

FOURTH ANNUAL Arthropod Genomics SYMPOSIUM

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

June 10-13, 2010 • Kansas City

POSTER ABSTRACTS

60

The transcriptome of *Phlebotomus papatasi*

*Abrudan, Jenica L.; Collins, Frank H.

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

Phlebotomus papatasi is one of seventy species belonging to Psychodidae, subfamily Phlebotomine that transmit disease to humans. 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. In spite of the medical importance, Phlebotomine sand flies species remain understudied and few genetic resources are available. The majority studies generating sand fly information investigate specific questions regarding vector-human interactions (EST analysis of salivary glands) or vector-parasite interactions (EST analysis of infected midgut); only one study has performed a global gene discovery analysis of a sand fly, *Lutzomyia longipalpis*. Here we complement this original study by analyzing the transcriptome of another sand fly: *Phlebotomus papatasi*. A normalized library was created using whole body *Ph. papatasi* (Israeli strain) sand flies. Total RNA was collected from the four larval stages, pupa, male and female at 1, 3, and 10 days post emergence. Additional RNA was included from females 6, 12, 24, 36, 48, 72, 94 and 120 hours post feeding on uninfected and *Leishmania major* infected mouse blood. Library construction was completed by Expressed Genomics and EST sequencing was performed by the Genome Center at Washington University, St. Louis. The resulting 37,487 high quality reads were assembled into 17,120 unique sequences (6,187 contigs and 10,993 singlets) using CAP3 with an average length of 550bp, range 100 - 3,340bp. Of these sequences 4,286 (25%) had no similarity to proteins available in NR or in UniProt databases, when searched for using BLAST (blastx, e-value: 10^{-5}), making them potentially sand fly specific proteins. Among the sequences with BLAST hits, 26% had a *Aedes aegypti* sequence as best hit. Functional annotation was done using GO terms with the blast2GO algorithm: 8,837 (50%) 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%). A comparison with similar data from *Lu. longipalpis* has revealed similar distribution for the GO subcategories (catabolic process, 13%; nucleotide binding, 16%; protein complex 19%) and specie of the best BLAST hit (26%). 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*.

Acknowledgements : I would like to thank Mary Ann McDowell and Neil Lobo for their assistance in the project.

15

The UDP-glycosyltransferase multigene family in *Helicoverpa armigera*

*Ahn, Seung-Joon; Vogel, Heiko; Heckel, David G.

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

Glycosylation in insects plays an important role in several processes, including xenobiotic detoxification, cuticle formation, pigmentation and olfaction. UDP-glycosyltransferase (UGT) enzymes detoxify reactive compounds and increase their solubility by transferring sugars from UDP-glycosides to a variety of small lipophilic aglycones. The UGT multigene family is composed of many paralogous genes, of which we identified 44 from *Helicoverpa armigera* (Lepidoptera: Noctuidae). A BLAST search of *B. mori* UGT protein sequences against our local *H. armigera* transcriptome database, consisting of EST sequences generated by both Sanger and 454-pyrosequencing of cDNAs from all developmental stages and tissues, revealed more than 160 matching contigs. These cDNA-derived contigs were used to BLAST against the *H. armigera* genomic dataset produced by 454-pyrosequencing and more than 300 genomic contigs were identified. Based on both the cDNA and genomic DNA searches we could identify 44 different UGT genes in *H. armigera*. All of the *H. armigera* UGT genes we found contained introns. The most common exon/intron structure is shown by genes with 3 introns and a very long (ca. 820 bp) first exon (28 genes), although there are other genes with 4, 5, 7 or 8 introns. A

multiple alignment and subsequent comparison of UGT protein sequences showed a high conservation in the C-terminal region (ca. 250 aa), as expected from its role as the common binding site to UDP-glycoside. The UGT signature motif in the C-terminus (ca. 30 aa) is highly conserved throughout viruses, bacteria, plants and animals. We constructed a consensus phylogenetic tree of *H. armigera* and *B. mori* UGT proteins with the neighbor-joining method using the conserved C-terminal region. Five major groups were defined with high bootstrap support. The largest group comprised of 19 *H. armigera* UGT genes showed higher within-group similarity (>70%) than the other groups, suggesting recent lineage-specific gene duplication events in this group. This group-specific gene family expansion might be driven by diversification in the N-terminus substrate binding domain, increasing the range of compounds that could be detoxified by glycosylation in this insect. Being highly diversified as well as being conserved in different regions, the UGT genes might shed light on the evolution of multigene families in combination with host adaptation in this generalist herbivorous insect.

43

The case for developing a lacewing genetic model organism

Allen, Margaret L.

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

Lacewings (Chrysopidae: Neuroptera) are a family of insect predators, also called aphidlions because of their voracious feeding on aphids. While lacewings 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 worth genetic investigation. The recent completion of the aphid genome almost demands 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 lacewing, *Chrysoperla rufilabris* is an overlooked insect model organism, and should be added to the list of organisms scheduled for rigorous genetic study and genomic sequencing. Lacewings belong to the order Neuroptera which has no representatives in nuclear genome projects. The order is genetically represented by very few nucleotide, EST, or protein sequences. Therefore a genetic model in this order would add substantially to the resources available to the systematics, phylogenetics, and evolution communities. The order is primitively holometabolous and should be studied as an evolutionary transition organism from the hemimetabolous insects (such as the pea aphid) to the apex holometabolous orders (Diptera, Lepidoptera, Coleoptera, and Hymenoptera). 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.

Lacewings may be ideal candidates as subjects of genetic study. They are amenable to mass rearing. The techniques used in mass rearing can be scaled down to small-scale rearing quite easily. While the lifecycle is not as short as *Drosophila*, under ideal conditions a generation of lacewings is about 30 days, and most species are multivoltine, requiring no diapause in culture. They fare well in small spaces if reasonably good humidity can be maintained and using a standard dark: light cycle (we use 8:16). Eggs of most lacewings, and certainly the commercially produced species, are laid individually on almost any surface, continually during the fecund portion of the female's adult lifespan, and have stalks. Thus they are both plentiful and easy to harvest. While eggs are not entirely transparent, they are sufficiently translucent to observe embryo development. Anterior and posterior are easily observed by the presence of stemmata or eyespots in later embryonic development. The egg stalk is attached at the posterior end of the egg, and there is a distinctive micropyle at the anterior end of the egg providing visible external reference at the very earliest possible developmental timing. Dorsal and ventral can also be identified on the developing egg because the embryonic germ band is easily identifiable. The chorion is soft and can be pierced with a microinjection needle. Embryonic development is rather slow, taking around 7 days. This would make the organism useful for developmental biologists interested in gene expression and signaling pathways. Larvae are not transparent, but can be handled easily. There are multiple larval instars (generally 3) leading up to a cessation of feeding, cocoon spinning and pupation. Pupae can be separated from the cocoon, using sufficient dexterity, and the pupa is moderately translucent. The adult stage is also translucent. Male and female adults can be easily identified and separated prior to mating for individual pair matings. Furthermore, the adult lifespan is long enough (often two months or more) that mating a parent to offspring is also possible. Adults fly, but weakly and are generally easy to handle. Once fertilized a gravid female produces eggs continually, sometimes for >30 days, producing up to 1200-1400 eggs. At least one highly inbred culture is available from our USDA ARS laboratory.

16

Molecular Evolution of Glutamate Ionotropic Receptors in 12 *Drosophila* Genomes

Almeida, Francisca C. *; Campos, José L.; and Rozas, Julio

Universitat de Barcelona, Facultat de Biologia, Dept de Genetica, Av. Diagonal 645, Barcelona, 08028, Spain

Olfaction, or the detection and recognition of airborne molecules, is an important mechanism animals use to perceive their environment. Success in living and reproduction are highly dependable on a proper identification of environmental clues such as for localization of food sources and identification of possible mates. In insects, olfaction was thought to be mediated mainly by two types of proteins: odorant-binding proteins (OBPs) and olfactory receptors (ORs). An analysis of ORs expressed in olfactory sensory neurons (OSNs) of *Drosophila* showed that not all sensory cells express these receptors, even though electrophysiological analysis revealed that these cells are sensitive to several types of olfactorial stimulus. These findings suggested that additional unidentified

olfactory receptors were probably present in *Drosophila*. Following this evidence, a bioinformatic approach was used to identify new candidate olfactory genes in *D. melanogaster*. This approach led to the discovery of a gene family with 67 genes related to the ionotropic glutamate receptors. These genes were dubbed ionotropic receptors (IR) and a major expression in the coeloconic OSNs of the antenna was confirmed for 15 of them.

While evolutionary aspects of OR and OBP genes have been studied in the genus *Drosophila* as well as in other insects, little is known about patterns of number of genes, gene conservation, dynamics of gene gain and loss, and evolutionary forces acting on IRs. Understanding the evolutionary processes governing a gene family allows for the proposition of hypothesis concerning function and ecological importance of the different genes in the family. A temporal perspective can be achieved by the inclusion of other species orthologs in the study of these processes. Here we present a study on the evolution of IRs in the 12 *Drosophila* species with complete genome data.

To identify the complete IR family in the 12 *Drosophila* species, several rounds of blast searches were done using first previously identified *D. melanogaster* IR genes and subsequently increasing the number of queries with genes identified in the previous searches. Both translation databases (of predicted genes) and raw genome sequences were used as subjects. HMM searches were also performed, using two profiles based on two major IR clusters. Results were gathered in gff tables and predicted gene annotations were manually inspected. Some new candidate genes, not predicted by automated methods, were identified in this process. More than 30% of all genes had some manual edition done. Several predicted genes were in fact two in tandem genes. Orthology assignment of the 937 revised genes was done in two steps. First, we obtained large clusters based on blast scores using the program CLANS. These clusters were aligned using MAFFT and a phylogenetic analysis was carried out on the alignment using RAXML. Sixty-nine ortholog groups were determined based on the resulting tree clades. These groups included not only IRs, but also closely related ionotropic glutamate receptors, which are in fact subfamilies of IRs.

We investigated the dynamics of gain and loss of IR genes, using three methods: the one implemented in program CAFE that assumes equal rates of gain and loss; a perl script that allows for independent estimates of gain and loss; and manual estimation based on individual phylogenetic tree for each ortholog group. Gene birth and death rates were obtained using previously published divergence times for the genus *Drosophila*. To study the evolutionary forces acting on IR genes, we estimated rates of synonymous and non-synonymous substitution using PAML. Different models were used to determine the best selection regime that fit the different ortholog groups of IRs. These results were analyzed and discussed in terms of the functional evidence of the genes in question and ecological differences of the *Drosophila* species.

61

Serine protease CLIPB9 prompts melanization in mosquitoes

*An, Chunju; Budd, Aidan; Kanost, Michael R.; Michel, Kristin

(CA, KM) Kansas State University, Division of Biology, Manhattan, KS 66506; (AB) European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany; (MRK) Kansas State University, Department of Biochemistry, Manhattan, KS 66506

Melanization is a prominent defense mechanism in arthropods to encapsulate and kill invading microorganisms or pathogens. It requires activation of prophenoloxidase (proPO) by proPO-activating protease (PAP), which is controlled by an extracellular protease cascade and its inhibitors. In some mosquito species, melanization can kill invading malaria parasites; however the molecular composition of this defense reaction in these important vectors of human disease agents is largely unknown. We previously identified the serine protease inhibitors, SRPN1 and SRPN2 as negative regulators of melanization in the African Malaria mosquito, *Anopheles gambiae*. We recently identified two serine proteases as putative targets on these serpins. CLIPB8 and B9 are serine proteases composed of a single amino-terminal clip domain and a carboxyl-terminal catalytic domain. We produced and purified recombinant mutants of both protease zymogens, in which the activation site was changed to permit cleavage by bovine Factor Xa. Activated CLIPB8 and B9 form SDS-stable serpin-enzyme complexes with SRPN1 and 2 *in vitro*. The association rate constants of SRPN1 and SRPN2 with CLIPB9 were determined, with a stoichiometry of inhibition similar to known serpin-serine protease interactions. Knockdown of SRPN2 and CLIPB9 significantly decreased the size and numbers of melanotic pseudotumors induced by silencing of SRPN2. Altogether, these results identify CLIPB9, and possibly CLIPB8, as novel members of the proPO activation cascade in *An. gambiae*.

1

BeeSpace: An Interactive System for Functional Analysis

*Arcoleo, David S.; *Khetani, Radhika S.; Schatz, Bruce R.

Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana 61801

The BeeSpace System is now available in production version 4 for the Arthropod Base Consortium at www.beespace.illinois.edu. The software supports an automatic analogue of the curator process for reading literature and extracting facts. Abstracts from the PubMed collection, updated weekly, can be used to make special targeted collections: including the functions of an organism, such as honey bee or flour beetle, or the functions of a behavior, such as foraging or defense. The semantic indexing extracts the content of literature terms, such as gene names, and their context embedded within the documents, such as their co-occurrence frequency with other terms.

The gene summarizer displays functional information about a gene, in the context of the special collection in use. The summarizer is similar to FlyBase in that gene-specific sentences extracted from literature collections are classified into functional categories. But the extraction takes place automatically from the special collection, producing the analogue to a BeeBase or a ForageBase. The Genelist Analyzer is another tool that produces the concepts most discriminating to a list of genes within the special collection. This tool is similar to Gene Ontology and can be used to analyze gene expressions from microarray or nextgen experiments. The BeeSpace version 4 system provides an easy-to-use interface with four unique subsystems: enabling the user to make collections via search, cluster collections into focused subcollections, summarize gene information, and analyze gene lists. This version of the system makes it very easy to switch between the subsystems for a comprehensive and flexible functional analysis from biological literature.

84

Identification of Novel Bombyxin Genes From the Genome of the Silkworm *Bombyx mori* and Their Expression Analysis

*Aslam, AFM; Kiya, T; Higuchi, T; Sakurai, S; Mita, K and Iwami, M.

(AFMA, TK, TH, SS, MI) Kanazawa University, Division of Life Sciences, Graduate School of Natural Science and Technology, Kanazawa 920-1192, Japan; (KM) National Institute of Agricultural Sciences, Tsukuba, 305-8602, Japan.

Insulin and insulin like growth factors (IGFs) play key roles in regulating growth, metabolism, and reproduction in higher organisms. Bombyxin is an insulin-like peptide of the silkworm *Bombyx mori*. Till now, 32 bombyxin genes have been identified in the *Bombyx* genome and grouped into six families. This suggests that very high degree of diversification in structure and genomic organization of bombyxin genes occurred through long time evolution. To reveal the genomic evolution of bombyxin genes, we analyzed recently sequenced full genome of *Bombyx mori*, identified six novel bombyxin genes and grouped to five novel families: bombyxin V through Z. All of these genes encoded preprobombyxin, the precursor molecule for bombyxin. Amino acid sequence comparison showed that these prepropeptides have 21 to 50% and 21 to 26% identities with previously characterized preprobombyxins and human preproinsulin, respectively. We characterized genomic organization and chromosomal location of these genes. Each of bombyxin V1 and V2 has one intervening intron at 5' untranslated region (UTR), whereas Z1 has three introns – two at 5' UTR and one in C peptide region. Human insulin and relaxin genes have intervening introns at almost the same position of these bombyxin genes. Bombyxin V1 and V2 are localized on chromosome 9; W1, Y1, and Z1 on chromosome 1 and X1 on chromosome 11.

We performed reverse transcription PCR and whole mount *in situ* hybridization in different tissues and life stages to observe temporal and spatial expression pattern. Newly identified bombyxin genes are expressed in diversified tissues. Bombyxin V1, W1 and Y1 genes are predominantly expressed in *Bombyx* brain and shown to localized in four pairs of the medial neurosecretory cells, which are the expression sites of other bombyxin genes. X1 gene has a pruned C peptide region and is solely expressed in fat body, a functional equivalent of the vertebrate liver and adipocytes, which are consistent with IGF. X1 gene expression in the fat body appeared in the non-feeding stage – fifth instar day eight. Z1 gene showed high expression in female fat body and ovary. *In situ* hybridization clearly demonstrated Z1 mRNA localization in the follicular cells of oocyte and thus suggests its function in reproduction. Overall, nearly similar structure, location of intervening intron, and high gasteroenteric expression propose an evidence of bombyxin gene evolution originated from a common ancestral gene like vertebrate insulin, IGFs and relaxin.

44

Initial characterization of the large, repetitive genome of the brine shrimp, *Artemia*, by Cot filtration and Next Gen sequencing.

*Avery, Brian J.; Peterson, Daniel G.

(BJA) Biology Program and Great Salt Lake Institute, Westminster College, Salt Lake City, UT, 84105; (DGP) Mississippi Genome Exploration Laboratory, Department of Plant & Soil Sciences, Life Sciences & Biotechnology Institute, Mississippi State University, Mississippi State, MS, 39762

The small crustacean *Artemia* (class Branchiopoda, order Anostraca, a.k.a. brine shrimp) is an extreme halophile that can thrive in aquatic environments uninhabitable by other animals. *Artemia* can survive in hypersaline habitats in large part through its ability to produce encysted embryos that are remarkably tolerant of UV exposure, years of anoxia, extreme temperatures, severe desiccation, and repeated cycles of hydration-dehydration. *Artemia* are found in salt lakes and salterns world wide and are highly polymorphic both within and between populations which makes them an interesting study system for population genetics and evolution. However, there are few genomic tools for *Artemia* since the genome is relatively large for a crustacean (average estimate of ~1.6 Gb) and contains a considerable amount of repetitive DNA. As a means to more efficiently explore the gene space of this unique organism and begin to develop better tools for population genetics and gene analysis, we have used Cot filtration to separate a fraction of *Artemia* genomic DNA that is enriched in single and low copy (SL) sequences. We have used Roche/454 pyrosequencing to generate 139 Mb of DNA sequence from our SL fraction and approximately 87 Mb of DNA sequence from unfiltered *Artemia* genomic DNA. We are currently comparing the sequences from each set (SL and whole genome) and analyzing the repeat and gene content of each set to produce a new set of genomic tools to aid in the study of the population genetics, evolution and biochemistry of this remarkable organism.

62

Insecticide Resistance Mechanisms and Gene Expression Profile in Malaria Vector - *Anopheles gambiae* s.s

Awolola, Taiwo S¹; *Oyewole, Isaac O²; Oduola, Adedayo, O³; Koekemoer, Lizette L⁴; Coetzee, Maureen⁵; Strode, Clare⁶; Ranson, Hilary⁷.

1,3. Nigerian Institute of Medical Research, Yaba, Lagos, Nigeria; 2. Biosciences and Biotechnology, Babcock University Ilisan Remo, Ogun State, Nigeria; 4. Vector Control Reference Unit, National Institute for Communicable Diseases, Johannesburg, South Africa; 5. Division of Virology and Communicable Disease Surveillance, School of Pathology of the National Health Laboratory Service and the University of the Witwatersrand, Johannesburg, South Africa; 6,7. Liverpool School of Tropical Medicine, Liverpool, UK.

The use ITNs and IRS are integral component of malaria control. In spite of the effectiveness, resistance remains a challenge. Pyrethroid insecticide is a class of insecticide approved for treating nets. Resistance has emerged in *Anopheles gambiae* & *An. funestus*. Resistance mechanisms in *Anopheles gambiae* s.s due to target-site insensitivity arising from a single point mutation, known as knockdown resistance (kdr) and metabolic-based resistance were investigated from 2002-2007. Bioassay performed on *An. gambiae* s.s. showed that more than 50% of the resistant strain possessed kdr mutation and about 3% susceptible strain. Resistant mosquitoes synergized using pyreperonyl butoxide before permethrin exposure showed a significant increase in mortality compared with the non-synergized. Biochemical assays showed an increased level of monooxygenase but not glutathione-S-transferase or esterase activities in the resistant mosquitoes. Microarray analysis using the *An. gambiae* detox-chip for expression of detoxifying genes performed in 2008 showed five over-expressed genes in the resistant strain viz. two cuticular genes (*CPLC8* and *CPLC#*), a cytochrome P450 (*CYP325A3*), a sigma class GST (*GSTSI-2*) and a thioredoxin peroxidase (*TPX4*) gene. *TPX2* was the only gene that showed significantly higher expression in the susceptible strain. The two cuticular genes, *CPLC8* and *CPLC#* could constitute a novel set of candidate genes for further study since they are not previously implicated in pyrethroid metabolism in *An. gambiae* s.s.

94

Transcriptomic and proteomic analysis of the first instar larvae of *Frankliniella occidentalis*, the insect vector of *Tomato spotted wilt virus*

*Badillo-Vargas, Ismael E.; Rotenberg, Dorith; Hiromasa, Yasuaki; Tomich, John M.; Whitfield, Anna E.

(IEBV, DR, AEW) Department of Plant Pathology, Kansas State University, Manhattan, KS 66506; (YH, JMT) Department of Biochemistry and Biotechnology Core/Proteomic Facilities, Kansas State University, Manhattan, KS 66506

The western flower thrips (WFT), *Frankliniella occidentalis* Pergande (order Thysanoptera, suborder Terebrantia, family Thripidae), is an economically important insect pest with a worldwide distribution. WFT is the most efficient vector of *Tomato spotted wilt virus* (TSWV), with the virus being transmitted by adult thrips only if it was acquired during the larval stages. Insects belonging to the order Thysanoptera have not been studied at the transcriptome or proteome level despite their agricultural importance. The goal of our study was to develop transcriptomic and proteomic tools for this important insect vector of plant viruses. A normalized cDNA library from first instar larval thrips was constructed and 13,839 high quality expressed sequence tags (ESTs) were obtained. The thrips ESTs assembled into 894 contigs and 11,806 singletons; 25% of these sequences have provisional annotations. For analysis of the thrips proteome, we collected larval thrips and separated the proteins using two dimensional gel electrophoresis. Following Coomassie brilliant blue staining, 194 protein spots were excised, trypsinized, and subjected to liquid chromatography with tandem mass spectrometers (LC-MS/MS). The mass-to-charge ratio as a function of time was then used to generate peptide sequences that were analyzed against the NCBI Metazoan database and the WFT EST collection using Mascot software. We found that 8% and 29% of the proteins spots had significant and marginally significant sequence matches, respectively, to the WFT EST collection. When compared to the Metazoan database, 21% and 16% of the protein spots had significant and marginally significant matches, respectively. The majority of the characterized proteins had similarity to proteins found in other insects and two proteins matched a *F. occidentalis* cysteine protease CP7 sequence in the NCBI GenBank database. Our findings indicate that the transcriptomic and proteomic tools developed for WFT thus far will enable us to study molecular components associated with the WFT-TSWV interaction. We are currently identifying WFT proteins that are differentially expressed in response to TSWV infection with future research initiatives to develop technologies that disrupt acquisition and/or transmission of the virus.

72

Combining next-generation sequencing strategies for rapid molecular resource development from an invasive aphid species, *Aphis glycines*.

Bai, Xiaodong; Zhang, Wei; Orantes, Lucia; Jun, Tae-Hwan; Mittapalli, Omkaprash; Mian, M. A. Rouf; *Michel, Andrew P.

(XB, WZ, LO, TJ, OM, APM) Department of Entomology, Ohio Agricultural Research and Development Center, The Ohio State University, 1680 Madison Ave., Wooster, OH, 44691, USA; (TJ, MARM) USDA-ARS and Department of Horticulture and Crop Sciences, The Ohio State University, Wooster, OH 44691, USA.

Aphids are one of the most important insect taxa in terms of ecology, evolutionary biology, genetics and genomics, and interactions with endosymbionts. Additionally, many aphids are serious pest species of agricultural and horticultural plants. Recent genetic and genomic research has expanded molecular resources for many aphid species, including the whole genome sequencing of the pea aphid, *Acyrthosiphon pisum*. However, the invasive soybean aphid, *Aphis glycines*, lacks in any significant molecular resources. Two next-

generation sequencing technologies (Roche-454 and Illumina GA-II) were used in a combined approach to develop both transcriptomic and genomic resources, including expressed genes and molecular markers. Over 278 million bp were sequenced among the two methods, resulting in 19,293 transcripts and 56,688 genomic sequences. From this data set, 635 SNPs and 1,382 microsatellite makers were identified. For each sequencing method, different soybean aphid biotypes were used which revealed potential biotype specific markers. In addition, we uncovered 39,822 bp of sequence that were related to the obligatory endosymbiont, *Buchnera aphidicola*, as well as sequences that confirm the presence of *Hamiltonella defensa*, a facultative endosymbiont. We uncovered significant molecular resources for an invasive, non-model aphid species, as well as associated bacterial endosymbionts. The resources presented here will complement ongoing molecular studies within the Aphididae (including the pea aphid whole genome) lead to better understanding of aphid adaptation and evolution, and help provide novel targets for soybean aphid control.

73

A Genomics Approach for Characterising Resistance to Neonicotinoid Insecticides in the Aphid *Myzus persicae*

Bass, Chris; Puinean, Alin M; Denholm, Ian; Field, Linda M; *Williamson, Martin S.
Centre for Pest and Disease Management, Rothamsted Research, Harpenden, AL5 2JQ, UK.

The aphid *Myzus persicae* is an important agricultural pest in the UK and other parts of the world, causing direct feeding damage and transmission of viral diseases on a variety of crops. It is normally controlled through the use of chemical insecticides, however field populations have now evolved resistance to most of the major classes. An important exception is the neonicotinoids that have retained effective field control, although there are recent indications that tolerance/resistance to this class may also be increasing. Work on other insect species has shown that resistance to neonicotinoids can arise through increased production of metabolic enzymes (such as cytochrome P450s) that detoxify the insecticide. In this study, we have made use of recent advances in the genomics of *M. persicae* to quantify the expression of all the genes encoding detoxification enzymes in a strain showing 50 fold resistance to imidacloprid. We found that resistance correlated with the over-expression of a single P450 gene and that this is caused by amplification of the gene coding for this enzyme. We also note that over-expression of certain cuticular proteins and reduced penetration of insecticide through the cuticle may also play a role in the resistance.

40

Gene discovery, linkage mapping and comparative genomics in Lepidoptera using next-generation RAD sequencing

*Baxter, Simon W.; Jiggins, Chris D.; Blaxter, Mark L.; Davey, John W.
(SWB, CDJ) Department of Zoology, University of Cambridge, Cambridge, CB2 3EJ, U.K.;
(MLB, JWD) Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, U.K.

