , 700 Chesterfield Parkway West, St Louis, MO, 63017.

Environmental uptake of double stranded RNA (dsRNA) has been demonstrated to occur in a number of invertebrate species including the pest western corn rootworm (WCR, *Diabrotica virgifera virgifera*, LeConte). The effect of dsRNA as a control agent for rootworm can be varied by choice of rootworm target gene and specific gene segment. Variation in target gene efficacy observed in artificial diet has been extended to in planta feeding studies in transgenic corn plants expressing inverted repeats. Targets with lower LC50s have created events reduced damage after WCR infestation. Sequence variation in rootworm populations is also being assessed to aid in target selection. Some genes show lower sequence diversity amongst

target rootworm populations and thus provide opportunities for enhancing durability of control. Combinations of RNAi and *Bacillus thuringiensis* protein toxins (Bt protein) provide two modes of action for prevention of insect resistance. The impact of two modes action is not unique to a particular RNAi/Bt protein combination. The RNAi component appears to bring a potent component to the combination as noted in reduction of WCR adult beetle emergence.

82

Towards the analysis of transcriptome from sexed *Ceratitis capitata* embryos: how to produce XX male embryos by transgene-mediated RNAi and identify male biased/male determining genes.

Salvemini¹, M., Sanges³, R., Nagaraju⁴, J., D'Amato¹, R., Petrella¹, V., Pane¹, A., Polito^{1,2}, L. C., and *Saccone¹, G.

¹Dipartimento delle Scienze Biologiche, Sezione Genetica, Università degli Studi di Napoli Federico II, Napoli, Italy.

²Istituto di Genetica e Biofisica "Adriano Buzzati Traverso, CNR, Napoli, Italy.

³Stazione Zoologica Anton Dohrn, Napoli, Italy.

⁴Laboratory of Molecular Genetics, Centre for DNA Fingerprinting & Diagnostics, Hyderabad, India.

In the agricultural pest insect *Ceratitis capitata*, the Mediterranean fruitfly, (Medfly) the sex determination is controlled by a Y-linked male determining factor which influences either directly or indirectly the state of activity of the female determining master gene *Cctra*^{ep} (*Cctra epigenetic*) at 5-7 hours of oviposition. We have developed a *Ceratitis* transgenic sexing strain which produces male-only progeny (95% efficiency) by transgene-mediated RNAi against the female determiner *Cctra*^{ep} gene. *C. capitata* XX males are fully fertile and could be successfully mated with XX females thus leading to the production of only female (XX) progeny. We have prepared polyA+ RNA from XX embryos and from mixed XX/XY embryos, both collected at 5-7 hours from oviposition. We have used a molecular subtractive approach to identify differentially expressed genes in XY versus XX at embryonic stages. We will outline the molecular strategy employed to approach this scientific problem and preliminary data describing the identification of 8 male-biased cDNA positive clones presently under analysis. We started in parallel a transcriptome analysis of the two samples to approach the identification of male-biased genes/male-determining genes by *in silico* subtractive analysis and the ongoing experimental analysis will be described.

83

Host race evolution in *Schizaphis graminum* (Hemiptera: Aphididae): nuclear DNA sequences

Shufran, Kevin A.

USDA-ARS, Wheat, Peanut, and Other Field Crops Research Unit, 1301 N. Western Rd., Stillwater, OK 74075, USA

The greenbug aphid, *Schizaphis graminum* (Rondani) was introduced into the US in the late 1880's and it established quickly as a pest on wheat, oat and barley. Sorghum was also a host, but it was not until 1968 that greenbug became a serious pest on it. The most effective control method is the planting of resistant varieties; however, the occurrence of greenbug biotypes has hampered the development and use of plant resistance as a management technique. Until the 1990s, the evolutionary status of greenbug biotypes was obscure. Four mtDNA *cytochrome oxidase subunit I* (COI) haplotypes were previously identified, suggesting that *S. graminum sensu lato* was comprised of host-adapted races. To elucidate the current evolutionary and taxonomic status of the greenbug and its biotypes, two nuclear genes and introns were sequenced; *cytochrome c* (CytC) and *elongation factor 1- α* (EF1- α). Phylogenetic analysis of CytC sequences were in complete agreement with COI sequences and demonstrated three distinct evolutionary lineages in *S. graminum*. EF1- α DNA sequences were in partial agreement with COI and CytC sequences, and demonstrated two distinct evolutionary lineages. Host-adapted races in greenbug are sympatric and appear reproductively isolated. Agricultural biotypes in *S. graminum* likely arose by genetic recombination via meiosis during sexual reproduction within host-races. The 1968 greenbug outbreak on sorghum was the result of the introduction of a host race adapted to sorghum, and not selection by host resistance genes in crops.

84

Virus-Induced Gene Silencing of Wheat Genes for Resistance to *Diuraphis noxia* (Hemiptera: Aphididae)

STARKUS, L*; MURUGAN, M; and SMITH, C.M.