Restriction-site associated DNA (RAD) sequencing is a relatively new method of sampling regions across the genomes of many individuals. This approach has broad potential for identifying genes of major effect, creating linkage maps to aid genome assembly, comparing chromosomal synteny between species and large scale population genetic studies. A RAD library was constructed using genomic DNA from a diamondback moth (*Plutella xylostella*) backcross. The library contained two parents plus 22 backcross progeny that were either resistant or susceptible to the insecticide spinosad. Sequencing was performed on a single illumina GAII lane (50 base paired end reads), and each RAD tag could be assigned to one of the 24 library individuals using a unique 5 bp index sequence. Polymorphisms in 6500 sequenced tags were found to be segregating among the backcross progeny. Linkage mapping assigned these tags to one of the 31 diamondback moth chromosomes, and the chromosome that carried a spinosad resistance mutation was successfully identified. The program Velvet was used to assemble the paired end reads for each RAD tag. This generated a larger genomic fragment pair (200-600 bp) for every RAD tag. BLAST results show that many of these sequences contain predicted protein coding regions and have single hits to the silkworm genome (*Bombyx mori*), enabling large scale comparative genomics.

54

Genomics Tools to Develop Control Strategies to the Coffee Berry Borer, *Hypothenemus hampei*, in Colombia

Benavides, Pablo*; Navarro, Lucio; Góngora, Carmenza; Hernandez, Eric; Acevedo, Flor; Machado, Carlos.
(PB, LN, CG, EH, FA) Colombian Center for Coffee Research – Cenicafe, Department of Entomology, Chinchina, Colombia; (CM) University of Maryland, Department of Biology, College Park, MD 20742.

Coffee berry borer (CBB), *Hypothenemus hampei*, is the major insect pest of coffee crops worldwide. In spite of the implementation of an integrated pest management approach in Colombia to mitigate its negative effects, CBB can still cause considerable coffee yield reduction and cup quality decrease. Given some limitations related to the implemented CBB control methods such as the effect of climate change on the insect dynamics, it is feasible to develop new strategies based on insect molecular data to ameliorate CBB damage on coffee crops. Hence, we have undertaken genomic approaches in order to figure out some insect physiological and molecular process leading the interaction between the CBB and its only host the coffee plant. Here, we describe a subtractive gene expression screening in CBB settled to identify defense genes against antibiosis *Coffea* plant compounds, while screening molecularly coffee plants in order to evaluate the plant genetic defense response against the CBB. We also show the efforts to get an entire view of the transcriptome of the CBB by using 454-pyrosequencing of all stages of its life cycle, and subsequently promote gene discovery and identification of candidate target genes. Thus, the first draft of the CBB transcriptome (~13.000 unigenes) is presented through EST assembly of 454-sequences (normalized and non-normalized cDNA libraries) and full-length cDNA libraries (midgut and whole

CBB adults) expressed under several conditions. Functional annotations covered a wide number of biological and molecular process, as well as allowed identification of candidate genes involved in basic process as host-plant recognition (odorant-binding proteins); allelochemical detoxification (cytochrome P450, glutathion s-transferases and esterases) and metabolism (proteases and glycosidases). Those genes are now candidate target to initiate insect transgenic approaches in future functional genomics studies. In addition, as a prerequisite to considerate efforts for CBB entire genome sequencing, we have done the first approximation to the insect genome size by using a real-time PCR method. This first estimation shows a ~110 Mb genome for the CBB, which places this insect as good candidate for whole genome sequencing. However, further estimations using cytogenetic techniques must be done to support our result.

4

Bee Pests and Pathogens: a unique resource within BeeBase

*Bennett, Anna K.; Cornman, R. Scott; Aronstein, Katherine A.; Evans, Jay D.; Elisk, Christine G. (AKB, CGE) Georgetown University, Department of Biology, Washington, DC 20057; (RSC, JDE) Bee Research Lab, USDA-ARS, Beltsville, MD 20705; (KAA) Honey Bee Research Unit, USDA-ARS, Weslaco, TX 78596

The "Bee Pests and Pathogens" (BPP) website (www.beebase.org/?q=bee_pathogens) is a unique collection of informatics resources within BeeBase for a growing number of sequenced honey bee, *Apis mellifera*, parasites. To our knowledge BeeBase is the only model organism genome database that integrates genomic resources for biologically and economically important parasites of the host. Honey bee health is a significant economic concern due to its impacts on agricultural pollination and is of research interest as a model for disease transmission in social groups. By combining genomic resources for honey bee and its parasites in a single location we enable researchers to more easily access data for the study of bee health and host-parasite interactions. BPP currently provides access to genomic resources for the bacterium *Paenibacillus larvae*, the fungus *Ascosphaera apis* and the fungus *Nosema ceranae*, causative agents of American Foulbrood disease, Chalkbrood disease and Nosema disease of honey bees, respectively. The NCBI phylogenetic classification for each species is provided along with links to the sequencing groups and genome publications. Genome browsers for each species display tracks for *ab initio* gene predictions and other available alignment data. BLAST is available for *P. larvae* and *N. ceranae* and will be expanded to other species in the future. Genomic resources for *Varroa destructor*, a mite, will be added to the site upon completion of the sequencing project. BPP also plans to integrate transcriptome data into the genome browsers as it becomes available and use it as evidence to improve gene predictions.

101

Control of Corn Rootworm by dsRNA expressed in planta

*Bolognesi, Renata; Clark, Tom; Flannagan, Ron; Heck, Gregory R; Ivashuta, Sergey; Wiggins, Elizabeth; Zhao, Suling; Segers, Gerrit
Monsanto Company, 800 North Lindbergh, Creve Coeur, MO 63167 USA

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*, Le Conte). 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 with less root injury. 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. Field studies have been conducted with transgenic plants expressing dsRNA to WCR target sequences. Under the dynamic pressure of rootworm herbivory in the field, higher levels of root damage were noted than in some growth chamber evaluations. 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 element appears to bring a potent component to the combination as noted in reduction of WCR adult beetle emergence.

45

Progress on the Whole Genome Sequencing of the Monarch Butterfly, *Danaus plexippus*

*Boore, Jeffrey L.; Fuerstenberg, Susan I.; Merlin, Christine; Reppert, Steven M.
(SIF, JLB) Genome Project Solutions, Hercules, CA 94547; (JLB) University of California, Department of Integrative Biology, Berkeley, CA 94720; (CM, SMR) University of Massachusetts Medical School, Department of Neurobiology, Worcester, MA 01604

We are determining a high-depth draft sequence of the 284 MB genome of the monarch butterfly, *Danaus plexippus*. This organism serves as a model for sun compass navigation that they use during their spectacular fall migration, for the molecular regulation of circadian rhythms, which they do by a novel mechanism having aspects of both the *Drosophila* and human clocks, and for other aspects of neurobiology. In addition, this work will contribute greatly to our understanding of genome evolution, because it will be the first butterfly genome sequenced and will complement work done on the genome of the moth, *Bombyx mori*. (Moths and butterflies comprise the two great lineages of the Lepidoptera.) Moreover, the monarch genome will be compared to the genomes already available for representatives of the Diptera, Coleoptera, and Hymenoptera, the crustacean, *Daphnia pulex*. and the chelicerate, *Ixodes scapularis*, as well as many more arthropod genomes now in progress.

We have generated 1x coverage using traditional Sanger sequencing and >16x coverage in Roche (454) sequence, with ~1x of this being in paired-end format with 20 kb separation. We have also generated deep coverage of the transcriptome using Illumina sequencing. We have made a preliminary assembly and this, plus all sequencing reads, are available from a web site listed on this poster. Within the next few months, we will complete this draft genome sequence using further paired-end Roche (to increase scaffold sizes) and Illumina (to correct any homopolymer errors) technologies, and then proceed with annotating genes and other structural features on the assembled genome, creating a gene catalog, making measures of the characteristics of the monarch genes and genome, comparing with other genomes, making specific scientific interpretations related to sun compass navigation, circadian rhythms, and other aspects of neurobiology and genome evolution, and presenting this to the scientific community.

We will interpret this genome using 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, biochemical pathway, protein characteristics, number of members in any gene family, and many others. This will include creating phylogenetic trees for all genes in all completely sequenced arthropod genomes using our PHRINGE (Phylogenetic Resources for Interpreting Genomes) pipeline, which allows transferring inference of gene function based on orthology rather than the error-prone BLAST matching in common use. Users will be able to view all evolutionary trees, multiple sequence alignments, comparative gene structures, and ortholog colinearity maps. (See more details of this at the poster “Evolutionary Analysis as the Basis for Interpreting, Comparing, and Presenting Genomes: The GATOR and PHRINGE System”.)

74

A functional genomics pipeline identifies candidate effectors from the aphid species *Myzus persicae*

*Bos, Jorunn I.B.¹; Prince, David¹; Pitino, Marco¹; Win, Joe²; Hogenhout, Saskia A.¹

¹Department of Disease and Stress Biology, The John Innes Centre, Norwich, NR4 7UH, United Kingdom; ²The Sainsbury Laboratory, Norwich, NR4 7UH, United Kingdom. jorunn.bos@bbsrc.ac.uk

Aphids are amongst the most devastating hemipteran sap-feeding insects of plants. Like most plant parasites, aphids require intimate associations with host plants to gain access to nutrients. Aphid feeding induces plant responses such as clogging of phloem sieve elements and callose formation, which are suppressed by unknown molecules, probably proteins, in aphid saliva. Thus, aphids, like plant pathogens, may deliver proteins (effectors) inside their hosts to modulate host cell processes to promote infestation. We exploited publicly available salivary gland expressed sequence tags (ESTs) to identify candidate effectors from the aphid species *Myzus persicae* (green peach aphid) based on common features of plant pathogen effectors. A total of 48 effector candidates were identified by bioinformatics, cloned, and subjected to transient overexpression in *Nicotiana benthamiana* to assay for: 1) elicitation of a phenotype, 2) suppression of Pathogen-Associated Molecular Pattern (PAMP)-mediated oxidative burst, and 3) effects on aphid reproductive performance (fecundity). We found an effector candidate that induced chlorosis and local cell death in *Nicotiana* species but not tomato suggesting activation of plant defenses. This same candidate suppressed the PAMP-mediated oxidative burst. Aphid fecundity assays identified three candidates that affected aphid performance, two of which reduced aphid fecundity, again suggesting that some effectors may indeed trigger plant defenses, whereas one increased aphid fecundity. Overall, we identified aphid salivary proteins that share features with plant pathogen effectors and therefore may function as aphid effectors by perturbing host cellular processes.

68

Dynamics of Evolutionary Conflict in the Germ Line of *Drosophila*

*Box, Kim; Castillo, Dean M.; Blumenstiel Justin P.

Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, Kansas 66045;

Parasites and their hosts are locked in an evolutionary arms race. The Piwi-interacting RNA (piRNA) machinery in *Drosophila* forms RNA-protein complexes that serve to protect the host against viruses and transposable elements in germ line cells. Previous studies have shown that the piRNA machinery appears to be undergoing a high rate of adaptive evolution, as predicted under a co-evolutionary arms race model. However, these studies were performed in species with relatively low TE content. It is not known whether high rates of evolution by adaptation generally shape the evolution of the piRNA machinery. Such a competition would seem to suggest that rates of evolution in piRNA genes would increase as transposable element, TE, abundance increases. However, other studies have shown that this positive correlation is not the case. Instead, higher levels of TE abundance are associated with lower rates of evolution of the piRNA machinery. If there is a negative correlation between rates of evolution and TE abundance, we might expect that under constraint, natural selection would favor an increase in gene expression in components of the piRNA machinery. Since codon bias is known to correlate with levels of gene expression we analyzed codon bias in the piRNA genes. Here we demonstrate that relationships between TE abundance and codon adaptive index is negative in piRNA genes and thus it appears that TE abundance has an effect on codon bias levels. This supports the idea that higher TE abundance induces constraint in the piRNA genes as opposed to higher levels of natural selection.

75

Genomic Approaches to Mapping the Pea Aphid Wing Dimorphism Locus

Brisson, Jenn A.

University of Nebraska, School of Biological Sciences, Lincoln, NE 68588

The pea aphid (*Acyrtosiphon pisum*) exhibits dramatically different adult phenotypes of winged or unwinged morphs that are induced by environmental conditions in asexual females (a polyphenism) and by a genetic locus in males (a polymorphism). Specifically, during the spring and summer months, females reproduce parthenogenetically, producing clonal daughters for ten to twenty generations over the course of the summer. These females are often unwinged. Under stressful conditions, however, such as when the host-plant becomes overcrowded, the mother gives birth to daughters that develop into the winged morph and are capable of dispersing to other host plants. Males are produced during a sexual generation in the fall. Males also exhibit winged and unwinged morphs, although in contrast to females, the male wing polymorphism is determined by a single unidentified locus on the X chromosome called *aphicarus* (*api*). Winged and unwinged male siblings are genetically identical except for their X-chromosomes, which carry either the *api*-unwinged or *api*-winged allele. Interestingly, genetic variation for the female wing polyphenism segregates with the male wing polymorphism. Therefore, the male polymorphism locus may play a role in the development of the asexual female wing dimorphism. If so, mapping and identifying the male polymorphism locus would ultimately inform us about the developmental genetic basis of the wing polyphenism.

I am using information from the pea aphid genome to map *api* using two approaches. First, I have sequenced genomic DNA from 20 unwinged and 20 winged males from natural populations using 18 lanes of Illumina sequencing. From these data, I have identified single nucleotide polymorphisms (SNPs) that segregate with the male phenotype. I have confirmed that several of them are within close proximity to *api*. Second, in collaboration with David Stern at Princeton University, we are using an Illumina genotyping approach to simultaneously discover and score SNPs in an F2 population segregating for *api*. By using these two complementary approaches, we hope to rapidly identify the genomic region containing the *api* locus.

85

A candidate gene similar to VAV1 oncogene from a lepidopteran, *Bombyx mori* by comparative genomics.

*Brown, Thomas M.

GeneCTAr Com LLC, Whitefish, MT 59937. brown_t@bellsouth.net

Human VAV1 oncogene encodes a cytosolic member of a tyrosine kinase signal transduction cascade. VAV1 maps to chromosome 19p13.2 and it is specifically expressed in the hematopoietic system. Six dipteran expression products were listed in NCBI UniGene. NCBI Blink taxonomy revealed 37 BLAST hits among arthropods of which 25 were dipteran. No lepidopteran sequences were found. Searching public databases, a candidate lepidopteran vav sequence was found in *Bombyx mori* and assembled from short contigs. This sequence is very similar to VAV1 and several hypothetical arthropod genes including one predicted from *Tribolium castaneum*. These assembled lepidopteran sequences will be presented as aligned to known and inferred insect vav sequences.

36

The genome of *Serratia symbiotica*, a symbiont of pea aphids

*Burke, Gaelen R.; Moran, Nancy A.

(GRB) The University of Arizona, The Department of Ecology and Evolutionary Biology, Tucson, AZ 85721; (NAM) Yale University, Department of Ecology and Evolutionary Biology, New Haven, CT 06520

Serratia symbiotica is a bacterial symbiont of aphids that provides heat tolerance to its hosts, seemingly through protection of the obligate symbiont *Buchnera*. This symbiont is facultative from the perspective of pea aphid hosts, and frequencies vary in populations of pea aphids. Despite its low titer in its host, metabolomic analysis of *S. symbiotica* revealed it has a large effect upon aphid metabolism, although many of the metabolites affected by this symbiont unidentified. Both the function and evolution of *S. symbiotica* has been explored using whole genome sequencing. *S. symbiotica* is in the early stages of transition to a symbiotic lifestyle, and this change is expected to be reflected in its genome. Until recently, purification of DNA for sequencing from this unculturable bacterium was limited by the symbiont's low density. We sequenced the genome of *S. symbiotica* from pea aphids, using selective whole genome amplification and next generation sequencing. When compared to a free-living *Serratia* genome, the genome exhibited size reduction, with streamlining through inactivation of many genes (>600), especially functional orthologs. This genome allowed us to sample a snapshot of molecular changes that occur in the process of evolution to a symbiotic lifestyle, and allowed us to survey the metabolic capabilities of *S. symbiotica*, and potential impacts upon host metabolism.

83

Desaturase Genes in the Newly Sequenced Genomes of the Red Harvester Ant, *Pogonomyrmex barbatus*, and the Argentine Ant, *Linepithema humile*

*Cash, Elizabeth I.; Gadau, Jürgen

Arizona State University, School of Life Sciences, Tempe, AZ 85287

Social insects are distinct from other insects in their ability to form complex communication systems. Inherent to their ability to effectively communicate, social insects must be able to identify and differentiate between social group members and non-members. In many ants, including *Pogonomyrmex barbatus* and *Linepithema humile*, species and colony recognition is largely based on cues produced by cuticular hydrocarbon (CHC) profiles. *P. barbatus* and *L. humile* each produce numerous different CHC compounds (34, and ~50 respectively) including: *n*-alkanes, methyl-branched alkanes, and alkenes. Until now, no genes have been identified or isolated that are known to influence the CHC patterns of any social insect. In *Drosophila*, however, earlier studies of CHC components show that carboxylases, elongases and desaturases each contribute considerably to CHC biosynthesis and variability. Of these, the best studied are the desaturases, which create carbon-carbon double bonds in linear alkanes to form alkenes and dienes. Currently, three $\Delta 9/\Delta 11$ desaturase genes in *Drosophila*, *desat1*, *desat2*, and *desatF* (syn. *Fad2*), are known to contribute specifically to alkene synthesis and phenotypic variation of cuticular hydrocarbons. Both *P. barbatus* and *L. humile* produce variable quantities of several CHC alkenes, making these three desaturase genes, *D. melanogaster desat1*, *desat2*, and *desatF*, excellent genome query candidates for study of CHC alkene genes. Preliminary results from our analyses of these two ant genomes show the presence of ten predicted $\Delta 9/\Delta 11$ desaturase genes in *P. barbatus*, and approximately twenty $\Delta 9/\Delta 11$ desaturase genes in *L. humile*. These results suggest that relatively larger expansions of $\Delta 9/\Delta 11$ desaturase genes in *P. barbatus* and *L. humile* may be useful in contributing to qualitative and quantitative differences in CHC alkene variation, which makes them likely candidates for experimental studies regarding their contribution to CHC profiles and recognition cues. We will present a full characterization and comparative analysis of $\Delta 9/\Delta 11$ desaturase genes from *P. barbatus* and *L. humile* at the symposium.

102

Evolution of the *Drosophila* piRNA machinery slows under increasing genome assault

*Castillo, Dean M.; Mell, Joshua Chang; Blumenstiel, Justin P.

(DMC, JPB) Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence

Kansas 66045; (JCM) Life Sciences Centre (Zoology), 2350 Health Sciences Mall, University of British Columbia, Vancouver, BC, Canada, V6T 3Z4

Hosts and their parasites are locked in a co-evolutionary struggle that can lead to various outcomes. Under certain conditions, cycles of adaptation and counter-adaptation can drive an evolutionary arms race. This arms race is often seen at the molecular level in the form of a high rate of adaptive evolution. Conversely, if the fixation of any single defense strategy comes at the cost of alternative beneficial strategies, a form of trench warfare may occur which tempers this cycle of rapid evolution. Here we demonstrate that as *Drosophila* genomes become increasingly burdened by transposable elements there is a corresponding increase in conservation of the piRNA machinery, consistent with a retreat into trench warfare.

86

Transcriptional fingerprinting of hormetic response towards gossypol in the cotton bollworm, *Helicoverpa armigera*

*Celorio-Mancera, Maria de la Paz; Ahn, Seung-Joon; Vogel, Heiko; Heckel, David G.

Max Planck Institute for Chemical Ecology, Department of Entomology, Beutenberg Campus, Hans-Knöll-Straße 8, 07745 Jena, Germany

Hormesis appears to be a universal biological response towards a chemical compound which is stimulatory in low dosages and detrimental in high dosages. Dose-dependent transcriptional response to gossypol, a polyphenolic binaphthyl dialdehyde found in the pigment glands of plant species belonging to Malvaceae, has been observed for a handful of genes including the expression of mitochondrial-related genes and P450 monooxygenases. However, there is no unbiased transcriptional profiling approach documented investigating the effects of different doses of gossypol. *Helicoverpa armigera*, an important heliothine pest species, damages a large number of plant species and cotton is one of its best suitable hosts. We investigated whether the hormetic effect of gossypol can be observed in *H. armigera* by exposing larvae to different doses of gossypol in their diet. Weight gain, pupal weight and larval development time data were documented after a three day feeding experiment. On the basis of net larval weight gain responses to gossypol, three treatments, that is, control (0), 0.016 and 0.16% gossypol, were selected for gene expression studies in a balanced interwoven loop design microarray experiment. The expression of genes involved in multiple biological processes was altered, including protein catabolism and ion transport. The transcriptional response of *H. armigera* to gossypol was found to be similar to the response by an annelid towards fluoroanthene, a polycyclic aromatic hydrocarbon, known to disrupt biological membranes due to its lipophilicity. Genes which predicted products are integral to membrane and associated with or located in the proteasome core complex and mitochondrial matrix were also significantly affected across tissue-gossypol treatments. In addition, oxidoreductase activity-related genes were observed to be significantly altered. The transcription level of members of the aryl hydrocarbon receptor gene family was found to be affected by the different gossypol dosages.

These findings provide an unbiased comprehensive view of how *H. armigera* transcriptome responds towards, not only detrimental gossypol concentrations but to lower, more ecologically meaningful doses.

55

Functional Analysis of the *Tribolium castaneum knickkopf* Gene Family in Organization of Cuticle

*Chaudhari, Sujata S; Arakane, Yasuyuki; Boyle, Daniel L; Moussian, Bernard; Kramer, Karl J; Beeman, Richard W; Muthukrishnan, Subbaratnam

(SC, YA, KJK, SM) Department of Biochemistry, Kansas State University, 141 Chalmers Hall, Manhattan, KS 66506, USA; (DB) Department of Biology, Kansas State University, 116 Ackert Hall, Manhattan, KS 66506, USA; (BM) Department of Genetics, Max-Planck-Institute for Developmental Biology, Tübingen, Germany; (KJK, RWB) Center for Grain and Animal Health Research ARS-USDA, 1515 College Avenue, Manhattan, KS 66502, USA

The insect cuticle is mainly composed of chitin and proteins, and forms a protective barrier that shields insects against biological and mechanical stresses. Proteins associated with chitin metabolism and/or cuticle assembly are attractive targets for biopesticide development. Recent studies of *Drosophila melanogaster* (*Dm*) have identified a gene, *knickkopf* (*knk*= broken head), whose expression is important for tracheal tube expansion and cuticle organization (Moussian et al., 2006). In the present study, we have identified three *Knk*-like genes in *Tribolium castaneum*, *TcKnk1*, *TcKnk2* and *TcKnk3*, which are orthologs of *DmKnk1*, *DmKnk2* and *DmKnk3*, respectively. All of these genes were differentially expressed during different developmental stages of the beetle, suggestive of distinct roles for each *TcKnk* in *Tribolium* development. All three *Knk* genes are expressed in the carcass (whole body without gut) but not in gut tissue. RNA interference (RNAi) of *TcKnk1* resulted in lethal phenotypes at larval-larval, larval-pupal and pupal-adult molts, whereas RNAi of *TcKnk2* and *TcKnk3* led to developmental arrest only at the pupal-adult molt with ~55% and 100% mortality, respectively. Interestingly, a lethal phenotype was observed only with dsRNAs specific for the C-terminal region of *TcKnk3* but not its N-terminal region, suggestive of alternatively spliced variants of *TcKnk3*. Confocal microscopic analysis of epidermal tissues indicated co-localization of *TcKnk1* with cuticular chitin in control insects. A significant reduction in chitin content was observed after RNAi for *TcKnk1*. Transmission electron microscopic analysis confirmed that *TcKnk1*-specific RNAi lethality is associated with defects in cuticle organization or synthesis. Collectively, our results suggest that *Knks* play crucial roles in the synthesis and/or organization of chitin in the cuticle of the red flour beetle and other species of insects.

References: Moussian, B. et al. (2006) Drosophila Knickkopf and Retroactive are needed for epithelial tube growth and cuticle differentiation through their specific requirement for chitin filament organization. Development 133, 163-71.

77

Vasa genes and germline development in aphids

Chung, Chen-Yo; Lin, Gee-Way; Lu, Hsiao-Ling; *Chang, Chun-che

Laboratory for Genetics and Development, Department of Entomology/Institute of Biotechnology, College of Bio-resources and Agriculture, National Taiwan University, No. 27, Lane 113, Roosevelt Road, Sec. 4, Taipei 106, TAIWAN

Molecular networks governing germline development share common components in animals. Among all germline genes, the *vasa* gene has been intensively studied and is regarded as one of the most conserved germline markers across animal phyla. We have identified four *vasa* homologues (*Apvasa1-4*) in the pea aphid *Acyrtosiphon pisum*, and have determined the expression patterns of their mRNAs in embryos produced during the asexual phase of the aphid life cycle. Of the four *Apvasa* genes, *Apvasa1* is germline specific but the others are not. Our finding of multiple copies of *vasa* in the pea aphid was unusual: the insects for which complete genomes are available: *Drosophila melanogaster*, *Tribolium castaneum*, *Apis mellifera* and *Bombyx mori* all have only a single *vasa* gene, and we did not know whether the multiple copies of *vasa* were unique to the pea aphid or shared with other taxa and if shared, whether their functions were shared as well. To explore this question we searched for *vasa* genes in the green peach aphid, *Myzus persicae*, which is distantly related to *A. pisum*. We have to date identified three *vasa* genes (*Mpvasa1-3*) in *M. persicae*. The amino acid sequences encoded by *Mpvasa1* and *Mpvasa2* have identity up to 90% with those encoded by *Apvasa1* and *Apvasa2*, respectively, but *MpVasa3*, although clearly an *ApVasa3* homologue, has lesser identity. We have not been able to identify by cloning or by nucleotide databank search, a potential *Apvasa4* homologue in *M. persicae*. We do observe that *Mpvasa1*, like *Apvasa1*, is specifically expressed in the embryonic germ cells. Our results suggest that the *vasa* gene has undergone a number of duplication events within the aphid lineage and that one copy of the gene retains its ancestral function as a germline marker. However, the number of genes within each taxon, the history of duplication events, and the functions of the extra *vasa* copies require further investigation.

63

A novel class of miniature inverted repeat transposable elements (MITEs) that contain hitchhiking (GTCY)_n microsatellites

*Coates Brad S.; Kroemer, Jeremy A.; Hellmich, Richard L.

(BSC, JAK, RLH) United States Department of Agriculture, Agricultural Research Service, Corn Insects & Crop Genetics Research Unit, Ames, IA 50011

The movement of miniature inverted repeat transposable elements (MITEs) affect the structure and function of a genome, and are major sources of genetic variation within natural populations and phenotypic traits with biological relevance. We show that a group of miniature inverted repeat transposable elements (MITEs) called the microsatellite-associated interspersed nuclear element family 2 (*MINE-2*) that integrate at consensus WTTT target sites, create dinucleotide TT target site duplications (TSDs), and form predicted MITE-like secondary structures; a 5' subterminal inverted repeat (SIRs; AGGGTTCCGTAG) that is partially complementary to 5' inverted repeats (IRs; ACGAAGCCCT) and 3' SIRs (TTACGGAACCT). A (GTCY)_n microsatellite is hitchhiking downstream of conserved 5' *MINE-2* secondary structures, causing flanking sequence similarity among mobile microsatellite loci that affect subsequent PCR assays. Transfection experiments show that secondary structures at *MINE-2* termini are sufficient to mediate the stable genome integration of reporter constructs in Sf9 and HighFiveTM insect cell lines, in contrast to transient expression of the constructs. These data provide insight into the transposition mechanism used by MITE-like element within arthropods, and allow future investigations of roles of genome plasticity in the evaluation of novel phenotypes.

87

A single nucleotide polymorphism (SNP) base linkage map for *Ostrinia nubilalis**Coates Brad S.; Kroemer, Jeremy A.; Bayles, Darrell O.; Hellmich, Richard L.; Wanner, Kevin W³; Robertson, Hugh M.

(BSC, JAK, RLH) United States Department of Agriculture, Agricultural Research Service, Corn Insects & Crop Genetics Research Unit, Ames, IA 50011; (KWW) Department of Plant Sciences and Plant Pathology, Montana State University, Bozeman, MT, 59717; (HMR) Department of Entomology, University of Illinois, Champaign, IL 61801

Single nucleotide polymorphisms (SNPs) are nucleotide substitution mutations that tend to be bi-allelic and at high densities within eukaryotic genomes, which make their development in assays to detect allelic variation attractive for genome mapping applications. Additionally, mutation rates at SNP are lower compared to microsatellites loci, and most loci represent stable bi-allelic lineages that can be traced with high fidelity in natural populations. A pipeline for the development of single locus markers from a total of 2742 putative SNPs within 7414 EST from the non-model crop pest species *Ostrinia nubilalis* (Lepidoptera: Crambidae) is described. Single base extension (SBE) assays were designed for 763 SNP loci SNP within 24 iPLEX GoldTM PCR multiplex reactions for subsequent detection on the SEQUENOM MassARRAY[®]. SNP polymorphism data was collected from 46 progeny of an *O. nubilalis* backcross pedigree, and linkage associations determined among segregating genotypes in order to construct a moderate density map constituting 32 linkage groups (LGs). Results indicate that SNPs are an abundant source of mutation data that can be developed into single locus genetic markers for non-model species, encompass a rapid entry point into genomic research, and may be valuable tools required for genome reassembly.

76

Analysis of aphid (Hemiptera: Aphididae) salivary proteins

W.R. Cooper, G.J. Puterka, J.W. Dillwith

We evaluated the salivary constituents of the Russian wheat aphid (*Diuraphis noxia*, Hemiptera: Aphididae) using 2-D electrophoresis and Orbitrap MS. Alkaline phosphatase was putatively identified in *D. noxia* saliva using Orbitrap MS, and alkaline phosphatase activity confirmed using biochemical assays. Alkaline phosphatase was only detected in probed 15% sucrose diet, and was not detected in probed diet consisting of amino acids (100 mM serine, 100 mM methionine, 100 mM aspartic acid, and 15% sucrose). Other peptides in sucrose diet weakly but significantly showed similarities to putative dehydrogenase and RNA helicase ESTs described from other aphids. We also compared salivary constituents between two *D. noxia* biotypes (RWA-1 and RWA-2) that differ in their responses to wheat resistance conferred by Dn-4 using gel electrophoresis and MALDI-TOF-MS. We did not detect differences, but analyses are ongoing. Lastly, we compared the salivary constituents between several grain aphids (*D. noxia*, *D. tritici*, *D. mexicana*, and *Schizaphis graminum*), and the pea aphid (*Acyrtosiphon pisum*) using protein electrophoresis and biochemical assays. Banding patterns were similar for the three *Diuraphis* species, but were variable across genera. All five species shared common a 130 kDa protein detected by SDS-PAGE and alkaline phosphatase activity.

103

Vector Biology Funding Opportunities at the National Institute of Allergy and Infectious Diseases

Costero, Adriana

NIH/NIAID

Comparative genomic studies on insects with holocentric chromosomes

d'Alençon, Emmanuelle; Sezutsu, Hideki; Legeai, Fabrice; Permal, Emmanuelle; Stanojčić, Slavica; Nègre, Nicolas; Bernard-Samain, Sylvie; Shimomura, Michihiko; Gimenez, Sylvie; Gagneur, Cyrill; Cousserans, François; Allassoeur, Benjamin; Léger, Alexandre; Abd-Alla, Adly; Juliant, Sylvie; Brun-Barale, Alexandra; Flutre, Timothée; Couloux, Arnaud; East, Peter; Gordon, Karl; Mita, Kasuei; Quesneville, Hadi; Feyereisen, René; *Fournier, Philippe.

(EDA, SS, SG, CG, FC, BA, AL, AA, SJ, PF) UMR1231, INRA & University, 34095 Montpellier, France; (AB, RF) UMR1301, INRA, CNRS, University, 06903 Sophia-Antipolis, France; (HS, MS, KM) Insect Genome Lab, NIAS, Tsukuba 305-8634, Ibaraki, Japan; (FL) UMR1099, INRA, INRIA, AgroCampus, 35042 Rennes, France; (EP, TF, HQ) UR1164, INRA, Versailles 78026, France; (SBS, AC) Genoscope, 91057 Evry, France; (PE, KG) Division of Entomology, CSIRO, Canberra, ACT 2601, Australia; (NN) UPR1142 CNRS, Montpellier Cedex, France.

In order to elucidate the molecular determinism and consequences of the holocentric nature of Lepidopteran chromosomes, we have developed a multidirectional approach.

First, this structure was expected to induce a specific syntenic organization. Indeed the holocentric nature of chromosomes suggests a scattered organization of centromeric determinants which may lead to a greater genomic plasticity, as fragments resulting from double-strand breaks may be maintained and reintegrated elsewhere. This hypothesis was studied by comparative analysis of 15 BACs of 3 Lep species, the two noctuid pests *Spodoptera frugiperda* and *Helicoverpa armigera*, and the silkworm *Bombyx mori*. A high degree of synteny conservation was observed between the 3 species. However conserved syntenic blocks of identified genes were very small, containing approximately 1.3 genes per block between *B. mori* and the two noctuid species and 2.0 genes per block between *S. frugiperda* and *H. armigera*. This corresponds to approximately two chromosome breaks per Mb DNA per MYA. This is a much higher evolution rate than among species of the *Drosophila* genus and may be related to the holocentric nature of the lepidopteran genomes. The analysis of gene clusters from rapidly evolving multigene families - such as the P450s - also shows evidence of recent and multiple gene duplication events that are specific for both noctuid genomes.

Second we analyzed the nature and frequency of repeated elements among these regions and found an abundance of LINES and TIRs rather than LTRs like in *Drosophila*'s. We also observed a biased distribution in some regions affected by synteny breakage.

Third, we looked for centromeric proteins in order to reveal features of the holocentric kinetochore. The discovery of an EST homolog to the human centromeric protein B in *S. frugiperda* prompted us to characterize that gene which has not yet been described in invertebrates. Like its human counterpart, the *Sf cenp-B* is related to the transposase of the *pogo* transposable element of *D. melanogaster*. We have evidences – among others from the syntenic comparison of its localization in the 3 Lep species -that the lepidopteran gene may have evolved by domestication of a transposase. The *Sf CENP-B* DNA binding activity *in vitro*, its nuclear location, its ability to bind to a retrotransposon derived sequence *in vivo* argue in favor of a functional homology to proteins of the CENP-B family.

Fourth, chromatin immunoprecipitation was achieved against heterochromatin markers and candidate regions were identified in the sequenced *S. frugiperda* BACs by qPCR with the precipitated genomic DNA as a template. From these data, arguments will be presented supporting hypotheses on the dispersed structure of heterochromatin regions. Further experiments are needed to establish a relationship with holocentromeres, and hopefully the future availability of entire genome sequences of *H. armigera* and *S. frugiperda* will bring light on this question.

These genomics data are gathered in the PestLepido-DB site (<http://www.inra.fr/lepidodb>), a public bioinformatics resource constructed using open source software tools from the GMOD database, including a Chado database, Gbrowse, the graphical tool CMAP, and a curation-visualization tool Apollo. Genes, transcripts, peptides, functional annotations, EST coverage, and repeats are already visible for the sequenced BACs and this site is able to welcome genomic and transcriptomic data that will soon become available for the two noctuids.

Exploring the *Nasonia vitripennis* 'Venome' by Two-dimensional Liquid Chromatography Electrospray Ionization Fourier Transform Ion Cyclotron Resonance.

*de Graaf, Dirk C.; Aerts, Maarten; Brunain, Marleen; Desjardins, Christopher A.; Jacobs, Frans J.; Werren, John H.; Devreese Bart (DGDC, BM, JFJ) Ghent University, Department of Physiology, B-9000 Ghent, Belgium
(AM, DB) Ghent University, Department of Biochemistry and Microbiology, B-9000 Ghent, Belgium
(DCA, WJH) University of Rochester, Department of Biology, NY 14627-0211, USA

The venom glands of parasitoid wasps manufacture various proteins and peptides that influence the arthropod host's physiology. In some cases, parasitoids do not kill their target instantly but instead keep them alive to serve as a fresh food supply for their offspring. The ectoparasitoid wasp *Nasonia vitripennis* (Hymenoptera: Pteromalidae) has been subjected to several studies aimed at understanding the roles of venom during parasitism. Ectoparasitoid wasps develop outside the host body, although they are frequently

attached or embedded in the host's tissues. With the *N. vitripennis* genome sequences available, we tackled its venom composition by a bioinformatic and a proteomic approach. Here we report on the proteomic analyses of crude *Nasonia* venom, implementing a two-dimensional liquid chromatography electrospray ionization Fourier transform ion cyclotron resonance (2D-LC-ESI-FT-ICR MS) setup. LC-ESI-FT-MS analysis on nine SCX fractions resulted in the identification of 91 unique peptides, accounting for 51 proteins. In addition, we also included the SCX flow-through fraction in the LC-ESI-FT MS analysis, which finally resulted in 990 MS/MS spectra identified as *N. vitripennis* peptides, of which only six were decoy hits. After removing redundant peptide identifications, 258 unique peptide sequences were retained, accounting for 76 unambiguously identified proteins with a less than 1% false positive rate. 2D-LC-ESI-FT-ICR MS was proven to be a powerful technique for the analysis of extremely small (venom) samples.

81

Annotation of male specific genes in the aphid *Myzus persicae*

Rebecca P. Duncan; Alex C.C. Wilson

University of Miami, Department of Biology, Coral Gables, FL, 33146

Aphids are cyclical parthenogens that reproduce asexually during most of the year, but give birth to genetically identical sexual females and males (except for the absence of one X chromosome in males) in the Fall in response to shorter days and lower temperatures. Their unusual life cycle makes them an interesting model for studying the genetic basis for polyphenism and the evolution of sex. The newly available genome sequence of the pea aphid, *Acyrtosiphon pisum*, and Expressed Sequence Tag (EST) libraries for another species, *Myzus persicae*, have opened opportunities to explore genome-level differential gene expression among the sexes and different morphs. We identified and annotated male-specific ESTs in *M. persicae* from a microarray comparing expression profiles in males, sexual females and asexual females. We found 167 ESTs that were upregulated in males relative to sexual and/or asexual females, 36 of which were not represented in female EST libraries. Twenty male-specific ESTs were upregulated in males relative to both female morphs. Of the 36 male-specific ESTs, we annotated 16 using *Drosophila melanogaster* and pea aphid gene models as a reference. The remaining 20 ESTs failed to match any known *D. melanogaster* gene. However, four of the 20 unannotated ESTs showed evidence of being present in the pea aphid genome, suggesting that they represent lineage-specific genes. We tested for upregulation of pea aphid orthologs in males using quantitative reverse transcriptase PCR. Interestingly, 10 ESTs correspond to pea aphid gene models that are nested within lineage-specific expansions of two - 19 paralogs. These gene family expansions include two amino acid transporters, a putative juvenile hormone esterase, and a putative peptidase. Male-specific paralogs nested within aphid gene expansions are suggestive of subfunctionalization of duplicated genes, a hypothesis that will be investigated in future work.

5

Prediction of miRNA targets in *Anopheles gambiae* using a word-based, position-independent comparative genomics approach.

*Dunn, W. Augustine; James, Anthony; Xie, Xiaohui

(WAD,AJ) University of California: Irvine, Molecular Biology and Biochemistry, Irvine, CA 92697

(XX) University of California: Irvine, Department of Computer Science, Irvine, CA 92697

MicroRNAs act as small, single-stranded RNAs in conjunction with protein complexes through binding to the 3'-untranslated regions (UTR) of target mRNAs to reduce the amount of target translation. They are generally ~20-22 nucleotides long and exhibit perfect Watson-Crick binding only along a seed region (~6-8 nt) located at the 5'-region of the miRNA. The remainder of its interactions with the target mRNA are rich in secondary structure and can vary from target to target aggravating efforts to predict miRNA targets *in silico*. Some strategies identify seed matches, then attempt to model the binding energies of possible secondary structures of the rest of the miRNA. Others use comparative genomics to predict conserved seed regions through global alignment of orthologous 3'-UTRs.

The currently sequenced species of mosquitoes present a challenge to alignment-based comparative genomics techniques due to their long divergence times (up to 200 million years to last common ancestor) and high levels of repetitive genomic elements like transposons (~47% of the *Aedes aegypti* genome consists of transposable elements). We present here an effort to use a method based on position-independent conservation of seed regions in orthologous 3'-UTRs to predict a subset of miRNA targets in the malaria vector *Anopheles gambiae*.

104

Assessing the Extent of Imprinted Gene Expression in the Honey Bee.

Emore, Christine; Arechavaleta-Velasco, Miguel E.; San Miguel, Phillip; Westerman, Rick; Robles-Rios, Carlos; *Hunt, Greg J.

(CE, GJH) Purdue University, Dept. of Entomology, West Lafayette, IN 47908; (PSM, RW) Purdue University, Dept. of Horticulture,

West Lafayette, IN 47908; (MEA-V, CR-R) Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias, Queretaro, MX

Social insects are considered examples of extreme cooperation between individuals but some social insects may also be examples of extreme intra-genomic conflict leading to selection for imprinting. The expression of genes in worker progeny can have different fitness consequences for honey bee males compared to females. Genomic imprinting results in the expression of alleles of particular genes that is dependent on whether they were maternally or paternally inherited. The honey bee is a promising model for looking at

intra-genomic conflict for several reasons. Honey bees have a fully functional CG methylation system. Honey bees also are polyandrous and have haploid males. Both of these traits theoretically increase selection for imprinting because 1) there is potential competition between the many subfamilies with different fathers for reproduction and 2) there are asymmetries in fitness consequences since the haploid males only have daughters but females have sons and daughters.

We obtained the first evidence for imprinting in honey bees with the discovery that colony-level stinging behavior shows a strong paternal effect. We now have evidence for a paternal effect on individual worker stinging behavior in response to electrical stimulus. For queens, a gene increasing the probability that an individual will sting may show a paternal effect because it would provide an advantage during inter-queen combats that decide which queen will become the sole reproductive female in the colony. In worker bees, paternal imprinting of defensive behavior would increase specialization of subfamilies, which may be advantageous to the colony. After selecting parents with the most extreme behavioral phenotypes for stinging behavior, we performed reciprocal F1 crosses between European and Africanized bees in Mexico. Preliminary sequencing of parents indicated that over half of all predicted genes could be screened for allele-specific expression in F1 workers. We have sequenced pools of cDNA from reciprocal F1 worker bees on the ABI SOLiD platform and are analyzing the sequence for numbers of reads for each allele in each family. The cDNA pools are from multiple individual first instar or adult honey bees. Comparisons of allele-specific expression of reciprocal crosses will be used to scan the genome for evidence of genomic imprinting. Candidate genes for defensive behavior quantitative trait loci will be of particular interest.

6

Evolutionary Analysis as the Basis for Interpreting, Comparing, and Presenting Genomes: The GATOR and PHRINGE System

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

(SIF, JLB) Genome Project Solutions, Hercules, CA 94547; (JLB) University of California, Department of Integrative Biology, Berkeley, CA 94720

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 \$20,000. 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.

78

Genes regulated in early embryogenesis during the switch from parthenogenetic to sexual reproduction in the pea aphid (*Acyrtosiphon pisum*)

* Gallot, Aurore ; Guernec, Grégory ; Colella, Stefano ; Charles, Hubert ; Jaubert-Possamai, Stéphanie ; Tagu, Denis

(AG, SJP, DT) INRA Rennes, UMR 1099 BiO3P, 35653 Le Rheu cedex, France;

(GG) INRA, UR1012 SCRIBE, IFR 140, GenOuest, 35000 Rennes, France;

(SC, HC) UMR 203 Biologie Fonctionnelle Insectes et Interactions (BF2I), INRA, INSA-Lyon, Université de Lyon, 69100 Villeurbanne, France

Aphids adapt easily to environmental changes by phenotypic plasticity. This faculty is one key of their demographic success on crops. A major illustration of aphid phenotypic plasticity is the ability to change of reproduction mode according to seasons. During spring and summer, parthenogenetic reproduction allows a rapid colonisation of the host plants. In the fall, the shortening of photoperiod induces a production of sexual males and females that will mate and produce overwintering eggs, resistant to frost. In sexual individuals germ cells undergo meiosis whereas in parthenogenetic females, the production of diploid oocyte-like cells occurs through a modified meiosis. Since the late 80's and the work of Roger Blackman (Blackman, 1987), nearly nothing new is known on the molecular bases that differentiate and regulate the formation of the oocytes in parthenogenetic viviparous and sexual oviparous aphid females. The aim of this work is to address this issue by a global transcriptomic approach using the newly developed pea aphid oligonucleotide microarray representing 23,855 pea aphid predicted genes. We first defined when the reproduction mode is determined during the development of sexual embryos: we adopted a pharmacological approach using kinoprene, an agonist to

juvenile hormones that reverses sexual reproduction to parthenogenetic reproduction. The results showed that the largest embryos (stage 17) (Miura *et al.*, 2003) within the abdomen of the fourth stage larvae are not yet determined to be sexual individuals, whereas the largest embryos (stage 20) within young adults are already engaged in meiosis and determined to be sexual individuals. We concluded that the determination of the reproduction mode in sexual embryos occurs between stage 17 and stage 20. Based on this observation, we compared the transcriptome of synchronous asexual and sexual embryos at 3 points along a developmental kinetic delimited by stages 17 and 20 as defined above. For each time point, we dissected 25 aphids and collected the 5 most developed embryos per aphid. This material was sampled in 5 biological independent replicates. The comparison of the transcriptomic profiles performed of the oligo-array between sexual and asexual development provided us with a list of 34 genes differentially expressed according to the reproduction mode of the developing embryos. Among these, 16 genes showed homology with *Drosophila melanogaster* genes involved in transcriptional, post-transcriptional and epigenetic regulations. We also found 9 genes homolog to *Drosophila* genes involved in germ cell development and early embryonic development. In particular the gene *nanos* involved into germ cell development in metazoans was identified as upregulated in the asexual reproduction mode. The expression of *nanos* has been previously localised by *in situ* hybridization into germ cells in pea aphid embryos (Chang *et al.*, 2009). These 34 differentially expressed genes identified by this global approach could be involved in the early step of the switch of reproduction mode of the pea aphid embryo in response to the seasonal modification of photoperiod.

7

Perfect(ing) Arthropod Genes with Next Gen Informatics

Gilbert, Don

Indiana University, Biology, Bloomington, IN 47405; gilbertd@indiana.edu

Next generation genome informatics. Software and statistical methods for gene finding and differential expression, using next generation RNA-seq and genome tiling data (NGS), have been developed and tested with arthropod genomes (*Daphnia*, *Drosophila*, and *Nasonia*). Abundant NGS transcription evidence is nearly sufficient to determine gene structures without predictive methods. RNA-seq and genome tiling methods find exon expression with similar power. RNA-seq, like ESTs, add intron splice information; Mate-pair RNA-seq often identifies gene boundaries, solving a major problem in gene modeling. Large amounts of transcriptional "dark matter" do not resolve as coding genes, but likely are long non-coding RNA. Accuracy and sensitivity of gene expression determination is increased with whole-gene NGS methods.

Arthropod gene orthology. Arthropods/euGenes is a comparative database of current gene sets, orthology and summary data, and is recently updated at <http://arthropods.eugen.es>. New OrthoMCL orthology for 263,000 genes of 14 species were computed on the TeraGrid. Web-searchable common gene summary pages describe these orthologs, along with GBrowse genome maps and BLAST services. Analytical summaries include gene structure statistics, EST assemblies, gene quality statistics, and species ortholog gain/loss tables. Certain features of gene structures are conserved (coding length), and others distinguish species (intron sizes and number, gene duplications). EST assemblies are useful to validate gene predictions and genome assemblies. *Daphnia pulex* maintains the strongest homology to human and other non-arthropod eukaryotes, followed by *Ixodes*. Among insects, *Tribolium* has the strongest non-insect homology, partly due to good quality gene modeling. Gene duplication rate is quite variable compared to singleton genes, and lacks a strong phylogenetic trend. As yet unpredicted orthologous genes exist in these genomes, and species gene sets differ in this.

Daphnia Genes 2.0. Gene modeling with NGS data has been applied to *Daphnia*, and is in progress for *Nasonia*, to produce improved, evidence-based gene sets. Evidence of exons (transcription data), introns (EST and RNA-seq), and gene boundaries (mate-pair RNA-seq, protein homology and tandem duplicates) are used with Augustus predictions. The best gene set is extracted from several predictions by combining all evidence with a heuristic method that mimics that of expert annotators. This new gene set for *Daphnia pulex* is much improved from the first set. Protein homology evidence is extensive, and indicates that *Daphnia's* gene set is one of the most complete among arthropods. Notable statistics: Daphnia2 recovers twice as much protein homology (82%) as Daphnia1, Daphnia2 recovers 90% of ESTs, versus 75% for Daphnia1, and Daphnia2 recovers 66% (5 Mb) of genome expression not in Daphnia1.

Discussion. The first *Daphnia* gene set left us with a quandry: most differential expression is in unannotated regions. Genes 2.0 resolve this quandry ... in favor of untranslated regions and non-coding RNA genes that have strong differential expression to environmental stresses. Gene models of your favorite arthropod can also be improved, using extensive and high quality NGS transcriptome data, combined with current arthropod homology, careful modeling of genes from this evidence, attention to non-coding expression, and species-specific evidence.

Transcriptional Changes Related To Maturation And Mating In Male And Female *Ceratitis Capitata* Wiedemann

***Gomulski, Ludvik M.; Dimopoulos, George; Xi, Zhiyong; Gabrieli, Paolo; Scolari, Francesca; Siciliano, Paolo; Malacrida, Anna R.; Gasperi, Giuliano**

(LMG, PG, FS, PS, ARM, GG) University of Pavia, Department of Animal Biology, Via Ferrata 1, 27100, Pavia, Italy; (GD, ZX) Johns Hopkins Bloomberg School of Public Health, Baltimore, USA

The medfly, *Ceratitis capitata*, is a highly invasive agricultural pest that has become a model insect for the development of biological control strategies. These control strategies are reliant on knowledge of the behaviour, physiology and molecular genetics of medfly reproduction. The recent availability of expressed sequence tags derived from adult medfly heads and embryos has permitted the development of microarrays for high throughput gene expression profiling. The genes that show changes in expression as a result of maturation and mating could potentially be used as targets for the development of novel medfly control strategies.

A high density oligonucleotide microarray was designed based on 11885 assembled expressed sequence tags from medfly cDNA libraries derived from 0 - 8 day old male and female heads and from 0 - 36 h embryos. Complementary RNA derived from virgin immature, mature and mated males and females of the ISPRA strain were labelled and hybridised to the microarrays using standard Agilent protocols. The microarray data were filtered, normalized and analysed using TIGR MIDAS, MeV and GEPAS software. Gene Ontology and GO Term Enrichment analysis analyses were performed using GO Slimmer and GOEAST, respectively. The microarray-derived gene expression results were confirmed by comparison with the expression patterns of 12 genes obtained using real-time quantitative PCR (qRT-PCR). Three reference genes were used for relative quantification normalization.

A microarray based gene expression approach was used to compare the head transcriptomes of adult male and female medfly at differential physiological stages, sexually immature and mature virgin individuals and mated individuals. Particular attention was focused on the expression patterns of transcripts belonging to reproduction, behaviour, sensory perception of chemical stimulus, and immune system processes. Compared to the broad transcriptional changes during the maturation of the female, post-mating changes in females were modest, suggesting that in the medfly mating does not trigger extensive transcriptional changes in the head. Of particular interest is the apparent lack of mating-induced immune responses in the medfly female, compared to *D. melanogaster*, which may be the result of the different reproductive strategies of these species.

Our study has revealed intriguing transcriptional changes during medfly maturation and mating. This study will encourage further investigations on numerous questions regarding the biology of this and related species. Apart from increasing our understanding of the molecular machinery behind these biological processes, the implicated genes may represent important targets that could be used to control populations of this pest species.

47

The Genome Sequence of the Leaf-cutter Ant *Atta cephalotes*

Grubbs, Kirk; Marsh, Sarah; Scott, Jarrod J.; Adams, Sandra M., Warren, Wes; Elsik, Christine G.; Clifton, Sandra W.; Taylor, James; Li, Lewyn; Bouffard, Pascal; Harkins, Tim; Slater, Steven; Robertson, Hugh; Viljakainan, Lumi; Smith, Christopher R.; Holt, Carson; Hu, Hao; Yandell, Mark; Smith, Christopher D.; Teiling, Clotilde; Weinstock, George; Gerardo, Nicole M.; Currie, Cameron R.;

*Suen, Garret

(KG, SM, JJS, SMA, CRC, GS) University of Wisconsin-Madison, Department of Bacteriology, Madison, WI 53706; (SS, CRC, GS) University of Wisconsin-Madison, Great Lakes Bioenergy Research Center, Madison, WI 53706; (WW, SWC, GW) Washington University at St. Louis, The Genome Center, St. Louis, MO 63108; (CGE) Georgetown University, Department of Biology, Washington, DC 20057; (JG, NMG) Emory University, Department of Biology, Atlanta, GA 30322; (HR) University of Illinois at Urbana-Champaign, Department of Entomology, Urbana, IL 61801; (LV) Cornell University, Department of Molecular Biology and Genetics, Ithaca, NY 14853; (CRS) Earlham College, Department of Biology, Richmond, IN 47374; (LL, PB, TT, CT) 454 Life Sciences, Branford, CT 06405; (CH, HH, 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

For ~10 million years, leaf-cutter ants have been farming fungi using leaf material. These ants perform the conspicuous behavior of cutting leaves off of trees and transporting this material in spectacular foraging trails back to their colony. These ants do not consume the leaves, but rather use it as a substrate to farm a fungus in specialized subterranean fungus gardens. The success of these ants is highlighted by their widespread distribution in the Neotropics and their massive colonies, which can reach upwards of 5 million workers. One of the keys to this success is the tight, obligate, and coevolved association they have with their cultivated fungus, which serves as the primary food source for the entire colony.