(L. S. and C. M. S.) Department of Entomology, Kansas State University, Manhattan, KS, 66506 USA

(M.M.) Department of Plant Molecular Biology and Biotechnology, Centre for Plant Molecular Biology, Tamil Nadu Agricultural University, Coimbatore 641003, Tamil Nadu, India

The Russian wheat aphid *Diuraphis noxia* (Kurdjumov) is a major pest of cereal grains and causes plant damage such as leaf rolling. The pseudo gall created by leaf rolling protects aphids and makes it difficult to treat infested plants with insecticides or biological control agents. The identification of new sources of barley and wheat resistance is necessary because of the development of virulent *D. noxia* biotypes. Virus-induced gene silencing (VIGS) is one method for screening candidate genes for resistance. The VIGS method was tested to determine the ability of *barley stripe mosaic virus* (BSMV) to serve as a VIGS vector in wheat plants containing the *Dnx* gene for resistance to *D. noxia*. Results indicated that BSMV VIGS silencing does not affect the resistance of the *Dnx* plants. Microarray and gene expression data were used to select the NBS-LRR type disease resistance protein genes TaAffx.104814.1.S1_at and TaAffx.28897.1.S1 - (NBS-LRR1 and NBS-LRR2, respectively), in order to assess their role in *Dnx* resistance. Aphids were allowed to feed on control and treated plants to assess *D. noxia* population development and mean weights of aphids surviving at the end of the experiment. There were no differences in *D. noxia* populations on NBS-LRR1 or NBS-LRR2 silenced leaves and control leaves, but mean *D. noxia* weights on NBS-LRR1 silenced plants were significantly greater than those on control *Dnx* plants or NBS-LRR2 silenced plants. *D. noxia* weights on NBS-LRR1 silenced plants were similar to weights of aphids from susceptible control plants.

85

Behavioral Resistance to *Varroa* mites: Two Genome-Wide Genotyping Studies in Honey Bees

*Tsuruda, Jennifer M.; Harris, Jeffrey W.; Bourgeois, Lanie; Danka, Robert G.; Arechavaleta-Velasco, Miguel E.; Alcalá-Escamilla, K. Itzel; Robles-Rios, Carlos; Hunt, Greg J.

(JMT, GJH) Purdue University, Department of Entomology, West Lafayette, IN 47907; (JWH, LB, RGD) USDA-ARS Honey Bee Breeding, Genetics and Physiology Lab, Baton Rouge, LA 70820; (MEA-V, KIA-E, CR-R) Nacional de Investigación en Fisiología y Mejoramiento Animal, INIFAP, km 1 Carretera a Colon, Ajuchitlan, Queretaro, Mexico

Varroa mite parasitism of honey bees is widely considered to be the greatest threat to beekeeping and has led to colony losses worldwide. *Varroa* not only feed on the hemolymph of developing and adult bees, but they are also capable of vectoring viruses. Typically, chemical treatments are utilized but the mites have developed resistance so alternative control methods are desired. Two behaviors have been identified as important traits for controlling the growth of *Varroa* mite populations in bee hives – grooming and *Varroa*-sensitive hygiene (VSH). When bees groom themselves or each other, they attempt to brush mites off their bodies and may chew the mite. Bees exhibiting VSH uncap and/or remove pupae from infested cells, thereby suppressing mite reproduction. We conducted one study for each of trait to look for associations between phenotype (grooming or VSH behavior) and genotype by mapping quantitative trait loci (QTL). In each project, high and low lines were crossed to produce F1 hybrid queens, which were then backcrossed to produce the mapping populations of worker bees. The queens were sequenced and aligned to the honey bee genome to identify single nucleotide polymorphisms (SNPs). Probes for 1536 SNPs were designed and 240 individuals were genotyped for each study to construct high-resolution genetic maps. Interval mapping was used to find QTL and candidate genes were identified. Gene expression levels of candidate genes will be investigated and gene knockdowns will be performed to verify gene function and association with *Varroa* resistance. These results may aid in the development of diagnostic measures for the traits and may be incorporated into honey bee breeding programs.

86

Gene expression profiles of Bt-resistant and susceptible European corn borer larvae (*Ostrinia nubilalis*) after ingestion of transgenic Cry1Ab corn leaves

*Yao, Jianxiu; Buschman, L Lawrent L; Zhu, Kun Yan

Kansas State University, Department of Entomology, Manhattan, KS 66506

Transgenic corn expressing insecticidal Cry1Ab from *Bacillus thuringiensis* (Bt) is very effective in controlling the European corn borer (ECB), *Ostrinia nubilalis*. However, it is necessary to understand how the ECB gut genes respond to transgenic corn because there is a threat the ECB could develop resistance to Cry toxins. In this study, we developed a cDNA microarray built upon the ECB gut-specific EST database. The cDNA microarray contains 15,000 cDNA elements representing 2,895 different genes. We used this microarray to construct the gene expression profiles for the early fourth-instar ECB larvae of the lab-selected resistant (Sky) strain and the susceptible (Meads) strain. Both strains were exposed to transgenic Cry1Ab corn (MON810-event) for 6-hrs. We identified 398 genes from the S strain and 264 genes from the R strain with a significantly increased or decreased expression (≥ 2.0 fold; p -value ≤ 0.05) in comparing with unexposed larvae. Among those genes, 154 in the S strain and 88 in the R strain had putative molecular functions based on gene ontology analysis. 45 differentially expressed genes were common between the R and S strains. Interestingly, two aminopeptidase genes (EST id: contig1398 and contig4776) were up-regulated in the R strain, and down-regulated in the S strain. The rest of the genes had similar regulation in the two strains. Especially, 17 genes in the S strain and 9 genes in the R strain had differential expression, which are potentially involved in Bt toxicity or/and Bt resistance. This study is monitoring of ECB gut transcripts for larvae exposed or unexposed to transgenic corn expressed Cry toxins. This study also provides a platform for further functional studies of transgenic corn-insect interactions.