In addition to the fungus they farm, these ants also engage in symbiosis with other microbes. The ant-fungus mutualism is parasitized by a fungus in the genus *Escovopsis*, which can attack and devastate the ants' cultivated fungus. To combat this, the ants form a second mutualism with Actinobacteria, which they maintain on their bodies. These bacteria produce an anti-fungal compound which can selectively inhibit the pathogenic fungus, but has little effect on the fungus the ants cultivate. A third fungus, a black yeast, is also found on the bodies of the ants, and is thought to exploit the ant-Actinobacteria mutualism by competing with the bacteria for

nutrients. Finally, recent work has revealed nitrogen-fixing bacteria that reside in the fungus garden and fix nitrogen for the ants. The combination of four mutualists and two pathogens represents one of the most complex symbioses ever described in nature.

Here, we have capitalized on next-generation 454 pyrosequencing through the Roche 10 Gigabase Sequencing and Transcriptome Analysis Program to sequence the genome of the leaf-cutter ant *Atta cephalotes*. Haploid males were collected from a mature *A. cephalotes* colony in Gamboa, Panama and sequenced at 454 using a combination of whole-genome shotgun, and 8kb and 20kb paired ends. Assembly of > 6 Gb of sequence resulted in a 18-20X coverage genome with a size of ~288 Mb. This assembly produced > 49,000 contigs with an average length of 5,866 bp. Scaffold sequence was also assembled, producing > 12,000 scaffolds with an average length of 27,219 bp and an N50 of > 6 Mbp.

A normalized transcript library for *A. cephalotes* was also generated from workers of various castes and ages, and sequenced. This generated > 38 Mbp of assembled data, and these transcripts were used by the program MAKER to identify 16,942 predicted ESTs for *A. cephalotes*. We are now manually annotating genes in *A. cephalotes* as part of the Ant Genomics Consortium, which includes the Argentine ant *Linepithema humile*, and the Harvester ant *Pogonomyrmex barbatus*. The preliminary annotation work we present here is focused on the areas of immune response, chemosensory, and caste differentiation.

In addition to this ant genome, we are sequencing the genomes of the other symbionts in this system and describe the genomes of the entire community of symbionts. These genomes will be indispensable for characterizing this complex symbiosis, including documenting this system during pathogen induction and the social immune response, understanding how the ant and their cultivated fungus communicate, and also how these ants signal to each other in order to orchestrate this peculiar insect agronomy lifestyle.

8

Comparative Genome Analysis of Microconserved Sequence Elements in Hymenopterans Identify Novel miRNAs

*Hagen, Darren E.; Elsik, Christine G.

Georgetown University, Department of Biology, Washington, DC 20057

MicroRNAs (miRNAs) are small endogenous RNA molecules involved in post-transcriptional regulation of gene expression by targeting messenger RNA (mRNA) causing mRNA destabilization or repression of translation. Insect miRNAs help regulate expression of proteins involved in development, metabolism, and many other important molecular mechanisms and cellular processes. To date, the ongoing genome sequencing and annotation efforts of the ant genome consortia bring the number of hymenopteran genomes available for study to five to date. Here we take advantage of these genomes to identify two sets of conserved miRNAs. The first set of miRNAs is comprised of sequences in the ant that are homologous to miRNA sequences previously identified in other organisms. The second set of miRNAs is the result of using a comparative genome approach to identify microconserved sequence elements (MCE), sequences 20 to 29 bases in length, between the honeybee (*Apis mellifera*), the jewel wasp (*Nasonia vitripennis*), and three ant species: *Pogonomyrmex barbatus*, *Linepithema humile*, and *Atta cephalotes*. Our analysis reveals that as many as 27 million sequences are shared in any three-way genome comparison. By increasing the length of the conserved sequences we are able to confirm sequence homology of a subset of the MCEs. Analysis of putative secondary structures using RNAfold reveals novel miRNA candidates. The software created as a result of this process allows for the rapid detection of novel miRNAs in the ever-expanding number of newly sequenced genomes.

17

Sociality Is Linked to Rates of Protein Evolution in a Highly Social Insect

*Hunt, Brendan G.; Wyder, Stefan; Elango, Navin; Werren, John H.; Zdobnov, Evgeny M.; Yi, Soojin V.; Goodisman, Michael A. D. (BGH, NE, SVY, MADG) Georgia Institute of Technology, School of Biology, Atlanta, GA 30332

(SW, EMZ) University of Geneva Medical School and Swiss Institute of Bioinformatics, 1211 Geneva, Switzerland

(JHW) University of Rochester, Department of Biology, Rochester, NY 14627

Eusocial insects exhibit unparalleled levels of cooperation and dominate terrestrial ecosystems. The success of eusocial insects stems from the presence of specialized castes that undertake distinct tasks. We investigated whether the evolutionary transition to societies with discrete castes was associated with changes in protein evolution. We predicted that proteins with caste-biased gene expression would evolve rapidly due to reduced antagonistic pleiotropy. We found that queen-biased proteins of the honeybee *Apis mellifera* did indeed evolve rapidly, as predicted. However, worker-biased proteins exhibited slower evolutionary rates than queen-biased or non-biased proteins. We suggest that distinct selective pressures operating on caste-biased genes, rather than a general reduction in pleiotropy, explain the observed differences in evolutionary rates. Our study highlights the interaction between highly social behavior and dynamics of protein evolution.

35

Dicistronic gene structures are conserved in several *Drosophilid* species

*Hunter, Henry C.; Smith, Christopher D.

San Francisco State University, Biology Department, 1600 Holloway Avenue, San Francisco, CA 94132

Encoding multiple genes in a single messenger RNA (mRNA) is a strategy common among prokaryotes and viruses, but this approach has only recently been observed in eukaryotic cells. High-quality cDNA-supported annotation of *Drosophila melanogaster* (fruit fly) has revealed over 100 dicistronic genes, where two non-overlapping proteins are expressed from a single mRNA transcript. The mechanisms that mediate downstream dicistronic gene expression in eukaryotes are poorly described, but evidence has shown that RNA structures may aid in translation of the internal gene. Recently, 11 other *Drosophilid* genomes have been completed, providing an excellent comparative genomic system to study the structure of dicistronic genes over evolutionary time. While full-length cDNA evidence supports the existence of these genes in *D. melanogaster*, to date there has been no direct evidence that these genes are also dicistronic in other *Drosophilid* species. We have identified and annotated orthologs of several *D. melanogaster* dicistronic genes and annotated their putative gene structures in multiple *Drosophilid* species. We used these annotations to design PCR primers to confirm the existence of each dicistronic transcript in species-specific cDNA samples. RT-PCR cDNAs were sequenced to verify and refine dicistronic gene annotations in *Drosophilid* genomes. Our results show that dicistronic gene structures are present in multiple fly species and provide clone-based support that will allow further studies to identify candidate RNA structures and cis-sequence motifs that may play a role in dicistronic gene expression. Our data will also provide a detailed comparative system to study the forces that cause merging and splitting of gene structures over evolutionary time.

56

Analysis of Functions of the Cuticular Protein Analogous to Peritrophins (*CPAP*) Gene Family in *Tribolium castaneum*.

*Jasrapuria, S.; Arakane, Y.; Beeman, R.W.; Kramer, K.J.; Muthukrishnan, S.

(SJ, YA, SM, KJK) Kansas State University, Department of Biochemistry, Manhattan, KS, 66506; (RWB) Center for Grain and Animal Health Research, ARS-USDA, Manhattan, KS, 66502.

Insect cuticle is composed mainly of chitin complexed with proteins that contain chitin-binding domains. The two major groups of chitin-binding domains found in cuticular proteins have either the cysteine-rich peritrophin-A motif or the histidine-rich Rebers & Riddiford motif. This study is focused on the identification and functional characterization of genes encoding Cuticular Protein Analogous to Peritrophins (CPAP) proteins with cysteine-containing ChtBD2 domains (Peritrophin A motif = CBM_14 = Pfam 01607) in the red flour beetle, *Tribolium castaneum*. The *CPAP* family of genes is expressed in cuticle-forming tissue and according to the number of ChtBD2 domains is divided into two subgroups. The *CPAP1* group of genes encodes an assortment of 10 proteins of diverse sizes, each with only one ChtBD2 domain with divergent developmental and tissue-specific expression. The *CPAP3* group of genes encodes 8 proteins with three ChtBD2 domains and a domain organization similar to the proteins encoded by the *gasp/obstructor* family of genes of *Drosophila melanogaster*. All the genes of the *CPAP3* family are expressed predominantly in epidermal tissues including the elytra and hindwings. RNAi for many of these genes results in lethal phenotypes that are manifested at different developmental stages, including embryonic, pupal and adult stages. RNAi for *CPAP3-C* results in death at the pharate adult stage. RNAi for other *CPAP3* genes results in abnormal elytra and wings, walking defects, and/or adult mortality. Scanning electron microscopic analysis of pharate adults following RNAi for the *CPAP3-D* gene shows that elytral cuticle is adversely affected. These results provide experimental support for specialization in the functions of individual CPAP proteins of *T. castaneum*.

18

Expansion of the miRNA pathway in the hemipteran insect *Acyrtosiphon pisum*Jaubert-Possamai, Stéphanie¹; Ortiz-Rivas, Benjamin¹; Rispe, Claude; Tanguy, Sylvie; Gordon, Karl; Walsh, Thomas; Edwards, Owain; *Tagu, Denis

(SJP, BOR, CR, ST, DT) INRA Rennes, UMR10S99 BiO3P, Domaine de la Motte, F-35653 Le Rheu, France ; (KG, TW, OE) CSIRO Entomology, GPO Box 1700, Canberra ACT 2601, Australia

¹ These authors contributed equally to this work

Comparative genomics studies of metazoans have shown the microRNA and siRNA pathways to be highly conserved overall; the major difference observed is that the vertebrate pathways overlap in sharing a single Dicer and a single Ago, while those of insects appear to be parallel, with distinct Dicers and Agos required for each pathway.

The genome of the pea aphid is the first available for a hemipteran insect. These aphids reproduce by viviparous parthenogenesis, but show an amazing phenotypic plasticity and are able to produce several adapted phenotypes from one single genotype. We found an unexpected expansion of the miRNA pathway in the pea aphid genome. It has two copies of the miRNA-specific Dicer-1 and Argonaute-1 genes, and four copies of Pasha, a cofactor of Drosha involved in miRNA biosynthesis.

Furthermore, for three of these duplication events (Dicer-1, Argonaute-1 and Pasha 3/4 vs. Pasha 1/2), one of the duplicated copies displays a much higher rate of molecular evolution than the other, slow-evolving copy. For a better evaluation of selective pressures and evolutionary divergence at both intra and inter-specific levels, we have obtained partial sequences of both copies of the gene Argonaute-1 for different *A. pisum* clones, as well as for other aphid species belonging to the same tribe Macrosiphini and the sister

tribe Aphidini. The analysis of the sequences suggests that the duplication of the gene occurred in the common ancestor of Aphidini and Macrosiphini, and that the copy Ago-1b had an accelerated rate of evolution since duplication. This copy shows a very high ratio of non-synonymous to synonymous substitutions, usually greater than 1, suggesting that positive selection has been the driving force during the evolution of Ago-1b. This could follow the acquisition of a new function for this copy. The analysis of synonymous distances suggests that the duplication of Argonaute-1 occurred concomitantly, within a brief period of time, with the ones affecting Dicer-1 and the copies 1 and 2 of Pasha, while the other duplications of Pasha were older.

Besides, we show by RT-PCR that all the expanded copies of the miRNA machinery genes are expressed in the different phenotypes of aphids. Investigating the function of these novel genes, identifying the small non-coding RNA associated with the duplicated proteins and determining if they reflect a new pathway offer an exciting new challenge in aphid biology.

100

Pyrosequencing of tick immune response genes

Jaworski, D.C.*; Zou, Z.; Bowen, C.J.; Wasala, N.B.; Madden, R.; Wang, Y.; Kocan, K.M.; Jiang, H.; Dillwith, J.W. (DCJ, ZZ, CJB, RM, YY, HJ, JWD) Oklahoma State University, Entomology and Plant Pathology; Veterinary Pathobiology (KMK), Stillwater OK 74078.

Ticks continue to be a threat to animal and human health, and new and novel control strategies are needed for ticks and tick-borne pathogens. The characterization of the tick-pathogen interface and the tick immune response to microbial infections is fundamental toward the formulation of new control strategies for ticks and the pathogens they transmit. Our overall hypothesis for this research is that the tick immune system manages the maintenance of pathogens. Therefore, discovery of tick immune response genes may provide targets for novel control strategies directed toward reducing vector competency and pathogen transmission. In these studies, 454 pyrosequencing, a high-throughput genomic sequencing method was used to discover tick genes expressed in response to bacterial and fungal infections. Expressed sequence tags (ESTs) were analyzed from *Dermacentor variabilis* ticks that had been injected with bacteria (*Escherichia coli*, *Bacillus subtilis*, *Micrococcus luteus*) or fungi (*Saccharomyces cerevisiae* and *Candida albicans*) and ticks that were naturally infected with the intracellular bacterium, *Anaplasma marginale*. By this approach, ESTs were assembled into 5995 contigs. Contigs fell into the five main functional categories of metabolism, genetic information processing, environmental information processing, cellular processes and human diseases. We identified more than 30 genes that are likely to encode for proteins involved in tick immune function. We further analyzed by RT-PCR the expression of 22 of these genes in each of our bacterial or fungal treatment groups and found seven were up-regulated. Upregulation of these seven genes was confirmed for bacterial, but not fungal treatment by qPCR. One of these products was novel, encoding for a new tick defensin. Our results clearly demonstrate the complexities of the tick immune system and mark new directions for further study and characterization of proteins that modulate microbial infections in the American dog tick.

66

Identification of a CTCF Binding Site in the Malaria Vector *Anopheles gambiae*

Michael Johanson*, Igor Sharakhov[‡], Maria Shradhova[‡], and Craig Coates*

[‡]Department of Entomology, Virginia Tech University, Blacksburg, VA 24061

*Disciplinary Program of Genetics & Department of Entomology, Texas A&M University, College Station, TX 77843

Transgenic mosquitoes are important in the study of vector pathogen interactions, insecticide resistance, and other aspects of mosquito biology. However, the effectiveness of transgenic mosquitoes is limited by position effect and position effect variegation caused by enhancer/silencer crosstalk and heterochromatin expansion respectively. CTCF is a DNA binding protein that has been identified in vertebrates, *Drosophila* and mosquito species. Studies in vertebrates and *Drosophila* demonstrate that this protein plays a role in insulating genes from enhancers, silencers and heterochromatin that would otherwise interfere with their normal expression profiles. Using the bioinformatics program, Patser, we have identified sequences throughout the *Anopheles gambiae* genome that are similar to the *Drosophila* CTCF binding site consensus (Bartkuhn *et al.*, 2009). These sequences are located in chromosomal regions that have been identified by CTCF immunostaining to polytene chromosomes of ovarian nurse cells and at euchromatin/heterochromatin boundaries. A clone of one of these binding sites has the potential to improve mosquito transgenics by cloning it into a transgene construct so that it can recruit endogenous CTCF that would insulate the transgene from potential position effect and position effect variegation. A potential binding site near a euchromatin/ heterochromatin boundary on the 2R chromosomal region has been validated using ChIP PCR. We plan to validate more of these sites using ChIP PCR, as well as identify more potential binding sites using ChIP-Seq of which a subset will be validated using ChIP PCR.

48

Rational for sequencing the genome of the flesh fly, *Sarcophaga crassipalpis*

Karl H Joplin^{1,2}

¹Department of Biological Sciences, ²Institute of Quantitative Biology, East Tennessee State University, Johnson City, TN, 37614 USA, joplin@etsu.edu, 423-439-6921 (ph) 423-439-5958 (fax)

Flesh flies have been valuable model organisms to study a variety of physiological, hormonal and developmental questions. Although they are not an important vector of disease, they are an important, if underappreciated, component of the recycling of organic carcasses and are used in forensic studies. From a biological standpoint, an important aspect of their lifecycle is their response to environmental cues to trigger a decision to enter the alternative life stage, diapause. Sequencing of the flesh fly will provide a crucial necessary to fully explore these processes. Diapause has important consequences to questions about control of developmental stages, aging, metabolic control, and immune response. An overview of the areas that would be amenable to genomic analysis include:

- Isolation of diapause differentially regulated genes and associated comparison with aging genes.
- Genetic control of metabolism associated with the inducible state of metabolic suppression during diapause.
- Molecular response to environmental cues where short daylength induces an optional developmental state.
- Genetic Regulatory Networks of developmental states.
- Pervasive Genomic Transcription and regulatory control of gene expression.
- Host-parasitoid interactions, comparison with *Nasonia* and *Musca*.
- Comparative genomic analysis with the growing genomic datasets of the Diptera clade such as the *Drosophila* group, *Musca*, *Aedes* and *Anopheles* genomes.

The community of researchers involved in flesh fly research is ready, prepared and eager to apply genomic analysis to their various research programs. The continued development of next generation sequencing and analysis platforms will enable the genomic projects of many model systems and the Sarcophaga project would be a leading candidate for sequencing. Acquisition of a complete genomic dataset will be the starting point for following investigations into the genomic transcriptional control and genetic regulatory networks that are becoming increasingly important components in the control of distinct biological states such as developmental life stages and gene expression and regulatory controls as physiological responses to the environment.

89

The Pyrokinin/PBAN Family of GPCRs: Evolutionary Trace Analysis Indicates Putative Ligand Recognition and Binding Sites

Jurenka, Russel A.; *Nusawardani, Tyasning

(RAJ, TN) Iowa State University, Department of Entomology, Ames, IA 50011

The pyrokinin/pheromone biosynthesis-activating neuropeptide (PBAN) family of G-protein coupled receptors (GPCRs) and their ligands have been identified in various insects, such as *Drosophila*, several moth species, mosquitoes, *Tribolium castaneum*, *Apis mellifera*, *Nasonia vitripennis*, and *Acyrtosiphon pisum*. Physiological functions of pyrokinin peptides include muscle contraction in many insects, whereas PBAN regulates pheromone production in moths; plus other functions indicating the pleiotropic nature of these ligands. Based on the alignment of annotated genomic sequences, the primary and secondary structures of the pyrokinin/PBAN family of receptors have similarity with the corresponding structures of the Cap2b-receptors of insects and the neuromedin U receptors found in vertebrates. Neuromedin U peptides regulate important physiological functions including feeding behavior in vertebrates. Studies of the pyrokinin/PBAN receptor structures are needed to understand the mechanisms of ligand-receptor binding specificity leading to receptor activation. In our study, evolutionary trace (ET) analysis on the identified insect receptor sequences was conducted to predict the putative ligand recognition and binding sites of the pyrokinin/PBAN family of receptors. The ET analysis of three class-specific receptors indicated several amino acid residues that are conserved in the transmembrane domains that are embedded in the cell membrane. The receptor extracellular domains exhibit some class-specific amino acid residues, which could indicate putative domains for activation of these receptors by ligand recognition and binding. The results of this research will help to engineer site-specific mutant receptors that target critical residues involved in ligand recognition and binding.

20

Geographical analysis on South-west islands for evolution of *Bombyx mandarina* (Japan) using rDNA sequence.

Kawanishi, Yuichi; Azimura, Hiroshi; Fujimoto, Hirofumi; Tu, Zhenli; Banno, Yutaka; Kamauchi, Yu; Nho, SiKab; Mita, Kazuei; Yamamoto, Kimiko; Nakajima, Yumiko; *Maekawa, Hideaki

(YK, YK, YN, HM: University of the Ryukyus, Nishihara, Okinawa, Japan 903-0213, HA, KM, KY: NIAS, YB:Kyushu Univ., HF:NIID, .ZT: Zhejiang Univ., S-KN: Kyungpook National Univ.)

The old Eurasian Continent had connected Japan islands, the Korean Peninsula, South-west islands and South-east China at Miocene about 7 my ago. Japanese *B. mandarina* contains characteristic rDNA sequences between 5.8S and 28S region different from *B. mori* type rDNA one. Two types might be separated 3 my ago and Japanese *B. mandarina* evolved independently. Sequences of rDNA cloned from genomic DNA of *B. mandarina* inhabiting China, Korea, Taiwan and Japan were compared and phylogenetically analyzed. Both Japanese *B. mandarina* type and *B. mori* type were isolated from almost Japanese *B. mandarina*. Phylogenetic analysis suggests that the both types must be separated before the period branched among Taiwan, Korea and China, respectively. Such *B. mori* type found in Japanese *B. mandarina* might be derived from which *B. mandarina* population with the *B. mori* type might be moved to Japanese islands area about 50 thousand years ago when Japan islands had re-connected to South-east Chinese area of the old Eurasian Continent lifted up on East China Sea. The *B. mori* has 28 chromosomes, however, this *B. mori* type rDNA population must be 27 chromosomes which might be derived from the 27 chromosomes population occurred around Korean Peninsula few million years ago. After the moving, the *B. mori* type population mated to Japanese *B. mandarina* type population with 27 chromosomes which has been

inhabiting Japan. Consequently, this speculation could explain that *B. mandarina* (Japan) populations with both types of rDNA are observed in present Japan.

21

Isolation and phylogenetic analysis of three types of *mariner*-like elements in the big-headed ant, *Pheidole megacephala* (Hymenoptera: Formicidae)

*Kawanishi, Y.; Ohnishi, H.; Kasahara, A.; Tsuji, K.; Maekawa, H.; Nakajima, Y.
(YK, AK, HM, YN) University of the Ryukyus, Center of Molecular Biosciences, Nishihara, Okinawa 903-0213; (HO, KT) University of the Ryukyus, Faculty of Agriculture, Nishihara, Okinawa 903-0213

The big-headed ant, *Pheidole megacephala*, is on the list of "100 of the World's Worst Invasive Alien Species" presented by Invasive Species Specialist Group. This ant is believed to originate in Africa and has been distributed over the tropical and subtropical area of the world with the expansion of human activities. It is important to elucidate route and time of invasion of this ant for estimation of an impact on the invaded ecosystem. We analyzed the origin of the ant inhabiting the South-West Islands of Japan by phylogenetic analysis of full-length *mariner*-like elements (MLEs) including the specimens from Africa.

We used the genomic DNA of the specimens of the ant collected from the South-West Islands (Okinawa main Is., Minami-daito Is., Miyako Is., Ishigaki Is. and Amami Is.), Gabonese Republic (Libreville) and French Reunion Island (Le Porto, Le Grand Brue) on the east of Madagascar. Partial sequences of MLE were isolated by preparatory experiment. These showed similarity to the MLE, *Myrmar2* of the ant, *Myrmica ruginodis*. We isolated full-length MLEs by PCR amplification using the sequence of the inverted terminal repeat (ITR) of *Myrmar2* as a primer. Three types of MLEs, *Pmmar1*, *Pmmar2* and *Pmmar3* were isolated from specimens. *Pmmar1* and *Pmmar2* were defective elements due to multiple frameshifts and nonsense mutation in their ORFs encoding the transposase, whereas *Pmmar3* inserted into *Pmmar1* was an intact element with a complete ORF. Phylogenetic analysis revealed that MLEs of the Reunion specimens were closely related to those of the South-West Is. specimens rather than to those of the Libreville specimens. This may suggest that the population of the ants in South-West Is. and Reunion have same origin. We infer that it is one of the possibilities ant invaded into the South-West Is. of Japan from Reunion along the old trading route about several hundred years ago. Furthermore, sequence comparison of the MLEs inserted into the specific site of the ant genome revealed that there is a unique allele to MLEs of specimens of the South-West Is. This may imply that there was different invasive route to the South-West Islands from the route via Reunion.

37

Metagenomic analysis of honey bees to identify stress/disease indicators

*Khetani, Radhika S.; Evans, Jay D.; Hudson, Matthew, E.
(RSK) University of Illinois at Urbana-Champaign, Institute for Genomic Biology, Urbana, IL 61801; (JDE) USDA-ARS Bee Research Laboratory, Beltsville, MD 20705; (MEH) University of Illinois at Urbana-Champaign, Department of Crop Sciences, Urbana, IL 61801

Honey bees, *Apis mellifera*, are the largest group of agriculturally-relevant pollinators across the world, and recent resounding losses in honey bee populations have dramatically impacted crop yields. The reasons for honey bee loss have been difficult to narrow down, but it is generally accepted that they result from multiple stress factors, including pathogens and environmental factors. Our research is focused on identification of biomarkers that can help in the identification of causes and also aid in the development of responses against stress and disease in honey bees. We are employing next-generation sequencing technology, specifically Illumina, to perform RNA metagenomic and RNA-seq analyses. For the metagenomic data analysis, we perform sequential alignments against several specific databases. This method reduces computational requirements and also readily separates the data into manageable pieces for analysis. Bees subjected to a specific stressor or multiple stressors in a lab environment, as well as bees collected from the field are utilized for these analyses. *Varroa destructor* is a common external parasitic mite and *Varroa* infestation impacts both individual bees and their colonies. As a result of physiological impacts and the efficiency of these mites as disease vectors, mite infestation is correlated with very high levels of RNA viruses in their bee hosts. Here I will present the methods employed, and the results obtained from the metagenomic analysis of bees from mite-infested colonies.

69

Transcriptome Analysis Identifies Over-Expressed Metabolic Genes in Pyrethroid-Resistant *Cimex lectularius*.

Kilcullen, Kathleen; Koganemaru, Reina; Anderson, Michelle A. E.; Miller, Dini; *Adelman, Zach N.
Virginia Polytechnic Institute and State University, Fralin Life Science Institute and Dept of Entomology, Blacksburg, VA 24061

Cimex lectularius, the common bed bug, has been rapidly expanding its range and abundance over the past ten years both within the US and abroad. This is thought to be due, at least in part, to the fact that bed bugs have become highly resistant to pyrethroid-based insecticides. While at least one case of target site resistance has been identified, other common forms of resistance, such as metabolism, have not yet been explored in detail. As no genes involved in the bed bug P450 or GST metabolic pathways have been described, we performed a transcriptomic analysis on both insecticide-resistant and susceptible bed bugs using the 454 Titanium platform. Following assembly of more than 2 million reads, we identified eleven putative GSTs and 34 putative P450 genes. Based on

a comparison between the number of reads obtained from each dataset, we observed that several of these genes were significantly up-regulated in the pyrethroid-resistant strain. Using real-time reverse-transcriptase polymerase chain reaction, we confirmed that several members of the CYP3 clan are indeed substantially up-regulated in resistant bed bugs. This data strongly suggests that pyrethroid resistance in US bed bugs has a metabolic component and provides a starting point for genetic and molecular studies in this insect.

92

Hemocyte transcriptome analysis of *Spodoptera exigua* using 454 pyrosequencing

Yonggyun Kim

Department of Bioresource Sciences, Andong National University, Andong 760-749, Korea

Hemocytes are the major tissue to express immune responses consisting of both cellular and humoral reactions. Transcriptome analysis of hemocytes of the beet armyworm, *Spodoptera exigua*, would be expected to propose immune-associated genes and their signal pathways. Total RNAs were extracted from the fifth instar larvae challenged with *Escherichia coli* and used for constructing cDNAs. Based on 100 bp cut off, 11,962 singletons were determined and used to produce 10,589 unique contigs. Only 18% clones were matched to functionally known genes, which were classified according to genetic ontology. About 50% clones matched to known genes were classified to immune-associated categories of cellular communication, defense, and interaction with the environment. Three immune signal genes (Toll, Relish, GPCR), two phenoloxidase genes (serpin and dopa decarboxylase), and several antimicrobial genes were further analyzed in their expression profiles in response to various microbial pathogens. This study suggests that *S. exigua* can express immune responses using Toll and Imd signal pathways.

23

Analysis of the *Tribolium* Genome Reveals Sequence Similarity to Bacterial Genes

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

Analysis of the *Tribolium castaneum* genome revealed the presence of a unique class of genes encoding proteins with high similarity to members of the bacterial PD-(D/E)XK nuclease superfamily. These genes, which encode restriction endonuclease-like domains found in many hypothetical bacterial proteins, have not been reported in other sequenced arthropods. Family members encode approximately 675 amino-acid residues each, spanning a recently reannotated N-terminal region that encodes an AAA-ATPase like protein domain (pfam09820), as well as a unique C-terminal region that encodes a bacterial Domain of Unknown Function (DUF) 1703 (pfam08011). 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. Phylogenetic analysis indicates that these genes have undergone expansion within the *Tribolium* genome, encompassing at least five subfamilies which are scattered throughout the genome (at least 8 different linkage groups). Tiling array data and RT-PCR indicate transcription of the largest subfamily (8 members). To explore whether these genes may be functional, dsRNA specific for one subfamily was synthesized, injected into penultimate instar larvae and knockout assessed by RNAi phenotype and RT-PCR. All insects receiving this dsRNA treatment died, indicating that at least one of these bacterial genes has acquired an essential function in a eukaryote. Gene expression patterns and functions of the other four subfamilies is currently being examined.

98

Searching the suitable internal control genes for real-time quantitative RT-PCR studies in *Ixodes scapularis* tick.

*Koci, J., Simo, L., Park, Y.

Kansas State University, Entomology Department, Manhattan, KS 66506

Obtaining reliable gene expression data using real-time quantitative RT-PCR (qRT-PCR) is highly dependent on the choice of the method for normalization. Although several house keeping genes were widely used for normalization in the quantifications of transcripts, systematic studies for identifying the best gene for normalization can be done for the samples in each experimental sets. We tested multiple candidate genes for their expression stability on two different tick tissues, salivary glands (SG) and synganglia (SYN), in multiple blood feeding phases. We chose a set of single copy house keeping genes predicted in the genome data of blacklegged tick *Ixodes scapularis*. Those six genes were: glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*), ribosomal protein L13A (*L13A*), TATA box-binding protein (*TBBP*), ribosomal protein S4 (*rRNAS4*), glucose 6-phosphate dehydrogenase (*GPDH*) and beta-glucuronidase (*GUSB*). The program geNorm (Vandesompele et al. 2002) was used for analysis of data from different feeding phases (daily samples in the 8 days from unfed to fully engorged females) in three biological replications. In the tests on SG, the rank of the six candidate reference genes from the most stable (normalized average lowest M) to the least stable was as follows: *rRNAS4+L13A* (0.15) > *GPDH* (0.26) and *TBBP* (0.31) > *GAPDH* (0.56) > *GUSB* (1) As for SYN, the rank was *rRNAS4+L13A* (0.37) > *TBBP* (0.73) > *GPDH* (0.81) > *GAPDH* (0.90) > *GUSB* (1). In both tissue samples, the most stable relative expression was found for *rRNAS4* and *L13A*, and the least stable expression was found for *GUSB*.

70

Transcriptional Analysis of Putative Cuticle Genes from a Field-Collected, Pyrethroid-Resistant Strain of the Common Bed Bug, *Cimex lectularius*.

*Koganemaru, Reina; Anderson, Michelle A. E.; Adelman, Zach N.; Miller, Dini M.

Virginia Polytechnic Institute and State University, Department of Entomology, Blacksburg, VA 24061

The common bed bug *Cimex lectularius* has become highly resistant to pyrethroid insecticides and has been growing in numbers over the past ten years in the U.S. A preliminary study of LD50 values revealed that changes in the cuticle might play a role in resistance in field-collected bed bugs. As no genes involved in the bed bug cuticle had been described, we performed a transcriptomic analysis using the 454 Titanium platform. Following assembly, we identified at least twenty four putative cuticle genes based on the presence of a chitin-binding domain (Pfam00379). Using two step real-time reverse-transcriptase polymerase chain reaction, we measured the level of transcription from each of these genes in both pyrethroid-resistant and susceptible bed bugs. Multiple genes from the field-collected, pyrethroid-resistant strain showed a substantial increase in transcription compared to a laboratory susceptible strain. In addition, the thickness of the abdominal cuticles from both strains of bed bugs was measured using scanning electron microscopy (SEM). This study reveals another method of pyrethroid resistance that was not reported previously in common bed bugs.

42

Use of Gal4-Mos1 and Gal4-piggyBac Chimeric TransposaseS for Transformation in insects & MAMMALIAN SYSTEMS

*Kommineni, Maragathavally; Coates, Craig

Texas A&M University, Department of Entomology, College Station, 77843-2475

Efficient integration of a functional gene, together with suitable transcriptional regulatory elements, is an essential prerequisite for successful gene integration in both insect and mammalian systems. Many current gene integration strategies are based on viral vectors; however, limited cargo capacity, host immune response, and the risk of insertional mutagenesis are limiting factors. Genetic transformation systems based on MosI and piggyBac transposable elements are used to achieve stable chromosomal integration. Particularly, the piggyBac transposase has been shown to be well suited for cell transfections and gene therapy approaches because of its flexibility for molecular modification, large cargo capacity and relatively high transposition activity. Since the piggyBac integration sites (TTAA) are randomly distributed in the genome, transgene expression can be influenced by position effects and position effect variegation.

We developed a novel technology to direct gene integration into a specific site on a target DNA molecule that utilizes a chimeric transposase in insect systems, which functioned subsequently in mammalian systems. The Gal4 DNA binding domain (DBD) is fused to the N-terminus of the Mos1 and piggyBac transposase coding regions (helper plasmids) and a target plasmid is created with upstream activating sequences (UAS). The Gal4 DBD binds to the UAS binding site with high affinity. The introduction of a DBD to the transposition machinery can potentially increase transformation efficiency and result in site-specific integrations (Maragathavally et al, 2006).

In insect systems, we obtained preliminary evidence of site-specific genomic integration with the Gal4-Mos1 and Gal4-pB chimeric transposase constructs. Transgenic lines of *Drosophila melanogaster* were obtained that contain a chromosomal UAS target array. *Drosophila* line UASLeoII#202 SA 126 9/2 has UAS sites on chromosome 3 and the UAS MycFSA 130 9/2lines has UAS site on chromosome 2. These strains were used as the recipients for a small pilot experiment using the Gal4-mariner and Gal4-piggyBac helper plasmid. Transgenic experiments with Gal4-pB resulted in 44-45% transformation efficiency and using Gal4-Mos1 resulted in 9-12% transformation efficiency, demonstrating that the helper plasmids are functional for producing germline transformants. EGFP expression was observed in the transgenic *Drosophila* lines and Southern blot analysis of these transgenic lines showed transgene integration in the genome. A number of lines showed a similar hybridization pattern, indicating a potential common insertion site or region. Inverse PCR analysis of a subset of transgenic lines reveals piggyBac and mariner insertions in genomic DNA. Interestingly, transgene insertions are in close proximity to the UAS binding site in the genome, indicating site-specific integration of the transgene. Preliminary studies demonstrate that the Gal4 chimeric transposases can result in site-specific integration.

To achieve targeted integration in mammalian systems, we have engineered chimeric ZnFn-mpB and GAL4-mpB constructs with the mouse codon-biased piggyBac transposase (mpB) gene. ZnFn-mpB and GAL4-mpB chimeric helper plasmids facilitate gene integrations in zinc finger binding sites and UAS landing sites respectively. We have also developed helper-independent single chimeric piggyBac plasmids with Gal4, mice and human zinc fingers fusions. Transected HEK293 cells and mice fibroblast cells with UAS target array revealed successful transgene integrations. EGFP expression and Southern blot analysis of transformed cell lines showed transgene integration in the human and mice genomes.

24

Linkage and expression of sex-linked genes on the *Ostrinia nubilalis* Z chromosome

*Kroemer, Jeremy A.; Coates, Brad S.; Nusawardani, Tyasning; Hellmich, Richard L.

(JAK, BSC, RLH) United States Department of Agriculture, Agricultural Research Service, Corn Insects & Crop Genetics Research Unit, Ames, IA 50011, (TN) Iowa State University, Department of Entomology, Ames, IA 50011

Genes that control voltinism (univoltine vs. bivoltine) and the male pheromone response are on the Z-chromosome in Lepidoptera. Deciphering the genetic basis for these ecological traits will enhance our ability to understand gene flow within and among populations that vary at these loci. We have designed 50 degenerate PCR primer pairs from the *Bombyx mori* Z chromosome scaffold and are analyzing their utility for gene expression analyses, sequence determination, and as polymorphic genetic markers for linkage mapping in *O. nubilalis* and other pest lepidopteran species. 36 of the degenerate primer pairs produced a single product from *O. nubilalis* genomic DNA, of which 10 markers were polymorphic and were integrated into a linkage map and positioned on a partial physical map constructed from bacterial artificial chromosome (BAC) library clones. Synteny and co-linearity of the *O. nubilalis* markers were analyzed via comparison to the *Bombyx mori* genome assembly for chromosome 1 (Z). Tissue and growth-stage specific gene expression profiles for *O. nubilalis* genes were compared with orthologs existing in *B. mori* microarray data. The higher resolution Z chromosome analysis from *Ostrinia nubilalis* will provide novel tools for eventual QTL mapping of sex-linked traits and identifying the specific genes that contribute to male pheromone response and voltinism differentiation in this and potentially other pest Lepidoptera.

65

Possible Immune Functions of Two Mosquito Multicopper Oxidases

*Lang, Minglin; Gorman, Maureen J.; Gardner, Stewart G.; Yungeberg, Sandi L.; Kanost, Michael R.

(ML, MJG, SGG, SLY, MRK) Department of Biochemistry, Kansas State University, Manhattan, KS 66506; (ML) College of Life Science, Agricultural University of Hebei, Baoding, China 071001

The multicopper oxidase (MCO) family of enzymes has been implicated in a diverse set of physiological processes; however, little is known about their function in insects, including the malaria mosquito, *Anopheles gambiae*. Our goal is to characterize the function of two *Anopheles* MCOs: MCO1 and MCO3. We found that these two genes are upregulated in the midguts and Malpighian tubules of mosquitoes following a blood meal and in other tissues in response to infection with bacteria or malaria parasites. These results suggest that MCO1 and MCO3 may function in defense against pathogens. Preliminary data suggest that double silencing of MCO1 and MCO3 increases the susceptibility of mosquitoes to malaria infection. Overexpression of MCO1 or MCO3 in *Drosophila* increases their sensitivity to paraquat exposure, suggesting that MCO activity may lead to ROS production. If this is true, the MCOs may participate in an ROS killing mechanism against the parasites. Alternatively, the MCOs may restrict pathogen growth by acting as ferroxidases that reduce the amount of available iron. Finally, MCO1 and/or MCO3 may be required for protein crosslinking in the peritrophic matrix, which acts as a natural barrier against parasite invasion of the gut. We will present the results of experiments designed to test these three hypotheses. The work was supported by NIH Grants R03AI057816 and R01AI070864.

3

Ensembl Genomes: extending Ensembl across arthropod taxonomy

*Lawson, D.; Derwent, P.S.; Haimel, M.; Herrero, J.; Keenan, S.; Kerhornou, A.; Koscielny, G.; Kähäri, A.; Kinsella, R.J.; Kulesha, E.; Maheswari, U.; Megy, K.; Nuhn, M.; Proctor, G.; Staines, D.; Valentin, F.; Vilella, A.J.; Yates, A.; Birney, E.; Kersey, P.J. 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.

We will describe the infrastructure of Ensembl Genomes, criteria for inclusion into the portal and plans for integrating further arthropod genomes in the coming year.

2

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.

Many more vector genomes will be sequenced in the coming years including new species (e.g. the vectors for Leishmaniasis (*Lutzomyia* and *Phlebotomus*) and Sleeping sickness (*Glossina*)) as well as more species/strains from clades already sampled (e.g. 12 genomes from the *Anopheles* species complex).

38

Meta-omic Profiling of Lignocellulose Degradation in the Hindgut of a Wood-feeding Cockroach

*Li, Xiangrui; Geib, Scott M.; Zhou, Xuguo

(XL, XZ) University of Kentucky, Department of Entomology, Lexington, KY 40546; (SMG) USDA-ARS-PBARC, 64 Nowelo St., Hilo, HI 96720.

The obligate relationship of both wood-feeding cockroaches in the genus *Cryptocercus* and basal lower termites with lignocellulolytic flagellate protists in their hindguts has long been a textbook example of symbiosis. *Cryptocercus* is sister group to Isoptera, and shares many similarities with basal termite groups. Similar to the lower termite, *Reticulitermes flavipes* (Kollar) (Isoptera: Rhinotermitidae), a wood-feeding cockroach, *Cryptocercus punctulatus* Scudder (Blattaria: Cryptocercidae), and their gut symbionts have co-evolved into an efficient mini bioreactor including a suite of specialized enzymes that synergistically break down lignocelluloses, the most abundant and sustainable solar batteries on earth. Here, we employed a meta-omics approach to inventory a comprehensive set of genes/proteins potentially involved in the lignocellulose degradation. In this report, we present the initial metaproteomics analysis of the microbial community residing in the hindgut paunch of *C. punctulatus*. Total protein was extracted from the luminal contents of adult woodroach hindguts. The protein sample was then subjected to trypsin digestion and followed by peptide identification using MALDI-TOF/TOF mass spectrometry. Several peptide libraries were created by translating EST gut and symbiont datasets from *R. flavipes*, and were used for peptide identification. The initial protein identification using the ProteinPilot Paragon algorithm yielded hundreds of significant matches in all MS analyses. These matches were then annotated by searching the corresponding EST against the NCBI database to provide annotation and putative classification of peptides. More thorough ID will be carried out upon the conclusion of a parallel metatranscriptomic sequencing of the exact sample submitted for the metaproteomic analysis. Given the fact that *Cryptocercus* and *Reticulitermes* share many flagellate symbionts such as oxymonadid and hypermastigid in their hindguts, it is germane to survey the homogeneity of lignocellulases across all known insect species with the most basal *C. punctulatus* as a reference using a comparative genomic approach.

22

Complete resequencing of 40 silkworms and single base-resolution methylome of the silkworm

*Li, Zhuo, Qingyou Xia

(LZ) BGI Shenzhen, Shenzhen 518083, China; (XQY) The Key Sericultural Laboratory of Agricultural Ministry, College of Biotechnology, Southwest University, Chongqing, 400715, China.

A single-base pair resolution silkworm genetic variation map was constructed from 40 domesticated and wild silkworms, each sequenced to approximately threefold coverage, representing 99.88% of the genome. We identified ~16 million single-nucleotide polymorphisms, many indels, and structural variations. We find that the domesticated silkworms are clearly genetically differentiated from the wild ones, but they have maintained large levels of genetic variability, suggesting a short domestication event involving a large number of individuals. We also identified signals of selection at 354 candidate genes that may have been important during domestication, some of which have enriched expression in the silk gland, midgut, and testis.

We also survey the methylome of the silkworm *Bombyx mori*, at single-base resolution using Illumina high-throughput bisulfite sequencing (MethylC-Seq). We conservatively estimate that 0.11% of genomic cytosines are methylcytosines, all of which probably occur in CG dinucleotides. CG methylation is substantially enriched in gene bodies and is positively correlated with gene expression levels, suggesting it has a positive role in gene transcription. We find that transposable elements, promoters and ribosomal DNAs are hypomethylated, but in contrast, genomic loci matching small RNAs in gene bodies are densely methylated. This work contributes to our understanding of epigenetics in insects, and in contrast to previous studies of the highly methylated genomes of *Arabidopsis* and human, demonstrates a strategy for sequencing the epigenomes of organisms such as insects that have low levels of methylation.

79

Analysis of the Soybean Aphid Transcriptome

Liu, Sijun, Nanasahab Chougule and Bryony B. Bonning

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

The soybean aphid, *Aphis glycines* Matsumura (Hemiptera: Aphididae) has caused significant economic loss in the Midwest since its detection in the United States in 2000. Genomic information is essential for in depth functional studies of soybean aphid physiology and biochemistry toward development of novel approaches for management of this economically important aphid. Transcriptomic data were generated from whole aphids and from 2,000 aphid guts using an Illumina GAII sequencer that generated 8 million reads of 75 nt per sample. Eland software (Illumina Corp.) used to align sequences to pea aphid contigs resulted in only 20% alignment to the pea aphid genome, and MAQ software only mapped 13.8% and 6.6% of the gut and whole aphid reads respectively to the pea aphid genome. The soybean aphid sequence data were therefore assembled de novo using the Velvet and ABySS assemblers. Multiple sets of contigs were generated by varying parameters (hash length k-mers, and coverage cut-off). Each contig set generated unique hits on BLASTX analysis against the NCBI nr database, with 3,300 to 8,500 genes identified per set. Of the 15,000 unique hits, 92% aligned to genes of the pea aphid or other aphid species, representing about half of the pea aphid genes. The remaining 8% of the hits aligned primarily to genes of other insects. On examination of functionally related pea aphid gene sets relating to development, immunity and stress, amino acid metabolism, sugar transport and ion channels, 80.1% of the 1,430 pea aphid genes examined hit putative homologs in the soybean aphid sequences.

As we are interested in the physiological basis for the lack of aphicidal activity of toxins derived from the bacterium *Bacillus thuringiensis* (Bt), we paid particular attention to genes encoding putative Bt toxin receptors. Aminopeptidase N is one of several proteins known to function in the insecticidal action of Bt toxins. Seven putative aminopeptidase N (APN) genes were identified in the soybean aphid. Contigs with the three most abundantly transcribed APN genes were confirmed by RT-PCR. This study highlights (1) the potential of the relatively inexpensive, short-read transcriptome sequencing for rapid identification of target genes and analysis of transcript abundance, and (2) the inadequacies of current bioinformatics software for transcriptome analyses.

25

Identification and developmental characterization of *Apanos* and *Apvasa* genes: insights into the roles of duplicated germline genes in the pea aphid *Acyrtosiphon pisum*

Lu, Hsiao-Ling; Lin, Gee-Way Lin; *Chang, Chun-che

Laboratory for Genetics and Development, Department of Entomology/Institute of Biotechnology, College of Bio-resources and Agriculture, National Taiwan University, No. 27, Lane 113, Roosevelt Road, Sec. 4, Taipei 106, TAIWAN

Germ cells are the only cell lineage that passes genetic information from generation to generation. In most animals, specification of germ cells occurs during embryogenesis. No matter how germ cells are specified, either by the maternal germ plasm during early development or by signal induction conferred by somatic cells from mid development onward, it has been demonstrated that a group of common genes--the "germline toolkit genes" (GTGs), are utilized for the development of germ cells in many species. Among all GTGs, *nanos* and *vasa* are recognized as the two most conserved germline markers across animal phyla. In the fruit fly *Drosophila melanogaster*, only a single copy of *nanos* and *vasa* that has been identified, but duplication of *nanos* or *vasa* has been observed in

some other insects. The pea aphid *Acyrtosiphon pisum*, however, is so far unique among the insects in having four copies of each gene—from apparently, dual duplication events. We hypothesize that the dual duplications of *nanos* and *vasa* might be required for germline development in the complex life cycle of aphids, which includes asexual and sexual phases. Here we expression patterns of *Apnanos1-4* and *Apvasa1-4* in the parthenogenetic and viviparous phase of the pea aphid life cycle, during which oogenesis and embryogenesis occur within a single ovarian tubule that allow us to trace gene expression throughout the entire developmental process in a single experimental preparation. Results of *in situ* hybridization show that two *nanos* genes (*Apnanos1* and *Apnanos2*) and one *vasa* gene (*Apvasa1*) are germline-specific whereas the other genes are not preferentially restricted to germ cells. This suggests that gene duplication has allowed new roles for some of the duplicated *nanos* and *vasa* genes in aphids. We will further explore the roles of *nanos* and *vasa* in sexual embryos in order to understand whether role shift of homologous genes occurs during the transition of life cycles.

26

Estimation of the Periods Some Different Transposable Elements Insert into and Move off from the Host Genome in the BmTNML Locus of *Bombyx mori* and *Bombyx mandarina*.

Mano, Shuhei; Kawanishi, Yuichi; Maekawa, Hideaki; *Nakajima, Yumiko

(SM) The Institute of Statistical Mathematics, 190-8562, Tachikawa, Tokyo, Japan; (YK, HM, YN) Tropical Biosphere Research Center, University of the Ryukyus, 903-0213, Okinawa, Japan

DNA-type transposons, mariner-like elements (MLEs) are thought to have moved horizontally. And some of them look like still be active. But it is ambiguous or not cleared when each of them invades into the specific site of the organism's genome. For taking advantage of transposons including MLEs as an evolutionary clock, reliable estimation for the time of each event is required. In this presentation, we demonstrate the inference of the insertion or excision time and the genealogy of some transposons in the complex of BmTNML locus of *Bombyx mori* and wild silkworm *Bombyx mandarina* by using the sequences of transposons itself and around the BmTNML locus (host genome) which assume polymorphism.

It has been cleared that the Inverted Terminal Repeat (ITR) of MLEs is useful primer for isolating the full length of MLEs from many species by PCR amplification. We have utilized the ITR of *Hyalophora cecropia* MLE as primers for Lepidopteran insects. We named the MLE that had been isolated by this method as “*Cecropia-Inverted-Mariner (CIM)*”. There are around twenty copies of CIMs in the total genome of *B. mori* and it exists by four copies on the Chr6. We named one locus of them as “BmTNML” locus in 1999 which made the tripartite structure composed of three different transposable elements, namely CIM, L1Bm, and BMC1 retrotransposable elements. Recently, it was understood that the unit that took this treble structure existed only one copy on the Chr6.

We isolated CIMs corresponding to the region of the BmTNML locus from *Bombyx mandarina* inhabiting in Japan, China, Korea, and Taiwan. Insertion patterns of the transposable element in BmTNML locus of *B. mandarina* are different from those of *B. mori* except for the point that all of them contain the CIM. And there are no tripartite structures, namely no BMC1 insertions are observed in *B. mandarina*. On the contrary, some of them were inserted by the maT type of transposable element at the 3' region of the CIM. We named it as BmamaT1 which is homologous to the *B. mori* MLE called Bmmar6 previously reported to be similar to a maT-family member, Bmmar1. We regard many of the BmamaT1s on the CIM in *B. mandarina* are active, because they are highly homologous each other, some of them had complete ORF for transposase, representative amino acid motif, and DD37D catalytic triad. Moreover we found out the footprint that the BmamaT1 might be cleaved from that site.

By using the sequences around the BmTNML locus, we inferred the genealogy of the *B. mori* and *B. mandarina*. It was estimated that speciation which should be older than the insertion event of L1Bm, is 6.2-8.4Mya. This estimate is in close accordance with a report based on the whole mitochondria genome comparison (7.3Mya, Yukuhiro, et al. 2002). The insertion event of CIM was estimated to be at least 15Mya, and the insertion events of BmamaT1 and BMC1 were estimated to be 4.7-5.9 Mya and 0.4-6.2 Mya, respectively. For all sequence analyses, we did not find any significant departure from the neutrality (Tajima's test, 5% significance level). It is remarkable that there were no significant difference between estimates by the host genome and the CIM. This finding shows that estimations of Most Recent Common Ancestor (MRCA) of CIM give good estimation of MRCA of the host genome. The maximum likelihood estimate is 0.42Mya, with the 95% confidence intervals that are 0.20-0.84 and 0.12-2.19 for the models whose variances are equal to and ten-times of the mean, respectively.

27

Selenoprotein Extinctions in Insects

*Mariotti, Marco; Chapple, Charles E.; Mateo, Andrea; Corominas, Montserrat; Guigó, Roderic

(MM, RG) Center for Genomic Regulation, Universitat Pompeu Fabra, Barcelona, Catalonia, Spain, c/ Dr. Aiguader 88; (CEC)

Center for Genomic Regulation, Universitat Pompeu Fabra and Institut Municipal d'Investigació Mèdica, Barcelona, Catalonia, Spain,

c/ Dr. Aiguader 88; (AM, MC) Universitat de Barcelona, Department of Genetics, Barcelona, Catalonia, Spain, a/ Diagonal 645

Until recently, it was believed that all known metazoan genomes encoded selenoproteins. In the framework of the 12 drosophilas genome project, *Drosophila willistoni* was the first selenoprotein-less animal identified. After that, other selenoprotein-less species were found among insects (*Bombix mori*, *Tribolium castaneum*, *Apis mellifera*, *Nasonia* species, *Acyrtosiphon pisum*, *Heliothis*

virescens, ...). In these species, selenoprotein genes have been lost or converted to cysteine homologues, and most of machinery proteins (the factors necessary for the production and insertion of selenocysteine) have been lost. Given their phylogenetic tree, this implies that several, independent extinction events occurred in the insects lineage. This suggests a insect-specific relaxation of the selective constraints acting across metazoan to maintain selenoproteins. Nonetheless, nothing is known yet about the mechanisms of the extinction, since we observe the absence of both selenoproteins and of most machinery proteins. What is disappearing first?

Through selenoprotein gene-targeted PCR sequencing, we recently identified a group of *Drosophila* species in which the selenoprotein extinction seems to be occurring, but not yet totally accomplished. We are currently sequencing the full genome of some species from this group. Our purpose is to perform a detailed comparison of the genomic content among species in the various states of selenoprotein extinction. This could shed light on the mechanisms driving this event, and also identify new potential factors involved in selenocysteine pathways.

57

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

Merzendorfer, Hans; *Meera Kumari, Kim, Hee S.; Brown, Susan J., Specht, Charlie A.; Beeman, Richard W.; Kramer, Karl J.; Muthukrishnan, Subbaratnam

(HM) University of Osnabrück, Department of Biology/Chemistry, 49069 Osnabrück, Germany; (HM, MK, KJK, SM) Department of Biochemistry, Kansas State University, Manhattan, KS 66506; (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 benzoylurea-derived insecticides such as diflubenzuron (DFB, Dimilin®) are in wide use to control various insect pests. Although this class of compounds is known to disrupt molting and to affect chitin content, their precise mode of action is still not understood. To gain more insight into the mechanism underlying the insect growth regulating effects of benzoylurea compounds, we are conducting a comprehensive study utilizing genomics, proteomics and immunolocalization in the model insect species and stored product pest, *Tribolium castaneum* (red flour beetle). In this study, DFB was added to a wheat flour-based diet and fed to larvae and adults at various concentrations. The effect on larval-larval and larval-pupal molting depended on the concentration and time of feeding on DFB-containing diet as well as the initial weight of the larvae. While younger larvae (starting weight ~1.5 mg) exhibited nearly 100% mortality upon exposure to 100 ppm DFB, some of the older larvae (starting weight ~2 mg) were able to survive through the pupal stage to become mature adults. However, 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 digestion with chitinolytic enzymes and quantitation of N-acetylglucosamine by the Morgan-Elson method. This result was consistent with confocal analysis of cuticle-forming tissues of pharate adults treated with a chitin-binding probe. In contrast, the chitosan content was elevated significantly after DFB treatment. To correlate the observed changes in chitin/chitosan content with changes in gene expression, we performed RT-PCR to examine transcript levels of several genes involved in chitin metabolism, as well as a gene encoding a putative receptor for DFB. Gene expression for Chs-A and Chs-B was only marginally up-regulated. However, transcript levels of chitin deacetylase 2 (Cda-2) and also for a molt-associated chitinase, Cht-5, were increased upon DFB treatment. Expression of the gene for a putative sulfonyleurea receptor a potential target of DFB, was unaffected.

Immunolocalization of chitin deacetylase in cryosections of larvae indicated that while CDA-1 is secreted and deposited in the cuticle in control animals, the cuticle of DFB-treated animals is only marginally stained. However, the epidermal cells accumulated intracellular granules, which were labeled with the CDA-1 antibodies. Immunolocalization of other proteins associated with chitin metabolism are in progress.

In summary, the red flour beetle is a good model organism for investigating the mode of action of insect growth regulators such as DFB. The study revealed that DFB causes an increase in chitosan and reduction in chitin and transcripts for specific genes involved in chitin metabolism, cuticle formation and insecticide metabolism. We also observed defects in the mobility (leg movement) of adults that eclosed following DFB-treatment of larvae. Hence, results from our study support the hypothesis that DFB affects the expression of genes involved in cuticle chitin metabolism, which are critical for normal insect growth and development.

71

Sampling the transcriptome of the bed bug, *Cimex lectularius*

Mittapalli, Omprakash*; Bai, Xiaodong; Jones, Susan C

(OM, XB) The Ohio State University/OARDC, Department of Entomology, Wooster, OH 44691; (SCJ) The Ohio State University, Department of Entomology, Columbus, OH 43210

Bed bugs (*Cimex lectularius*) are poised to become one of the major pests in households throughout the U.S. Despite their high impact status, very little is known about this blood-feeding insect at the molecular level. A functional genomics approach will not only allow us to decipher knowledge at the gene level but also will provide a better understanding of the physiology-driven molecular processes

in bed bugs. We have applied the newer-generation Roche-454 sequencing strategy and obtained 35,646 bed bug transcriptome sequences with the total length of 12,882,069 bp. Based on sequence similarity with an E value cutoff of $1e-5$, 10,944 bed bug transcriptome sequences matched to protein sequences in GenBank nr database, with the majority to sequences from insects, including various species of aphids, *Drosophila*, mosquitoes, beetles, Lepidoptera, and others. The remaining 24,702 sequences are potentially bed bug-specific sequences. Interestingly, we have identified 59 *Wolbachia* sequences suggesting the presence of *Wolbachia* endosymbiont, which is consistent with the observations of female-skewed natural populations. *Wolbachia* endosymbiont may serve as a potential target for bed bug control. Finally, the EST database revealed 295 single nucleotide polymorphisms (SNPs) in 125 bed bug transcriptome sequences and 370 microsatellite loci in 345 bed bug transcriptome sequences. These genetic resources can be used for bed bug population genetic studies that may identify bed bug distribution patterns and facilitate control efforts. Overall, the current study provides a solid foundation for both basic and applied studies on bed bugs.

58

Functional genomics of an invasive insect pest: insights at the tissue Level

Mittapalli, Omprakash*; Bai, Xiaodong; Herms, Daniel A.; Bonello Pierluigi

(OM, XB, DAH) The Ohio State University/OARDC, Department of Entomology, Wooster, OH 44691; (PB) The Ohio State University, Department of Plant Pathology, Columbus, OH 43210

Very little is known at the molecular level as to how invasive insect pests adapt to plant defenses and/or synthetic chemicals. However, with the advent of second-generation sequencing strategies, genome-wide information for ‘non-model’ insect systems is rapidly becoming a dream come true. We are studying the midgut and fat body of the emerald ash borer (*Agrilus planipennis*), an exotic invasive insect pest that has killed millions of ash trees (*Fraxinus* spp.) across North America. Both tissues represent major interfaces that deal with several vital physiological functions including digestion, detoxification and immune response. Our initial efforts in sampling the transcriptomes of the midgut and fat body of *A. planipennis* on a Roche-454 platform yielded 63,374 unigenes (contigs plus singletons).

We obtained 126,185 transcriptomic reads for the midgut and 240,848 reads for the fat body. Post assembly of the sequences, 25,713 high quality ESTs (2,218 contigs and 23,495 singletons) were obtained for the midgut and 37,661 ESTs (5,376 contigs and 32,285 singletons) for the fat body. Among the midgut ESTs recovered, 9,066 (36%) matched to proteins in the GenBank nr database while 14,253 (38%) of the fat body ESTs showed similarity. A vast majority (95%) of the top matches were to arthropod proteins. The rest of the top matches of *A. planipennis* ESTs were to proteins of non-insect eukaryotes, archaea (midgut only), bacteria, viruses, and artificial sequences. Gene Ontology (GO) terms were assigned to 7,304 midgut ESTs and 10,074 fat body ESTs. The GO assignment for the midgut ESTs included 2,527 Biological Process terms, 585 Cellular Component terms and 1,169 Molecular Function terms; similarly the GO assignment for the fat body ESTs included 2,827 Biological Process terms, 646 Cellular Component terms and 1,454 Molecular Function terms. We identified 122 KEGG metabolic pathways for the fat body database and 115 for the midgut database. Some of the fat body-specific metabolic pathways were steroid biosynthesis, ubiquinone and other terpenoid-quinone biosynthesis, tetracycline biosynthesis, D-Arginine and D-ornithine metabolism, O-Mannosyl glycan biosynthesis, insect hormone biosynthesis etc. In contrast only the beta-Lactam resistance pathway was specific to the midgut. A pfam domain search yielded 5,876 domains in 4,925 sequences of the midgut. Among these midgut domains, 248 (4.2%) were predicted as chitin binding Peritrophin-A domains. On the other hand, the pfam domain search yielded 8,852 domains in 7,205 sequences of the fat body; of these only 16 (0.2%) chitin binding Peritrophin-A domains were predicted.

We identified 137 putative SNPs in 45 ESTs of the midgut and 347 SNPs in 143 ESTs of the fat body. A total of 888 microsatellite loci were predicted for *A. planipennis*. For (317 in the midgut ESTs and 571 in the fat body ESTs). The vast majority of the microsatellite loci (91% for fat body and 89% for midgut) were di- or tri-nucleotide repeats. Finally, a comparative analysis of the derived *A. planipennis* midgut and fat body ESTs with protein sequences in the draft genomes of *Anopheles gambiae*, *Drosophila melanogaster*, and *Tribolium castaneum* revealed most sequence similarity (36.7%, 23,236 out of 63,374) with *T. castaneum*. There were 16,610 sequences that were shared among the four insect species compared. About 61.9% (39,228 out of 63,374) of the *A. planipennis* transcriptomic sequences were specific to *A. planipennis*. This study represents one of its first kinds to elucidate tissue-specific genomic analysis using second-generation sequencing technology. Results will allow for a better understand of the physiology-driven molecular mechanisms in an invasive insect pest and provide a resource to identify alternate control targets.

9

The Hymenoptera Genome Database

*Munoz-Torres, Monica C.; Reese, Justin T.; Childers, Christopher P.; Bennett, Anna K.; Sundaram, Jaideep P.; Vile, Donald C.; Elsik, Christine G.

Georgetown University, Department of Biology, Washington, DC 22209.

With over 115,000 described species, the order Hymenoptera comprises approximately 10% of the species diversity on Earth. Species in this group directly affect human health and agriculture through diverse roles such as pollinators, bioindicators, parasitoids and agents of biological control, to name a few. The Hymenoptera Genome Database (HGD, <http://HymenopteraGenome.org>) is an informatics resource supporting genomics research for the group. Users access genome sequences and annotation tools through

viewers that include tracks for predicted genes, manually annotated genes, linkage markers, transposable elements, SNPs, and alignments of protein homologs and ESTs. Currently, HGD contains up-to-date versions of the genome annotations for honey bee, *Apis mellifera* (<http://BeeBase.org>), jewel wasp, *Nasonia vitripennis* (<http://NasoniaBase.org>) and three bee pathogens, *Ascosphaera apis*, *Nosema ceranae*, and *Paenibacillus larvae*. HGD supports community gene annotation of the bee and wasp genomes using a remote connection to its Chado database by Apollo Genome Annotation client software. The ants, another group of Hymenoptera, account for enormous biomass in tropical and subtropical systems and play important roles as decomposers, herbivores and predators. Extending HGD to serve as data repository for all completed genome annotations of species of Hymenoptera, we are currently adding the genomes of three species of ants, *Atta cephalotes*, *Linepithema humile* and *Pogonomyrmex barbatus* (each from separate genome consortia) as well as three species of bees, *Apis florea*, *A. scutellata* and *Bombus terrestris* (Baylor College of Medicine Human Genome Sequencing Center). Combining these genomes into a single resource will enhance the value of genomic data for each species by facilitating cross-species comparisons. We plan to implement a computational pipeline based on ortholog comparisons to assign Gene Ontology categories to each generated sequence.

49

Tools and Resources for Whitefly Biology defined by Next Generation Sequencing

Patel, Mitul; Collins, Carl; Colvin, John; Seal, Susan; *Bailey, David

(MP, DB) Greenwich University Enterprises Ltd, University of Greenwich, United Kingdom; (CC, JC, SS) Natural Resources Institute, University of Greenwich, United Kingdom

New tools and technologies for agrochemical discovery are urgently needed as the world faces major challenges to food security and sustainability. At the University of Greenwich in London, we are using the whitefly, *Bemisia tabaci* (Hemiptera:Aleyrodidae) as a target for developing such resources.

The whitefly is of global economic importance as a direct pest and virus-disease vector of numerous agricultural and horticultural crops. It is also of major scientific interest because resistance to a wide range of insecticides has already been described for this species. Unfortunately, no genome sequence for any *Bemisia* biotype is currently available. However, comparative genomic studies with related insects such as the pea aphid (*Acyrtosiphon pisum*; Hemiptera:Aphididae) can inform functional annotation of specific genes, while transcriptomic studies using high-throughput next generation sequencing (htNGS) can enable a detailed understanding of the dynamic basis of whitefly biology. Together, these approaches are already defining new molecular targets for agrochemical and agribiotech intervention, and enabling the development of new diagnostics for monitoring population divergence and spread.

Our initial research has focussed on transcriptomic studies of a *B. tabaci* population from India (Asia I mtCOI clade), selected for homozygosity over multiple generations, and confirmed as genetically homogeneous using partial mtCOI gene sequencing. High quality RNA preparations have been produced from several tissues: in this poster we focus on the transcriptome from adult females.

cDNA libraries, produced using the Evrogen Mint cDNA synthesis kit, were subjected to Roche 454 Titanium DNA sequencing. An initial 301,000 sequences were assembled into 3,822 core contigs using the Genomics Workbench from CLCbio. These 'Asia 1' contigs have been compared in detail to a previously published dataset of 1017 contigs from the *B. tabaci* B biotype. Primary sequence information and contig assemblies from these and other whitefly biotypes have now been integrated within a unified *Bemisia* transcriptome database, and annotated against other genomes using the software programmes BLAST, BLAST2GO, GOanna, and Exonerate.

htNGS gives excellent coverage of the most abundantly expressed whitefly mitochondrial and nuclear genes (including well characterised genes such as vitellogenin, NADH dehydrogenase, ATPase, succinate dehydrogenase, as well as many lower abundance genes). From this information, we have already identified informative sequence polymorphisms which may serve as a basis for identification of field isolates using technologies such as real-time PCR (RT-PCR).

The development of tools for monitoring biotype complexity, especially for newly emerging biotypes, is important to underpin future pest control strategies. To facilitate this, we are also designing methodologies for the rapid analysis of specific transcripts. Initial studies have focussed on the genes involved in vitellogenin (Vg) synthesis, and an RT-PCR assay has been developed to study their expression.

We recognize the power of genome sequence comparisons in both diagnostic development and resistance monitoring. We have a strong interest in sequencing the complete genome of *B. tabaci* and are actively seeking partnerships to achieve this.

Characterization of Multicopper Oxidase Related Protein (MCORP) in Two Insect Species

Peng, Zeyu*; Gorman, Maureen J.; Arakane, Yasuyuki; Kanost, Michael R.
Kansas State University, Department of Biochemistry, Manhattan, KS 66506

Multicopper oxidases (MCOs) have diverse functions but are evolutionarily and structurally related. Ten histidines and one cysteine form two highly conserved copper binding centers, which are required for oxidase activity. We discovered a gene that we named multicopper oxidase related protein (MCORP), which looks somewhat like a MCO, but lacks the conserved cysteine and many of the histidines. We identified MCORP orthologs in many insect species, including *Tribolium castaneum*, *Anopheles gambiae*, *Aedes aegypti*, *Culex pipiens*, *Bombyx mori*, and *Acyrtosiphon pisum*. Phylogenetic analysis of MCORPs and MCOs indicated that MCORPs belong to a specific cluster. We cloned full-length cDNAs of *A. gambiae* and *T. castaneum* MCORP and studied their expression patterns by RT-PCR. Both genes are highly expressed in late developmental stages, particularly in pupae. In adults, AgMCORP is expressed predominantly in reproductive tissues. We purified recombinant AgMCORP and TcMCORP to make antibodies that we will use to determine the location of MCORP within an insect. RNAi mediated knock down of MCORP in *T. castaneum* resulted in a pupal lethal phenotype, which demonstrated that TcMCORP is essential to pupal development. To try to learn the function of MCORP, we will expand our RNAi studies and determine whether recombinant MCORPs have enzymatic activity.

This research was supported by a grant from NIH (R01 AI070864).

Annotation and transcriptional analysis of pea aphid amino acid transporters

*Price, Daniel R.G.; Duncan, Rebecca P.; Wilson, Alex C.C.
University of Miami, Department of Biology, 1301 Memorial Dr, Coral Gables, FL 33146

Aphids feed on plant phloem sap, a nutritionally unbalanced diet that is low in essential amino acids. The ability of aphids to thrive on a phloem sap diet critically depends on their obligate symbiosis with the intracellular bacterium, *Buchnera aphidicola*. Central to the aphid/*Buchnera* symbiosis is the exchange of amino acids between symbiotic partners with *Buchnera* receiving non-essential amino acids from host aphids, and then synthesizing essential amino acids that are returned to the host. Populations of *Buchnera* are housed in specialized aphid cells called bacteriocytes. Transport of non-essential amino acids from the aphid host to *Buchnera* and export of essential amino acids from *Buchnera* to the host is reliant on transport across two aphid membrane systems: (i) the bacteriocyte cell membrane and (ii) the symbiosomal membrane, which surrounds each *Buchnera* cell. Currently, the molecular mechanisms of amino acid transport in the symbiosis are largely unknown. To date we have identified in the pea aphid genome 69 predicted genes encoding proteins belonging to the amino acid/polyamine/organocation (APC) superfamily. Nineteen of these predicted genes belong to family aa_trans (Pfam, Aa_trans, PF01490) and 25 to the aa_permease family (Pfam, AA_permease, PF00324). Phylogenetic reconstruction of these putative aphid amino acid transporters shows several aphid specific gene expansion events. Intriguingly, many of the transporters from these expanded families are expressed in bacteriocyte tissues (as judged by the presence of ESTs from bacteriocyte libraries) and therefore are implicated in amino acid transport across the bacteriocyte membrane and/or the symbiosomal membrane. Quantitative PCR is being used to validate the expression profiles of these transporter genes and will identify a subset of genes that play an important role in amino acid transport at the interface of the aphid/*Buchnera* symbiosis.

Activation of Antimicrobial Peptide Genes by Bacterial Components in the Tobacco Hornworm *Manduca sexta*

Rao, Xiang-Jun; *Yu, Xiao-Qiang
University of Missouri – Kansas City, School of Biological Sciences, Division of Cell Biology and Biophysics, Kansas City, MO 64110

Insects can produce a variety of antimicrobial peptides (AMPs). Some AMPs, such as Attacin and Cecropin, are commonly found among different insects, other AMPs, like Moricin, Lebocin and Gloverin, have been identified only in lepidopteran insects. Activation of AMPs is an important defense mechanism in insects. AMP expression is regulated by signal transduction pathways, such as the Toll and immune deficiency (IMD) pathways in *Drosophila melanogaster*. The Toll pathway is activated mainly by Lys (Lysine)-type peptidoglycan (PG), a major component of Gram-positive bacteria, and fungi, while the IMD pathway is triggered by the DAP (Diaminopimelic acid)-type PG of Gram-negative bacteria and Gram-positive Bacilli species. But lipopolysaccharide (LPS), a major component of Gram-negative bacteria, does not activate the Toll or IMD pathway in adult flies. However, highly purified LPS can activate immune responses in the fat body of *Bombyx mori*, a lepidopteran insect, although the activation level by highly purified LPS is lower than that by crude LPS and PG. In addition, very little is known about lipoteichoic acid (LTA), another major component of Gram-positive bacteria, in activation of AMPs in insects. Thus, it is necessary to confirm activation of AMPs by LPS in another lepidopteran insect, and to investigate whether LTA can activate AMP expression.

In this study, we investigated activation of AMPs by three major bacterial components, LPS (including both smooth and rough mutants of LPS), LTA, and PG from the same bacterial strains, in the larvae of *Manduca sexta*, a lepidopteran insect. We found that transcription levels of AMPs, such as Attacin, Lebocin and Moricin, in the hemocytes and fat body of *M. sexta* larvae were

significantly induced by smooth LPS and rough mutants of LPS (Ra and Re forms), synthetic lipid A, LTA, and PG to various levels. LPS-K12 from *Escherichia coli* K12 and LTA-BS from *Bacillus subtilis* activated AMP expression to significantly higher levels than PG-K12 and PG-BS, respectively, from the same bacterial strains, but LTA-SA from *Staphylococcus aureus* activated AMP expression to a significantly lower level than PG-SA did. In addition, smooth LPS were more potent than lipid A and rough mutants of LPS in activation of AMP expression. Our results demonstrated for the first time that LTA-BS can activate AMP expression, and different moieties of LPS may synergistically activate AMP expression in *M. sexta*.

10

A general bioinformatics framework for exploring diversity and divergence between genomic samples

Regier, Allison*; Thrasher, Andrew; Besansky, Nora; Emrich, Scott

(AR, AT, SE) University of Notre Dame, Computer Science and Engineering, Notre Dame, IN 46556; (NB) University of Notre Dame, Biological Sciences, Notre Dame, IN 46556

Genome sequencing is becoming an important tool for arthropod genetics. It can be used to find and genotype genetic markers, to examine diversity and divergence across populations, and to compare closely related species, all at very high resolution. As the cost of genome sequencing falls, it is becoming available to research communities beyond well-studied arthropods such as *Drosophila* and *Anopheles*, both dipterans. These non-model arthropods are important for advancing basic science as well as applications to global health and agriculture. However, tools for analyzing this flood of genomic data have lagged behind the demand for many scientific communities.

For example, resequencing is a common paradigm for using genome sequence data. In resequencing, sequence reads from one or more samples are aligned to a closely related reference genome. Single nucleotide polymorphisms, small insertions and deletions, as well as larger structural variation can be detected by comparing the reads that align to the reference genome. The exact requirements of downstream analysis depend on the type of sample. For example, if two inbred genomes are being compared, each position is expected to be homozygous within a sample. On the other hand, if pooled samples are being compared, polymorphism may be expected. Additionally, different types of analysis may have different requirements. One type of analysis may look for fixed differences, or positions with no within-sample polymorphism but different base calls between samples, whereas another analysis may count positions with shared polymorphism. The sensitivity and specificity of the analysis can be tuned by requiring different thresholds (e.g. coverage and/or minimum quality score).

Here, we describe a flexible and scalable bioinformatics framework based on the Genome Analysis Toolkit (http://www.broadinstitute.org/gsa/wiki/index.php/The_Genome_Analysis_Toolkit) to allow investigators to explore diversity and divergence within any resequencing data. This tool has a flexible interface that allows users to define multiple types of positions of interest by specifying rules such as coverage and quality thresholds for each sample, zygosity requirements, and sample comparison methods. Each position along the reference sequence is classified based on whether or not it satisfied a given set of rules. Positions of interest are counted and output for further analysis. Any sequence alignment tool that encodes underlying read-reference alignments in SAM format can be used in our framework. Further, we show that distributed and parallel computing can be used in this general framework both in computing read alignments and processing the results as desired.

This tool has been used to examine diversity and divergence across two recently sequenced genomes of the malaria mosquito, *Anopheles gambiae* and is being used to analyze diversity in multiple arthropod transcriptome studies including ecologically focused sequencing of butterflies. Significantly, this framework allows us to explore how incorporating sequence reads from the reference strain (when available) into the analysis could improve the assessment of diversity and divergence between the two sample genomes.

11

Sequence Generation and Assembly Progress on Current Arthropod Genomes at the Baylor College of Medicine Human Genome Sequencing Center

Richards, Stephen*; Scherer, Steven; Jhangiani Shalini N; Qu, Carson; Jiang, Huaiyang; Worley, Kim; Gibbs, Richard A.

Baylor College of Medicine Human genome Sequencing Center, Houston, TX 77030

The Human genome Sequencing Center at Baylor College of Medicine is currently sequencing the genomes of multiple arthropods in collaboration with many researchers from all over the world. These include the Hessian Fly (*Mayetiola destructor*), in collaboration with Jeff Stuart, Sue Brown, and Ming Chen; the Cotton Boll worm (*Helicoverpa amigera*) with Karl Gordon, John Oakenshot and Phil Batterham; the *Drosophila* Genetic Reference Panel (~ 200 *Drosophila melanogaster* inbred lines) in collaboration with Trudy Mackay, Eric Stone, and Julien Aryroles; the Sand fly (*Lutzomyia longipalpis*), in collaboration with Rod Dillon, and Many Ann McDowell; the Centipede (*Stigamia maritima*), in collaboration with Michael Akams laboratory; the Bumble Bee (*Bombus terrestris*) and the Dwarf Honey bee (*Apis florea*) in collaboration with the bee genome consortium – a list of many including but not limited to Gene Robinson, Hugh Robertson and Chris Elisk; the Med Fly (*Ceratitis capitata*), in collaboration with Alfred Handler and Ludvik Gomulski; the Postman butterfly (*Heliconius melpomene*) in collaboration with the *Heliconius* Genome Consortium organised by Chris Jiggins, and Owen McMillian; the Tobacco hornworm (*Manduca sexta*) in collaboration with Mike Kanost and Gary Blissard; a Predatory Mite (*Metaseiulus occidentalis*) in collaboration with Marjory Hoy;

and 7 additional comparative *Drosophila* species as part of the Mod Encode project in collaboration with Michael Eisen, Peter Cherbas, and Artyom Kopp

We here present our ongoing progress on these projects – some are nearing the release of the final assembly, some are in the midst of data collection, whereas others are just at the DNA collection stage. We also present our experience with Newbler assembly of these organisms and an idea on gap filling to improve assemblies.

12

The *Drosophila* Genetic Reference Panel: Whole Genome Association Mapping of Quantitative Traits, and a new tool for *Drosophila* genetics.

Richards, Stephen^{1*}; Zhu, Dianhui¹; Han, Yi¹; Ayroles, Julien²; Carbone, Mary Anna²; Stone, Eric²; Mackay, Trudy²; Gibbs, Richard A.¹

(SR, DZ, YH, RAG) Baylor College of Medicine Human genome Sequencing Center, Houston, TX; (JA, MAC, ES, TM) Department of Genetics, N.C. State University, Raleigh, NC

The *Drosophila* Genetic Reference Panel is a community resource of 192 sequenced inbred *Drosophila melanogaster* lines with measured quantitative traits. Forty lines have been sequenced to a minimum of 12X coverage using both 454 and Illumina sequencing platforms. The remaining 152 lines are being sequenced using the Illumina platform. Alignments have been generated using BWA, and polymorphisms identified using Atlas SNP2. We find approximately 500,000 SNPs, and 50,000 indels per inbred line, relative to the reference sequence. Extensive quality control genotyping has ensured both sequence and strain sample integrity, matching and true homozygosity for all of the lines. Lines with excess heterozygosity have been replaced to optimize the power of association studies.

Initial whole genome association analyses have identified between 20-100 polymorphisms associated with complex traits with p-values of 10⁻⁵ or less. Interestingly, 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 induced mutational screens. It also provides guidance for human complex trait studies. We present multiple methods for the validation of associations.

As a community resource we are working to facilitate the use of the reference panel in as many settings as possible. The fly stocks are currently available from the Bloomington *Drosophila* Stock Center. We are also preparing a website for genome wide associations allowing input of phenotypes measured on the lines, and outputting associated polymorphisms and 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 high school students counting bristles.

50

More hymenopteran genomes than you can shake a stick at

Robertson, Hugh M.

Department of Entomology, University of Illinois, Urbana-Champaign, IL 61801

Several genome projects on various Hymenoptera are underway, with many more planned. I will describe details for an ant genome and a bumble bee genome, both generated using next-generation DNA sequencing platforms.

1. The Baylor Human Genome Sequencing Center is undertaking Honey Bee Genome Project II (funded by the NHGRI), which is a multi-pronged effort to improve the existing honey bee v4 assembly and annotation. The genome has been re-sequenced with 454 and SOLiD to improve the assembly, along with manual superscaffolding. Multiple tissues and life stages have been subjected to EST sequencing using 454. The genome of the dwarf honey bee *Apis florea* and the European buff-tailed bumble bee *Bombus terrestris*, have been sequenced using 454 and ILLUMINA to provide comparative data for annotation. This effort is hoped to result in assembly v5 and a new gene model set by the end of this year.

2. Six ant genomes have been sequenced, all using next-generation sequencing, either 454 alone, ILLUMINA alone, or a mix of the two. I will provide details of the community-funded invasive argentine ant *Linepithema humile* genome assembly and analysis. Our analysis of the insect chemoreceptor superfamily indicates considerable expansion of the GR family (>100 genes), independently of the expansions seen in *Nasonia* (73 genes), and reveals additional GR lineage losses from *Apis* (10 genes and ~50 pseudogenes). The OR family is larger than *Apis* (174 genes), and perhaps larger than *Nasonia* (301 genes), with multiple large gene subfamily expansions.

3. Under the aegis of a NIH Pioneer Award to Gene Robinson, at Illinois we are sequencing multiple additional bee and other hymenopteran genomes, using all ILLUMINA GAI sequencing, and assembly with the BGI assembler SOAPdenovo. We have assembled the common Eastern USA bumble bee *Bombus impatiens* into several thousand scaffolds with N50 of ~1 Mbp and total size ~230 Mbp. Quality checks indicate that most genes are present and buildable. This genome used a whole flowcell and cost

\$25,000, but we are able to perform a comparably good assembly with half the data, so are currently attempting hymenopteran genome sequencing with half a flowcell, bringing the cost down to \$12,000 for the sequencing, and perhaps \$3,000 for assembly.

Hymenopteran genomes have several advantages for genome sequencing, primarily that they can be done mostly from a single haploid male, ensuring that a single genome sequence is present. They are also commonly reasonably small (<300 Mbp). Finally, the bee genomes commonly only have short DNA transposons like mariners, and few long retrotransposons, facilitating both sequencing (long PE reads are not essential), and assembly. Whether this approach will work for diploid and larger more complex genomes will be tested later this year.

Within a year approximately 100 arthropod genomes will be available. This community has the potential for generating 1000 genomes in short order, and in comparison to the vertebrate community might contemplate a 10,000, 100,000 or even 1 million arthropod genome project.

82

Single insect and tissue compartment specific response to RNAi microinjection in aphids: each individual matters

*Sapountzis, Panagiotis; Dupont, Gabrielle; Gaget, Karen; Horn, Thomas; Boutros, Michael; Febvay, Gérard; Jaubert-Possamai, Stéphanie; Charles, Hubert; Rahbé, Yvan; Colella, Stefano and Calevro, Federica (PS, GD, KG, FG, HC, YR, SC, FC) UMR203 Biologie Fonctionnelle Insectes et Interactions (BF2I), INRA, INSA-Lyon, Université de Lyon, F-69621 Villeurbanne France; (TH, MB) Boveri-Group Signaling and Functional Genomics, German Cancer Research Center Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany; (SJP) Institut de la Recherche Agronomique-INRA, UMR1099 BiO3P, Domaine de la Motte, F-35653 Le Rheu, France

Aphids (Hemiptera, Aphididae) belong to one of the three major orders of insect agricultural pests and cause billions of dollars in damage annually. So far, the treatment against aphids has been solely chemical insecticides. Novel pest control strategies might greatly benefit from the recent development of genomics tools to study aphid's biology. Post-transcriptional gene silencing using RNAi could be used not only to discover gene functions in the pea aphid model organism but also, in the future, as a method to target selected genes for aphid biocontrol. Although there have been studies on dsRNAi on aphids, several important questions, such as the individual reaction of individual aphid and tissues to RNAi, still remain open.

In the present study, we investigated the effects of RNAi by microinjection of cathepsin-L double-stranded RNA (dsRNA) in the pea aphid *Acyrtosiphon pisum*. The target sequence was selected based on the E-RNAi webtool for dsRNA design recently implemented for the pea aphid genome. Aphids were monitored for a five-day period (24, 48, and 120 hours after microinjection) and the efficiency of the RNAi knockdown was analysed by quantitative RT-PCR (qRT-PCR). Compared to previous studies, the novelty of our approach is that we analysed the response of single individual aphids in 5 different body compartments: bacteriocytes, gut, embryos, head and the remaining carcass.

Time-course analysis of the RNAi silencing revealed a huge variation of the gene expression knockdown level, ranging from 0 to 80%, depending on tissue, individual and time after injection. Interestingly, an altered phenotype (body shape changes,...) was observed in 13 out of the 75 aphids injected with cathepsin-L dsRNA, 24 and 48 hours after the injections. The qRT-PCR analysis showed that this phenotype was correlated with the reduction of cathepsin-L expression levels in those individuals. In conclusion, our study shows the need of individual analysis to better characterize the real phenotypic effect of the RNAi treatment.

RNAi webtool URL: <http://www.e-rnai.org/>

51

Embryonic transcriptome analysis of the Caribbean fruit fly, *Anastrepha suspensa*

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

USDA-ARS, Center for Medical, Agricultural and Veterinary Entomology, 1700 SW 23rd Drive, Gainesville, FL 32608, USA.

The embryonic transcriptome of the Caribbean fruit fly, *Anastrepha suspensa*, was sequenced by 454 pyrosequencing in an effort to isolate embryonic promoters and genes involved in programmed cell death. A cDNA library was constructed from total RNA pooled from various time points in embryogenesis using the SMART cDNA synthesis kit (BD Biosciences), normalized using the Kamchatka crab duplex-specific nuclease (Evrogen), titrated and sequenced on a 454-FLX platform. To specifically isolate genes promoting programmed cell death, a second transcriptome was constructed from total RNA isolated from embryos treated with gamma-irradiation known to promote premature cell death. In total, 58 million bases were obtained from 342,052 reads with an average length of 200 bp that formed 16,288 contigs ranging from 91 to 1,019 nt in length and 313,668 singlets. In vitro functional assays were performed on genes selected from the cell death-induced library. Putative *A. suspensa* cognates to the *Drosophila* hid and reaper pro-apoptotic genes were isolated and tested. These genes were also modified to include or delete binding and regulatory regions known to be important to *Drosophila* gene function, and were comparatively tested with their *Drosophila* homologs to verify function.

Sequence information from the ongoing Medfly whole genome sequencing project also will be used to identify and compare cell death genes between different tephritid species. Those shown to be most effective may then be tested for use as lethal-effectors in pest species to develop conditional lethality strains for use in new biologically-based control strategies.

95

Cloning and Characterizing Antiviral Defense Pathway Genes in *Frankliniella occidentalis*, the Vector Host of Tomato spotted wilt virus

*Schneweis, Derek; Whitfield, Anna; Rotenberg, Dorith

Kansas State University, Department of Plant Pathology, Manhattan, KS 66506

The western flower thrips, *Frankliniella occidentalis* (Order Thysanoptera), is an efficient vector of the plant-pathogenic virus Tomato spotted wilt virus (TSWV). TSWV is transmitted by *F. occidentalis* in a circulative and propagative manner, infecting and replicating in several types of insect tissues with no apparent pathogenic effect on the vector. Thrips molecules that recognize and/or respond to TSWV infection have not been characterized. Our primary research goal is to investigate and develop an understanding of the molecular basis of innate immune response of *F. occidentalis* to TSWV infection. We hypothesize that genes associated with small RNA-mediated gene silencing pathways, namely the eukaryotic antiviral defense strategy (siRNA pathway), are important modulators of the amount of TSWV harbored by the thrips vector. The specific objectives of our research are to i) generate full-length mRNA sequences of Dicer-2 (pattern recognition receptor of double-stranded viral RNA) and Argonaute-2 (effector of silencing of mRNA viral transcripts) and ii) quantify the expression of these siRNA pathway genes in *F. occidentalis* exposed to TSWV. Using an expressed sequence tag (EST) collection of first instar larvae (L1) of *F. occidentalis* (3,240 annotated sequences /12,700 ESTs) developed and annotated in our laboratory, we identified five sequences with putative homology to key components of the siRNA pathway in other well-characterized insects. Of these five EST sequences, we identified one putative Dicer-2 (Dcr-2) and one Argonaute-2 (Ago-2) sequence based on sequence analysis, each of which are approximately 1000 nucleotides in length from the 3' terminal end of the transcript. Dcr and Ago transcripts are generally > 4000 nucleotides in length. I isolated total RNA from TSWV-infected and non-infected L1s, pooled the RNA, cloned and sequenced Dcr-2 transcripts by performing sequential 5'RACE. To date, I generated a consensus sequence of 2,515 nucleotides containing two predicted RIBOc conserved domains and a portion of a PAZ super-family domain. I plan to continue 'walking' to the 5' terminal end of the Dcr-2 transcript, and perform 5'RACE for the Ago-2 transcript. The sequence data will be used to design specific real-time reverse transcriptase PCR (qRT-PCR) primers required to quantify or measure expression of these two thrips defense transcripts and TSWV RNA (titer) after exposure to the virus. Our in-depth study of the expression of two central genes in the antiviral defense pathway and their association with virus titer will enable us to gain insight into the interaction resulting from years of co-evolution of the virus and its vector.

30

Genome mapping and characterization of the *Anopheles gambiae* heterochromatin

*Sharakhov, Igor V.; Sharakhova, Maria V.; George, Phillip; Brusentsova, Irina V.; Leman, Scotland C.; Bailey, Jeffrey A.; Smith, Christopher D.

(IVS, MVS, PG) Department of Entomology, Virginia Tech, Blacksburg, VA 24061; (IVB) Department of Molecular and Cellular Biology, Institute of Chemical Biology and Fundamental Medicine Siberian Branch of Russian Academy of Sciences, Novosibirsk 630090, Russia; (SCL) Department of Statistics, Virginia Tech, Blacksburg, VA 24061; (JAB) Program in Bioinformatics and Integrative Biology and Department of Medicine, Division of Transfusion Medicine, University of Massachusetts Medical School, Worcester, MA 01605; (CDS) Department of Biology, San Francisco State University, San Francisco, CA 94132.

Heterochromatin plays an important role in chromosome function and gene regulation. Despite the availability of polytene chromosomes and genome sequence, the heterochromatin of the major malaria vector *Anopheles gambiae* has not been mapped and characterized. To determine the extent of heterochromatin within the *An. gambiae* genome, genes were physically mapped to the euchromatin-heterochromatin transition zone of polytene chromosomes. The study revealed that a minimum of 232 genes reside in 16.6 Mb of mapped heterochromatin. Gene ontology analysis revealed that heterochromatin is enriched in genes with DNA-binding and regulatory activities. Immunostaining of the *An. gambiae* chromosomes with antibodies against *Drosophila melanogaster* heterochromatin protein 1 (HP1) and the nuclear envelope protein lamin Dm0 identified the major invariable sites of the proteins' localization in all regions of pericentric heterochromatin, diffuse intercalary heterochromatin, and euchromatic region 9C, but not in the compact intercalary heterochromatin. To better understand the molecular differences among chromatin types, novel Bayesian statistical models were developed to analyze genome features. We provide evidence that the diffuse intercalary heterochromatin has a higher coverage of DNA transposable elements, minisatellites, and satellites than does the compact intercalary heterochromatin. Our results demonstrate that *Anopheles* polytene chromosomes and whole-genome shotgun assembly render the mapping and characterization of a significant part of heterochromatic scaffolds a possibility. These results reveal the strong association between characteristics of the genome features and morphological types of chromatin. Initial analysis of the *An. gambiae* heterochromatin provides a framework for its functional characterization and comparative genomic analyses with other organisms.

Draft Transcriptome Assembly of the Tobacco budworm, *Heliothis virescens* (L.)

*Shelby, Kent S.; Perera, Omaththage P.

(KSS) USDA-ARS-BCIRL, Columbia, MO 65201; (OPP) USDA-ARS-SIMRU, Stoneville, MS, 38776.

Massively parallel next generation sequencing technologies have revolutionized the acquisition and analysis of transcriptomes from beneficial and pest insects. Heliothine moths are major polyphagous pests of commodity crops such as maize, cotton, soybeans and vegetables throughout the world. Control of larvae of the North American pest moth *Heliothis virescens*, also known as the tobacco budworm, and other closely related heliothines would be fundamentally advanced by understanding--and disrupting--the basic mechanisms underlying resistance to insecticides, entomopathogens and parasitoids. An iterative de novo budworm exome assembly was constructed by combining Illumina and Sanger reads. For this first assembly of a budworm exome a total of 43,201,749 high quality input sequences assembled into 69,643 contigs with an average length of 339 bases (average coverage of 21). Cumulative contig length and the total sequence lengths were 21,984,026 and 698,317,284 nucleotides, respectively. 1831 contigs were greater than 2000 bases in length. This newly constructed exome will constitute the reference assembly for all future functional genomic studies of this pest species utilizing next-generation RNA-seq digital transcript profiling and other methods.

KEGG Bioinformatics Resource for Arthropod Genomics

*Shigenobu, S.; Kanehisa, M.; Goto S.

(SS) National Institute for Basic Biology, NIBB Core Research Facilities, Okazaki Aichi 444-8585, Japan; (KM, GS) Kyoto University, Institute for Chemical Research, Bioinformatics Center, Uji Kyoto 611-0011, Japan

KEGG (Kyoto Encyclopedia of Genes and Genomes) is a bioinformatics resource for understanding biological function from a genomic perspective. It is a multi-species, integrated resource consisting of genomic, chemical and network information with cross-references to numerous outside databases. With the recent release of the draft genome sequences of several insect species, there has been a growing demand to place these insect genomic information into a metabolic pathway and other biological contexts. We have launched the KEGG Insect Project so that insect researchers can take advantage of the rich resource in KEGG.

We first surveyed official gene sets released by various insect genome projects and found inconsistency with NCBI RefSeq in most cases. It is mainly because RefSeq takes a conservative policy to store only reliable gene models with experimental evidences, while many communities allow ab initio gene models in order to cover gene repertoire as many as possible. To deal with this double-standard situation, we are building database entries from both of RefSeq and the official gene sets released by the consortium for each genome project. We then stored them into the regular GENES database and DGENES (Draft GENES) database, respectively. For example, pea aphid, *Acyrtosiphon pisum*, has two gene sets in KEGG: "api" set is generated from RefSeq composed of 10,362 genes and "dapi" set is from AphidBase composed of 34,603 genes. We are also developing a tool to work with community members to incorporate the manual annotations into the KEGG database.

A new feature recently added is KEGG Organism Groups Search, where an organism group is defined as a combination of KEGG organisms, enabling the analysis of combined pathway maps for the group, for example, in symbiosis or pathogenesis. In this poster, we show the example of aphid and the bacterial symbiont *Buchnera aphidicola*, where the complementary relationship of amino acid metabolic pathway in the symbiosis is relevant.

The ultimate goal of KEGG Insect will be to bring new insect species into KEGG from genome sequences as well as EST data; and add new PATHWAY maps specific to insects such as insect hormone metabolism and chitin synthesis. KEGG is publicly available at <http://www.genome.jp/kegg/>. Any comments and requests are welcome.

Dopamine receptor D1 in the salivary secretion of the blacklegged tick *Ixodes scapularis*

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

Kansas State University, Entomology Department, Manhattan, KS 66506

Dopamine (DA) has been shown to be the major neurotransmitter that activates salivary secretion of ticks. The direct DA action on the salivary glands were thought to be through the D1-type (excitatory) receptor, a G protein-coupled receptor (GPCR). Genome sequences of the black legged ticks *Ixodes scapularis* yielded the gene encoding D1. The tissue specific PCR showed that D1 is expressed in salivary glands (SG) and also in synganglion of partially fed female tick. The quantitative reverse transcriptase PCR (qPCR) for the mRNAs isolated from the SG of different feeding-phases showed that the D1 receptor mRNA constitutively existed in the SG from pre-blood feeding to repletion of the tick after the full engorgement. Using immunohistochemistry we identified 17 pairs of neurons expressed D1 in the *Ixodes* synganglion. We also found the D1 immuno-reactivity in the luminal surface of salivary gland acini II and III of unfed and five days fed ticks. A reporter system using heterologous expression of the D1 showed strong responses of the receptor to DA in mobilization of intracellular calcium (EC50= 248 nM) and in elevation of cAMP (EC50= 138 pM). . Also

butaclamol a stereotypic antagonist of dopamine receptors was active on the D1 receptor at a high concentration in suppression of calcium mobilization pathway. This study suggested the important function of D1 receptor and presence of a complex DA responsive systems in the salivary control in tick.

28

Genomic Analysis of Post Zygotic Sex Determination in the Hessian Fly

*Stuart, Jeff; Zhao, Chaoyang; Benatti, Thiago; Chen, Ming-Shun; Richards, Stephen (JS, CZ, TB,) Purdue University, Department of Entomology, West Lafayette, IN 47907; (M-SC) USDA-ARS, Kansas State University, Department of Entomology, Manhattan, KS 66506; (SR) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030

Gall midges (Diptera: Cecidomyiidae) use post-zygotic chromosome elimination to establish sex-determining chromosome numbers. Embryos that retain a diploid number of chromosomes establish a female determining karyotype, typically A1A2X1X2/A1A2X1X2. Embryos that eliminate paternally derived chromosomes establish a male determining karyotype, typically A1A2X1X2/A1A2OO. In the majority of gall midges so far examined, the number of chromosomes eliminated is maternally controlled so that some females of any given species produce only female offspring whereas others produce only male offspring.

We have investigated this phenomenon in the Hessian fly (*Mayetiola destructor*), a gall midge and important pest of wheat (*Triticum* spp.), using cytogenetics, fluorescence in situ hybridization (FISH), molecular genetic mapping, gene expression analyses, and whole genome sequencing. An unusual neo-sex chromosome (called a W' chromosome) was discovered by this analysis and has been partially characterized. The W' consists of one or two small inversions on the long arm of Hessian fly chromosome A1. When the W' is present in a female, her offspring retain a diploid number of chromosomes and develop as females. When the W' is lacking in a female, her offspring eliminate the paternally derived X1 and X2 chromosomes and develop as males. The genes in the inverted region of the W' chromosome have been identified and the expression of some of these genes have been examined in male-producing females, female-producing females, male eggs, and female eggs. Although the analysis is not yet complete, it suggests that one gene disrupted by one inversion breakpoint may be the sex determination master switch that controls the elimination or retention of paternally derived chromosomes in Hessian fly embryos.

29

Temporal genetic variation, and bloodmeal identification of *Anopheles funestus* and *An. gambiae* mosquitoes from Mozambique.

*Szalanski, Allen.L.; Trout, Rebecca; Austin, James W.; Magnus, Roxane; Charlwood, J. Derek. (ALS, RM) University of Arkansas, Department of Entomology, Insect Genetics Lab, Fayetteville, AR 72701; (RT) University of California-Davis, School of Veterinary Medicine, Davis, CA 95616; (JWA) BASF Corporation, Research Triangle Park, NC 27709; (JDC) MOZDAN Project, Maputo, Mozambique.

The anopheline mosquitoes *Anopheles funestus* and *An. gambiae* are important vectors of malaria in Africa. We conducted a study on genetic diversity of the mtDNA ND5 gene of these two species collected from 2002 to 2009 in a remote village located in Mozambique. In addition, host bloodmeals were also identified using DNA sequencing of a mtDNA cytochrome b marker. From the 78 *An. funestus* subjected to DNA sequencing, a total of 18 haplotypes were observed, of which, seven were singletons. Haplotype MZ148CI was the most common overall, but it was most prevalent in 2005, and not appreciably represented in the 2002 and 2009 samples. Wright's FST statistic revealed high levels of genetic differentiation between the 2002 and 2005 and 2005 and 2009 populations, but not between the 2002 and 2009 populations. Of the 18 *An. gambiae* sequenced, a total of 7 haplotypes were observed, with four of them being singletons. There was no apparent structure in the temporal variation of the *An. gambiae* haplotypes, which can be attributed to the limited sample size. Molecular diagnostics of blood meals show that of the 93 samples evaluated, seven were confirmed to have fed on *Homo sapiens* or *Bos taurus*. This study demonstrates significant population structure differences in *An. funestus* from the same location over time, and that the samples available for evaluation in this study provide clues into the species and seasonal flux of important Malaria vectors for the inhabitants of this area.

103

Manual Annotation of miRNA, RNAi and piRNA Pathway Genes in Ant Genomes

*Tao, Shu; Placek, Jennifer E.; Munoz-Torres, Monica C.; Hagen, Darren E.; Smith, Christopher D.; Elisk, Christine G. (ST, MCMT, DEH, CGE) Georgetown University, Department of Biology, Washington, DC 20057; (JEP, CDS) San Francisco State University, Department of Biology, San Francisco, CA 94132

Here we report the manual annotation of the genes involved in microRNA (miRNA), RNA interference (RNAi) and piwi-interacting RNA (piRNA) pathways from three species of ants native to the Americas: *Pogonomyrmex barbatus*, *Linepithema humile*, and *Atta cephalotes*. *P. barbatus* is found in arid regions of the southwest US and northern Mexico, *L. humile* is native to Argentina but invasive globally, and *A. cephalotes* ranges from southern Mexico to Brazil. Central to miRNA and RNAi pathways are miRNAs and small-interfering RNAs (siRNAs), which are 20~25 nt long and lead to translational arrest and/or target mRNA cleavage depending on the complementarity between the mature miRNA strand or siRNA strand and the target mRNA sequence. piRNAs, on the other hand,

are typically 26–31 nt in length and involved in transposon silencing in germline cells, especially during spermatogenesis. Using the MAKER pipeline and Apollo annotation software to visualize and manage all the computational gene evidence provided by the three ant genome consortiums, we characterized a complete repertoire of genes potentially involved in the miRNA, RNAi and piRNA pathways in all three species. Annotations were performed by separately querying each ant genome with curated amino acid sequences from other insect genomes available (e.g., *Drosophila melanogaster* and *Apis mellifera*). We will present the gene repertoire for these pathways in the ant genomes and compare them with pathways in other sequenced insect genomes.

59

A Naturally-occurring Silencer of Medea Selfish Activity in *Tribolium castaneum*

Thomson, M. Scott

University of Wisconsin-Parkside, Department of Biological Sciences, Kenosha, WI 53141

Selfish genetic elements of the Medea family possess a maternal lethality phenotype, such that all non-Medea progeny fail to develop. This selfish activity should favor spread of Medea through *Tribolium castaneum* populations, counterbalanced by general negative effects associated with Medea. The naturally-occurring hybrid incompatibility factor, H, produces negative effects in combination with Medea. H and Medea are distinct loci mapping to different autosomes. In general, H male crossed to Medea female produces a poorly viable, reduced fecundity phenotype that is temperature-sensitive (more severe at lower temperature). The reciprocal cross distinguishes members of the Medea family: M1 male crossed to H female results in failure of all progeny to develop past egg eclosion; M4 male crossed to H female produces progeny with temperature-sensitive poor viability that is significantly less severe than the reciprocal cross, and these H-M4 heterozygotes are fertile. This presentation describes the nature of Medea lethality observed in fertile H-M4 heterozygotes. The conclusion is that H blocks expression of M4 maternal lethality. Thus, naturally-occurring H may counter the spread of Medea M4 through populations via two mechanisms: the negative effect of poor viability/fecundity; and the silencing effect that deprives Medea of its selfish lethality. High-resolution recombinational genetic mapping is underway with the goal of cloning H by position.

14

Using the “CycADS” annotation management system to develop a BioCyc metabolic network database for sequenced arthropods genomes: first steps towards “ArthropodsCyc”

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

(PBP, GF, FC, YR, HC, SC) UMR203 BF2I, 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; (AFV, GF, FC, YR, HC, MFS, SC) BAMBOO, INRIA Rhône-Alpes, France; (JHC, TG) Center for Genomic Regulation, Barcelona, Spain.

The recent sequencing of the genome of the pea aphid (*Acyrtosiphon pisum*), together with the already available sequence of its primary symbiont (*Buchnera aphidicola*) genome, prompted us - during the genome annotation phase - to develop “AcypiCyc”: a BioCyc database dedicated to the pea aphid. This metabolic reconstruction was driven by the development of CycADS (Cyc Annotation Database System): an automated annotation management system that allows the seamless integration of the latest sequence information into metabolic networks reconstruction.

CycADS is centred on an ad hoc SQL database, complemented by a set of Java programs to import and export relevant information. Data from GenBank and from different gene annotation tools (such as KAAS, PRIAM, Blast2GO, PhylomeDB, etc) are collected into the database. Using an evidence score based filter system, data are later extracted to generate a complete input file to build and/or update BioCyc databases using the ‘Pathway tools’ software. Furthermore, CycADS allows the integration in a BioCyc database of specific links to custom information resources (in AcypiCyc: AphidBASE, KEGG orthology and PhylomeDB) together with already available databases cross-references (e.g. GenBank, BRENDA).

BioCyc databases offer a framework for the integrated analysis of metabolic networks, together with the possibility to browse an organism metabolism. Thanks to its open platform formats (BioPAX, SBML), the BioCyc data can be integrated into other tools to perform both metabolic network and complex genomic data analyses. The CycADS pipeline, beyond allowing an easy update of the AcypiCyc database over time, can be applied to other newly sequenced genomes.

In fact, AcypiCyc can be considered as a prototype of future generation BioCyc databases for other arthropods and such databases will contribute to the build-up of resources for arthropods genomics research. Using the CycADS system we have already reconstructed the metabolism of *Tribolium castaneum* and *Drosophila melanogaster*, and these databases are now included into the AcypiCyc web interface. These first three insects metabolism reconstructions are the first steps towards the generation of “ArthropodsCyc”: a metabolic BioCyc database for arthropods whose genome is (or will be) sequenced. Such database, generated using our CycADS pipeline, will allow researchers to browse their model organism metabolism and to perform comparative analyses.

URLs: <http://pbil.univ-lyon1.fr/software/cycads/acypicyc/home>; <http://cycadsys.org>

31

Exploring the Evolution of Social behaviour using Transcriptomic Data

*Venkat, Aarti; Varala, Kranthi; Woodard, Hollis; Fischman, Brielle; Clark, Andrew; Cameron, Sydney; Robinson, Gene; Hudson, Matthew.

(AV, KV, MH) University of Illinois at Urbana-Champaign, Department of Crop Sciences, IL 61820; (HW, BF, SC, GR) University of Illinois at Urbana-Champaign, School of Integrative Biology, IL 61820; (AC) Cornell University, Department of Molecular Biology and Genetics, Ithaca, NY 14853.

Sociality is at the root of tremendous ecological success of several taxa, including humans, ants, bees, wasps and termites. The degree and type of sociality varies greatly across taxa. The evolution of complex social behaviour can be studied by performing comparative analyses across a phylogeny of taxa showing a wide diversity in social lifestyles. Bees serve as excellent model systems as varying degrees of sociality ranging from solitary to eusocial is observed in extant bees. Sociality has evolved at multiple independent times in the bees. Our aim is to identify adaptive changes in the protein coding regions of bees showing diverse social behaviour patterns. To this end, we used next generation 454 FLX sequencing technology to generate the brain expressed sequence tags (ESTs) of 9 socially diverse bees. The ESTs were assembled into species specific non redundant contigs and singletons. We generated multiple sequence alignments of putative orthologous regions using the *Apis mellifera* gene models as the reference. The evolutionary changes associated with these alignments was statistically inferred using maximum likelihood methods that use sophisticated codon substitution models to detect adaptive changes in the protein coding genes. The rapidly evolving genes were then annotated for their functions using Gene Ontology.

53

The first butterfly genome? Process and progress in the complete genomic sequencing of *Heliconius melpomene*, the long-wing passion-flower butterfly.

*Walters, James R.; Heliconius Genome Consortium

University of Cambridge, Department of Zoology, Cambridge, 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 was initiated less than 2 years ago and reflects a collaboration between the broader *Heliconius* genetics research community and the Baylor Human Genome Sequencing Center. The 295 Mbp genome of *H. melpomene* is being sequenced entirely with 'next generation' technologies, primarily via 454 but with some Solexa data also to be included.

As of April, 2010 we have generated 6.7 Gbp of 454 sequences which corresponds to ~20x coverage. A Newbler assembly of the data yielded ~14,000 scaffolds with N50 = 62 Kbp and ~105 contigs with N50 = 8.1 Kbp. Ongoing 454 sequencing of an 8 Kbp mate-pair library and additional Solexa runs are expected to substantially improve the assembly in the near future. Scaffolds will be anchored and ordered along the chromosomes via RAD-tag genotyping of a F2 mapping family generated from the sequenced *H. melpomene* strain. RNA-seq data from several tissues will complement existing Sanger and 454 ESTs from wing tissues to guide genome annotation efforts, which will be conducted in collaboration with EBI's Ensembl Genomes initiative.

32

Evolutionary Genomics of *Colias* Phosphoglucose Isomerase (PGI) Introns

*Wang, Baiqing; DePasse, J. Mason; Watt, Ward B.

Dept. of Biology, Stanford University, Stanford, CA 94305-5020

*Corresponding/presenting author: bqwang@stanford.edu

Protein-coding genes of eukaryotes often include introns – non-coding intervening sequences which are transcribed but later removed from precursor mRNA. Introns usually comprise a large portion of eukaryotic genomes. They are becoming increasingly important in studies of genome evolution as more regulatory functions, and roles in gene evolution, have been found for them. The availability of several fully sequenced eukaryotic genomes allows study of general patterns of intron properties and the exon-intron structure of eukaryotic genes. Here, we obtain and analyze intron sequences from the glycolytic enzyme phosphoglucose isomerase, PGI (E.C. 5.3.1.9), of *Colias* butterflies in order to gain understanding of intron diversity and evolution within a single gene whose coding region displays persistent, naturally selected polymorphism. The effects of *Colias* PGI variants on protein function, flight biophysics, and organismal fitness have been well characterized, and recent work has begun to offer mechanistic molecular evolutionary explanations of these effects. Here we study the 11 introns of 30 PGI coding-region alleles in *Colias*, finding that all of these introns are highly variable, diversified by a large number of point mutations and small and large indels. The extent of genetic variation of these introns is

affected by intron position, intron length, and adjacent exon sequences. Some of them exhibit substantial within-intron homology of sequence, while others show almost none except for their first few and last few base pairs. Intragenic recombination and hitchhiking also appear to have shaped the divergence and variation of PGI introns. Further, comparison of the *Colias* PGI sequences to those of the silkworm (*Bombyx mori*) genome shows very little sequence homology between the PGI introns of these two Lepidopteran taxa, in sharp contrast to the complete conservatism of intron positions within their coding regions. We consider the evolutionary implications of these results, and so do note the important utility of well-established genomes of particular species for support of molecular-evolutionary study in related species which as yet do not have full genomic information available.

39

Metagenomic Profiling of Gut Bacterial Community in *Anopheles gambiae*

Wang, Ying; Gilbreath III, Thomas M.; Yan, Guiyun; *Xu, Jiannong

(YW, JX) New Mexico State University, Biology Department, Las Cruces NM 88003; (TMG, GY) University of California, Irvine, College of Health Sciences, Irvine CA 92697

Mosquito *Anopheles gambiae* is a major malaria vector. Vector competence is determined by the tripartite eco-interactions among microbes, malaria and mosquito immunity. Mosquito gut harbors diverse microbial communities. However, little is known about the dynamics of the gut microbiota from larva to adult and its impact on the gut eco-symbiotic interplays. Early studies of the gut bacteria relied on culture- and/or cloning-based low throughput techniques that are able to characterize only a small fraction of the microbiota. Using pyrosequencing approach targeting the V1-3 hypervariable region of the 16S rRNA gene, we gained an unprecedented view of the diversity present in the gut microbiota. Here we showed the dynamics of the gut microbial community from larva to adult and the effect of blood feeding on the gut community.

Pyrosequencing yielded 73315 sequences from 13 samples of a lab reared *An. gambiae*. The sequences were assigned to 260 genera belonging to 11 phyla with dominances of Proteobacteria, Bacteroidetes, Actinobacteria and Firmicutes. The community structure changes along the life stages. Among the tags obtained from larva, more than half (54.3%) belong to the Family Enterobacteriaceae (unable to classify them in a genus). *Microbacterium* (23.7%) is abundant only in larva. Pupa harbors a complex community with the dominances of *Elizabethkingia* (26.1%), *Acidovorax* (10.5%) and Enterobacteriaceae (10.8%). In newly emerged adults, the community becomes less complex, dominated by *Elizabethkingia* (55.9%), *Staphylococcus* (8%), Enterobacteriaceae (7.5%) and *Leucobacter* (6.9%). At least 7 out of 20 (35%) genera that were present in pupa did not rise in the adult, likely due to the “sterilization” process during the metamorphosis. *Fingoldia*, *Serratia* and *Pantoea* seem to be newly established in the adult. Bloodmeal reduces the diversity of gut microbiota. After blood feeding, 6 out of 17 (35.3%) genera went undetectable. *Elizabethkingia* and Enterobacteriaceae dominate the community, accounting for up to 77% of the total tags. The diversity of gut community reduced with the drop of operational taxonomic unit (OTUs) from 2080 (before feeding) to 1630 (after feeding). Species richness was estimated by abundance-based coverage (ACE) and the Chao estimator of species diversity, which were implemented by the software Mothur. Reduced oviposition was observed in mosquitoes fed with broad spectrum antibiotics, suggesting that the gut bacteria are beneficial to the mosquito fecundity. Gut community fluctuation may also affect gut-microbe homeostasis, which was suggested by the expression dynamics of antimicrobial genes throughout the different life stages.

We also surveyed bacterial communities in semi-natural habitats and in larval guts reared in the habitats in Kenya. Diverse bacteria were identified in the habitats and some of them were found in the gut. And the gut community structures are different between the natural settings and the lab conditions.

33

Correlating Traits of Gene Retention, Sequence Divergence, Duplicability and Essentiality in Arthropods Compared to Vertebrates and Fungi.

*Waterhouse, Robert M.; Zdobnov, Evgeny M.; Kriventseva, Evgenia V.

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

The recent availability of a growing number of sequenced genomes allowed us to examine gene genealogies across 23 arthropod species and compare these to 40 vertebrates and 32 fungi. The OrthoDB hierarchical orthology delineation procedure employs all-against-all Smith-Waterman sequence comparisons to classify orthologous groups of genes descended from the last common ancestor at each species radiation across the phylogeny. For the arthropod, vertebrate, and fungal lineages, the procedure classified ~86% of a total of more than 1.36M protein-coding genes, over 90% of which were classified into groups linked to Gene Ontology (GO) or InterPro functional annotations.

Taking advantage of these data, we quantified properties of ortholog phyletic retention, copy-number variation, and sequence divergence to examine correlations with gene essentiality and functional traits in arthropods compared to vertebrates and fungi. We find that more than half of orthologous groups are universally distributed across each lineage. These universal orthologs are either

predominantly single-copy or predominantly multi-copy, and sequence evolution of single-copy orthologs is markedly more constrained. Arthropods do not exhibit the accumulation of multi-copy genes that characterize the vertebrates, suggesting a lower tolerance of slightly deleterious mutations in arthropods that prevents the accumulation of redundant gene copies that may occur under weaker selection pressures in vertebrates. Gene essentiality data from representative model organisms, *Drosophila melanogaster*, *Mus musculus*, and *Saccharomyces cerevisiae*, reveal that essential genes are significantly enriched in universal orthologs in each lineage, which consistently exhibit greater sequence conservation. Universal arthropod orthologous groups with essential genes are mostly single-copy while in vertebrates essential-gene-containing groups are mostly multi-copy. Orthologous groups with genes involved in immune and stress responses as well as adhesion and reproductive processes are relatively divergent in both arthropods and vertebrates, while divergence of DNA metabolism, signal transduction and cell communication are common to arthropods and fungi. Orthologous groups with numerous homologs are characterized by protein domains that have independently proliferated in each lineage, revealing the highly abundant Zinc finger proteins in both arthropods and vertebrates, and the proliferation of Rhodopsin-like G-protein-coupled receptors (GPCRs) in vertebrates that contrasts the expanded superfamily of arthropod peptidases.

All orthology data with corresponding protein sequences and associated functional annotations are available from the OrthoDB website at <http://cegg.unige.ch/orthodb>. OrthoDB provides intuitive querying facilities enabling logical searches with annotation keywords or GO, InterPro, and protein identifiers, as well as queries based on user-specified phyletic profiles of orthologous group member genes and BLAST similarity searches. In addition, the results have been integrated into FlyBase and UniProtKB to provide linkouts to OrthoDB from these popular resources. With 23 arthropod species including 12 drosophilids, 3 mosquitoes, silkworm, flour beetle, honeybee, *Nasonia* wasp, body louse, pea aphid, *Daphnia* waterflea, and the deer tick, OrthoDB represents a comprehensive resource for comparative arthropod genomics.

97

Evolutionary genomics of an enzyme, PEPCK, evolving in two cell compartments at once

Ward B. Watt*, Eddie Wang, and Baiqing Wang

Dept. of Biology, Stanford University, Stanford, CA 94305-5020

*Corresponding/presenting author: wbwatt@stanford.edu

Two important uses of genomics information from particular species are to support focused studies of specific genetic systems and to support study of related species for which genomic information is not (yet) available. We use the genome of the silkworm, *Bombyx mori* (Lepidoptera, Bombycidae), as such a "bridging" information source to support study of natural genetic variation of the enzyme phosphoenolpyruvate carboxykinase, PEPCK, in *Colias* butterflies (Lepidoptera, Pieridae), and to place it in context of comparative genome-level structure of this genetic system. Guanine-nucleotide-using PEPCK (E.C. 4.1.1.32) is known mainly in animals but also in a few prokaryotes. It converts phosphoenolpyruvate (PEP) plus guanosine diphosphate and a bicarbonate ion to (and from) oxaloacetic acid plus guanosine triphosphate. PEPCK offers an alternative route besides the pyruvate kinase reaction for carbon skeletons to move between glycolysis in the cytosol and the Krebs cycle in the mitochondrial matrix. Thus it is expressed in both these cellular compartments. In vertebrates and in *Drosophila*, two complete compartment-specific, paralogous genes occur; we show by phylogenetic analysis that all the vertebrate paralogs studied represent one gene-duplication event but the *Drosophila* paralogs arise from an independent duplication. In contrast, analysis of genomic structure and an expressed-sequence library in *Bombyx* reveals compartment-specific PEPCK enzymes that differ only in their first, 5', exons, which are alternatively spliced onto a single following series of exons 2-13 (comprising the vast majority of the protein sequence) to target the two compartments. This exon-splicing strategy and mechanism is repeated in PEPCK of *Colias*. In all three *Colias* taxa examined, the cDNA sequence of exons 2-13, in common between the splice forms, is extremely variable at both nonsynonymous and synonymous basepair sites, to a degree matched so far only by a strongly selected polymorphism previously studied in *Colias* at the PGI, phosphoglucose isomerase, gene. In all three taxa, a major (second allele frequency > 0.1) amino acid polymorphism, Gly ↔ Ser, is found at the same codon (346 of the cytosolic form, 335 of the mitochondrial form) in the midst of the common exons block of PEPCK. In the lowland species complex, the mitochondrion-specific 5'-exon 1 shows a major Lys ↔ Arg polymorphism as well. We consider possible evolutionary explanations for these findings and propose strategies to test alternative working hypotheses derived from them.

19

Resourceful Expressed Sequence Tags dataset from a mighty predatory mite

Ju-Lin Weng, David Margolies, Yoonseong Park

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

Expressed sequence tag data mining strategy has multi-fold application for an important biological control agent, *Phytoseiulus persimilis* (Acari, Mesostigmata, Phytoseiidae), a specialist predatory mite of spider mite species (Acari, Prostigmata, Tetranychidae). Genetic enhancement for pest control efficacy in the predatory mite may be possible. Custom development of specific cDNA libraries is a cost-effective approach for gene discovery and build-up of resourceful databases (i.e. for genetic marker development). Two cDNA libraries were combined and analyzed with ArthropodEST pipeline (Arthropod Genomics Center- KSU). A total of 27,237 sequences were clustered into 678 consigns and 3,467 remained as singlets (ranging from 100- 1,013 bases). Blast2Go analysis yielded 36.6% of annotated sequences from 10,251 EST submitted. A significant number of putative genes involved in reproduction, development, locomotion, and response to stress were found. The black legged tick, *Ixodes scapularis* (Acari, Ixodida, Ixodidae), was

identified as the taxon with largest number of top-hits (1,920 top hits), and the western predatory mite *Galandromus* (= *Metaseilus*) *occidentalis* (Acari, Mesostigmata, Phytoseiidae) had 132 top hits. Both arachnid species belong to the Superorder Parasitiformes, same as *P. persimilis*.

Molecular data are especially valuable when paleontological records are scarce. We took a multigene approach for temporal reconstruction of the Acari clade using our EST database. Sequences of high similarity to insects/arthropod genes were being targeted for phylogenetic analysis: 325 top-hit gene fragments (e-value < E⁻¹⁰⁰) were used to search for single orthologs in other arthropods and the nematode *Caenorhabditis elegans*. Predatory mite sequences were blast searched against *Drosophila melanogaster* protein database (Flybase) for conceptual translation. Amino acid sequences from the predatory mite were searched for orthologs among 11 arthropod species using OrthoDB (The Hierarchical Catalog of Eukaryotic Orthologs- Zdobnov's Computational Evolutionary Genomics group). Wormbase was used to blast for single and top hit (e-value < E⁻⁶⁰) orthologous protein sequences in *C. elegans*. The groups containing paralogy caused by the error in the search of OrthoDB were screened out. The tree for single gene analysis with more than 60% bootstrapping support for the grouping of *P. persimilis* and *C. elegans* was used as parameter for exclusion. True orthologs of 74 genes from the predatory mite were aligned individually and trimmed out the flanking positions in the other 12 species. Phylogenetic analyses were performed for 18,416 amino acids containing 10,106 variable sites (6,059 parsimony informative sites). The divergence time of the Acari clade was estimated based on the distance using four amino acid substitution models, assuming a global clock. Our result suggests that the divergence time of Acari is much earlier (average ranging from 482 - 656 Mya) compared to the recent findings (Devonian).

34

Applying Large-Scale Sequence Information from the Transcriptome to Reveal Life History Variation Maintained by the Dynamics of a Butterfly Metapopulation

Wheat, Christopher W.; *Fescemyer, Howard W.; Kvist, J.; Vera, J. Cristobal; Frilander, Mikko J.; Hanski, Ilkka; Marden, James H. (CWW, HWF, JCV, JHM) The Pennsylvania State University, Department of Biology, 208 Mueller Lab, University Park, PA 16802 USA; (CWW, IH) University of Helsinki, Department of Department of Ecology and Evolutionary biology, PL 65, Viikinkaari 1, 00014 Finland; (JK, MJF) University of Helsinki, Institute of Biotechnology, PL 56, Viikinkaari 9, 00014 Finland

Genomic approaches in wild populations have the potential to reveal the molecular bases of traits affecting fitness and may ultimately lead to a synthesis of population biology and genomics. There also is increasing evidence that functionally important developmental and phenotypic variation is governed primarily by regulation of gene expression. We examined gene expression in an ecologically well-studied metapopulation to obtain insights into how extinction-colonization dynamics cause diversifying selection and maintain genetic variation with large fitness effects.

The Glanville fritillary butterfly (*Melitaea cinxia* L.; Lepidoptera: Nymphalidae) has a large metapopulation in the Åland Islands in Finland that persists in a balance between stochastic local extinctions and establishment of new populations. Previously documented strong single-gene effects concerning phosphoglucose isomerase (*Pgi*) on butterfly performance and fitness components raise questions about the possibility of more extensive interaction between genetic variation and metapopulation processes.

In 2008, we reported at this meeting and in *Molecular Ecology* the de novo assembly of GS20 454 sequence data (transcriptome assembly 1) into much of the transcriptome from the Glanville fritillary butterfly. Subsequent FLX 454 sequencing at the University of Helsinki expanded this transcriptome (transcriptome assembly 2) and its annotation (<http://cinxiabase.vhost.psu.edu>). A custom oligonucleotide microarray covering over 60% of the transcriptome was constructed from transcriptome assembly 1. This array was used to test the hypothesis that establishment of new populations in currently unoccupied habitat patches selects systematically for different genetic variation affecting life history traits than is present in old populations.

Butterflies were sampled from newly established and old local populations (> 5 years continuously occupied) interspersed across the Åland metapopulation. Transcript level was quantified separately in abdomens (N=18) and thoraces (N=34) from common garden-reared 2-day old virgin females originating from 12 new and 12 old populations. The abdomen contains the ovaries and fat body, whereas the thorax contains mostly flight muscles. Hence, these body parts are informative about gene expression underlying fecundity and dispersal. Quantitative PCR of selected genes was used to confirm findings based on microarrays.

New-population females had higher expression of transcripts for genes involved in egg provisioning (e.g., larval serum proteins, lipid transporters, angiotensin converting enzyme, vitellogenin) and maintenance of flight muscle proteins (proteasome and chaperones). Physiological studies confirmed that new-population butterflies had accelerated egg maturation and enhanced flight metabolism. The microarray results and additional cloning and sequencing revealed variation in the 3'-UTR of succinate dehydrogenase subunit D (*Sdhd*). Expression of genes involved in oogenesis and central metabolism varied between allelic forms of the metabolic genes *Pgi* and *Sdhd*, which themselves were associated with differences in flight metabolic rate, population age and population growth rate. By applying transcriptome analysis to natural populations we discovered that genetic polymorphism in particular alleles, gene expression phenotypes and life history traits are associated with large fitness effects. We suggest that the rapid turnover of populations in metapopulations maintains genetic variation.

Changes in gene expression in European corn borer larvae (*Ostrinia nubilalis*) in response to *Bacillus thuringiensis* Cry1Ab protoxin exposure

Jianxiu Yao, Chitvan Khajuria, Nanyan Lu, Lawrent L Buschman, Brenda Oppert, Ming-Shun Chen, Kun Yan Zhu
Kansas State University, Department of Entomology, Manhattan, KS

The European corn borer (ECB), *Ostrinia nubilalis*, is the target pest for corn varieties expressing insecticidal toxins from *Bacillus thuringiensis* (Bt). We developed a cDNA microarray based on 15,000 cDNA elements representing 2,755 unique genes identified from ECB larval gut. This microarray was used to monitor gene expression in early fourth instar Bt-susceptible ECB larvae after 6 h feeding on diet with or without Cry1Ab protoxin. We identified 169 genes from the gut for which the expression was changed more than 2-fold for larvae exposed to Cry1Ab, in which 57 annotated genes were classified based on their molecular functions: 35 genes are involved in catalytic activity, 13 for binding activity, and the remaining genes for other functions. Thirty-six up-regulated transcripts may be involved in insect defense and signaling, and 21 down-regulated genes may have roles in Cry1Ab protoxin activation, and toxin binding processes. Ten of these genes are potentially involved in Bt toxicity and resistance, which include 7 serine protease, 1 alkaline phosphatase, 1 aminopeptidase, and 1 cadherin-like genes. The expression of serine protease genes was variable (one was up-regulated and six were down-regulated), but the three Bt-binding proteins, aminopeptidase, alkaline phosphatase and cadherin, were down regulated under Cry1Ab protoxin exposure. This study is the first large-scale survey of Cry1Ab protoxin induced transcriptional responses in ECB gut tissue and also provides a platform for functional studies of toxin-insect interactions.

Deletion of an N-acetyltransferase gene of silkworm confers the new black pattern in insects

Shuai Zhan^{1*}, Qihong Guo^{1*}, Minghui Li¹, Muwang Li², Jianyong Li³, Xuexia Miao¹, Yongping Huang^{1†}

¹Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, 200032, China; ²Sericultural Research Institute, Chinese Academy of Agriculture Sciences, Zhenjiang, 212018, China; ³Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

Body color variation is of great important for insects in adapting to environment change, such as industrial pollution or in mimicry for surviving. Many coloration mutants are maintained for the silkworm, *Bombyx mori*, a model Lepidoptera insect additionally with significant economic value. Within these, the phenotype of melanism (mln) is interesting for the overall black coloration in adults while just regional black in the head and tail spot of larvae. This is an ideal model for investigating the molecular mechanism of color pattern polymorphism. Here we firstly report the identification of an adult trait for silkworm with its progresses in genomics. Our positional cloning result indicates that the mutation of an arylalkylamine N-acetyltransferase (AANAT) gene confers the black coloration. About 100bp deletion was found in mln genome which causes two alternative splicing transcripts with premature terminations. Enzymatic assay revealed a significant loss of AANAT activity in the mutant proteins. Furthermore we performed RNA interference of AANAT in wild-type pupae and a certain proportion of adults with ectopic black coloration were observed subsequently. In the established insect pigmentation metabolism, most key nodes were observed resulting in the ectopic phenotypes just except AANAT, which was considered involving in an alternative colorless downstream pathway. Our mln mutant result is, to our knowledge, the first characterized AANAT mutant model that determines the insect color pattern.

Pyrosequencing-Based Expression Profiling and Identification of Differentially Regulated Genes in *Manduca sexta*

*Zhang, Shuguang; Zhang, Xiufeng; Gunaratna, Ramesh; Najjar, Fares; Wang, Yang; Roe, Bruce; Jiang, Haobo

(SZ, XZ, RG, YW, HJ) Oklahoma State University, Department of Entomology and Plant Pathology, Stillwater, OK 74078; (FN, BR) University of Oklahoma, Department of Chemistry and Biochemistry, University of Oklahoma, Norman, OK 73019

Although *Manduca sexta* has significantly contributed to our knowledge on various insect physiological processes, the lack of its genome sequence hampers large-scale gene discovery, transcript profiling, and proteomics analysis in this biochemical model species. Here we report a simple experimental approach based on massively parallel pyrosequencing, which allowed us to categorize transcripts based on their relative abundances and to discover process- or tissue-specifically regulated genes at the same time. We obtained 1,821,652 reads at an average size of 289 bp per read from fat body (F) and hemocytes (H) of naïve control (C) and bacteria-injected (I) *M. sexta* larvae. Most of the reads (92.1%) were assembled into 19,020 contigs and, for each contig, numbers of the annihilated reads from groups CF, IF, CH, and IH were normalized for ratio calculations. Using IF/CF and IH/CH, we identified 521 contigs whose read numbers increased at least five folds in either tissue after the immune challenge. A majority of polypeptides encoded by these contigs are similar in sequence to defense-related proteins, whereas others may participate in immune responses in unknown ways. Additionally, we found 277 contigs that were down-regulated to less than 1/10 of the control levels in either tissue. These include prophenoloxidasases, hemocyte protease-1, lacunin, hemocytin, and hemolectin. Using CF/CH and IF/IH, we discovered 429 and 241 contigs that were preferentially expressed in hemocytes (i.e. C or I ratio <0.1) and fat body (i.e. C or I ratio >10), respectively. The numbers of total and tissue-specific contigs suggest a higher functional diversity of hemocytes than fat body. We also integrated the new sequence data with those from the previous study and updated our EST database to support future genome assembly, gene annotation, and proteomics analysis in *M. sexta*. In summary, we have established a cost-effective approach for gene discovery and expression profiling in organisms with unknown genome sequences.

Genome-wide analysis of chitin synthase and chitinase gene families in African malaria mosquito, *Anopheles gambiae*

*Zhang, Xin; Zhang, Jianzhen; Zhu, Kun Y.

Kansas State University, Entomology Department, Manhattan, KS 66506

As arthropod growth and development are strictly dependent on precisely toned chitin synthesis and degradation, chitin metabolism represents an attractive target site for combating insect pests as this pathway is absent in humans and other vertebrates. We report the identification and characterization of two key gene families encoding chitin synthases and chitinases that are responsible for chitin biosynthesis and degradation, respectively, in African malaria mosquito, *Anopheles gambiae*. Totally, we identified and comprehensively characterized 2 chitin synthase and 20 chitinase genes which have been assigned into 2 (1 and 2, or A and B) and 8 (I-VIII) groups, respectively. Immunohistochemistry analysis and the stage- and tissue-specific gene expression analysis by various approaches including RT-PCR, quantitative real-time PCR, and in situ hybridization have significantly advanced our understanding of chitin synthesis and degradation in this insect. Furthermore, the functional analysis of each CHS gene has been studied by silencing each gene in mosquito larvae through the oral-delivered gene specific dsRNA/chitosan nanoparticles. Repression of the *AgCHS1* and the reduced chitin content resulting from the oral-delivered RNAi suggested the systemic nature of RNAi in mosquito larvae. In addition, we demonstrated that silencing of *AgCHS1* increased larval susceptibilities to diflubenzuron, whereas silencing of *AgCHS2* enhanced the peritrophic matrix disruption and thus increased larval susceptibilities to calcofluor white or dithiothreitol.

THANK YOU FOR PARTICIPATING IN THE
2010 ARTHROPOD GENOMICS SYMPOSIUM
POSTER SESSIONS.

An electronic version of the
poster abstracts is available at:

<http://www.k-state.edu/agc/symp2010/Abstracts-2010.pdf>